Compressed sensing in quantum tomography without *a priori* information

Yong Siah Teo

CMQC, Frontier Physics Research Division, SNU

Department of Optics, Palacký University, October 13, 2017

1. Known results
   - Economical quantum tomography
   - Compressed Sensing

2. New Results
   - Adaptive compressed sensing
   - Algorithm
   - Graphs

3. Conclusions

Informationally incomplete tomography

For an unknown quantum state ρ (D-dimensional):

- Informationally complete (IC) measurement ($M \geqslant D^2$ outcomes) $\Rightarrow$ unique reconstruction (ρ for noiseless data).
- A non-IC measurement gives a convex set $\mathcal{C}$ of infinitely many estimators that are consistent with the data (max. likelihood (ML) probabilities).

Informationally incomplete tomography

For an unknown quantum state ρ (D-dimensional):

- Informationally complete (IC) measurement ($M \geqslant D^2$ outcomes) $\Rightarrow$ unique reconstruction (ρ for noiseless data).
- A non-IC measurement gives a convex set $\mathcal{C}$ of infinitely many estimators that are consistent with the data (max. likelihood (ML) probabilities).

  Can we get ρ with incomplete data + some more information?

Compressed sensing (CS) in quantum tomography

Can we get ρ with incomplete data + some more information? ✓

If ρ is known to have a rank of at most r (r-sparse):

- A specialized set of IC CS measurement + rank minimization $\Rightarrow$ unique reconstruction ρ (non-convex l_0 problem).
- A specialized set of IC CS measurement + feasible reconstruction scheme with positivity constraint $\Rightarrow$ unique reconstruction ρ.

Can CS be more versatile?

In standard CS schemes:

- Either the r-sparsity assumptions is needed to construct CS measurements, or certain CS random measurements (like random Pauli bases) are employed.
- The r-sparsity assumption requires more experimental justifications.

Can CS be more versatile?

In standard CS schemes:

- Either the r-sparsity assumptions is needed to construct CS measurements, or certain CS random measurements (like random Pauli bases) are employed.
- The r-sparsity assumption requires more experimental justifications.

Is CS achievable without any *a priori* information, and with other more experimentally convenient (deterministic) measurements?

Adaptive compressed sensing (ACS)—tackling the crux of the problem

Yes, CS can be asymptotically achieved without any *a priori* information about ρ and with deterministic measurements. ✓

- **The heart of CS:** rapidly shrink the convex set $\mathcal{C}$ to the point ρ.
- **Rephrasing the problem:** *For a given ρ, find the set of measurement outcomes $\mathcal{A}$ of minimal cardinal M such that $\mathcal{C}$ is a single point.*

  There are three issues:

  (a) Choosing the optimal measurement set $\mathcal{A}$.

  (b) Computing the size s of $\mathcal{C}$

  (c) Dealing with a completely unknown ρ

(a) Choosing the optimal measurement set $\mathcal{A}$

- Without any information about ρ, optimal $\mathcal{A}$ depends on *a posteriori* information given by data.
- The optimal $\mathcal{A}$ may be found through an adaptive scheme.
- Experimental observers frequently pick $\mathcal{A}$ to be a set of measurement bases $\mathcal{A} = \{\mathcal{B}_1, \mathcal{B}_2 \ldots\}$, so that $\mathcal{B}_{k+1}$ depends on $\{\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_k\}$.
- So $\mathcal{B}_1$ affects the next choice of $\mathcal{B}_2$, which in turn affects the next choice of $\mathcal{B}_3$, and so on.

(b) Computing the size s of $\mathcal{C}$

- A standard definition of size of a (convex) set, say for $\mathcal{C}$, is the integral

$$s = \int_{\mathcal{C}} (\mathrm{d}\rho) \leqslant 1, \quad (\mathrm{d}\rho) : \text{ prior of } \rho\,.$$

- s is hard to compute (especially for higher dimensions).
- We need a feasible indicator of s.

