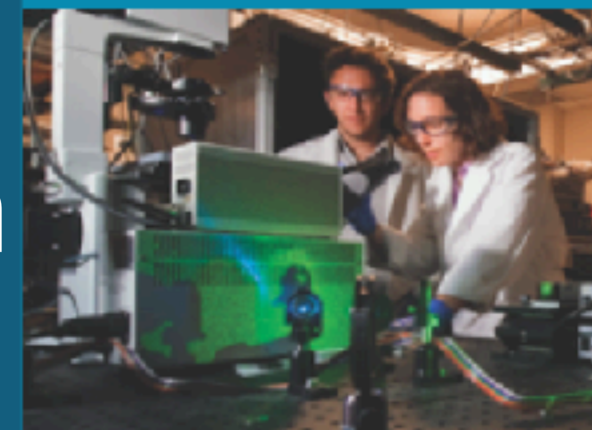
SAND2018-2024PE

Sandia National Laboratories



On the edge: Geometry, model selection, and quantum compressed sensing



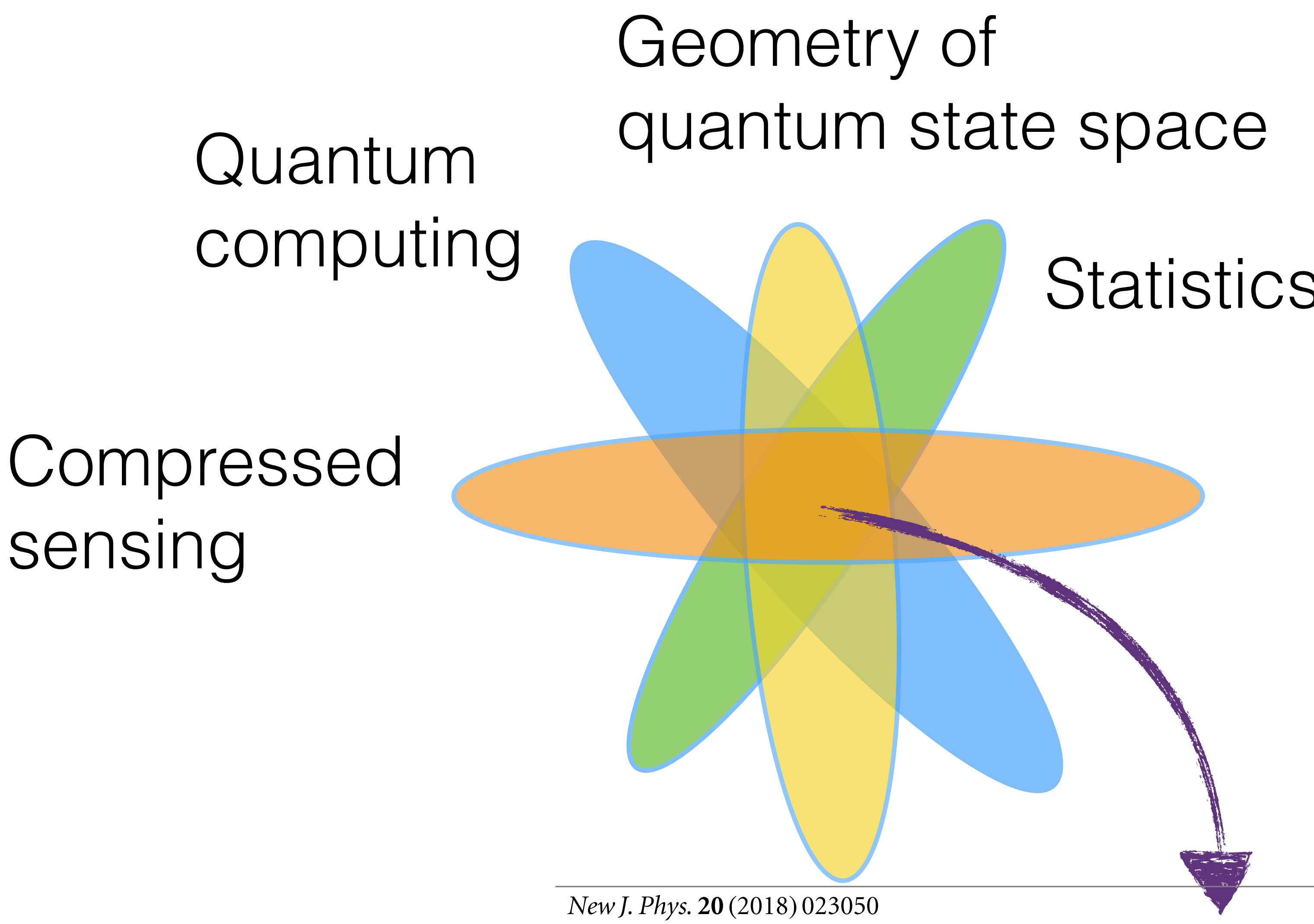
PRESENTED BY

Travis L Scholten @Travis_Sch

Center for Quantum Information and Control, UNM

Center for Computing Research, Sandia National Laboratories

This talk lies at the intersection of several topics.



New J. Phys. **20** (2018) 023050 https://doi.org/10.1088/1367-2630/aaa7e2

New Journal of Physics
The open access journal at the forefront of physics

Deutsche Physikalische Gesellschaft DPG
IOP Institute of Physics

Published in partnership with: Deutsche Physikalische Gesellschaft and the Institute of Physics

Scholten &
Blume-Kohout,
NJP **20** 023050 (2018)

PAPER

Behavior of the maximum likelihood in quantum state tomography

Travis L Scholten[1,2] **and Robin Blume-Kohout**[1,2]

[1] Center for Computing Research (CCR), Sandia National Laboratories, United States of America
[2] Center for Quantum Information and Control (CQuIC), University of New Mexico, United States of America

Characterizing the behavior of noisy, intermediate-scale quantum information processors can be hard.

Suppose we have an n-qubit NISQ device.

The number of parameters to be estimated in various tomographic protocols scales poorly:

State tomography - $p = \mathcal{O}(4^n)$

Process tomography - $p = \mathcal{O}(16^n)$

Gate set tomography - $p = \mathcal{O}(M * 16^n)$

How do we reduce the number of parameters necessary to characterize the device?

In practice, we usually impose constraints on the estimates to reduce the number of parameters.

State tomography - $p = \mathcal{O}(4^n)$

"State has known rank": $p = \mathcal{O}(r * 2^n)$

Process tomography - $p = \mathcal{O}(16^n)$

"Process is unitary": $p = \mathcal{O}(4^n)$

Gate set tomography - $p = \mathcal{O}(M * 16^n)$

"Error generators act on one or two qubits": $p = \mathcal{O}\left(M\left(12n + 120n^2\right)\right)$

Our work: identify how to use statistical model selection to choose a good *model*.

State tomographers have been doing model selection all along!

For tomography, a *model* is a set of density matrices.

Trivial model selection:

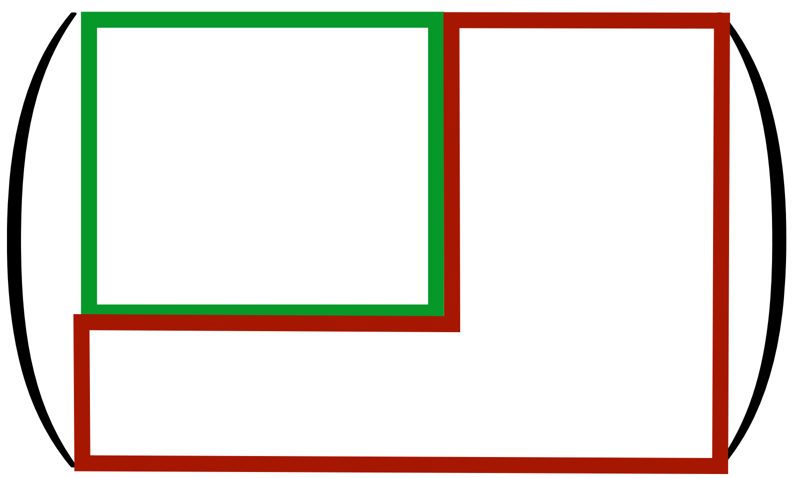
$$\hat{\rho} = \left(\boxed{} \right)$$

Pick a Hilbert space by fiat.
("Of course it's a qubit!")

State tomographers have been doing model selection all along!

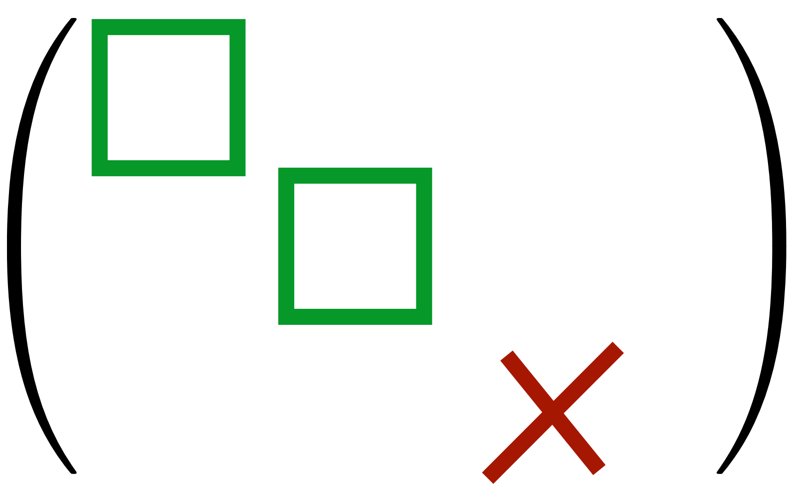
For tomography, a *model* is a set of density matrices.

Non-trivial model selection:

$\hat{\rho} =$ 

Restrict estimate to a subspace.

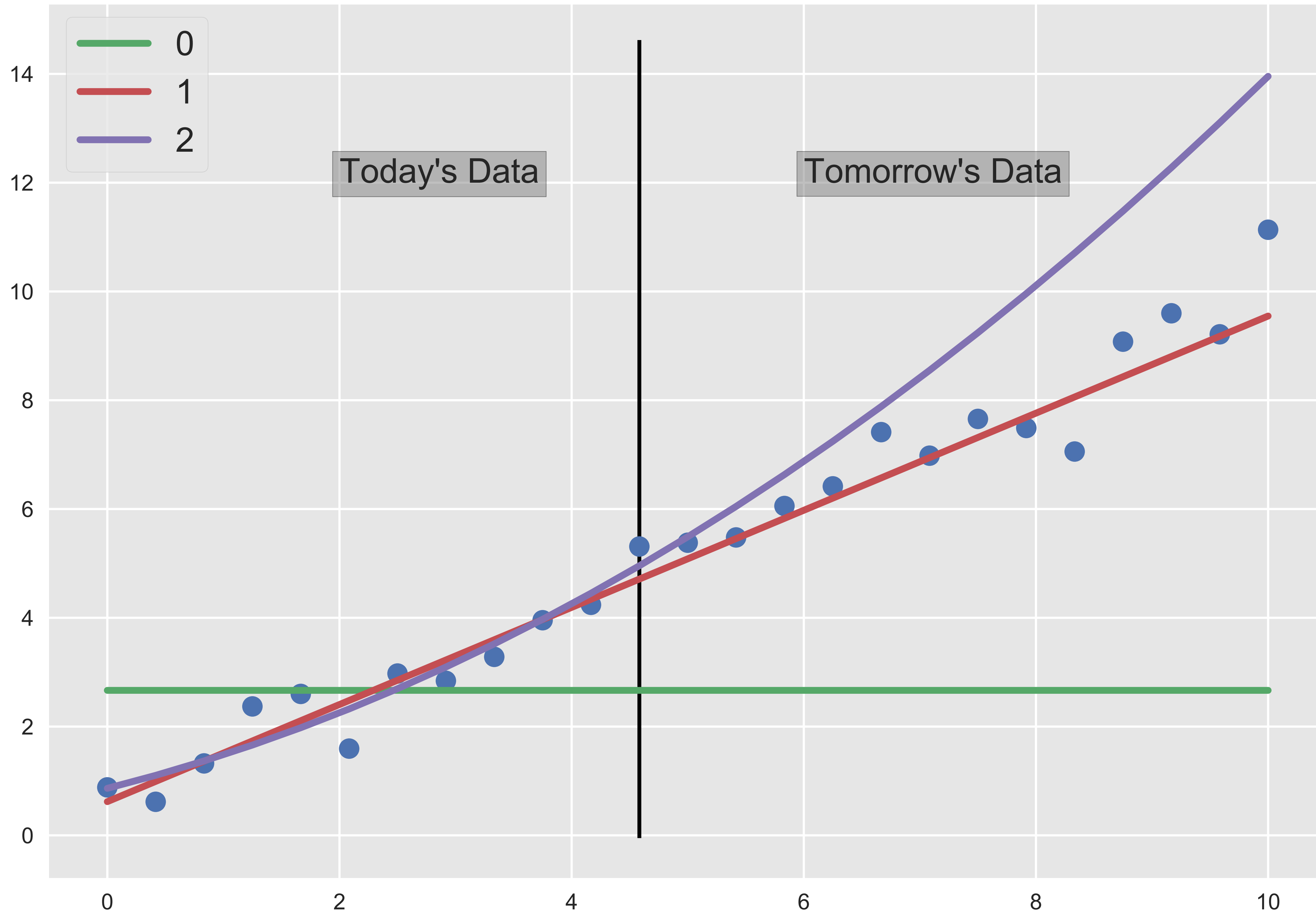
$$\hat{\rho} = \sum_{j,k=0}^{N-1} \rho_{jk} |j\rangle\langle k|$$

$\hat{\rho} =$ 

Restrict the rank of the estimate.

$$\hat{\rho} = \sum_{j=0}^{r-1} \lambda_j |\lambda_j\rangle\langle\lambda_j|$$

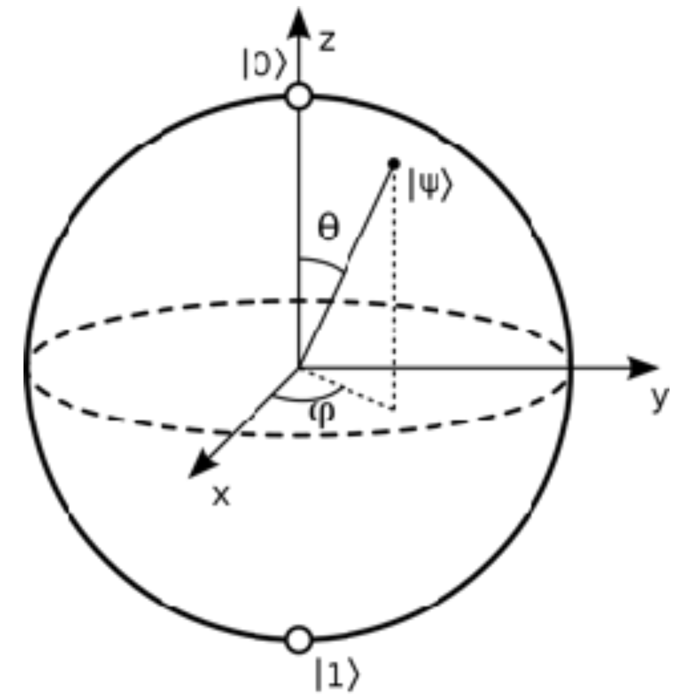
Model selection is used to identify which model fits the data well, and is also useful for prediction.



