Electronic Colloquium on Computational Complexity, Report No. 51 (2011)

A concentration inequality for the overlap of a vector on a large set

With application to the communication complexity of the Gap-Hamming-Distance problem

Thomas Vidick[*]

March 22, 2011

Abstract

Given two sets $A, B \subseteq \mathbb{R}^n$, a measure of their dependence, or correlation, is given by the expected squared inner product between random $x \in A$ and $y \in B$. We prove an inequality showing that no two sets of large enough Gaussian measure (at least $e^{-\delta n}$ for some constant $\delta > 0$) can have correlation substantially lower than would two random sets of the same size. Our proof is based on a concentration inequality for the overlap of a random vector on a large set.

As an application, we show how our result can be combined with the partition bound of Jain and Klauck to give a simpler proof of a recent linear lower bound on the randomized communication complexity of the Gap-Hamming-Distance problem due to Chakrabarti and Regev.

1 Introduction

Let $A, B \subseteq \mathbb{R}^n$, and let γ be the n-dimensional Gaussian measure. Denote by $\gamma_{|A\times B}$ the measure corresponding to the normalized restriction of $\gamma \times \gamma$ to $A \times B$, and let

$$\nu(A,B) \;:=\; \mathrm{E}_{(x,y)\sim\gamma_{|A\times B}}\left[\,(x\cdot y)^2\,\right].$$

The quantity $\nu(A,B)$ can be interpreted as a measure of correlation between A and B, in the sense that a large ν indicates sets with mostly aligned vectors, while a small ν indicates sets that are close to being orthogonal.

We study the following question: How small can $\nu(A,B)$ be for arbitrary sets A, B? If A contains a single vector x of norm $\sqrt{n}$, then the minimizing B of fixed measure is the fattened equator $B = \{y \in \mathbb{R}^n : -t\sqrt{n} \leq y \cdot x \leq t\sqrt{n}\}$, for which $\nu(\{x\}, B) \leq t^2 n$ and $\gamma(B) \sim \sqrt{2/\pi}t$ for small t. Hence for any fixed $\delta > 0$ there exists a set B of *constant* measure such that $\nu(\{x\}, B) = \delta n$, an arbitrarily small fraction of its expected value for a pair of vectors in $\mathbb{R}^n$ chosen at random according to $\gamma \times \gamma$. In this note we show that in case A is restricted to not being too small (i.e. $\gamma(A) \geq e^{-\delta n}$), then no set B of measure at least $e^{-\delta n}$ can significantly bias $\nu(A,B)$ below its expectation for random vectors. More precisely we show the following:

Theorem 1. *For any $\eta > 0$, there exists a $\delta > 0$ such that for all large enough n, if A, B both have measure $\gamma(A), \gamma(B) \geq e^{-\delta n}$ then*

$$\nu(A,B) \;\geq\; (1-\eta)\,\nu(\mathbb{R}^n,\mathbb{R}^n) \;=\; (1-\eta)\,n, \tag{1}$$

[*] Computer Science division, UC Berkeley, USA. Supported by ARO Grant W911NF-09-1-0440 and NSF Grant CCF-0905626.

ISSN 1433-8092

Note that one may not hope for such a strong inequality in the opposite direction, as the spherical caps $A = B = \{x \in \mathbb{R}^n,\, x_1 \geq \sqrt{2\delta n}\}$ have measure approximately $e^{-\delta n}$ but correlation $\nu(A, B) = \Omega(\delta^2 n^2)$. The proof of the theorem is based on a concentration inequality for the random variable $\nu(\{y\}, S)$, where $y \sim \gamma$ and S is a fixed large enough set, which is described in Lemma 7 below.

