CHICAGO JOURNAL OF THEORETICAL COMPUTER SCIENCE 2012, Article 01, pages 1–12
http://cjtcs.cs.uchicago.edu/

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*

Received: October 8, 2011; revised: October 25, 2011; published: July 1, 2012.

Abstract: Given two sets $A, B \subseteq \mathbb{R}^n$, a measure of their 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 Gaussian 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.

Key words and phrases: Gaussian concentration inequality, isoperimetry, Gap-Hamming Distance

1 Introduction

Let A, B be two non-empty measurable subsets of $\mathbb{R}^n$ equipped with the n-dimensional Gaussian measure γ. Denote by $\gamma_{|A\times B}$ the probability measure corresponding to the normalized restriction of $\gamma \times \gamma$ to $A \times B$, and let

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

*Supported by the National Science Foundation under Grant No. 0844626.

DOI: 10.4086/cjtcs.2012.001

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

A vector $x \in \mathbb{R}^n$ distributed according to γ has squared norm tightly concentrated around n (precisely, it follows a $\chi^2(n)$ distribution, with expectation n, variance $\sqrt{2n}$ and sub-exponential tails). By rotation invariance of γ, for any fixed vector $y \in \mathbb{R}^n$ the inner-product $x \cdot y$ is distributed as a centered Gaussian with variance $\|y\|^2$. Hence for any non-empty measurable set A it holds that $\nu(A,\mathbb{R}^n) = n = \nu(\mathbb{R}^n,\mathbb{R}^n)$.

We study the following question: How much smaller than the average value $\nu(\mathbb{R}^n,\mathbb{R}^n)$ can $\nu(A,B)$ be for arbitrary sets A,B of given measure? If we allow both sets A, B to be arbitrarily small then ν can also be arbitrarily small: take $A = \{x\}$, $B = \{y\}$, with x,y orthogonal, as the limiting example. If we allow A to be arbitrarily small, but constrain B to have measure $\gamma(B) \geq t$, where t is a small constant, then $\nu(A,B)$ can still be quite small. Indeed, for a fixed vector x (of norm $\sqrt{n}$) choose B as the fattened equator $B = \{y \in \mathbb{R}^n : -t\sqrt{\pi n/2} \leq y \cdot x \leq t\sqrt{\pi n/2}\}$, of measure $\gamma(B) \approx t$. Then for A an infinitesimal ball centered at x we get $\nu(A,B) \leq t^2 n$, an arbitrarily small fraction of $\nu(\mathbb{R}^n,\mathbb{R}^n) = n$.

In this note we show that in case both A and B are restricted to not being too small, then $\nu(A,B)$ cannot be much lower than $\nu(\mathbb{R}^n,\mathbb{R}^n)$. More precisely we show the following:

Theorem 1.1. *For any $\eta > 0$, there exists[1] 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*

$$\nu(A,B) \geq (1-\eta)\,\nu(\mathbb{R}^n,\mathbb{R}^n) = (1-\eta)\,n. \tag{1.1}$$

We remark that an analogue of Theorem 1.1 can also be proved for subsets of the unit sphere $\mathbb{S}^{n-1} \subset \mathbb{R}^n$, with the Haar measure playing the role of the Gaussian measure: indeed, our proof of Theorem 1.1 relies on concentration properties of the n-dimensional Gaussian measure which also hold for the Haar measure on the sphere.

Choosing $A = B = \{x \in \mathbb{R}^n, \|x\|^2 \leq (1-\delta)n\}$, of measure at least $e^{-c\delta n}$ for some fixed $c > 0$, shows that the dependence of δ on η in Theorem 1.1 should be at least linear. Our proof only achieves a weaker dependence $\delta = \Omega(\eta^4)$.

Note that one may not hope for such a strong inequality as the one proven in Theorem 1.1, but in the opposite direction: the half-spaces $A = B = \{x \in \mathbb{R}^n, x_1 \geq \sqrt{2\delta n}\}$ have measure approximately $e^{-\delta n}$ but correlation $\nu(A,B) = \Omega(\delta^2 n^2)$.

Application: the communication complexity of the Gap-Hamming-Distance problem. The motivation for, and main application of, Theorem 1.1 is to give a new, simpler proof of a recent breakthrough result by Chakrabarti and Regev [2], who proved a linear lower bound on the bounded-error randomized communication complexity of the Gap-Hamming-Distance (GHD) problem. In this problem, Alice is given an n-bit string x, and Bob an n-bit string y such that either $\Delta(x,y) \geq n/2 + \sqrt{n}$ or $\Delta(x,y) \leq n/2 - \sqrt{n}$, where $\Delta(x,y)$ denotes the Hamming distance. The goal is to decide which holds. Proving an $\Omega(n)$ lower

[1] An explicit estimate of δ (as a function of η) can easily be extracted from our proof, although in general we do not attempt to optimize the exact numerical constants that we give.

bound for this problem was a long-standing open problem in communication complexity (see [2] for a detailed history of the problem).

