Limitations on quantum dimensionality reduction

Aram Harrow[1,2], **Ashley Montanaro**[3] and Tony Short[3]

[1]Department of Computer Science & Engineering, University of Washington

[2]Department of Mathematics, University of Bristol

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

arXiv:1012.2262

Engineering and Physical Sciences Research Council

Dimensionality reduction

- High-dimensional data is ubiquitous in computer science.
- However, algorithms operating on such data are often inefficient (e.g. having runtime exponential in the dimension).
- This problem can be mitigated using a result known as the Johnson-Lindenstrauss Lemma.

Dimensionality reduction

- High-dimensional data is ubiquitous in computer science.
- However, algorithms operating on such data are often inefficient (e.g. having runtime exponential in the dimension).
- This problem can be mitigated using a result known as the Johnson-Lindenstrauss Lemma.

JL Lemma

Given a set S of n d-dimensional real vectors, there is a linear map $\mathcal{E} : \mathbb{R}^d \to \mathbb{R}^{O(\log n/\epsilon^2)}$ that preserves all Euclidean distances in S, up to a multiple of $1 - \epsilon$. Further, there is an efficient randomised algorithm to find and implement $\mathcal{E}$.

Dimensionality reduction

The JL Lemma is in fact a corollary of the following result.

JL Lemma

For all dimensions d, e, there is a distribution $\mathcal{D}$ over linear maps $\mathcal{E} : \mathbb{R}^d \to \mathbb{R}^e$ such that, for all real vectors v, w,

$$\Pr_{\mathcal{E} \sim \mathcal{D}} [(1-\epsilon)\|v-w\|_2 \leqslant \|\mathcal{E}(v)-\mathcal{E}(w)\|_2 \leqslant \|v-w\|_2] \geqslant 1-\exp(-\Omega(\epsilon^2 e)).$$

Such a distribution $\mathcal{D}$ is known as an embedding with distortion $1/(1-\epsilon)$.

Dimensionality reduction

The JL Lemma is in fact a corollary of the following result.

JL Lemma

For all dimensions d, e, there is a distribution $\mathcal{D}$ over linear maps $\mathcal{E} : \mathbb{R}^d \to \mathbb{R}^e$ such that, for all real vectors v, w,

$$\Pr_{\mathcal{E}\sim\mathcal{D}}[(1-\epsilon)\|v-w\|_2 \leqslant \|\mathcal{E}(v)-\mathcal{E}(w)\|_2 \leqslant \|v-w\|_2] \geqslant 1-\exp(-\Omega(\epsilon^2 e)).$$

Such a distribution $\mathcal{D}$ is known as an embedding with distortion $1/(1-\epsilon)$.

Note the following interesting aspects of this result:

- The target dimension e does not depend on the source dimension d at all.
- The distribution $\mathcal{D}$ does not depend on the vectors whose dimensionality is to be reduced.

Uses of the JL Lemma and other norm embeddings in quantum information theory

- [Cleve et al '04] used the JL Lemma to give an upper bound on the amount of shared entanglement required to win a class of nonlocal games.
- [Gavinsky, Kempe and de Wolf '06] used the JL Lemma to simulate arbitrary quantum communication protocols by quantum SMP protocols.
- [Aubrun, Szarek and Werner '10] have used a version of Dvoretzky's theorem on "almost-Euclidean" subspaces of matrices under Schatten norms to give counterexamples to the additivity conjectures of quantum information theory.
- [Fawzi, Hayden and Sen '10] have used embeddings of the "$\ell_1(\ell_2)$" norm to prove the existence of strong entropic uncertainty relations.

Quantum embeddings?

- We would like to generalise the JL Lemma to the quantum world. What should such a generalisation look like?

- Classically, an embedding is a distribution over linear maps which approximately preserves distances between vectors with high probability.

- Analogously, a quantum embedding should be a distribution over physically implementable operations which approximately preserves distances between quantum states with high probability.

Definitions

- (Mixed) quantum states are positive semidefinite matrices with trace 1.