Comparison with [2]. Chakrabarti and Regev recently settled the long-standing open problem of the randomized communication complexity of the Gap-Hamming-Distance (GHD) problem, showing a $\Omega(n)$ lower bound for n-dimensional inputs. Their proof is based on a variant of the smooth rectangle bound [4], and at its core is an inequality similar to the one we prove in Theorem 1, except that it applies to the $\cosh$ function, intead of the square function. More precisely, if one defines

$$\tilde{\nu}_\alpha(A, B) := \mathrm{E}_{(x,y)\sim\gamma_{|A\times B}}\left[\cosh(\alpha\, x \cdot y)\right]$$

for any $\alpha > 0$, then the key step in the proof of Theorem 3.5 from [2] consists in showing that, for every $c, \eta > 0$ there is a $\delta > 0$ such that for every $0 \leq \alpha \leq c/\sqrt{n}$ and A, B of measure at least $e^{-\delta n}$,

$$\tilde{\nu}_\alpha(A, B) \;\geq\; (1-\eta)\, \tilde{\nu}_\alpha(\mathbb{R}^n, \mathbb{R}^n). \tag{2}$$

The proof of (2) is based on a powerful result, Theorem 3.1 in [2] which shows that, if A is large enough then for almost all $y \in \mathbb{R}^n$ the distribution of $\langle x, y\rangle$ for $x \sim \gamma_{|A}$ is close to a mixture of translated Gaussians. Theorem 3.1 can be seen to imply both (2) and our Theorem 1. The proof of Theorem 3.1, though, is quite involved, and the main contribution of our work consists in giving a direct proof of our Theorem 1, which we show is strong enough to imply a linear lower bound on the randomized communication complexity of GHD.

2 Preliminaries

Distributions. Let $N(0, \sigma^2)$ denote the distribution of a normal random variable with mean 0 and variance σ^2. Let χ^2 be the distribution of the square of a random variable distributed as $N(0,1)$, and $\chi^2(k)$ the distribution of the sum of the squares of k independent $N(0,1)$ random variables. γ is the n-dimensional Gaussian measure on $\mathbb{R}^n$, with density $\gamma(x) = (2\pi)^{-n/2} e^{-\|x\|^2/2}$. We sometimes abuse notation and also denote by γ the $2n$-dimensional distribution $\gamma \times \gamma$. If $S \subseteq \mathbb{R}^n$, $\gamma_{|S}$ denotes the normalized restriction of γ to S: $\gamma_{|S}(x) = \gamma(x)/\gamma(S)$ if $x \in S$ and 0 otherwise.

Concentration bounds. We will use the following large deviation bounds.

Fact 2 (Gaussian tail bound)**.** *Let X be a standard normal random variable. Then for every $t \geq 0$,*

$$\Pr\big(|X| \geq t\big) \;\leq\; e^{-t^2/2}.$$

Proof. Bound the upper tail as

$$\begin{aligned}
\Pr\big(X \geq t\big) &= \frac{1}{\sqrt{2\pi}} \int_t^\infty e^{-x^2/2}\mathrm{dx} \\
&= \frac{1}{\sqrt{2\pi}} \int_0^\infty e^{-(x+t)^2/2}\mathrm{dx} \\
&\leq \frac{e^{-t^2/2}}{\sqrt{2\pi}} \int_0^\infty e^{-x^2/2}\mathrm{dx} \;=\; \frac{1}{2} e^{-t^2/2}.
\end{aligned}$$

A similar bound holds for the lower tail. □

Fact 3 (Bernstein's inequality, see, e.g., Prop. 16 in [6]). *Let $X_1, \ldots, X_N$ be independent random variables such that for every i, $\mathrm{E}[X_i] = 0$, and there exists $K > 0$ such that, for all i and $t \geq 0$, $\Pr(|X_i| \geq t) \leq e^{1-t/K}$. Then for every $a \in \mathbb{R}^N$ and $t \geq 0$, we have*

$$\Pr\Big(\Big| \sum_i a_i X_i \Big| \geq t \Big) \leq 2e^{-\frac{1}{4e} \min\left(\frac{t^2}{2eK^2\|a\|_2^2}, \frac{t}{K\|a\|_\infty} \right)}.$$

As a corollary, one can obtain the following bound for the tail of the χ^2 distribution.