Chakrabarti and Regev's 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 1.1, except that it applies to the cosh function, instead of the square function. More precisely, if one defines

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

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 $0 \le \alpha \le c/\sqrt{n}$ and A, B of measure at least $e^{-\delta n}$,

$$\tilde{\nu}_\alpha(A,B) \;\ge\; (1-\eta)\,\tilde{\nu}_\alpha(\mathbb{R}^n,\mathbb{R}^n). \tag{1.2}$$

Given that the cosh function has a quadratic behavior around 0, $\cosh x = \frac{x^2}{2} + O(x^4)$, our theorem may not be so surprising once one knows of (1.2). However, for large values of $\alpha\, x\cdot y$ the behavior of the two functions, x^2 and $\cosh x$, is different enough that we do not see how one could deduce Theorem 1.1 from (1.2) or vice-versa.

The proof of (1.2) 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 (1.2) 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.

Related work. After the completion of our work, Sherstov [6] provided yet another proof of Chakrabarti and Regev's lower bound for GHD. His proof is shorter and combinatorial, while the one in [2], as well as ours, uses geometric arguments and concentration of measure in $(\mathbb{R}^n,\gamma)$. The main innovation of Sherstov's proof is to consider a problem equivalent to GHD, called *gap orthogonality*, for which a linear lower bound can be proved using the corruption method [8], while the proof in [2] goes through the more powerful but also more involved partition bound of Jain and Klauck [4]. We note that key to both Sherstov's and Chakrabarti and Regev's proofs of an anti-concentration result similar to our Theorem 1.1 is a technical argument showing that any large enough set A must contain a linear number of *almost-orthogonal* vectors. We completely avoid that step and instead work directly with a matrix representation of the set A (cf. Section 3 for a definition).

Organization. We review useful concentration bounds in Section 2. Our main result, Theorem 1.1, is further discussed and proved in Section 3. The application to the communication complexity of GHD is detailed in Section 4.

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 $S \subseteq \mathbb{R}^n$ is measurable with positive measure, $\gamma_{|S}$ denotes the normalized restriction of γ to S: for any measurable A, $\gamma_{|S}(A) = \gamma(A \cap S)/\gamma(S)$.

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

Fact 2.1 (Gaussian tail bound)**.** *Let X be a standard normal random variable. Then for every $t \geq 0$,*

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

Proof. Bound the upper tail as

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

A similar bound holds for the lower tail. □

Fact 2.2 (Bernstein's inequality, see, e.g., Prop. 16 in [7])**.** *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 $a = (a_1, \ldots, a_N) \in \mathbb{R}^N$ and $t \geq 0$, we have*

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

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

Claim 2.3 (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 $a = (a_1, \ldots, a_N) \in \mathbb{R}^N$ and $t \geq 0$,*

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

Proof. By Fact 2.1, 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 2.2 with $K = 2$, which leads to the claimed bound. □

The bound in Claim 2.3 becomes very weak as soon as even one of the coefficients a_i is large compared to the others. 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 2.4. *Let $N \in \mathbb{N}$, let $X_1, \dots, X_N$ be i.i.d. standard normal random variables, $a_1 \geq \cdots \geq a_N \geq 0$ non-negative reals sorted in non-increasing order, and $M = \sum_{i=1}^N a_i$. Then for every integer $1 \leq k \leq N$ and $t \geq 0$,*

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

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

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

Hence applying Claim 2.3 to $X_{k+1}, \dots, X_N$ yields that for every $t \geq 0$,

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

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

We will also use the Berry-Esseen theorem.

Fact 2.5 (Berry-Esseen Theorem, see, e.g., [3], Chapter XVI)**.** *Let $X_1, \dots, X_N$ 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 1.1 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 non-empty measurable subset of $\mathbb{R}^n$. Given such a set, it will be convenient to work with the positive semidefinite 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

$$\forall y \in \mathbb{R}^n \qquad y^T\mathbf{S}y = \mathrm{E}_{x\sim\gamma_S}\left[y^T xx^T y\right] = \mathrm{E}_{x\sim\gamma_S}\left[(x\cdot y)^2\right]. \tag{3.1}$$

As we will see, (3.1) lets us relate the concentration properties of $\nu(\{y\}, S)$, for $y \sim \gamma$, to the eigenvalues of $\mathbf{S}$. The following simple claim will be useful.

