Multireference alignment

An invariant features approach

Nicolas Boumal
Princeton University

With Tamir Bendory, Chao Ma, Zhizhen Zhao and Amit Singer

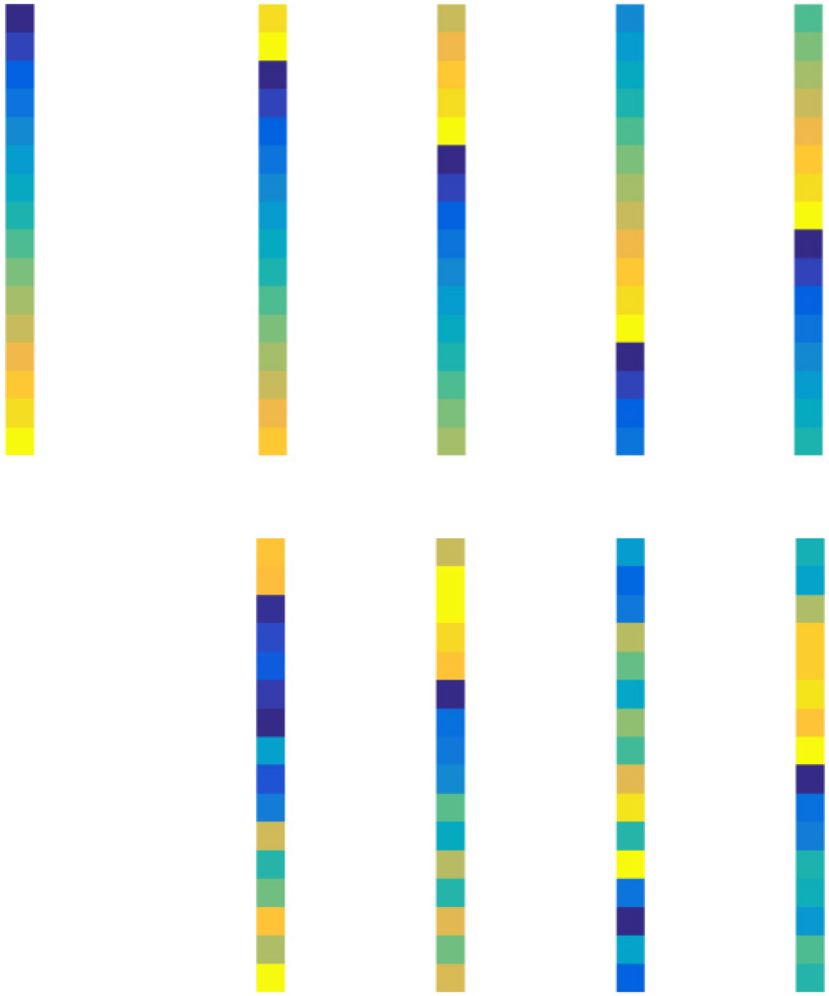
The multireference alignment model

True signal: $x \in \mathbf{C}^L$

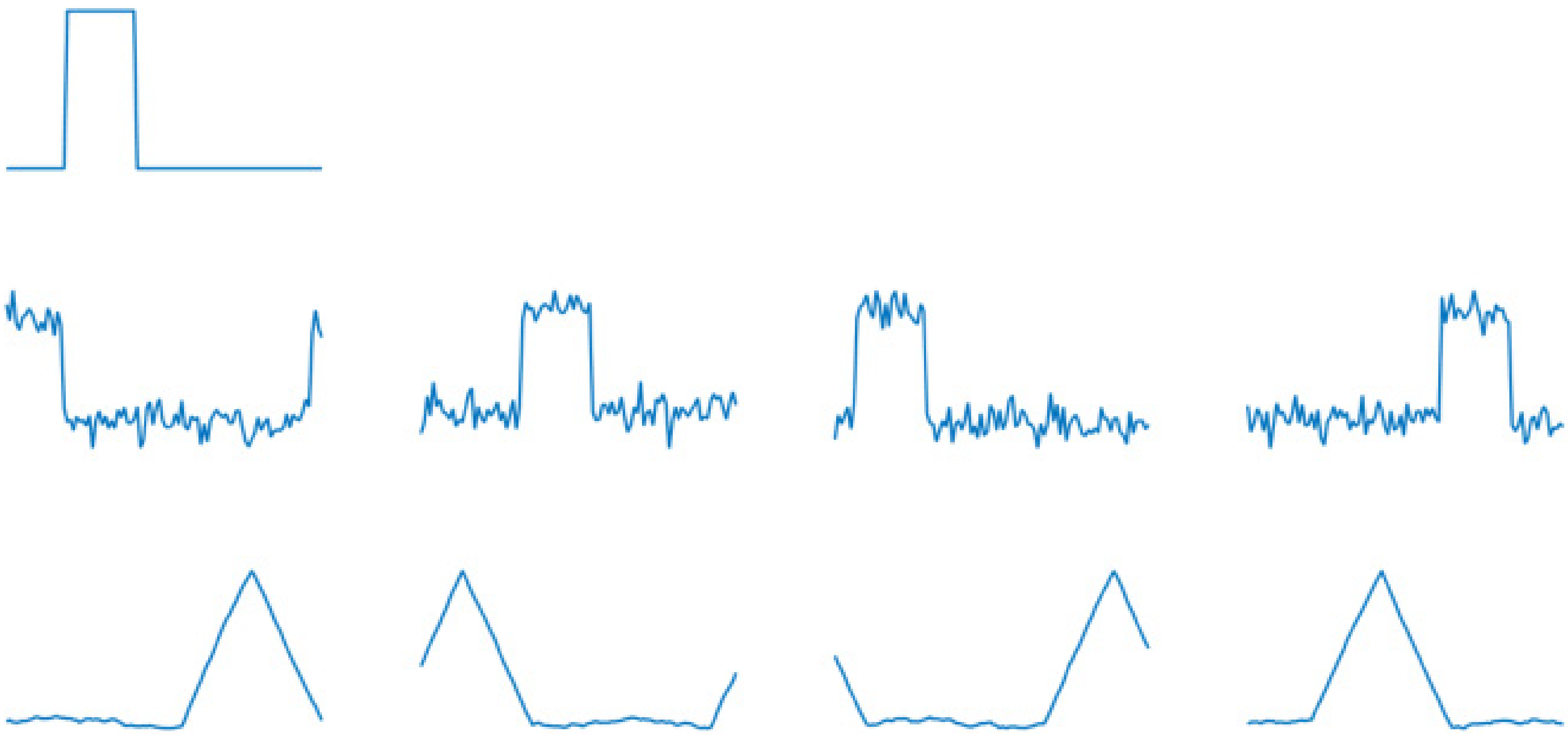
M observations:

