Fourier analysis of boolean functions in quantum computation

Ashley Montanaro

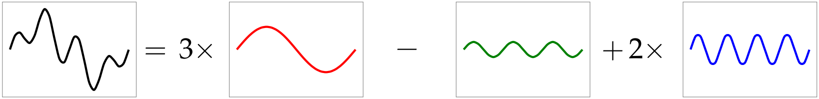
Centre for Quantum Information and Foundations,
Department of Applied Mathematics and Theoretical Physics,
University of Cambridge

arXiv:1007.3587 and arXiv:0810.2435

EPSRC Engineering and Physical Sciences Research Council

Fourier analysis

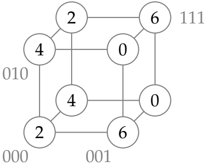
...traditionally looks like this:



- Given some (periodic) function $f : \mathbb{R} \to \mathbb{R}$...
- ...we expand it in terms of trigonometric functions $\sin(kx)$, $\cos(kx)$...
- ...in an attempt to understand the structure of f.

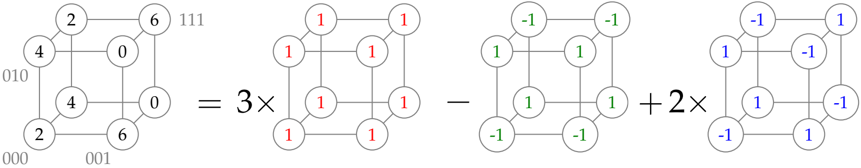
Fourier analysis

In computer science, it's natural to consider functions on the set of *n*-bit strings – also known as the boolean cube $\{0,1\}^n$:



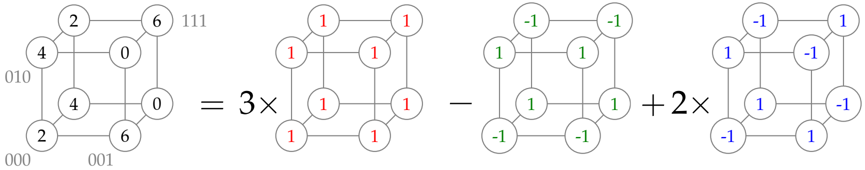
Fourier analysis

In computer science, it's natural to consider functions on the set of *n-bit strings* – also known as the boolean cube $\{0, 1\}^n$:



Fourier analysis

In computer science, it's natural to consider functions on the set of *n*-bit strings – also known as the boolean cube $\{0,1\}^n$:



- Given some function $f : \{0,1\}^n \to \mathbb{R}$...
- ...we expand it in terms of parity functions...
- ...in an attempt to understand the structure of f.

Fourier analysis on the boolean cube

- We expand functions $f : \{0,1\}^n \to \mathbb{R}$ in terms of the parity functions

  $$\chi_S(x) = (-1)^{\sum_{i \in S} x_i},$$

  also known as the characters of $\mathbb{Z}_2^n$.

Fourier analysis on the boolean cube

- We expand functions $f : \{0,1\}^n \to \mathbb{R}$ in terms of the parity functions
$$\chi_S(x) = (-1)^{\sum_{i \in S} x_i},$$
also known as the characters of $\mathbb{Z}_2^n$.
- There are 2^n of these functions, indexed by subsets $S \subseteq \{1, \ldots, n\}$. $\chi_S(x) = -1$ if the no. of bits of x in S set to 1 is odd.

Fourier analysis on the boolean cube

- We expand functions $f : \{0,1\}^n \to \mathbb{R}$ in terms of the parity functions
$$\chi_S(x) = (-1)^{\sum_{i \in S} x_i},$$
also known as the characters of $\mathbb{Z}_2^n$.
- There are 2^n of these functions, indexed by subsets $S \subseteq \{1, \ldots, n\}$. $\chi_S(x) = -1$ if the no. of bits of x in S set to 1 is odd.
- Any $f : \{0,1\}^n \to \mathbb{R}$ has the expansion
$$f = \sum_{S \subseteq \{1,\ldots,n\}} \hat{f}(S)\, \chi_S$$
for some $\{\hat{f}(S)\}$ – the Fourier coefficients of f.
- The degree of f is $\max\{|S| : \hat{f}(S) \neq 0\}$, which is just the degree of f as a real n-variate polynomial.

Applications of Fourier analysis on the boolean cube

This approach has led to new results in many areas of classical computer science, including:

- Probabilistically checkable proofs [Håstad '01; Dinur '07; . . .]
- Decision tree complexity [Nisan & Szegedy '94]
- Influence of voters and fairness of elections [Kahn, Kalai, Linial '88; Kalai '02]
- Computational learning theory [Goldreich & Levin '89; Kushilevitz & Mansour '91; . . .]
- Property testing [Bellare et al '95; Matulef et al '09; . . .]

This talk

This talk is about applying and generalising Fourier analysis on the boolean cube in quantum computation.

- Quantum vs. classical communication complexity
- Hypercontractivity and low-degree polynomials
- Generalising Fourier analysis to quantum computation
- Spectra of k-local operators

One-way communication complexity

- The field of communication complexity studies the amount of communication between parties required for them to compute some function of their joint inputs.

One-way communication complexity

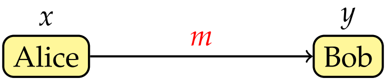
- The field of communication complexity studies the amount of communication between parties required for them to compute some function of their joint inputs.

- One of the simplest models of communication complexity is the one-way model.

One-way communication complexity

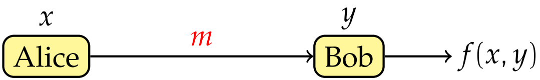
- The field of communication complexity studies the amount of communication between parties required for them to compute some function of their joint inputs.

- One of the simplest models of communication complexity is the one-way model.



One-way communication complexity

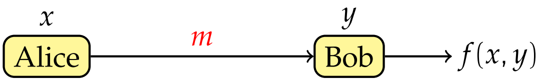
- The field of communication complexity studies the amount of communication between parties required for them to compute some function of their joint inputs.
- One of the simplest models of communication complexity is the one-way model.



One-way communication complexity

- The field of communication complexity studies the amount of communication between parties required for them to compute some function of their joint inputs.

- One of the simplest models of communication complexity is the one-way model.