Claim 3.1. *Let $0 < \delta < 1/2$, and S a measurable subset of $\mathbb{R}^n$ such that $\gamma(S) \geq e^{-\delta n}$. Then for all large enough n it holds that*

$$\left|\mathrm{Tr}\,\mathbf{S} - n\right| \leq 50e\sqrt{\delta}\,n.$$

Proof. For any measurable set S,

$$\begin{aligned}\mathrm{Tr}\,\mathbf{S} &= \mathrm{E}_{y\sim\gamma}\left[y^T\mathbf{S}y\right] \\ &= \mathrm{E}_{x\sim\gamma_S, y\sim\gamma}\left[(x\cdot y)^2\right] \\ &= \mathrm{E}_{x\sim\gamma_S, y_1\sim\gamma}\left[\|x\|^2 y_1^2\right] \\ &= \mathrm{E}_{x\sim\gamma_S}\left[\|x\|^2\right],\end{aligned}$$

where the third equality follows from the rotation invariance of γ, and the last uses independence of x and y. Clearly, the set S of fixed measure $e^{-\delta n}$ which maximizes $|\mathrm{Tr}\mathbf{S} - n|$ is then either $S = \{x \in \mathbb{R}^n,\ \|x\|^2 > n + t\}$ or $S' = \{x \in \mathbb{R}^n,\ \|x\|^2 < n - t'\}$, where t (resp. t') is chosen so that $\gamma(S) = e^{-\delta n}$ (resp. $\gamma(S') = e^{-\delta n}$). By Claim 2.3, for S or S' to have measure at least $e^{-\delta n}$ it is necessary that $t, t' \leq t_0 = 6e\sqrt{\delta}\,n$. Using

$$\mathrm{E}[X] \leq \alpha + \int_{u=\alpha}^{\infty} \Pr(X \geq u)\mathrm{du}$$

for any non-negative random variable X and non-negative α, we can bound

$$\begin{aligned}\left|\mathrm{Tr}\,\mathbf{S} - n\right| &\leq \mathrm{E}_{x\sim\gamma_S}\left[\left|\|x\|^2 - n\right|\right] \\ &\leq 8t_0 + e^{\delta n}\int_{u=8t_0}^{\infty} \Pr_{x\sim\gamma}\left(\left|\|x\|^2 - n\right| > u\right)\mathrm{du} \\ &\leq 8t_0 + e^{\delta n}\int_{u=8t_0}^{4en} 2e^{-u^2/(32e^2n)}\mathrm{du} + e^{\delta n}\int_{u=4en}^{\infty} 2e^{-u/(8e)}\mathrm{du} \\ &\leq 8t_0 + 2e^{\delta n}\int_{u=2\sqrt{\delta n}}^{\infty} e^{-u^2/2}\mathrm{du} + \frac{1}{4e}e^{-(1-2\delta)n/2} \\ &\leq 8t_0 + 1 + \sqrt{2\pi}e^{-\delta n} \\ &\leq 8t_0 + 2,\end{aligned}$$

where the 3rd inequality uses the bound from Claim 2.3, the 5th inequality uses the Gaussian tail bound proved in Fact 2.1, and the 5th and 6th inequalities hold for all large enough n. Finally, the same bound holds for $\left|\mathrm{Tr}\,\mathbf{S}' - n\right|$. □

We show the following concentration bound.

Lemma 3.2. *There exists a constant $c > 0$ such that the following holds. Let $\delta > 0$ and $S \subseteq \mathbb{R}^n$ a non-empty measurable set such that $\gamma(S) \geq e^{-\delta n}$. Then for all $\alpha > (50e)\,\delta$ and all large enough n,*

$$\Pr_{y\sim\gamma}\left(y^T\mathbf{S}y \leq \mathrm{Tr}\,\mathbf{S} - \alpha n\right) \leq e^{-c\alpha^4 n}. \tag{3.2}$$

As shown in Claim 3.1 above, if $\gamma(S) \geq e^{-\delta n}$ then $\mathrm{Tr}\mathbf{S}$ is within a factor $\approx (1 \pm O(\sqrt{\delta}))$ of n, so that for small α the αn factor in (3.2) corresponds to a small fraction of $\mathrm{Tr}\mathbf{S}$.

