COMMUNICATION COMPLEXITY AND THE LOG-RANK CONJECTURE

JAKE WELLENS

In the setting of (two-party) communication complexity, we want to compute the value of a known function $f : X \times Y \rightarrow \{0, 1\}$ on a particular input $(x, y)$, which has been split up between two parties, using as little communication as possible between the parties, which we’ll call Alice and Bob.

Techniques and results from this area have found applications in (i) proving time-space lower bounds for Turing machines, (ii) bounds on other complexity measures for boolean functions like decision tree and circuit depth, monotone circuit depth, (iii) area/time lower bounds for VLSI chips, (iv) extended formulations of LPs.

**Def:** A deterministic protocol computing a function $f(x, y)$ is a binary tree $T$ whose internal nodes specify which party speaks and the value of the bit they communicate, as a function of their input. The leaves of the tree are labelled with 0 or 1, in such a way that if Alice and Bob’s path through the tree given inputs $(x, y)$ ends up in that leaf, the label on the leaf is $f(x, y)$. The cost of such a protocol is the length of the longest root-leaf path in $T$. We denote by $D(f)$ the minimum cost among all deterministic protocols computing $f$.

$D(f) \leq n$ for all $f$; indeed, Alice can just send Bob $x$ and he can compute $f(x, y)$.

$\text{MED}(x, y)$ given sets $x, y \subset [n]$, find the median of the multiset $x \cup y$: assuming both lists have the same length (we can achieve this by padding), Alice and Bob can each output a bit indicating which side of $n/2$ their set’s median lies on. If these bits match, they know that $\text{MED}(x \cup y)$ lives in a set of size $n/2$; if the bits differ, they can each throw away half of their list (the half on the correct side of their median), so after $O(\log n)$ steps, either they know the median or each of their lists has only one element, in which case they can wrap up with another $O(\log n)$ bits of communication.

Polylogarithmic communication complexity is the gold standard, in fact:

$$P^{cc} := \{ f : D(f) = O(\text{polylog}(n)) \}$$ (1)
What do you think is the communication complexity of the following functions? \(\text{EQ}_n, \text{DISJ}_n, \text{GT}_n\)?

The number of leaves of in \(T\) is \(\leq 2^c(T)\). The set of inputs which end up in any one leaf is a rectangle – indeed, if \((x, y)\) and \((x', y')\) induce all of the same communication, then so must \((x, y')\), since Alice can’t tell the difference from \((x, y)\) and Bob can’t tell the difference from \((x', y')\). These two observations can be used to prove

\[
D(\text{EQ}_n) = D(\text{DISJ}_n) = D(\text{GT}_n) = n\tag{2}
\]

since for \(\text{EQ}_n\), no two inputs \((x, x)\) and \((x', x')\) can end up in the same leaf, otherwise \((x, x')\) would also end up in that 1-leaf. Similarly, \((A, A^c)\) and \((B, B^c)\) cannot end up in the same leaf, otherwise \((A, B^c)\) and \((A^c, B)\) would also have to be disjoint, which implies \(A = B\). Thus, any protocol \(T\) computing either function must have \(2^n\) leaves.

Fix a graph \(G\) with \(n\) vertices and let \(\text{CIS}_G(C, I)\) be the partial function which takes in the characteristic vectors of a clique \(C\) and an independent set \(I\) in \(G\) and outputs 1 iff \(|C \cap I| \neq 0\). I claim \(D(\text{CIS}_G) = O(\log^2 n)\):

Maintain and update a set \(G_i\) of active nodes in \(G\), starting with \(G_0 = G\), let \(n_i = |G_i|\):

Alice looks for a vertex \(v\) in \(C\) with degree \(< n_i/2\), sends index to Bob if found (log \(n\) bits)

Bob looks for a vertex \(v\) in \(I\) with degree \(\geq n_i/2\), sends index to Alice if found

If both fail, \(C \cap I = \emptyset\). If either passes, can remove half of \(G_i\)’s vertices, since either they cannot be in \(C\) or cannot be in \(I\). Terminates in \(O(\log n)\) steps, with a total of \(O(\log^2 n)\) communication.

