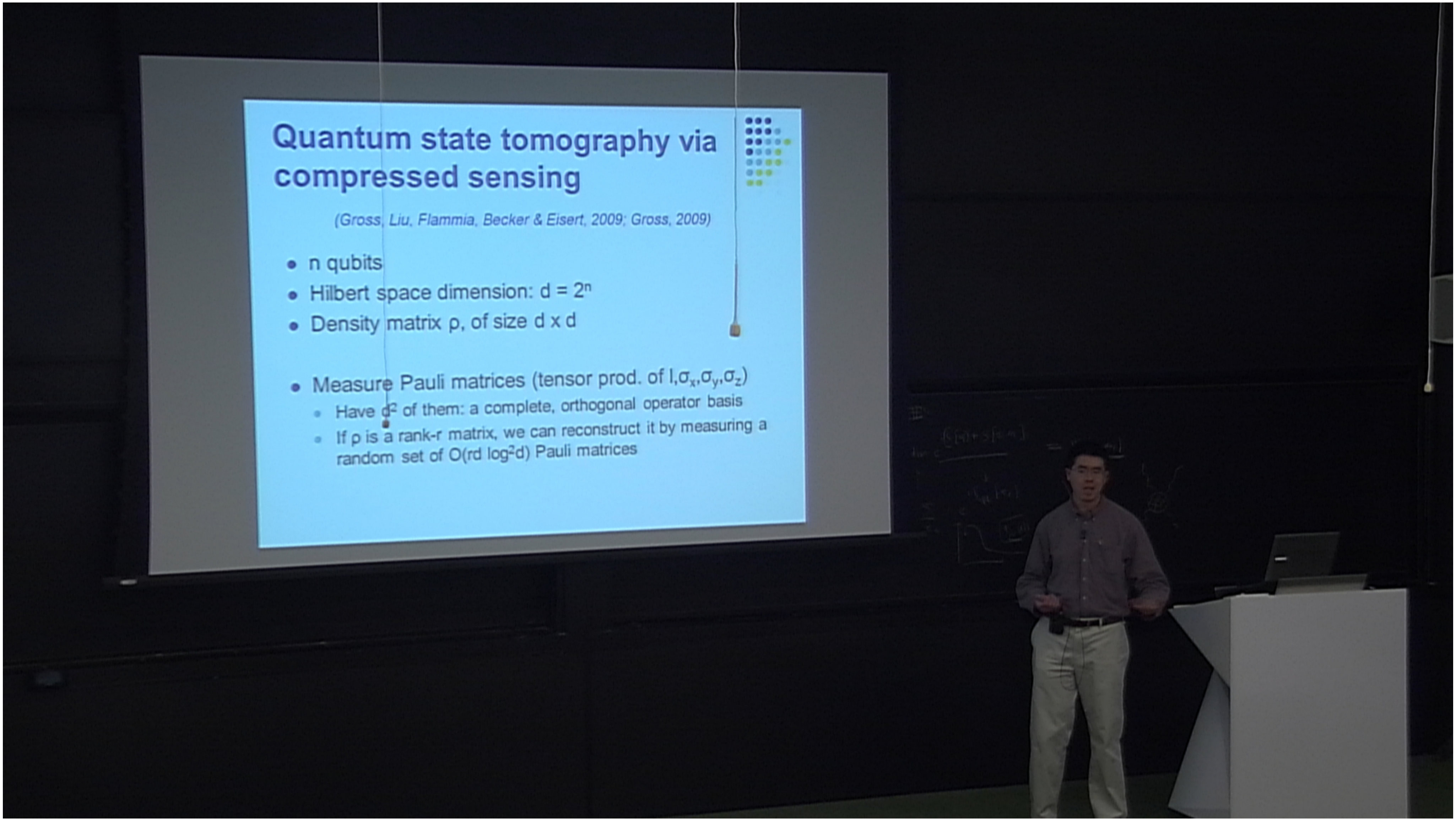
Title: Universal Low-rank Matrix Recovery from Pauli Measurements

