High-Dimensional Expanders imply Agreement Expanders
Second Session

September 25, 2019

Contents

1 Main Theorem: Formal Statement

We continue our exposition of the Dinur and Kaufman result establishing that HDXs are agreement expanders [2]. First, we introduce a new piece of notation.

Definition 1.1 (Restriction)**.** *Let X be a simplicial complex and $T \in X$. Then the restriction $X|_T$ of X to the vertices of T is defined as*

$$X|_T \;\coloneqq\; \{S \in X \mid S \subset T\}.$$

Now we have all the elements define our agreement tester more precisely. For the collection of subsets of $[n]$, we take a γ-HDX X (with dimension and expansion to be specified latter). For the tester distribution we take the following (we assume the dimension of X to be at least $2k$).

Definition 1.2 ($\mathcal{D}_{k,2k}$)**.** *Tester Distribution $\mathcal{D}_{k,2k}$*

- *Choose $r \sim \Pi_{2k}$,*
- *Choose independently $S, S' \in X|_r(k)$ uniformly in r, and*

- *Output* (S, S').

Recall that we defined the rejection probability of the tester as $\mathrm{disagree}_{(X,D)}(F)$. Sometimes it is also convenient to work with its dual notion.

Definition 1.3 (Agreement Probability)**.** *We define the agreement of a tester* (X, D) *as*

$$\mathrm{agree}_{(X,D)}(F) \;\coloneqq\; \Pr_{(S,S')\sim D}\left[f_S \equiv f_{S'}\right],$$

i.e., $\mathrm{agree}_{(X,D)}(F) = 1 - \mathrm{disagree}_{(X,D)}(F)$.

Remark 1.1. *To make the notation lighter we drop* X *from* $\mathrm{agree}_{(X,D)}$ *and* $\mathrm{disagree}_{(X,D)}$ *since* X *is usually clear from the context.*

The formal version of the main theorem of Dinur and Kaufman [2] is given below.

Theorem 1.2 (Dinur and Kaufman Main [2])**.** *Let* X *be a* d*-dimensional* γ*-HDX with* $d > (k+1)^2$, $\gamma < 1/d$ *and* k *a sufficiently large* [1]*. Let* $F = \{f_S\}_{S\in X(k)}$ *be an ensemble. If* $\mathrm{agree}_{\mathcal{D}_{k,2k}}(F) = 1 - \epsilon$, *then there is* $g\colon X(0) \to \{0,1\}$ *such that*

$$\Pr_{S\sim\Pi_k}\left[f_S \equiv g|_S\right] \;=\; 1 - O(\epsilon).$$

Furthermore, g *is defined according to the majority (weighted by* Π_k*), namely,*

$$g(x) \;\coloneqq\; \mathrm{majority}_{S\ni x} f_S(x),$$

for every $x \in X(0)$.

The proof of Theorem 1.2 crucially relies on the sampling properties of a HDX and the various local complete complexes contained in it. Note that when $a \in X(d)$ is fixed the restriction of X to a is a complete complex for which agreement expansion is known to hold 1.3.

Theorem 1.3 (Dinur and Steurer [3] (Abridged))**.** *Let* $X = \Delta_k(d)$*, i.e.,* X *is the complete* k*-dimensional complex with* $d > k^2$*. Let* $F = \{f_S\}_{S\in X(k)}$ *be an ensemble. If* $\mathrm{agree}_{\mathcal{D}_{k,2k}}(F) = 1 - \epsilon$, *then there is* $g\colon X(0) \to \{0,1\}$ *such that*

$$\Pr_{S\sim\Pi_k}\left[f_S \equiv g|_S\right] \;=\; 1 - O(\epsilon).$$

Furthermore, g *is defined according to the majority (weighted by* Π_k*), namely,*

$$g(x) \;\coloneqq\; \mathrm{majority}_{S\ni x} f_S(x),$$

for every $x \in X(0)$.