We can also have non-deterministic protocols: in this setting, Alice and Bob get access to a shared non-deterministic string \(z\), and decide (after some communication) whether to accept or reject. As per usual with non-determinism, if \(f(x, y) = 1\), we require that \(\exists z\) which causes the protocol to accept; if \(f(x, y) = 0\), we require that \(\forall z\) the protocol rejects. The cost of the protocol is \(|z| + (\text{communication})\), although if we include the communication transcript \(\pi\) as a part of \(z\), Bob and Alice can just check to see if that’s what they would’ve said, and so we don’t need any more than 1 bit of communication in this model. Let \(N^1(f)\) denote the minimal cost of a non-deterministic protocol computing \(f\) and define \(\text{NP}^{\text{cc}} := \{ f : N^1(f) = O(\text{polylog} n) \}\).

We already know \(\text{P}^{\text{cc}} \neq \text{NP}^{\text{cc}}\): indeed, we know that the function \(\text{EQ}_n \notin \text{P}^{\text{cc}}\), yet it is certainly in \(\text{co} - \text{NP}^{\text{cc}}\), since if \(x \neq y\), the proof \(z\) can just be an index in which they differ.
Let $M_f$ denote the $2^n \times 2^n$ matrix representation of $f$. Let $C^1(f)$ (cover number) denote the size of the smallest cover of $M_f$'s 1-entries by monochromatic rectangles.

Claim: $N^1(f) = \log C^1(f)$.

Proof: Given a fixed optimal cover of the 1's by rectangles, the proof $z$ can be the index of the 1-rectangle containing $(x, y)$, which takes $\log C^1(f)$ bits to describe. Conversely, given an optimal protocol, for each of the $2^{N^1(f)}$ possible strings $z$ we have a (possibly empty) 1-rectangle $A_z \times B_z$ of inputs for which the protocol accepts – these form a cover of the 1-entries of $M_f$. □

The leaves of any protocol for $f$ can be thought of as a partition of $M_f$ into monochromatic rectangles. Let $\chi_i(f)$ (partition number) denote the size of the optimal partition of $f^{-1}(i)$ into monochromatic rectangles, and $\chi(f) := \chi_0(f) + \chi_1(f)$:

Claim: $\log(\text{rank}_{\mathbb{F}_2}(M_f)) \leq \log(\chi(f)) \leq D(f) \leq \text{rank}_{\mathbb{F}_2}(M_f) \leq \text{rank}_{\mathbb{R}}(M_f)$.

Proof: Note that $M_f$ can be written as the sum of $\chi_1(f)$-many rank-one matrices. By the sub-additivity of rank, we then have $\text{rank}(M_f) \leq \chi_1(f)$, and hence $\log(\text{rank}(M_f)) \leq \log(\chi_1(f)) \leq \log(\chi(f))$. Since the leaves in any protocol give monochromatic 0 and 1 covers of $M_f$, we have $D(f) \geq \log(\chi_0(f) + \chi_1(f))$. For the remaining inequality, observe that if we fix an $\mathbb{F}_2$-basis for the rowspace of $M_f$, then Alice can tell Bob which row she’s in by simply sending him the coordinates of her row in that basis. □

Corollary: The inner-product function $\text{IP}_n$ has deterministic communication complexity $n$, since $M_{\text{IP}_n}$ is essentially a Hadamard matrix with full rank.

Corollary: The partition number $\chi_1$ and the cover number $C^1$ can differ exponentially. Indeed, $C^1(\neg \text{EQ}_n) = 2^{N^0(\text{EQ}_n)} = n$ while $\chi_1(\neg \text{EQ}_n) \geq \text{rank}(J - I) = 2^n$. However, this is the most they can differ, since

$$\log(\chi_1(f)) \leq D(f) \leq 2^{N^1(f)} = C^1(f).$$

Theorem [Aho-Ullman-Yannakakis ’83]: $D(f) = O(N^1(f)N^0(f))$. In particular, $\text{P}^{cc} = \text{NP}^{cc} \cap \text{co} - \text{NP}^{cc}$.

