COMP760, LECTURES 12-13: REVERSE BONAMI-BECKNER AND EXPANSION, GAUSSIAN SPACE

HAMED HATAMI

In this lecture we are going to address a question related to expansion. We choose an element in a product space and change each coordinate with a small probability. How large is the probability that starting in a given small set A, the new point lands in another small set B? We are going to prove a lower bound on this probability that depends on the relative densities of such sets. To this end we introduce new tools concerning $\|\cdot\|_p$ with $p < 1$.

1. Noise and expansion

A ρ-correlated copy of an element x is obtained by changing some of its coordinates.

Definition 1.1. *For $0 \leq \rho \leq 1$, let x be a uniform random variable taking values in $\{0,1\}^n$, and let $y \in \{0,1\}^n$ be a random variable such that for each i independently*

$$Pr[y_i = x_i] = \frac{1}{2}(1+\rho),$$

or equivalently,

$$\mathbb{E}(-1)^{x_i}(-1)^{y_i} = \rho.$$

In this case y is called a ρ-correlated copy of x.

Recall that the noise operator can be expressed in terms of a correlated copy of x,

$$T_\rho f(x) = \mathbb{E}_y f(y).$$

Consider $A, B \subseteq \{0,1\}^n$ with relative densities α, β. That is

$$\frac{|A|}{2^n} = \alpha, \qquad \frac{|B|}{2^n} = \beta.$$

Note that α, β are small but may not be constant. We are interested in the following question: Pick a random $x \in \{0,1\}^n$ and ρ-correlated copy y of x; How small can the following probability can be?

$$\Pr[x \in A, y \in B] = \alpha \Pr[y \in B | x \in A].$$

If we want to minimize this probability intuitively we would like to choose two opposite corners of the cube. The probability in this case can be upper-bounded using the following lemma whose proof we omit.

Lemma 1.2. *Fix $a, b > 0$ and let $A, B \subseteq \{0,1\}^n$ be*

$$A = \left\{x | \sum x_i \leq \frac{n}{2} - a\sqrt{n}\right\},$$
$$B = \left\{x | \sum x_i \leq \frac{n}{2} - b\sqrt{n}\right\}.$$

These notes are scribed by Laura Eslava.

Let $x \in \{0,1\}^n$ be uniform and y be a correlated copy of x. Then we have the following upper bound

$$\lim_{n\to\infty} Pr[x \in A, y \in B] \le \frac{\sqrt{1-\rho^2}}{2\pi a(\rho a + b)} e\left\{-\frac{1}{2}\cdot\frac{a^2+b^2+2\rho ab}{1-\rho^2}\right\}$$

The main term in the bound above is the exponential one and it involves the relative densities of A and B as

$$\lim_{n\to\infty} \frac{|A|}{2^n} = \frac{1}{\sqrt{2\pi a}} e^{-a^2/2},$$
$$\lim_{n\to\infty} \frac{|B|}{2^n} = \frac{1}{\sqrt{2\pi b}} e^{-b^2/2}.$$

We are going to establish a lower-bound in Theorem 2.5 that almost matches the upper-bound of Lemma 1.2.

Let us first try the straight forward Fourier analytic approach that correspond to a spectral gap method as the eigenvectors of T_ρ are the characters because

$$T_\rho \chi_S = \rho^{|S|}\chi_S.$$

Therefore, the eigenvalues of T_ρ are $\rho^{|S|}$, the largest is 1, corresponding to the principal character ($S=\emptyset$), and the second largest is ρ, recall $\rho < 1$. To compute the Fourier expansions, fix x and average over y,

$$\Pr[x \in A, y \in B] = \mathbb{E}1_A(x)1_B(y) = \mathbb{E}1_A(x)T_\rho 1_B(x).$$

We can use an spectral gap method, that is, to remove the first coefficient and bound the other ones by the second largest, finally we use Cauchy-Schwarz to derive the following,

$$\begin{aligned}\sum \widehat{1_A}(S)\widehat{1_B}(S)\rho^{|S|} &\ge \widehat{1_A}(\emptyset)\widehat{1_B}(\emptyset) - \rho\sum_{S\neq\emptyset}\left|\widehat{1_A}(S)\widehat{1_B}(S)\right| \\ &\ge \alpha\beta - \rho\left(\sum_{S\neq\emptyset}\left|\widehat{1_A}(S)\right|^2\right)^{1/2}\left(\sum_{S\neq\emptyset}\left|\widehat{1_A}(S)\right|^2\right)^{1/2} = \alpha\beta - \rho\sqrt{\alpha-\alpha^2}\sqrt{\beta-\beta^2}.\end{aligned}$$

The last term is large and the bound is negative (and useless) unless ρ is very small. So we need a deeper approach.

