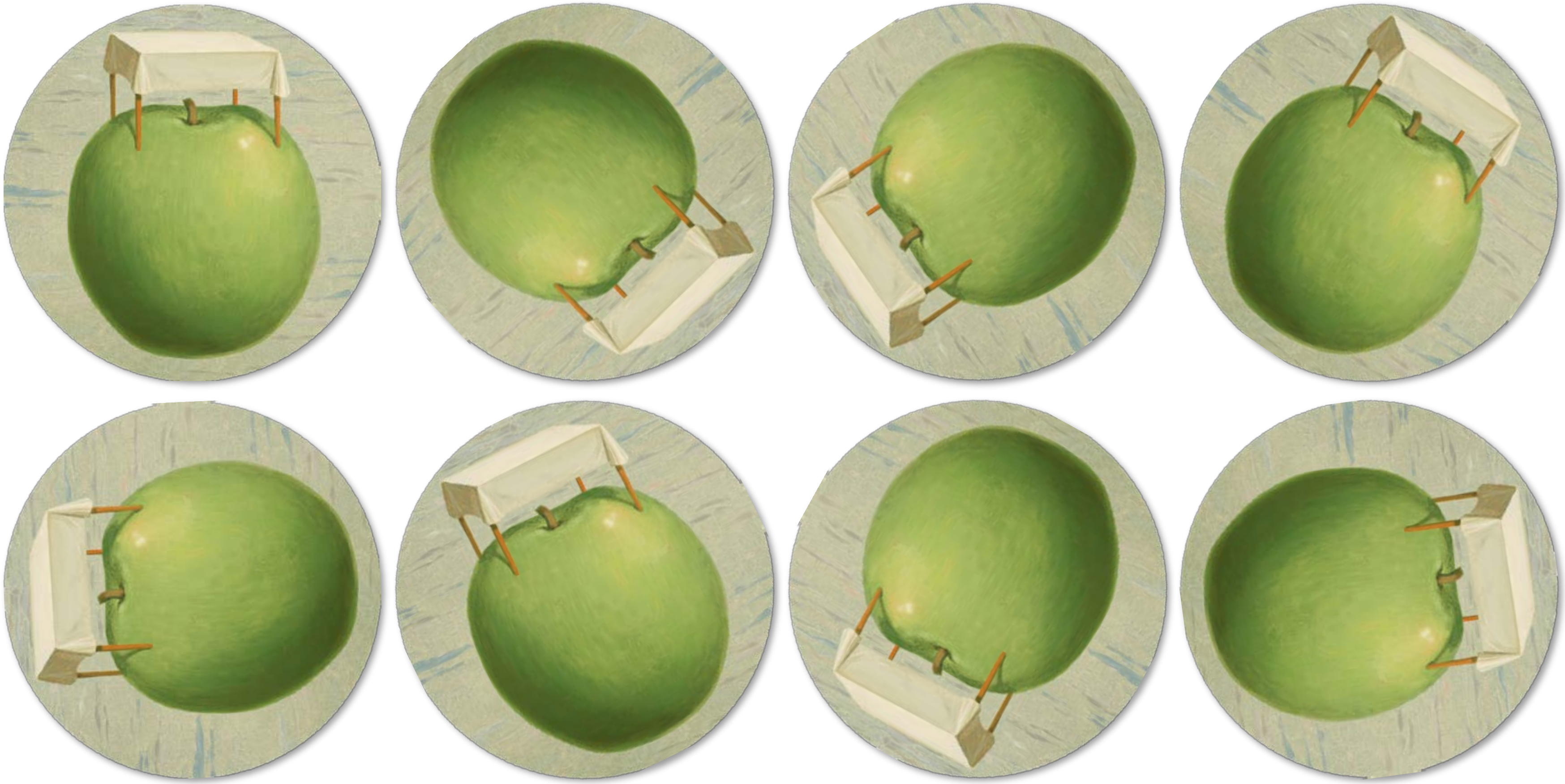
Near optimal bounds for Phase synchronization