Proof: Let $L_0$ and $L_1$ be optimal covers by 0 and 1 rectangles, respectively. Alice looks for an $R \in L_1$ such that $x \in \pi_A(R)$, and $\pi_A(R) \cap \pi_A(R') = \emptyset$ for at least $1/2$ of all active $R' \in L_0$. If she can find one, then she uses $\log(|L_0|)$ to tell Bob which one it is, and the set of active 0-rectangles is cut in half. If she can’t, then
Bob does the analogous thing: he tries to find an $R \in L_1$ such that $y \in \pi_B(R)$ and $\pi_B(R) \cap \pi_B(R') = \emptyset$ for at least $1/2$ of all active $R' \in L_0$. If either Alice (resp. Bob) succeeds, they can prune out of $L_0$ all rectangles which are row (resp. column)-disjoint from $R$. If neither one can find such an $R$, then it must be that $f(x, y) = 0$. Indeed, otherwise $(x, y) \in R \in L_1$ for some $R$, and since for any $R' \in L_0$, $R$ and $R'$ are disjoint rectangles, they must be disjoint in at least one of their Alice/Bob projections. Thus, $R$ is disjoint from at least half of the rectangles in any collection of 0-rectangles in either the Alice or Bob projection. Thus, this protocol is valid, and takes at most $O(\log |L_0| \log |L_1|) = O(N^1(f)N^0(f))$ bits of communication.

As a corollary, note that since $N^1(f) = \log \chi_1(M_f) \leq \log \chi(M_f)$ and $N^0(f) \leq \log \chi_0(f) \leq \log \chi(f)$, we then have $\log \chi(f) \leq D(f) \leq O(\log^2 \chi(f))$. One might hope something similar holds if we replace the partition number with the rank:

**The Log-Rank Conjecture** [Lovasz, Saks ’88]: For any function $f$,

$$D(f) \leq (\log(\text{rank}_\mathbb{R}(M_f)))^{O(1)}$$

(Note: The conjecture is false for rank$_{\mathbb{F}_2}$, since, for example, the Hadamard matrix $M_{\mathbb{F}_2}$ has rank $2^n$ over $\mathbb{R}$ but only $n$ over $\mathbb{F}_2$.) If all this communication stuff doesn’t interest you, at least the following purely algebraic form might:

**Equivalent form 1**: For any graph $G$ with chromatic number $\chi_G$ and adjacency matrix $A_G$, we have

$$\log \chi_G \leq (\log \text{rank}(A_G))^{O(1)}$$

**Equivalent form 2**: The partition number of a boolean matrix is quasi-polynomial in the rank. Equivalently, every boolean matrix of rank $r$ can be written as a sum of $2^{\log(r)^{O(1)}}$ boolean rank-one matrices.

**Proof of equivalence**: If log-rank holds, then let $M_i$ be the 1-rectangles corresponding to the 1-leaves of a $\log(r)^{O(1)}$-protocol for $M_f$. Conversely, if $M_f$ can be written in such a way, then we get a 1-cover of $M_f$ of size $2^{\log(r)^{O(1)}}$, and since $M_{1-f}$ has pretty much the same rank as $M_f$, we also have a 0-cover of $M_f$ by as many rectangles. Thus, $N^1(f), N^0(f)$ are both polylogarithmic in $r$, and by Aho-Ullman-Yannakakis, so must be $D(f)$. □

In fact, since a result from the 90’s says $D(f) \leq O(N^1(f) \log \text{rank}(M_f))$, the log-rank conjecture is actually equivalent to the cover number $C^1(f)$ being quasi-polynomial in the rank.

**Equivalent form 3**: Let (non-negative rank) $\text{rank}^+(M)$ be the least $r$ such that $M = \sum_{i=1}^{r} x_i y_i^t$ for non-negative vectors $x_i, y_i$. Then for any boolean matrix $M$, we
have
\[
\log \text{rank}^+(M) \leq (\log \text{rank}(M))^{O(1)}
\]

Proof of equivalence: Note that \(\log(\text{rank}^+(M_f)) \leq D(f)\) since \(\text{rank}^+(M_f) \leq \chi_1(f)\), and hence 3 is implied by the log-rank conjecture. Conversely, it is a theorem of Lovasz [1990] that
\[
D(f) \leq O(\log \text{rank}^+(M_f) \log \text{rank}(M_f))
\]