2. Reverse Bonami-Beckner

We are going to use "L_p-norms" for $p < 1$, and obtain a Reverse Bonami-Beckner inequality. It is similar to the hypercontractivity theorem but the direction of the inequality is reversed and applies to $-\infty < q \le p < 1$. Also unlike the original Bonami-Beckner inequality, it only applies to non-negative functions. In fact all of the next 4 theorems and lemmas require the functions to be non-negative.

Theorem 2.1 (Inverse Hölder Inequality). *If $f, g \ge 0$ are measurable functions with respect to a measure space then*

$$\langle f, g\rangle \ge \|f\|_p \|g\|_q,$$

where $-\infty < q, p < 1$ and $\frac{1}{p}+\frac{1}{q} = 1$.

Remark 2.2. When $p < 1$, the function $\|\cdot\|_p$ is not a norm, and it is a notation to denote

$$\left(\int |f|^p\right)^{1/p}.$$

In fact if $f, g \geq 0$ then the triangle inequality is reversed for $-\infty < p < 1$,

$$\|f + g\|_p \geq \|f\|_p + \|g\|_p.$$

To see this note that by Inverse Hölder Inequality

$$\begin{aligned} \|f+g\|_p^p &= \int (f+g)^p = \int (f+g)^{p-1} f + \int (f+g)^{p-1} g \\ &\geq \left(\int (f+g)^p\right)^{\frac{p-1}{p}} \|f\|_p + \left(\int (f+g)^p\right)^{\frac{p-1}{p}} \|g\|_p, \end{aligned}$$

which simplifies to the desired inequality. ∎

Theorem 2.3 (Reverse Bonami-Beckner inequality)**.** *Let $f : \{0,1\}^n \to \mathbb{R}^+$, then*

$$\|T_\rho f\|_q \geq \|f\|_p,$$

for $0 \leq \rho \leq \sqrt{\frac{p-1}{q-1}}$ and $-\infty < q \leq p < 1$.

The proof is similar to the Bonami-Beckner inequality. First one proves it for the 1-dimensional case and then induction establishes the general case.

Corollary 2.4. *Let $f, g : \{0,1\}^n \to \mathbb{R}^+$ and $x \in \{0,1\}^n$ uniform and a ρ-correlated y copy of x, then*

$$\mathbb{E} f(x) g(y) \geq \|f\|_p \|g\|_q,$$

where $0 < \rho \leq \sqrt{(1-p)(1-q)} \leq 1$ and $-\infty < q, p < 1$.

Proof. Let $p' = \frac{p}{1-p}$, so that p, p' are conjugate exponents. We use the reverse Hölder's inequality and then apply the inverse Bonami-Beckner inequality,

$$\begin{aligned} \mathbb{E} f(x) g(y) &= \mathbb{E} f(x) T_\rho g(x) \\ &\geq \|f\|_p \|T_\rho g\|_{p'} \\ &\geq \|f\|_p \|g\|_q, \end{aligned}$$

where the last inequality requires $0 < \rho \leq \sqrt{\frac{1-q}{1-p'}} = \sqrt{(1-p)(1-q)}$. □

Now we can prove the main theorem of the lecture, regarding the lower bound on the probability that a ρ-correlated copy of a uniform element that is in A lands in B.