$$\xi_i = R_{\ell_i} x + \sigma n$$

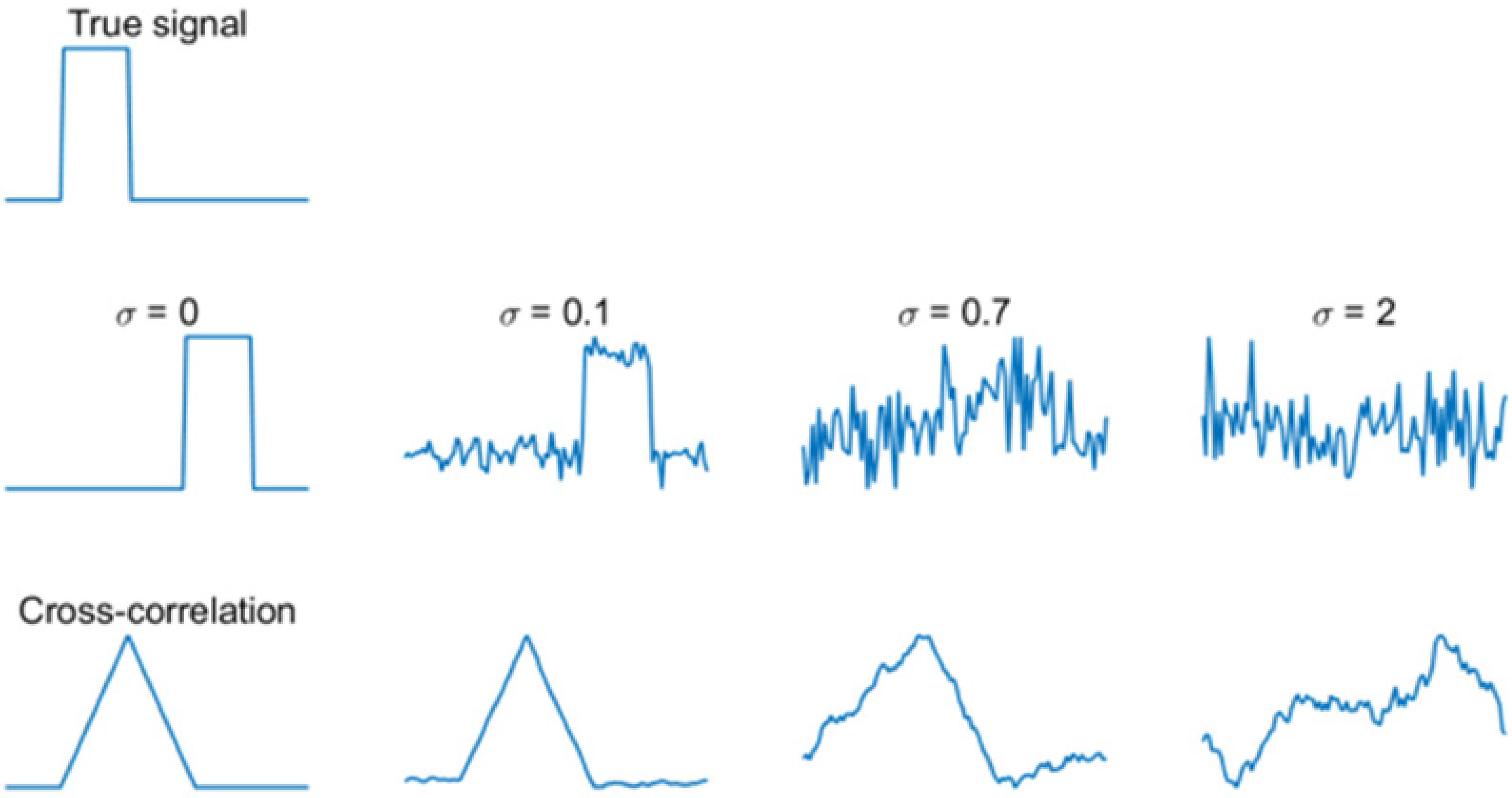
With $n \sim N(0, I_L)$, i.i.d.