Claim 4 (χ^2 tail bound). *Let $N \in \mathbb{N}$, and $X_1, \ldots, X_N$ be i.i.d standard normal random variables. Then for every $a_1, \ldots, a_N \in \mathbb{R}$ and $t \geq 0$,*

$$\Pr\Big(\Big| \sum_{i=1}^N a_i\, X_i^2 - \sum_{i=1}^N a_i \Big| \geq t \Big) \;\leq\; 2\, e^{-\frac{1}{8e} \min\left(\frac{t^2}{4e\|a\|_2^2}, \frac{t}{\|a\|_\infty} \right)}.$$

Proof. By Fact 2, for every i the X_i satisfy that for every $t \geq 0$,

$$\begin{aligned} \Pr(|X_i^2 - 1| \geq t) &= \Pr(X_i^2 \geq t+1) + \Pr(X_i^2 \leq 1-t) \\ &\leq e^{1-(t+1)/2} \end{aligned}$$

where the extra factor e ensures that the bound is trivial whenever the second term $\Pr(X_i^2 \leq 1-t)$ is nonzero. Hence the $Y_i := X_i^2 - 1$ satisfy the hypothesis of Fact 3 with $K = 2$, which leads to the claimed bound. □

The bound in Claim 4 becomes very weak as soon as even one of the coefficients a_i is very large. In the case where the a_i are non-negative and most are small we can still keep a good control over the *lower* tail, as the following claim shows.

Claim 5. *Let $N \in \mathbb{N}$, let $X_1, \ldots, X_N$ be i.i.d standard normal random variables, $a_1 \geq \ldots \geq a_N \geq 0$ non-negative reals sorted in decreasing order, and $M = \sum_{i=1}^N a_i$. Then for every $\beta > 0$ and $t \geq 0$,*

$$\Pr\Big(\sum_{i=1}^N a_i\, X_i^2 - M \leq -\sum_{i=1}^{\beta M} a_i - t \Big) \;\leq\; 2\, e^{-\frac{\beta t}{8e} \min\left(\frac{\beta t}{4eM}, 1 \right)}.$$

Proof. Since the a_i are sorted, for every $i > \beta M$ we have $a_i \leq 1/\beta$, so that

$$\|a_{>\beta M}\|_2^2 \;:=\; \sum_{i=\beta M+1}^N a_i^2 \;\leq\; N/\beta^2 \qquad \text{and} \qquad \|a_{>\beta M}\|_\infty \;:=\; \max_{i > \beta M} |a_i| \;\leq\; 1/\beta$$

Hence applying Claim 4 to $X_{\beta M+1}, \ldots, X_M$ yields that for every $t \geq 0$,

$$\Pr\Big(\Big| \sum_{i=\beta M+1}^N a_i\, X_i^2 - \sum_{i=\beta M+1}^N a_i \Big| \geq t \Big) \;\leq\; 2\, e^{-\frac{1}{8e} \min\left(\frac{\beta^2 t^2}{4eN}, \beta t \right)},$$

which proves the claim since $\sum_{i=\beta M}^N a_i\, X_i^2 \leq \sum_{i=1}^N a_i\, X_i^2$. □

We will also use the Berry-Esseen theorem.

Fact 6 (Berry-Esseen Theorem, see, e.g., [3], Chapter XVI). *Let $X_1, \ldots, X_N$ be i.i.d such that $\mathrm{E}[X_i] = 0$, $\mathrm{E}[X_i^2] = \sigma^2$ and $\mathrm{E}[|X_i|^3] = \rho$, define $Y = (X_1 + \cdots + X_N)/(\sqrt{N}\sigma)$ and let Z be distributed as $N(0,1)$. Then for all $t \geq 0$,*

$$\big| \Pr(Y \geq t) - \Pr(Z \geq t) \big| \;\leq\; \frac{3\,\rho}{\sigma^3 \sqrt{N}}$$

Communication complexity. For a partial function $f : X \times Y \to \{0, 1, \star\}$, we let $R_\varepsilon(f)$ be the ε-error randomized communication complexity of the function f (we refer to [5] for more background on communication complexity). Here we allow X, Y to be infinite subsets of $\mathbb{R}^n$ and measure input size by the dimension n alone.

