Lecture 5: Cosystolic expansion and property testing

Irit Dinur

December 12, 2022

In this lecture we will continue to study coboundary and cosystolic expansion and its relation to property testing. We prove coboundary expansion of the complete complex, and then we will prove the linearity testing theorem of Blum Luby and Rubinfeld, and describe it as coboundary expansion of a certain chain complex.

1 Cosystolic expansion

Recall that given a d-dimensional simplicial complex, we define the chain of coboundary maps to be

$$C_{-1} \overset{\delta_{-1}}{\to} C_0 \overset{\delta_0}{\to} C_1 \overset{\delta_1}{\to} \cdots \overset{\delta_{i-1}}{\to} C_i \overset{\delta_i}{\to} C_{i+1} \to \cdots$$

where $C_i = C_i(X, \mathbb{Z}_2) = \mathbb{Z}_2^{X(i)}$ is the space of all i-chains, and where $\delta_i : C_i \to C_{i+1}$ is given by

$$\forall f \in C_i, \forall s \in X(i+1), \quad \delta_i f(s) = \sum_{t<s} f(t) \mod 2$$

We set the coboundaries to be $B^i = Im(\delta_{i-1})$ and the cocycles to be $Z^i = Ker(\delta_i)$ and noted that $B^i \subseteq Z^i \subseteq C^i$ and furthermore we let the i-th cohomology be $H^i = Z^i/B^i$. When $B^i = Z^i$ this is trivial. We defined coboundary expansion to be

$$h^i(X, \mathbb{Z}_2) = \min_{f \in C_i \setminus B^i} \frac{wt(\delta_i f)}{\mathrm{dist}(f, B^i)}.$$

In this definition we are comparing two measures for the distance of a given chain f to B^i. The numerator is a distance that is easy to calculate and even to estimate: simply check for each $i+1$-face s, whether $\delta_i f(s) = 0$. The denominator involves the distance of f from a given set, and may require exponential time to compute (indeed, in some cases this is NP-hard). Having a constant coboundary expansion means that these two measures are comparable, and that we can estimate the denominator by the numerator. Indeed, a convenient equivalent way to think about coboundary expansion is this

$$\forall f \in C^i, \qquad \mathrm{dist}(f, B^i) \leqslant wt(\delta_i f) \cdot \frac{1}{h^i}.$$

How large or small can h^i be? It can be zero when $Z^i B^i$: take any $f \in Z^i \setminus B^i$, it has $wt(\delta_i f) = 0$. Even if $h^i \neq 0$ it can be as small as $O(1)/|X(i)|$. For example, check this on $i = 0$ and, say, the cycle graph.

The same definition that pertains to coboundaries, can be made regarding cocycles.

Definition 1.1. The cosystolic expansion of a d-dimensional simplicial complex X, at level $i < d$, is defined to be

$$h^i(X, \mathbb{Z}_2) = \min_{f \in C_i \setminus Z^i} \frac{wt(\delta_i f)}{\mathrm{dist}(f, Z^i)}.$$

We are using the same notation h^i for coboundary and cosystolic expansion simply because in case $B^i = Z^i$ these are the same, and in case $B^i \subsetneq Z^i$, then the coboundary expansion is zero and the quantity of interest will be cosystolic expansion.

We say that a family $(X_n)_{n=1}^{\infty}$ of simplicial complexes is a family of *ε-cosystolic expanders* if for all n, $h^i(X_n) > \varepsilon$. Are there any coboundary expanders (sparse or not)?

We will see that there are families of *bounded-degree* cosystolic expanders.

2 Coboundary expansion of the complete complex

A first question to ask is whether the complete complex is a coboundary expander.

The definitions of link expansion and random walk expansion were developed by comparison to the complete complex, and the fact that the complete complex is an expander according to these definitions is immediately obvious.

In contrast, the definition of coboundary/cosystolic expansion is derived by phrasing the 0-dimensional case in cohomological terms, and then generalizing syntactically. It turns out that even the expansion of the most naturally expanding complex is not obviously clear.

Theorem 2.1 ([3, 4, 5])**.** *Let $X = \Delta^{n-1}$ be the complete complex on n vertices. Then $h^k(X) \geqslant \frac{n}{n-k-1}$.*

Before proving the theorem, let us consider the case $k = 1$. We need to show that for any $f \in C_1$,

$$\mathrm{dist}(f, B^0) \leqslant wt(\delta_1 f) \cdot \frac{n-2}{n}.$$

In other words, assuming that δf is small, we must find some chain $g \in C_0$ such that $f \approx \delta_0 g$.