A *model* is a parameterized family of probability distributions.

Common model for state tomography:

$$\mathcal{M}_{\mathcal{H}} = \{\rho \mid \rho \in \mathcal{B}(\mathcal{H}),\ \mathrm{Tr}(\rho) = 1,\ \rho \geq 0\}$$



Probabilities via the Born rule: $p_j = \mathrm{Tr}(\rho E_j)$

Changing the state changes the probability!

$$\mathrm{POVM} = \{|0\rangle\langle 0|, |1\rangle\langle 1|\}$$

$$\rho_0 = |0\rangle\langle 0| \implies \Pr(\text{“0”}) = 1$$

$$\rho_0 = |+\rangle\langle +| \implies \Pr(\text{“0”}) = 1/2$$

$$\rho_0 = |1\rangle\langle 1| \implies \Pr(\text{“0”}) = 0$$

Maximum likelihood estimation is a common way to infer which parameters of a model can explain your data best.

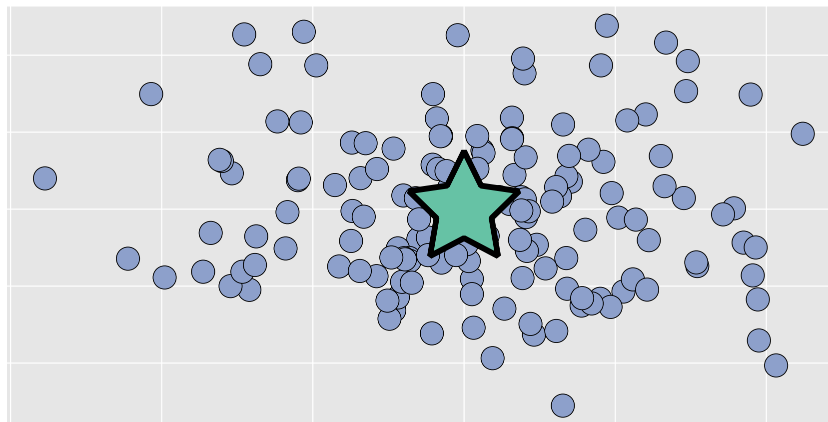
Given data, *likelihood* is

$$\mathcal{L}(\rho) = \mathrm{Pr}(\mathrm{Data}|\rho)$$

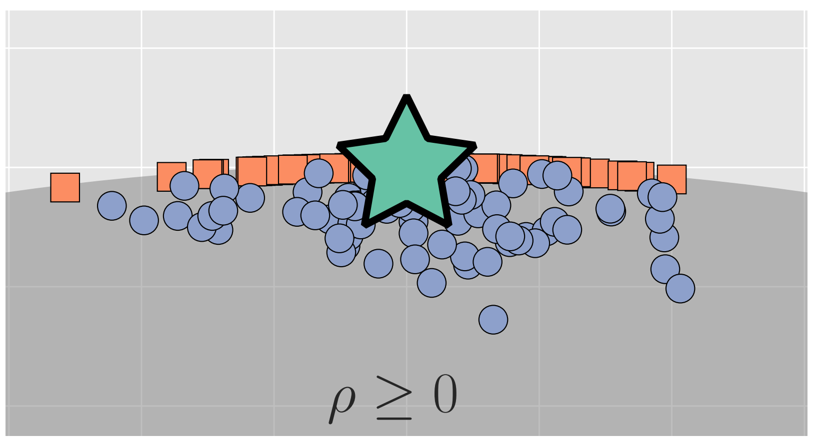
The maximum likelihood estimate is computed as

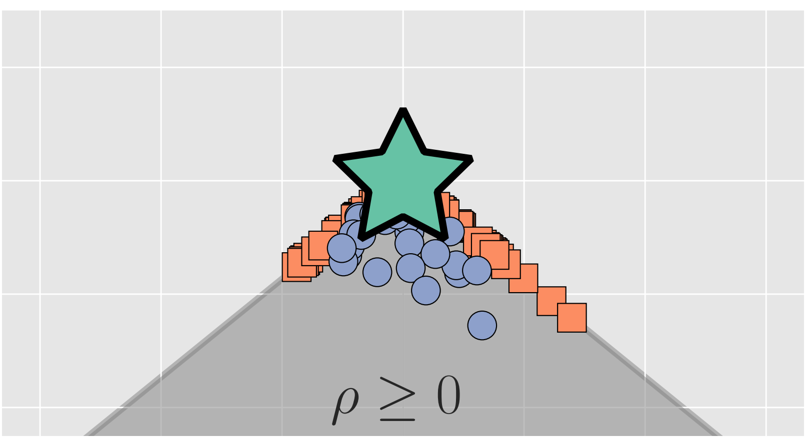
$$\hat{\rho}_{\mathrm{ML},\mathcal{M}} = \max_{\rho\in\mathcal{M}} \mathcal{L}(\rho)$$

Quantum state space has boundaries, posing some challenges for tomography & model selection.



Easy to reason about (many known results)



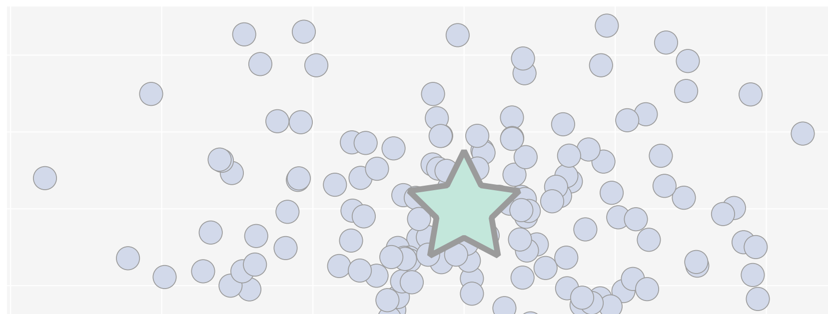


Hard to reason about (known results don't apply!)

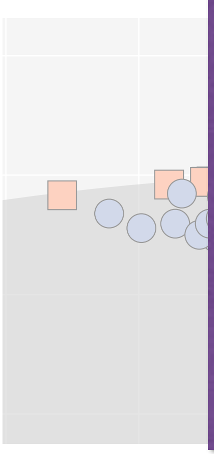
Tomography:
Boundaries *distort the distribution* of maximum likelihood estimates (makes reasoning about their properties hard).

Model selection:
Common techniques (Wilks theorem, information criteria) cannot be used!

Quantum state space has boundaries, posing some challenges for tomography & model selection.



Easy to reason about
(many known results)



on about
ts don’t

These issues stem from the fact that the models used in tomography do not satisfy *Local Asymptotic Normality* (LAN).

Tomography:
Boundaries *distort the distribution* of maximum likelihood estimates (makes reasoning about their properties hard).

Model selection:
Common techniques (Wilks theorem, information criteria) cannot be used!

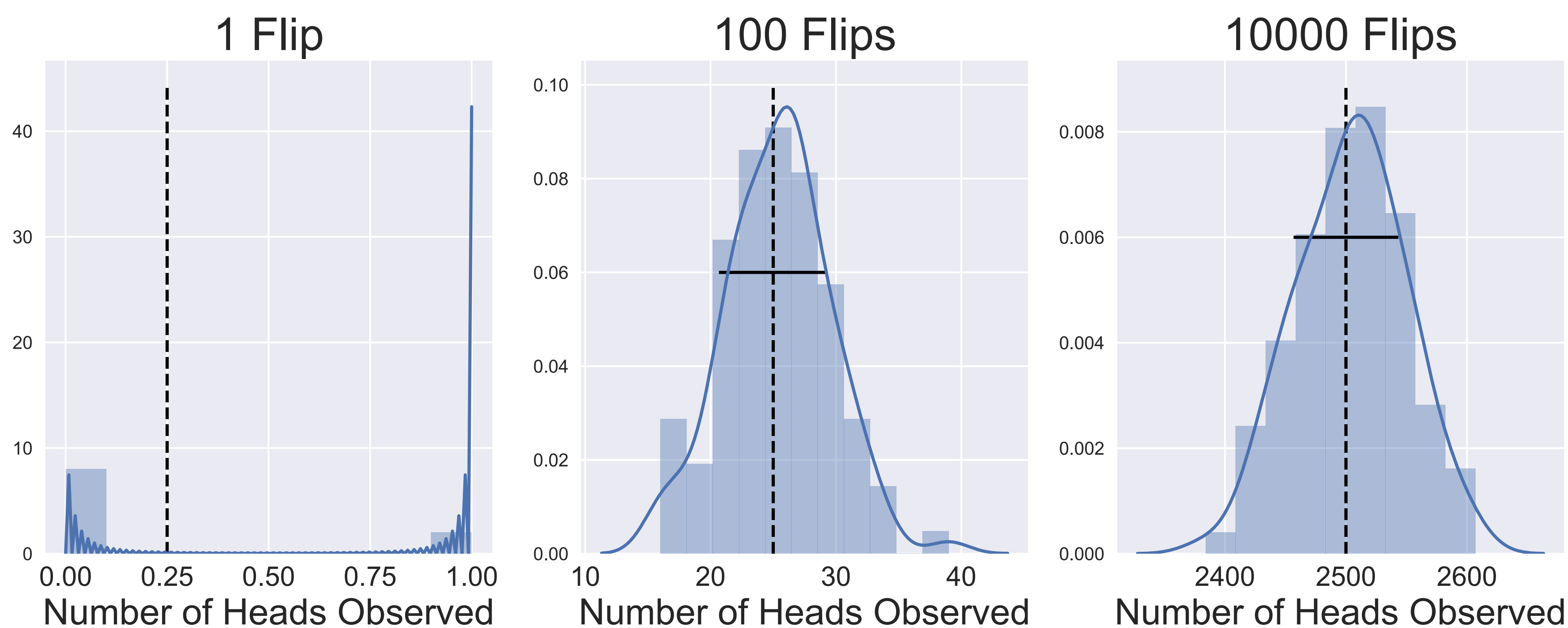
In classical statistics, models often satisfy *local asymptotic normality (LAN)*.

Local = fix a value for the parameters θ_0

Asymptotic = the number of samples goes to infinity

Normality = the probability distribution function can be approximated by a Gaussian

Example: coin flips



$\left(P_{\theta_0+u/\sqrt{n}} : u \in \mathbb{R}^m\right)$ & $\left(N(u, I_{\theta_0}^{-1}) : u \in \mathbb{R}^m\right)$ have similar statistical properties

Le Cam L., Yang G.L. (2000) Local Asymptotic Normality. In: Asymptotics in Statistics. Springer Series in Statistics. Springer, New York, NY

If LAN is satisfied by a model, then several properties follow.

If LAN is satisfied, then asymptotically:

Likelihoods are Gaussian:

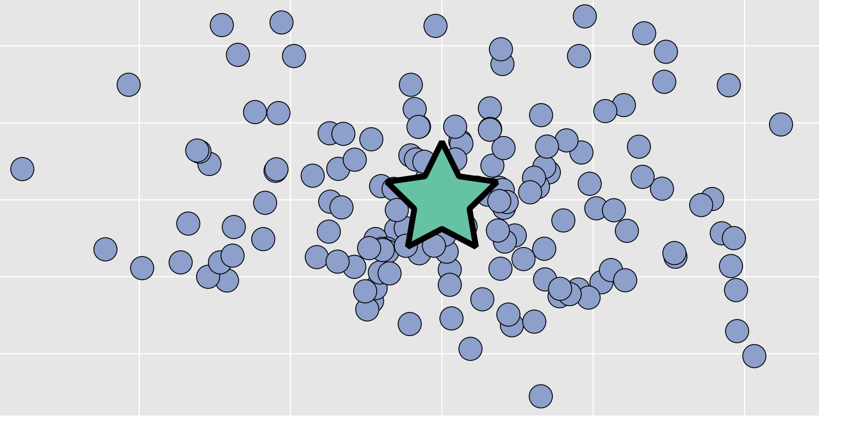
$$\mathcal{L}(\rho) \equiv \Pr(\mathrm{Data}|\rho)$$

$$\underset{N\to\infty}{\propto} \mathrm{Exp}\left[-\frac{1}{2}\mathrm{Tr}(\rho - \hat{\rho}_{\mathrm{ML},\mathcal{M}})F(\rho - \hat{\rho}_{\mathrm{ML},\mathcal{M}})\right]$$

Maximum likelihood (ML) estimates are normally distributed:

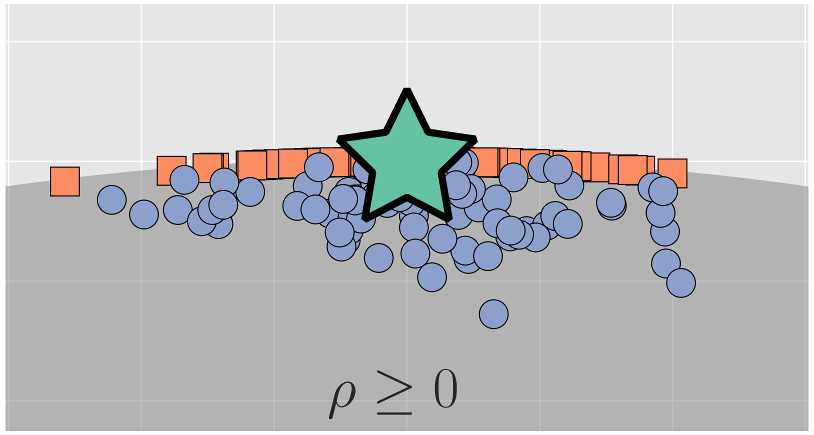
$$\hat{\rho}_{\mathrm{ML},\mathcal{M}} \equiv \underset{\rho\in\mathcal{M}}{\mathrm{argmax}}\ \mathcal{L}(\rho)$$

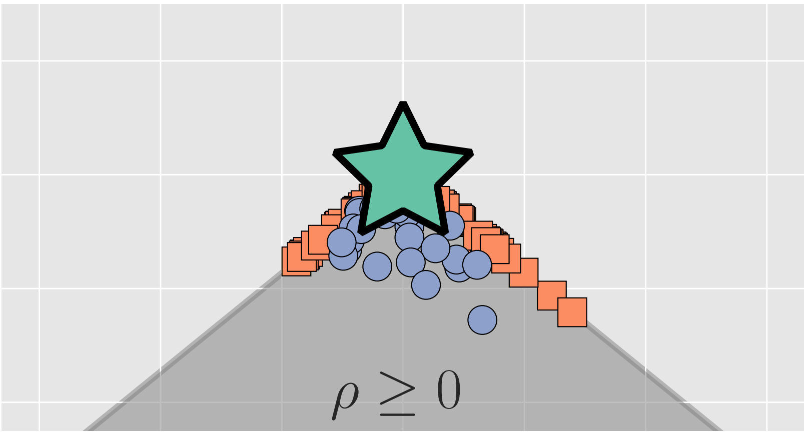
$$\xrightarrow{\mathrm{d}} \mathcal{N}(\rho_0, \mathcal{F}^{-1})$$



Key implication: $\mathcal{M}$ satisfies LAN $\implies \hat{\rho}_{\mathrm{ML},\mathcal{M}} \sim \mathcal{N}(\rho_0, F^{-1})$

We know ML estimates in state tomography are not always normally distributed, implying LAN is not satisfied.