3 Proof of the main inequality

The proof of Theorem 1 is based on a concentration bound for the average squared inner product between a vector $y \in \mathbb{R}^n$ and a random $x \in S$, where S is a fixed subset of $\mathbb{R}^n$. Given such a set, it will be convenient to work with the positive matrix $\mathbf{S} = \mathrm{E}_{x \sim \gamma_{|S}}\left[xx^T\right]$, where the expectation is taken entrywise. This matrix satisfies the following key relation

$$\forall y \in \mathbb{R}^n \qquad y^T \mathbf{S} y \;=\; \mathrm{E}_{x \sim \gamma_{|S}}\left[y^T x x^T y\right] \;=\; \mathrm{E}_{x \sim \gamma_{|S}}\left[(x \cdot y)^2\right]. \tag{3}$$

As we will see, (3) lets us relate the concentration properties of $\nu(\{y\}, S)$, for $y \sim \gamma$, to the spectrum of $\mathbf{S}$. We show the following concentration bound.

Lemma 7. *There exists constants $c, c' > 0$ such that the following holds. Let $\delta > 0$ and $S \subseteq \mathbb{R}^n$ such that $\gamma(S) \geq e^{-\delta n}$. Then for all $\alpha > c'\delta$,*

$$\Pr_{y \sim \gamma}\left(y^T \mathbf{S} y \leq \mathrm{Tr}\,\mathbf{S} - \alpha n\right) \leq e^{-c\alpha^4 n}. \tag{4}$$

Note that, for any set S, $\mathrm{E}_{y \sim \gamma}\left[y^T \mathbf{S} y\right] = \mathrm{Tr}\,\mathbf{S} \approx n$,[1] so that (4) shows that $y^T \mathbf{S} y$ cannot be much lower than its expectation. Before turning to the proof of the lemma, and showing how it implies our main theorem, we give an example showing that the constraint $\alpha > c'\delta$ is necessary (for some $c' > 0$). The same example also shows that one cannot hope for a similar bound on the probability that $y^T \mathbf{S} y$ is *greater* than $\mathrm{Tr}\,\mathbf{S} + \alpha n$, even for relatively large α.

Example. Fix a parameter $\alpha > 0$ (think of α as a small constant), and consider the spherical cap $S_\alpha = \{x \in \mathbb{R}^n : x_1 \geq \sqrt{\alpha n}\}$. Since for $X \sim N(0, 1)$, $\Pr(X \geq \sqrt{\alpha n}) \leq e^{-\alpha n/2}$, the measure of S_α can be upper-bounded as $\gamma(S_\alpha) \leq e^{-\alpha n/2}$. Here the matrix $\mathbf{S}_\alpha$ is diagonal, with the first eigenvalue approximately equal to αn, and the remaining $(n-1)$ each equal to 1, so that its trace is $\mathrm{Tr}\,\mathbf{S}_\alpha \approx \alpha n + (n-1)$. For $y \sim \gamma$, the distribution of y_1^2 is χ^2 with expectation 1 and standard deviation $\sqrt{2}$; in particular with constant probability it is less than $3/4$. Conditioning on this event,

$$y^T \mathbf{S}_\alpha y \approx \alpha n y_1^2 + (y_2^2 + \cdots + y_n^2) \leq (3\alpha/4)n + (y_2^2 + \cdots + y_n^2)$$

which is less than $\mathrm{Tr}\,\mathbf{S}_\alpha - (\alpha/4)n$ with constant probability. Hence one should expect that in (4) it is necessary to allow the overlap $y^T \mathbf{S}_\alpha y$ to be moderately smaller than its expectation $\mathrm{Tr}\,\mathbf{S}_\alpha$, since this can hold even with constant probability. Moreover, it is not hard to show that for any $\beta > 0$ we have $\Pr(y_1^2 > \sqrt{2\beta n}) = \Omega(n^{-1/2} e^{-\beta n})$, and if this holds then the overlap $y^T \mathbf{S}_\alpha y$ is at least $\alpha\sqrt{2\beta}\, n^{3/2}$, which is much larger than $\mathrm{Tr}\,\mathbf{S}_\alpha = (\alpha + 1)\, n - 1$ for any $\beta = \omega(1/n)$: there cannot be any strong concentration in the direction opposite to the one we are claiming.

Before proving Lemma 7, we show that it implies Theorem 1.

[1] We will get back to this approximation later, but it is not hard to show that $\mathrm{Tr}\,\mathbf{S}$ is within a factor $\approx (1 \pm \sqrt{\delta})$ of n by using the bound given in Claim 4.