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 $\delta > 0$ (think of δ as a small constant), and consider the halfspace $S_\delta = \{x \in \mathbb{R}^n : x_1 \geq \sqrt{\delta n}\}$, which has measure $\gamma(S_\delta) \approx e^{-\delta n/2}$. By definition, the (i,j)-th coefficient of the matrix $\mathbf{S}_\delta$ associated to S_δ takes the value

$$\left(\mathbf{S}_\delta\right)_{i,j} = \mathrm{E}_{x\sim\gamma_{|S_\delta}}\left[x_i x_j\right] = \begin{cases} \mathrm{E}_{x\sim\gamma_{|S_\delta}}\left[x_i^2\right] & \text{if } i = j, \\ \left(\mathrm{E}_{x\sim\gamma_{|S_\delta}}\left[x_i\right]\right)\left(\mathrm{E}_{x\sim\gamma_{|S_\delta}}\left[x_j\right]\right) = 0 & \text{if } i \neq j, \end{cases}$$

since $\mathrm{E}_{x\sim\gamma_{|S_\delta}}\left[x_i\right] = 0$ whenever $i > 1$. Hence $\mathbf{S}_\delta$ is diagonal, with first diagonal entry equal to $\mathrm{E}_{x\sim\gamma_{|S_\delta}}\left[x_1^2\right] \approx \delta n$, and the remaining $(n-1)$ each equal to 1. In particular the trace of $\mathbf{S}_\delta$ is

$$\mathrm{Tr}\mathbf{S}_\delta \approx \delta n + (n-1).$$

Now take a random $y \in \mathbb{R}^n$, distributed according to γ. The distribution of y_1^2 is standard χ^2, which has constant probability of being less than $1/2$. Conditioning on this event,

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

which is less than $\mathrm{Tr}\mathbf{S}_\delta - (\delta/2)n$ with constant probability. This shows that in (3.2) it is necessary to allow the overlap $y^T\mathbf{S}_\delta y$ to be moderately smaller than its expectation $\mathrm{Tr}\mathbf{S}_\delta$, since this can hold even with constant probability.

To show that one cannot hope to prove a bound similar to (3.2) for the upper tail of $y^T\mathbf{S}_\delta y$, observe that if $y_1 \sim N(0,1)$ then $\Pr(y_1 > n^{1/4}) \approx \Omega(n^{-1/2}e^{-\sqrt{n}/2})$. In case this holds, the overlap $y^T\mathbf{S}_\delta y$ is at least $\delta n^{5/4}$, which is much larger than $\mathrm{Tr}\mathbf{S}_\delta \approx (\delta+1)n$.

Before proving Lemma 3.2, we show that it implies Theorem 1.1.

Proof of Theorem 1.1. 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 3.1.
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 3.2.

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 (3.1). □

We turn to the proof of Lemma 3.2. Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$ be the eigenvalues of $\mathbf{S}$, sorted in non-increasing order. For any $y \in \mathbb{R}^n$ one can 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 γ is rotation-invariant, the y_i are independently distributed according to the standard normal distribution. However, as shown in the example of the halfspace $S_{2\delta} = \{x \in \mathbb{R}^n : x_1 \geq \sqrt{2\delta n}\}$ 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 2.3 would lead to a rather poor bound. Rather, we will use Claim 2.4. 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 3.3. *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 non-increasing order. Let $\lceil \delta n \rceil \leq k \leq n$ be an integer. Then for all n large enough,*

$$\sum_{i=1}^{k} \lambda_i \leq (25e)\, k. \tag{3.3}$$

Proof. Let $\lambda_1 \geq \cdots \lambda_n$ be the eigenvalues of $\mathbf{S}$ sorted in non-increasing order, and $u_1, \ldots, u_n$ the corresponding eigenvectors. For $i \in [n]$ and $x \in \mathbb{R}^n$, let $x_i = x \cdot u_i$ be the i-th coordinate of x in the basis given by the u_i. By definition

$$\sum_{i=1}^{k} \lambda_i = \sum_{i=1}^{k} u_i^T \mathbf{S} u_i = \mathrm{E}_{x \sim \gamma_{|S}} \Big[\sum_{i=1}^{k} (x \cdot u_i)^2 \Big] = \mathrm{E}_{x \sim \gamma_{|S}} \Big[\sum_{i=1}^{k} x_i^2 \Big].$$

For any $\beta \geq 0$, Claim 2.3 gives the bound

$$\Pr_{x \sim \gamma} \left(x_1^2 + \cdots + x_k^2 \geq (1+\beta)\, k \right) \leq 2 e^{-\frac{k}{8e} \min\left(\frac{\beta^2}{4e}, \beta\right)},$$

so that, letting $\beta' = \beta - 8e$ we have that for every $\beta' \geq -4e$,

$$\Pr_{x \sim \gamma_{|S}} \left(x_1^2 + \cdots + x_k^2 \geq (1 + 8e + \beta')\, k \right) \leq 2\, e^{-\frac{k}{8e}(\beta' + 8e)} e^{\delta n} \leq 2\, e^{-\frac{k\beta'}{8e}},$$

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

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

which proves the claim. □

We finish by showing how Claim 3.3 implies Lemma 3.2.