Application: (extended formulations) The extension complexity \(xc(P)\) of a polyhedron \(P\) is the smallest number of facets of a higher dimensional polyhedron \(Q\) such that there exists a linear transformation \(\pi\) with \(\pi(Q) = P\). Given any description of \(P = \{x : Ax \leq b\}\) with \(f\) inequalities and \(v\) vertices \(x_1, \ldots, x_v\), then the slack matrix \(S \in \mathbb{R}^{f \times v}\) is defined by \(S_{ij} = b_i - A_ix_j\). Then

**Theorem** [Yannakakis ’91]: \(xc(P) = \text{rank}^+(S)\). If \(M_S\) is the zero-one version of \(S\), then \(\text{rank}^+(S) \geq C^1(M_S) = 2^{\chi_1(f_S)}\).

Examples: parity polytope, permutahedron, matching polytope in planar graphs, all have polynomial size extensions; can show exponential lower bounds for the correlation polytope

Question: How big does the \(O(1)\) have to be?

**Theorem** [Goos, Pitassi, Watson]: There exists an \(F\) with \(D(F) = \tilde{\Omega}(\log^2 \chi_1(F))\), and hence the log-rank conjecture cannot hold with constant less than 2.

Their function \(F\) is the composition of a complicated outer function with a small inner gadget... but in any case we get hardness for a more natural problem via a reduction:

**Corollary**: There exists a graph \(G\) with \(D(\text{CIS}_G) = \tilde{\Omega}(\log^2(n))\).

Proof: Given a function \(f\) and a 1-cover \(L_1\) of \(M_f\) by disjoint 1-rectangles, consider the graph \(G\) on \(L_1\), where two rectangles are adjacent iff they intersect in their \(\pi_A\)-projections. Then any input \((x, y)\) corresponds to a pair \((C, I)\) on \(G\) – namely \(C\) is all of the 1-rectangles in \(L_1\) whose \(\pi_A\) projections contain \(x\), and \(I\) is the set of all \(R \in L_1\) such that \(y \in \pi_B(R)\). Then \(C \cap I \neq 0 \iff f(x, y) = 1\). Apply this to the function \(F\) from the previous theorem. \(\square\)

Towards proving the conjecture, the best known upper bound was, until recently, due to Kotlov [1997]:
\[
D(f) \leq \log(4/3)\text{rank}(M_f)
\]
In 2013, this was improved by Ben-Sasson, Ron-Zewi and Lovett to

\[ D(f) \leq O(\text{rank}(M_f) / \log \text{rank}(M_f)) \]

although this result was conditional on the polynomial Freiman-Rusza conjecture, which says roughly that if \( f : \mathbb{F}_2^n \to \mathbb{F}_2^m \) is approximately linear in the sense that \( f(x + y) - f(x) - f(y) \) only takes on \( s \) values, then there is an actual linear function \( g \) such that \( g - f \) takes on at most \( s^{O(1)} \) values.

**Main Theorem** [Lovett ’14]: For any function \( f \), we have

\[ D(f) \leq O(\sqrt{\text{rank}(M_f) \log_2 \text{rank}(M_f)}) \]

Lovett’s proof relies on a discrepancy lower bound, namely, if we define

\[ \text{disc}(f) := \min_{\mu} \max_R |\mu(f^{-1}(1) \cap R) - \mu(f^{-1}(0) \cap R)| \]

then a result of Kushilevitz-Nisan [1997] says if \( \text{rank}(M_f) = r \), then \( \text{disc}(f) \geq \frac{1}{8\sqrt{r}} \).

