Optimal rates of estimation for the multi-reference alignment problem

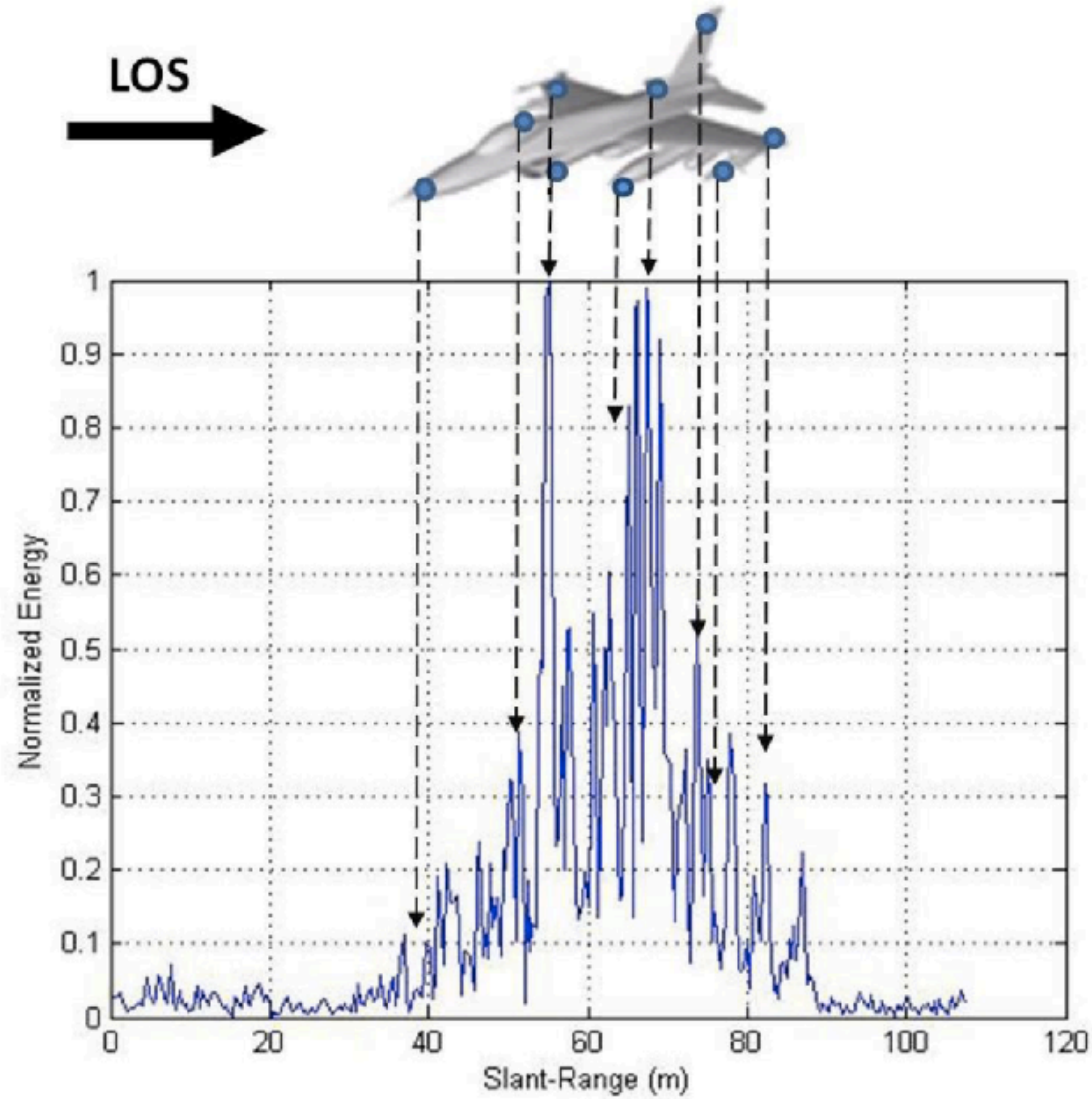
Jonathan Weed
Massachusetts Institute of Technology