Connection to graph property testing. The 0-chains $g \in C_0$ are arbitrary $0-1$ functions on the vertices, namely every g corresponds to a subset S of the vertices. The coboundaries $\delta g \in B^0$ are indicators of the edges crossing the cut between S and $\bar{S}$. So, for $k = 1$, the question of coboundary expansion boils down to a property testing question called *biclique testing*. This is the question of testing if a given graph is a biclique (i.e. the vertices can be partitioned so that the graph is a complete bipartite graph between the two parts). Given an arbitrary graph $f \in C_1$, test whether it is close to a biclique, i.e. to a graph that indicates a cut between S and $\bar{S}$. Indeed, in the literature of property testing [2] this is a well known result, stating that the triangle test is a good so-called proximity-oblivious tester (POT) for the property of being a biclique. This result appears in [2, Proposition 8.6].

To warm up to the proof, assume first that $\varepsilon = 0$. In this case we can "reconstruct" the cut $S, \bar{S}$ from f by fixing some v, putting v in $\bar{S}$, and then deciding for every other vertex u whether it is in S according to $f(uv)$. Namely, we set $S = \{u \in V \mid \ f(uv) = 1\}$. We can now check that for every edge uw,

$$f(uw) = f(uv) + f(vw) = \mathbf{1}_S(u) + \mathbf{1}_S(w) = \delta\mathbf{1}_S(uw).$$

To prove the theorem we do the same thing, but robustly.

Proof. We first analyze the case $k = 1$. Fix $v \in X(0)$. Let $g \in C_0$ be defined by $g(v) = 0$ and $g(u) = f(uv)$ for all $u \neq v$. By definition, for every edge $e \ni v$, $\delta g(e) = f(e)$. For edges uw where $v \neq u, w$,

$$\delta g(uw) = g(u) + g(w) = f(uv) + f(vw) = f(uw) + \delta f(uvw).$$

Since $\delta f(uvw)$ has weight ε in total, there must be some vertex v such that the weight on triangles touching v is also no more than ε. Choosing this as out initial v we get

$$\begin{aligned}\mathop{\mathbb{P}}_e[f(e) \neq \delta g(e)] &= \mathop{\mathbb{P}}_e[f(e) \neq \delta g(e) \mid e \ni v] \cdot \mathbb{P}[e \ni v] + \mathop{\mathbb{P}}_e[f(e) \neq \delta g(e) \mid e \not\ni v] \cdot \mathbb{P}[e \not\ni v] \\ &\leqslant 0 \cdot \frac{2}{n} + \varepsilon \cdot \frac{n-2}{n} \\ &= \varepsilon \cdot \frac{n-2}{n}.\end{aligned}$$

For general k the proof is nearly the same. Let $f \in C_k$ and let $\varepsilon = wt(\delta f)$. We will find $g \in C_{k-1}$ such that $\delta g \approx f$.

Choose some $v \in V$, and define $g \in C_{k-1}$ by setting

$$\forall t \in X(i-1), \quad g(t) = \begin{cases} 0, & t \ni v \\ f(t \cup \{v\}), & t \not\ni v \end{cases}.$$

Fix first some $s \in X(i)$ that contains v. We have by definition

$$\delta g(s) = \sum_{t<s} g(t) = g(s \setminus \{v\}) = f(s \setminus \{v\} \cup \{v\}) = f(s).$$

Now assume $s \not\ni v$.

$$\delta g(s) = \sum_{t<s} g(t) = \sum_{t<s} f(t \cup \{v\}) = \delta f(s \cup \{v\}) + f(s)$$

where the last equality is by definition of $\delta f(r)$ at $r = s \cup \{v\}$. We see that $\delta g(s) = f(s)$ whenever $\delta f(s \cup \{v\}) = 0$. So let us choose the vertex v such that δf is non zero on the smallest number of sets $r \ni v$. By averaging this is at most ε. Like before we get

$$\begin{aligned}\mathop{\mathbb{P}}_{r \in X(k)}[f(r) \neq \delta g(r)] &= \mathop{\mathbb{P}}_r[f(r) \neq \delta g(r) \mid r \ni v] \cdot \mathbb{P}[r \ni v] + \mathop{\mathbb{P}}_r[f(r) \neq \delta g(r) \mid r \not\ni v] \cdot \mathbb{P}[r \not\ni v] \\ &\leqslant 0 \cdot \frac{k+1}{n} + \varepsilon \cdot \frac{n-k-1}{n} \\ &= \varepsilon \cdot \frac{n-k-1}{n}.\end{aligned}$$

□

3 Linearity Testing

Given a function $f : \mathbb{Z}_2^n \to \mathbb{Z}_2$, how can we test if it is a linear function? By linear we mean a function of the form $f(x) = \sum_{i=1}^n a_i x_i \mod 2$ for some coefficients $(a_1, \ldots, a_n)$.