Nicolas Boumal
Princeton University, Math & PACM

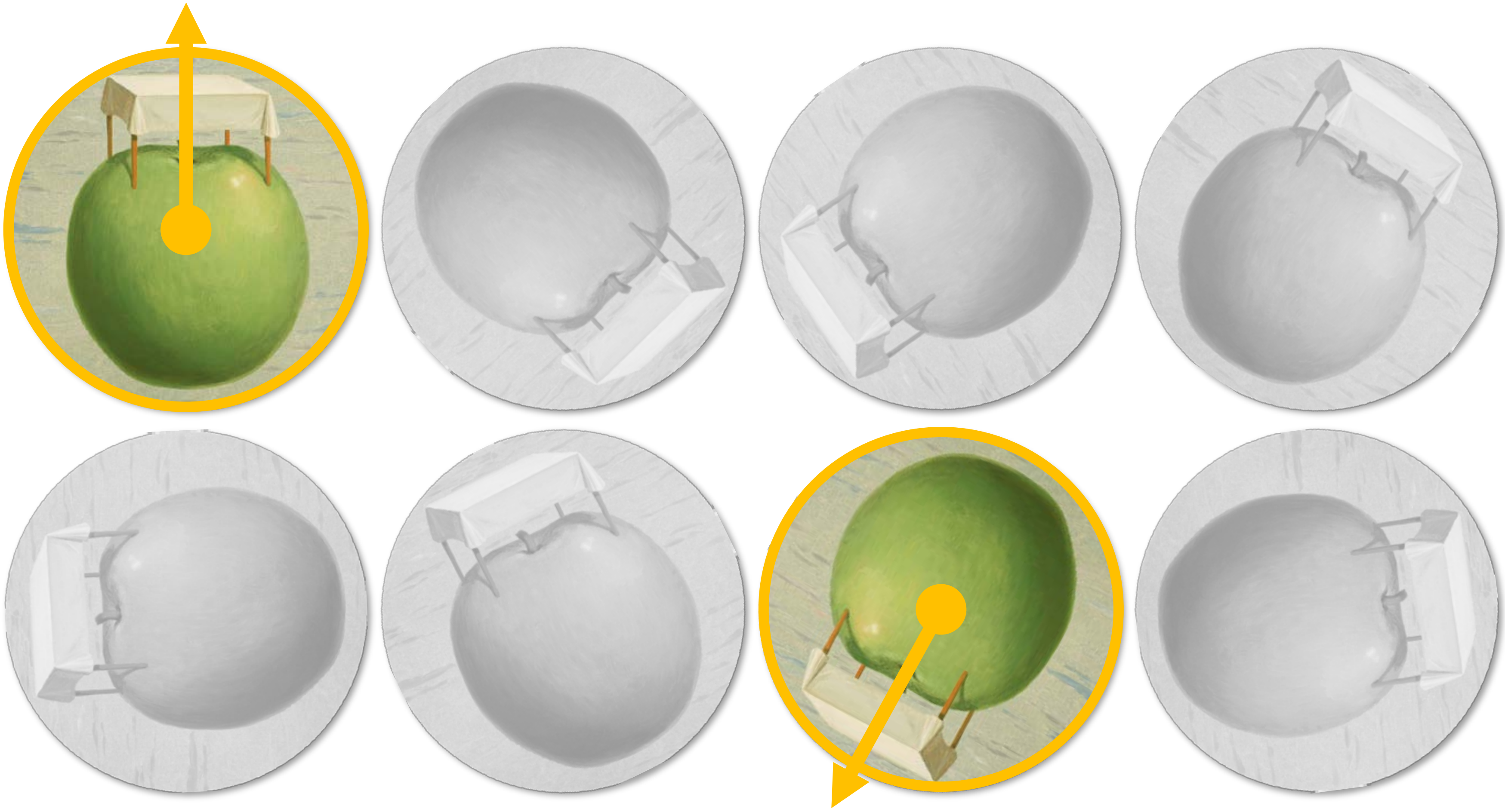
Joint work with
Afonso Bandeira, Amit Singer and Yiqiao Zhong

Alan Turing Institute, Dec. 20, 2017

The goal: estimate individual angles, from pairwise comparisons (up to global shift)



The goal: estimate individual angles,
from pairwise comparisons (up to global shift)