Definitions

- (Mixed) quantum states are positive semidefinite matrices with trace 1.
- The distance between quantum states ρ, σ can be measured using the Schatten p-norm

$$\|\rho - \sigma\|_p = \left(\sum_i |\lambda_i(\rho - \sigma)|^p \right)^{1/p},$$

where $\lambda_i(X)$ is the i'th eigenvalue of X. The case $p = 1$ is known as the trace norm.

Definitions

- (Mixed) quantum states are positive semidefinite matrices with trace 1.
- The distance between quantum states ρ, σ can be measured using the Schatten p-norm

$$\|\rho - \sigma\|_p = \left(\sum_i |\lambda_i(\rho - \sigma)|^p \right)^{1/p},$$

  where $\lambda_i(X)$ is the i'th eigenvalue of X. The case $p = 1$ is known as the trace norm.
- Physically implementable operations in quantum theory (maps taking quantum states to quantum states) are known as quantum channels.

Definitions

- (Mixed) quantum states are positive semidefinite matrices with trace 1.
- The distance between quantum states ρ, σ can be measured using the Schatten p-norm

$$\|\rho - \sigma\|_p = \left(\sum_i |\lambda_i(\rho - \sigma)|^p \right)^{1/p},$$

  where $\lambda_i(X)$ is the i'th eigenvalue of X. The case $p = 1$ is known as the trace norm.
- Physically implementable operations in quantum theory (maps taking quantum states to quantum states) are known as quantum channels.
- A quantum channel is a completely positive, trace-preserving map from $\mathcal{B}(d)$ to $\mathcal{B}(e)$ (for some d, e), where $\mathcal{B}(d)$ denotes the set of d-dim. Hermitian matrices.

Quantum embeddings

Quantum embeddings

A quantum embedding from $S \subseteq \mathcal{B}(d)$ to $\mathcal{B}(e)$ in the Schatten p-norm, with distortion $1/(1-\epsilon)$ and failure probability δ, is a distribution $\mathcal{D}$ over quantum channels $\mathcal{E} : \mathcal{B}(d) \to \mathcal{B}(e)$ such that, for all $\rho, \sigma \in S$,

$$\Pr_{\mathcal{E}\sim\mathcal{D}}\left[(1-\epsilon)\|\rho-\sigma\|_p \leqslant \|\mathcal{E}(\rho)-\mathcal{E}(\sigma)\|_p \leqslant \|\rho-\sigma\|_p\right] \geqslant 1-\delta.$$

Quantum embeddings

Quantum embeddings

A quantum embedding from $S \subseteq \mathcal{B}(d)$ to $\mathcal{B}(e)$ in the Schatten p-norm, with distortion $1/(1-\epsilon)$ and failure probability δ, is a distribution $\mathcal{D}$ over quantum channels $\mathcal{E} : \mathcal{B}(d) \to \mathcal{B}(e)$ such that, for all $\rho, \sigma \in S$,

$$\Pr_{\mathcal{E}\sim\mathcal{D}}\left[(1-\epsilon)\|\rho-\sigma\|_p \leqslant \|\mathcal{E}(\rho)-\mathcal{E}(\sigma)\|_p \leqslant \|\rho-\sigma\|_p\right] \geqslant 1-\delta.$$

- Note that we generalise the definition of classical embeddings to allow embeddings that only work for subsets S of quantum states.
- We will be particularly interested in unitarily invariant subsets: sets S where $\rho \in S$ implies $U\rho U^\dagger \in S$ for all unitary operators U.
- An interesting such subset is the set of all d-dimensional pure states.

Results in this talk

1. Quantum dimensionality reduction in the 2-norm is very limited.

2. Two operational meanings of the 2-norm.

3. Dimensionality reduction in the trace norm: upper and lower bounds.

Dimensionality reduction in the 2-norm

Theorem

Let $\mathcal{D}$ be a distribution over quantum channels $\mathcal{E} : \mathcal{B}(\mathbb{C}^d) \to \mathcal{B}(\mathbb{C}^e)$ such that, for fixed quantum states $\rho \neq \sigma$ and for all unitary operators $U \in U(d)$,

$$\Pr_{\mathcal{E}\sim\mathcal{D}}[\|\mathcal{E}(U\rho U^\dagger) - \mathcal{E}(U\sigma U^\dagger)\|_2 \geqslant (1-\epsilon)\|U\rho U^\dagger - U\sigma U^\dagger\|_2] \geqslant 1-\delta$$

for some $0 \leqslant \epsilon, \delta \leqslant 1$. Then $e \geqslant (1-\delta)(1-\epsilon)^2 d$.