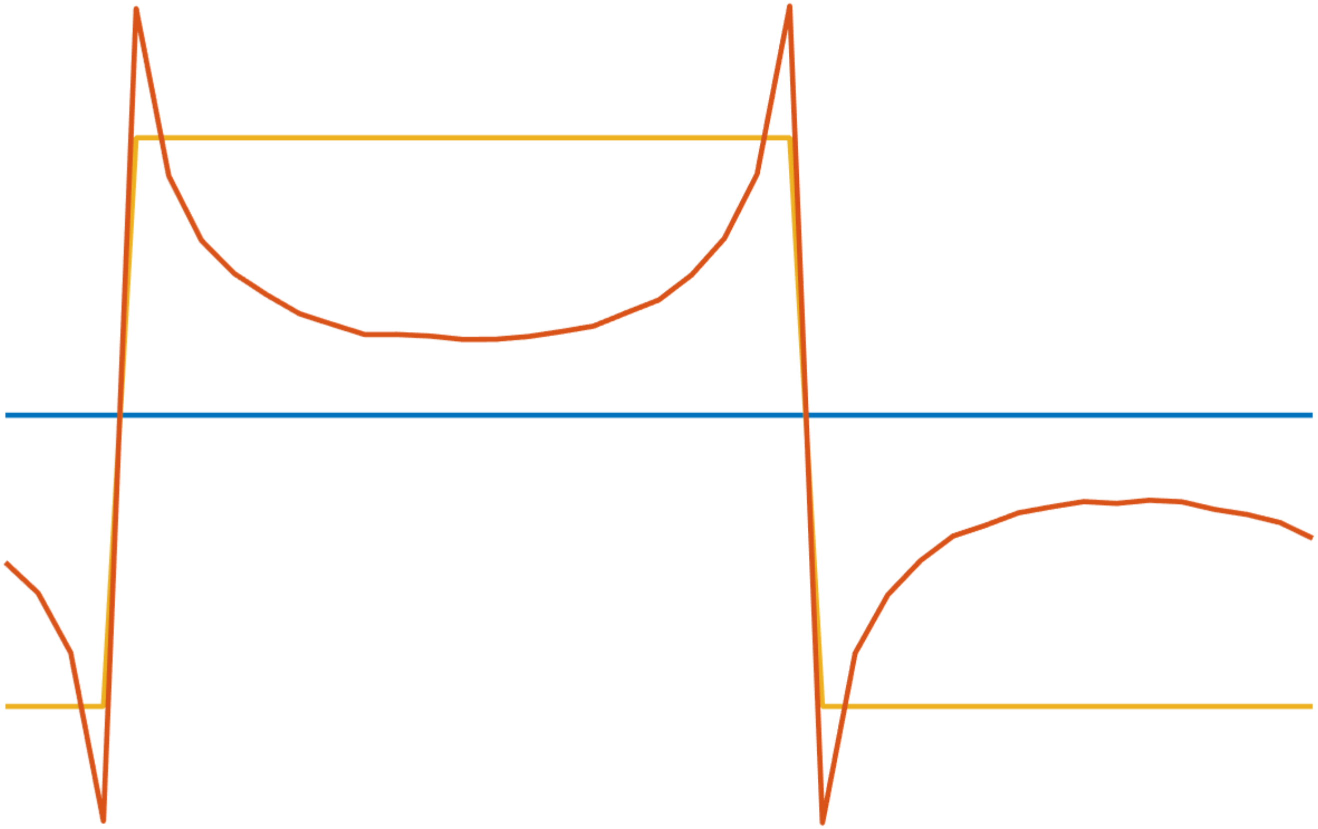
High SNR is easy



There is a fundamental limit to our ability to align one pair of observations

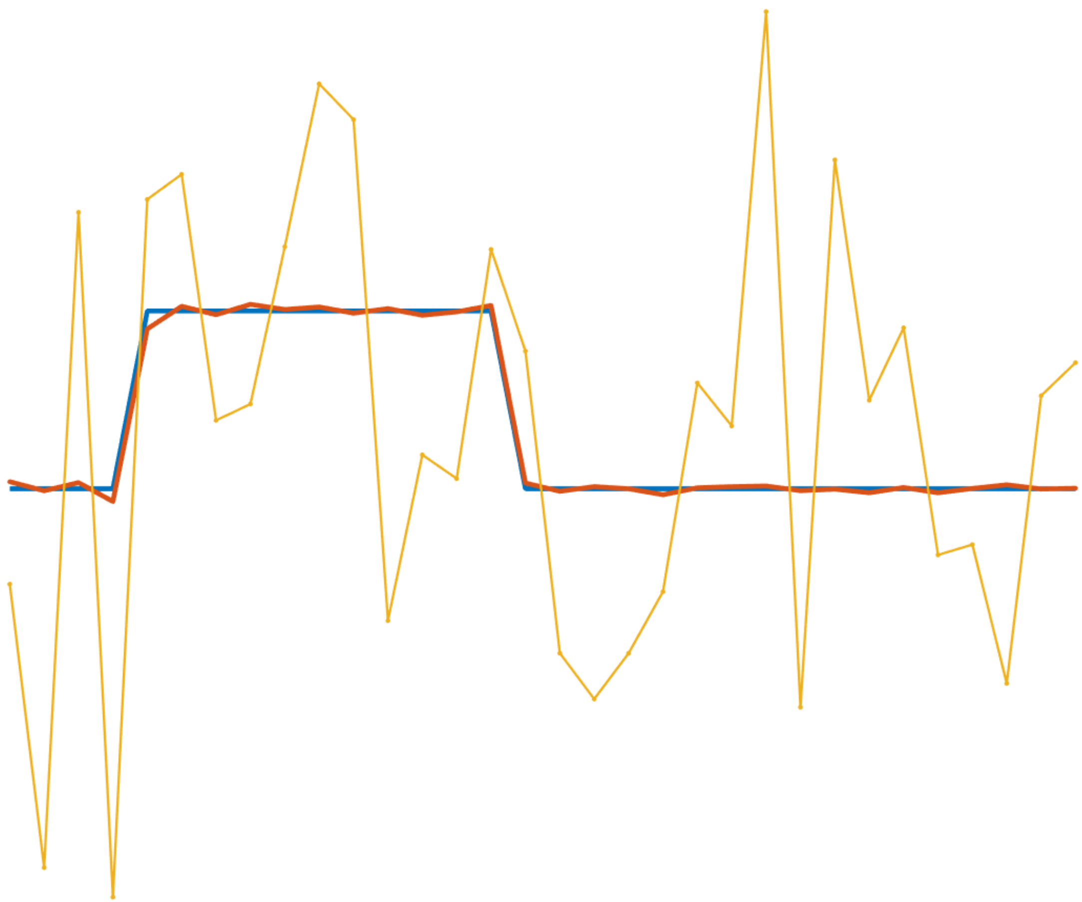


Einstein-from-noise



It is preferable to avoid the need for an initial guess.

Importantly, we don't need to estimate the shifts—only the signal



Length $L = 32$, #observations $M = 100{,}000$, noise level $\sigma = 1$
(window height is 1)

$\xi_i = R_{\ell_i} x + \sigma n$

The invariant features approach

Estimate the signal from statistics which are invariant under cyclic shifts.

For example, to estimate the mean of the signal, we do not need to estimate the shifts:

$$\hat{\mu}_x = \frac{1}{M} \sum_{i=1}^{M} \mu_{\xi_i}$$

Estimates first Fourier coefficient, variance $O\left(\frac{\sigma^2}{M}\right)$.

$\xi_i = R_{\ell_i}x + \sigma n$

Invariant feature 2: power spectrum

Let $y = \mathrm{DFT}(x)$ (discrete Fourier transform).

The power spectrum of x is:

$$P_x = |y|^2$$

Invariant because: $\mathrm{DFT}(R_t x)_k = y_k e^{-\frac{2\pi i t}{L}k}$.