Date: Apr 04, 2012 04:00 PM

URL: http://pirsa.org/12040101

Abstract: We study the problem of reconstructing an unknown matrix M, of rank r and dimension d, using O(rd poly log d) Pauli measurements. This has applications to compressed sensing methods for quantum state tomography. We give a solution to this problem based on the restricted isometry property (RIP), which improves on previous results using dual certificates. In particular, we show that almost all sets of O(rd log^6 d) Pauli measurements satisfy the rank-r RIP. This implies that M can be recovered from a fixed ("universal") set of Pauli measurements, using nuclear-norm minimization (e.g., the matrix Lasso), with nearly-optimal bounds on the error. Our proof uses Dudley's inequality for Gaussian processes, together with bounds on covering numbers obtained via entropy duality.



Quantum state tomography via compressed sensing
(Gross, Liu, Flammia, Becker & Eisert, 2009; Gross, 2009)
n qubits
Hilbert space dimension: d = 2ⁿ
Density matrix ρ, of size d x d
Measure Pauli matrices (tensor prod. of I,σx,σy,σz)
Have d² of them: a complete, orthogonal operator basis
If ρ is a rank-r matrix, we can reconstruct it by measuring a random set of O(rd log²d) Pauli matrices

"Universal" Low-rank Matrix Recovery using Pauli Measurements

Yi-Kai Liu
National Institute of Standards and Technology
Gaithersburg, MD, USA

Quantum state tomography via compressed sensing

(Gross, Liu, Flammia, Becker & Eisert, 2009; Gross, 2009)

- n qubits
- Hilbert space dimension: $d = 2^n$
- Density matrix ρ, of size d x d

- Measure Pauli matrices (tensor prod. of $I, \sigma_x, \sigma_y, \sigma_z$)
  - Have d^2 of them: a complete, orthogonal operator basis
  - If ρ is a rank-r matrix, we can reconstruct it by measuring a random set of $O(rd \log^2 d)$ Pauli matrices

Quantum state tomography via compressed sensing

(Gross, Liu, Flammia, Becker & Eisert, 2009; Gross, 2009)

- n qubits
- Hilbert space dimension: $d = 2^n$
- Density matrix ρ, of size d x d

- Measure Pauli matrices (tensor prod. of $I, \sigma_x, \sigma_y, \sigma_z$)
  - Have d^2 of them: a complete, orthogonal operator basis
  - If ρ is a rank-r matrix, we can reconstruct it by measuring a random set of $O(rd \log^2 d)$ Pauli matrices

Quantum state tomography via compressed sensing

(Gross, Liu, Flammia, Becker & Eisert, 2009; Gross, 2009)

- For any matrix ρ (of dimension d and rank r):
  - Choose a random set Ω of $O(rd \log^2 d)$ Pauli matrices
  - Then with high probability (over the choice of Ω),…
  - One can uniquely reconstruct ρ from the measurements Ω
    - Estimate the expectation values Tr(Pρ) (for all P in Ω)
    - Solve a convex program – takes time poly(d)

This talk: "universal" low-rank matrix recovery

(Liu, 2011)

- For any matrix ρ (of dimension d and rank r):
  - Choose a random set Ω of $O(rd \log^6 d)$ Pauli matrices
  - Then with high probability (over the choice of Ω),…
  - One can uniquely reconstruct ρ from the measurements Ω
    - Estimate the expectation values Tr(Pρ) (for all P in Ω)
    - Solve a convex program – takes time poly(d)
- Can fix the set Ω once and for all!
  - That Ω will work for every rank-r matrix ρ – it is "universal"
  - Actually, most choices of Ω will have this property!