Theorem 2.5. *Let $A, B \subseteq \{0,1\}^n$ have relative densities*

$$\frac{|A|}{2^n} = e^{-a^2/2} \qquad \frac{|B|}{2^n} = e^{-b^2/2},$$

and let $x \in \{0,1\}^n$ be uniform and y be a ρ-correlated copy of x. Then

$$Pr[x \in A, y \in B] \geq e\left\{-\frac{1}{2} \cdot \frac{a^2 + b^2 + 2\rho ab}{1-\rho^2}\right\}.$$

Proof. Let p, q be such that $\rho^2 = (1-p)(1-q)$, by corollary 2.4 we have that
$$\Pr[x \in A, y \in B] = \mathbb{E}1_A(x)1_B(y) \geq \|1_A\|_p \|1_B\|_q.$$
Now our task is to optimize p so that the the R.H.S. is maximized. Note that the L_p norm can be expressed in term of the relative density because we are dealing with an indicator function
$$\|1_A\|_p = e^{-a^2/2p} \qquad\qquad \|1_B\|_p = e^{-b^2/2q}.$$
To simplify computations, write $p = 1 - \rho r$ and $q = 1 - \frac{\rho}{r}$ with $r > 0$, where
$$r = \frac{1-p}{\rho} = \frac{\rho}{1-q}.$$
Then the optimal solution is achieved when
$$r = \frac{\frac{b}{a} + \rho}{1 + \rho\frac{b}{a}}.$$
This gives the claimed lower bound as for the optimal value of r,
$$\frac{a^2}{p} + \frac{b^2}{q} = \frac{a^2 + b^2 + 2\rho ab}{1-\rho^2}.$$
□

We obtain the following corollary.

Corollary 2.6. *Let $A, B \subseteq \{0,1\}^n$ with relative densities $\alpha > 0$ and $\alpha^\sigma > 0$ respectively, where $\sigma > 0$. Let $x \in \{0,1\}^n$ be uniform and y be a ρ-correlated copy of x. Then*
$$Pr[x \in A, y \in B] \geq \alpha\alpha^{(\sqrt{\sigma}+\rho)^2/(1-\rho^2)}.$$
In particular, if $|A| = |B|$, the this probability is at least $\alpha^{(1+\rho)/(1-\rho)}$.

Another interesting corollary of the inverse Bonami-Beckner inequality is that we can measure how flat T_ρ can be, since we expect T_ρ to smooth f. That is, we can bound $\Pr[T_\rho f(x) > 1 - \delta]$.

Theorem 2.7. *Let $f : \{0,1\}^n \to [0,1]$ with $\mathbb{E}f = \alpha$. Then for any $0 < \rho < 1$ and $0 \leq \epsilon \leq 1 - \alpha$ we have*
$$Pr[T_\rho f > 1 - \delta] < \epsilon$$
provided that $0 \leq \delta < \epsilon^{\rho^2/(1-\rho^2)+O(\kappa)}$, where $\kappa = \sqrt{\frac{\alpha \log(e/1-\alpha)}{1-\rho}}$.

Proof. Define indicator functions
$$g : x \to \begin{cases} 1 & \text{if } T_\rho f(x) > 1 - \delta \\ 0 & \text{otherwise} \end{cases}$$
$$h : x \to \begin{cases} 1 & \text{if } f(x) > b \\ 0 & \text{otherwise,} \end{cases}$$
where $b = \frac{1}{2}(1 + \alpha)$. We need to show that $\epsilon' := \mathbb{E}g \leq \epsilon$. By the first moment method,
$$\alpha = \mathbb{E}f \geq (1 - \mathbb{E}h)b,$$
then
$$\mathbb{E}h > 1 - \frac{\alpha}{b} = \frac{1-\alpha}{1+\alpha},$$

and therefore support of h is not very small. Now, when $g(x) = 1$ we have $T_\rho(1 - f(x)) \le \delta$, so

$$T_\rho[(1-b)h(x)] < \delta$$

and so

$$T_\rho[h(x)] \le \frac{\delta}{1-b}.$$

Thus

$$\mathbb{E}[gT_\rho h] \le \frac{\delta\epsilon'}{1-b} = \frac{2\delta\epsilon'}{1-\alpha}. \tag{1}$$

Meanwhile, By Corollary 2.6

$$\mathbb{E}[gT_\rho h] \ge \epsilon' \cdot \epsilon'^{\frac{(\sqrt{\beta}+\rho)^2}{1-\rho^2}} \tag{2}$$

where $\beta = \frac{\log \mathbb{E}h}{\log \epsilon'}$. Now (1) and (2) together with our assumption on δ leads to the desired bound $\epsilon' \le \epsilon$. ◻