Estimate phases from relative info

Unknowns

$z_1, \ldots, z_n \in \mathbb{C}$, with $|z_1| = \cdots = |z_n| = 1$

$\mathbb{C}_1^n$

Data

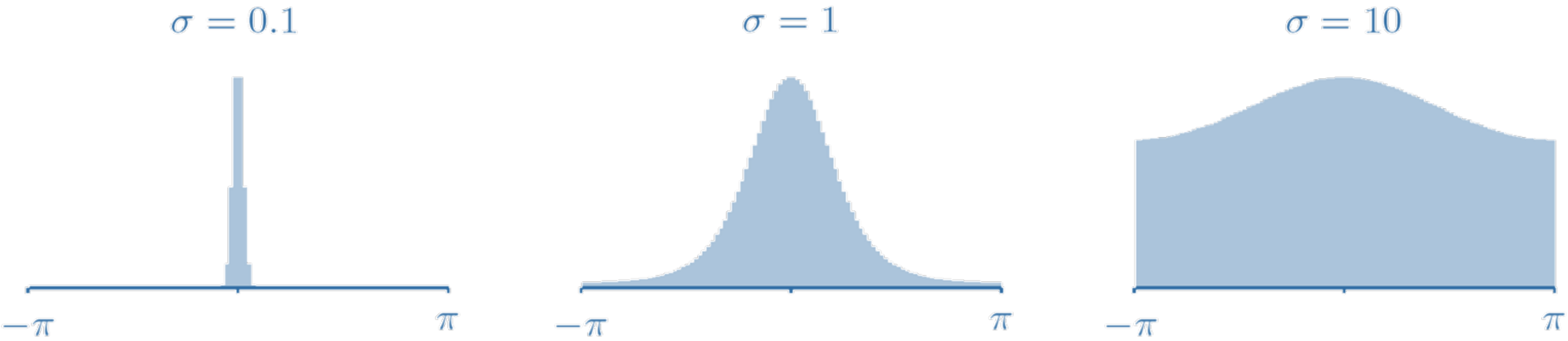
Noisy measurements of relative phases:

$C_{ij} = z_i z_j^* + \sigma W_{ij}$

What does additive Gaussian noise mean here?

Noise affects the phase of the measurement

$$C_{ij} = z_i z_j^* + \sigma W_{ij}$$



Maximum likelihood estimation

$$z \in \mathbb{C}_1^n, \qquad C = zz^* + \sigma W$$

Under Gaussian noise, the MLE solves a least-squares

$$\min_{x \in \mathbb{C}_1^n} \|C - xx^*\|_F^2 \quad \equiv \quad \max_{x \in \mathbb{C}_1^n} x^* C x$$

The MLE is NP-hard to compute

The optimization problem

$$\max_{x \in \mathbb{C}_1^n} \; x^* C x$$

has a quadratic cost $x^* C x$, and

nonconvex quadratic constraints $|x_i|^2 = 1$.

And yet, it's pretty easy to solve… in the right regime

Through SDP relaxation	Via generalized power method
With Riemannian optimization	Using dominant eigenvector