Remark 1.4. *Note that* $X(i)$ *and* Π_i *are not (necessarily) the same in the preceding two theorems.*

Remark 1.5. *Contrary to the graph case where results for the complete graph tend to be straightforward, for higher dimensions this is not necessarily the case. The result of Dinur and Steurer 1.3 is one such example where considerable work is required to prove agreement expansion of complete complexes.*

[1] Taking $k > 200$ suffice.

Remark 1.6. *Since there exist γ-HDXs with linear (in n) number of hyperedges (assuming $d = O(1)$ and $\gamma = \Theta(1)$), Dinur and Kaufman result 1.2 is a derandomization/sparsification of the Dinur and Steurer 1.3.*

Remark 1.7. *In [1], Dinur et al. studied "higher order" agreement testing.*

We proceed to prove Theorem 1.2 assuming Theorem 1.3. At a high-level the strategy can be described as follows. For every $a \in X(d)$, we use Theorem 1.3 to construct local majorities $g_a \colon a \to \{0,1\}$ on $X|_a$. Since the total error in our tester is "small", namely ϵ, we will be able to deduce that most of these local majorities agree with most f_S for $S \in X|_a(k)$. The main difficulty is to prove that these local majorities $\{g_a\}_{a \in X(d)}$ can be "glued" together to form the global majority $g \colon X(0) \to \{0,1\}$. This part of the proof is somewhat involved (and technical) and heavily uses several expander graphs naturally arising as substructures of a HDX.

2 Definitions

The analysis involves a multitude of probability measures. For this reason, we adopt somewhat strict conventions. In a probability statement of the form $\Pr[\cdot]$ where x is drawn form a set E, we convention that $\Pr_{x \in E}[\cdot]$ means that x is selected uniformly in E. In addition, the statement $\Pr_{x,y \in E}[\cdot]$ means that x and y are sampled independently (and also uniformly by the previous convention). When the underlying distribution is not necessarily uniform we use $\Pr_{x \sim E}[\cdot]$. We extend these conventions to expectations. Oftentimes for clarity we indicate the nature of the drawing explicitly.

It will be important to consider top faces $a \sim \Pi_d$ giving rise to $r \sim \Pi_{2k}$ in the tester since $X|_a$ is a complete complex (of dimension d) with well known agreement expansion 1.3.

Definition 2.1 ($\mathcal{D}_{k,2k}$)**.** *Tester Distribution $\mathcal{D}_{k,2k}$*

- *Choose $a \sim \Pi_d$,*
- *Choose r uniformly in $X|_a(2k)$,*
- *Choose independently $S, S' \in X|_r(k)$ uniformly in r, and*
- *Output (S, S').*

Remark 2.1. *The distribution $\mathcal{D}_{k,2k}$ is still the same.*

For $a \in X(d)$ we will consider its "local" majority function g_a on the complete k-dimensional complexes $X|_a$.

Definition 2.2 (Local Majority g_a)**.** *Given $a \in X(d)$, we define the local majority g_a as*

$$g_a(x) \;\coloneqq\; \text{majority}_{S\colon a \supset S \ni x} f_S(x).$$

In this case, the majority is weighted according to the uniform distribution since $X|_a$ is a complete complex (all measures are uniform).

2.1 Bad Sets and Bad Local Complexes

The crux of the analysis lies in bounding the measure of a variety of "bad" sets. We collect all of them in this subsection. They are mostly quite natural, but their precise role will be clear in the upcoming proofs.

Definition 2.3 (Global Bad Collection B_x for x)**.** *Given $x \in X(0)$, we define the bad sets for x in the ensemble F as*

$$B_x \;\coloneqq\; \{S \in X_x(k) \mid f_{S \sqcup \{x\}}(x) \neq g(x)\}.$$

Remark 2.2.

$$\Pr_{S \sim X_x(k)} [S \in B_x] \;\leq\; \frac{1}{2}.$$

Definition 2.4 (Local Bad Collection $B_{x,a}$ for x and a)**.** *Given $x \in X(0)$ and $a \in X(d)$, we define the bad sets for x and a in the ensemble F as*

$$B_{x,a} \;\coloneqq\; \{S \in X_x|_{a'}(k) \mid f_{S \sqcup \{x\}}(x) \neq g_a(x)\},$$

where $a' = a \setminus \{x\}$.