Key issue: $\hat{\rho}_{\mathrm{ML},\mathcal{M}} \not\sim \mathcal{N}(\rho_0, F^{-1}) \implies \mathcal{M}$ does not satisfy LAN

Because LAN is not satisfied, the assumptions necessary for many model selection tools are violated!!

How do we fix this?

New Journal of Physics
The open access journal at the forefront of physics

Deutsche Physikalische Gesellschaft DPG
IOP Institute of Physics

Published in partnership with: Deutsche Physikalische Gesellschaft and the Institute of Physics

PAPER

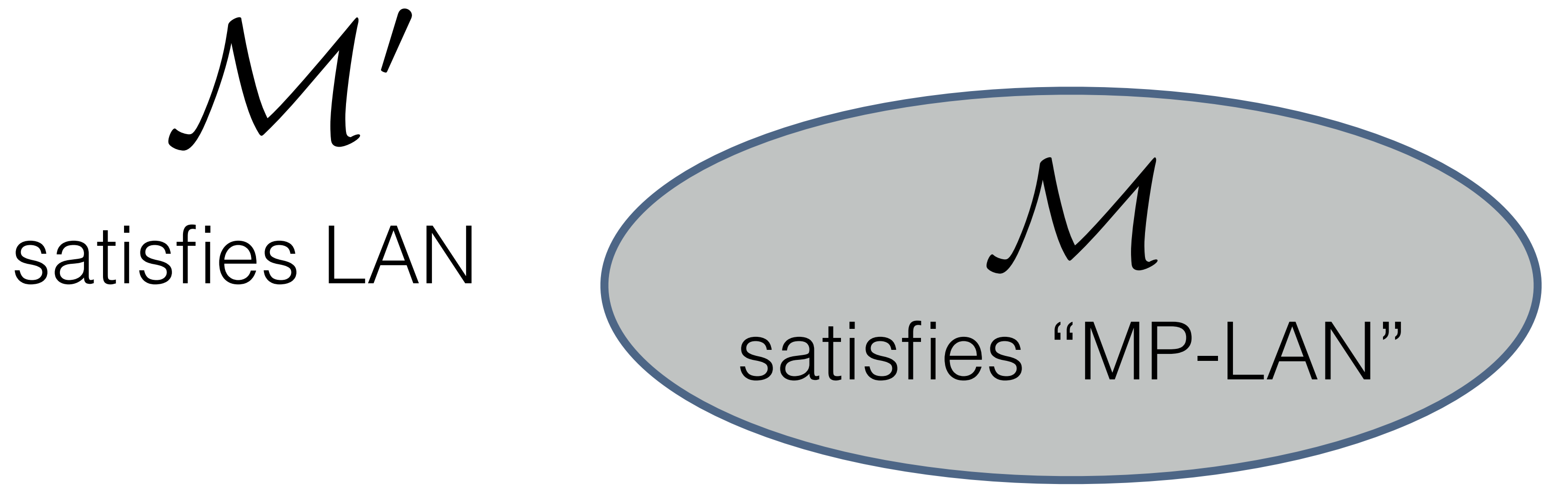
Behavior of the maximum likelihood in quantum state tomography

Travis L Scholten[1,2] **and Robin Blume-Kohout**[1,2]

[1] Center for Computing Research (CCR), Sandia National Laboratories, United States of America
[2] Center for Quantum Information and Control (CQuIC), University of New Mexico, United States of America

We show how to **generalize LAN** for models with convex boundaries.

We define a new generalization of LAN for models with convex boundaries.



Definition 1 (Metric-projected local asymptotic normality, or MP-LAN). *A model $\mathcal{M}$ satisfies MP-LAN if, and only if, $\mathcal{M}$ is a convex subset of a model $\mathcal{M}'$ that satisfies LAN.*

We show that quantum state space satisfies MP-LAN.

In state tomography,

$$\mathcal{M}_{\mathcal{H}} = \{\rho \mid \rho \in \mathcal{B}(\mathcal{H}),\ \mathrm{Tr}(\rho) = 1,\ \rho \geq 0\}$$ (all density matrices)

Define

$$\mathcal{M}'_{\mathcal{H}} = \{\sigma \mid \sigma \in \mathcal{B}(\mathcal{H}), \mathrm{Tr}(\sigma) = 1\}$$ (lift positivity constraint)

(Likelihood is twice continuously differentiable,
so LAN is satisfied.)

New J. Phys. **20** (2018) 023050 https://doi.org/10.1088/1367-2630/aaa7e2

New Journal of Physics
The open access journal at the forefront of physics

Deutsche Physikalische Gesellschaft DPG
IOP Institute of Physics

Published in partnership with: Deutsche Physikalische Gesellschaft and the Institute of Physics

PAPER

Behavior of the maximum likelihood in quantum state tomography

Travis L Scholten[1,2] **and Robin Blume-Kohout**[1,2]

[1] Center for Computing Research (CCR), Sandia National Laboratories, United States of America
[2] Center for Quantum Information and Control (CQuIC), University of New Mexico, United States of America

We show how to **generalize LAN** for models with convex boundaries.

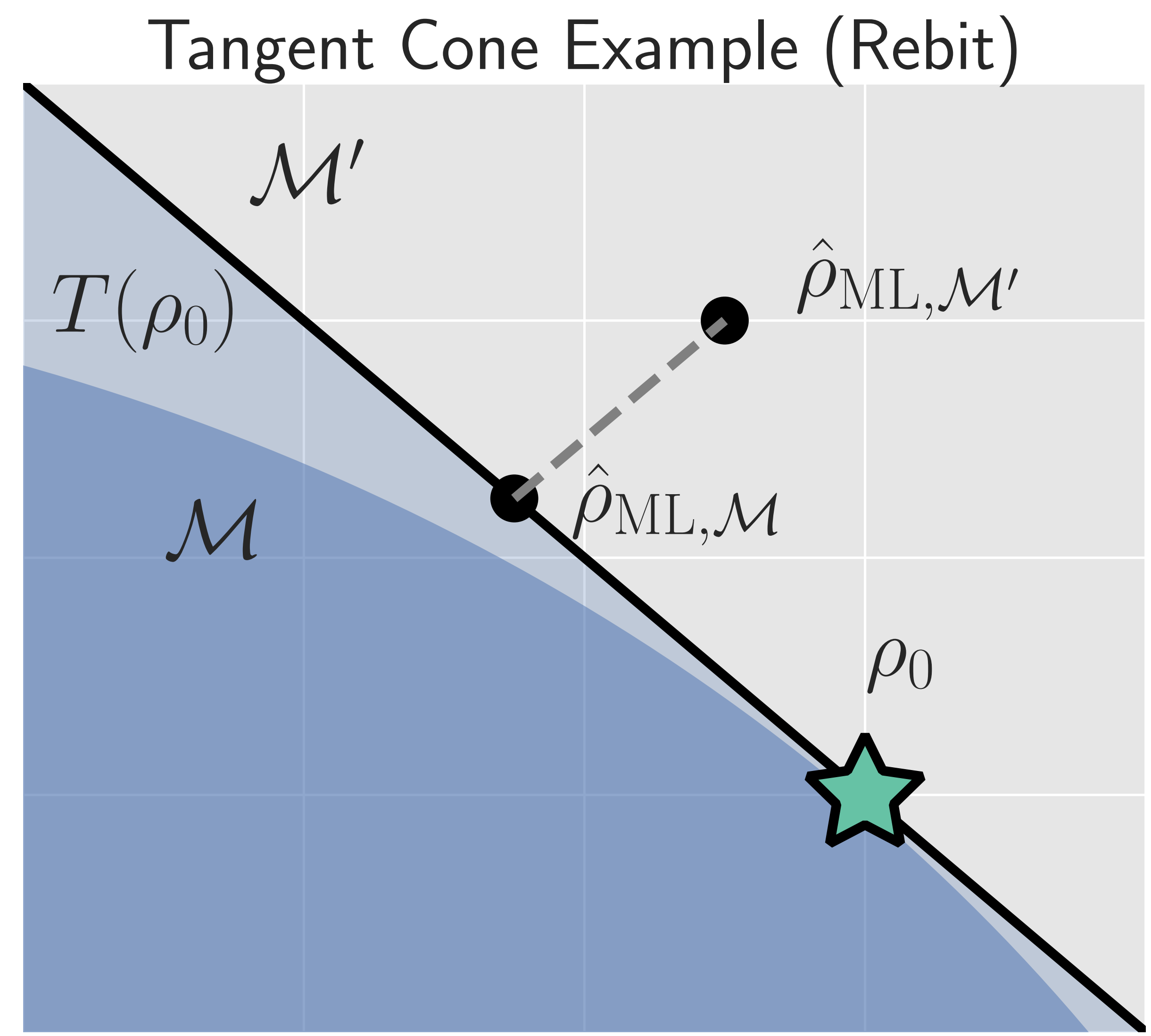
We **derive asymptotic properties** of models that satisfy MP-LAN.

Suppose a model satisfies MP-LAN. Then asymptotically,…

…the *local state space* is the *tangent cone*.

We can *zoom in* on the region of state space around the true state to determine the behavior of ML estimates.

Asymptotically, all the ML estimates are contained in a (shrinking) ball around the true state.



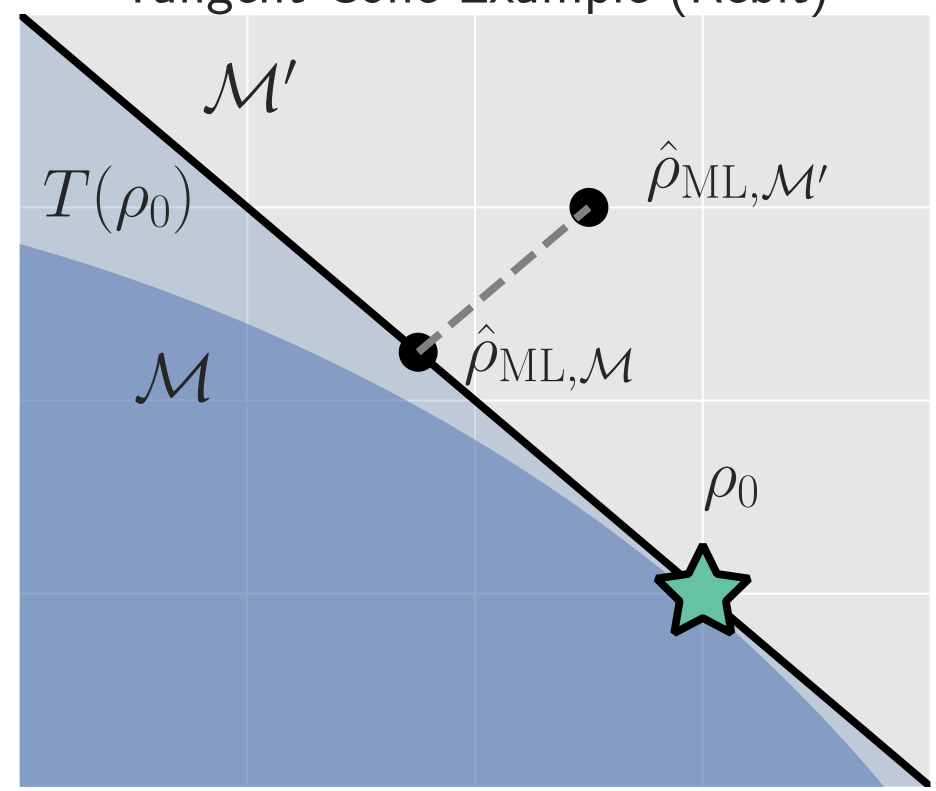
Suppose a model satisfies MP-LAN. Then asymptotically,…

…the ML estimate in the constrained model is
the *metric projection* of the ML estimate in the larger model.

Because $\mathcal{M}'$ satisfies LAN:

$$\mathcal{L}(\rho) \propto \mathrm{Exp}\left[-\tfrac{1}{2}\mathrm{Tr}(\rho - \hat{\rho}_{\mathrm{ML},\mathcal{M}'})F(\rho - \hat{\rho}_{\mathrm{ML},\mathcal{M}'})\right]$$

Tangent Cone Example (Rebit)



Maximize the likelihood over $\mathcal{M}$:

$$\hat{\rho}_{\mathrm{ML},\mathcal{M}} = \underset{\rho \in \mathcal{M}}{\mathrm{argmax}}\ \mathcal{L}(\rho)$$

Asymptotically, equal to minimizing
Fisher-adjusted distance
over tangent cone

$$\hat{\rho}_{\mathrm{ML},\mathcal{M}} = \underset{\rho \in T(\rho_0)}{\mathrm{argmin}}\ \mathrm{Tr}[(\rho - \hat{\rho}_{\mathrm{ML},\mathcal{M}'})F(\rho - \hat{\rho}_{\mathrm{ML},\mathcal{M}'})]$$

“Metric projection onto the tangent cone”

Suppose a model satisfies MP-LAN. Then asymptotically,…

…the increase in goodness of fit (as measured by loglikelihood)
is equal to increase in squared error (as measured by Fisher information).