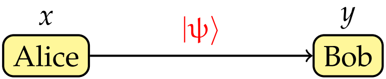
- The classical one-way communication complexity (1WCC) of a boolean function f is the length of the shortest message m sent from Alice to Bob that allows Bob to compute $f(x,y)$ with constant probability of success $> 1/2$.

One-way quantum communication complexity

Can we do better by sending a quantum message?

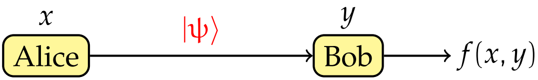
One-way quantum communication complexity

Can we do better by sending a quantum message?



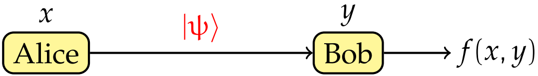
One-way quantum communication complexity

Can we do better by sending a quantum message?



One-way quantum communication complexity

Can we do better by sending a quantum message?



- The quantum 1WCC of f is the smallest number of qubits sent from Alice to Bob that allows Bob to compute $f(x, y)$ with constant probability of success $> 1/2$.

- We don't allow Alice and Bob to share any prior entanglement or randomness.

Quantum one-way communication complexity

The model of quantum one-way communication complexity is not (very) well understood. We know that:

Quantum one-way communication complexity

The model of quantum one-way communication complexity is not (very) well understood. We know that:

- If $f(x, y)$ is allowed to be a partial function (i.e. there is a promise on the inputs), there can be an exponential separation between quantum and classical 1WCC [Gavinsky et al '08].

Quantum one-way communication complexity

The model of quantum one-way communication complexity is not (very) well understood. We know that:

- If $f(x, y)$ is allowed to be a partial function (i.e. there is a promise on the inputs), there can be an exponential separation between quantum and classical 1WCC [Gavinsky et al '08].
- In fact, for partial functions, quantum one-way communication is exponentially stronger than even two-way classical communication [Klartag and Regev '10].

Quantum one-way communication complexity

The model of quantum one-way communication complexity is not (very) well understood. We know that:

- If $f(x, y)$ is allowed to be a partial function (i.e. there is a promise on the inputs), there can be an exponential separation between quantum and classical 1WCC [Gavinsky et al '08].
- In fact, for partial functions, quantum one-way communication is exponentially stronger than even two-way classical communication [Klartag and Regev '10].
- If $f(x, y)$ is a total function, the best separation we have is a factor of 2 for equality testing [Winter '04].

Quantum one-way communication complexity

The model of quantum one-way communication complexity is not (very) well understood. We know that:

- If $f(x, y)$ is allowed to be a partial function (i.e. there is a promise on the inputs), there can be an exponential separation between quantum and classical 1WCC [Gavinsky et al '08].
- In fact, for partial functions, quantum one-way communication is exponentially stronger than even two-way classical communication [Klartag and Regev '10].
- If $f(x, y)$ is a total function, the best separation we have is a factor of 2 for equality testing [Winter '04].

Today: I'll talk about a (slight) improvement on the separation of [Gavinsky et al '08], based on Fourier-analytic techniques.

The problem

Perm-Invariance

- Alice gets an n-bit string x.
- Bob gets an $n \times n$ permutation matrix M.
- Bob has to output $\begin{cases} 1 & \text{if } Mx = x \\ 0 & \text{if } d(Mx, x) \geqslant \beta|x| \\ \text{anything} & \text{otherwise,} \end{cases}$

where β is a constant, $|x|$ is the Hamming weight of x and $d(x, y)$ is the Hamming distance between x and y.

The problem

Perm-Invariance

- Alice gets an n-bit string x.
- Bob gets an $n \times n$ permutation matrix M.
- Bob has to output $\begin{cases} 1 & \text{if } Mx = x \\ 0 & \text{if } d(Mx, x) \geqslant \beta|x| \\ \text{anything} & \text{otherwise,} \end{cases}$

where β is a constant, $|x|$ is the Hamming weight of x and $d(x, y)$ is the Hamming distance between x and y.

This is a natural (?) generalisation of the Subgroup Membership problem where Alice gets a subgroup $H \leqslant G$, Bob gets a group element $g \in G$, and they have to determine if $g \in H$.

Main result

Theorem

- There is a quantum protocol that solves Perm-Invariance with constant success probability and communicates $O(\log n)$ bits.

Main result

Theorem

- There is a quantum protocol that solves Perm-Invariance with constant success probability and communicates $O(\log n)$ bits.
- Any one-way classical protocol that solves Perm-Invariance with a constant success probability strictly greater than 1/2 must communicate at least $\Omega(n^{7/16})$ bits (for $\beta = 1/8$).

Main result

Theorem

- There is a quantum protocol that solves Perm-Invariance with constant success probability and communicates $O(\log n)$ bits.
- Any one-way classical protocol that solves Perm-Invariance with a constant success probability strictly greater than $1/2$ must communicate at least $\Omega(n^{7/16})$ bits (for $\beta = 1/8$).

Therefore, there is an exponential separation between quantum and classical one-way communication complexity for this problem.

The lower bound has since been improved to $\Omega(n^{1/2})$ by [Verbin and Yu '11].

The quantum protocol

The quantum protocol is simple:

The quantum protocol

The quantum protocol is simple:

- Alice prepares two copies of the $\log n$ qubit state $|\psi_x\rangle := \sum_{i,x_i=1} |i\rangle$ and sends them to Bob.

The quantum protocol

The quantum protocol is simple:

- Alice prepares two copies of the $\log n$ qubit state $|\psi_x\rangle := \sum_{i, x_i = 1} |i\rangle$ and sends them to Bob.

- Bob performs the unitary operator corresponding to the permutation M on one of the states, to produce the state $|\psi_{Mx}\rangle$, and then uses the swap test to check whether the states are equal.

The quantum protocol

The quantum protocol is simple:

- Alice prepares two copies of the $\log n$ qubit state $|\psi_x\rangle := \sum_{i,x_i=1} |i\rangle$ and sends them to Bob.

- Bob performs the unitary operator corresponding to the permutation M on one of the states, to produce the state $|\psi_{Mx}\rangle$, and then uses the swap test to check whether the states are equal.