Remark 2.3.

$$\Pr_{S\colon a \supset S \ni x} [S \in B_{x,a}] \;\leq\; \frac{1}{2}.$$

A key notion for a vertex is *confused*.

Definition 2.5 (Confused Vertex)**.** *We say that $x \in X(0)$ is confused provided*

$$\Pr_{S \sim X_x(k-1)} [S \in B_x] \;\geq\; \frac{3}{10}.$$

The notion of confused can be applied locally to $X|_a$.

Definition 2.6 (Confused Vertex in a)**.** *Given $a \in X(d)$, we say that $x \in a$ is confused in a provided*

$$\Pr_{S\colon a \supset S \ni x} [S \in B_{x,a}] \;\geq\; \frac{2}{10}.$$

From the perspective of a vertex $x \in X(0)$, some local complete complexes $X|_a$ have a majority g_a that disagrees with global majority g at x, i.e., $g_a(x) \neq g(x)$. These "terrible" local complexes pose difficulties in "gluing" the local majorities into the global majority.

Definition 2.7 (Terrible Complete Complexes)**.** *Given $x \in X(0)$, we define*

$$T_x \coloneqq \{a \ni x \mid a \in X(d) \;\wedge\; g_a(x) \neq g(x)\}.$$

Claim 2.4. *If $a \in T_x$, then*

$$\Pr_{S\colon a \supset S \ni x} [S \in B_x] \;\geq\; \frac{1}{2}.$$

To make the notation more compact we use $\mu(B_x)$ and $\mu(T_x)$ to indicate the measure the distributions on the link X_x assign to B_x and T_x. More precisely, we have

$$\mu(B_x) \coloneqq \Pr_{S\sim X_x(k-1)} [S \in B_x],$$

and

$$\mu(T_x) \coloneqq \Pr_{a\sim X_x(d-1)} [a \sqcup \{x\} \in T_x],$$

Since we will consider what happens in $X|_a$, it will be important to quantify how the agreement test performs on it.

Definition 2.8 (Local Disagreement)**.** *Given $a \in X(d)$, we define the local disagreement of the tester to a as*

$$\epsilon_a \ \coloneqq\ 1 - \mathrm{agree}_{\mathcal{D}_{k,2k}|_a}(F).$$

Remark 2.5.

$$\mathbb{E}_a \epsilon_a \ =\ \epsilon.$$

3 Proof of Main Theorem

The proof of the Main Theorem follows a hierarchical structure. It uses two "major" lemmas (Lemma 3.1 and Lemma 4.6) establishing non-trivial "dumping", as k increases, in the measure of "bad" sets. In turn, proving these lemmas will take us to consider three auxiliary claims.

Lemma 3.1 (First Major Lemma)**.**

$$\mathbb{E}_{x\sim\Pi_0} \mu(T_x) \cdot \mathbb{1}_{[x \text{ is not confused}]} \ =\ O\left(\frac{\epsilon}{k+1}\right).$$

Lemma 3.2 (Second Major Lemma)**.**

$$\Pr_{x\sim\Pi_0} [x \text{ is confused}] \ =\ O\left(\frac{\epsilon}{k+1}\right).$$

Assuming the two major lemmas we are ready to prove the Main Theorem. In general, we follow the original proof of Dinur and Kaufman, but we try to provide further details in some steps.