An interesting related result by Gavinsky and Lovett [2013] shows that one may use randomized protocols in order to prove the log-rank conjecture:

**Theorem** [GL ’13]: If \( R(f) \) denotes the randomized communication complexity of \( f \), then

\[ D(f) \leq O(R(f) \log^2(\text{rank}(M_f))) \]
1. Rothvoss’s Proof

A quick overview of the proof: First we reduce the bound to the task of showing that a low rank matrix contains a large monochromatic rectangle. Since \( M \) has rank \( r \), we can find \( u_x, v_y \in \mathbb{R}^r \) with \( M_{xy} = \langle u_x, v_y \rangle \). We want subsets \( X' \subset X \) and \( Y' \subset Y \) such that \( \langle u_x, v_y \rangle = 1 \) for all \((x, y) \in X' \times Y'\) – to find them, we’ll use hyperplane rounding, that is, choose a bunch of random half spaces through the origin and set \( X', Y' \) to be the vectors lying in their intersection. The basic idea is that if \( \|u_x\|_2, \|v_y\|_2 \) aren’t too big, and \( \langle u_x, v_y \rangle = 1 \), then the angle between \( u_x, v_y \) is significantly smaller than when \( \langle u_x, v_y \rangle = -1 \), and hence \( u_x \) and \( v_y \) are more likely to end up on the same side of a random hyperplane.

First we reduce the theorem to showing that any matrix with low rank has a large monochromatic subrectangle:

**Lemma** [Nisan-Wigderson ’94]: Suppose any rank \( r \) matrix \( M \in \{\pm 1\}^{X \times Y} \) has a monochromatic rectangle \( R \) of size at least \( 2^{-c(r)}|X \times Y| \). Then for any \( f \),

\[
D(f) \leq O(\log^2(\text{rank}(M_f)) + \sum_{i=1}^{\log(\text{rank}(M_f))} c(\text{rank}(M_f)/2^i))
\]

(Note: if one could show the hypothesis of the above lemma held for \( c(r) = \text{polylog}(r) \), then the log-rank conjecture would follow. Indeed, this is equivalent to the log-rank conjecture, since then \( M_f \) is covered by \( 2^{\text{polylog}(r)} \) disjoint rectangles, and hence at least one of them has size \( 2^{-\text{polylog}(r)}|X \times Y| \).)

**Proof:** Let \( R \) be such a rectangle and

\[
M = \left( \begin{array}{ccc} R & A \\ B & C \end{array} \right)
\]

Since \( \text{rank}(A) + \text{rank}(B) \leq r + 1 \), one of them must have rank \( \leq r/2 + 1 \). WLOG suppose its \( B \). Then Bob sends a bit to indicate whether his input \( y \) lies in the shadow of \( B \) – if it does, then both parties know that \( (x, y) \) lies in a rank \( r/2 + 2 \) submatrix, while if it doesn’t, then both parties know that \( (x, y) \) lives in a submatrix of size at most \( (1 - 2^{-c(r)})|X \times Y| \).

The protocol defined in this way that stops once the rank has been reduced to \( r/2 \) has at most \( O(2^c(r) \log |X \times Y|) \) leaves – by a standard balancing act on protocol trees, this can be achieved by a protocol with depth \( O(c(r) + \log \log |X \times Y|) = O(c(r) + \log(r)) \), since assuming redundant rows and columns have been omitted from \( M \), the \( \mathbb{F}_2 \)-span of \( r \) vectors contains at most \( 2^r \) distinct vectors and hence \( |X \times Y| \leq 2^{2r} \). Resuming the process until the rank reaches \( r/4 \) and then balancing,
By construction, comparable size, and hence we can take entries are 1’s. Then the GL lemma gives us a truly monochromatic rectangle of columns leaves behind a set $B$ of $A$-0-entries. By Markov’s inequality, etc, a total of $\log(r)$ times, gives the result.

Next we use an easy lemma (essentially due to Gavinsky and Lovett, 2014) to show that it actually suffices to exhibit a large rectangle which is only almost monochromatic:

**Lemma [GL ’14]:** Suppose a boolean matrix $M$ with rank $r$ has a sub-rectangle $R$ with $(1 - \frac{1}{2^r})|R|$ of its entries equal to 1. Then $R$ contains a monochromatic rectangle $R'$ with $|R'| \geq \frac{1}{8}|R|$.