The loglikelihood ratio statistic
comparing two models is

$$\lambda(\mathcal{M}_1, \mathcal{M}_2) = -2\log\left(\frac{\mathcal{L}(\hat{\rho}_{\mathrm{ML},\mathcal{M}_1})}{\mathcal{L}(\hat{\rho}_{\mathrm{ML},\mathcal{M}_2})}\right)$$

"How much better does one model do
in fitting the data compared to another?"

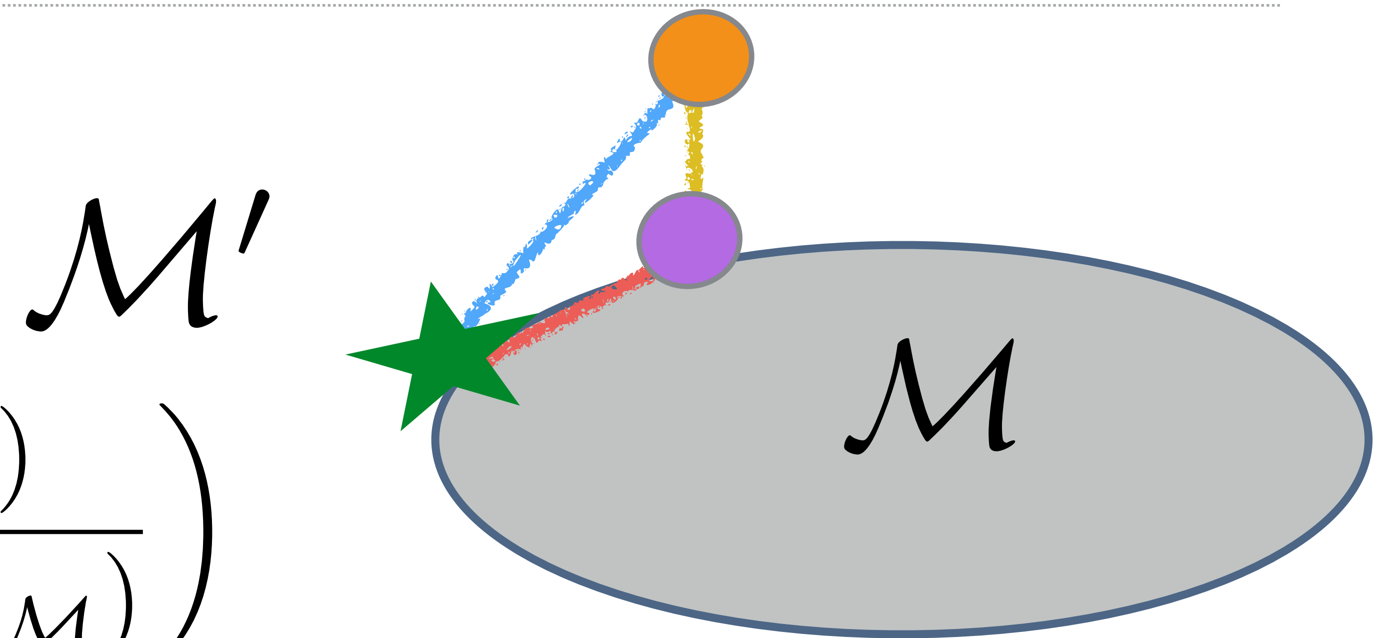
For analysis purposes: introduce a *reference model*

$$\lambda(\mathcal{M}_1, \mathcal{M}_2) = \lambda(\rho_0, \mathcal{M}_2) - \lambda(\rho_0, \mathcal{M}_1)$$

Suppose a model satisfies MP-LAN. Then asymptotically,…

…the increase in goodness of fit (as measured by loglikelihood) is equal to increase in squared error (as measured by Fisher information).

Because $\mathcal{M}$ satisfies MP- LAN,



$$\lambda(\rho_0, \mathcal{M}) = -2\log\left(\frac{\mathcal{L}(\rho_0)}{\mathcal{L}(\hat{\rho}_{\mathrm{ML},\mathcal{M}})}\right)$$

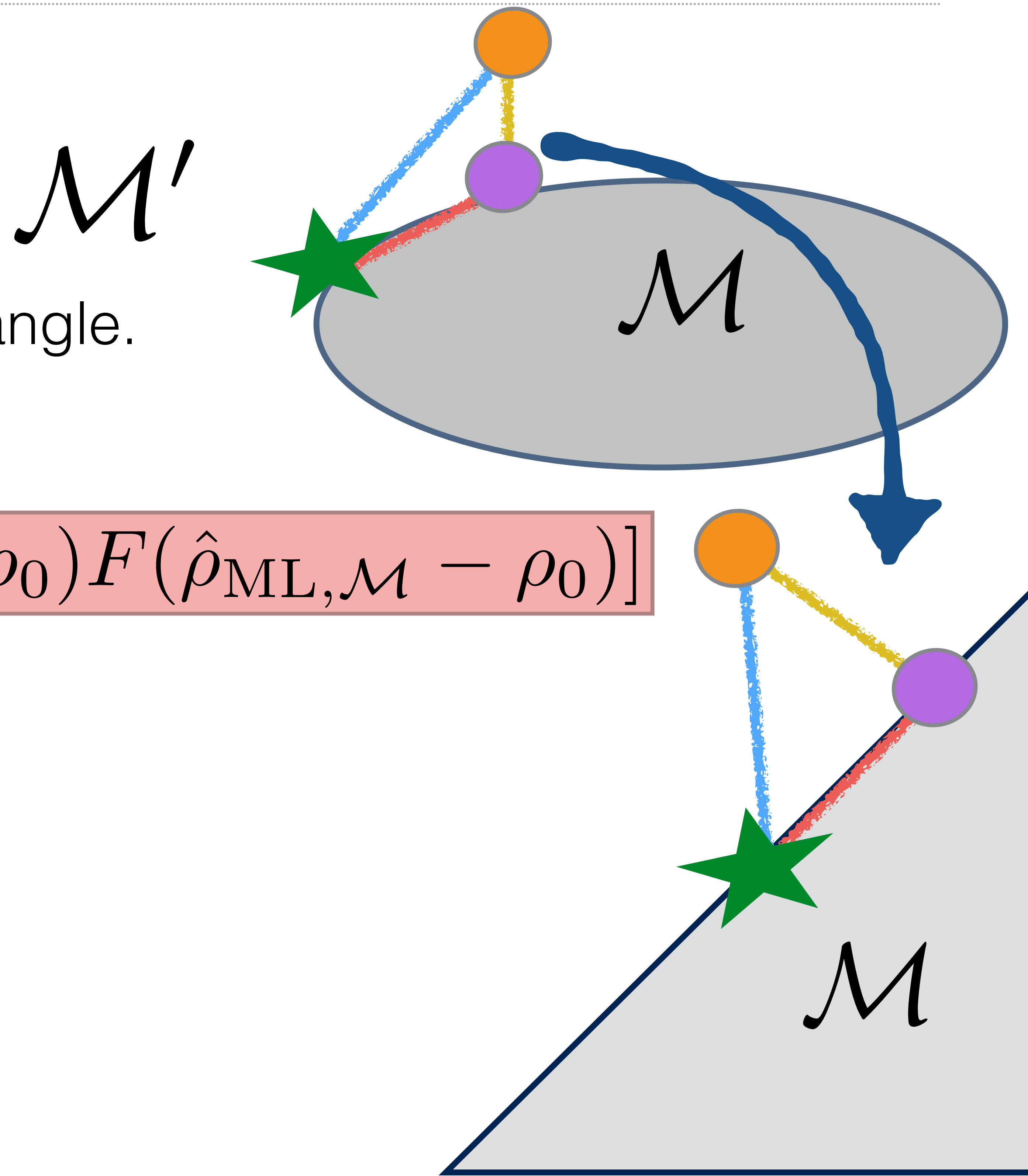
$$\underset{\mathrm{LAN}}{\longrightarrow} \mathrm{Tr}[(\rho_0 - \hat{\rho}_{\mathrm{ML},\mathcal{M}'})F(\rho_0 - \hat{\rho}_{\mathrm{ML},\mathcal{M}'})]$$

$$- \mathrm{Tr}[(\hat{\rho}_{\mathrm{ML},\mathcal{M}} - \hat{\rho}_{\mathrm{ML},\mathcal{M}'})F(\hat{\rho}_{\mathrm{ML},\mathcal{M}} - \hat{\rho}_{\mathrm{ML},\mathcal{M}'})]$$

Suppose a model satisfies MP-LAN. Then asymptotically,…

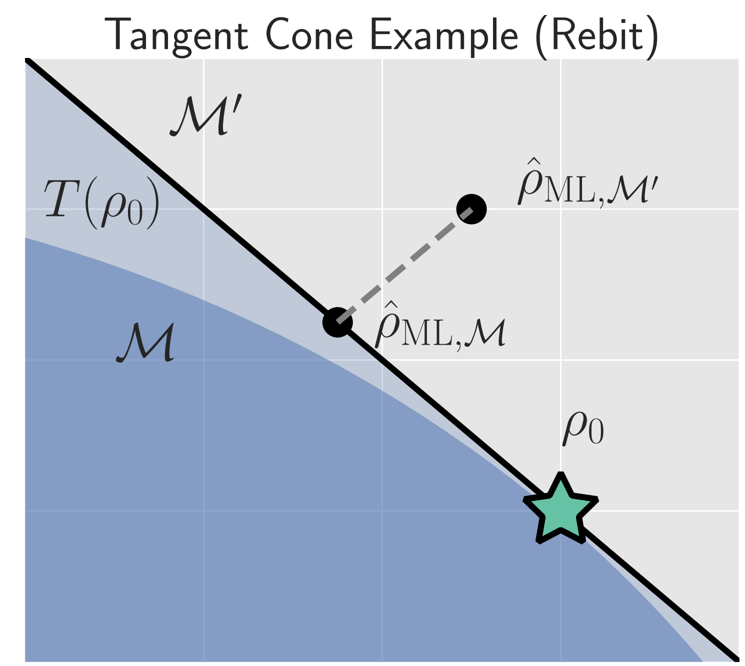
…the increase in goodness of fit (as measured by loglikelihood) is equal to increase in squared error (as measured by Fisher information).

Because the *local state space* is the *tangent cone*, the metric projection must complete a right triangle.



$$\lambda(\rho_0, \mathcal{M}) = \mathrm{Tr}[(\hat{\rho}_{\mathrm{ML},\mathcal{M}} - \rho_0)F(\hat{\rho}_{\mathrm{ML},\mathcal{M}} - \rho_0)]$$

Suppose a model satisfies MP-LAN. Then asymptotically,…



…the *local state space* is the *tangent cone*.

…the ML estimate in the constrained model is the *metric projection* of the ML estimate in the larger model.

$$\hat{\rho}_{\mathrm{ML},\mathcal{M}} = \underset{\rho \in T(\rho_0)}{\mathrm{argmin}}\ \mathrm{Tr}[(\rho - \hat{\rho}_{\mathrm{ML},\mathcal{M}'})F(\rho - \hat{\rho}_{\mathrm{ML},\mathcal{M}'})]$$

…the increase in goodness of fit (as measured by loglikelihood) is equal to increase in squared error (as measured by Fisher information).

$$\lambda(\rho_0, \mathcal{M}) \ = \ \mathrm{Tr}[(\hat{\rho}_{\mathrm{ML},\mathcal{M}} - \rho_0)F(\hat{\rho}_{\mathrm{ML},\mathcal{M}} - \rho_0)]$$


East Sandia Mountains - 2017 September 24

New J. Phys. **20** (2018) 023050 https://doi.org/10.1088/1367-2630/aaa7e2

New Journal of Physics
The open access journal at the forefront of physics

Deutsche Physikalische Gesellschaft DPG
IOP Institute of Physics

Published in partnership with: Deutsche Physikalische Gesellschaft and the Institute of Physics

PAPER

Behavior of the maximum likelihood in quantum state tomography

Travis L Scholten[1,2] **and Robin Blume-Kohout**[1,2]

[1] Center for Computing Research (CCR), Sandia National Laboratories, United States of America
[2] Center for Quantum Information and Control (CQuIC), University of New Mexico, United States of America

We show how to **generalize LAN** for models with convex boundaries.

We **derive asymptotic properties** of models that satisfy MP-LAN.

We **provide a replacement** to the classical Wilks theorem for models that satisfy MP-LAN.

A canonical model selection rule uses the *loglikelihood ratio statistic.*

Recall the loglikelihood ratio statistic comparing two models is

$$\lambda(\mathcal{M}_1, \mathcal{M}_2) = -2 \log \left(\frac{\mathcal{L}(\hat{\rho}_{\mathrm{ML}}, \mathcal{M}_1)}{\mathcal{L}(\hat{\rho}_{\mathrm{ML}}, \mathcal{M}_2)} \right)$$

Tells us how much better one model fits the data than the other.

Because of extra parameters, one model might fit better because it's fitting *noise* - how to correct for that?

Need to know the null behavior —
what happens when both models are equally good?

The *Wilks theorem* describes the null behavior of this statistic.

$$\lambda(\mathcal{M}_1, \mathcal{M}_2) = -2\log\left(\frac{\mathcal{L}(\hat{\rho}_{\mathrm{ML},\mathcal{M}_1})}{\mathcal{L}(\hat{\rho}_{\mathrm{ML},\mathcal{M}_2})}\right)$$

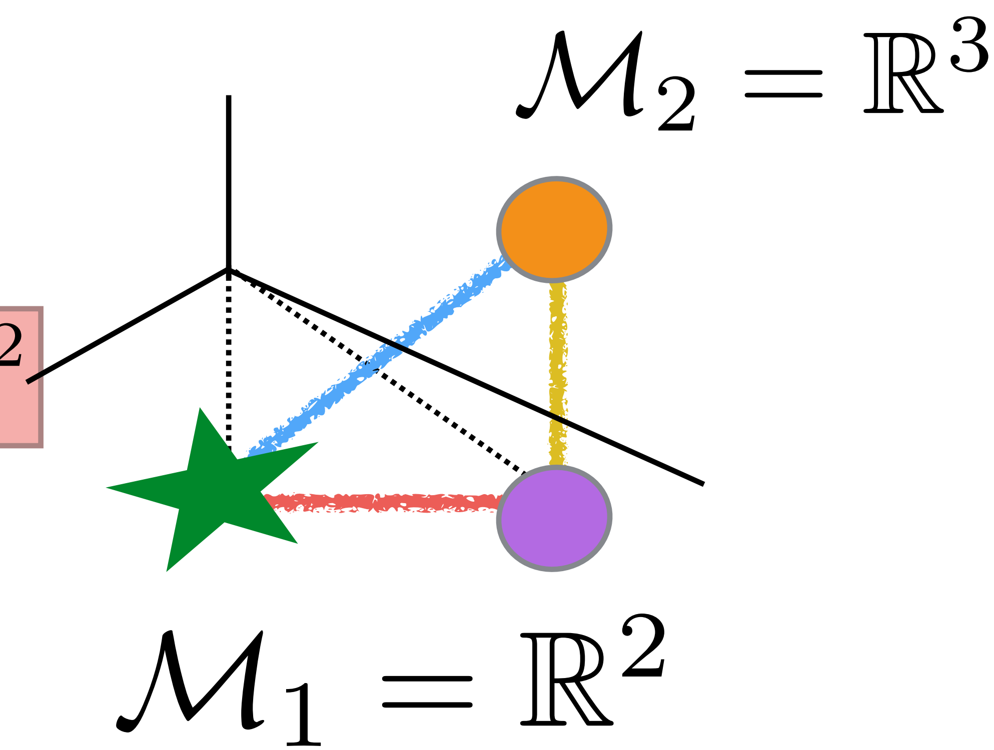
<u>Wilks theorem (1938):</u>

Assume that $\rho_0 \in \mathcal{M}_1, \mathcal{M}_2$, that $\mathcal{M}_1 \subset \mathcal{M}_2$, and that $\mathcal{M}_1, \mathcal{M}_2$ satisfy LAN. Then $\lambda \sim \chi^2_{\dim(\mathcal{M}_2)-\dim(\mathcal{M}_1)}$.