Proof of Theorem 1.2. Suppose $F = \{f_S\}_{S\in X(k)}$ is such that $\mathrm{agree}_{\mathcal{D}_{k,2k}}(F) = 1 - \epsilon$. We need to prove that the majority $g\colon X(0) \to \{0,1\}$ satisfy

$$\Pr_{S\sim\Pi_k} [f_S \equiv g|_S] \ =\ 1 - O(\epsilon).$$

We have

$$\begin{aligned}
\Pr_{S\sim\Pi_k} [f_S \equiv g|_S] \ &\geq\ \Pr_{a\sim\Pi_d, S\in X|_a(k)} [f_S \equiv g_a|_S \equiv g|_S] \\
&\geq\ 1 - \Pr_{a\sim\Pi_d, S\in X|_a(k)} \Bigg[\underbrace{f_S \not\equiv g_a|_S}_{E_1}\Bigg] - \Pr_{a\sim\Pi_d, S\in X|_a(k)} \Bigg[\underbrace{g_a|_S \not\equiv g|_S}_{E_2}\Bigg]. \quad \text{(Union Bound)}
\end{aligned}$$

Hence it is enough to bound $\Pr_{a\supset S}[E_1]$ and $\Pr_{a\supset S}[E_2]$ where $\Pr_{a\sim\Pi_d, S\subset X(k)|_a}[\cdot]$ was abbreviated to $\Pr_{a\supset S}[\cdot]$. We start with the former which is easier once we assume agreement expansion of complete complexes 1.3. For this we consider the conditional probability

$$\Pr_{a\supset S}[E_1|a] \;=\; \Pr_{S\in X|_a(k)}[E_1].$$

Since the disagreement on a is ϵ_a (by definition), invoking Theorem 1.3 we obtain

$$\Pr_{S\in X|_a(k)}[E_1] \;=\; O(\epsilon_a).$$

Hence

$$\Pr_{a\supset S}[E_1] \;=\; \mathbb{E}_{a\sim\Pi_d}\Pr_{a\supset S}[E_1|a] \;=\; \mathbb{E}_{a\sim\Pi_d}O(\epsilon_a) \;=\; O(\epsilon).$$

Now we bound $\Pr_{a\supset S}[E_2]$ which is considerably trickier but follows somewhat easily assuming the two major lemmas above 3.1 and 4.6. We have

$$\begin{aligned}
\Pr_{a\supset S}[E_2] &= \mathbb{E}_{a\sim\Pi_d}\mathbb{E}_{S\in X|_a(k)}\mathbb{1}_{[\exists x\in S \;|\; g_a(x)\neq g(x)]} \\
&\leq \mathbb{E}_{a\sim\Pi_d}\mathbb{E}_{S\in X|_a(k)}\sum_{x\in S}\mathbb{1}_{[g_a(x)\neq g(x)]} && \text{(Union Bound)} \\
&= \mathbb{E}_{a\sim\Pi_d}\mathbb{E}_{S\in X|_a(k)}\sum_{x\in S}\mathbb{1}_{[a\in T_x]} \\
&= (k+1)\cdot\mathbb{E}_{x\sim\Pi_0}\mu(T_x).
\end{aligned}$$

We split the expectation $\mathbb{E}_{x\sim\Pi_0}\mu(T_x)$ over confused and non confused $x\in X(0)$, that is,

$$\begin{aligned}
\mathbb{E}_{x\sim\Pi_0}\mu(T_x) &= \mathbb{E}_{x\sim\Pi_0}\mu(T_x)\cdot\mathbb{1}_{[x \text{ is confused}]} + \mathbb{E}_{x\sim\Pi_0}\mu(T_x)\cdot\mathbb{1}_{[x \text{ is not confused}]} \\
&\leq \Pr_{x\sim\Pi_0}[x \text{ is confused}] + \mathbb{E}_{x\sim\Pi_0}\mu(T_x)\cdot\mathbb{1}_{[x \text{ is not confused}]} \\
&\leq O\left(\frac{\epsilon}{k+1}\right) + \mathbb{E}_{x\sim\Pi_0}\mu(T_x)\cdot\mathbb{1}_{[x \text{ is not confused}]} && \text{(By Lemma 4.6)} \\
&\leq O\left(\frac{\epsilon}{k+1}\right) + O\left(\frac{\epsilon}{k+1}\right). && \text{(By Lemma 3.1)}
\end{aligned}$$

Therefore, $\Pr_{a\supset S}[E_2] = O(\epsilon)$ concluding the proof. □

4 Proof of Two Major Lemmas

From the "self-similarity" definition of γ-HDX, the following fact is immediate.