(P_x is the DFT of the auto-correlation of x.)

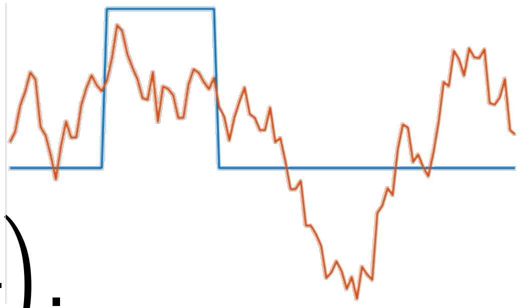
$\xi_i = R_{\ell_i} x + \sigma n$

Invariant feature 2: power spectrum

$y = \mathrm{DFT}(x), P_x = |y|^2$

$$\mathbf{E}\{P_{R_t x + \sigma n}\} = P_x + L\sigma^2$$

$$\hat{P}_x = \frac{1}{M} \sum_{i=1}^{M} P_{\xi_i} - L\sigma^2$$



Unbiased estimator with variance $O\left(\frac{\sigma^4}{M}\right)$.

Sufficient to estimate the Fourier moduli.

$\xi_i = R_{\ell_i} x + \sigma n$

Invariant feature 3: bispectrum

Let $y = \text{DFT}(x)$. The bispectrum of x is:

$$(B_x)_{k\ell} = y_k \overline{y_\ell} y_{\ell-k}$$

Invariant because: $\text{DFT}(R_t x)_k = y_k e^{-\frac{2\pi i t}{L} k}$.

(B_x is the 2D-DFT of the triple-correlation of x.)

$\xi_i = R_{\ell_i} x + \sigma n$

Invariant feature 3: bispectrum

$y = \mathrm{DFT}(x)$, bispectrum $(B_x)_{k\ell} = y_k \overline{y_\ell} y_{\ell-k}$

If x has zero mean, then

$$\mathbf{E}\left\{B_{R_t x + \sigma n}\right\} = B_x$$

$$\hat{B}_{x-\mu_x} = \frac{1}{M} \sum_{i=1}^{M} B_{\xi_i - \hat{\mu}_x}$$

Unbiased estimator with variance $O\left(\frac{\sigma^6}{M}\right)$.

Can we recover Fourier phases from the bispectrum?

$$\xi_i = R_{\ell_i} x + \sigma n$$

Invertibility (up to shifts)

The bispectrum alone is sufficient to recover the signal, under support conditions on the DFT.

If $x \in \mathbf{C}^L$, this is sufficient:

$y_k \neq 0$ for $k \in \{1, \ldots, K\}$ with $K \geq \frac{L+1}{2}$

$y_k = 0$ for $k \in \{K+1, \ldots, L-1\}$

y_0 free

In principle, there is no limit on σ

The map $\boldsymbol{\phi}: x \mapsto (\mu_x, P_x, B_x)$ is smooth.

If $y = \mathrm{DFT}(x)$ is non-vanishing, $\mathrm{J}_{\boldsymbol{\phi}}$ is invertible at x.

Thus, $\mathrm{J}_{\boldsymbol{\phi}^{-1}}$ exists, that is, sensitivity is finite.

Variance on (μ_x, P_x, B_x) is $O\left(\frac{\sigma^6}{M}\right)$, so,
if M grows as σ^6, arbitrary precision in principle.

$\xi_i = R_{\ell_i} x + \sigma n$

MRA from invariant features

Input M observations ξ_i, noise level σ
Output estimate of the signal, $\hat{x}$

1. Estimate invariant features $O(ML^2)$
   Compute $\hat{\mu}_x, \hat{P}_x, \hat{B}_{x-\mu_x}$

2. Combine to estimate $y = \mathrm{DFT}(x)$ $O(\mathrm{fun}(L))$
   a. Estimate y_0 from $\hat{\mu}_x$
   b. Estimate $|y_k|$ from $\hat{P}_x$ for $k \neq 0$
   c. Estimate phases of y_k from $\hat{B}_{x-\mu_x}$ for $k \neq 0$

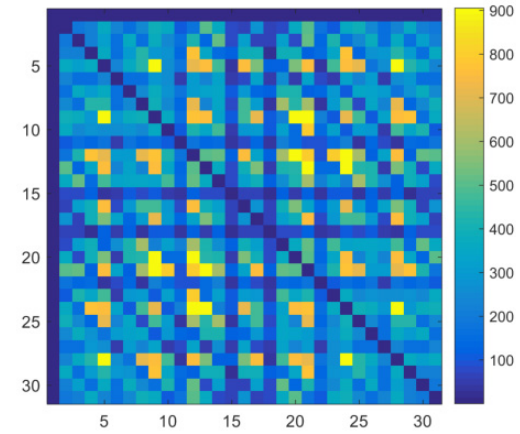
3. Return $\hat{x}$, the inverse DFT of the estimated y

Bispectrum inversion (phases only)

This is the part we focus on:

Given a bispectrum estimator

$$\hat{B}_{k\ell} \approx y_k \overline{y_\ell} y_{\ell-k},$$

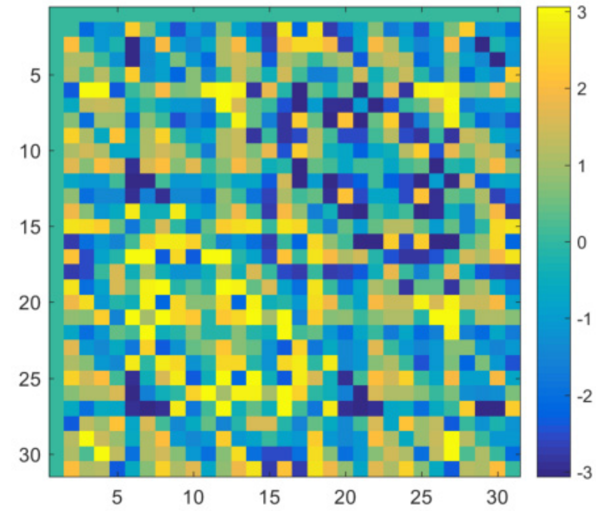


how can we recover the phases of y?