Les Houches, 10.04.17

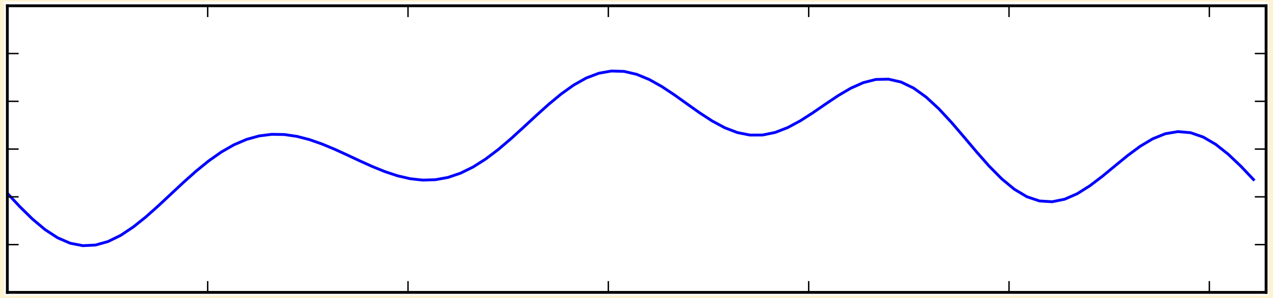
Joint work with

Afonso S. Bandeira
New York University

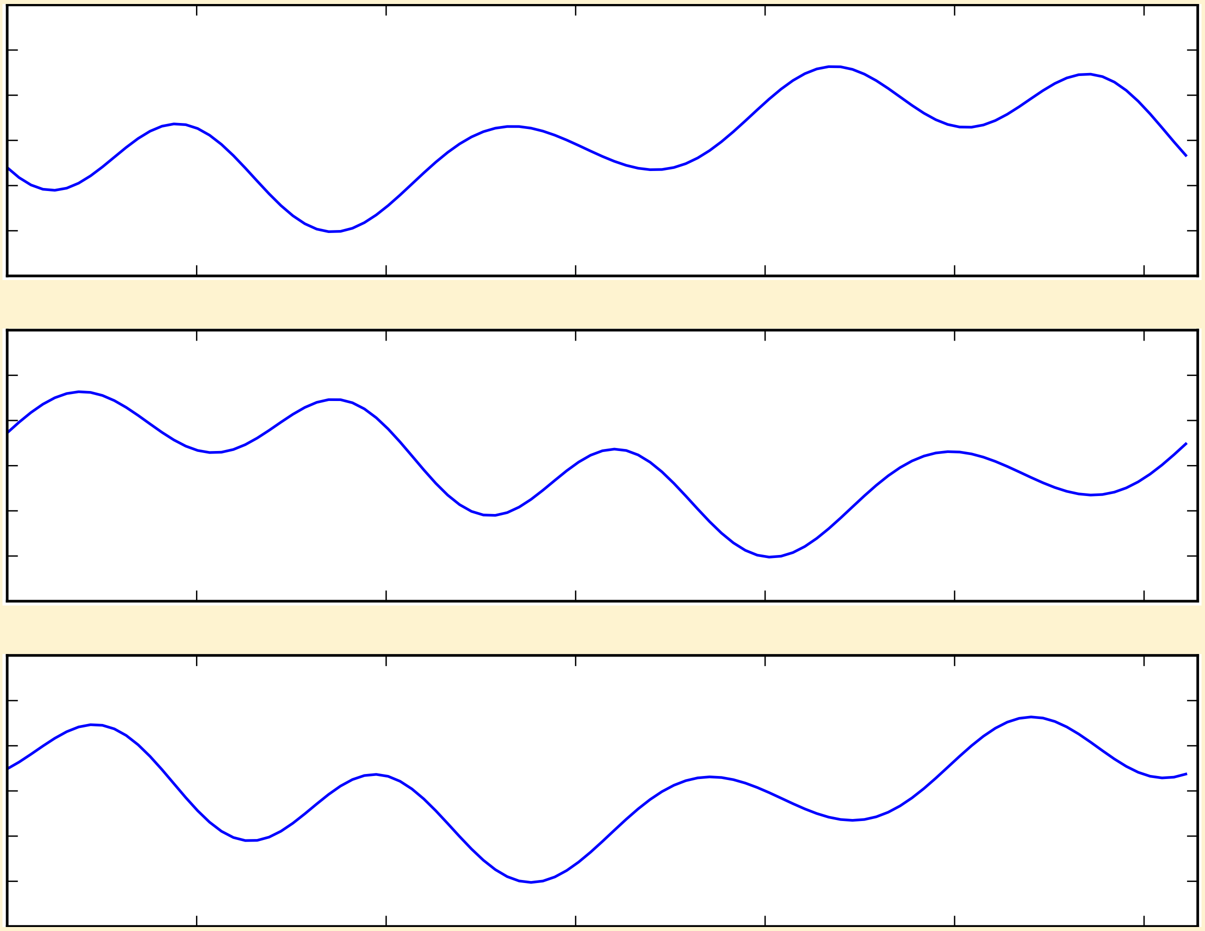
Philippe Rigollet
Massachusetts Institute of Technology



