Lower bounds on quantum query complexity (Draft notes)

1 Polynomial method

The key observation in the application of the polynomial method to quantum query complexity is that, after t many queries, the amplitudes in the state of the algorithm on input x are polynomials in x of degree at most t. More precisely, let

$$|\psi_x^t\rangle = |x\rangle \sum_{i,w} \alpha_{i,w}(x)|i\rangle|w\rangle \quad .$$

Then each $\alpha_{i,w}(x)$ is a polynomial of degree at most t. This follows from the following key lemma of Beals et al. [BBC+01].

Lemma 1 ([BBC+01]). *Let $\{|\psi_x\rangle\}_{x\in S}$ be a set of states where*

$$|\psi_x\rangle = \sum_{i,w} \alpha_{i,w}(x)|i\rangle|w\rangle$$

each $\alpha_{i,w}(x)$ is polynomial of degree at most t. Then

$$O_x|\psi_x\rangle = \sum_{i,w} \alpha'_{i,w}(x)|i\rangle|w\rangle$$

where each $\alpha'_{i,w}(x)$ is a polynomial of degree at most $t+1$ in x. Further, for any linear transformation A,

$$A|\psi_x\rangle = \sum_{i,w} \beta_{i,w}(x)|i\rangle|w\rangle$$

where each $\beta_{i,w}(x)$ is again a polynomial of degree at most t in x.

Proof. To see the first statement of the lemma, note that

$$O_x|\psi_x\rangle = \sum_{i,w}(2x_i - 1)\alpha_{i,w}(x)|i\rangle|w\rangle \quad .$$

Thus if $\alpha_{i,w}(x)$ is a polynomial of degree at most t, then $\alpha'_{i,w}(x) = (2x_i - 1)\alpha_{i,w}(x)$ is a polynomial of degree at most $t+1$.

For the second statement,

$$A|\psi_x\rangle = \sum_{i,w} \beta_{i,w}(x)|i\rangle|w\rangle$$

where $\beta_{i,w}(x) = \sum_{j,v} A_{(i,w),(j,v)}\alpha_{j,v}(x)$. As $\beta_{i,w}(x)$ is a linear combination of degree t polynomials, it is again a degree t polynomial. □

Corollary 2. *Let $\{|\psi_x^t\rangle\}_{x\in S}$ be the set of states of a quantum query algorithm on input x after t queries. Then*

$$|\psi_x^t\rangle = \alpha_{i,w}(x)|i\rangle|w\rangle$$

where each $\alpha_{i,w}(x)$ is a polynomial of degree at most t.

Proof. The proof is by induction. Initially, $|\psi_x^0\rangle = |0\rangle|0\rangle$ for every x. Thus $\alpha_{0,0}(x) = 1$ and $\alpha_{i,w}(x) = 0$ for every $(i,w) \neq (0,0)$. Thus each $\alpha_{i,w}(x)$ is a polynomial of degree at most 0.

By the induction hypothesis assume that after t queries the coefficients of $|\psi_x^t\rangle$ are degree at most t polynomials. We have $|\psi_x^{t+1}\rangle = O_x U|\psi_x^t\rangle$. By Lemma 1 the coefficients of $|\psi_x^{t+1}\rangle$ will be degree at most $t+1$ polynomials in x. □

Now let us look at the probability that the algorithm outputs 1 on input x. Let $\{\Pi_0, \Pi_1\}$ be the measurement used at the end of the algorithm, where Π_0, Π_1 are projectors summing to the identity. In particular, Π_1 is positive semidefinite and has a decomposition $\Pi_1 = W^*W$.

Let $\{|\psi_x^T\rangle\}_{x\in S}$ be the final states of a T-query quantum algorithm. The probability $p_{\mathcal{A}}(x)$ that the algorithm $\mathcal{A}$ outputs 1 on input x is

$$p_{\mathcal{A}}(x) = \langle\psi_x^T|\Delta_1|\psi_x^T\rangle = \|W|\psi_x^T\rangle\|^2 \ .$$

By Lemma 1 the coefficients of $W|\psi_x^T\rangle$ are degree at most T polynomials in x. Thus $p_{\mathcal{A}}(x)$ is a sum of squares of degree at most T polynomials. This gives the following theorem.