Frequency marching

Phase estimation is trivial in the absence of noise.
Let $\tilde{y} = \text{phase}(y)$ and $\tilde{B} = \text{phase}(B)$, then:

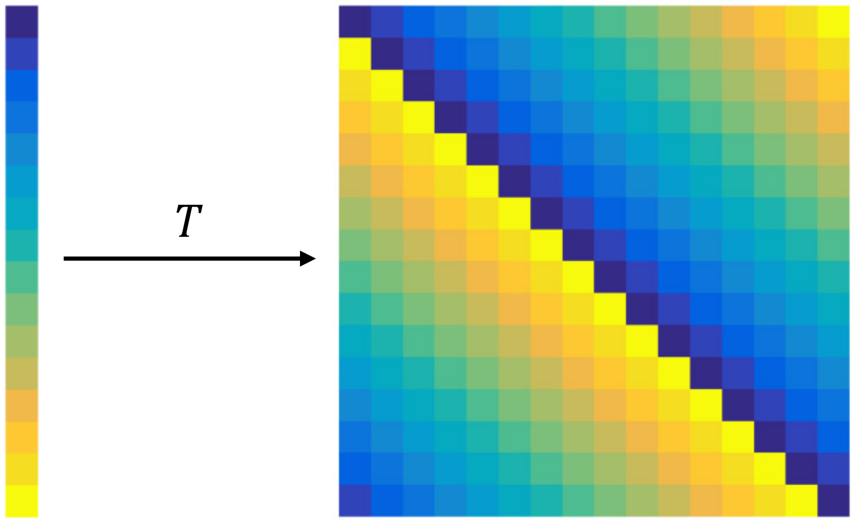
$$\tilde{B}_{k\ell} = \tilde{y}_k \overline{\tilde{y}_\ell} \tilde{y}_{\ell-k}$$



1. Pick $\tilde{y}_1$ in $e^0 \sim e^{\frac{2\pi i}{L}}$
2. To find $\tilde{y}_2$, use $\tilde{B}_{12} = \tilde{y}_1 \overline{\tilde{y}_2} \tilde{y}_1$
3. To find $\tilde{y}_3$, use $\tilde{B}_{13} = \tilde{y}_1 \overline{\tilde{y}_3} \tilde{y}_2$
4. To find $\tilde{y}_4$, use $\tilde{B}_{14} = \tilde{y}_1 \overline{\tilde{y}_4} \tilde{y}_3$ and $\tilde{B}_{24} = \tilde{y}_2 \overline{\tilde{y}_4} \tilde{y}_2$
5. Etc.

$$\tilde{B}_{k\ell} \approx \tilde{y}_k \overline{\tilde{y}_\ell} \tilde{y}_{\ell-k}$$

$T(\tilde{y})_{k\ell} = \tilde{y}_{\ell-k}$ (circulant)



$\tilde{B} \approx \tilde{y}\tilde{y}^* \circ T(\tilde{y})$

Phases-from-bispectrum as non-convex optimization

Since $\tilde{B} \approx \tilde{y}\tilde{y}^* \circ T(\tilde{y})$, try to solve:

$$\min_{z \in \mathbf{C}^L} \left\| W \circ \left(\tilde{B} - zz^* \circ T(z) \right) \right\|_\mathrm{F}^2$$

subject to: $|z_k| = 1$ for each k

$$\left\| W \circ \left(\tilde{B} - zz^* \circ T(z) \right) \right\|_\mathrm{F}^2 = \underbrace{\| \ldots \|_\mathrm{F}^2 + \| \ldots \|_\mathrm{F}^2}_{\text{constant}} - 2 \left\langle W^{(2)} \circ \tilde{B}, zz^* \circ T(z) \right\rangle$$

Phases-from-bispectrum as non-convex optimization

With $M(z) = W^{(2)} \circ \tilde{B} \circ \overline{T(z)}$:

$$\max_{z \in \mathbf{C}^L} \langle z, M(z) z \rangle$$

$$\text{subject to: } |z_k| = 1 \text{ for each } k$$

Closely related to phase synchronization.

Difference: cubic data instead of quadratic.

Commercial: phase synchronization

$$\max_{z \in \mathbf{C}^L} \langle z, Mz \rangle$$

subject to: $|z_k| = 1$ for each k

Near optimal bounds with Joe Zhong (ORFE)
arXiv:1703.06605

The proof involves ℓ_∞ perturbation bounds for eigenvectors.

Solve locally using the geometry of the set of phases (torus)

With $M(z) = B \circ \overline{T(z)}$:

$$\max_{z \in \mathbf{C}^L} \langle z, M(z) z \rangle$$
$$\text{subject to: } |z_k| = 1 \text{ for each } k$$

Apply Riemannian trust-regions

- Random initialization
- Converges to a second-order KKT point

Non-convexity and convergence

$$\max_{z \in \mathbf{C}^L} \langle z, M(z)z \rangle$$
subject to: $|z_k| = 1$ for each k

- In the noiseless case, *empirically*, there are L second-order KKT points: global optima corresponding to each shift

- Local convergence is quadratic
- Global convergence: $O\left(\frac{1}{\varepsilon^3}\right)$ worst-case