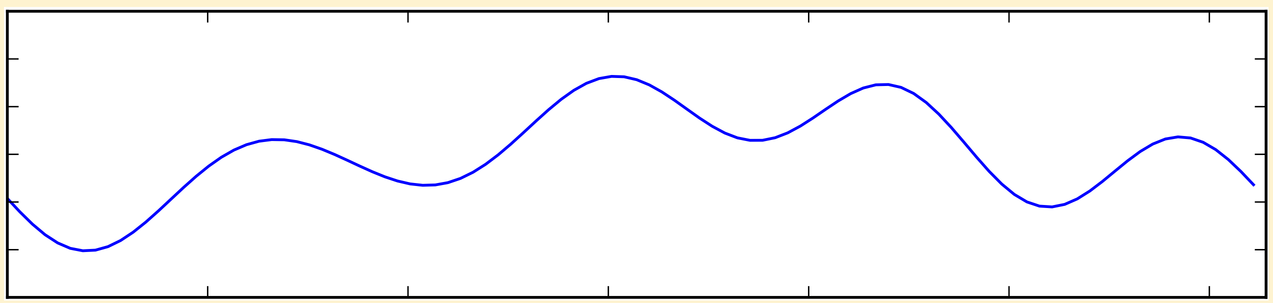
"Computational Burden Resulting from Image Recognition of High Resolution Radar Sensors," López-Rodríguez et al. 2013