Key insight: $\hat{\rho}_{\mathrm{ML},\mathcal{M}_2} = \hat{\rho}_{\mathrm{ML},\mathcal{M}_1} \oplus \sigma$

$\sigma \sim \mathcal{N}(0, \mathcal{I})$

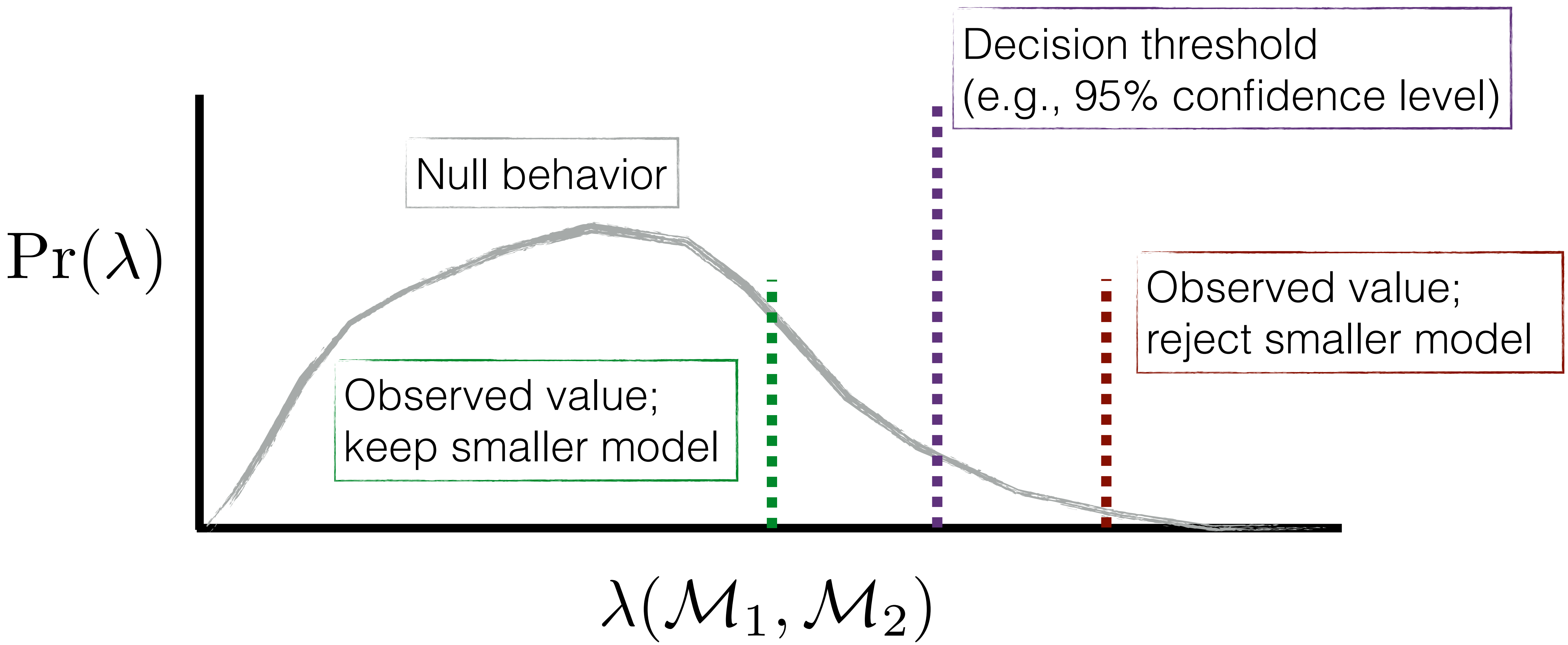
$$\lambda = ||\hat{\rho}_{\mathrm{ML},\mathcal{M}_2} - \rho_0||^2 - ||\hat{\rho}_{\mathrm{ML},\mathcal{M}_1} - \rho_0||^2$$
$$= ||\sigma||^2$$



Knowing the null behavior allows us to formulate a *decision rule* for choosing between two models.

If both models were equally good, would I expect to see the value of the statistic that I actually observed?

Set a *threshold* for judging when to reject smaller model.

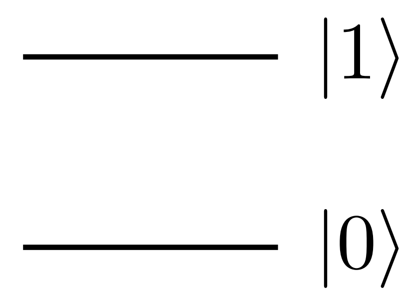


For models that might be useful in state tomography, the Wilks theorem fails spectacularly.

Define $\mathcal{M}_d = \{\rho \in \mathcal{B}(\mathcal{H}_d) \mid \mathrm{Tr}(\rho) = 1, \rho \geq 0\}$

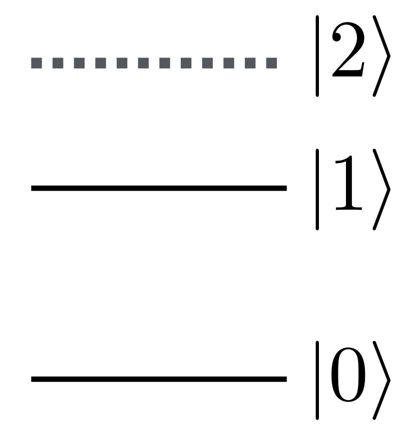
(d-dimensional density matrices)

Useful for, e.g., determining additional degrees of freedom are present.



"I built a qubit"

$\mathcal{M}_2$



"I built a qutrit"

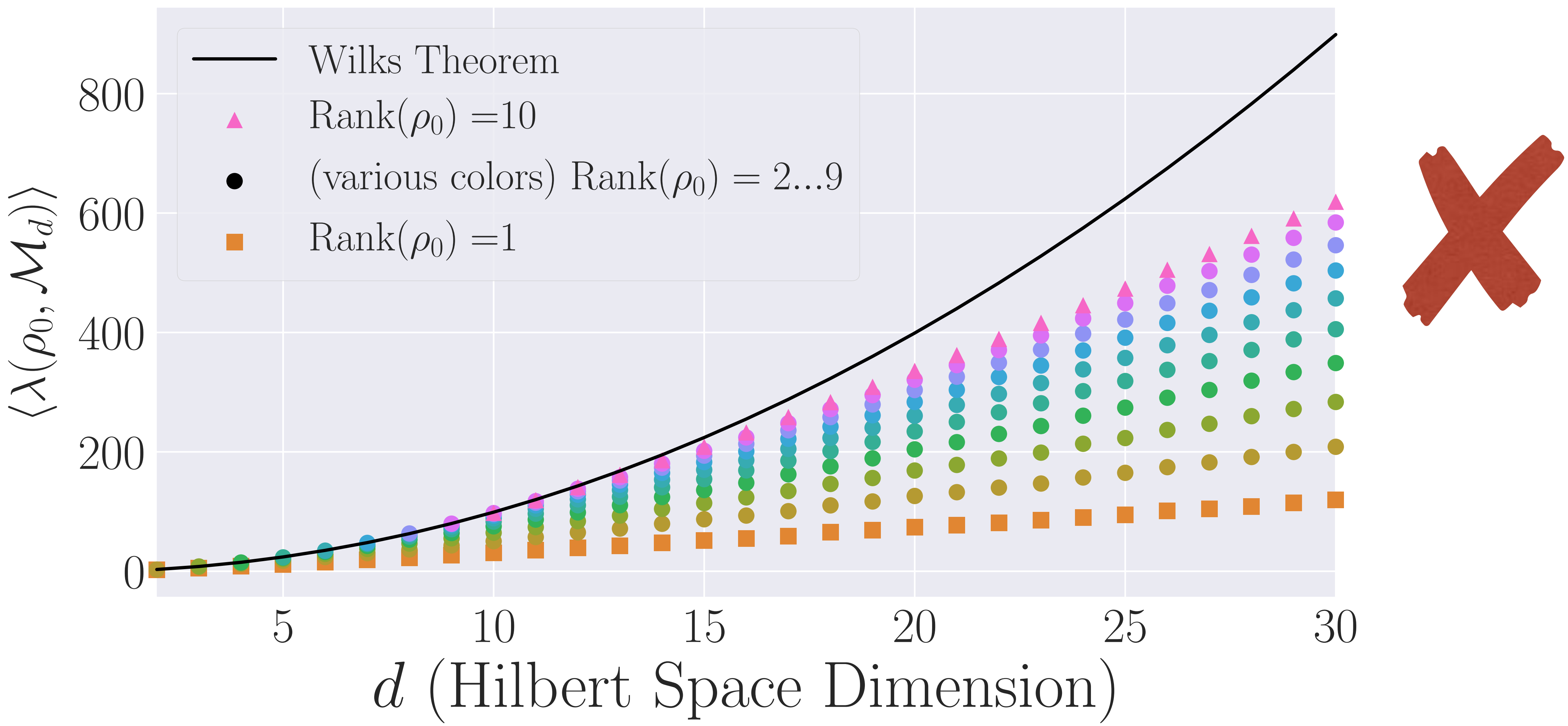
$\mathcal{M}_3$

Practical concern for most physical architectures (superconductors, ions, etc) for detecting *leakage*.

For models that might be useful in state tomography, the Wilks theorem fails spectacularly.

Define $\mathcal{M}_d = \{\rho \in \mathcal{B}(\mathcal{H}_d) \mid \mathrm{Tr}(\rho) = 1, \rho \geq 0\}$ (d-dimensional density matrices)

Wilks theorem says $\langle \lambda(\rho_0, \mathcal{M}_d) \rangle = d^2 - 1$



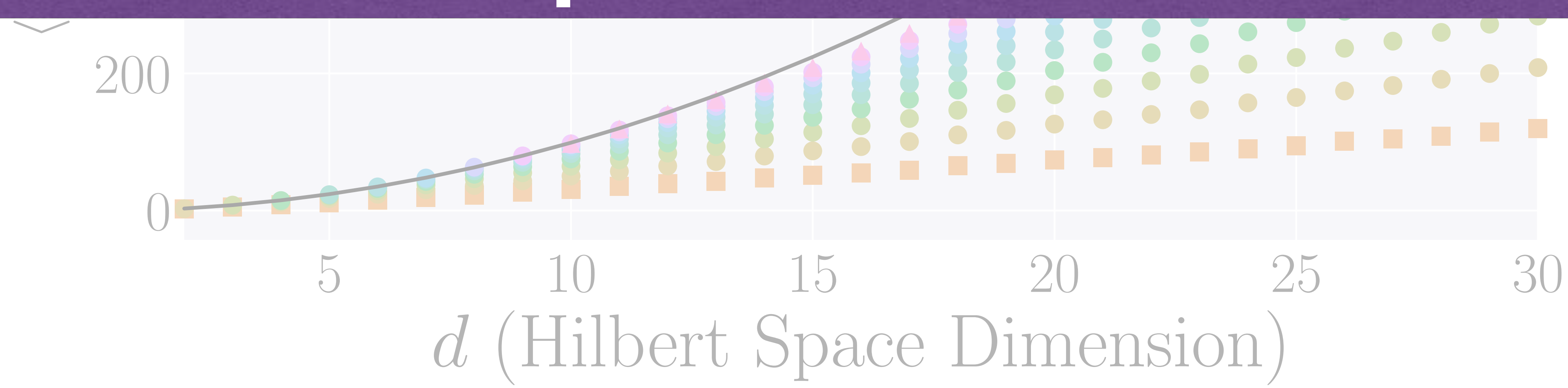
For models that might be useful in state tomography, the Wilks theorem fails spectacularly.

Define $\mathcal{M}_d = \{\rho \in \mathcal{B}(\mathcal{H}_d) \mid \mathrm{Tr}(\rho) = 1, \rho \geq 0\}$ (d-dimensional density matrices)

Wilks theorem says $\langle \lambda(\rho_0, \mathcal{M}_d) \rangle = d^2 - 1$

Quantum state space **doesn't satisfy LAN**, so **Wilks theorem cannot be applied**!

Can we derive a **replacement** using the fact **state space satisfies MP-LAN**?



Our replacement for Wilks approximates the expected value of the loglikelihood ratio statistic.

Because state space satisfies MP-LAN,

$$\lambda(\rho_0, \mathcal{M}_d) = \mathrm{Tr}[(\rho_0 - \hat{\rho}_{\mathrm{ML},\mathcal{M}_d})F(\rho_0 - \hat{\rho}_{\mathrm{ML},\mathcal{M}_d})]$$

where $\mathcal{M}_d = \{\rho \in \mathcal{B}(\mathcal{H}_d) \mid \mathrm{Tr}(\rho) = 1, \rho \geq 0\}$

$$\langle \lambda(\rho_0, \mathcal{M}_d) \rangle = \ ??$$

Our replacement for Wilks approximates the expected value of the loglikelihood ratio statistic.