(b) Computing the size s of $\mathcal{C}$

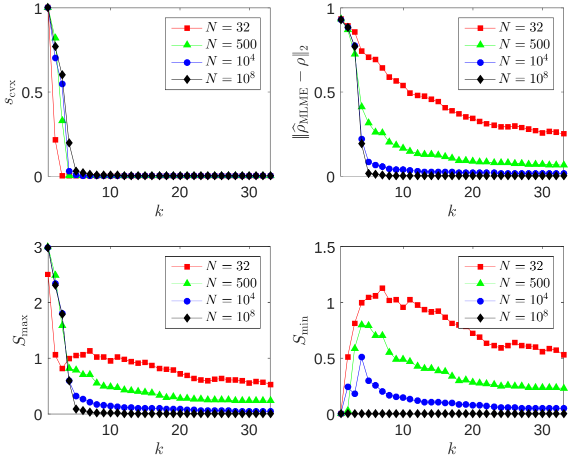
ONE recipe for a feasible indicator—*size monotone*

- Pick a concave (convex) function $f(\rho)$ that has a unique maximum (minimum) to characterize convex set $\mathcal{C}$ for some measurement $\mathcal{A}$.
- Define $s_{\text{CVX}} = \text{const.} \times (f_{\max} - f_{\min})$ over $\mathcal{C}$.
  One normalization: $s_{\text{CVX},1} \equiv 1$ and $s_{\text{CVX},k_{\text{IC}}} \equiv 0$.
- s_{CVX} is a size monotone when the sufficient condition holds

  Data are noiseless so that as $\mathcal{B}_1, \mathcal{B}_2, \ldots$ are progressively measured, $\mathcal{C}_1 \supseteq \mathcal{C}_2 \supseteq \ldots$. This condition ensures that $f_{\max,1} \geqslant f_{\max,2} \geqslant \ldots$ and $f_{\min,1} \leqslant f_{\min,2} \leqslant \ldots$ due to concavity of $f(\rho)$.
- Since $\mathcal{C}$ is a convex set, $s_{\text{CVX}} = 0 \Rightarrow s = 0$ whenever f has a unique maximum.

Von Neumann entropy $S(\rho) = -\text{tr}\{\rho \log \rho\}$ is one example of such $f(\rho)$.

(b) Computing the size s of $\mathcal{C}$: Monotonicity of $s_{\rm CVX}$

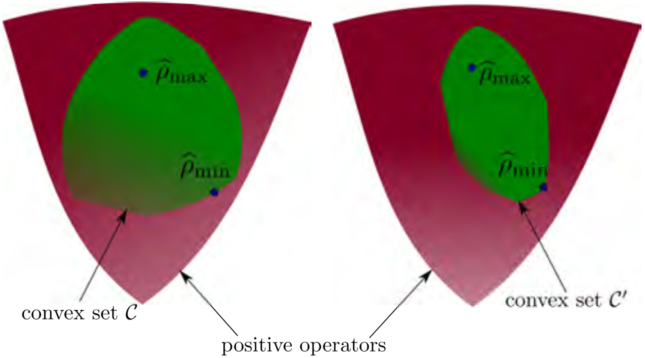


- N: Number of data copies
- k: Number of measured bases
- $\widehat{\rho}_{\rm MLME}$: Max. entropy state over $\mathcal{C}_k$ consistent with all ML probabilities for each k

A fixed sequence of mutually unbiased bases (MUB) are measured on a randomly generated 5-qubit pure state.

(b) Computing the size s of $\mathcal{C}$

- If the sufficient condition holds, smaller s_{CVX} implies smaller s.
- $\therefore$ size monotone $\equiv$ size-reduction witness under this condition.
- We judge the quality of measurements by the rate at which s_{CVX} approaches zero.



(c) Dealing with a completely unknown ρ

- With no information about ρ, CS should then depend only on *a posteriori* information from measured data.
- With the concave function $f(\rho)$ assigned, the unique maximum $\widehat{\rho}_{\max}$ that gives $f_{\max}$ may act as the *a posteriori* information.
- For $f(\rho) = S(\rho)$, $\widehat{\rho}_{\mathrm{MLME}}$ may be used as the *a posteriori* guide to find the next optimal measurement basis.
- $\widehat{\rho}_{\mathrm{MLME}} \to \rho$ as number of measured bases increases.