Theorem 3. *Let $\mathcal{A}$ be a T-query quantum algorithm. Then the probability $p_{\mathcal{A}}(x)$ that the algorithm outputs 1 on input x is a sum of squares of degree at most T polynomials in x. In particular, $p_{\mathcal{A}}(x)$ is a polynomial of degree at most $2T$.*

2 Positive adversary bound

In the classical adversary method, the adversary adaptively determines the input in response to the queries made by the algorithm. In this process, the adversary need not choose a particular input to answer according to initially, but rather maintains a set of possible inputs that are consistent with the queries made so far. The adversary is free to continue answering queries according to any input in this set.

Ambainis developed a lower bound for quantum query complexity coined the quantum adversary method, that is in some sense a quantum analogue of the classical adversary method [Amb02]. We will refer to this as the *positive adversary method* to distinguish it from a later adaptation called the negative adversary method.

The quantum adversary method feeds a quantum algorithm a *superposition* of inputs. As a quantum algorithm is a linear operation, if it is successful on each input individually, it must also be successful on any superposition of inputs. This is analogous in the randomized case to constructing a hard distribution of inputs.

Imagine first the simplified case of a perfect algorithm A for a boolean function f that on input $|x\rangle|0\rangle$ produces $A|x\rangle|0\rangle = |x\rangle|f(x)\rangle$. If we run the input on the superposition $|\psi_0\rangle = \sum_x \alpha_x |x\rangle|0\rangle$, then $A|\psi_0\rangle = \sum_x \alpha_x |x\rangle|f(x)\rangle$.

In the initial state $|\psi_0\rangle$, the input register is in tensor product with the workspace: $|\psi_0\rangle = (\sum_x \alpha_x |x\rangle) \otimes |0\rangle$. For a nontrivial function f, in the final state $\sum_x \alpha_x |x\rangle|f(x)\rangle$ the input register will be entangled with the workspace register. The quantum adversary method upper bounds how much this entanglement can increase with a single query.

To implement this idea, we must take some measure of the "entanglement" between the input and workspace registers at a specific point of the algorithm. Let $|\psi_x^t\rangle = \mathsf{U}_t \mathsf{O}_x \mathsf{U}_{t-1} \cdots \mathsf{U}_1 \mathsf{O}_x \mathsf{U}_0 |0\rangle|0\rangle$ be the state of the algorithm on input x after t queries. The measure used by the positive adversary method looks at the change in inner products $|\langle \psi_x^t | \psi_y^t \rangle|$ for x, y with $f(x) \neq f(y)$ as queries are made. In what follows, let $\mathsf{P}_i = \sum_z |i\rangle|z\rangle\langle z|\langle i|$ be the projector onto the subspace querying the i^{th} input bit, for $i \in \{0, \ldots, n\}$.

1. $\langle \psi_x^0 | \psi_y^0 \rangle = 1$ for all x, y.

2. If the algorithm successfully computes f with T queries then $|\langle \psi_x^T | \psi_y^T \rangle| \leq 2\sqrt{\varepsilon(1-\varepsilon)}$ for every x, y with $f(x) \neq f(y)$.

3. $\langle \psi_x^t | \psi_y^t \rangle - \langle \psi_x^{t+1} | \psi_y^{t+1} \rangle = 2 \sum_{i: x_i \neq y_i} \langle \psi_x^t | \Pi_i | \psi_y^t \rangle$.

Item 1 This is because $|\psi_x^0\rangle = \mathsf{U}_0|0\rangle|0\rangle$ is independent of the input x.

Item 2 Say that $f(x) = 0$ and $f(y) = 1$. Recall that if the algorithm is correct with error probability ε, then there are orthogonal projectors Π_0, Π_1 such that $\|\Pi_0 |\psi_y^T\rangle\|^2 \leq \varepsilon$ and $\|\Pi_1 |\psi_x^T\rangle\|^2 \leq \varepsilon$. Furthermore, as these projectors are orthogonal $1 = \|\Pi_0 |\psi_z^T\rangle\|^2 + \|\Pi_1 |\psi_z^T\rangle\|^2$ for $z \in \{x, y\}$. Thus

$$\begin{aligned}
|\langle \psi_x^T | \psi_y^T \rangle| &= |\langle \psi_x^T | \Pi_0 + \Pi_1 | \psi_y^T \rangle| \\
&\leq |\langle \psi_x^T | \Pi_0 | \psi_y^T \rangle| + |\langle \psi_x^T | \Pi_1 | \psi_y^T \rangle| \\
&\leq \|\Pi_0 |\psi_x^T\rangle\| \cdot \|\Pi_0 |\psi_y^T\rangle\| + \|\Pi_1 |\psi_x^T\rangle\| \cdot \|\Pi_1 |\psi_y^T\rangle\| \\
&\leq 2\sqrt{\varepsilon(1-\varepsilon)}
\end{aligned}$$