Because state space satisfies MP-LAN,

$$\lambda(\rho_0, \mathcal{M}_d) = \mathrm{Tr}[(\rho_0 - \hat{\rho}_{\mathrm{ML},\mathcal{M}_d})F(\rho_0 - \hat{\rho}_{\mathrm{ML},\mathcal{M}_d})]$$

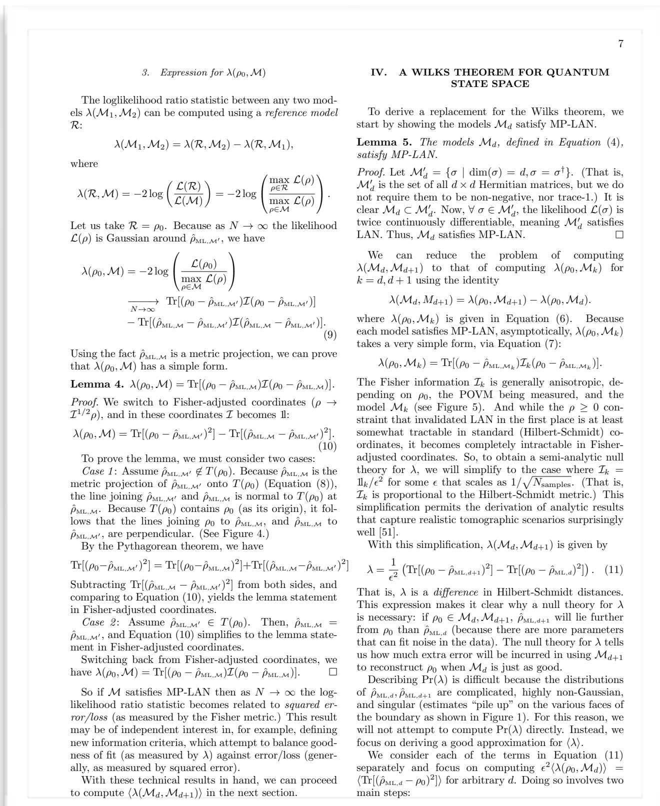
where $\mathcal{M}_d = \{\rho \in \mathcal{B}(\mathcal{H}_d) \mid \mathrm{Tr}(\rho) = 1, \rho \geq 0\}$

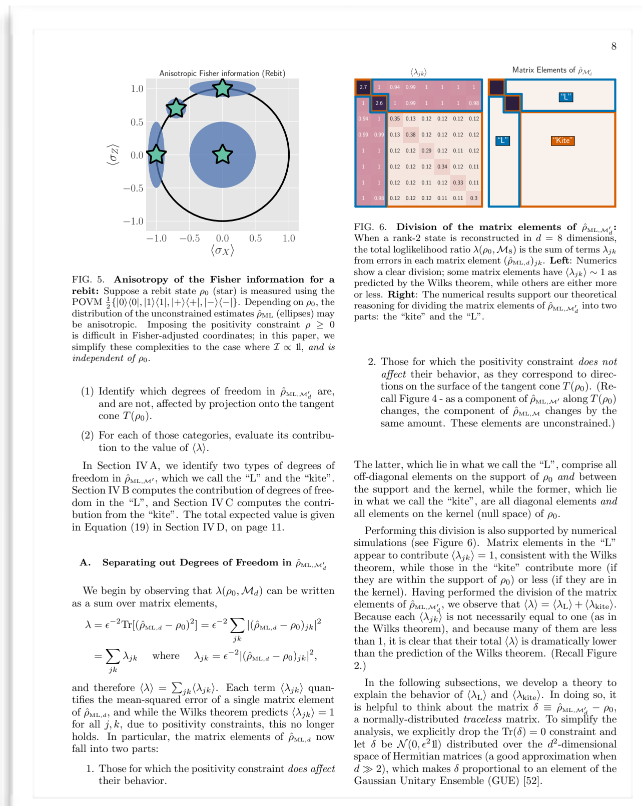
$$\langle \lambda(\rho_0, \mathcal{M}_d) \rangle = \ ??$$

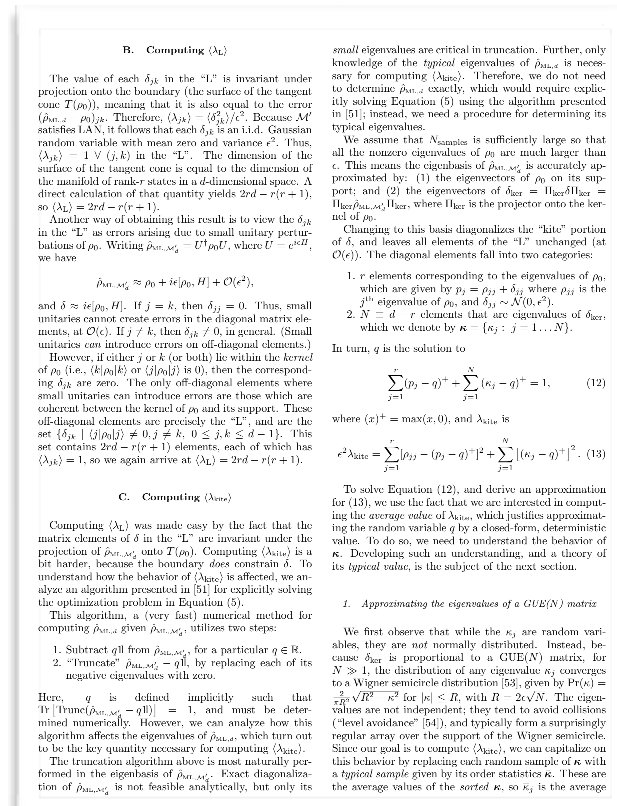
To make progress, we **assume the Fisher information is isotropic**.

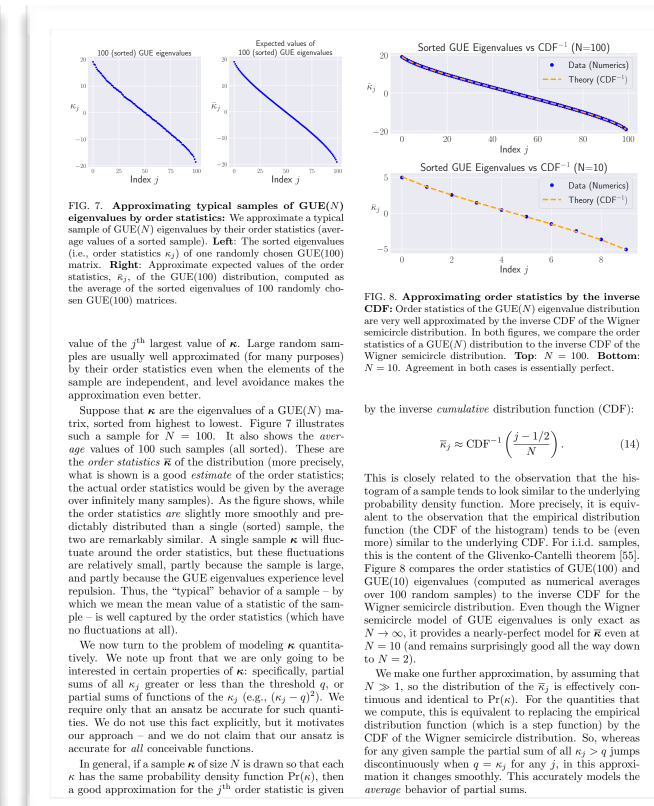
(Never actually happens…except in trivial cases)

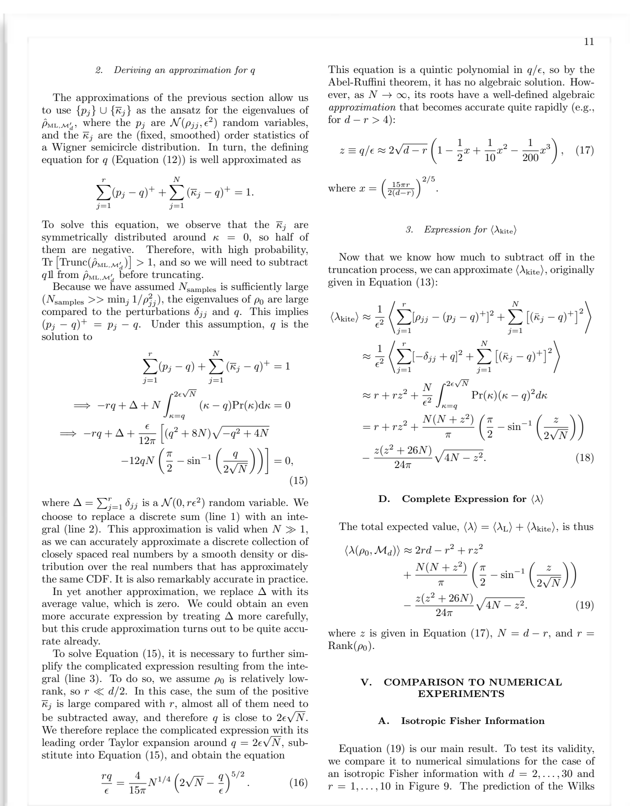
Even with that assumption, the calculation* was non-trivial…











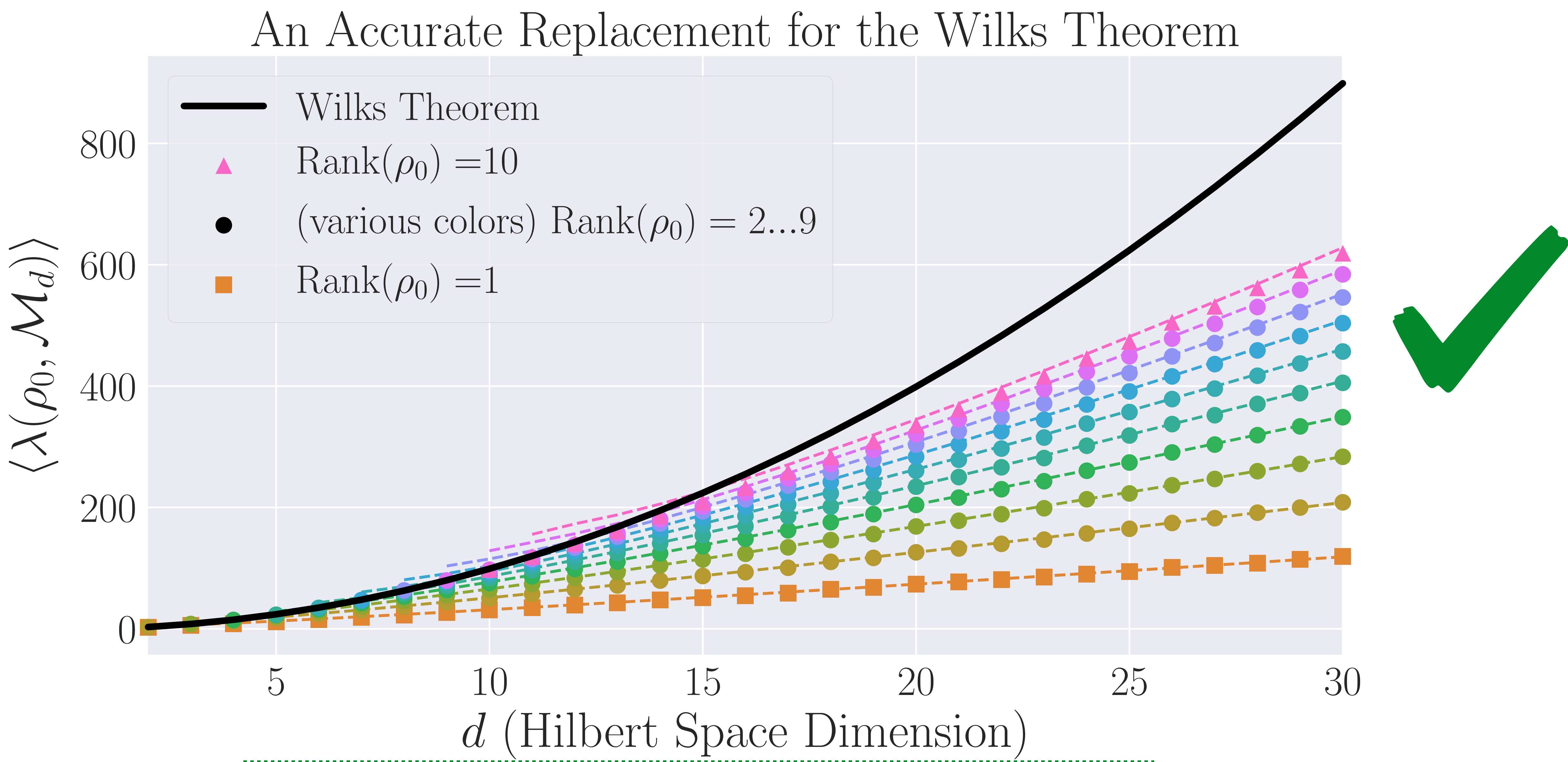
Geometry of the tangent cone
("L" and the "kite")

Truncating unconstrained ML estimates
(IBM algorithm)

Random matrix theory
(Gaussian Unitary Ensemble)

*Scholten & Blume-Kohout, NJP **20** 023050 (2018)

…but our result had much better agreement!

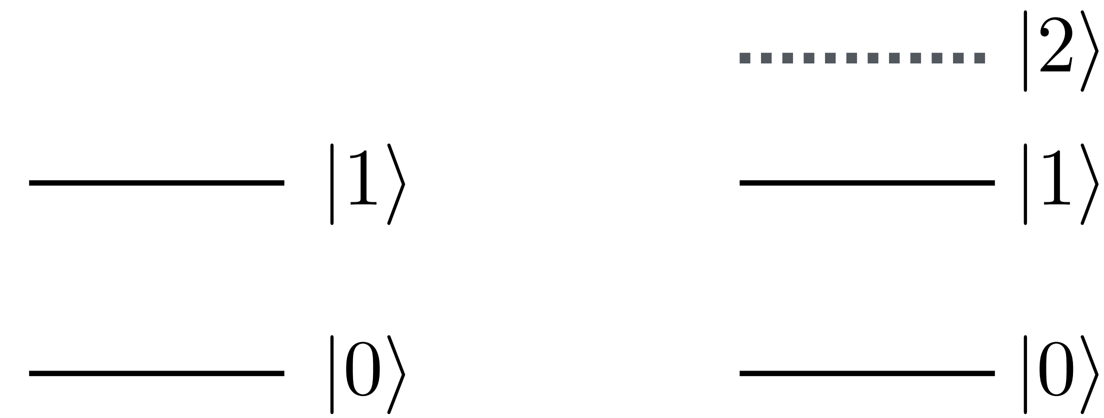


$$\langle\lambda(\rho_0, \mathcal{M}_d)\rangle \approx 2rd - r^2 + rz^2 + \frac{N(N+z^2)}{\pi}\left(\frac{\pi}{2} - \sin^{-1}\left(\frac{z}{2\sqrt{N}}\right)\right) - \frac{z(z^2+26N)}{24\pi}\sqrt{4N - z^2}.$$

$$z \approx 2\sqrt{d-r}\left(1 - \frac{1}{2}x + \frac{1}{10}x^2 - \frac{1}{200}x^3\right) \quad x = \left(\frac{15\pi r}{2(d-r)}\right)^{2/5} \quad N = d - r$$

What can we do with this result?