Two different pictures of state space

- Original matrix completion results
  - "Dual certificates"
  - Local properties of state space around a point ρ
- This talk – "universal" matrix recovery
  - "Restricted isometry property" (RIP)
  - Global properties: whole state space can be embedded (w/ small distortion) into R^m, $m = O(rd \text{ polylog } d)$

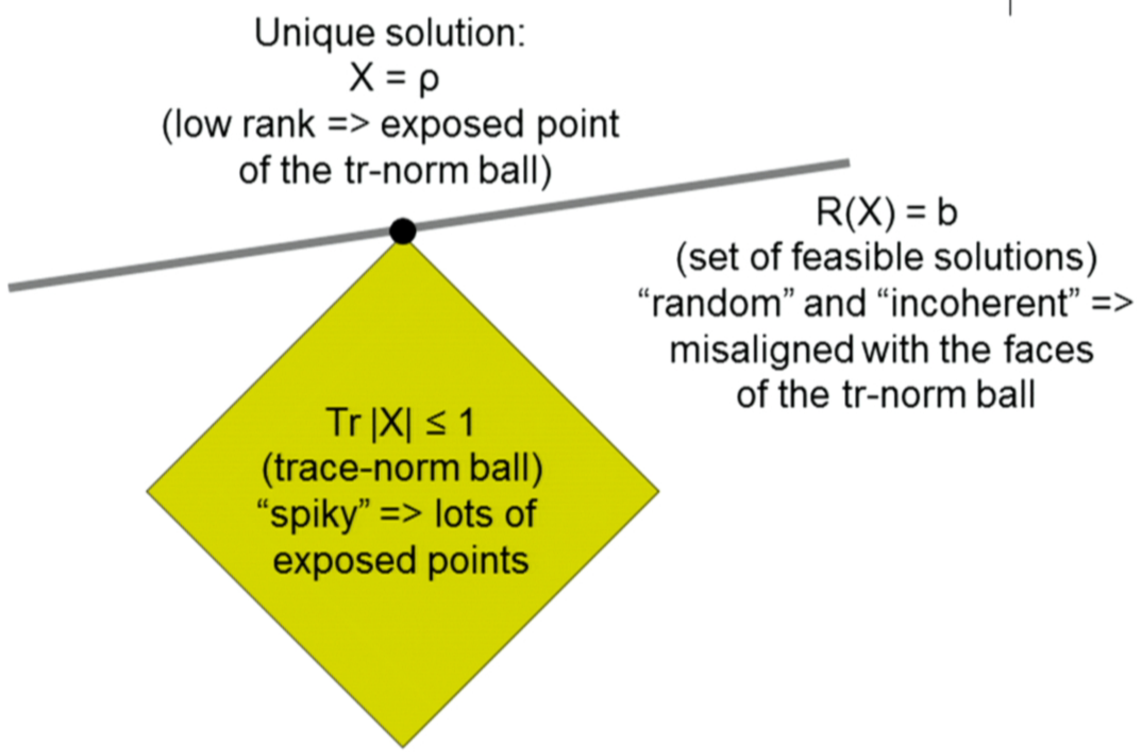
Reconstructing ρ

- ρ = unknown state, dim. d, rank $r << d$
- Ω = subset of Pauli operators, $|\Omega| = O(rd \text{ polylog } d)$

- $R(\rho) = [Tr(P\rho)]_{P \text{ in } \Omega}$, vector of Pauli expectation values
  - R = "sampling operator"
  - In a real experiment, after measuring P in Ω, we get $b \approx R(\rho)$

- Solve: $\text{argmin}_X \, Tr|X|$ s.t. $||R(X)-b||_2 \leq \varepsilon$, $X \geq 0$

What happens around ρ



What happens around ρ

- Hyperplane {X : R(X) = b} is "misaligned" with the faces of the trace-norm ball
  - Any perturbation X = ρ+δ either changes the value of R(X), or increases the trace norm of X
    - "Dual certificate"
- Key facts
  - Measurements are "incoherent": $||P|| \leq d^{-1/2} ||P||_2$
    - E.g., Pauli matrices, Gaussian random matrices
  - For each ρ, we choose a random hyperplane
    - It's likely to be good