Fact 4.1. *The link X_x is also a γ-HDX.*

4.1 Proof of First Major Lemma

To prove the first major lemma we will need the following auxiliary claim.

Claim 4.2 (First Auxiliary Claim)**.**

$$\mathbb{E}_{x\sim\Pi_0}\mu(B_x) \;=\; O(\epsilon).$$

Lemma 4.3 (First Major Lemma (restatement))**.**

$$\mathbb{E}_{x\sim\Pi_0}\mu(T_x)\cdot \mathbb{1}_{[x \text{ is not confused}]} \;=\; O\left(\frac{\epsilon}{k+1}\right).$$

Proof. If x is not confused, then by definition $\Pr_{S\sim X_x(k-1)}[S \in B_x] < 0.3$. Since X_x is a γ-HDX and $d > (k+1)^2$, the bipartite graph between $X_x(k-1)$ and $X_x(d-1)$ is a sampler with largest non-trivial singular value λ satisfying $\lambda^2 \le 1/(k+1)$. By the baby sampling fact, we get

$$\begin{aligned}
100\cdot\lambda^2\cdot \Pr_{S\sim X_x(k-1)}[S\in B_x] \;&\ge\; \Pr_{a\sim X_x(d-1)}\left[\Pr_{S\sim X_x|_a(k-1)}[S\in B_x] > \Pr_{S\sim X_x(k-1)}[S\in B_x] + 0.1\right]\\
&\ge\; \Pr_{a\sim X_x(d-1)}\left[\Pr_{S\sim X_x|_a(k-1)}[S\in B_x] > 0.4\right]\\
&\ge\; \Pr_{a\sim X_x(d-1)}[a\sqcup\{x\}\in T_x] =: \mu(T_x),
\end{aligned}$$

which simplifies to

$$\mu(T_x) \;\le\; \frac{100}{k+1}\cdot\mu(B_x).$$

Applying Claim 4.2 yields

$$\mathbb{E}_{x\sim\Pi_0}\mu(T_x)\cdot\mathbb{1}_{[x \text{ is not confused}]} \;\le\; \frac{100}{k+1}\cdot\mathbb{E}_{x\sim\Pi_0}\mu(B_x) \;=\; O\left(\frac{\epsilon}{k+1}\right).$$

□

4.2 Proof of Second Major Lemma

The proof of the second major lemma relies on the following two auxiliary claims. In these claims, $\epsilon_0 > 0$ is a sufficiently small constant to be defined later.

Claim 4.4 (Second Auxiliary Claim)**.** *Assuming $\epsilon < \epsilon_0/2$, then at most $O(\lambda^2\cdot\epsilon)$ of the a's in $X(d)$ have $\epsilon_a > \epsilon_0$.*

Claim 4.5 (Third Auxiliary Claim)**.** *For every $a\in X(d)$, if $\epsilon_a \le \epsilon$ then*

$$\Pr_{x\sim a}[x \text{ confused in } a] \;=\; O\left(\frac{\epsilon_a}{k+1}\right).$$

Lemma 4.6 (Second Major Lemma (restatement))**.**

$$\Pr_{x\sim\Pi_0}[x \text{ is confused}] \;=\; O\left(\frac{\epsilon}{k+1}\right).$$