Classic lifting trick: rewrite everything in terms of $X = xx^*$

$$\max_{x \in \mathbb{C}_1^n} x^* C x$$

The cost

$$x^* C x = \mathrm{Trace}(x^* C x) = \mathrm{Trace}(CX)$$

The constraints

$$|x_i|^2 = 1 \Leftrightarrow x_i x_i^* = 1 \Leftrightarrow X_{ii} = 1$$

The link

$$\exists x: X = xx^* \Leftrightarrow X \succcurlyeq 0, \mathrm{rank}(X) = 1$$

Suggests a semidefinite relaxation

Recast our problem

$$\max_{x \in \mathbb{C}_1^n} x^* C x$$

Into

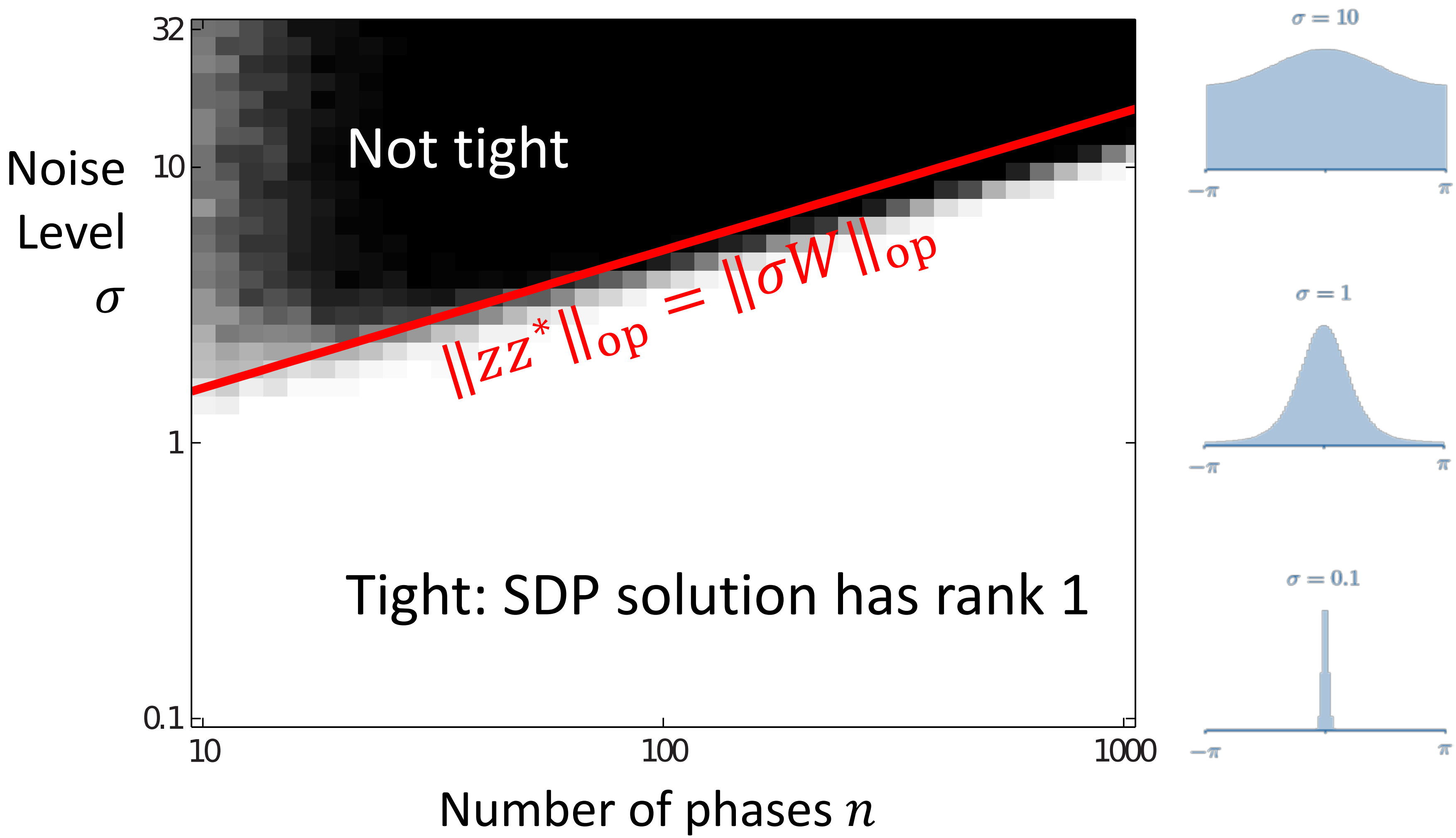
$$\max_{X \in \mathbb{C}^{n \times n}} \operatorname{Trace}(CX)$$

$$\operatorname{diag}(X) = \mathbf{1}$$

$$X \succcurlyeq 0$$

$$\sout{\operatorname{rank}(X) = 1}$$

The SDP seems tight for $\sigma \leq \sim\sqrt{n}$



We want to explain this theoretically; First, via a simpler relaxation.

Through SDP relaxation	Via generalized power method
With Riemannian optimization	Using dominant eigenvector

A spectral relaxation

Relax our problem from

$$\max_{x \in \mathbb{C}^n} x^* C x \quad \text{subject to} \quad |x_1| = \cdots = |x_n| = 1$$

to

$$\max_{x \in \mathbb{C}^n} x^* C x \quad \text{subject to} \quad |x_1|^2 + \cdots + |x_n|^2 = n$$

That is: compute dominant eigenvector of C.

The eigenvector estimator is not the MLE, but it is a good estimator

We prove (with high probability for large n):

$$\|x - z\|_2 \leq 12\sigma$$

$$\|x - z\|_\infty \leq \sim\sigma\sqrt{\frac{\log n}{n}} \qquad \text{—if } \sigma \leq \sim\sqrt{\frac{n}{\log n}}$$

(With global phase fixed by: $z^* x = |z^* x|$.)

Eigenvector x: error in 2-norm

The eigv. beats the signal: $z^*Cz \le x^*Cx$.