Main idea of ACS, with $f(\rho) = S(\rho)$

- After measuring $k = 1$ basis $\mathcal{B}_1$ (reference basis $\mathcal{B}_0 \equiv \mathcal{B}_1$), we look for $\widehat{\rho}_{\mathrm{MLME},1}$ and use this as an *a posteriori* estimate of ρ.
- Choose the optimal U_2 that gives the smallest $s_{\mathrm{CVX},2}$ to rotate $\mathcal{B}_0$ according to measured data **combined with** data predicted by $\widehat{\rho}_{\mathrm{MLME},1}$. Then measure this U_2-rotated basis.
- Then with these $k = 2$ measured bases $\{\mathcal{B}_1, \mathcal{B}_2\}$, choose the next optimal U_3 that gives the smallest $s_{\mathrm{CVX},3}$ according to **all** measured data **combined with** data predicted by $\widehat{\rho}_{\mathrm{MLME},2}$, and measure this U_3-rotated basis.
- Continue until $s_{\mathrm{CVX},k}$ is small enough.

Structure of U is flexible; *E.g.* U **may take local tensor-product structure (common in experiments), or other device-dependent structure in some degrees of freedom.**

Pseudocode

STEP 1 Set $k = 1$ and measure basis $\mathcal{B}_1$ and set it as the computational basis.

STEP 2 Perform MLME and obtain $\widehat{\rho}_{\text{MLME},1}$ and $S_{\max,1}$. That $S_{\min,1} = 0$ is clear, and thus $s_{\text{CVX},1} = 1$.

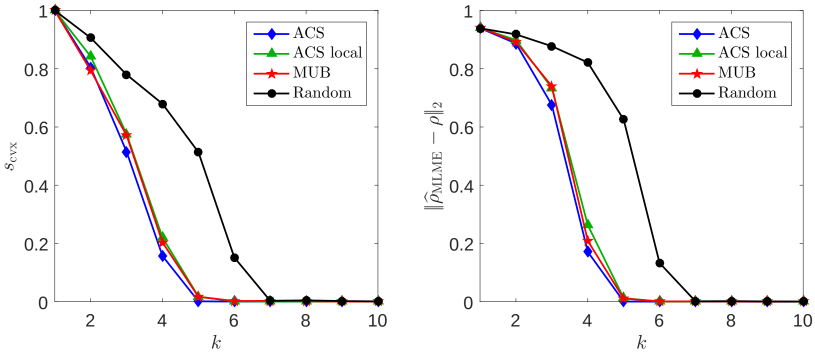
STEP 3 Search for the unitary operator U_{k+1} that defines $\mathcal{B}_{k+1} = U_{k+1}\mathcal{B}_k U_{k+1}^\dagger$ such that $s_{\text{CVX},k+1}$ is minimized with the cumulatively measured bases $\{\mathcal{B}_1, \ldots, \mathcal{B}_k\}$ by using $\widehat{\rho}_{\text{MLME},k}$ as an *a posteriori* estimator of ρ to generate simulated data for $\mathcal{B}_{k+1}$. Minimization of $s_{\text{CVX},k+1}$ may be done with a nonlinear optimization routine.

STEP 4 Measure the basis $\mathcal{B}_{k+1}$ and perform MLME to obtain $\widehat{\rho}_{\text{MLME},k+1}$ with the cumulatively measured data.

STEP 5 Raise k by 1.

STEP 6 Repeat STEP 3 through STEP 5 until $s_{\text{CVX},k}$ is below certain pre-chosen threshold value.

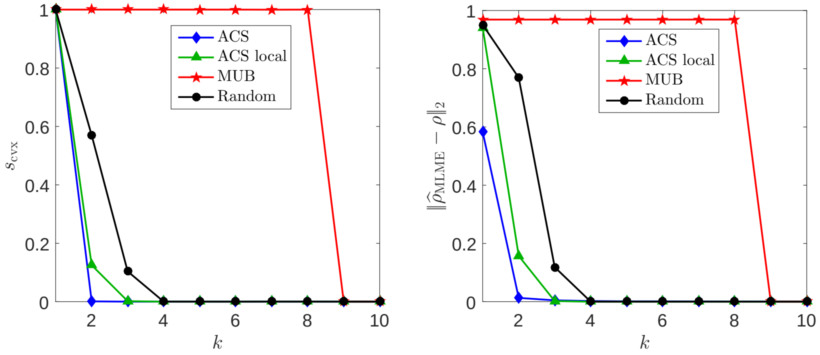
Rank-1 states



- ACS: Adaptive CS over arbitrary U space
- ACS local: Adaptive CS over tensored U space
- MUB: Optimization over a set of MUB ($D = 2^5 = 32$)