Proof. Suppose $x \in X(0)$ is confused, i.e., $\Pr_{S\sim X_x(k)}[S \in B] \geq 0.3$. Recall that the bipartite graph between $X_x(k-1)$ and $X_x(d-1)$ has largest non-trivial singular value λ satisfying $\lambda^2 < 1/(k+1) < 1/200$. By the baby sampling fact, we get

$$\begin{aligned}
100 \cdot \lambda^2 \cdot \Pr_{S\sim X_x(k-1)}[S \in B_x] &\geq \Pr_{a\sim X_x(d-1)}\left[\Pr_{S\sim X_x|_a(k-1)}[S \in B_x] < \Pr_{S\sim X_x(k-1)}[S \in B_x] - 0.1\right] \\
&\geq \Pr_{a\sim X_x(d-1)}\left[\Pr_{S\sim X_x|_a(k-1)}[S \in B_x] < 0.2\right] && (x \text{ is confused}) \\
&= \Pr_{a\sim X_x(d-1)}[x \text{ is not confused in } a],
\end{aligned}$$

or equivalently

$$\Pr_{a\sim X_x(d-1)}[x \text{ is confused in } a] \geq \frac{1}{2}.$$

Then on one hand we have

$$\begin{aligned}
\Pr_{a\sim \Pi_d, x\in a}[x \text{ is confused in } a] &\geq \Pr_{a\sim X_x(d-1):\, x \text{ confused}}[x \text{ is not confused in } a] \Pr_{x\sim\Pi_0}[x \text{ is confused}] \\
&\geq \frac{1}{2} \cdot \Pr_{x\sim\Pi_0}[x \text{ is confused}].
\end{aligned}$$

On the other hand

$$\begin{aligned}
\Pr_{a\sim \Pi_d, x\in a}[x \text{ is confused in } a] &= \mathbb{E}_{a\sim\Pi_d} \Pr_{x\in a}[x \text{ is confused in } a|a] \\
&= \mathbb{E}_{a\sim\Pi_d} \Pr_{x\in a}[x \text{ is confused in } a|a] \cdot \mathbb{1}_{[\epsilon_a > \epsilon_0]} \\
&\quad + \mathbb{E}_{a\sim\Pi_d} \Pr_{x\in a}[x \text{ is confused in } a|a] \cdot \mathbb{1}_{[\epsilon_a \leq \epsilon_0]} \\
&= O(\lambda^2 \cdot \epsilon) + \mathbb{E}_{a\sim\Pi_d} \Pr_{x\in a}[x \text{ is confused in } a|a] \cdot \mathbb{1}_{[\epsilon_a \leq \epsilon_0]} && \text{(By Claim 4.4)} \\
&\leq O(\lambda^2 \cdot \epsilon) + \mathbb{E}_{a\sim\Pi_d} O\left(\frac{\epsilon_a}{k+1}\right) && \text{(By Claim 4.5)} \\
&\leq O\left(\frac{\epsilon}{k+1}\right).
\end{aligned}$$

Combining both bounds for $\Pr_{a\sim\Pi_d, x\in a}[x \text{ is confused in } a]$ concludes the proof. ☐

5 Proof of Auxiliary Claims

5.1 First Auxiliary Claim

The proof of the first auxiliary claim takes place in yet another graph inside our HDX, namely, the two-step random walk from $X_x(k-1)$ to $X_x(2k-1)$. Its proof will also clarify the choice of the tester distribution since it is chosen so that this graph is an expander.

Definition 5.1 (Two-Step Random Walk). *Given $x \in X(0)$, we define the two step random walk graph $G_x = (V, E)$ where $V = X_x(k-1)$ and $E = \{\{S, S'\} \mid S, S' \in X_x(2k-1)\}$. Furthermore, to each edge $\{S, S'\}$ we assign weight*

$$\mathbb{E}_{r\sim X_x(2k-1)} \Pr_{T\in X_x|_r(k-1)}[T = S] \Pr_{T\in X_x|_r(k-1)}[T = S'].$$

Let $U_{i,x}$ and $D_{i,x}$ be the up and down operators defined in the link X_x. Set

$$M_x \;:=\; D_{k,x} \cdots D_{2k-1,x} U_{2k-2,x} \cdots U_{k-1,x}.$$

Claim 5.1. *The two step walk $G_x = (V, E)$ is the random walk given by the operator M_x. Moreover, the second non-trivial singular value λ of M_x satisfy*

$$\lambda \;\leq\; \frac{1}{2}.$$

Proof. The first part is a simple verification. The second follows from the Spectral Boosting Lemma from the first session. □