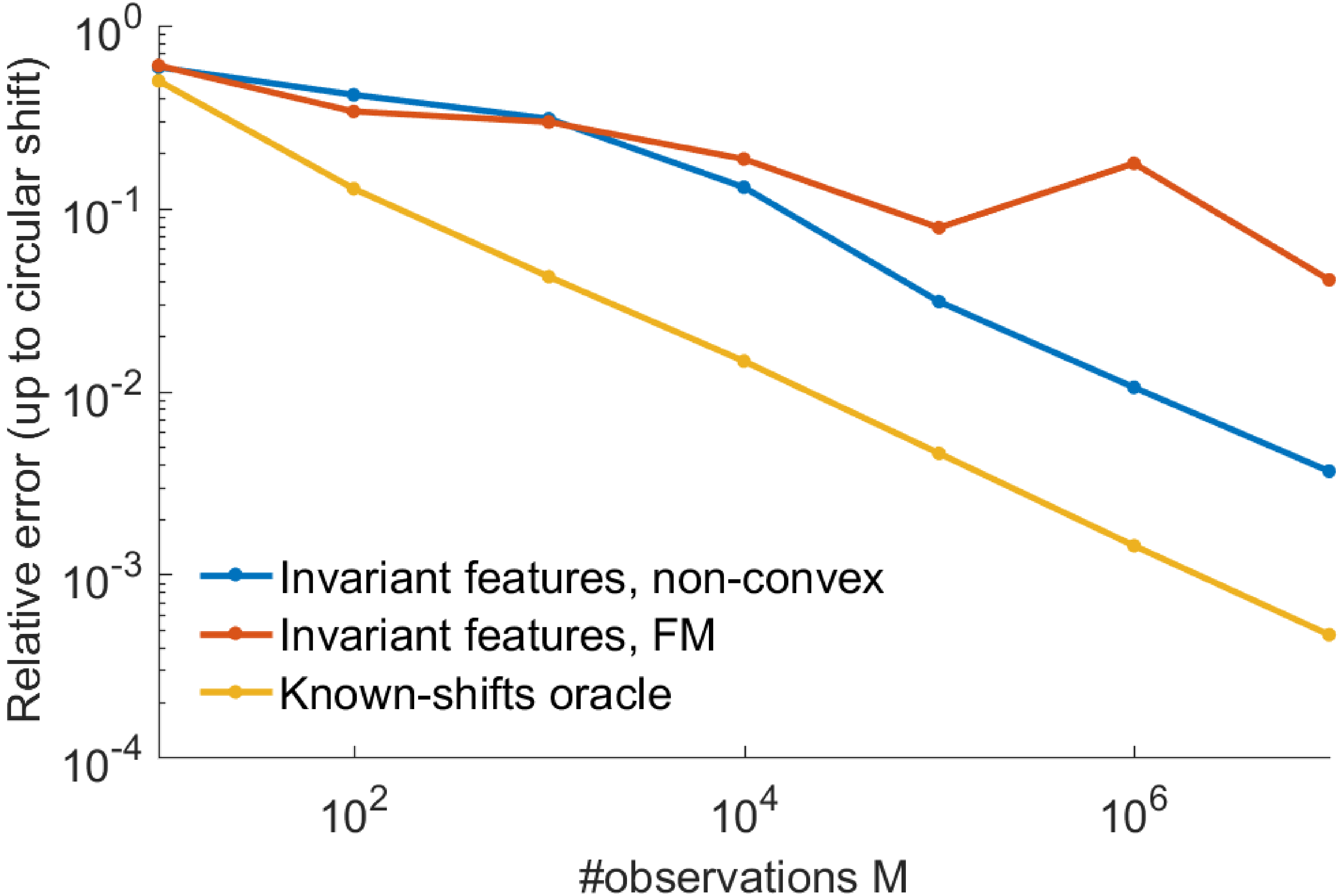
arXiv:1605.08101

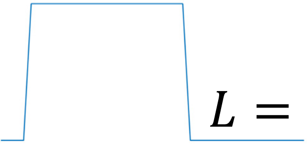
Numerical experiments

Relative error metric:

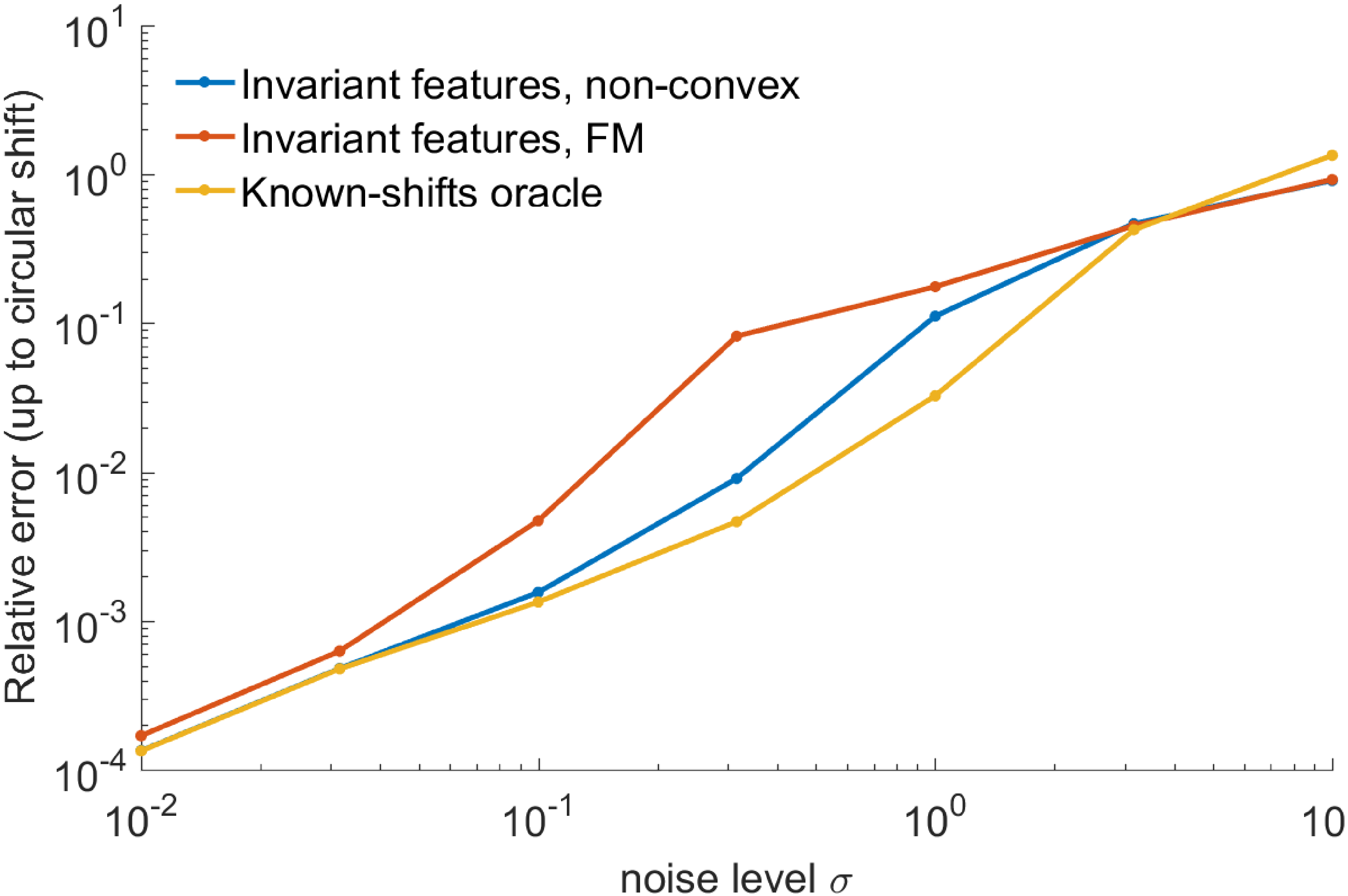
$$\mathrm{RMSE}(x_{\mathrm{true}}, x_{\mathrm{est}}) = \min_{\mathrm{shift}\ t} \frac{\|R_t x_{\mathrm{est}} - x_{\mathrm{true}}\|_2}{\|x_{\mathrm{true}}\|_2}$$