- By the promise that either $|\psi_{Mx}\rangle = |\psi_x\rangle$, or $\langle\psi_{Mx}|\psi_x\rangle \leqslant 1/8$, these two cases can be distinguished with a constant number of repetitions.

The classical lower bound

We prove a lower bound for a special case of Perm-Invariance.

PM-Invariance

- Alice gets a $2n$-bit string x such that $|x| = n$.
- Bob gets a $2n \times 2n$ permutation matrix M, where the permutation entirely consists of disjoint transpositions (i.e. corresponds to a perfect matching on the complete graph on $2n$ vertices).
- Bob has to output $\begin{cases} 1 & \text{if } Mx = x \\ 0 & \text{if } d(Mx, x) \geqslant n/8 \\ \text{anything} & \text{otherwise.} \end{cases}$

Plan of attack

- Imagine Alice and Bob have a randomised protocol that uses a small amount of communication.

Plan of attack

- Imagine Alice and Bob have a randomised protocol that uses a small amount of communication.

- Fixing a distribution on the inputs, this corresponds to a partition of Alice's inputs into large subsets, each corresponding to a short message.

Plan of attack

- Imagine Alice and Bob have a randomised protocol that uses a small amount of communication.

- Fixing a distribution on the inputs, this corresponds to a partition of Alice's inputs into large subsets, each corresponding to a short message.

- Fix two "hard" distributions: one on Alice & Bob's zero-valued inputs, and one on their one-valued inputs.

Plan of attack

- Imagine Alice and Bob have a randomised protocol that uses a small amount of communication.

- Fixing a distribution on the inputs, this corresponds to a partition of Alice's inputs into large subsets, each corresponding to a short message.

- Fix two "hard" distributions: one on Alice & Bob's zero-valued inputs, and one on their one-valued inputs.

- Show that the induced distributions on Bob's inputs are close to uniform whenever Alice's subset is large.

Plan of attack

- Imagine Alice and Bob have a randomised protocol that uses a small amount of communication.
- Fixing a distribution on the inputs, this corresponds to a partition of Alice's inputs into large subsets, each corresponding to a short message.
- Fix two "hard" distributions: one on Alice & Bob's zero-valued inputs, and one on their one-valued inputs.
- Show that the induced distributions on Bob's inputs are close to uniform whenever Alice's subset is large.
- This means they're hard for Bob to distinguish.

Proof idea: one-valued inputs

We want to show that Bob's induced distribution on inputs such that $Mx = x$ is close to uniform (the argument for zero-valued inputs is similar but easier).

Proof idea: one-valued inputs

We want to show that Bob's induced distribution on inputs such that $Mx = x$ is close to uniform (the argument for zero-valued inputs is similar but easier).

- Fix distribution $\mathcal{D}_1$ to be uniform over all pairs (M, x) such that $Mx = x$.

Proof idea: one-valued inputs

We want to show that Bob's induced distribution on inputs such that $Mx = x$ is close to uniform (the argument for zero-valued inputs is similar but easier).

- Fix distribution $\mathcal{D}_1$ to be uniform over all pairs (M, x) such that $Mx = x$.
- Let p_M be the probability under $\mathcal{D}_1$ that Bob gets M, given that Alice's input was in A, for an arbitrary set A.

Proof idea: one-valued inputs

We want to show that Bob's induced distribution on inputs such that $Mx = x$ is close to uniform (the argument for zero-valued inputs is similar but easier).

- Fix distribution $\mathcal{D}_1$ to be uniform over all pairs (M, x) such that $Mx = x$.
- Let p_M be the probability under $\mathcal{D}_1$ that Bob gets M, given that Alice's input was in A, for an arbitrary set A.
- Let N_{2n} be the number of partitions of $\{1, \ldots, 2n\}$ into pairs. Then

$$p_M = \frac{\binom{2n}{n}}{N_{2n}\binom{n}{n/2}} \Pr_{x \in A}[Mx = x].$$

Proof idea

We want to show that Bob's induced distribution on inputs such that $Mx = x$ is close to uniform.

Proof idea

We want to show that Bob's induced distribution on inputs such that $Mx = x$ is close to uniform.

- Upper bounding the 1-norm by the 2-norm, we have

$$\|\mathcal{D}_1^A - U\|_1 \leqslant \sqrt{N_{2n} \sum_M p_M^2 - 1}$$

where U is the uniform distribution on Bob's inputs.

Proof idea

We want to show that Bob's induced distribution on inputs such that $Mx = x$ is close to uniform.

- Upper bounding the 1-norm by the 2-norm, we have

$$\|\mathcal{D}_1^A - U\|_1 \leqslant \sqrt{N_{2n} \sum_M p_M^2 - 1}$$

where U is the uniform distribution on Bob's inputs.

- We can now calculate

$$N_{2n} \sum_M p_M^2 = \frac{\binom{2n}{n}^2}{N_{2n} \binom{n}{n/2}^2 |A|^2} \left(\sum_{x,y \in A} \sum_M [Mx = x, My = y] \right).$$

Proof idea

- It turns out that the sum over M only depends on the Hamming distance $d(x, y)$:

$$\sum_M [Mx = x, My = y] = h(x + y)$$

  where $h : \{0, 1\}^{2n} \to \mathbb{R}$ is a function such that $h(z)$ only depends on the Hamming weight $|z|$.

Proof idea

- It turns out that the sum over M only depends on the Hamming distance $d(x, y)$:

$$\sum_M [Mx = x, My = y] = h(x + y)$$

  where $h : \{0, 1\}^{2n} \to \mathbb{R}$ is a function such that $h(z)$ only depends on the Hamming weight $|z|$.

- So

$$N_{2n} \sum_M p_M^2 = \frac{\binom{2n}{n}^2}{N_{2n} \binom{n}{n/2}^2 |A|^2} \left(\sum_{x,y} f(x) f(y) h(x + y) \right),$$

  where f is the characteristic function of A.

Proof idea

- It turns out that the sum over M only depends on the Hamming distance $d(x, y)$:

$$\sum_M [Mx = x, My = y] = h(x + y)$$

  where $h : \{0, 1\}^{2n} \to \mathbb{R}$ is a function such that $h(z)$ only depends on the Hamming weight $|z|$.