Corollary

Any embedding of a unitarily invariant set of states from d to e dimensions which has constant distortion in the 2-norm and succeeds with constant probability must satisfy $e = \Omega(d)$.

Proof idea

The theorem is essentially immediate from the following

Lemma

Let ρ and σ be quantum states and let $\mathcal{E} : \mathcal{B}(\mathbb{C}^d) \to \mathcal{B}(\mathbb{C}^e)$ be a quantum channel. Then

$$\int \|\mathcal{E}(U\rho U^\dagger) - \mathcal{E}(U\sigma U^\dagger)\|_2^2 \, dU \leqslant \frac{d(e^2-1)}{e(d^2-1)} \|\rho - \sigma\|_2^2.$$

Proof idea

The theorem is essentially immediate from the following

Lemma

Let ρ and σ be quantum states and let $\mathcal{E} : \mathcal{B}(\mathbb{C}^d) \to \mathcal{B}(\mathbb{C}^e)$ be a quantum channel. Then

$$\int \|\mathcal{E}(U\rho U^\dagger) - \mathcal{E}(U\sigma U^\dagger)\|_2^2 \, dU \leqslant \frac{d(e^2-1)}{e(d^2-1)} \|\rho - \sigma\|_2^2.$$

The left-hand side of this inequality can be written out explicitly in terms of the flip (swap) operator F as

$$\int \|\mathcal{E}(U\rho U^\dagger) - \mathcal{E}(U\sigma U^\dagger)\|_2^2 \, dU = \frac{\|\rho - \sigma\|_2^2}{d^2-1} \operatorname{tr}\left[F\,\mathcal{E}^{\otimes 2}\left(F - \frac{I_{d^2}}{d}\right)\right]$$

and bounded using a new (?) inequality $\operatorname{tr}[F\,\mathcal{E}^{\otimes 2}(F)] \leqslant de$.

Operational meanings of the 2-norm

How meaningful/useful is it to preserve the 2-norm distance between quantum states?

Operational meanings of the 2-norm

How meaningful/useful is it to preserve the 2-norm distance between quantum states?

- The trace distance $\|\rho - \sigma\|_1$ between two quantum states ρ and σ has the following operational interpretation.
- Imagine we are given a state promised to be either ρ or σ, and want to determine which is the case. The optimal success probability we can achieve is precisely $\frac{1}{2} + \frac{1}{4}\|\rho - \sigma\|_1$.

Operational meanings of the 2-norm

How meaningful/useful is it to preserve the 2-norm distance between quantum states?

- The trace distance $\|\rho - \sigma\|_1$ between two quantum states ρ and σ has the following operational interpretation.
- Imagine we are given a state promised to be either ρ or σ, and want to determine which is the case. The optimal success probability we can achieve is precisely $\frac{1}{2} + \frac{1}{4}\|\rho - \sigma\|_1$.
- It is less obvious that the 2-norm distance $\|\rho - \sigma\|_2$ has a nice operational interpretation, and this distance measure is usually only used for calculational simplicity.
- However, it turns out that there are (at least) two operational interpretations of this distance measure.

Equality testing without a reference frame

- We are given a description of two different states ρ and σ. An adversary prepares two systems in one of the states $\rho \otimes \rho$, $\sigma \otimes \sigma$, $\rho \otimes \sigma$ or $\sigma \otimes \rho$, with equal probability of each.

Equality testing without a reference frame

- We are given a description of two different states ρ and σ. An adversary prepares two systems in one of the states $\rho \otimes \rho$, $\sigma \otimes \sigma$, $\rho \otimes \sigma$ or $\sigma \otimes \rho$, with equal probability of each.
- He then applies an unknown unitary U to each system (i.e. he applies $U \otimes U$ to the joint state).