**Proof:** Write $R = A \times B$. Let $A' \subseteq A$ be the set of rows which have at most $\frac{1}{2^r}|B|$ 0-entries. By Markov’s inequality, $|A'| \geq \frac{1}{2}|A|$. The matrix $A' \times B$ still has rank at most $r$, and hence there are $r$ rows $x_1, \ldots, x_r \in A'$ which span the whole rowspace of $A' \times B$. For each $i$, let $B_i$ be the set of columns $y \in B$ such that $f(x_i, y) \neq 1$. By construction, $|B_i| \leq \frac{1}{2^r}|B|$. Hence, $\bigcup_{i=1}^r B_i \subseteq \frac{1}{2}|B|$, and so removing all these columns leaves behind a set $B'$ of at least $\frac{1}{2}|B|$ columns such that the rows $x_1, \ldots, x_r$ are all constant on $B'$, and since they span the rows of $A' \times B'$, every row in $A' \times B'$ must be constant. Taking the majority, we get a monochromatic rectangle $R'$ of size $\geq \frac{1}{2}|A' \times B'| \geq \frac{1}{8}|R|$.

Thus, our main goal will be to prove the following theorem:

**Theorem:** Given any boolean function $f : X \times Y \to \{\pm 1\}$ with rank$(M_f) = r$, and any measure $\mu$ on $X \times Y$ with $\mu(f^{-1}(1)) \geq \delta > 0$, there exists a rectangle $R \subseteq X \times Y$ with $\mu(R) \geq 2^{-\Theta(\sqrt{r} \log(\frac{1}{\delta}))}$ with $\mathbb{E}_{(x,y) \sim R}[f(x, y)] \geq 1 - \delta$.

In particular, one can set $\mu$ to be the uniform measure, $\delta = \frac{1}{2^r}$, to obtain a rectangle $|R| \geq 2^{-\Theta(\sqrt{r} \log r)}$ which is almost monochromatic – that is, a $(1 - \frac{1}{4^r})$ fraction of $R$’s entries are 1’s. Then the GL lemma gives us a truly monochromatic rectangle of comparable size, and hence we can take $c(r) = \Theta(\sqrt{r} \log(r))$ in the NW lemma to conclude $D(f) \leq O(\sqrt{r} \log(r))$ as desired.

Now let’s prove the theorem. Given a rank $r$ matrix $M \in \{\pm 1\}^{X \times Y}$, we know by one definition of rank that there exist vectors $u_x, v_y \in \mathbb{R}^r$ such that $M_{(x,y)} = \langle u_x, v_y \rangle$ and $\text{span}\{u_x : x \in X\} = \mathbb{R}^r$. For any invertible transformation $T : \mathbb{R}^r \to \mathbb{R}^r$, we can transform this set of vectors $u_x \mapsto Tu_x$ and $v_y \mapsto (T^{-1})^T v_y$ into another one with the same property, since pairwise inner products are conserved. For the proof of the theorem, we’ll want to our vectors to have small norm – $\|u_x\|_2, \|v_y\|_2 \leq r^{1/4}$ is the best we can and will do – to find a transformation $T$ which accomplishes this, we’ll use an old theorem from convex geometry by Fritz John:
John’s Theorem [John 1948]: For any full dimensional, compact, symmetric convex set \( K \subset \mathbb{R}^r \) and any ellipsoid \( E \subset \mathbb{R}^r \) centered at the origin, there is an invertible linear map \( T \) such that \( E \leq T(K) \subseteq \sqrt{r}E. \)

Now apply the theorem to \( K = \text{conv}\{\pm u_x : x \in X\} \) and \( E = r^{-1/4}B \) to find a \( T \) such that

\[
T^{-1/4}B \subseteq T(K) \subseteq r^{1/4}B
\]

and replace \( u_x \) by \( Tu_x \) and \( v_y \) by \((T^{-1})^Tv_y\). Then \( K \) becomes \( T(K) \) and we have \( \|u_x\|_2 \leq r^{1/4} \). Then

\[
\|v_y\|_2 = r^{1/4} \langle v_y, \frac{v_y}{r^{1/4}\|v_y\|_2}\rangle \leq r^{1/4} \max_{w \in K} \langle v_y, w\rangle = r^{1/4} \max_{x \in X} |\langle v_y, u_x\rangle| = r^{1/4}
\]
as desired. If we let \( \overline{u}_x, \overline{v}_y \) be the unit vectors in these directions, then

\[
\langle \overline{u}_x, \overline{v}_x \rangle = \begin{cases} \geq \frac{1}{\sqrt{r}} & \text{if } M_{xy} = 1 \\ \leq -\frac{1}{\sqrt{r}} & \text{if } M_{xy} = -1 \end{cases}
\]