The linear functions are the functions

$$\mathcal{L} = \left\{ f_a : \mathbb{Z}_2^n \to \mathbb{Z}_2 \;\middle|\; f_a(x) = \sum_{i=1}^n a_i x_i \mod 2 \right\}$$

and the question is about membership of f in $\mathcal{L}$. A possible test for linearity is this: choose a random pair of points $x, y \in V$ and accept it

$$f(x) + f(y) = f(x + y).$$

Clearly if f is linear, the test will pass with probability 1. Blum Luby and Rubinfeld proved [1] that it is a "good test" in the following sense,

Theorem 3.1 (BLR Linearity Testing). *Let $f : \mathbb{Z}_2^n \to \mathbb{Z}_2$. Then*

$$Prob_{x,y \in \mathbb{Z}_2^n}[f(x) + f(y) \neq f(x + y)] \geqslant \min(\frac{2}{9}, \mathrm{dist}(f, \mathcal{L})).$$

In particular this means that the only functions satisfying all of these tests (for all $x, y \in \mathbb{Z}_2^n$) are the linear functions.

We will prove this theorem shortly, but first we would like to rephrase it in cohomological terms. We introduce the following chain of linear maps

$$V = \mathbb{Z}_2^n \xrightarrow{\delta_0} \mathbb{Z}_2^V \xrightarrow{\delta_1} \mathbb{Z}_2^{V \times V} \tag{3.1}$$

where δ_0 maps a sequence of coefficients $a \in V$ to a linear function $\delta_0 a = f_a$; and where δ_1 maps a function $f \in \mathbb{Z}_2^V = \mathbb{Z}_2^{\mathbb{Z}_2^n}$ to $\delta_1 f \in \mathbb{Z}_2^{V \times V}$ defined by

$$\forall x, y \in \mathbb{Z}_2^n, \quad \delta_1(f)(x, y) = f(x) + f(y) + f(x + y).$$

This presentation follows a very nice lecture by Uli Wagner, [6]. The following claim is pretty obvious,

Claim 3.2. Every $f \in \mathcal{L}$ satisfies, for all x, y, $f(x) + f(y) = f(x + y)$. Namely, $\delta_1 \circ \delta_0 = 0$

Proof. The first part is clear, to see that it implies the second part, observe that $\mathcal{L} = Im(\delta_0)$, and that $Ker\delta_1$ is the set of functions that satisfy $f(x) + f(y) = f(x + y)$ for all x, y. ◻

The above claim implies that we are looking at a chain complex with two maps, also known as a 2-chain. In general,

Definition 3.3. A chain complex is a sequence of linear maps

$$V_0 \xrightarrow{d_0} V_1 \xrightarrow{d_1} V_2 \xrightarrow{d_2} \cdots$$

such that for all $i \geqslant 0$, $d_{i+1} \circ d_i = 0$.

We define the cosystolic expansion of a chain complex at level i in the natural way,

Definition 3.4 (Expansion of a chain complex)**.** Given a chain complex

$$V_0 \xrightarrow{d_0} V_1 \xrightarrow{d_1} V_2 \xrightarrow{d_2} \cdots$$

we define

$$h^i = \min_{f \in V_i \setminus Ker(d_i)} \frac{wt(d_i f)}{\mathrm{dist}(f, Ker(d_i))}.$$

Returning to our chain (3.1). It is easy to see that $Ker(\delta_1) = \mathcal{L} = Im(\delta_0)$. Moreover, Theorem 3.1 implies that

$$h^1 \geqslant 2/9.$$

Proof. Let us now prove the theorem in three steps. Fix $f : \mathbb{Z}_2^n \to \mathbb{Z}_2$, and denote $\varepsilon = wt(\delta_1 f) = Prob_{x,y\in\mathbb{Z}_2^n}[f(x) + f(y) \neq f(x + y)]$. Assume that $\varepsilon < 2/9$, otherwise we are done. The proof will find a function $g \in \mathcal{L}$ and show it is close to f. We define $g : \mathbb{Z}_2^n \to \mathbb{Z}_2$ by

$$\forall x \in \mathbb{Z}_2^n, \quad g(x) =_{y\in\mathbb{Z}_2^n} f(y) + f(x + y).$$

Every y "votes" for the value of $g(x)$, and the final value is decided by majority. The proof proceeds in three steps, captured by the following three claims.

Claim 3.5. $\mathrm{dist}(f, g) = \mathbb{P}_x[g(x) \neq f(x)] \leqslant \varepsilon$.

In the next step we will see that the value $g(x)$ is obtained by a relatively vast majority

Claim 3.6. For all x, $\mathbb{P}_y[g(x)] \neq f(x + y) + f(y)] < 1/3$.

Finally, we will prove that g is linear,

Claim 3.7. g is linear, namely $g \in \mathcal{L}$.

