Universal Low-rank Matrix Recovery using Pauli Measurements

Yi-Kai Liu
Applied and Computational Mathematics, NIST

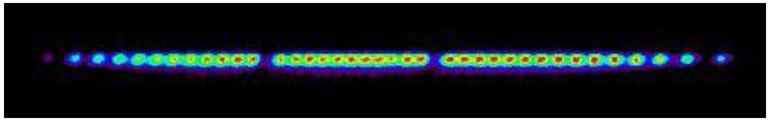
Joint work with: Steve Flammia, David Gross, Stephen Becker, Brielin Brown, Jens Eisert

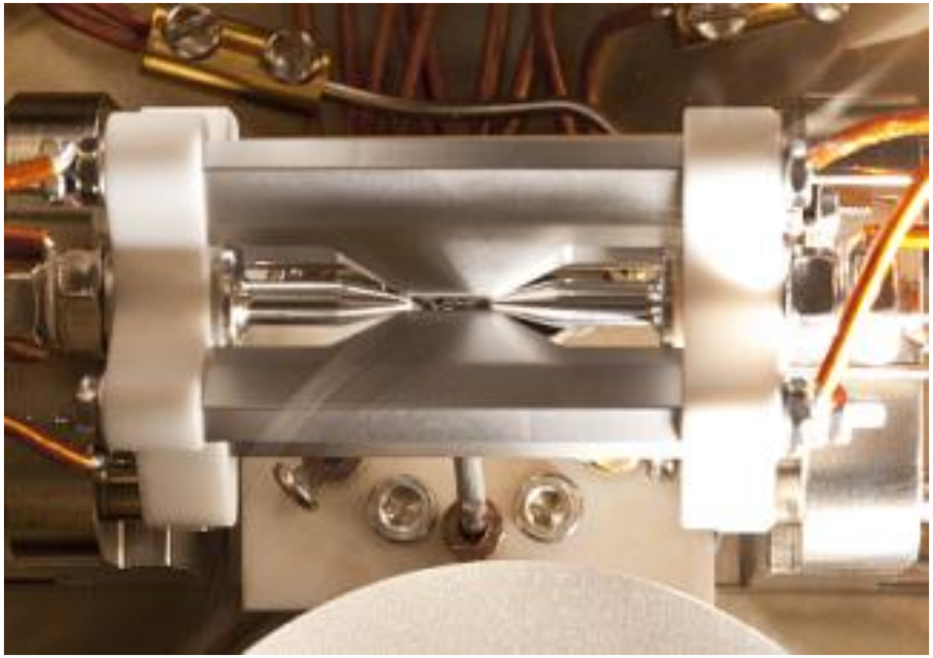
This talk

- A measurement problem: quantum state tomography
  - Solution using compressed sensing
- New result: “universal” low-rank matrix recovery
  - Why it works: geometric intuition
  - Proof ideas