- So

$$N_{2n} \sum_M p_M^2 = \frac{\binom{2n}{n}^2}{N_{2n} \binom{n}{n/2}^2 |A|^2} \left(\sum_{x,y} f(x) f(y) h(x + y) \right),$$

  where f is the characteristic function of A.

- This means that it's convenient to upper bound $N_{2n} \sum_M p_M^2$ using Fourier analysis over the group $\mathbb{Z}_2^{2n}$.

Fourier analysis to the rescue

- For any functions $f, g : \{0,1\}^n \to \mathbb{R}$,

$$\sum_{x,y \in \{0,1\}^n} f(x)f(y)g(x+y) = 2^{2n} \sum_{S \subseteq [n]} \hat{g}(S)\hat{f}(S)^2.$$

Fourier analysis to the rescue

- For any functions $f, g : \{0,1\}^n \to \mathbb{R}$,

$$\sum_{x,y \in \{0,1\}^n} f(x)f(y)g(x+y) = 2^{2n} \sum_{S \subseteq [n]} \hat{g}(S)\hat{f}(S)^2.$$

- This allows us to write

$$N_{2n} \sum_M p_M^2 = \frac{\binom{2n}{n}^2 2^{4n}}{N_{2n}\binom{n}{n/2}^2} \frac{1}{|A|^2} \sum_{S \subseteq [n]} \hat{h}(S)\hat{f}(S)^2,$$

where f is the characteristic function of A, and h is as on the previous slide.

Upper bounding this sum

We can upper bound this sum using the following crucial inequality.

Lemma

Let A be a subset of $\{0,1\}^n$, let f be the characteristic function of A, and set $2^{-\alpha} = |A|/2^n$. Then, for any $1 \leqslant k \leqslant (\ln 2)\alpha$,

$$\sum_{x,|x|=k} \hat{f}(x)^2 \leqslant 2^{-2\alpha} \left(\frac{(2e \ln 2)\alpha}{k} \right)^k .$$

Upper bounding this sum

We can upper bound this sum using the following crucial inequality.

Lemma

Let A be a subset of $\{0,1\}^n$, let f be the characteristic function of A, and set $2^{-\alpha} = |A|/2^n$. Then, for any $1 \leqslant k \leqslant (\ln 2)\alpha$,

$$\sum_{x,|x|=k} \hat{f}(x)^2 \leqslant 2^{-2\alpha} \left(\frac{(2e \ln 2)\alpha}{k} \right)^k .$$

- This inequality is based on a result of Kahn, Kalai and Linial (the KKL Lemma), which in turn is based on a "hypercontractive" inequality of Bonami, Gross and Beckner.

Upper bounding this sum

We can upper bound this sum using the following crucial inequality.

Lemma

Let A be a subset of $\{0,1\}^n$, let f be the characteristic function of A, and set $2^{-\alpha} = |A|/2^n$. Then, for any $1 \leqslant k \leqslant (\ln 2)\alpha$,

$$\sum_{x,|x|=k} \hat{f}(x)^2 \leqslant 2^{-2\alpha} \left(\frac{(2e \ln 2)\alpha}{k} \right)^k .$$

- This inequality is based on a result of Kahn, Kalai and Linial (the KKL Lemma), which in turn is based on a "hypercontractive" inequality of Bonami, Gross and Beckner.
- Here α ends up (approximately) measuring the length of Alice's message in bits.

Finishing up 1WCC

To summarise:

- We calculate and upper bound the Fourier transform $\hat{h}(x)$, which turns out to be exponentially decreasing with $|x|$.

Finishing up 1WCC

To summarise:

- We calculate and upper bound the Fourier transform $\hat{h}(x)$, which turns out to be exponentially decreasing with $|x|$.
- We upper bound the "Fourier weight at the k'th level" of f, $\|f^{=k}\|_2^2$, using the KKL Lemma.

Finishing up 1WCC

To summarise:

- We calculate and upper bound the Fourier transform $\hat{h}(x)$, which turns out to be exponentially decreasing with $|x|$.
- We upper bound the "Fourier weight at the k'th level" of f, $\|f^{=k}\|_2^2$, using the KKL Lemma.
- Combining the two upper bounds, we end up with something that's smaller than a constant unless $|A| \leqslant 2^{2n-\Omega(n^{7/16})}$.

Finishing up 1WCC

To summarise:

- We calculate and upper bound the Fourier transform $\hat{h}(x)$, which turns out to be exponentially decreasing with $|x|$.
- We upper bound the "Fourier weight at the k'th level" of f, $\|f^{=k}\|_2^2$, using the KKL Lemma.
- Combining the two upper bounds, we end up with something that's smaller than a constant unless $|A| \leqslant 2^{2n-\Omega(n^{7/16})}$.
- Thus, unless Alice sends at least $\Omega(n^{7/16})$ bits to Bob, he can't distinguish his induced distribution from uniform with probability better than a fixed constant.

Finishing up 1WCC

To summarise:

- We calculate and upper bound the Fourier transform $\hat{h}(x)$, which turns out to be exponentially decreasing with $|x|$.
- We upper bound the "Fourier weight at the k'th level" of f, $\|f^{=k}\|_2^2$, using the KKL Lemma.
- Combining the two upper bounds, we end up with something that's smaller than a constant unless $|A| \leqslant 2^{2n-\Omega(n^{7/16})}$.
- Thus, unless Alice sends at least $\Omega(n^{7/16})$ bits to Bob, he can't distinguish his induced distribution from uniform with probability better than a fixed constant.
- So the classical 1WCC of PM-Invariance is $\Omega(n^{7/16})$.

Hypercontractivity and noise: an interlude

- The KKL Lemma is fundamentally based on understanding the application of noise to functions $f : \{0,1\}^n \to \mathbb{R}$.

Hypercontractivity and noise: an interlude

- The KKL Lemma is fundamentally based on understanding the application of noise to functions $f : \{0,1\}^n \to \mathbb{R}$.
- We now define the noise operator $\mathcal{D}_\rho$ with noise rate ρ. For a given string $x \in \{0,1\}^n$, define the distribution $y \sim_\rho x$ as follows. Each coordinate $y_i = x_i$ with probability $1/2 + \rho/2$, and $y_i = 1 - x_i$ with probability $1/2 - \rho/2$.