A global picture

- Sampling operator $R(\rho) = [\mathrm{Tr}(P\rho)]_{P \text{ in } \Omega}$, $|\Omega| \sim rd \log^6 d$

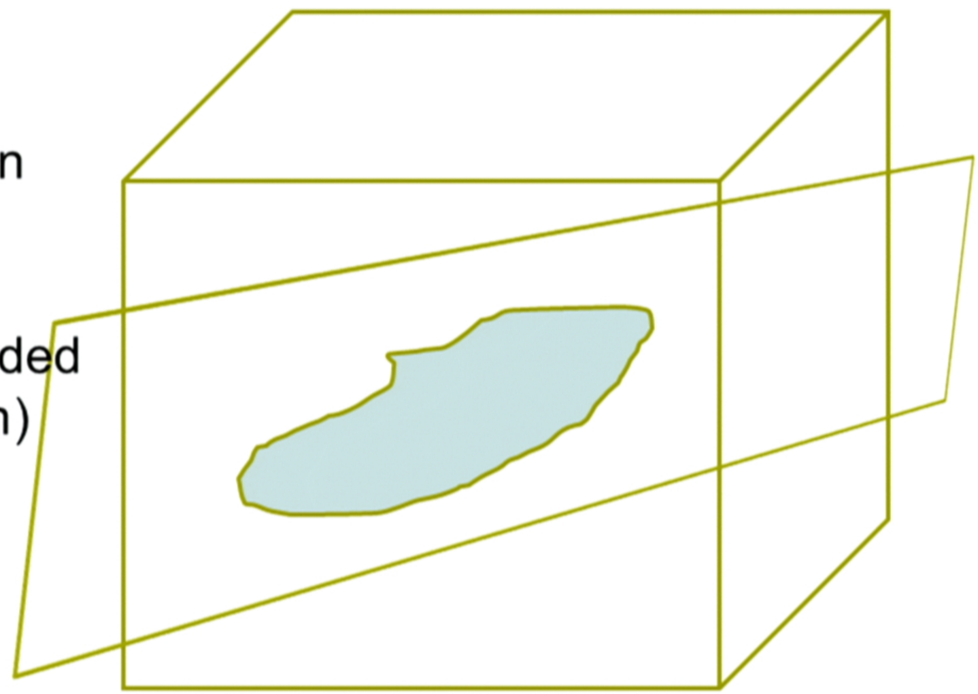
- Restricted isometry property (RIP) (w/ rank r, error δ): for all X with dim. d and rank r,

$$(1-\delta)\ ||X||_2 \leq ||R(X)||_2 \leq (1+\delta)\ ||X||_2$$

  - "Embedding the manifold of low-rank matrices into a low-dimensional linear space"
  - This implies universal low-rank matrix recovery

A global picture

- The manifold of pure states
  - A curved surface, w/ real dim. ~d
  - Naturally defined in Euclidean space w/ dim. d^2
  - But can be embedded (w/ minor distortion) in a subspace w/ dim. $O(d \log^6 d)$



A global picture

- Why is this embedding possible?
  - Measurements are "incoherent": $||P|| \le d^{-1/2} ||P||_2$
    - E.g., Pauli matrices, Gaussian random matrices
  - For any low-rank state, the Pauli coefficients are fairly uniform (not peaked)
    - So it's enough to sample a random subset of them
    - Hard part: showing that this is true "uniformly" over all low-rank states
    - Covering the trace-norm ball – "entropy argument"

The rest of this talk

- Why “universality” is useful
  - Error bounds, sample complexity of tomography
- Proof ideas
  - Entropy argument
- Compressed tomography for continuous-variable quantum systems
  - Homodyne detection

Error bounds for compressed tomography

(Liu, 2011)

- Reconstructing a full-rank state ρ
  - Intuition: if we measure $O(rd \log^6 d)$ Pauli's, we should be able to reconstruct the first r eigenvectors of ρ (call this ρ_r)
  - Theorem: we obtain an estimate σ such that $||\rho - \sigma||_2^2 \leq (\text{polylog } d)\, ||\rho - \rho_r||_2^2$
    - Much stronger than error bounds using dual certificate
    - Combining RIP result (Liu, 2011) with error bound from (Candes and Plan, 2011)