Choose a Hilbert space dimension for a quantum system (with prior information about rank).



"I built a qubit" $\mathcal{M}_2$

"I built a qutrit" $\mathcal{M}_3$

Reason about the *effective* number of parameters in the model.

Classically: $$\langle \lambda(\rho_0, \mathcal{M}) \rangle = \dim(\mathcal{M})$$

"Quantumly": $$\langle \lambda(\rho_0, \mathcal{M}) \rangle \sim \text{``}\dim(\mathcal{M})\text{''}$$

Connections to compressed sensing?

What can we do with this result?

Cho
dime
syst
abou

$|2\rangle$
$|1\rangle$
$|0\rangle$
trit”

Reason about the *effective* number of parameters in the model.

Classically: $$\langle \lambda(\rho_0, \mathcal{M}) \rangle = \dim(\mathcal{M})$$

“Quantumly”: $$\langle \lambda(\rho_0, \mathcal{M}) \rangle \sim \text{“}\dim(\mathcal{M})\text{”}$$

Connections to compressed sensing?

Recent results in classical compressed sensing show how the *geometry of convex optimization* affects performance.

Suppose we acquire data of the form $\mathbf{z}_0 = A\mathbf{x}_0$

Estimate the signal using convex optimization:

$$\hat{\mathbf{x}}_0 = \underset{\mathbf{x}\in\mathcal{M}}{\operatorname{argmin}} \ f(\mathbf{x}) \quad \text{s.t.} \quad \mathbf{z}_0 = A\mathbf{x}$$

To reason about properties of the estimate, look at *descent cone*:

$$D(f, \mathbf{x}) = \bigcup_{\tau>0} \ \{\mathbf{y} \in \mathcal{M} \ : \ f(\mathbf{x} + \tau\mathbf{y}) \leq f(\mathbf{x})\}$$

Living on the edge: Phase transitions in convex programs with random data
D Amelunxen, M Lotz, M McCoy, & J Tropp
Information and Inference, 2014

Recent results in classical compressed sensing show how the *geometry of convex optimization* affects performance.

"Interaction" of descent cone and null space of A determines whether we can uniquely recover the signal:

Fact: $\mathbf{x}_0$ is the unique optimal point of minimizing a proper convex function if, and only if, $D \cap \mathrm{null}(A) = \{\mathbf{0}\}$.

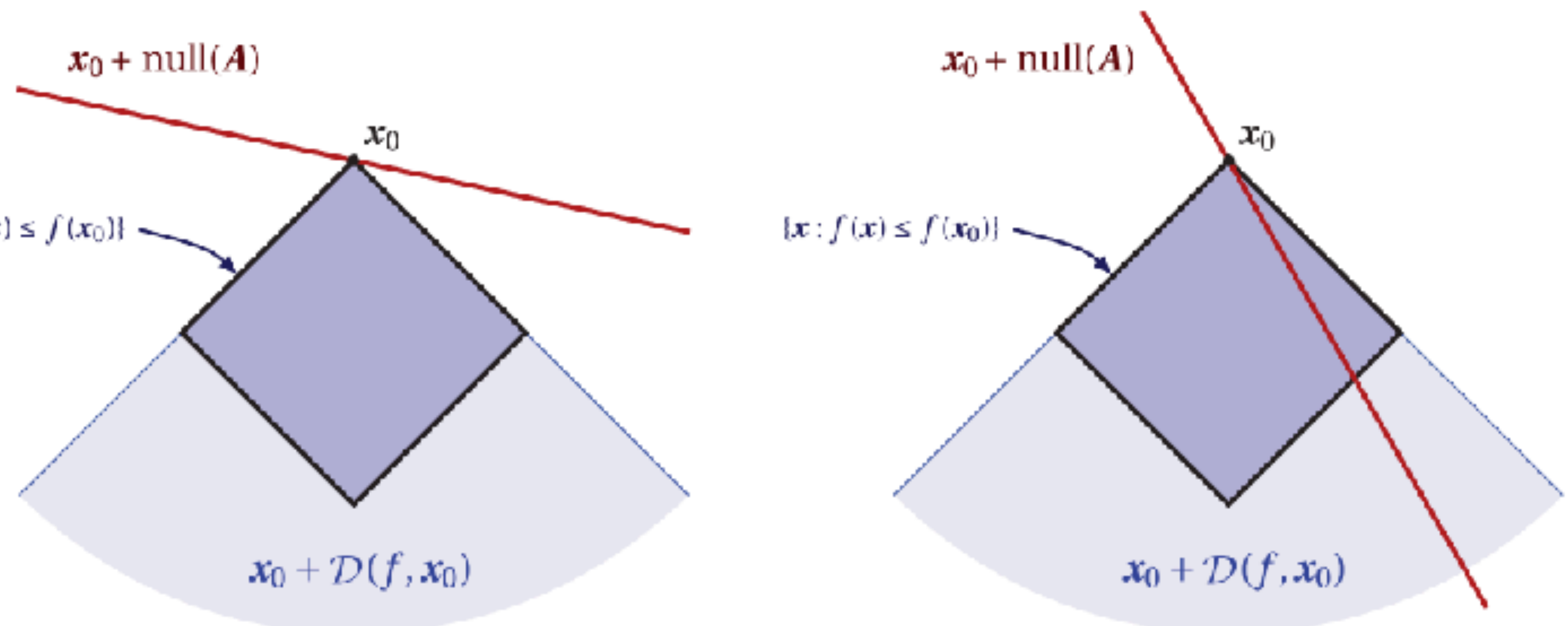


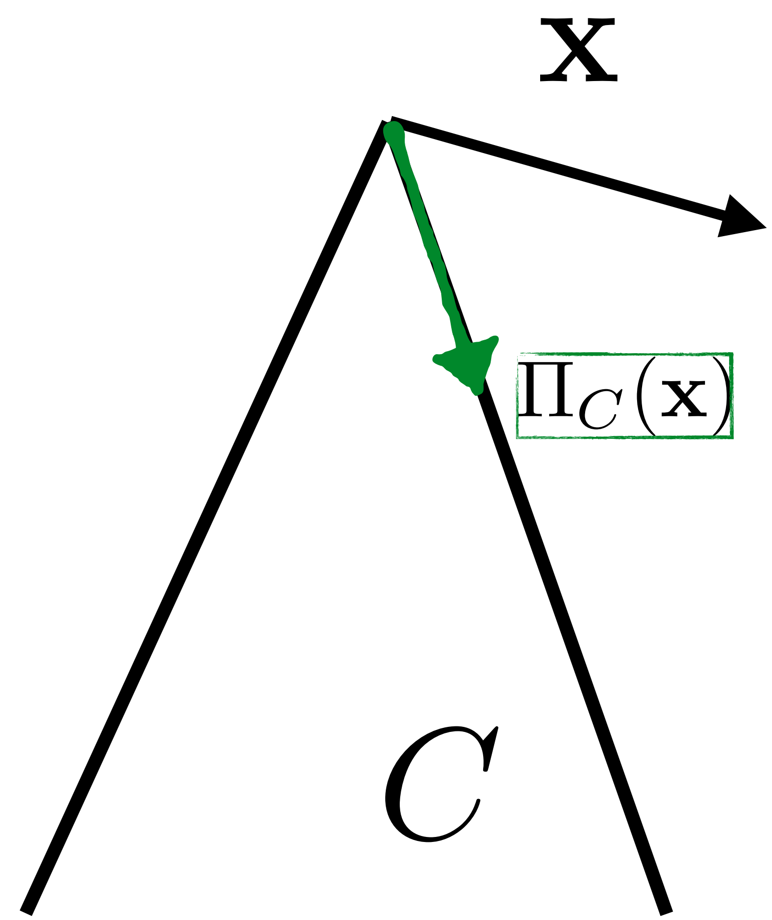
FIG. 4. The optimality condition for a regularized inverse problem. The condition for the regularized linear inverse problem (2.4) to succeed requires that the descent cone $\mathcal{D}(f, x_0)$ and the null space $\mathrm{null}(A)$ do not share a ray. [Left] The regularized linear inverse problem succeeds. [Right] The regularized linear inverse problem fails.

Living on the edge: Phase transitions in convex programs with random data
D Amelunxen, M Lotz, M McCoy, & J Tropp
Information and Inference, 2014

Computing the *statistical dimension* of the descent cone tells us when unique recovery is possible.

Given cone C, define the *metric projection* of a point onto C as

$$\Pi_C(\mathbf{x}) = \underset{\mathbf{y}\in C}{\operatorname{argmin}}\ ||\mathbf{x}-\mathbf{y}||$$



The statistical dimension of the cone is

$$\delta(C) = \langle ||\Pi_C(\mathbf{x})||^2 \rangle \quad \mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathcal{I})$$

If C is an L-dimensional subspace, $\delta(C) = L$

Living on the edge: Phase transitions in convex programs with random data
D Amelunxen, M Lotz, M McCoy, & J Tropp
Information and Inference, 2014

Computing the *statistical dimension* of the descent cone tells us when unique recovery is possible.

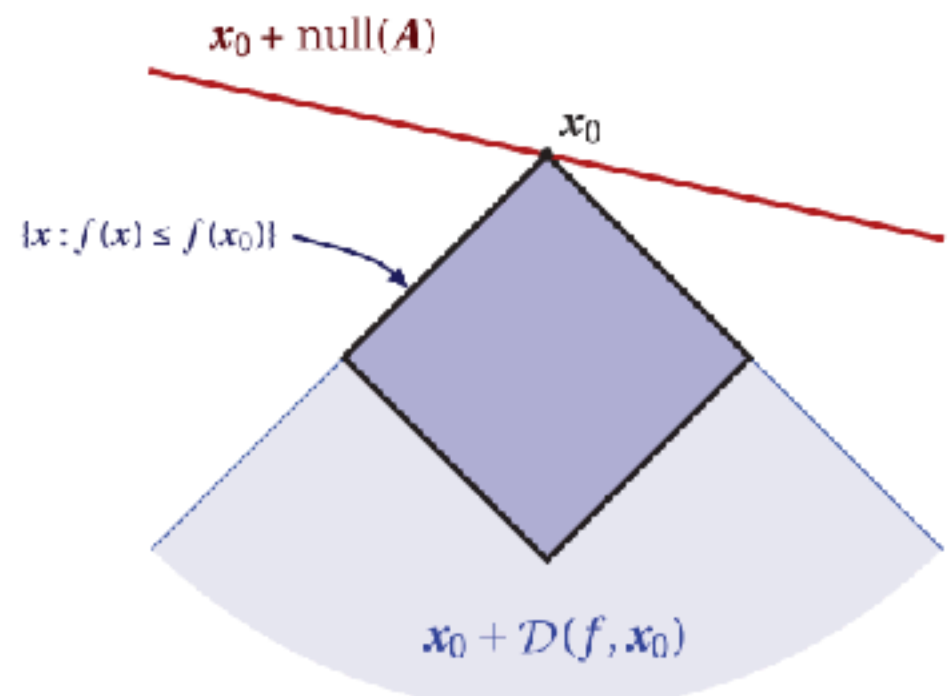
Theorem: Suppose $A \in \mathbb{R}^{m \times d}$, with i.i.d $\mathcal{N}(0,1)$ entries.

If $m \geq \delta(D(f, \mathbf{x}_0)) + \sqrt{8 \log(4/\eta)}\sqrt{d}$,

then recovery is possible with probability $\geq 1 - \eta$.

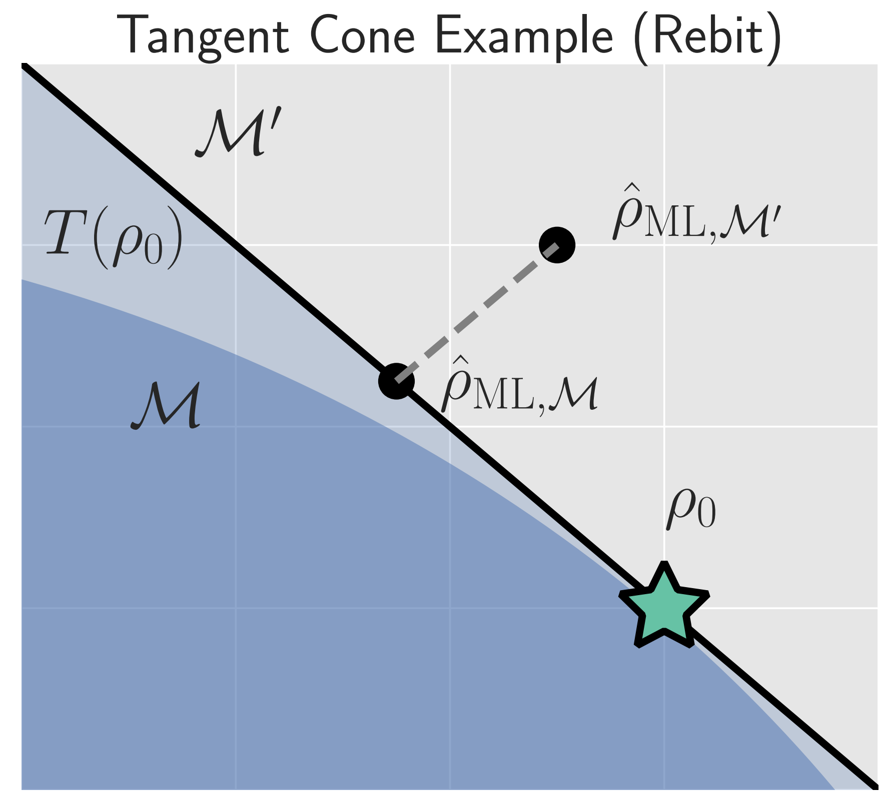
With enough constraints, the null space doesn't intersect the descent cone.

"Skinnier" descent cones have lower statistical dimension, meaning fewer measurements are necessary.



Living on the edge: Phase transitions in convex programs with random data
D Amelunxen, M Lotz, M McCoy, & J Tropp
Information and Inference, 2014

Our replacement for the Wilks theorem gives the statistical dimension of the tangent cone!