Claim 5.2 (First Auxiliary Claim (restatement))**.**

$$\mathbb{E}_{x \sim \Pi_0} \mu(B_x) \;=\; O(\epsilon).$$

Proof. Fix $x \in X(0)$. Let ϵ_x be

$$\epsilon_x \;:=\; \Pr_{\{S,S'\} \sim E(G_x)} \left[f_{S \sqcup \{x\}}(x) \neq f_{S' \sqcup \{x\}}(x) \right].$$

Since we are working with Boolean functions, if $f_{S \sqcup \{x\}}(x) \neq f_{S' \sqcup \{x\}}(x)$ then either

$$f_{S \sqcup \{x\}}(x) \;=\; g(x),$$

or

$$f_{S' \sqcup \{x\}}(x) \;=\; g(x).$$

Let $G_x = (V, E)$ be the two step graph. This allows us to express ϵ_x equivalently as

$$\epsilon_x \;=\; \mu(E(B_x, V \setminus B_x)).$$

Applying Cheeger's inequality gives

$$\frac{\lambda_{\text{Laplacian}}}{2} \;\leq\; \frac{\mu(E(B_x, V \setminus B_x))}{\min(\mu(B_x), \mu(V \setminus B_x))} = \frac{\epsilon_x}{\mu(B_x)},$$

where we used the fact $\mu(B_x) \leq 1/2$. Simplifying

$$\frac{\lambda_{\text{Laplacian}}}{2} \cdot \mu(B_x) \;\leq\; \epsilon_x.$$

Using Claim 5.1, $\lambda_{\text{Laplacian}} \geq 1/2$. Thus,

$$\mathbb{E}_{x \sim \Pi_0} \mu(B_x) \;\leq\; \mathbb{E}_{x \sim \Pi_0} O(\epsilon_x) \;=\; O(\epsilon),$$

where the last equality follows from

$$\begin{aligned}
\mathbb{E}_{x \sim \Pi_0} O(\epsilon_x) \;&=\; \mathbb{E}_{x \sim \Pi_0} \mathbb{E}_{\{S,S'\} \sim E(G_x)} \mathbb{1}_{\left[f_{S \sqcup \{x\}}(x) \neq f_{S' \sqcup \{x\}}(x) \right]} \\
&\leq\; \mathbb{E}_{x \sim \Pi_0} \mathbb{E}_{\{S,S'\} \sim E(G_x)} \mathbb{1}_{\left[f_{S \sqcup \{x\}} \not\equiv f_{S' \sqcup \{x\}} \right]} \\
&=\; \mathbb{E}_{r \sim \Pi_{2k}} \mathbb{E}_{T,T' \in X|_r(k)} \mathbb{1}_{[f_T \not\equiv f_{T'}]} \\
&=\; \text{disagree}_{\mathcal{D}_{k,2k}}(F) \;=\; O(\epsilon).
\end{aligned}$$

□

5.2 Second and Third Auxiliary Claims

Recall that $\epsilon_0 > 0$ is a sufficiently small constant. More precisely it can be extracted as the minimum required valued in both auxiliary claims below.

Claim 5.3 (Second Auxiliary Claim (restatement)). *Assuming $\epsilon < \epsilon_0/2$, then at most $O(\lambda^2 \cdot \epsilon)$ of the a's in $X(d)$ have $\epsilon_a > \epsilon_0$.*