Lecture 13

In this lecture we define the Noise Stability of boolean function. We are interested in finding the Boolean functions that have largest noise stability. It was conjectured by Subash Khot that the majority function is the stablest Boolean function. The analogous statement in the Gaussian setting was proved in 1983 by Borell. Recently Mossel, O'Donnell, Oleszkiewicz found a method to deduce the discrete case from Borell's result. In this lecture we introduce the basic notions of gaussian space. This will give us an analogous setup. The proofs are going to be shown in the following lecture.

We start with the definition of noise operator, this measures the proportion of elements x that remains in the support of a function f when we add some noise to them.

Definition 2.8. *For $0 \le \rho \le 1$, the noise stability of $f : \{0,1\}^n \to \mathbb{R}$ is*

$$\mathbb{S}_\rho(f) := \mathbb{E}f(x)f(y),$$

where y is a ρ-correlated copy of x and $x \in \{0,1\}^n$ is uniform.

Note that for fixed x, $\mathbb{E}f(y) = T_\rho f(x)$, so we can look at $\mathbb{S}_\rho f$ as a correlation between f and $T_\rho f$, then

$$\mathbb{S}_\rho(f) = \mathbb{E}f(x)T_\rho f(x) = \sum_{S\subseteq[n]} \rho^{|S|}|\hat{f}(S)|^2.$$

Now, for a balanced function $f : \{0,1\}^n \to \{0,1\}$ what is the largest possible value of $\mathbb{S}_\rho f(x)$? A first approach is to separate the principal coefficient and upper bound $\rho^{|S|}$ by ρ, so we get

$$\begin{aligned}\mathbb{S}_\rho(f) &= \sum_{S\subseteq[n]} \rho^{|S|}|\hat{f}(S)|^2 \\ &\le |\mathbb{E}f|^2 + \rho\sum_{S\neq\emptyset}|\hat{f}(S)|^2 = \frac{1}{4} + \frac{\rho}{4},\end{aligned}$$

where in the last equality we used the assumption that f is balanced and boolean. On the other hand, half-cubes achieve this upper bound, for example, if $f(x) = x_1$, then the fourier expansion is

$$f = \frac{1}{2} + \frac{1}{2}\chi_{\{i\}},$$

and so

$$\mathbb{S}_\rho(f) = \frac{1}{4} + \frac{\rho}{4}.$$

But the influence of half-cubes is concentrated in one coordinate. So, if we assume that all the influences are small, we overrule the simple case of half-cubes.

In contrast with the half-cubes where the fourier coefficients are concentrated in the first level, the next theorem states that even when all the influences are small, if one looks at the levels above d, the sum of the squared fourier coefficients is still large.

Theorem 2.9 (Bourgain 2000)**.** *If $f : \{0,1\}^n \to \{0,1\}$ is balanced and $I_i \le 10^{-d}$ for all $i \in [n]$, then*

$$\|f^{\ge d}\|_2^2 = \sum_{|S| \ge d} |\hat{f}(S)|^2 \ge d^{-1/2 - o\left(\sqrt{\frac{\ln\ln d}{\ln d}}\right)},$$

which is $d^{-1/2-o(1)}$.

As an exercise, prove the following corollary of Theorem 2.9

Corollary 2.10. *If $f : \{0,1\}^n \to \{0,1\}$ is balanced and $I_i(f) = 2^{-O(1/\epsilon)}$ for all $i \in [n]$, then*

$$\mathcal{S}_{1-\epsilon}(f) \le \frac{1}{2} - \epsilon^{1/2+o(1)}$$

Note that this is a great improvement compared with the first bound, that would be $\frac{1}{2} - \frac{\epsilon}{4}$. However, this is not sharp as the majority function has an even larger noise stability and it is conjecture that it achieves essentially the maximum noise stability among balanced functions. The majority function $\mathrm{Maj}_n : \{0,1\}^n \to \{0,1\}$ is defined as

$$\mathrm{Maj}_n : x \to \begin{cases} 1 & \sum x_i \ge n/2 \\ 0 & \sum x_i < n/2. \end{cases}$$