Proof of Theorem 1. Let $\eta > 0$ be given, and let $\mathbf{A} := \mathrm{E}_{x\sim\gamma_{|A}}\left[xx^T\right]$. Fix a $\delta > 0$ small enough so that both the following hold:

1. $|\mathrm{Tr}\,\mathbf{A} - n| \leq \eta\, n/4$. This is made possible by Claim 4.
2. The set of y for which $y^T\mathbf{A}y \leq \mathrm{Tr}\,\mathbf{A} - \eta n/4$ has measure less than $(\eta/4)e^{-\delta n}$. This can be obtained from Lemma 7.

Combining these two estimates, we obtain

$$\begin{aligned}\mathrm{E}_{y\sim\gamma_{|B}}\left[y^T\mathbf{A}y\right] &\geq \frac{1}{\gamma(B)}\left(\gamma(B) - (\eta/4)e^{-\delta n}\right)(\mathrm{Tr}\mathbf{A} - \eta n/4)\\ &\geq (1-\eta/4)(n - \eta\, n/2)\\ &\geq (1-\eta)\, n,\end{aligned}$$

which proves the theorem in light of (3), after noting that $\mathrm{E}_{(x,y)\sim\gamma}\left[\,(x\cdot y)^2\,\right] = n$. □

We turn to the proof of Lemma 7. Let $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq 0$ be the eigenvalues of $\mathbf{S}$, sorted in decreasing order. For any $y \in \mathbb{R}^n$ one can re-write

$$y^T\mathbf{S}y = \sum_i \lambda_i\, y_i^2,$$

where the y_i are y's coefficients in the eigenbasis of $\mathbf{S}$. Since the distribution γ is rotation-invariant, the y_i are distributed according to the standard normal distribution. However, as shown in the example of the cap $S_{2\delta} = \{x \in \mathbb{R}^n : x_1 \geq \sqrt{2\delta n}\}$ discussed above, some of the λ_i can be quite large: $S_{2\delta}$ has measure $\gamma(S) \approx e^{-\delta n}$, but the corresponding matrix $\mathbf{S}_{2\delta}$ has $\lambda_1 \approx 2\delta n$. Hence a direct use of Claim 4 would lead to a rather poor bound. Rather, we will use Claim 5. For this to be effective, we need to show that, while the largest eigenvalues of $\mathbf{S}$ can be quite large, its spectrum must still be relatively spread out. This is made precise in the following claim.

Claim 8. *For any $\delta > 0$, let $S \subseteq \mathbb{R}^n$ be of measure $\gamma(S) \geq e^{-\delta n}$, and let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of $\mathbf{S}$ sorted in decreasing order. Then for any $\alpha \geq \delta$ and all n large enough,*

$$\sum_{i=1}^{\alpha n} \lambda_i \leq (25e)\,\alpha\, n. \tag{5}$$

Proof. If $P_{\alpha n}$ is the projection on the span of the eigenvectors corresponding to the largest αn eigenvalues of $\mathbf{S}$, their sum is $\mathrm{Tr}(P_{\alpha n}\mathbf{S}) = \mathrm{E}_{x\sim\gamma_{|S}}\left[x_1^2 + \cdots + x_{\alpha n}^2\right]$, where the x_i are the coordinates of x in the eigenbasis of $\mathbf{S}$. For any $t \geq 0$, Claim 4 gives the bound

$$\Pr_{x\sim\gamma}\left(x_1^2 + \cdots + x_{\alpha n}^2 \geq (1+t)\,\alpha n\right) \leq 2e^{-\frac{\alpha n}{8e}\min\left(\frac{t^2}{4e},t\right)},$$

so that, letting $t' = t - 8e$ we have that for every $t' \geq 4e$,

$$\Pr_{x\sim\gamma_{|S}}\left(x_1^2 + \cdots + x_{\alpha n}^2 \geq (1+8e+t')\,\alpha n\right) \;\leq\; 2\,e^{-\frac{\alpha n}{8e}(t'+8e)}e^{\delta n} \;\leq\; 2\,e^{-\frac{\alpha n t'}{8e}},$$

where we used our assumption $\alpha \geq \delta$. Since for any non-negative random variable X, $\mathrm{E}[X] = \int_{t=0}^{\infty}\Pr(X \geq t)$, we get

$$\mathrm{E}_{x\sim\gamma_{|S}}\left[x_1^2 + \cdots + x_{\alpha n}^2 - (1+8e)\alpha\, n\right] \leq 16e + 4e\,\alpha n$$

which proves the claim. □

We finish by showing how Claim 8 implies Lemma 7.