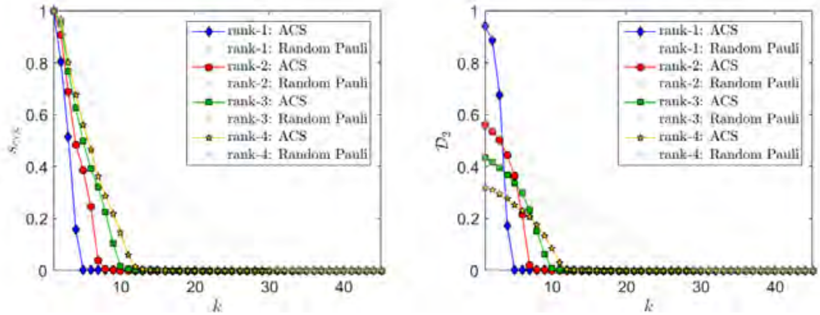
$\ast$ *Averaged over 8 Haar-distributed 5-qubit pure states.*

Mutual unbiasedness: double-edge sword in ACS



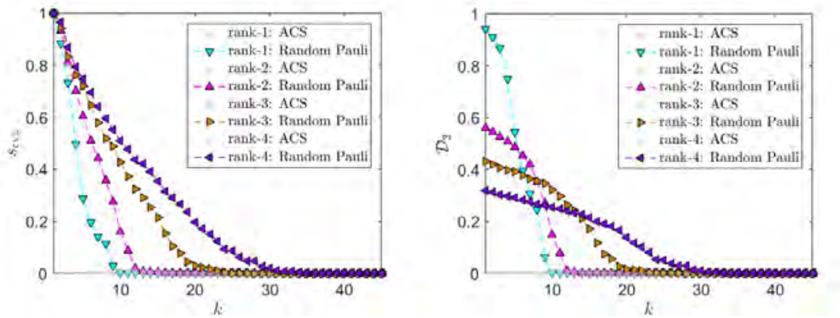
- MUB are asymptotically optimal in minimal-basis tomography
- But can perform extremely badly in ACS when ρ is one of their eigenstates.
- ACS schemes can, on the other hand, adapt to any ρ and $f(\rho)$ to improve the *a posteriori* information and size monotone.

Rank-r states ($r = 1, 2, 3, 4$)



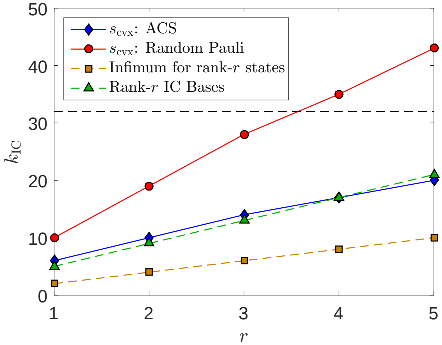
- ACS plots averaged over 8 random 5-qubit states distributed according to the Hilbert-Schmidt measure for each r.
- IC $k = k_{\mathrm{IC}}$ never exceeds $D + 1$.

Rank-r states $(r = 1, 2, 3, 4)$



- Random-Pauli-bases (RPB) measurement plots averaged over 8 random 5-qubit states for each r.
- Avg. k_{IC} can exceed $D + 1$ due to overcompleteness of the RPB. (*i.e.* minimal number of 2-qubit RPBs for $r = D = 4$ is $6 > 5$.)
- Minimal sets of RPBs are highly specific, not random.

Rank-r states ($r = 1, 2, 3, 4$)



- Avg. scaling for ACS is better than that for RPB.
- Known state-of-the-art rank-r IC bases: rank-r Goyeneche-type bases $(4r + 1)$.
- Avg. k_{IC} for ACS is comparable and can beat the Goyeneche-type bases for larger r.

Implications

- **Golden lesson:** When no *a priori* information is available, tomography schemes that depend on data *a posteriori* information should be adaptive and not restricted to specific measurement sets.
- ACS can achieve CS behavior with size monotones constructed from any convex function, for any unknown ρ and no *a priori* information.
- ACS beats current known RPB measurements in terms of the average k_{IC}, and eventually beats even the state-of-the-art Goyeneche-type rank-r bases which are strictly-IC.
- Adaptation is versatile (accepts any forms of $\mathcal{A}$).