We’ll now use hyperplane rounding to find our nearly monochromatic rectangle \( R \). More explicitly, let \( g_1, \ldots, g_T \sim N^r(0,1) \) be \( T \) independent random Gaussian vectors and define rectangles

\[
R_t := \{x \in X : \langle u_x, g_t \rangle \geq 0\} \times \{y \in Y : \langle v_y, g_t \rangle \geq 0\}
\]

and then set \( R := R_1 \cap \cdots \cap R_T \). We’ll see that when \( T = \frac{7 \ln(2/\delta)\sqrt{r}}{\delta} \), this rectangle \( R \) satisfies the desired properties (namely that \( \mu(R) \geq 2^{-\Theta(\sqrt{r} \log \frac{1}{\delta})} \) and \( \mu(R \cap Q_{-1}) \leq \delta \cdot \mu(R) \)) in expectation.

Recall that if the angle between \( u \) and \( v \) is \( \theta \), then

\[
\Pr[\langle g, u \rangle \geq 0 \text{ and } \langle g, v \rangle \geq 0] = \frac{1}{2} - \frac{\theta}{2\pi}
\]

Let \( Q_i = \{(x, y) : M_{xy} = i\} \). Then if \( (x, y) \in Q_1 \), the angle between \( u_x \) and \( v_y \) must be \( \leq \pi/2 \), and hence \( \Pr[(x, y) \in R_t] \geq \frac{1}{4} \), while if \( (x, y) \in Q_{-1} \), we use the inequality \( \frac{1}{2} - \frac{\theta}{2\pi} \leq \frac{1}{4} - \frac{\lfloor \cos \theta \rfloor}{7} \) for \( \theta \in (\pi/2, \pi) \) to conclude \( \Pr[(x, y) \in R_t] \leq \frac{1}{4} - \frac{\theta}{2\pi} \).

By independence, we have

\[
\mathbb{E}[\mu(R \cap Q_1)] \geq \mu(Q_1) \frac{1}{4^T} \geq \frac{\delta}{4^T} \text{ and } \mathbb{E}[\mu(R \cap Q_{-1})] \leq \mu(Q_{-1}) \left(\frac{1}{4^T} - \frac{1}{7\sqrt{r}}\right)^T
\]

and in particular

\[
\frac{\mathbb{E}[\mu(R \cap Q_{-1})]}{\mathbb{E}[\mu(R \cap Q_1)]} \leq \frac{1}{\delta} \left(1 - \frac{4}{7\sqrt{r}}\right)^T \leq \frac{1}{\delta} \exp(-\frac{4T}{7\sqrt{r}}) \leq \frac{\delta}{2}
\]
or in other words

\[
\mathbb{E}[\mu(R \cap Q_1) - \frac{1}{\delta} \mu(R \cap Q_{-1})] \geq \frac{1}{2} \mathbb{E}[\mu(R \cap Q_1)] \geq \delta/4^T \geq 2^{-\Theta(\sqrt{r} \log \frac{1}{\delta})}
\]
Letting $R$ be any rectangle attaining this expectation, we have $\mu(R) \geq 2^{-\Theta(\sqrt{\tau \log \frac{1}{\epsilon}})}$ and $\mu(R \cap Q_{-1}) \leq \delta \cdot \mu(R)$.

Note that this approach can also show the discrepancy lower bound $\text{disc}(f) \geq \Omega(1/\sqrt{r})$. Indeed, take any measure $\mu$ and WLOG assume $\mu(Q_1) \geq 1/2$, then for a single Gaussian $g \sim N^r(0, 1)$, $R = \{x \in X : \langle u_x, g \rangle \geq 0\} \times \{y \in Y : \langle v_y, g \rangle \geq 0\}$, we have

$$E[\mu(R \cap Q_1) - \mu(R \cap Q_{-1})] \geq \frac{1}{2} \cdot \frac{1}{4} - \frac{1}{2} \cdot \left(\frac{1}{4} - \frac{1}{7\sqrt{r}}\right) = \frac{1}{14\sqrt{r}}$$