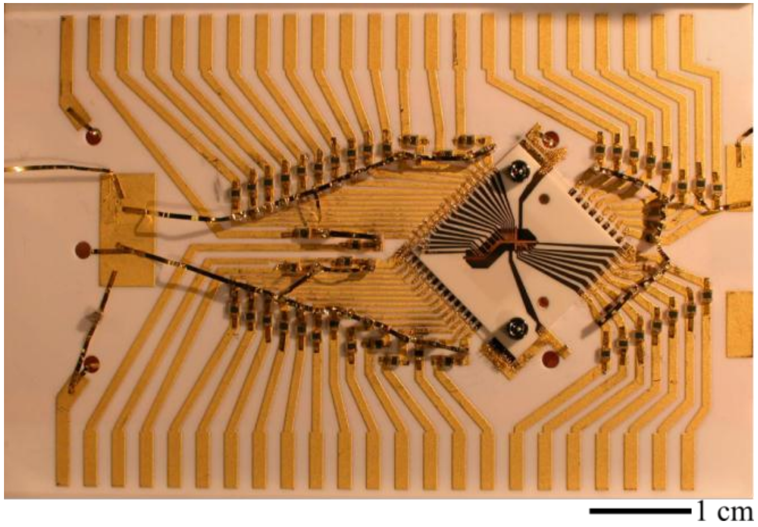
Quantum state tomography

- Want to characterize the state of a quantum system
- Example: ions in a trap





Blatt group, Univ. Innsbruck



Wineland group, NIST-Boulder

Quantum state tomography

- n ions = n qubits
  - Current experiments: 8 to 14 qubits in a single trap
  - Future goal: 50-100 qubits, multiple interconnected traps

- State of n qubits is described by a *density matrix* ρ
  - Dimension d x d, where $d = 2^n$
  - Positive semidefinite matrix w/ trace 1
  - Challenges: large dimension, most matrix elements are small (~1/sqrt(d))

Quantum state tomography

- We can measure Pauli matrices
  - Tensor products of 2x2 matrices
  - $I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \sigma_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \sigma_y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \sigma_z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$
  - $A \times B = \begin{bmatrix} A_{11}B & A_{12}B \\ A_{21}B & A_{22}B \end{bmatrix}$
- For any Pauli matrix P, we can estimate the "expectation value" Tr(Pρ)
  - Prepare the quantum state ρ, measure P, observe ±1, repeat many times, average the results

Quantum state tomography

- Pauli matrices form an *orthogonal basis* for C^{dxd}
- Simple tomography:
  - For all Pauli's P, estimate expectation values Tr(Pρ)
  - Reconstruct ρ by linear inversion, or maximum likelihood
- This is very slow!
  - $O(d^3)$ time – measure d^2 Pauli matrices, ~d times
  - Takes hours, for an ion trap with 8-10 qubits
  - Some details omitted…

Quantum state tomography via compressed sensing

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

- For many interesting quantum states, ρ is *low-rank*
  - Pure states => rank 1
  - Pure states w/ local noise => “effective” rank d^{ε}
- O(rd) parameters, rather than d^2 (where r = rank(ρ))
  - Can we do tomography more efficiently? – Yes!
  - Using an incomplete set of O(rd) Pauli matrices? – Yes!
  - How to choose this set? – At random!
  - How to reconstruct ρ? – Convex optimization!

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 Ω), one can uniquely reconstruct ρ:
  - Estimate $b(P) \approx \mathrm{Tr}(P\rho)$ (for all P in Ω)
  - Solve a convex program:
    $\mathrm{argmin}_X \, \mathrm{Tr}(X)$ s.t. $X \geq 0$ and $|\mathrm{Tr}(PX) - b(P)| \leq \varepsilon$ (for all P in Ω)

Favors low-rank solutions

Where did this idea come from?

- Medical imaging (CAT scans)
  - Reconstruct an image from a (rather incomplete) subset of its Fourier components
  - Naive reconstruction produces lots of artifacts; regularize by minimizing the L1 norm
  - Works well when the true image F is piecewise constant, so its derivative F' is sparse
  - Need O(k polylog n) Fourier components, when F' has k spikes and dimension n
  - Fourier vectors are "incoherent" wrt sparse vectors: $||f||_\infty \leq (1/\sqrt{d})\ ||f||_2$ (Candes, Romberg & Tao, 2004)

Where did this idea come from?

- From sparse vectors to low-rank matrices
  - L1 norm => nuclear norm
    - Sum of singular values, aka, trace norm, Schatten 1-norm
    - (Recht, Fazel & Parrilo, 2007)
  - See also work on "matrix completion"
    - Reconstruct a low-rank matrix M from a subset of entries
    - Assume singular vectors of M are "incoherent" wrt std basis
    - (Candes & Recht, 2008; Candes & Tao, 2009)
  - Fourier vectors => Pauli matrices
    - Pauli matrices are "incoherent" wrt low-rank matrices: $||P|| \leq (1/\sqrt{d})\ ||P||_F$
    - (Gross, Liu, Flammia, Becker & Eisert, 2009; Gross, 2009)