Item 3 First note that

$$\begin{aligned}
\langle \psi_x^{t+1} | \psi_y^{t+1} \rangle &= \langle \psi_x^t | \mathsf{O}_x \mathsf{U}_t^* \mathsf{U}_t \mathsf{O}_y | \psi_y^t \rangle \\
&= \langle \psi_x^t | \mathsf{O}_x \mathsf{O}_y | \psi_y^t \rangle \\
&= \sum_{i \in \{0,1,\ldots,n\}} (-1)^{x_i + y_i} \langle \psi_x^t | \mathsf{P}_i | \psi_y^t \rangle.
\end{aligned}$$

This means

$$\langle\psi_x^t|\psi_y^t\rangle - \langle\psi_x^{t+1}|\psi_y^{t+1}\rangle = 2\sum_{i:x_i\neq y_i}\langle\psi_x^t|\mathsf{P}_i|\psi_y^t\rangle.$$

Example: Lower bound for OR_n For this example we only need to consider inputs with Hamming weight zero or one. Let 0_n^i be the all-zero string with the i^{th} bit flipped. We define the potential measure

$$W(t) = \sum_{i\in[n]}\langle\psi_{0_n}^t|\psi_{0_n^i}^t\rangle.$$

Initially, $W(0) = n$ by Item 1. If a quantum query algorithm computes OR_n with error probability at most ε with T many queries, then $W(T) \leq 2n\sqrt{\varepsilon(1-\varepsilon)}$ by Item 2. We now bound $W(t) - W(t+1)$, how much the potential function can change with a single query. To do this, we apply Item 3 to obtain

$$\begin{aligned}
W(t) - W(t+1) &= \sum_{i\in[n]}\langle\psi_{0_n}^t|\psi_{0_n^i}^t\rangle - \langle\psi_{0_n}^{t+1}|\psi_{0_n^i}^{t+1}\rangle \\
&= 2\sum_{i\in[n]}\langle\psi_{0_n}^t|\mathsf{P}_i|\psi_{0_n^i}^t\rangle \\
&\leq 2\sum_{i\in[n]}\|\mathsf{P}_i|\psi_{0_n}^t\rangle\|\cdot\|\mathsf{P}_i|\psi_{0_n^i}^t\rangle\| \\
&\leq 2\sum_{i\in[n]}\|\mathsf{P}_i|\psi_{0_n}^t\rangle\|
\end{aligned}$$

In the final step we have simply used the bound $\|\mathsf{P}_i|\psi_{0_n^i}^t\rangle\| \leq 1$. We know that $\sum_{i\in[n]}\|\mathsf{P}_i|\psi_{0_n}^t\rangle\|^2 \leq \||\psi_{0_n}^t\rangle\|^2 = 1$. Thus applying the Cauchy-Schwarz inequality shows

$$W(t) - W(t+1) \leq 2\sqrt{n}\ .$$

Finally, $W(0) - W(T) \geq (1 - 2\sqrt{\varepsilon(1-\varepsilon)})n$ and

$$W(0) - W(T) = W(0) - W(1) + W(1) - W(2) + \cdots + W(T-1) - W(T) \leq 2T\sqrt{n}\ .$$

Combining these gives $T \geq (1 - 2\sqrt{\varepsilon(1-\varepsilon)})\sqrt{n}/2$.

We next go over the unweighted adversary method from Ambainis' original paper. This is a very easy to apply special case of the positive adversary bound that still can be used to give tight lower bounds for interesting problems, for example $\mathrm{AND}_n \circ \mathrm{OR}_n$.

Theorem 4 (Ambainis [Amb02]). *Let $S \subseteq \{0,1\}^n$ and $f : S \to \{0,1\}$ be a boolean function. If there are sets $X \subseteq f^{-1}(0)$ and $Y \subseteq f^{-1}(1)$ and a relation $R \subseteq X \times Y$ such that*

1. *for every $x \in X$ there are at least m many y with $(x,y) \in R$.*

2. *for every $y \in Y$ there are at least m' many x with $(x,y) \in R$.*

3. for every $x \in X$ and $i \in [n]$ there are at most ℓ many y with $(x,y) \in R$ and $x_i \neq y_i$.