Theorem 2.11. *The limit as n tends to ∞ of the noise stability of the majority function is*

$$\lim_{n\to\infty} \mathbb{S}_\rho(\mathrm{Maj}_n) = \frac{1}{4} + \frac{\arccos\rho}{2\pi}.$$

Theorem 2.12 (Majority is stablest)**.** *For $0 < \rho < 1$, if $f : \{0,1\}^n \to \{0,1\}$ is balanced and $I_i(f) \le \epsilon$ for all $i \in [n]$, then*

$$\mathbb{S}_\rho(f) \le \frac{1}{4} + \frac{\arccos\rho}{2\pi} + O\left(\frac{\log\log 1/\epsilon}{\log 1/\epsilon}\right).$$

Note that $O\left(\frac{\log\log 1/\epsilon}{\log 1/\epsilon}\right) = o(\epsilon)$. This theorem together with the so called "unique games conjecture" imply strong results about hardness of approximation.

For the proof of Theorem 2.12, we actually have to use geometry. The rest of the lecture we will define gaussian random variables in $\mathbb{R}^n$, state some of their basics properties and settle an analogous setup for the noise operators. In the next lecture we will prove the analogue of Theorem 2.11 in $\mathbb{R}^n$ and then translate it back to the discrete case.

Definition 2.13. *The normal distribution on* $\mathbb{R}$ *is the probability distribution* γ_1 *on* $\mathbb{R}$ *with density function*

$$\frac{e^{-x^2/2}}{\sqrt{2\pi}},$$

that is

$$\gamma_1([a,b]) = \int_a^b \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx.$$

A random variable g with distribution γ_1 is called a gaussian, these random variables have the property that $\mathbb{E}g = 0$ and $\mathbb{E}g^2 = 1$.

Definition 2.14. *Let* γ_n *denote the corresponding product probability distribution on* $\mathbb{R}^n$*. In other words,*

$$\gamma_n(\{x \in \mathbb{R}^n | a_i \leq x_i \leq b_i\}) = \prod_{i=1}^{n} \gamma_1([a_i, b_i]). \tag{3}$$

The density function of γ_n *is*

$$\Phi_n(x) = \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{\frac{-\|x\|^2}{2}},$$

that is, if $A \in \mathbb{R}^n$*, then* $\gamma_n(A) = \int_A \Phi_n(x)dx$*.*

Remark 2.15. The way that we defined the gaussian measure on $\mathbb{R}^n$ as the product space in (3) it is natural to expect that gaussians depend on the coordinates but they do not. The measure γ_n is uniformly distributed in spheres centered at the origin; when one fixes a sphere, the function Φ_n becomes constant and therefore independent of the coordinates.

In particular if $g_1, \ldots, g_n$ are i.i.d. gaussians and $\alpha, \beta \in \mathbb{R}^n$ with $\|\alpha\|_2 = \|\beta\|_2$, then the random variables

$$\alpha_1 g_1, + \cdots + \alpha_n g_n \qquad \text{and} \qquad \beta_1 g_1 + \cdots + \beta_n g_n,$$

have the same distribution. In particular, $\sum \alpha_i g_i$ has the same distribution as $\|\alpha\|_2 g$, where g is a one dimensional gaussian. ■

Now we consider the characters of $\mathbb{Z}_2^n$ in the gaussian space. Let $S \subseteq [n]$, then we define

$$\omega_S : x \mapsto \prod_{i \in S} x_i.$$

Lemma 2.16. *The functions* $\omega_S : (\mathbb{R}^n, \gamma_n) \to \mathbb{R}$ *are orthonormal.*