New result: "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 Ω),…
  - One can uniquely reconstruct ρ:
    - Estimate the expectation values Tr(Pρ) (for all P in Ω)
    - Solve a convex program

- 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 results on matrix completion / compressed tomography
  - "Dual certificates"
  - Local properties of state space around a point ρ

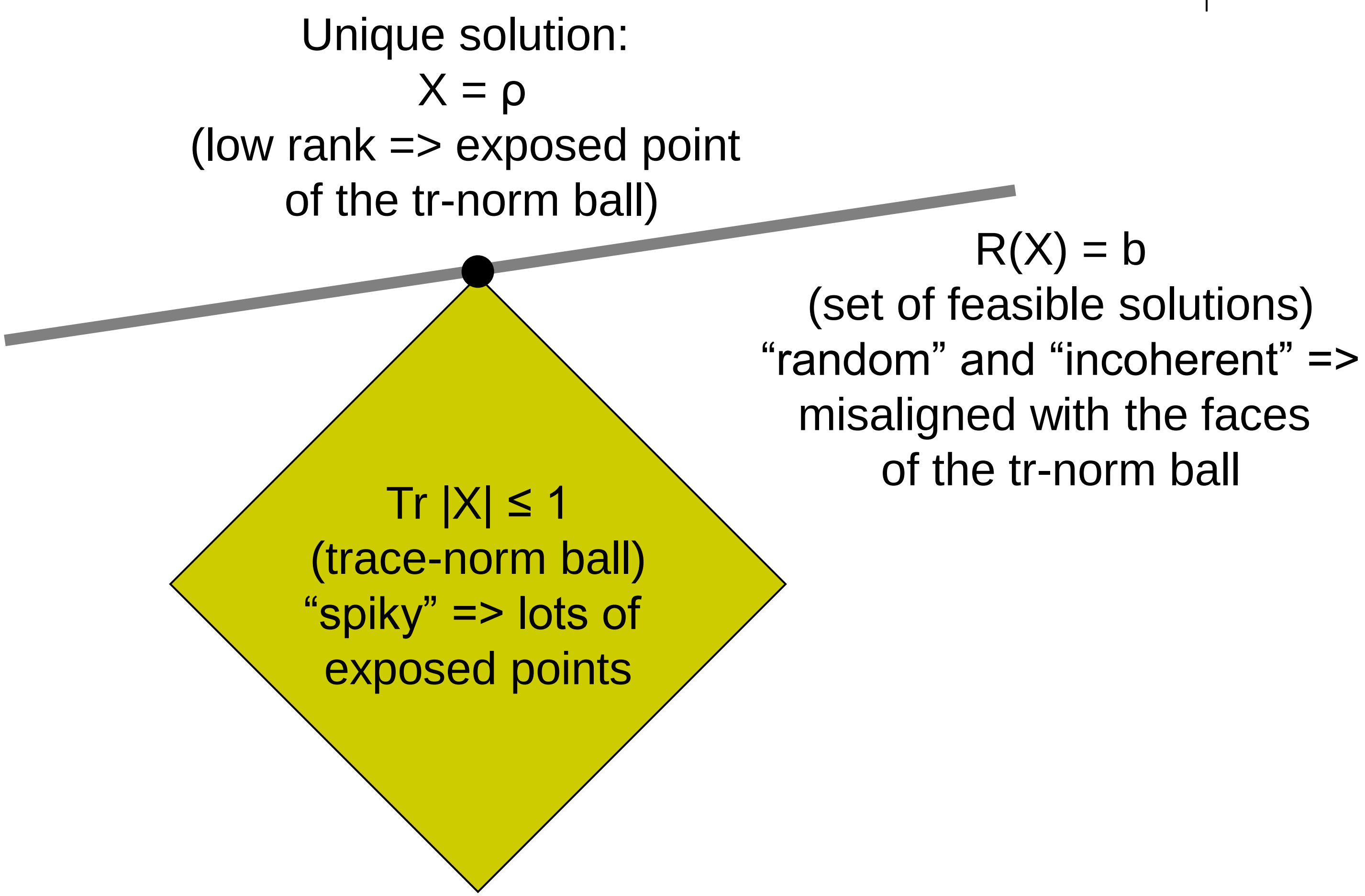
- New result – "universal" matrix recovery
  - "Restricted isometry property" (RIP)
  - Global properties: whole state space can be embedded (w/ small distortion) into R^m, $m = O(rd \text{ polylog } d)$