4. for every $y \in Y$ and $i \in [n]$ there are at most ℓ' many x with $(x,y) \in R$ and $x_i \neq y_i$.

Then

$$Q_\varepsilon(f) \geq \frac{1-2\sqrt{\varepsilon(1-\varepsilon)}}{2}\sqrt{\frac{mm'}{\ell\ell'}}$$

The intuition behind this bound is that a successful algorithm must be able to distinguish every pair $(x,y) \in R$. If no index i distinguishes many of these pairs, then the algorithm will have to make many queries. We now give the proof.

Proof. The idea of the proof is similar to that for OR_n. Now consider the weight function

$$W(t) = \sum_{(x,y)\in R} \langle \psi_x^t | \psi_y^t \rangle \ .$$

If there is a T-quantum query algorithm to compute f with error probability ε then we have $W(0) - W(T) \geq (1-2\sqrt{\varepsilon(1-\varepsilon)})|R|$ by Items 1 and 2. By the assumption of the theorem, $|R| \geq |X|m$ and $|R| \geq |Y|m'$, and thus also $|R| \geq \sqrt{mm'|X||Y|}$.

Now let us consider how the weight function changes with a single query.

$$\begin{aligned}
W(t) - W(t+1) &= \sum_{(x,y)\in R} \langle \psi_x^t | \psi_y^t \rangle - \langle \psi_x^{t+1} | \psi_y^{t+1} \rangle \\
&= 2 \sum_{(x,y)\in R} \sum_{i:x_i \neq y_i} \langle \psi_x^t | \mathsf{P}_i^* \mathsf{P}_i | \psi_y^t \rangle \\
&= 2 \sum_{i\in[n]} \sum_{(x,y)\in R, x_i\neq y_i} \langle \psi_x^t | \mathsf{P}_i^* \mathsf{P}_i | \psi_y^t \rangle
\end{aligned}$$

Let $\alpha_{z,i} = \|\mathsf{P}_i|\psi_z^t\rangle\|$ for $z \in X \cup Y$ and note that $\sum_{i\in[n]} \alpha_{z,i}^2 = 1$. Applying the inequality $2ab \leq \gamma^2 a^2 + b^2/\gamma^2$ with $\gamma = (\frac{\ell'|Y|}{\ell|X|})^{1/4}$ we have

$$\begin{aligned}
W(t) - W(t+1) &\leq \sum_{i:x_i\neq y_i} \sum_{(x,y)\in R, x_i\neq y_i} \sqrt{\frac{\ell'|Y|}{\ell|X|}}\alpha_{x,i}^2 + \sqrt{\frac{\ell|X|}{\ell'|Y|}}\alpha_{y,i}^2 \\
&\leq \sum_{i\in[n]} \sum_{x\in X} \sqrt{\frac{\ell\ell'|Y|}{|X|}}\alpha_{x,i}^2 + \sum_{i\in[n]} \sum_{y\in Y} \sqrt{\frac{\ell\ell'|X|}{|Y|}}\alpha_{y,i}^2 \\
&= \sum_{x\in X} \sqrt{\frac{\ell\ell'|Y|}{|X|}} \sum_{i\in[n]} \alpha_{x,i}^2 + \sum_{y\in Y} \sqrt{\frac{\ell\ell'|X|}{|Y|}} \sum_{i\in[n]} \alpha_{y,i}^2 \\
&= 2\sqrt{\ell\ell'|X||Y|} \ .
\end{aligned}$$

Putting everything together, as $W(0) - W(T) \geq (1 - 2\sqrt{\varepsilon(1-\varepsilon)})\sqrt{mm'|X||Y|}$ and $W(t) - W(t+1) \leq 2\sqrt{\ell\ell'|X||Y|}$ we find that

$$T \geq \frac{1 - 2\sqrt{\varepsilon(1-\varepsilon)}}{2}\sqrt{\frac{mm'}{\ell\ell'}} \ ,$$