Proof of Lemma 7. Let α be given, $\beta := \alpha/(100e)$, and let $y_i \sim N(0,1)$ be i.i.d. By Claim 5, using a crude bound $\operatorname{Tr}\mathbf{S} \le 2n$ (which follows from Claim 4 for all large enough n), we get that for any $t \ge 0$,

$$\Pr\Big(\sum_{i=1}^{n} \lambda_i y_i^2 \le \operatorname{Tr}\mathbf{S} - t - \sum_{i=1}^{2\beta n} \lambda_i\Big) \;\le\; 2\, e^{-\frac{\beta t}{8e} \min\left(\frac{\beta t}{4en}, 1\right)}. \tag{6}$$

By Claim 8, $\sum_{i=1}^{2\beta n} \lambda_i \le (25e)2\beta n = \alpha n/2$, provided the condition $2\beta \ge \delta$ is satisfied, which we can ensure by setting $c' = 50e$ in the statement of the lemma. Choosing $t = \alpha n/2$ in (6) finishes the proof. □

4 Application to communication complexity

In this section we explain how Theorem 1 leads to a lower bound on the communication complexity of the GHD problem. In fact, we will show a lower bound for its continuous analogue, the Gap-Inner-Product (GIP) problem, defined on $\mathbb{R}^n \times \mathbb{R}^n$ by

$$\mathrm{GIP}_{n,t,g}(x,y) = \begin{cases} 1 & \text{if } x \cdot y \ge t + g, \\ 0 & \text{if } x \cdot y \le t - g, \\ \star & \text{otherwise.} \end{cases}$$

For us, the parameters of interest (and arguably the most natural[2]) are $t, g = \Theta(\sqrt{n})$. A lower bound on GIP is easily seen to imply an equivalent lower bound for GHD (see e.g. Proposition 3 in [1] for a proof that the two problems have essentially the same randomized communication complexity).

The proof of the lower bound is based on a technique introduced in [2], and is closely related to the "partition bound" of [4]. For the reader's convenience we cite a "meta-theorem" from [2], which we will combine with the results of the previous section to re-prove the linear lower bound on the randomized communication complexity of the GIP problem first proved in [2], also through the following meta-theorem, but using a much more involved technical argument than ours.

Theorem 9 (Theorem 2.2 in [2])**.** *For all α_0, α_1, α_+, $\varepsilon > 0$ such that $\varepsilon < (\alpha_1 - \alpha_+)/(\alpha_0 + \alpha_1)$, there exist $\beta \in \mathbb{R}$ and $\varepsilon' > 0$ such that the following holds. Let $f : X \times Y \to \{0, 1, \star\}$ be a partial function. Let $A_0 = f^{-1}(0)$ and $A_1 = f^{-1}(1)$. Suppose that there exist distributions μ_0, μ_1, μ_+ on $X \times Y$, and a real number $m > 0$ such that*

1. *for $i \in \{0,1\}$, μ_i is mostly supported on A_i, i.e., $\mu_i(A_i) \ge 1 - \varepsilon$, and*
2. *the following holds for all rectangles $R \subseteq X \times Y$:*

$$\alpha_1 \mu_1(R) - \alpha_+ \mu_+(R) \le \alpha_0 \mu_0(R) + 2^{-m}.$$

Then $R_{\varepsilon'}(f) \ge m + \beta$.

[2] Note that two random vectors taken according to γ have expected inner product 0, with a standard deviation of $\sqrt{n}$.