Hypercontractivity and noise: an interlude

- The KKL Lemma is fundamentally based on understanding the application of noise to functions $f : \{0,1\}^n \to \mathbb{R}$.
- We now define the noise operator $\mathcal{D}_\rho$ with noise rate ρ. For a given string $x \in \{0,1\}^n$, define the distribution $y \sim_\rho x$ as follows. Each coordinate $y_i = x_i$ with probability $1/2 + \rho/2$, and $y_i = 1 - x_i$ with probability $1/2 - \rho/2$.
- In other words, each bit of x is flipped with probability $1/2 - \rho/2$.

Hypercontractivity and noise: an interlude

- The KKL Lemma is fundamentally based on understanding the application of noise to functions $f : \{0,1\}^n \to \mathbb{R}$.
- We now define the noise operator $\mathcal{D}_\rho$ with noise rate ρ. For a given string $x \in \{0,1\}^n$, define the distribution $y \sim_\rho x$ as follows. Each coordinate $y_i = x_i$ with probability $1/2 + \rho/2$, and $y_i = 1 - x_i$ with probability $1/2 - \rho/2$.
- In other words, each bit of x is flipped with probability $1/2 - \rho/2$.
- Then write

$$(\mathcal{D}_\rho f)(x) = \mathbb{E}_{y \sim_\rho x}[f(y)].$$

Hypercontractivity and noise: an interlude

- The KKL Lemma is fundamentally based on understanding the application of noise to functions $f : \{0,1\}^n \to \mathbb{R}$.
- We now define the noise operator $\mathcal{D}_\rho$ with noise rate ρ. For a given string $x \in \{0,1\}^n$, define the distribution $y \sim_\rho x$ as follows. Each coordinate $y_i = x_i$ with probability $1/2 + \rho/2$, and $y_i = 1 - x_i$ with probability $1/2 - \rho/2$.
- In other words, each bit of x is flipped with probability $1/2 - \rho/2$.
- Then write
$$(\mathcal{D}_\rho f)(x) = \mathbb{E}_{y \sim_\rho x}[f(y)].$$
- Crucially, noise "smoothes out" high-order Fourier coefficients:
$$\widehat{\mathcal{D}_\rho f}(S) = \rho^{|S|} \hat{f}(S).$$

Hypercontractivity of the noise operator

Define the normalised p-norm of f by

$$\|f\|_p = \Big(\frac{1}{2^n} \sum_{x \in \{0,1\}^n} |f(x)|^p\Big)^{1/p}.$$

This family of norms is non-decreasing with p.

However, we have the following (non-trivial!) inequality.

Bonami-Gross-Beckner hypercontractive inequality

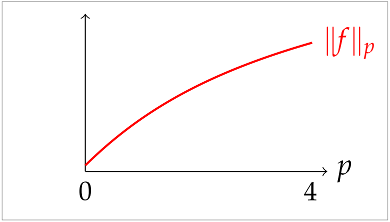
Let $f : \{0,1\}^n \to \mathbb{R}$ be a function on the boolean cube. Then, for any $1 \leqslant p \leqslant q$, provided that $\rho \leqslant \sqrt{\frac{p-1}{q-1}}$, we have

$$\|\mathcal{D}_\rho f\|_q \leqslant \|f\|_p.$$

In other words, noise smoothes f out in a formal sense: note that if f is constant, $\|f\|_p$ is constant wrt p.

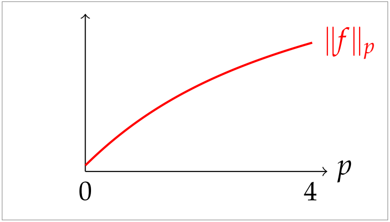
Hypercontractivity of the noise operator

p-norms of a random function f increase with p:

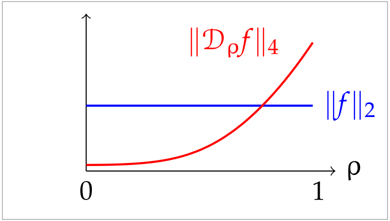


Hypercontractivity of the noise operator

p-norms of a random function f increase with p:



Applying noise smooths f by reducing its higher norms:



Why should we care?

Applications! For example, the KKL Lemma follows from:

Different norms of low-degree polynomials are close

Let $f : \{0,1\}^n \to \mathbb{R}$ be a function on the boolean cube with degree at most d. Then, for any $q \geqslant 2$, $\|f\|_q \leqslant (q-1)^{d/2} \|f\|_2$.

Why should we care?

Applications! For example, the KKL Lemma follows from:

Different norms of low-degree polynomials are close

Let $f : \{0,1\}^n \to \mathbb{R}$ be a function on the boolean cube with degree at most d. Then, for any $q \geqslant 2$, $\|f\|_q \leqslant (q-1)^{d/2} \|f\|_2$.

Armed with the hypercontractive inequality, the proof is simple. Writing $f^{=k} = \sum_{S,|S|=k} \hat{f}(S) \chi_S$,

$$\|f\|_q^2 \quad = \quad \left\| \sum_{k=0}^{d} f^{=k} \right\|_q^2$$

Why should we care?

Applications! For example, the KKL Lemma follows from:

Different norms of low-degree polynomials are close

Let $f : \{0,1\}^n \to \mathbb{R}$ be a function on the boolean cube with degree at most d. Then, for any $q \geqslant 2$, $\|f\|_q \leqslant (q-1)^{d/2}\|f\|_2$.

Armed with the hypercontractive inequality, the proof is simple. Writing $f^{=k} = \sum_{S,|S|=k} \hat{f}(S)\chi_S$,

$$\|f\|_q^2 \;=\; \left\|\sum_{k=0}^{d} f^{=k}\right\|_q^2 = \left\|\mathcal{D}_{1/\sqrt{q-1}}\left(\sum_{k=0}^{d}(q-1)^{k/2} f^{=k}\right)\right\|_q^2$$

Why should we care?

Applications! For example, the KKL Lemma follows from:

Different norms of low-degree polynomials are close

Let $f : \{0,1\}^n \to \mathbb{R}$ be a function on the boolean cube with degree at most d. Then, for any $q \geqslant 2$, $\|f\|_q \leqslant (q-1)^{d/2}\|f\|_2$.