Sample complexity

(B. Brown, S. Flammia, D. Gross, Y. Liu, in preparation, 2011)

- How many copies of the unknown state ρ are needed?
  - Compressed tomography uses fewer measurement settings, but might have higher sample complexity
    - RIP implies that this is not so!
    - For pure states, as we vary m from d polylog d to d^2 : sample complexity is the same (up to log factors)
  - Practical consideration: in some devices, repeating a measurement is much faster than changing measurement settings

Sample complexity

(B. Brown, S. Flammia, D. Gross, Y. Liu, in preparation, 2011)

- How many copies of the unknown state ρ are needed?
  - Compressed tomography uses fewer measurement settings, but might have higher sample complexity
    - RIP implies that this is not so!
    - For pure states, as we vary m from d polylog d to d^2 : sample complexity is the same (up to log factors)
  - Practical consideration: in some devices, repeating a measurement is much faster than changing measurement settings

Sample complexity

(Brown, Flammia, Gross & Liu, in preparation)
(da Silva, Landon-Cardinal & Poulin, 2011; Flammia & Liu, 2011)

	Compressed tomography (unknown state is approx. low-rank)	**Fidelity estimation** (target state is pure)
# of parameters to be learned	O(rd)	1
# of Pauli operators ("meas. settings")	O(rd polylog d)	O(1)
# of copies of unknown state ("sample complexity")	$O(r^2d^2$ polylog d)	O(d)

Proof ideas

- Restricted isometry property (RIP)
- RIP implies low-rank matrix recovery
- Pauli measurements obey RIP

RIP implies low-rank matrix recovery (1)

- Proof ideas: *(Recht, Fazel & Parrilo, 2007)*
  - Unknown state ρ, has rank r
  - Measure R(ρ) (noiseless case)
  - Solve $X = \text{argmin}_X \, \text{Tr}|X|$ s.t. $R(X) = R(\rho)$
  - Let $S = X - \rho$, note $R(S) = 0$; want to show $S = 0$

  - Look at the tail of S:
  - $S = S_0 + S_c$ where S_0 has rank 2r, S_c has zero "overlap" with row and column spaces of ρ
    - X has minimum trace => $||S_c||_{tr} \leq ||S_0||_{tr} \leq (2r)^{1/2} \, ||S_0||_2$

RIP implies low-rank matrix recovery (2)

- Proof ideas:
  - Let $S = X - \rho$, note $R(S) = 0$; want to show $S = 0$
  - Write the tail of S as a sum of low-rank terms:
  - $S_c = \Sigma_{j\geq 1} S_j$ (blocks of 3r consecutive singular values)
    - "Telescoping series": $||S_j||_2 \leq (3r)^{-1/2} ||S_{j-1}||_{tr}$
    - Apply RIP term-by-term: $||R(S_j)||_2 \approx ||S_j||_2$
  - Combining these bounds, we get that $S = 0$

Operators that obey RIP

- Proof ideas:
  - Previous work: RIP for Gaussian random matrices: use "union bound" over all rank-r matrices *(Recht et al, 2007)*
  - Our work: RIP for random Pauli matrices: use "entropy argument" – improve on union bound, by keeping track of correlations *(Rudelson & Vershynin, 2006)*
  - Prove bounds on covering numbers, using entropy duality *(Guedon et al, 2008)*

(Liu, 2011)

Pauli measurements obey RIP (1)