We will apply this theorem to $f = GIP_{n,t,g}$, with parameters $t = -(d+c)\sqrt{n}/2$ and $g = (d-c)\sqrt{n}/2$, where $c = 0.5$ and $d = 0.6$ (note that Lemmas 4.1 and 4.2 in [2] show that the exact choice of t and g does not affect the randomized communication complexity too much, as long as say $t, g = \Theta(\sqrt{n})$). We instantiate μ_1 as the $2n$-dimensional standard Gaussian distribution γ. For μ_0 we choose the distribution with density

$$\mu_0(x, y) = \begin{cases} 0 & \text{if } x \cdot y > 0, \\ \frac{2}{n(2\pi)^n}(x \cdot y)^2 e^{-\|x\|^2/2} e^{-\|y\|^2/2} & \text{otherwise,} \end{cases}$$

while μ_+ is chosen with density $\mu_+(x, y) = \mu_0(-x, y)$. All these distributions are invariant under arbitrary simultaneous rotations of x and y; their densities are represented on Figure 1 for a fixed $y = y_0$, as a function of $x = t\, y_0$, $t \in \mathbb{R}$.

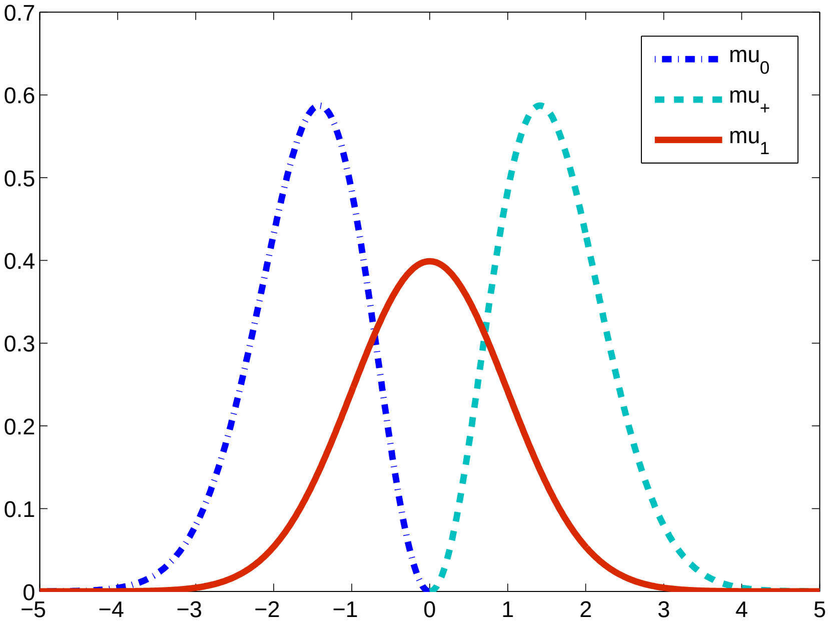

Figure 1: The one-dimensional densities obtained from μ_0 (dotted, left), μ_+ (dotted, right) and μ_1 (plain) by conditioning on $y = y_0$ and projecting x on $\mathbb{R}y_0$.

We first verify Condition 1 of Theorem 9, which intuitively states that μ_0 should be mostly supported on 0-inputs, and μ_1 on 1-inputs, as one can observe graphically in Figure 1. For this we will use that for large n, for $x, y \in \mathbb{R}^n$ distributed independently according to γ, the inner product $x \cdot y$ is essentially distributed as a Gaussian with standard deviation $\sqrt{n}$. This follows from the Berry-Esseen theorem (Fact 6) applied to

$X_i = x_i \cdot y_i$, which are i.i.d. with variance $\sigma^2 = 1$ and third moment $\rho = 2\sqrt{2/\pi}$. This lets us write

$$\begin{aligned}\mu_1(A_1) &= \Pr_{(x,y)\sim\gamma}\big(x\cdot y > -c\sqrt{n}\big)\\ &\geq \frac{1}{\sqrt{2\pi}}\int_{-c}^{\infty} e^{-t^2/2}\mathrm{dt} - \Omega\big(\frac{1}{\sqrt{n}}\big)\\ &\geq \frac{1}{2} + \frac{c}{\sqrt{2\pi}}e^{-c^2/2} - \Omega\big(\frac{1}{\sqrt{n}}\big) \;\geq\; 0.76\end{aligned}$$