Start with unconstrained ML estimates

$$\hat{\rho}_{\mathrm{ML},\mathcal{M}'_d} \sim \mathcal{N}(\rho_0, \mathcal{I}/N)$$

Compute metric projections onto tangent cone

$$\hat{\rho}_{\mathrm{ML},\mathcal{M}_d} = \Pi_{T(\rho_0)}(\hat{\rho}_{\mathrm{ML},\mathcal{M}'_d})$$

Expected value of loglikelihood ratio statistic is the statistical dimension

$$\delta(T(\rho_0)) = \langle \mathrm{Tr}[(\Pi_{T(\rho_0)}(\hat{\rho}_{\mathrm{ML},\mathcal{M}_d}) - \rho_0)^2] \rangle$$

Does this result provide new insight into quantum compressed sensing?

We understand how the positivity constraint in state tomography affects reconstruction.

Kalev, et. al, *npj Quantum Information* **1**, 15018 (2015)
Quantum tomography protocols with positivity are compressed sensing protocols

Theorem 1. Let P_0 be a positive semidefinite matrix with rank $(P_0) \leqslant r$, and let $\mathbf{y} = \mathcal{A}[P_0]$ be the measurement record obtained by a sensing map $\mathcal{A}$ that corresponds to compressing measurements for a rank-r Hermitian matrix. Then P_0 is the unique matrix within the set of positive semidefinite matrices of any rank that is consistent with the measurement record.

Requires restrictions on the measurement map

For the case of *Pauli measurements*, we can compute the number of outcomes necessary for reconstruction.

Gross, et. al, PRL **105**, 150401 (2010)
Quantum State Tomography via Compressed Sensing

Theorem 1 (low-rank tomography)—Let ρ be an arbitrary state of rank r. If $m = cdr\log^2 d$ randomly chosen Pauli expectations are known, then ρ can be uniquely reconstructed by solving the convex optimization problem (1) with probability of failure exponentially small in c.

For the case of *Gaussian measurements*, we can compute the number of outcomes necessary for reconstruction.

Chandrasekaran et. al, Foundations of Computational Mathematics (2012) 12:805–849
The Convex Geometry of Linear Inverse Problems

Proposition 3.11 *Let $\mathbf{x}^\star$ be an $m_1 \times m_2$ rank-r matrix with $m_1 \leq m_2$. Letting $\mathcal{A}$ denote the set of unit-Euclidean-norm rank-one matrices, we have that*

$$w\big(T_{\mathcal{A}}(\mathbf{x}^\star) \cap \mathbb{S}^{m_1 m_2 - 1}\big)^2 \leq 3r(m_1 + m_2 - r).$$

Thus $3r(m_1 + m_2 - r) + 1$ random Gaussian measurements suffice to recover $\mathbf{x}^\star$ via nuclear norm minimization with high probability.

Consequence:

In state tomography, $6rd - 3r^2$ measurements are sufficient.

In the limit of large dimension, our result for the statistical also yields a similar conclusion.

$$\langle\lambda\rangle \xrightarrow[d\to\infty]{} rd\left[6-\frac{20}{7}\left(\frac{15\pi r}{2d}\right)^{2/5}+\frac{20}{21}\left(\frac{15\pi r}{2d}\right)^{4/5}\right]-5r^2.$$

Statistical dimension of tangent cone	⟷	Number of measurements for quantum compressed sensing (Gaussian model)
Tangent cone in state space	⟷	Descent cone of some convex function??

Wrap up: geometry, model selection, and quantum compressed sensing

New J. Phys. **20** (2018) 023050 https://doi.org/10.1088/1367-2630/aaa7e2

New Journal of Physics
The open access journal at the forefront of physics

Deutsche Physikalische Gesellschaft DPG
IOP Institute of Physics
Published in partnership with: Deutsche Physikalische Gesellschaft and the Institute of Physics

PAPER

Behavior of the maximum likelihood in quantum state tomography

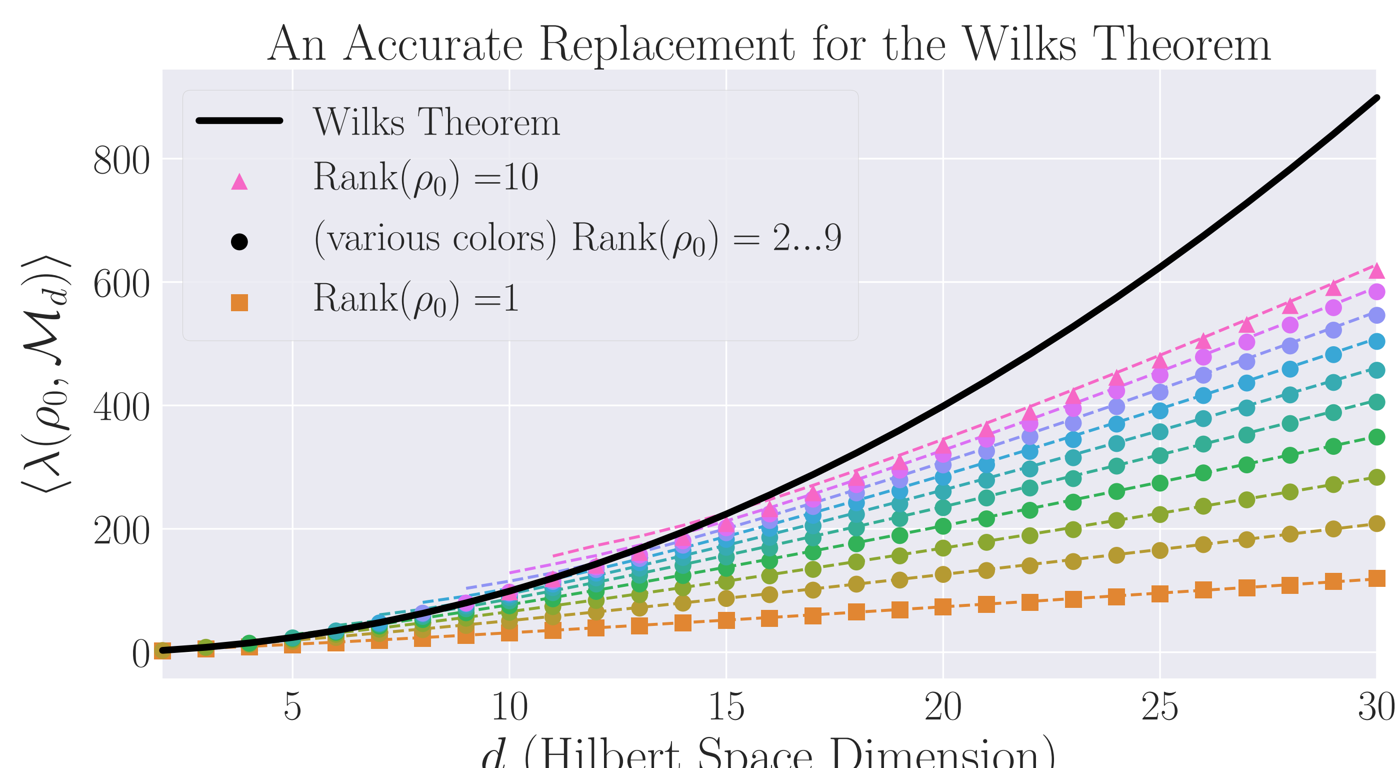
Travis L Scholten[1,2] **and Robin Blume-Kohout**[1,2]
[1] Center for Computing Research (CCR), Sandia National Laboratories, United States of America
[2] Center for Quantum Information and Control (CQuIC), University of New Mexico, United States of America

$\mathcal{M}'$ satisfies LAN

$\mathcal{M}$ satisfies “MP-LAN”

New generalization of LAN (applicable to quantum models)

Replacement for the classical Wilks theorem (model selection for state-space dimension)



Wrap up: geometry, model selection, and quantum compressed sensing

New J. Phys. **20** (2018) 023050 https://doi.org/10.1088/1367-2630/aaa7e2

New Journal of Physics
The open access journal at the forefront of physics

Deutsche Physikalische Gesellschaft DPG
IOP Institute of Physics
Published in partnership with: Deutsche Physikalische Gesellschaft and the Institute of Physics

PAPER

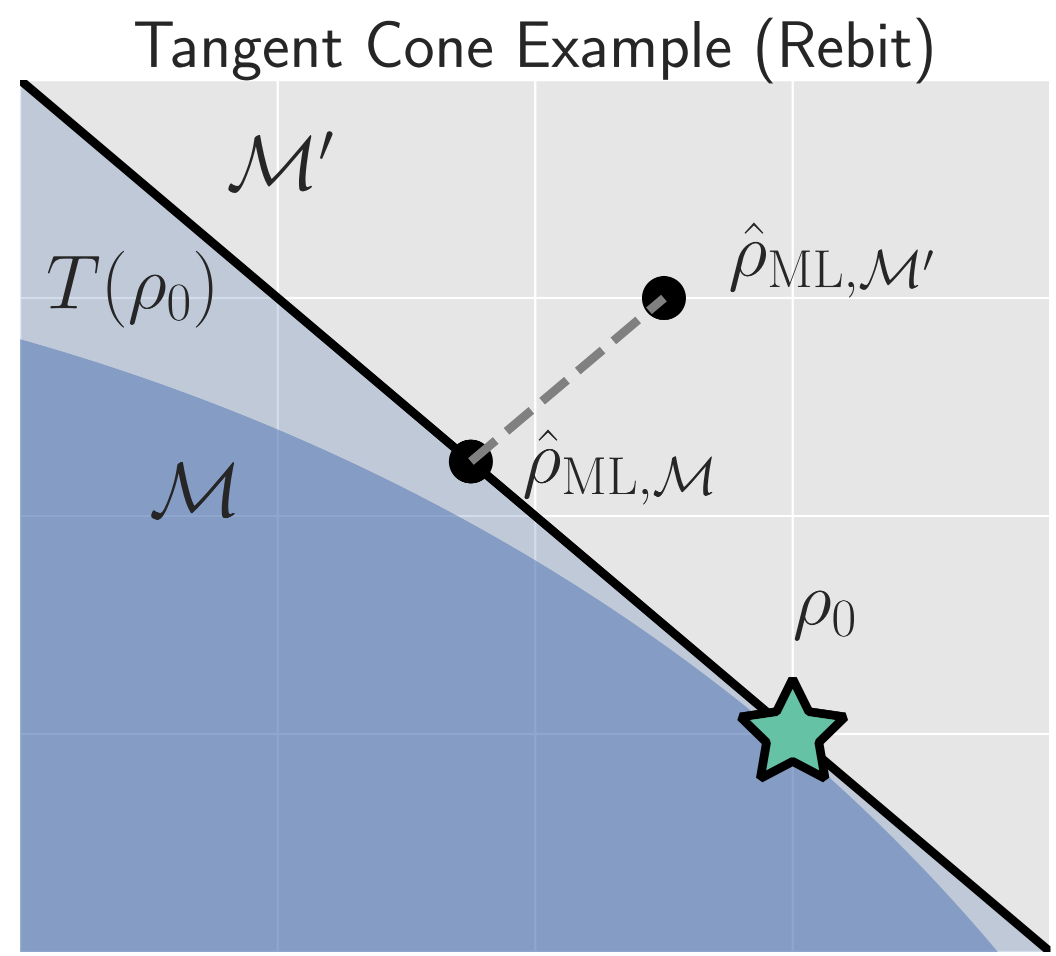
Behavior of the maximum likelihood in quantum state tomography

Travis L Scholten[1,2] and **Robin Blume-Kohout**[1,2]

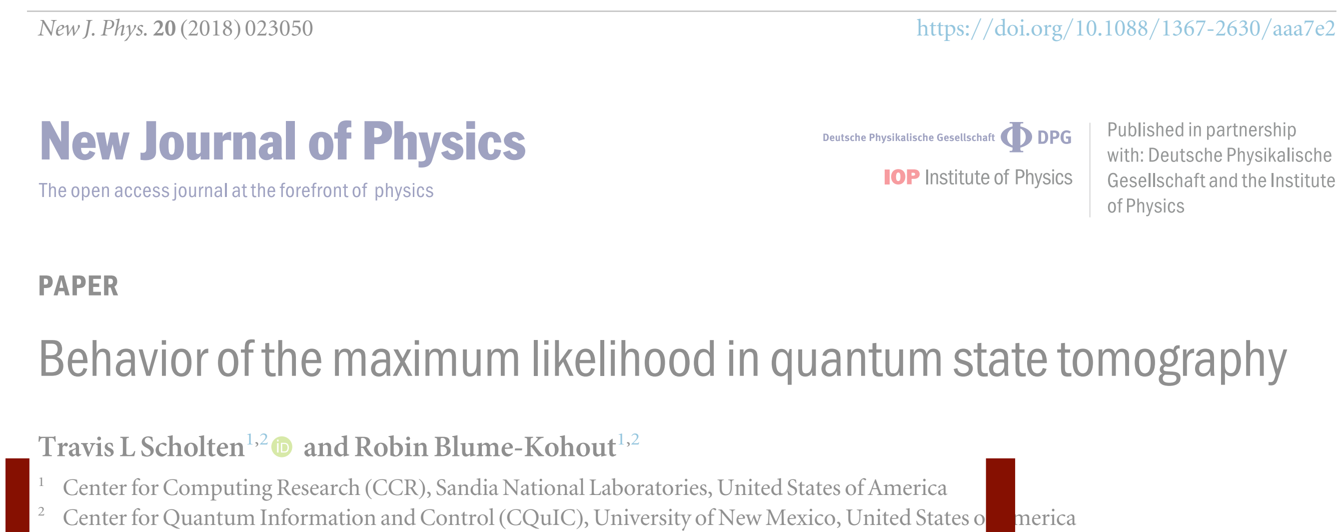
[1] Center for Computing Research (CCR), Sandia National Laboratories, United States of America
[2] Center for Quantum Information and Control (CQuIC), University of New Mexico, United States of America

Understanding the geometry of convex optimization

Connections to quantum compressed sensing



Wrap up: geometry, model selection, and quantum compressed sensing


New J. Phys. **20** (2018) 023050 https://doi.org/10.1088/1367-2630/aaa7e2

New Journal of Physics

The open access journal at the forefront of physics

Deutsche Physikalische Gesellschaft DPG
IOP Institute of Physics

Published in partnership with: Deutsche Physikalische Gesellschaft and the Institute of Physics

PAPER

Behavior of the maximum likelihood in quantum state tomography

Travis L Scholten[1,2] and Robin Blume-Kohout[1,2]

[1] Center for Computing Research (CCR), Sandia National Laboratories, United States of America

[2] Center for Quantum Information and Control (CQuIC), University of New Mexico, United States of America

Thank you!

Understanding the geometry of convex optimization

Connections to quantum compressed sensing