Proof. Let $h\colon X(2k) \to [0,1]$ be defined as

$$h(r) \;\coloneqq\; \mathrm{agree}_{\mathcal{D}_{k,2k}|_r}(F),$$

for every $r \in X(2k)$. This implies that

$$\mathbb{E}_{r\sim X(2k)} h(r) \;=\; \epsilon.$$

Let $M \coloneqq U_{d-1} \cdots U_{2k}$. By definition

$$(Mh)(a) \;=\; \mathbb{E}_{r\in X|_a(2k)} h(r) \;=\; \epsilon_a,$$

for $a \in X(d)$. Set $E \coloneqq \{a \in X(d) \mid \epsilon_a > \epsilon_0\}$. Then by expander mixing lemma

$$\langle Mh, \mathbb{1}_E\rangle \;\le\; \epsilon \cdot \Pr_{a\sim X(d)}[E] + \lambda\sqrt{\epsilon \cdot \Pr_{a\sim X(d)}[E]}.$$

Since $\epsilon_a > \epsilon_0$ by assumption, we get

$$\langle Mh, \mathbb{1}_E\rangle \;\ge\; \Pr_{a\sim X(d)}[E] \cdot \epsilon_0$$

Assuming $\epsilon < \epsilon_0/2$ and combining both bounds on $\langle Mh, \mathbb{1}_E\rangle$ gives

$$\Pr_{a\sim X(d)}[E] \;\le\; \left(\frac{2}{\epsilon_0}\right)^2 \cdot \lambda^2 \cdot \epsilon \;\le\; O\left(\lambda^2 \cdot \epsilon\right).$$

$\square$

Claim 5.4 (Third Auxiliary Claim (restatement)). *For every $a \in X(d)$, if $\epsilon_a \le \epsilon$ then*

$$\Pr_{x\sim a}[x \text{ confused in } a] \;=\; O\left(\frac{\epsilon_a}{k+1}\right).$$

Proof. Fix $a \in X(d)$ with $\epsilon_a \le \epsilon_0$. Note that $X|_a$ contains a complete k-dimensional complex. Let

$$C_a \;\coloneqq\; \{x \in a \mid x \text{ is confused in } a\}.$$

Applying the agreement expansion theorem for complete complexes 1.3, we get that the local majority g_a satisfy

$$\Pr_{S\in X|_a(k)}\left[\underbrace{g_a|_S \not\equiv f_S}_{E}\right] \;=\; O(\epsilon_a) \;\le\; K \cdot \epsilon_a,$$

for some constant $K > 0$ (that was hidden in the big O).

Let M be a Markov operator corresponding to the bipartite graph between $X|_a(0)$ $(= a)$ and $X|_a(k)$. We know that the second non-trivial singular value λ of M satisfy $\lambda^2 \leq 1/(k+1)$. The expander mixing lemma gives

$$\langle M\mathbb{1}_{C_a}, \mathbb{1}_E\rangle \;\leq\; K \cdot \Pr_{x\in a}[x \in C_a] \cdot \epsilon_a + \lambda\sqrt{K \cdot \Pr_{x\in a}[x \in C_a] \cdot \epsilon_a}.$$

On the other hand since each $x \in C_a$ has least 0.2 fraction of neighbors in E (x is confused), we get

$$\langle M\mathbb{1}_{C_a}, \mathbb{1}_E\rangle \;\geq\; 0.2 \cdot \Pr_{x\in a}[x \in C_a].$$

Assuming $\epsilon_0 \leq 0.1/K$, we conclude

$$\Pr_{x\in a}[x \in C_a] \;\leq\; \frac{\lambda^2}{0.01} \cdot \epsilon_a,$$

and we are done. $\square$

References

[1] Irit Dinur, Yuval Filmus, and Prahladh Harsha. Analyzing boolean functions on the biased hypercube via higher-dimensional agreement tests: [extended abstract]. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2019, San Diego, California, USA, January 6-9, 2019*, pages 2124–2133, 2019.

[2] Irit Dinur and Tali Kaufman. High dimensional expanders imply agreement expanders. In *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017, Berkeley, CA, USA, October 15-17, 2017*, pages 974–985, 2017.

[3] Irit Dinur and David Steurer. Direct product testing. In *Proceedings of the 2014 IEEE 29th Conference on Computational Complexity*, CCC '14, pages 188–196, 2014.