for large enough n. Similarly, we compute

$$\begin{aligned}\mu_0(A_0) &= 1 - \Pr_{(x,y)\sim\mu_0}\big(x\cdot y > -d\sqrt{n}\big)\\ &= 1 - \frac{2}{n(2\pi)^n}\iint_{-d\sqrt{n}<x\cdot y\leq 0}(x\cdot y)^2 e^{-\|x\|^2/2}e^{-\|y\|^2/2}\,\mathrm{dx\,dy}\\ &\geq 1 - \frac{2d^2}{(2\pi)^n}\iint_{-d\sqrt{n}<x\cdot y\leq 0} e^{-\|x\|^2/2}e^{-\|y\|^2/2}\,\mathrm{dx\,dy}\\ &\geq 1 - 2d^2\frac{1}{\sqrt{2\pi}}\int_{-d}^{0} e^{-t^2/2}\,\mathrm{dt} - \Omega\big(\frac{1}{\sqrt{n}}\big)\\ &= 1 - 2d^2\frac{1}{\sqrt{2\pi}}\int_{0}^{d} e^{-t^2/2}\,\mathrm{dt} - \Omega\big(\frac{1}{\sqrt{n}}\big) \;\geq 0.78\end{aligned}$$

for large enough n, so by setting $\varepsilon := 0.3$ we make sure that Condition 1. is satisfied. In order to verify Condition 2., observe that for any rectangle R,

$$(\mu_0+\mu_+)(R) = \frac{2}{n(2\pi)^n}\iint_{(x,y)\in R}(x\cdot y)^2 e^{-\|x\|^2/2}e^{-\|y\|^2/2}\,\mathrm{dx\,dy} = \frac{2}{n}\gamma(R)\mathrm{E}_{(x,y)\sim\gamma_{|R}}\big[\,(x\cdot y)^2\,\big]\,,$$

so that by setting $\eta = 0.05$, Theorem 1 implies the existence of a $\delta > 0$ such that that $(\mu_0(R)+\mu_+(R))/2 \geq (1-\eta)\gamma(R)$, as long as $\gamma(R) \geq e^{-\delta n}$. Choosing $\alpha_0 = \alpha_+ = 1/2$, $\alpha_1 = 0.95$ and $m = (\ln 2)\,\delta n$, Condition 2. reads

$$\frac{\mu_0(R)+\mu_+(R)}{2} \geq 0.95\,\gamma(R) - e^{-\delta n},$$

which is trivially satisfied by any R with $\gamma(R) < e^{-\delta n}$, and for all R such that $\gamma(R) \geq e^{-\delta n}$ by the previous arguments. Note also that with our choice of coefficients α the inequality on ε is satisfied.

As a consequence, Theorem 9 directly implies the existence of $\varepsilon' > 0$ and $\beta \in \mathbb{R}$ such that

$$R_{\varepsilon'}(GIP_{n,-.55\sqrt{n},.05\sqrt{n}}) \geq (\ln 2)\,\delta n + \beta.$$

Acknowledgments. I thank Oded Regev and Ronald de Wolf for helpful discussions, and Anindya De, Andrew Drucker and Oded Regev for comments that greatly improved the presentation of this manuscript.

References

[1] J. Brody, A. Chakrabarti, O. Regev, T. Vidick, and R. De Wolf. Better gap-hamming lower bounds via better round elimination. *In Proc. of 13th APPROX-RANDOM*, pp. 476–489, 2010

[2] A. Chakrabarti and O. Regev. An Optimal Lower Bound on the Communication Complexity of Gap-Hamming-Distance. *To appear in STOC*, 2011.

[3] W. Feller. An introduction to probability theory and its applications, volume II. John Wiley & Sons, Inc., New York-London-Sydney, 1971.

[4] R. Jain and H. Klauck. The Partition Bound for Classical Communication Complexity and Query Complexity. *In Proc. of 25th IEEE CCC*, pp.247–258, 2010.

[5] E. Kushilevitz and N. Nisan. Communication complexity. *Cambridge University Press*, 1997.

[6] R. Vershynin. Introduction to the non-asymptotic analysis of random matrices. Lecture notes available on the author's webpage, 2010.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de