- Let **R** be the random Pauli sampling operator
- Proof ideas:
- Random variables taking values in a Banach space
  - Consider self-adjoint linear operators **M**: $C^{dxd} \rightarrow C^{dxd}$
  - Define the norm $||\mathbf{M}||_{(r)} = \sup_{X \text{ in } U} |Tr(X^{+}\mathbf{M}(X))|$
  - $U = \{ X \text{ in } C^{dxd} \text{ s.t. } ||X||_2 \leq 1, \text{rank}(X) \leq r \}$
- We want to show that $||\mathbf{R}^*\mathbf{R} - \mathbf{1}||_{(r)} < 2\delta - \delta^2$
  - Construct **R** by sampling Pauli matrices iid at random
  - **R*R** is a sum of iid random variables
  - Bound **R*R** – **1** in expectation, then use tail bound

(Liu, 2011)

Pauli measurements obey RIP (2)

- Dudley's inequality:
  - Gaussian process: family of rv's $G(X)$ (for all X in U)
  - $U = \{ X \text{ in } C^{dxd} \text{ s.t. } ||X||_2 \leq 1, \text{rank}(X) \leq r \}$

- $E[\sup_{X \text{ in } U} G(X)] \leq (\text{const}) \cdot \int_{\varepsilon \geq 0} \log^{1/2} N(U,d_G,\varepsilon)\, d\varepsilon$
  - d_G is a metric: $d_G(X,Y) = (E[(G(X)-G(Y))^2])^{1/2}$ (measures strength of correlation b/w $G(X)$ and $G(Y)$)
  - $N(U,d_G,\varepsilon)$ is a covering number: # of balls of radius ε needed to cover U
  - Integrate over different scales $0 < \varepsilon < \infty$

(Liu, 2011)

Pauli measurements obey RIP (1)

- Let **R** be the random Pauli sampling operator
- Proof ideas:
- Random variables taking values in a Banach space
  - Consider self-adjoint linear operators **M**: $C^{dxd} \rightarrow C^{dxd}$
  - Define the norm $||\mathbf{M}||_{(r)} = \sup_{X \text{ in } U} |Tr(X^{+}\mathbf{M}(X))|$
  - $U = \{ X \text{ in } C^{dxd} \text{ s.t. } ||X||_2 \leq 1, \text{rank}(X) \leq r \}$
- We want to show that $||\mathbf{R^*R} - \mathbf{1}||_{(r)} < 2\delta - \delta^2$
  - Construct **R** by sampling Pauli matrices iid at random
  - **R*R** is a sum of iid random variables
  - Bound **R*R** – **1** in expectation, then use tail bound

(Liu, 2011)

Pauli measurements obey RIP (2)

- Dudley's inequality:
  - Gaussian process: family of rv's G(X) (for all X in U)
  - $U = \{ X \text{ in } C^{dxd} \text{ s.t. } \|X\|_2 \leq 1, \text{rank}(X) \leq r \}$
- $E[\sup_{X \text{ in } U} G(X)] \leq (\text{const}) \cdot \int_{\varepsilon \geq 0} \log^{1/2} N(U, d_G, \varepsilon)\, d\varepsilon$
  - d_G is a metric: $d_G(X,Y) = (E[(G(X)-G(Y))^2])^{1/2}$ (measures strength of correlation b/w G(X) and G(Y))
  - $N(U,d_G,\varepsilon)$ is a covering number: # of balls of radius ε needed to cover U
  - Integrate over different scales $0 < \varepsilon < \infty$

(Liu, 2011)

Pauli measurements obey RIP (3)

- Bounding the covering numbers $N(U, d_G, \varepsilon)$
  - Let B_1 be the trace-norm ball
  - Define a semi-norm on $C^{d\times d}$, $||M||_X = \max_{P \text{ in } \Omega} |Tr(P^+M)|$
  - Problem reduces to bounding $N(B_1, ||\cdot||_X, \varepsilon)$
  - Trivial bound:
    $N(B_1, ||\cdot||_X, \varepsilon) \leq$ (polynomial in $1/\varepsilon$, exponential in d^2)
  - Clever bound:
    $N(B_1, ||\cdot||_X, \varepsilon) \leq$ (exponential in $1/\varepsilon^2$, quasipolynomial in d)