Proof of Lemma 3.2. Let α be given, $\beta := \alpha/(100e)$, and let $y_i \sim N(0,1)$ be i.i.d. By Claim 2.4, using a crude bound $\mathrm{Tr}\mathbf{S} \leq 2n$ (which follows from Claim 3.1 for all large enough n), we get that for any $t \geq 0$,

$$\Pr\Big(\sum_{i=1}^{n} \lambda_i y_i^2 \leq \mathrm{Tr}\mathbf{S} - t - \sum_{i=1}^{2\beta n} \lambda_i\Big) \leq 2\,e^{-\frac{\beta t}{8e}\min\left(\frac{\beta t}{8en},1\right)}. \tag{3.4}$$

By Claim 3.3, $\sum_{i=1}^{2\beta n} \lambda_i \leq (25e)2\beta n = \alpha n/2$, provided the condition $2\beta \geq \delta$ is satisfied, which follows from our assumption that $\alpha > (50e)\delta$. Choosing $t = \alpha n/2$ in (3.4) finishes the proof. □

4 Application to communication complexity

In this section we explain how Theorem 1.1 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

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

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 4.1 (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., $\mu_i(A_i) \geq 1 - \varepsilon$, and*
2. *the following holds for all rectangles $R \subseteq X \times Y$:*

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

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

[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 $x = t\frac{y_0}{\|y_0\|}$, $t \in \mathbb{R}$.

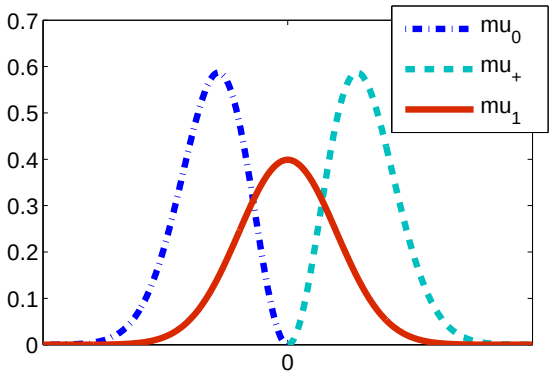

Figure 1: The one-dimensional densities obtained from μ_0 (dotted, left), μ_+ (dotted, right) and μ_1 (plain) by conditioning on $y = y_0$ and projecting x on $\mathbb{R}y_0$.

We first verify Condition 1 of Theorem 4.1, 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 2.5) 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

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

for large enough n. Similarly, we compute

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

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,

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

so that by setting $\eta = 0.05$, Theorem 1.1 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 4.1 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, Oded Regev and an anonymous referee 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. 13th APPROX-RANDOM*, pp. 476–489, 2010. 9

[2] A. CHAKRABARTI AND O. REGEV: An optimal lower bound on the communication complexity of gap-Hamming-distance. In *Proc. 43rd ACM STOC*, pp. 51–60, 2011. 2, 3, 9, 10

[3] W. FELLER: *An Introduction to Probability Theory and its Applications, Volume II*. John Wiley & Sons, Inc., 1971. 5

[4] R. JAIN AND H. KLAUCK: The partition bound for classical communication complexity and query complexity. In *Proc. 25th IEEE CCC*, pp. 247–258, 2010. 3, 9

[5] E. KUSHILEVITZ AND N. NISAN: *Communication Complexity*. Cambridge University Press, 1997. 5

[6] A. SHERSTOV: The communication complexity of gap Hamming distance. *Theory of Computing*, 8(1):197–208, 2012. 3

[7] R. VERSHYNIN: *Introduction to the non-asymptotic analysis of random matrices*. Cambridge University Press, 2012. Chapter 5 of Compressed Sensing, Theory and Applications. 4

[8] A. C.-C. YAO: Lower bounds by probabilistic arguments. In *Proc. 24th IEEE FOCS*, pp. 420–428, 1983. 3

AUTHOR

Thomas Vidick
Postdoctoral associate
Massachussetts Institute of Technology, Cambridge, MA
vidick@csail.mit.edu
http://people.csail.mit.edu/~vidick