Hence, with $C = zz^* + \sigma W$ and $\|W\|_{\mathrm{op}} \le 3\sqrt{n}$,

$$n^2 + \sigma z^* W z \le |z^* x|^2 + \sigma x^* W x$$

$$\begin{aligned} n^2 - |z^* x|^2 &\le \sigma[x^* W x - z^* W z] \\ &= \sigma \Re\{(x - z)^* W (x + z)\} \\ &\le \sigma \|x - z\|_2 \|W\|_{\mathrm{op}} \|x + z\|_2 \end{aligned}$$

Divide by $n + |z^* x| \ge n$, use $\|x - z\|_2^2 = 2(n - |z^* x|)$:

$$\|x - z\|_2 \le 12\sigma$$

Eigenvector x: error in ∞-norm

$$C = zz^* + \sigma W \qquad Cx = \lambda x$$

$$|x_m - z_m| = \left|\frac{(Cx)_m}{\lambda} - z_m\right|$$

$$= \left|\frac{z^* x}{\lambda} z_m + \frac{\sigma}{\lambda}(Wx)_m - z_m\right|$$

$$\leq \left|\frac{|z^* x|}{\lambda} - 1\right| + \frac{\sigma}{\lambda}|(Wx)_m|$$

The first part is easy

$$\left|\frac{|z^* x|}{\lambda} - 1\right| \leq \frac{72\sigma^2 + 3\sigma\sqrt{n}}{n - 3\sigma\sqrt{n}} \leq \sim \frac{\sigma}{\sqrt{n}}$$

From 2-norm work:

$$n - 72\sigma^2 \leq |z^* x| \leq n$$

From $C = zz^* + \sigma W$ and $\|W\|_{\mathrm{op}} \leq 3\sqrt{n}$:

$$n - 3\sigma\sqrt{n} \leq \lambda \leq n + 3\sigma\sqrt{n}$$

$$\sigma \leq \sim \sqrt{\frac{n}{\log n}}$$

It comes down to $\|Wx\|_\infty \leq \sim\sqrt{n \log n}$

This is delicate because x is a dominant eigenvector of $C = zz^* + \sigma W$:

x and W are statistically dependent

Something that doesn't work but is informative:

$$\|Wx\|_\infty \leq \|Wz\|_\infty + \|W(x - z)\|_\infty$$

W and z *are* independent: this is small enough!

x and z can be too far apart for a 2-norm bound to work here…

Lessons learned from the failed attempt

$$\|Wx\|_\infty \leq \|Wz\|_\infty + \|W(x-z)\|_\infty$$

This could work if instead of comparing x to z we compared x to another vector, independent of W as well, yet much closer.

The main idea:

Introduce auxiliary problems

Auxiliary problems

Let W^m be W with 0's in row and column m.

Let $C^m = zz^* + \sigma W^m$.

Let x^m be the eigenvector estimator for C^m.

x^m is independent of the mth row of W, and
x^m is very close to x

Using the auxiliary problems to bound $\|Wx\|_\infty \leq \sim\sqrt{n \log n}$

$$|(Wx)_m| = |w_m^* x| \leq |w_m^* x^m| + |w_m^*(x - x^m)|$$

$|w_m^* x^m| \leq \sim\sqrt{n}$ owing to independence

$$|w_m^*(x - x^m)| \leq \sim\sqrt{n}\, \|x - x^m\|_2$$

Show $\|x - x^m\|_2 \leq \sim 1$, then take maximum over m.

$$\sigma \leq \sim \sqrt{\frac{n}{\log n}}$$

Showing x and x^m are close is a job for Davis-Kahan

$$C = zz^* + \sigma W^m + \sigma \Delta^m$$

x is dominant eigenvector of C

x^m is dominant eigenvector of C^m

Davis-Kahan: δ is eigengap $\lambda_1 - \lambda_2$ of C^m,

$$\|x - x^m\|_2 \leq \sqrt{2}\,\frac{\|\sigma \Delta^m x^m\|_2}{\delta - \|\sigma \Delta^m\|_2} \leq \sim \frac{\sigma}{n}\|\Delta^m x^m\|_2 \leq \sim \frac{\sigma}{\sqrt{n}}$$

$$\sigma \leq \sim \sqrt{\frac{n}{\log n}}$$

Putting it all together

$$|x_m - z_m| \leq \sim \frac{\sigma}{\sqrt{n}} + \frac{\sigma}{n}|(Wx)_m|$$

$$|(Wx)_m| \leq \sim\sqrt{n} + \sigma$$

Take union bound over m; still with high probability:

$$\|x - z\|_\infty \leq \sim\sigma\sqrt{\frac{\log n}{n}}$$

$$\sigma \leq \sim\sqrt{\frac{n}{\log n}}$$

Showed eigenvector method is great, but still doesn't explain SDP tightness.