Equality testing without a reference frame

- We are given a description of two different states ρ and σ. An adversary prepares two systems in one of the states $\rho \otimes \rho$, $\sigma \otimes \sigma$, $\rho \otimes \sigma$ or $\sigma \otimes \rho$, with equal probability of each.
- He then applies an unknown unitary U to each system (i.e. he applies $U \otimes U$ to the joint state).
- Our task is to determine whether the two systems have the same state (i.e. were originally $\rho \otimes \rho$ or $\sigma \otimes \sigma$) or different states.
- This models equality testing in a two-party scenario in which the preparer and tester do not share a reference frame (local basis).

Equality testing without a reference frame

- We are given a description of two different states ρ and σ. An adversary prepares two systems in one of the states $\rho \otimes \rho$, $\sigma \otimes \sigma$, $\rho \otimes \sigma$ or $\sigma \otimes \rho$, with equal probability of each.
- He then applies an unknown unitary U to each system (i.e. he applies $U \otimes U$ to the joint state).
- Our task is to determine whether the two systems have the same state (i.e. were originally $\rho \otimes \rho$ or $\sigma \otimes \sigma$) or different states.
- This models equality testing in a two-party scenario in which the preparer and tester do not share a reference frame (local basis).

Theorem

The maximal probability of success is $\frac{1}{2} + \frac{1}{8}\|\rho - \sigma\|_2^2$.

State discrimination with a random measurement

- We are given a state which is promised to be either ρ or σ, with equal probability of each, and we wish to determine which is the case.
- We are allowed to perform a projective measurement in a random basis (i.e. to apply a random unitary operator and measure in the computational basis), and have to decide whether the state is ρ or σ based on the outcome.

State discrimination with a random measurement

- We are given a state which is promised to be either ρ or σ, with equal probability of each, and we wish to determine which is the case.
- We are allowed to perform a projective measurement in a random basis (i.e. to apply a random unitary operator and measure in the computational basis), and have to decide whether the state is ρ or σ based on the outcome.

Theorem

The expected optimal probability of success p satisfies

$$\frac{1}{2} + \frac{1}{6}\|\rho - \sigma\|_2 \leqslant p \leqslant \frac{1}{2} + \frac{1}{2}\|\rho - \sigma\|_2.$$

The lower bound was originally shown by [Ambainis and Emerson '07]; also see the proof by [Matthews, Wehner and Winter '09].

Dimensionality reduction in the trace norm: upper bound

The situation in the trace norm is somewhat better.

Theorem

For any pair of rank r quantum states ρ, σ, there is a distrib. $\mathcal{D}$ on quantum channels $\mathcal{E} : \mathcal{B}(\mathbb{C}^d) \to \mathcal{B}(\mathbb{C}^{O(\sqrt{rd/\epsilon})})$ such that

$$\Pr_{\mathcal{E}\sim\mathcal{D}}[\|\mathcal{E}(\rho) - \mathcal{E}(\sigma)\|_1 \geqslant (1-\epsilon)\|\rho - \sigma\|_1] \geqslant 1 - d\exp(-K\epsilon d)$$

for some universal constant K.

Dimensionality reduction in the trace norm: upper bound

The situation in the trace norm is somewhat better.

Theorem

For any pair of rank r quantum states ρ, σ, there is a distrib. $\mathcal{D}$ on quantum channels $\mathcal{E} : \mathcal{B}(\mathbb{C}^d) \to \mathcal{B}(\mathbb{C}^{O(\sqrt{rd/\epsilon})})$ such that

$$\Pr_{\mathcal{E}\sim\mathcal{D}}[\|\mathcal{E}(\rho) - \mathcal{E}(\sigma)\|_1 \geqslant (1-\epsilon)\|\rho - \sigma\|_1] \geqslant 1 - d\exp(-K\epsilon d)$$

for some universal constant K.

- Note that this is in contrast to the classical case, where dimensionality reduction in the ℓ_1 norm is known to be considerably harder than the ℓ_2 norm.
- [Winter '04] had previously shown this theorem for $r = 1$ (pure states) using essentially the same distribution $\mathcal{D}$.

Dimensionality reduction in the trace norm: upper bound

- The embedding we will use is conceptually simple: apply a random unitary and trace out (discard) a subsystem.