Armed with the hypercontractive inequality, the proof is simple. Writing $f^{=k} = \sum_{S,|S|=k} \hat{f}(S)\chi_S$,

$$\begin{aligned} \|f\|_q^2 &= \left\| \sum_{k=0}^{d} f^{=k} \right\|_q^2 = \left\| \mathcal{D}_{1/\sqrt{q-1}} \left(\sum_{k=0}^{d} (q-1)^{k/2} f^{=k} \right) \right\|_q^2 \\ &\leqslant \left\| \sum_{k=0}^{d} (q-1)^{k/2} f^{=k} \right\|_2^2 \end{aligned}$$

Why should we care?

Applications! For example, the KKL Lemma follows from:

Different norms of low-degree polynomials are close

Let $f : \{0,1\}^n \to \mathbb{R}$ be a function on the boolean cube with degree at most d. Then, for any $q \geqslant 2$, $\|f\|_q \leqslant (q-1)^{d/2}\|f\|_2$.

Armed with the hypercontractive inequality, the proof is simple. Writing $f^{=k} = \sum_{S,|S|=k} \hat{f}(S)\chi_S$,

$$\begin{aligned}
\|f\|_q^2 &= \left\| \sum_{k=0}^{d} f^{=k} \right\|_q^2 = \left\| \mathcal{D}_{1/\sqrt{q-1}} \left(\sum_{k=0}^{d} (q-1)^{k/2} f^{=k} \right) \right\|_q^2 \\
&\leqslant \left\| \sum_{k=0}^{d} (q-1)^{k/2} f^{=k} \right\|_2^2 = \sum_{k=0}^{d} (q-1)^k \sum_{S \subseteq [n], |S|=k} \hat{f}(S)^2
\end{aligned}$$

Why should we care?

Applications! For example, the KKL Lemma follows from:

Different norms of low-degree polynomials are close

Let $f : \{0,1\}^n \to \mathbb{R}$ be a function on the boolean cube with degree at most d. Then, for any $q \geqslant 2$, $\|f\|_q \leqslant (q-1)^{d/2}\|f\|_2$.

Armed with the hypercontractive inequality, the proof is simple. Writing $f^{=k} = \sum_{S,|S|=k} \hat{f}(S)\chi_S$,

$$\begin{aligned}
\|f\|_q^2 &= \left\|\sum_{k=0}^{d} f^{=k}\right\|_q^2 = \left\|\mathcal{D}_{1/\sqrt{q-1}}\left(\sum_{k=0}^{d}(q-1)^{k/2}f^{=k}\right)\right\|_q^2 \\
&\leqslant \left\|\sum_{k=0}^{d}(q-1)^{k/2}f^{=k}\right\|_2^2 = \sum_{k=0}^{d}(q-1)^k \sum_{S\subseteq[n],|S|=k}\hat{f}(S)^2 \\
&\leqslant (q-1)^d \sum_{S\subseteq[n]}\hat{f}(S)^2
\end{aligned}$$

Why should we care?

Applications! For example, the KKL Lemma follows from:

Different norms of low-degree polynomials are close

Let $f : \{0,1\}^n \to \mathbb{R}$ be a function on the boolean cube with degree at most d. Then, for any $q \geqslant 2$, $\|f\|_q \leqslant (q-1)^{d/2}\|f\|_2$.

Armed with the hypercontractive inequality, the proof is simple. Writing $f^{=k} = \sum_{S,|S|=k} \hat{f}(S)\chi_S$,

$$
\begin{aligned}
\|f\|_q^2 &= \left\| \sum_{k=0}^{d} f^{=k} \right\|_q^2 = \left\| \mathcal{D}_{1/\sqrt{q-1}} \left(\sum_{k=0}^{d} (q-1)^{k/2} f^{=k} \right) \right\|_q^2 \\
&\leqslant \left\| \sum_{k=0}^{d} (q-1)^{k/2} f^{=k} \right\|_2^2 = \sum_{k=0}^{d} (q-1)^k \sum_{S\subseteq[n],|S|=k} \hat{f}(S)^2 \\
&\leqslant (q-1)^d \sum_{S\subseteq[n]} \hat{f}(S)^2 = (q-1)^d \|f\|_2^2.
\end{aligned}
$$

A generalisation of Fourier analysis

- We would like to generalise these classical results to a "truly quantum" (noncommutative) setting.

A generalisation of Fourier analysis

- We would like to generalise these classical results to a "truly quantum" (noncommutative) setting.

- Our generalisation (others are possible): instead of decomposing functions $f : \{0, 1\}^n \to \mathbb{R}$, we decompose Hermitian operators on the space of n qubits.

- It turns out that a natural analogue of the characters of $\mathbb{Z}_2$ are the Pauli matrices.

“Fourier analysis” for qubits

Write

$$\sigma^0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \ \sigma^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \ \sigma^2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \ \text{and } \sigma^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

“Fourier analysis” for qubits

Write

$$\sigma^0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \; \sigma^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \; \sigma^2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \text{ and } \sigma^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

We write a tensor product of Paulis as
$\chi_{\mathbf{s}} := \sigma^{s_1} \otimes \sigma^{s_2} \otimes \cdots \otimes \sigma^{s_n}$, where $s_j \in \{0, 1, 2, 3\}$.

"Fourier analysis" for qubits

Write

$$\sigma^0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \; \sigma^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \; \sigma^2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \text{ and } \sigma^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

We write a tensor product of Paulis as
$\chi_{\mathbf{s}} := \sigma^{s_1} \otimes \sigma^{s_2} \otimes \cdots \otimes \sigma^{s_n}$, where $s_j \in \{0, 1, 2, 3\}$.

Any n qubit Hermitian operator f has an expansion

$$f = \sum_{\mathbf{s} \in \{0,1,2,3\}^n} \hat{f}_{\mathbf{s}} \, \chi_{\mathbf{s}}.$$

for some real $\{\hat{f}_{\mathbf{s}}\}$ – the Pauli coefficients of f. This is our analogue of the Fourier expansion of a function $f : \{0, 1\}^n \to \mathbb{R}$.

Note that f is a k-local operator if $\max\{|\mathbf{s}| : \hat{f}_{\mathbf{s}} \neq 0\} \leqslant k$.

A quantum noise operator

The right quantum generalisation of the noise operator turns out to be the qubit depolarising channel!