as desired. $\square$

Example: Lower bound for $f = \mathrm{OR}_n \circ \mathrm{AND}_n$ We will decompose an input $\bar{x} \in \{0,1\}^{n^2}$ as $\bar{x} = (x_1, \ldots, x_n)$ with each $x_i \in \{0,1\}^n$. We will let the set $X \subseteq f^{-1}(0)$ be the maxterms of the $\mathrm{OR}_n \circ \mathrm{AND}_n$. That is, the set of all inputs $(x_1, \ldots, x_n)$ where each x_i has exactly one zero. We will let $Y \subseteq f^{-1}(1)$ be the set of inputs $(y_1, \ldots, y_n)$ where $y_i = 1_n$ for exactly one $i \in [n]$, and every other string y_j has exactly one zero, for $j \neq i$. Each $(x_1, \ldots, x_n) \in X$ will be in relation with exactly n many $\bar{y} \in Y$, namely $(1_n, x_2, \ldots, x_n), \ldots, (x_1, \ldots, x_{n-1}, 1_n)$. Each $(y_1, \ldots, y_n) \in Y$ will be in relation with n many $\bar{x} \in X$. Namely, if the i^{th} block of $\bar{y}$ is all ones then $\bar{y}$ will be in relation with the strings $(y_1, \ldots, y_{i-1}, \bar{1}_n^j, y_{i+1}, \ldots, y_n)$ for $j = 1, \ldots, n$. Note that every pair $(\bar{x}, \bar{y}) \in R$ has Hamming distance 1. This means that for each $x \in X$ and $i \in [n^2]$, there is at most one y with $(x, y) \in R$ and $x_i \neq y_i$. A similar fact holds for each $y \in Y$, thus $\ell = \ell' = 1$ in the application of Theorem 4. As $m = m' = 1$, we obtain a bound of $\Omega(n)$ on the bounded-error quantum query complexity of $\mathrm{OR}_n \circ \mathrm{AND}_n$.

3 General form of positive adversary method

Theorem 4 is very easy to use and effective for many problems. The idea behind this bound can be generalized, however, to give an elegant and general lower bound, the positive adversary bound.

So far we have only seen applications of this bound with uniform weights. Instead of a relation on $X \times Y \subseteq f^{-1}(0) \times f^{-1}(1)$, we can instead consider giving a *weight* $\Gamma(x, y) \geq 0$ (possibly zero) to every element of $f^{-1}(0) \times f^{-1}(1)$. We will think of Γ as a $|f^{-1}(0)|$-by-$|f^{-1}(1)|$ matrix. Intuitively $\Gamma(x, y)$ should be given more weight the harder x and y are to distinguish, i.e. if they are close in Hamming distance.

Definition 5. Let $S \subseteq \{0,1\}^n$ and $f : S \to \{0,1\}$. The positive-weight adversary bound of f, denoted $\mathrm{ADV}(f)$, is defined as

$$\mathrm{ADV}(f) = \underset{\Gamma \geq 0}{\mathrm{maximize}} \frac{\|\Gamma\|}{\max_{i \in \{1,\ldots,n\}} \|\Gamma \circ D_i\|} \ .$$

Here $\Gamma, D_1, \ldots, D_n$ are $|f^{-1}(0)|$-by-$|f^{-1}(1)|$ matrices and D_i is defined as $D_i(x, y) = 1$ if $x_i \neq y_i$ and $D_i(x, y) = 0$ otherwise.

Theorem 6. *Let $S \subseteq \{0,1\}^n$ and $f : S \to \{0,1\}$. Then*

$$Q_\varepsilon(f) \geq \frac{1 - 2\sqrt{\varepsilon(1-\varepsilon)}}{2}\mathrm{ADV}(f) \ .$$

Proof. Let $X = f^{-1}(0)$ and $Y = f^{-1}(1)$ We show that for any $|X|$-by-$|Y|$ matrix Γ that is nonnegative it holds that

$$Q_\varepsilon(f) \geq \frac{1-2\sqrt{\varepsilon(1-\varepsilon)}}{2} \frac{\|\Gamma\|}{\max_{i\in\{1,\ldots,n\}} \|\Gamma \circ D_i\|} \ .$$

Let $u \in \mathbb{R}^{|X|}, v \in \mathbb{R}^{|Y|}$ be vectors such that $u^T \Gamma v = \|\Gamma\|$. As Γ is nonnegative we can take u, v to be as well.

Now consider a successful ε-error quantum algorithm for f making T queries. Let $|\psi_x^t\rangle$ be the state of the algorithm on input x after t queries.

We define a potential function based on Γ, u, v as follows:

$$W(t) = \sum_{x\in X, y \in Y} \Gamma(x,y) u_x v_y \langle \psi_x^t | \psi_y^t \rangle \ .$$

As $\langle \psi_x^0 | \psi_y^0 \rangle = 1$ for all x, y we have $W(0) = \|\Gamma\|$.