Dimensionality reduction in the trace norm: upper bound

- The embedding we will use is conceptually simple: apply a random unitary and trace out (discard) a subsystem.
- However, when the target dimension e doesn't divide d, we are forced to consider random isometries instead.
- Recall that an isometry is a norm-preserving linear map, i.e. a map taking an orthonormal basis of one space to an orthonormal set of vectors in another (potentially larger) space.

Dimensionality reduction in the trace norm: upper bound

- The embedding we will use is conceptually simple: apply a random unitary and trace out (discard) a subsystem.

- However, when the target dimension e doesn't divide d, we are forced to consider random isometries instead.

- Recall that an isometry is a norm-preserving linear map, i.e. a map taking an orthonormal basis of one space to an orthonormal set of vectors in another (potentially larger) space.

- The embedding thus consists of applying a fixed isometry that maps $\mathbb{C}^d \mapsto \mathbb{C}^e \otimes \mathbb{C}^{\lceil d/e \rceil}$, applying a random unitary operator, and discarding the second subsystem.

Proof idea

We extend the techniques of [Winter '04].

- Let V be the isometry that was randomly picked, and let $\mathcal{E}_V$ be the corresponding quantum channel implemented.
- In order for it to hold that $\|\mathcal{E}_V(\rho - \sigma)\|_1 \geqslant (1-\epsilon)\|\rho - \sigma\|_1$, it suffices to exhibit an operator M with $0 \leqslant M \leqslant I$ and

$$\mathrm{tr}[M(\mathcal{E}_V(\rho - \sigma))] \geqslant (1-\epsilon)\|\rho - \sigma\|_1/2.$$

Proof idea

We extend the techniques of [Winter '04].

- Let V be the isometry that was randomly picked, and let $\mathcal{E}_V$ be the corresponding quantum channel implemented.
- In order for it to hold that $\|\mathcal{E}_V(\rho - \sigma)\|_1 \geqslant (1-\epsilon)\|\rho - \sigma\|_1$, it suffices to exhibit an operator M with $0 \leqslant M \leqslant I$ and

$$\mathrm{tr}[M(\mathcal{E}_V(\rho - \sigma))] \geqslant (1-\epsilon)\|\rho - \sigma\|_1/2.$$

- To find such an operator, expand

$$V(\rho - \sigma)V^\dagger = \sum_{i \in S^+} \lambda_i |\psi_i\rangle\langle\psi_i| - \sum_{i \in S^-} \mu_i |\psi_i\rangle\langle\psi_i|,$$

  where $\lambda_i, \mu_i \geqslant 0$, and note that $P_V := \sum_{i \in S^+} |\psi_i\rangle\langle\psi_i|$ is the projector onto a random subspace.
- Now take $M = \mathrm{tr}_B P_V$.

Proof idea

- We have

$$\mathrm{tr}[M(\mathcal{E}_V(\rho-\sigma))] = \sum_{i\in S^+} \lambda_i \,\mathrm{tr}[M\,\mathrm{tr}_B\,|\psi_i\rangle\langle\psi_i|] - \sum_{i\in S^-} \mu_i \,\mathrm{tr}[M\,\mathrm{tr}_B\,|\psi_i\rangle\langle\psi_i|]$$

and want to show that this is high (whp) relative to $\|\rho-\sigma\|_1/2 = \sum_{i\in S^+} \lambda_i$.

Proof idea

- We have

$$\mathrm{tr}[M(\mathcal{E}_V(\rho-\sigma))] = \sum_{i\in S^+} \lambda_i \, \mathrm{tr}[M \, \mathrm{tr}_B \, |\psi_i\rangle\langle\psi_i|] - \sum_{i\in S^-} \mu_i \, \mathrm{tr}[M \, \mathrm{tr}_B \, |\psi_i\rangle\langle\psi_i|]$$

  and want to show that this is high (whp) relative to $\|\rho - \sigma\|_1/2 = \sum_{i\in S^+} \lambda_i$.

- By the definition of M, $\mathrm{tr}[M \, \mathrm{tr}_B \, |\psi_i\rangle\langle\psi_i|] = 1$ for all $i \in S^+$.

- What remains is to show that $\mathrm{tr}[M \, \mathrm{tr}_B \, |\psi_i\rangle\langle\psi_i|]$ is small (whp) for all $i \in S^-$.