A quantum noise operator

The right quantum generalisation of the noise operator turns out to be the qubit depolarising channel!

- Let $\mathcal{D}_\epsilon$ be the qubit depolarising channel with noise rate $1 - \epsilon$, i.e.

$$\mathcal{D}_\epsilon(\rho) = \frac{(1-\epsilon)}{2}\,\mathrm{tr}(\rho)\mathbb{I} + \epsilon\,\rho.$$

A quantum noise operator

The right quantum generalisation of the noise operator turns out to be the qubit depolarising channel!

- Let $\mathcal{D}_\epsilon$ be the qubit depolarising channel with noise rate $1-\epsilon$, i.e.

$$\mathcal{D}_\epsilon(\rho) = \frac{(1-\epsilon)}{2}\,\mathrm{tr}(\rho)\mathbb{I} + \epsilon\,\rho.$$

- Then

$$\mathcal{D}_\epsilon^{\otimes n}(\rho) = \sum_{\mathbf{s}\in\{0,1,2,3\}^n} \epsilon^{|\mathbf{s}|}\,\hat{\rho}_{\mathbf{s}}\,\chi_{\mathbf{s}}.$$

(this connection goes back at least a decade [Bruss et al ’99], and was used in [Kempe et al ’08] to give upper bounds on fault-tolerance thresholds)

A quantum noise operator

The right quantum generalisation of the noise operator turns out to be the qubit depolarising channel!

- Let $\mathcal{D}_\epsilon$ be the qubit depolarising channel with noise rate $1-\epsilon$, i.e.

$$\mathcal{D}_\epsilon(\rho) = \frac{(1-\epsilon)}{2}\,\mathrm{tr}(\rho)\mathbb{I} + \epsilon\,\rho.$$

- Then

$$\mathcal{D}_\epsilon^{\otimes n}(\rho) = \sum_{\mathbf{s}\in\{0,1,2,3\}^n} \epsilon^{|\mathbf{s}|}\,\hat{\rho}_{\mathbf{s}}\,\chi_{\mathbf{s}}.$$

(this connection goes back at least a decade [Bruss et al '99], and was used in [Kempe et al '08] to give upper bounds on fault-tolerance thresholds)

Can we prove an equivalent hypercontractive result for this channel?

Quantum hypercontractivity

Theorem

Let H be a Hermitian operator on n qubits and assume that $1 \leqslant p \leqslant 2 \leqslant q$. Then, provided that $\epsilon \leqslant \sqrt{\frac{p-1}{q-1}}$, we have

$$\|\mathcal{D}_\epsilon^{\otimes n}(H)\|_q \leqslant \|H\|_p.$$

Quantum hypercontractivity

Theorem

Let H be a Hermitian operator on n qubits and assume that $1 \leqslant p \leqslant 2 \leqslant q$. Then, provided that $\epsilon \leqslant \sqrt{\frac{p-1}{q-1}}$, we have

$$\|\mathcal{D}_\epsilon^{\otimes n}(H)\|_q \leqslant \|H\|_p.$$

- The proof relies on the Pauli expansion and a non-commutative generalisation of Hanner's inequality by King.
- It isn't a simple generalisation of the classical proof, but would be if the maximum output $p \to q$ norm were multiplicative!

"Application": Spectra of k-local operators

The proof of the classical corollary of the hypercontractive inequality goes through without change.

Different norms of k-local operators are close

Let H be a k-local Hermitian operator on n qubits. Then, for any $q \geqslant 2$, $\|H\|_q \leqslant (q-1)^{k/2}\|H\|_2$.

"Application": Spectra of k-local operators

The proof of the classical corollary of the hypercontractive inequality goes through without change.

Different norms of k-local operators are close

Let H be a k-local Hermitian operator on n qubits. Then, for any $q \geqslant 2$, $\|H\|_q \leqslant (q-1)^{k/2}\|H\|_2$.

This easily implies the following bound.

Spectral concentration for k-local operators

Let H be a k-local Hermitian operator on n qubits with eigenvalues (λ_i) and $\|H\|_2 = 1$. Then, for any $t \geqslant (2e)^{k/2}$,

$$\Pr[|\lambda_i| \geqslant t] \leqslant \exp(-kt^{2/k}/(2e)).$$

"Application": Spectra of k-local operators

The proof of the classical corollary of the hypercontractive inequality goes through without change.

Different norms of k-local operators are close

Let H be a k-local Hermitian operator on n qubits. Then, for any $q \geqslant 2$, $\|H\|_q \leqslant (q-1)^{k/2} \|H\|_2$.

This easily implies the following bound.

Spectral concentration for k-local operators

Let H be a k-local Hermitian operator on n qubits with eigenvalues (λ_i) and $\|H\|_2 = 1$. Then, for any $t \geqslant (2e)^{k/2}$,

$$\Pr[|\lambda_i| \geqslant t] \leqslant \exp(-kt^{2/k}/(2e)).$$

Note that we have not constrained the topology of H's k-locality at all. Stronger results can be proven (e.g. [Hartmann et al '04]'s "central limit theorem") with additional constraints.

Conclusions

- Fourier analysis on the boolean cube is a powerful technique in classical computer science which is now finding applications in quantum computation. Fourier analysis can be generalised to the quantum regime.

Conclusions

- Fourier analysis on the boolean cube is a powerful technique in classical computer science which is now finding applications in quantum computation. Fourier analysis can be generalised to the quantum regime.
- Can there be any asymptotic separation between quantum and classical 1WCC for a total function?
- Can we find any (real!) applications of quantum hypercontractivity? e.g. quantum k-SAT, fault tolerance, …
- There are many results in the classical theory of boolean functions which might be generalisable to the quantum regime.

Thanks!

arXiv:1007.3587v3

arXiv:0810.2435 (joint work with Tobias Osborne)

More formally

For any distribution $\mathcal{D}$ on Alice and Bob's inputs, let $\mathcal{D}^S$ be the induced distribution on Bob's inputs, given that Alice's input was in set S.

Lemma (e.g. [Gavinsky et al '08])

- Let $f : \{0,1\}^m \times \{0,1\}^n \to \{0,1\}$ be a function of Alice and Bob's distributed inputs.

More formally

For any distribution $\mathcal{D}$ on Alice and Bob's inputs, let $\mathcal{D}^S$ be the induced distribution on Bob's inputs, given that Alice's input was in set S.