$L = 41, \sigma = 1$, averaged over 3 repeats, Gaussian noise

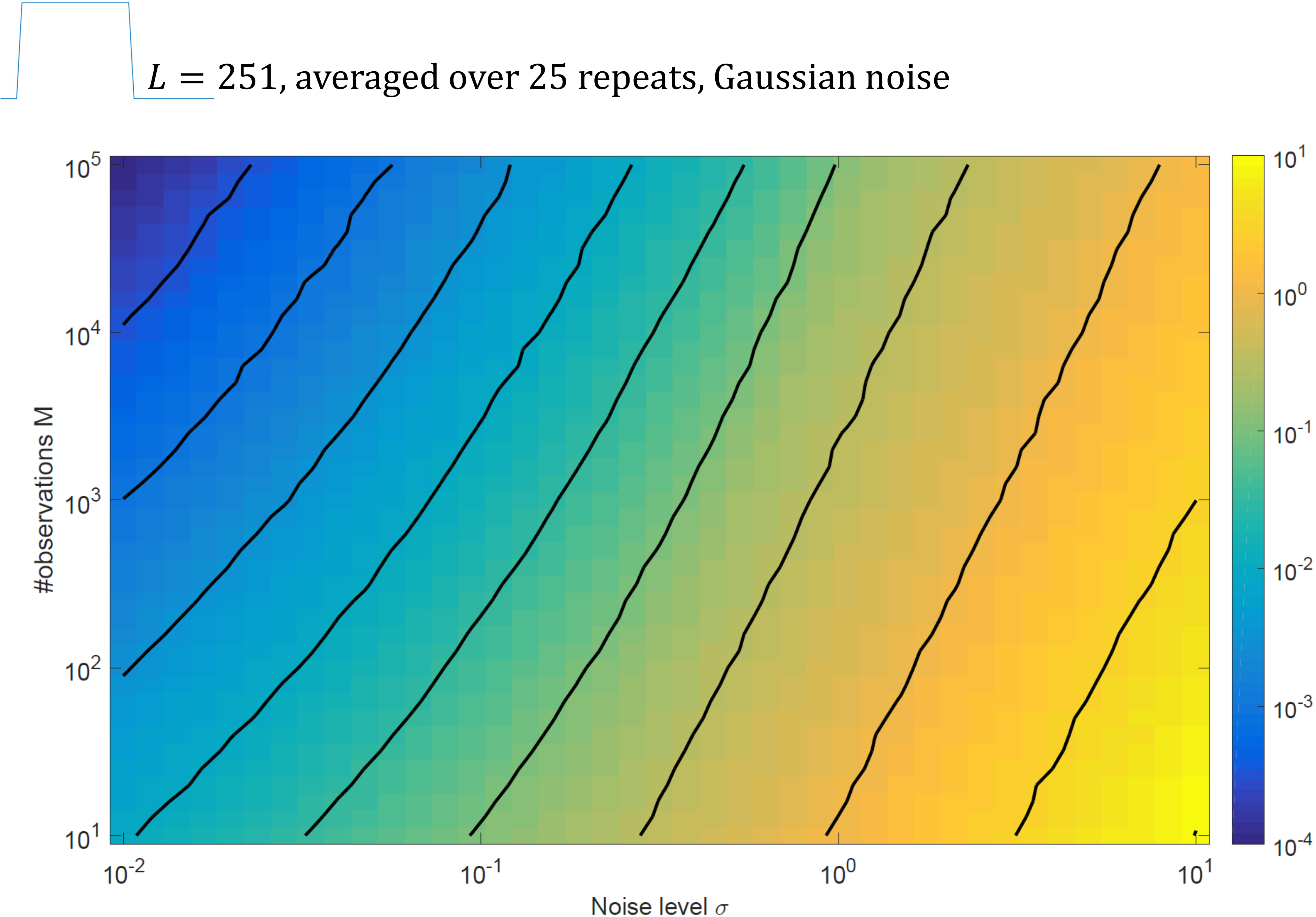




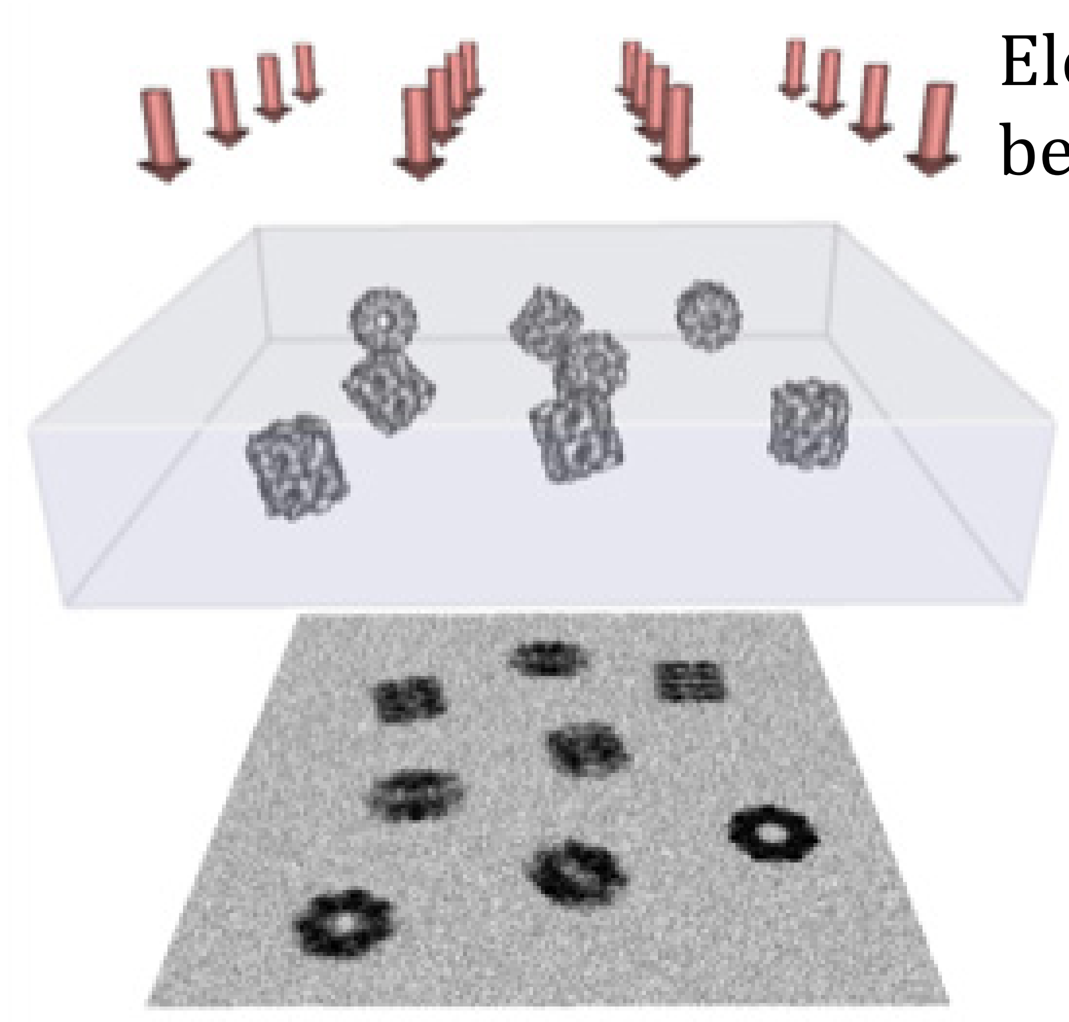
$L = 41, M = 10\,000$, averaged over 10 repeats, Gaussian noise



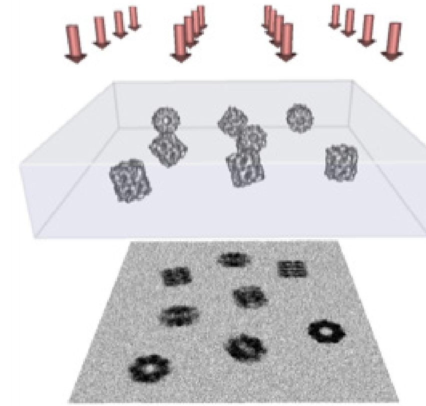
$L = 251$, averaged over 25 repeats, Gaussian noise



Looking ahead: Cryo-EM

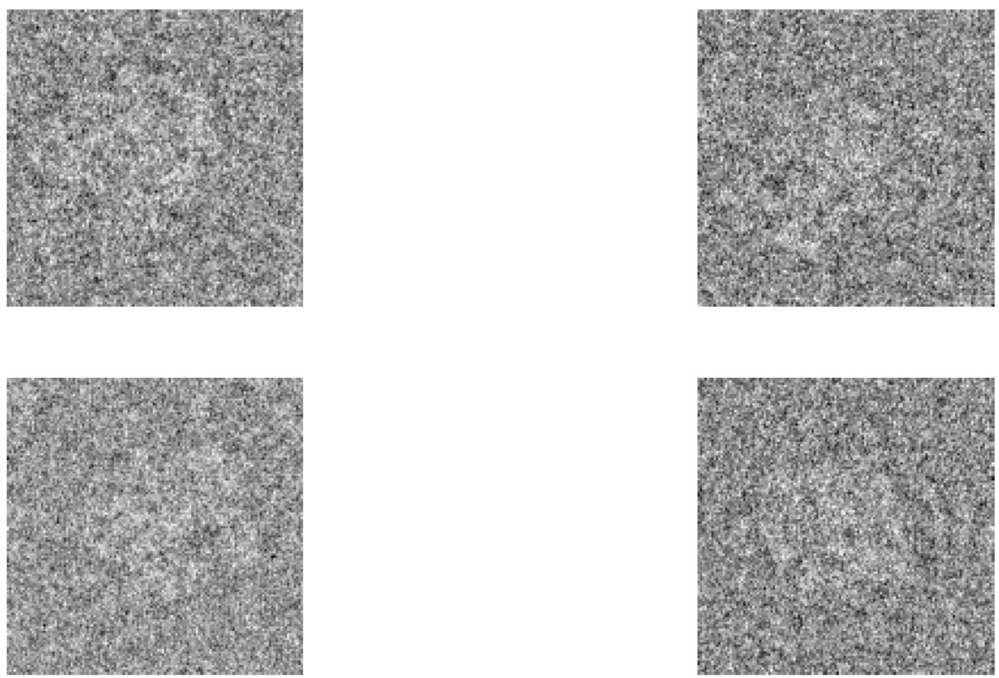


Looking ahead: Cryo-EM



E. coli 50S ribosomal subunit

Images provided by Dr. Fred Sigworth, Yale Medical School



Kam’s method

Expand the molecule’s 3D-DFT in spherical harmonics:

$$\hat{\phi}(k,\theta,\varphi) = \sum_{\ell=0}^{L} \sum_{m=-\ell}^{\ell} A_{\ell,m}(k) Y_\ell^m(\theta,\varphi)$$

Zvi Kam showed in ‘70s how to compute an analog of the power spectrum without estimating viewing directions:

$$C_\ell(k_1,k_2) = \sum_{m=-\ell}^{\ell} A_{\ell,m}(k_1) A_{\ell,m}^*(k_2)\text{, or: } C_\ell = A_\ell A_\ell^*$$