Proof. The inner product of any two function ω_S, ω_T can be express as the expected value of its product with respect to γ_n, so

$$\langle \omega_S, \omega_T \rangle \;\; = \int \omega_S(x)\omega_T(x) d\gamma_n(x) = \mathbb{E}\omega_S(g_1, \ldots, g_n)\omega_S(g_1, \ldots, g_n),$$

where g_i are i.i.d. gaussians, so by independence

$$\mathbb{E}\omega_S(g_1, \ldots, g_n)\omega_S(g_1, \ldots, g_n) = \mathbb{E}\prod_{i \in S} g_i \prod_{i \in T} g_i = \left(\prod_{i \in S \cap T} \mathbb{E}g_i^2\right)\left(\prod_{i \in S \triangle T} \mathbb{E}g_i\right) = \begin{cases} 0 & S \triangle T \neq \emptyset \\ 1 & \text{otherwise} \end{cases}$$

Therefore, the inner product of any two of those functions is zero unless they are the same functions, and the norm of all of them is 1. □

Remark 2.17. $\{\omega_S\}_{S \subseteq [n]}$ do not generate all of $L_2(\mathbb{R}^n, \gamma_n)$ but one can extend them using the so called Hermité polynomials to a basis for $L_2(\mathbb{R}^n, \gamma_n)$. ■

To define the noise stability we have to define what a ρ-correlated of a gaussian is:

Definition 2.18. *Let $0 \leq \rho \leq 1$, two gaussians g, h are ρ-correlated if*

$$g = \rho h + \sqrt{1-\rho^2} g',$$

where g' is a gaussian independent of g, h.

Here g and h have ρ correlation. To see this use the definition of g to get

$$\mathbb{E}(g(x)h(x)) = \mathbb{E}[\rho h^2 + \sqrt{1-\rho^2} h g'] = \mathbb{E}\rho = \rho,$$

we again used that the expected value of a gaussian is zero and the second moment is 1. On the other hand, the coefficients are chosen so that g is a gaussian, note that $g = \rho h + \sqrt{1-\rho^2} g'$ has the same distribution as

$$(\rho^2 + \sqrt{1-\rho^2}) g'' = g'',$$

where g'' is a gaussian. Now, we define the analogue of the "noise operator" for the functions on the gaussian space.

Definition 2.19. *Let $0 \leq \rho \leq 1$, then the Ornestein-Uhlenbeck operator acting on $L_2(\mathbb{R}^n, \gamma_n)$ is defined as*

$$U_\rho f(x) = \mathbb{E} f(y),$$

where $y = \rho x + \sqrt{1-\rho^2} g$ is a ρ-correlated copy of x.

Lemma 2.20. *We have $U_\rho \omega_S = \rho^{|S|} \omega_S$.*

Proof. For fixed x_i's we have

$$U_\rho \omega_S(x) = \mathbb{E} \prod_{i \in S} (\rho x_i + \sqrt{1-\rho^2} g_i) = \rho^{|S|} \prod_{i \in S} x_i = \rho^{|S|} \omega_S(x).$$

□

Finally, we get to define the noise stability for function on the guassian space.

Definition 2.21. *The noise stability of $f : (\mathbb{R}^n, \gamma_n) \to \mathbb{R}$ is defined as*

$$\mathbb{S}_\rho(f) := \mathbb{E}(f U_\rho f) = \mathbb{E} f(x) f(y),$$

where x has distribution γ_n and y is a ρ-correlated copy of x.

The analogues of Theorem 2.11 and Theorem 2.12 in the setting of gaussians are respectively

Theorem 2.22. *Let $\mathrm{Maj}_n : \mathbb{R}^n \to \{0, 1\}$ be defined as*

$$\mathrm{Maj}_n(x) = \begin{cases} 1 & \text{if } \sum x_i \geq 0 \\ 0 & \text{if } \sum x_i < 0 \end{cases}$$

then, its noise stability is given by

$$\mathbb{S}_\rho(\mathrm{Maj}_n) = \frac{1}{4} + \frac{\arccos \rho}{2\pi}.$$

and

Theorem 2.23. *If $A \subseteq \mathbb{R}^n$ satisfies $\gamma_n(A) = 1/2$, then*

$$\mathbb{S}_\rho(1_A) = \mathbb{S}_\rho(A) \leq \frac{1}{4} + \frac{\arccos \rho}{2\pi}.$$

Note that unlike in Theorem 2.12, the previous theorem does not require any conditions on the influences.

School of Computer Science, McGill University, Montréal, Canada
E-mail address: hatami@cs.mcgill.ca