Proof. Proof of Claim 3.5 Write

$$\mathop{\mathbb{P}}_{x,y}[g(x) = f(x + y) + f(y)] = \mathop{\mathbb{E}}_x \mathop{\mathbb{E}}_y[\mathbf{1}_{g(x)=f(x+y)+f(y)}] \geqslant \mathop{\mathbb{E}}_x \mathop{\mathbb{E}}_y[\mathbf{1}_{f(x)=f(x+y)+f(y)}] = \varepsilon,$$

where the inequality holds by choice of $g(x)$. ◻

Proof. Proof of Claim 3.6 Let $p_x = \mathbb{P}_y[f(y) + f(x + y) = g(x)$. We know that $p_x \geqslant 1/2$, and need to show that $p_x > 2/3$. If we choose y_1, y_2 independently, then the probability that they vote in the same way is $p_x^2 + (1 - p_x)^2$. We lower bound this by

$$\begin{aligned} p_x^2 + (1 - p_x)^2 &= \mathop{\mathbb{P}}_{y_1,y_2}[f(y_1) + f(x + y_1) = f(y_2) + f(x + y_2)] \\ &= \mathop{\mathbb{P}}_{y_1,y_2}[f(y_1) + f(x + y_2) = f(y_2) + f(x + y_1)] \geqslant 1 - 2\varepsilon \geqslant 5/9. \end{aligned}$$

Here the inequality is because y_1 and $x + y_2$ are distibuted as two independent points so with probability at least $1 - \varepsilon$, $f(y_1) + f(x + y_2) = f(x + y_1 + y_2)$. Similarly with probability at least $1 - \varepsilon$, $f(y_2) + f(x + y_1) = f(x + y_1 + y_2)$. By union bound, the probability that both events hold is at least $1 - 2\varepsilon$. We deduce that $p_x^2 + (1 - p_x)^2 > 5/9$, which implies $p_x > 2/3$. ◻

Proof. Proof of Claim 3.7 Fix some arbitrary x, y, and choose z uniformly at random. Now, by Claim 3.6,

$$\begin{aligned} \mathop{\mathbb{P}}_z[g(x) \neq f(z) + f(x + z)] &< 1/3 \\ \mathop{\mathbb{P}}_z[g(y) \neq f(y + z) + f(z)] &< 1/3 \\ \mathop{\mathbb{P}}_z[g(x + y) \neq f(x + z) + f(y + z)] &< 1/3 \end{aligned}$$

There is at least one z for which all three equalities hold. Summing over the three equations, the right hand side becomes 0 because everything cancels, so we are left with $g(x) + g(y) = g(x + y)$ ◻

Together, the three claims imply the theorem ◻

We end this lecture with an interesting observation. The proof of Claim 3.7 really shows one more step in our chain complex: $\delta_2 : \mathbb{Z}_2^{V\times V} \to \mathbb{Z}_2^{V\times V\times V}$, defined by

$$\forall F \in \mathbb{Z}_2^{V\times V},\ \forall x,y,z \in \mathbb{Z}_2^n, \qquad \delta_2 F(x,y,z) = F(z, x+z) + F(y+z, z) + F(x+z, y+z).$$

Observe that for every $f : \mathbb{Z}_2^n \to \mathbb{Z}_2$, if we choose $F = \delta_1 f$, then $\delta_2 F = 0$ (this was used in the proof above). This means that $\delta_2 \circ \delta_1 = 0$ so indeed the entire sequence of maps

$$V = \mathbb{Z}_2^n \xrightarrow{\delta_0} \mathbb{Z}_2^V \xrightarrow{\delta_1} \mathbb{Z}_2^{V\times V} \xrightarrow{\delta_2} \mathbb{Z}_2^{V\times V\times V}$$

is a chain complex. We used the extra step in the chain to prove expansion of 1-chains. This phenomenon will repeat itself when we prove cosystolic expansion of high dimensional expanders.

References

[1] Manuel Blum, Michael Luby, and Ronitt Rubinfeld. Self-testing/correcting with applications to numerical problems. In *Proc. 22nd ACM Symp. on Theory of Computing*, pages 73–83, 1990. 4

[2] Oded Goldreich. *Introduction to Property Testing*. Cambridge University Press, 2017. 2

[3] Mikhail Gromov. Singularities, expanders and topology of maps. part 2: from combinatorics to topology via algebraic isoperimetry. *Geometric and Functional Analysis*, 20(2):416–526, 2010. 2

[4] Nathan Linial and Roy Meshulam. Homological connectivity of random 2-complexes. *Combinatorica*, 26(4):475–487, 2006. 2

[5] R. Meshulam and N. Wallach. Homological connectivity of random k-dimensional complexes. *Random Structures and Algorithms*, 34:408–417, 2008. 2

[6] Uli Wagner. *Expansion from a Cohomological Viewpoint*, 2019. `https://simons.berkeley.edu/talks/cohomological-viewpoint-expansion`. 4