Some notation

- Sampling operator: $R(\rho) = [\mathrm{Tr}(P\rho)]_{P \text{ in } \Omega}$
  - Returns a vector of Pauli expectation values
  - ρ = unknown state
  - Ω = subset of Pauli operators
  - In a real experiment, after measuring P in Ω, we get $b \approx R(\rho)$

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

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": $||P|| \leq d^{-1/2} ||P||_F$
    - E.g., Pauli matrices, Gaussian random matrices
  - For each ρ, we choose a random hyperplane
    - It's likely to be good

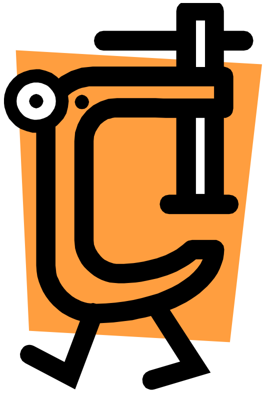
A global picture

- 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,

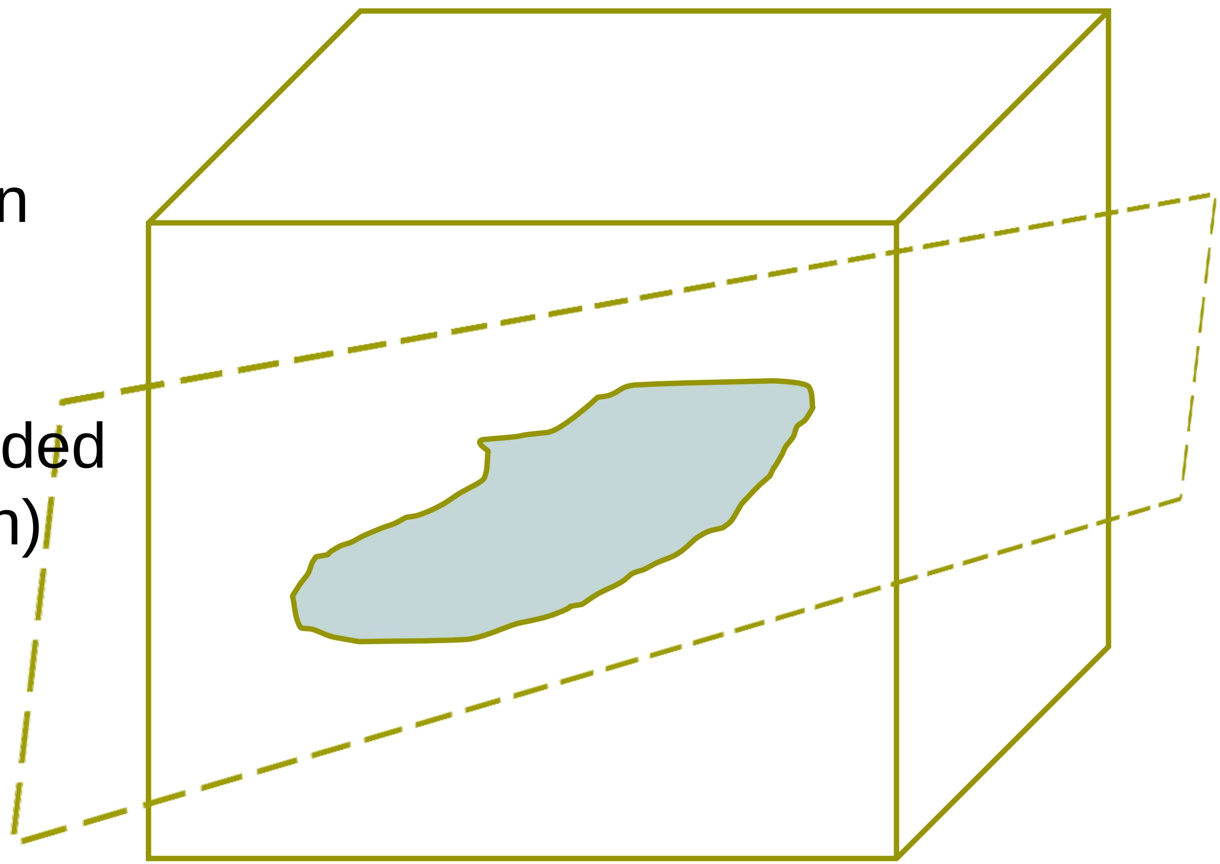
$$(1-\delta)\ \|X\|_2 \leq \|R(X)\|_2 \leq (1+\delta)\ \|X\|_2$$