(Liu, 2011)

Pauli measurements obey RIP (4)

- Bounding $N(B_1, ||\cdot||_X, \varepsilon)$ via entropy duality
  - Rewrite it as:
    $N[\ S : (C^{dxd}, \text{trace norm}) \rightarrow (C^m, L_\infty \text{ norm})\ ]$
  - This is related to the dual covering number:
    $N[\ S^* : (C^m, L_1 \text{ norm}) \rightarrow (C^{dxd}, \text{operator norm})\ ]$
  - Which we can bound by known techniques… (B. Maurey)

Continuous-variable systems

(Ohliger, Nesme, Gross, Liu & Eisert, 2011)

- Instead of an orthonormal operator basis, use a tight frame $\{w_a\}$ (w.r.t. a probability measure μ):

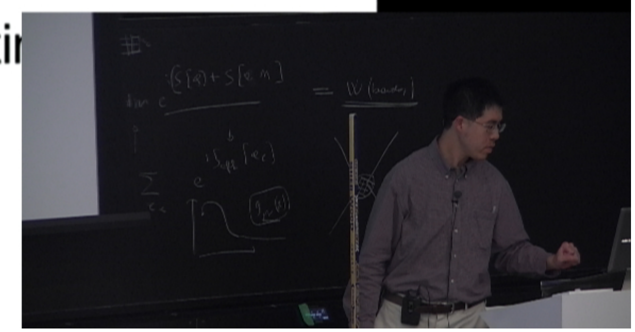
  $\rho = \int w_a \, \mathrm{Tr}(w_a^+ \rho) \, d\mu(a)$, for all ρ

- Incoherence condition: $||w_a|| \leq O(1)$

Continuous-variable systems

(Ohliger, Nesme, Gross, Liu & Eisert, 2011)

- Example: states with up to n photons (in a single mode)
  - Let the w_a be weighted displacement operators
    - Sample a from a Gaussian of width ~sqrt(n)
    - These form a tight frame
  - The w_a are incoherent!
    - Truncating to low-energy subspace
  - Expectation values $Tr(w_a^+\rho)$ can be esti using homodyne measurements
    - Fourier transform of the Wigner function



Continuous-variable systems

(Ohliger, Nesme, Gross, Liu & Eisert, 2011)

- Example: states with up to n photons (in a single mode)
  - Let the w_a be weighted displacement operators
    - Sample a from a Gaussian of width ~sqrt(n)
    - These form a tight frame
  - The w_a are incoherent!
    - Truncating to low-energy subspace
  - Expectation values $Tr(w_a^+ \rho)$ can be estimated using homodyne measurements
    - Fourier transform of the Wigner function

Some practical issues

(B. Brown, S. Flammia, D. Gross, Y. Liu, in preparation, 2011)

- Different estimators:
  - Trace min: $\mathrm{argmin}_X\ \mathrm{Tr}(X)$ s.t. $X \geq 0$, $\|R(X)-b\|_2 \leq \varepsilon$
  - Dantzig selector: $\mathrm{argmin}_X\ \mathrm{Tr}(X)$ s.t. $X \geq 0$, $\|R^*(R(X)-b)\| \leq \varepsilon$
  - Lasso: $\mathrm{argmin}_X\ \|R(X)-b\|_2^2 + \lambda \mathrm{Tr}(X)$ s.t. $X \geq 0$
  - Regularized MLE: $\mathrm{argmin}_X\ -\log L(X|b) + \lambda \mathrm{Tr}(X)$ s.t. $X \geq 0$
  - Other kinds of measurements (besides expectation values)?

?

Some practical issues

(B. Brown, S. Flammia, D. Gross, Y. Liu, in preparation, 2011)

- Different algorithms:
  - Interior-point SDP solvers
  - Singular-value thresholding
  - Gradient descent on the Grassmannian

Open questions

- Different motivations for compressed sensing?
  - Fewer quantum measurements?
  - Less classical postprocessing?
- Can we use these methods to do other things?