Through SDP relaxation	Via generalized power method
With Riemannian optimization	Using dominant eigenvector

General idea: dual certification

$$\max_{X\in\mathbb{C}^{n\times n}} \ \mathrm{Trace}(CX)$$
$$\mathrm{diag}(X) = \mathbf{1}$$
$$X \succcurlyeq 0$$

Lemma:

X solves the SDP if

$$S(X) = \Re\{\mathrm{ddiag}(CX)\} - C \succcurlyeq 0$$

Proof: $0 \leq \mathrm{Tr}(S(X)X') \leq \mathrm{Tr}(CX) - \mathrm{Tr}(CX')$.

$S(xx^*) \succcurlyeq 0 \Rightarrow x$ is optimal *and* computable.

General idea: dual certification

Let x be the MLE, solution of

$$\max_{x \in \mathbb{C}_1^n} x^* C x$$

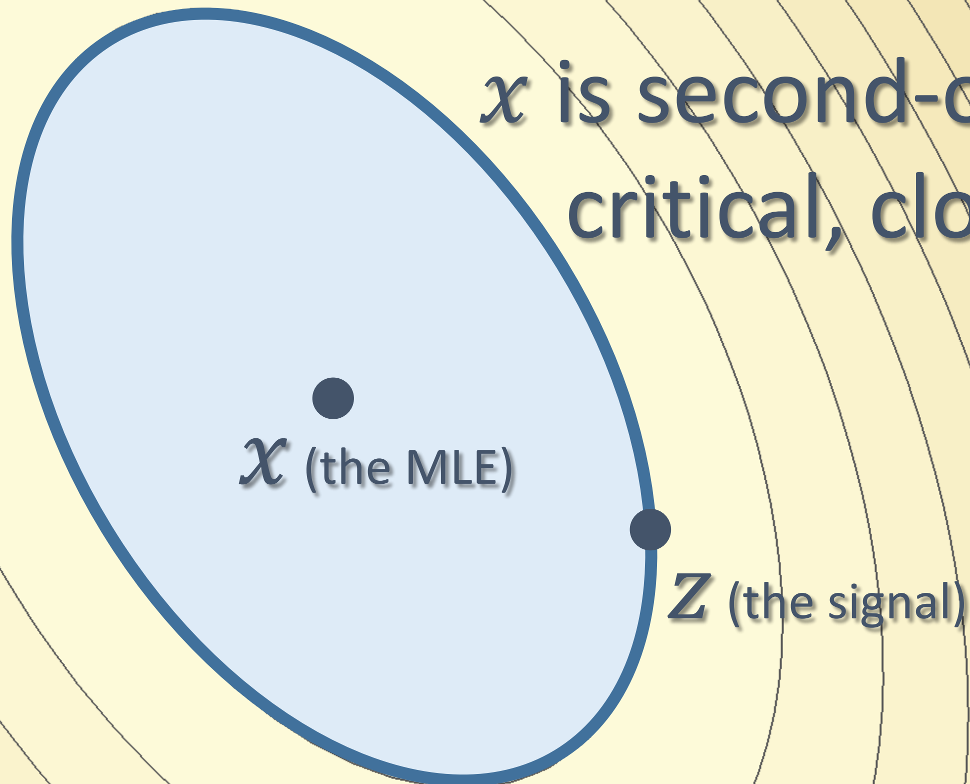
We aim to prove that $S(xx^*) \succcurlyeq 0$.

Challenge: we don't know x.