  - “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": $||P|| \leq d^{-1/2} ||P||_2$
    - 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: what happens when ρ is full-rank?
  - Sample complexity: how many copies of ρ are needed for tomography?

- Proof ideas
  - Entropy argument

- Some practical issues

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

(Flammia, Gross, Liu & Eisert, 2012)

- Compressed tomography uses fewer *measurement settings* m

- But maybe we pay a price in higher *sample complexity*?
  - In practice, answer seems to be no!
  - Total sample complexity stays the same for all m in the range: rd polylog d ≤ m ≤ d^2
  - RIP-based analysis confirms this (up to log factors)!
  - Convenient when it is easier to repeat a measurement than to change measurement settings

Sample complexity

(Flammia, Gross, Liu & Eisert, 2012)
(da Silva, Landon-Cardinal & Poulin, 2011; Flammia & Liu, 2011)

- Using Pauli measurements:

	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
  - (Recht, Fazel & Parrilo, 2007; Candes & Plan, 2010)
- Pauli measurements obey RIP
  - (Liu, 2011)

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 $||\mathbf{R^*R} - \mathbf{1}||_{(r)} < 2\delta - \delta^2$
  - Construct **R** by sampling Pauli matrices iid at random
  - **R*R** is a sum of iid random variables, E(**R*R**) = 1
  - Bound $E(||\mathbf{R^*R} - \mathbf{1}||_{(r)})$, 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^{dxd}, $||M||_X = \max_{P \text{ in } \Omega} |Tr(P^+M)|$
  - Problem reduces to bounding $N(B_1, ||\cdot||_X, \varepsilon)$

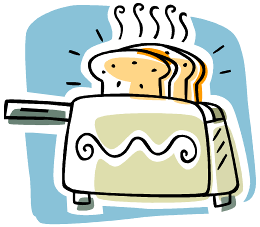
  - Trivial bound:
    $N(B_1, ||\cdot||_X, \varepsilon) \leq$ (polynomial in $1/\varepsilon$, exponential in d^2)

  - 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):

  $$\int w_a \, \mathrm{Tr}(w_a^{+}\rho) \, d\mu(a) = \rho/d^2, \text{ for all } \rho$$

- Incoherence condition: $\|w_a\| \leq O(1/\sqrt{d})$

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

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

?

Some practical issues

- How to solve the trace-minimization convex program?
- Interior-point SDP solvers
  - Very accurate, fast enough for 6 qubits
- First-order methods
  - Can handle very large instances, but less accurate?
  - Careful: objective function is not smooth!
  - E.g., 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?
  - Higher-order tensors?
  - Machine learning: matrix completion, learning HMM's