By Item 2 above, for the algorithm to have error at most ε we must have $\langle \psi_x^T | \psi_y^T \rangle \leq 2\sqrt{\varepsilon(1-\varepsilon)}$. As all of $\Gamma(x,y), \alpha_x, \beta_y$ are nonnegative, this means

$$W(T) = \sum_{x\in X, y\in Y} \Gamma(x,y) u_x v_y \langle \psi_x^T | \psi_y^T \rangle \leq 2\sqrt{\varepsilon(1-\varepsilon)} \|\Gamma\| \ .$$

Thus $W(0) - W(T) \geq (1 - 2\sqrt{\varepsilon(1-\varepsilon)})\|\Gamma\|$.

Next we upper bound the change in the potential function with a single query $W(t) - W(t+1)$. Using Item 3 above we have

$$\begin{aligned} W(t) - W(t+1) &= 2 \sum_{x\in X, y\in Y} \Gamma(x,y) u_x v_y \sum_{i: x_i \neq y_i} \langle \psi_x^t | P_i^* P_i | \psi_y^t \rangle \\ &= 2 \sum_{i\in[n]} \sum_{x\in X, y \in Y} \Gamma(x,y) D_i(x,y) u_x v_y \langle \psi_x^t | P_i^* P_i | \psi_y^t \rangle \end{aligned}$$

Let us now phrase this in terms of the trace product of matrices. Let $\Gamma \circ D_i$ denote the entrywise product of Γ and D_i. Let B_i be a $|X|$-by-$|Y|$ matrix where $B_i(x,y) = u_x v_y \langle \psi_x^t | P_i^* P_i | \psi_y^t \rangle$, and let $\|B_i\|_{tr}$ be the trace norm of B_i. Then we have

$$\begin{aligned} W(t) - W(t+1) &= 2 \sum_{i\in[n]} \mathrm{Tr}((\Gamma \circ D_i) B_i^T) \\ &\leq 2 \sum_{i \in [n]} \|\Gamma \circ D_i\| \|B_i\|_{tr} \\ &\leq 2 \max_{i\in[n]} \|\Gamma \circ D_i\| \sum_{j\in[n]} \|B_j\|_{tr} \ . \end{aligned}$$

Now we proceed to show that $\sum_{j\in[n]} \|B_j\|_{tr} = 1$. To do this we use the inequality $\|QR^T\|_t r \leq \frac{1}{2}(\|Q\|_F^2 + \|R\|_F^2)$. Note that $B_j = Q_j R_j^T$ where Q_j is the matrix with $Q_j(x,:) = u_x \mathsf{P}_i|\psi_x^t\rangle$ for $x \in X$ and R_j is the matrix with $R_j(y,:) = v_y \mathsf{P}_i|\psi_y^t\rangle$ for $y \in Y$. Then

$$\begin{aligned}
\sum_{j\in[n]} \|Q_j\|_F^2 &= \sum_{j\in[n]} \sum x \in X v_x^2 \|\mathsf{P}_i|\psi_x^t\rangle\|^2 \\
&= \sum x \in X v_x^2 \sum_{j\in[n]} \|\mathsf{P}_i|\psi_x^t\rangle\|^2 \\
&= \sum x \in X v_x^2 \\
&= 1 \quad .
\end{aligned}$$

A similar computation shows $\sum_{j\in[n]} \|R_j\|_F^2 = 1$. This implies $W(t) - W(t+1) \leq 2 \max_{i\in[n]} \|\Gamma \circ D_i\|$.

Putting everything together $W(0) - W(T) \geq (1 - 2\sqrt{\varepsilon(1-\varepsilon)})\|\Gamma\|$ and

$$\begin{aligned}
W(0) - W(T) &= W(0) - W(1) + W(1) - W(2) + \cdots W(T-1) - W(T) \\
&\leq 2T \max_{i\in[n]} \|\Gamma \circ D_i\| \quad .
\end{aligned}$$

Combining these gives the theorem. □

References

[Amb02] Andris Ambainis. Quantum lower bounds by quantum arguments. *J. Comput. Syst. Sci.*, 64(4):750–767, 2002.

[BBC+01] Robert Beals, Harry Buhrman, Richard Cleve, Michele Mosca, and Ronald de Wolf. Quantum lower bounds by polynomials. *J. ACM*, 48(4):778–797, 2001.