$S(X) = \Re\{\text{ddiag}(CX)\} - C$

Step 1: characterize the MLE x

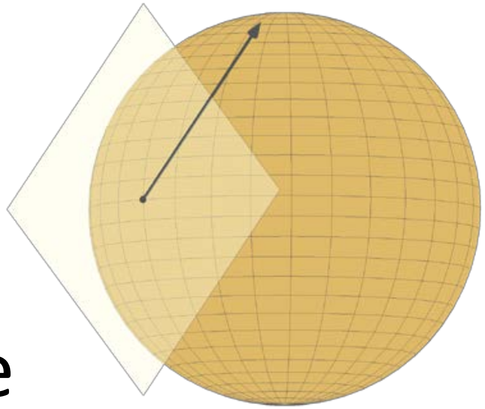
x is second-order critical, close to z.



level sets of x^*Cx

Step 1: characterize the MLE x

Standard necessary optimality conditions:



1. Gradient equals zero

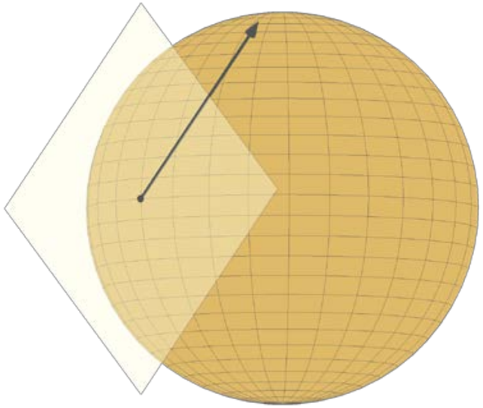
2. Hessian is positive semidefinite

Here: it's the same, with Riemannian derivatives

Step 1: characterize the MLE x

Gradient is zero iff $Sx = 0$.

Hessian is positive semidefinite iff
S is positive semidefinite *on tangent space*.



$$S(X) = \Re\{\mathrm{ddiag}(CX)\} - C$$

Step 2: certify

Assuming x is second-order critical and close to z, show that

$$S = \mathrm{ddiag}(Cxx^*) - C \succcurlyeq 0.$$

For all $u \in \mathbb{C}^n$ with $u^*x = 0$, a few lines give:

$$u^*Su \geq \|u\|_2^2\left(n - \sigma\left[216\sigma + 3\sqrt{n} + \|Wx\|_\infty\right]\right)$$

(Used $\|W\|_{\mathrm{op}} \leq 3\sqrt{n}$ again.)

Step 2: certify

Sufficient condition for tightness of the SDP:

$$n \geq \sigma\left[216\sigma + 3\sqrt{n} + \|Wx\|_\infty\right]$$

Target: show $\|Wx\|_\infty \leq \sim\sqrt{n \log n}$.

Similar to eigenvector situation, need to control $\|Wx\|_\infty$ where this time x is the MLE.

We want to control $\|Wx\|_\infty$

For x the dominant eigenvector of C,
we can fully characterize x:

$Cx = \lambda x$ with $\lambda \approx n$, and we have Davis-Kahan.

For x the MLE, we have neither.

Strategy: track generalized power method.

Generalized power method

$$\begin{aligned} x^0 &= \text{phase}\left(x^{\text{eig}}\right) \\ x^{k+1} &= \text{phase}\left(Cx^k\right), \qquad k = 0,1, \ldots \end{aligned}$$

Set up auxiliary sequences same as before.

Iterates of aux. sequences remain close to each other, close to z in 2-norm, and such that $\|Wx\|_\infty$ is small enough.

Show convergence to MLE: deduce $\|Wx\|_\infty$ is small.

Shows SDP is tight, and generalized power method solves it efficiently.

Through SDP relaxation	Via generalized power method
With Riemannian optimization	Using dominant eigenvector

There is still a gap for the Riemannian optimization approach

Best theorem so far (despite better numerics):

For the MLE, if $\sigma \leq \sim n^{1/4}$, second-order necessary optimality conditions are sufficient for optimality, whp.

See Huikang Liu, Man-Chung Yue and Anthony Man-Cho So, SIOPT 2017

On the Estimation Performance and Convergence Rate of the Generalized Power Method for Phase Synchronization

Take home messages

Under white noise, phases are easy to synchronize.

Auxiliary problems (or replicas) can be an efficient proof technique to handle statistical dependence.

More work is needed for the Riemannian approach.

For details, from older to newer

1. With Bandeira & Singer, Math. Prog. 2017
   Tightness of the maximum likelihood semidefinite relaxation for angular synchronization

2. SIOPT 2016
   Nonconvex phase synchronization

3. **With Joe Zhong, SIOPT 2018**
   Near-optimal bounds for phase synchronization