Lemma (e.g. [Gavinsky et al '08])

- Let $f : \{0,1\}^m \times \{0,1\}^n \to \{0,1\}$ be a function of Alice and Bob's distributed inputs.
- Let $\mathcal{D}_0$, $\mathcal{D}_1$ be distributions on the zero/one-valued inputs, respectively, that are each uniform over Alice's inputs, when averaged over Bob's inputs.

More formally

For any distribution $\mathcal{D}$ on Alice and Bob's inputs, let $\mathcal{D}^S$ be the induced distribution on Bob's inputs, given that Alice's input was in set S.

Lemma (e.g. [Gavinsky et al '08])

- Let $f : \{0,1\}^m \times \{0,1\}^n \to \{0,1\}$ be a function of Alice and Bob's distributed inputs.
- Let $\mathcal{D}_0$, $\mathcal{D}_1$ be distributions on the zero/one-valued inputs, respectively, that are each uniform over Alice's inputs, when averaged over Bob's inputs.
- Assume there is a one-way classical protocol that computes f with success probability $1 - \epsilon$, for some $\epsilon < 1/3$, and uses c bits of communication.

More formally

For any distribution $\mathcal{D}$ on Alice and Bob's inputs, let $\mathcal{D}^S$ be the induced distribution on Bob's inputs, given that Alice's input was in set S.

Lemma (e.g. [Gavinsky et al '08])

- Let $f : \{0,1\}^m \times \{0,1\}^n \to \{0,1\}$ be a function of Alice and Bob's distributed inputs.
- Let $\mathcal{D}_0$, $\mathcal{D}_1$ be distributions on the zero/one-valued inputs, respectively, that are each uniform over Alice's inputs, when averaged over Bob's inputs.
- Assume there is a one-way classical protocol that computes f with success probability $1 - \epsilon$, for some $\epsilon < 1/3$, and uses c bits of communication.
- Then there exists $S \subseteq \{0,1\}^m$ such that $|S| \geqslant \epsilon\, 2^{m-c}$, and $\|\mathcal{D}_0^S - \mathcal{D}_1^S\|_1 \geqslant 2(1 - 3\epsilon)$.

Relation to previous work

This is equivalent to the following problem.

PM-Invariance

- Alice gets a $2n$-bit string x.
- Bob gets an $n \times 2n$ matrix M over $\mathbb{F}_2$, where each row contains exactly two 1s, and each column contains at most one 1.
- Bob has to output $\begin{cases} 0 & \text{if } Mx = 0 \\ 1 & \text{if } |Mx| \geqslant n/16 \\ \text{anything} & \text{otherwise.} \end{cases}$

Relation to previous work

A similar problem was used by [Gavinsky et al '08] to separate quantum and classical 1WCC.

α-Partial Matching

- Alice gets an n-bit string x.
- Bob gets an $\alpha n \times n$ matrix M over $\mathbb{F}_2$, where each row contains exactly two 1s, and each column contains at most one 1, and a string $w \in \{0,1\}^{\alpha n}$.
- Bob has to output $\begin{cases} 0 & \text{if } Mx = w \\ 1 & \text{if } Mx = \bar{w} \\ \text{anything} & \text{otherwise.} \end{cases}$

So the main difference is the relaxation of the promise by removing this second string from Bob's input.

Proof sketch

- The proof is by induction on n. The case $n = 1$ follows immediately from the classical proof.

[1] C. King, "Inequalities for trace norms of 2x2 block matrices", 2003

Proof sketch

- The proof is by induction on n. The case $n = 1$ follows immediately from the classical proof.
- For $n > 1$, expand ρ as $\rho = \mathbb{I} \otimes a + \sigma^1 \otimes b + \sigma^2 \otimes c + \sigma^3 \otimes d$, and write it as a block matrix.

[1] C. King, “Inequalities for trace norms of 2x2 block matrices”, 2003

Proof sketch

- The proof is by induction on n. The case $n = 1$ follows immediately from the classical proof.
- For $n > 1$, expand ρ as $\rho = \mathbb{I} \otimes a + \sigma^1 \otimes b + \sigma^2 \otimes c + \sigma^3 \otimes d$, and write it as a block matrix.
- Using a non-commutative Hanner's inequality for block matrices[1], can bound $\|\mathcal{D}_\epsilon^{\otimes n}(\rho)\|_q$ in terms of the norm of a 2×2 matrix whose entries are the norms of the blocks of $\mathcal{D}_\epsilon^{\otimes n}(\rho)$.

[1] C. King, "Inequalities for trace norms of 2x2 block matrices", 2003

Proof sketch

- The proof is by induction on n. The case $n = 1$ follows immediately from the classical proof.
- For $n > 1$, expand ρ as $\rho = \mathbb{I} \otimes a + \sigma^1 \otimes b + \sigma^2 \otimes c + \sigma^3 \otimes d$, and write it as a block matrix.
- Using a non-commutative Hanner's inequality for block matrices[1], can bound $\|\mathcal{D}_\epsilon^{\otimes n}(\rho)\|_q$ in terms of the norm of a 2×2 matrix whose entries are the norms of the blocks of $\mathcal{D}_\epsilon^{\otimes n}(\rho)$.
- Bound the norms of these blocks using the inductive hypothesis.

[1] C. King, "Inequalities for trace norms of 2x2 block matrices", 2003

Proof sketch

- The proof is by induction on n. The case $n = 1$ follows immediately from the classical proof.
- For $n > 1$, expand ρ as $\rho = \mathbb{I} \otimes a + \sigma^1 \otimes b + \sigma^2 \otimes c + \sigma^3 \otimes d$, and write it as a block matrix.
- Using a non-commutative Hanner's inequality for block matrices[1], can bound $\|\mathcal{D}_\epsilon^{\otimes n}(\rho)\|_q$ in terms of the norm of a 2×2 matrix whose entries are the norms of the blocks of $\mathcal{D}_\epsilon^{\otimes n}(\rho)$.
- Bound the norms of these blocks using the inductive hypothesis.
- The hypercontractive inequality for the base case $n = 1$ then gives an upper bound for this 2×2 matrix norm.

[1] C. King, "Inequalities for trace norms of 2x2 block matrices", 2003