Proof idea

- We have

$$\mathrm{tr}[M(\mathcal{E}_V(\rho-\sigma))] = \sum_{i\in S^+} \lambda_i \, \mathrm{tr}[M\, \mathrm{tr}_B |\psi_i\rangle\langle\psi_i|] - \sum_{i\in S^-} \mu_i \, \mathrm{tr}[M\, \mathrm{tr}_B |\psi_i\rangle\langle\psi_i|]$$

  and want to show that this is high (whp) relative to $\|\rho-\sigma\|_1/2 = \sum_{i\in S^+} \lambda_i$.

- By the definition of M, $\mathrm{tr}[M\, \mathrm{tr}_B |\psi_i\rangle\langle\psi_i|] = 1$ for all $i \in S^+$.

- What remains is to show that $\mathrm{tr}[M\, \mathrm{tr}_B |\psi_i\rangle\langle\psi_i|]$ is small (whp) for all $i \in S^-$.

- This can be done using concentration of measure ideas: the states $|\psi_i\rangle$, $i \in S^-$ are random, subject to the constraint $\mathrm{tr}[P_V|\psi_i\rangle\langle\psi_i|] = 0$.

- We end up with a bound that works as long as $e^2 \gtrsim rd$.

Dimensionality reduction in the trace norm: lower bound

Using the inequality $\|X\|_1 \leqslant \sqrt{e}\|X\|_2$ for e-dimensional operators X, the following result is essentially immediate from the 2-norm lower bound.

Theorem

Let $\mathcal{D}$ be a distribution over quantum channels $\mathcal{E} : \mathcal{B}(\mathbb{C}^d) \to \mathcal{B}(\mathbb{C}^e)$ such that, for fixed quantum states $\rho \neq \sigma$ and for all unitary U,

$$\Pr_{\mathcal{E}\sim\mathcal{D}}[\|\mathcal{E}(U\rho U^\dagger) - \mathcal{E}(U\sigma U^\dagger)\|_1 \geqslant (1-\epsilon)\|U\rho U^\dagger - U\sigma U^\dagger\|_1] \geqslant 1-\delta$$

for some $0 \leqslant \epsilon, \delta \leqslant 1$. Then $e \geqslant (1-\delta)(1-\epsilon)\sqrt{d}\frac{\|\rho-\sigma\|_1}{\|\rho-\sigma\|_2}$.

Dimensionality reduction in the trace norm: lower bound

In particular, we have the following corollary which implies that the upper bound is optimal for some sets of states.

Corollary

- Any embedding which preserves trace norm distances between any pair of rank r mixed states (up to a constant) must have target dimension $\Omega(\sqrt{rd})$.
- In particular, any embedding which preserves trace norm distances between any pair of d-dimensional pure states (up to a constant) must have target dimension $\Omega(\sqrt{d})$.

Summary

- Any embedding of a unitarily invariant set of d-dimensional quantum states that achieves constant distortion in the 2-norm must have target dimension $\Omega(d)$.
- In some situations (e.g. when the basis in which the states were prepared is unknown or the measurement apparatus does not depend on the states to be distinguished) the 2-norm is the “right” measure of distinguishability between quantum states.
- d-dimensional mixed states with rank r can be embedded in $O(\sqrt{rd})$ dimensions with constant distortion in the trace norm.

Open questions

- There is a gap between our upper and lower bounds for the trace norm. Can it be removed?

Open questions

- There is a gap between our upper and lower bounds for the trace norm. Can it be removed?
  - Tighter analysis of the embedding used in the upper bound might allow this...

Open questions

- There is a gap between our upper and lower bounds for the trace norm. Can it be removed?
  - Tighter analysis of the embedding used in the upper bound might allow this...
  - ...possibly at the expense of making the bound more complicated.

Open questions

- There is a gap between our upper and lower bounds for the trace norm. Can it be removed?
  - Tighter analysis of the embedding used in the upper bound might allow this...
  - ...possibly at the expense of making the bound more complicated.

- What is the situation when we have multiple copies of the input state or additional classical information?
  - For some results in this direction, see [Fawzi, Hayden and Sen '10].

Thanks!

arXiv:1012.2262