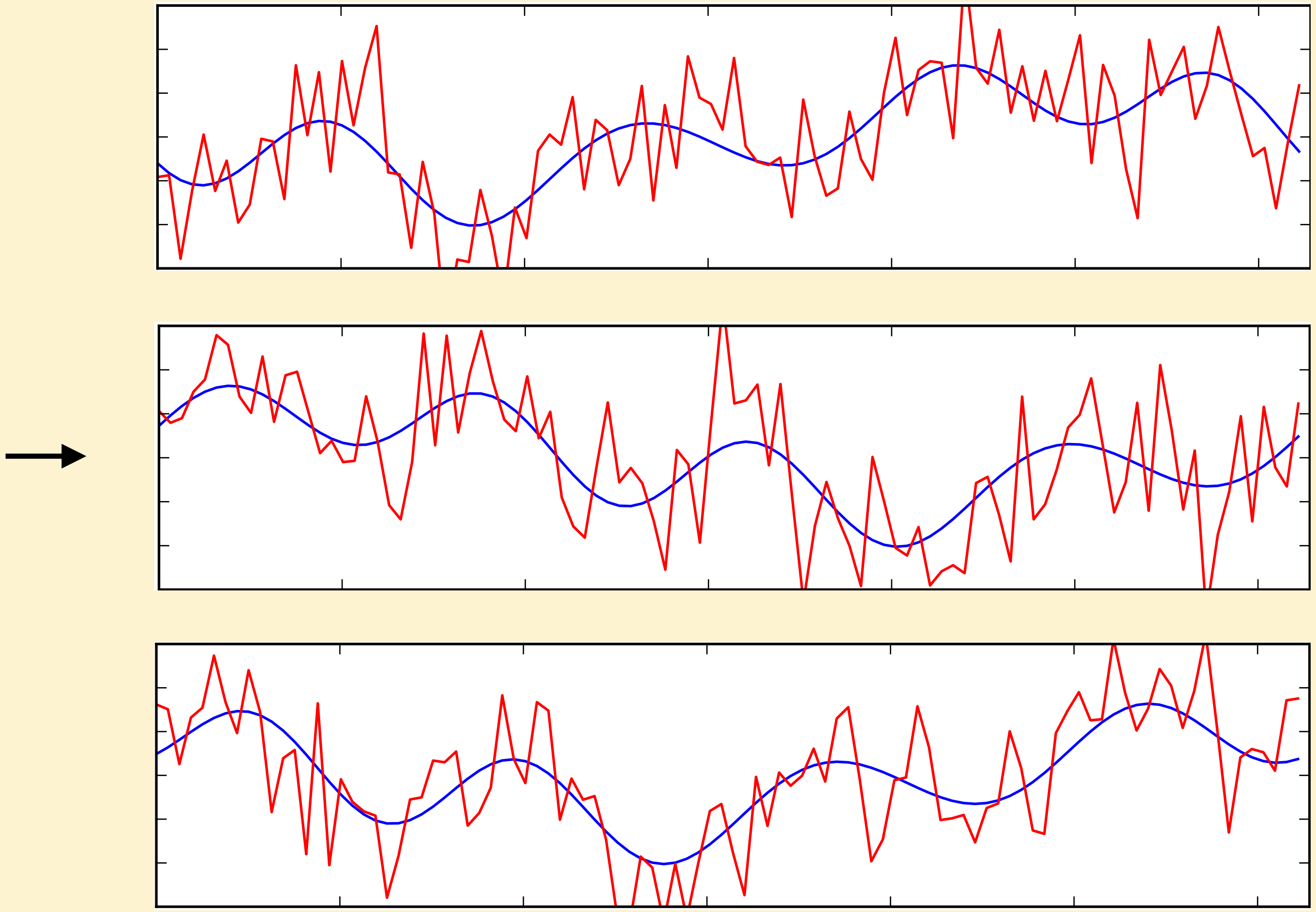
true signal



shifted signals



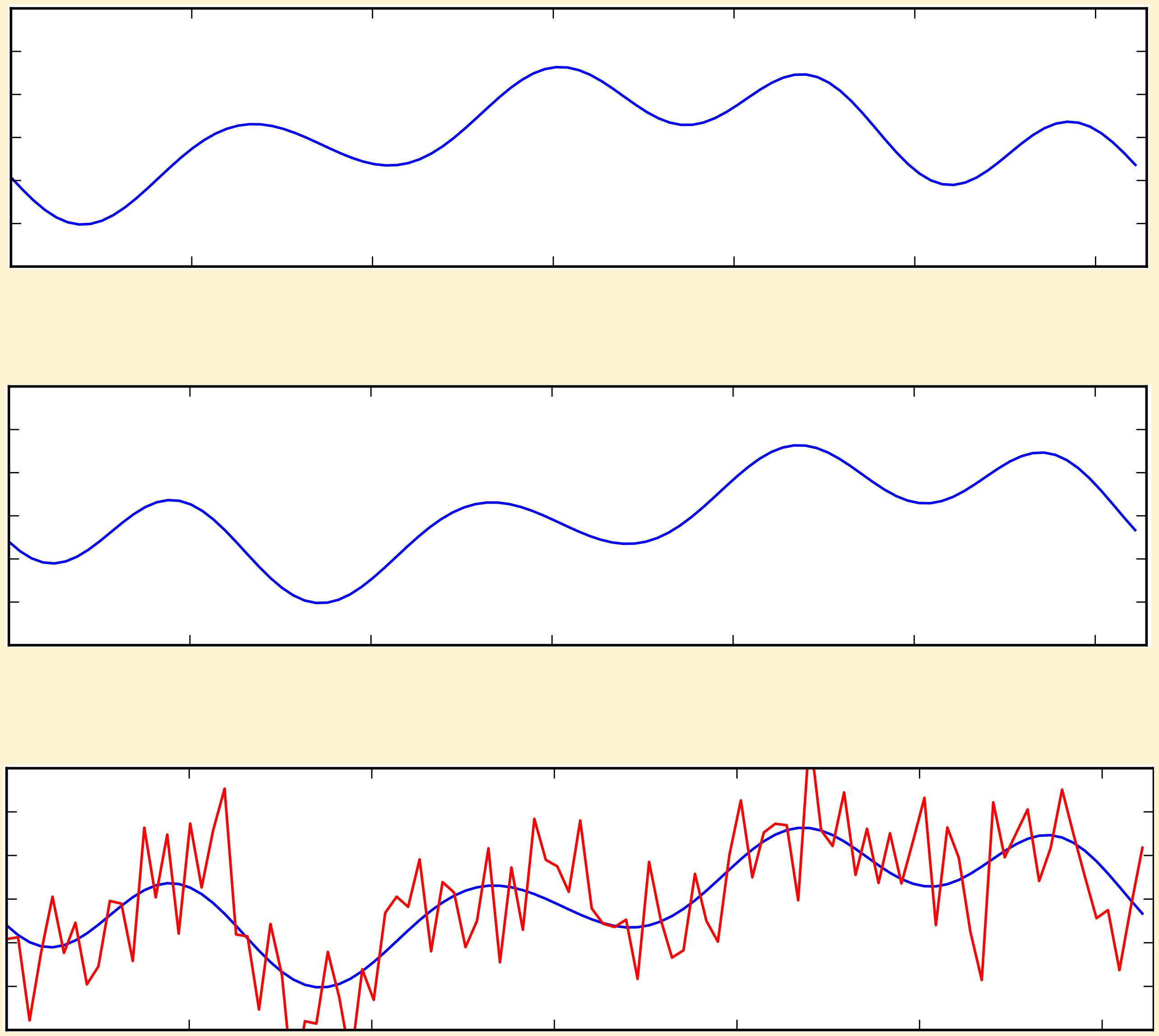
true signal



noisy data

Multi-reference alignment

[BCSZ'14]



true signal: $\theta \in \mathbb{R}^d$

rotated signal: $R_{\ell_i}\theta$

noisy data: $Y_i = R_{\ell_i}\theta + \sigma\xi_i$

"Multireference alignment using semidefinite programming," Bandeira, Charikar, Singer, Zhu. 2014.

Multi-reference alignment

[BCSZ'14]

$$Y_i = R_{\ell_i}\theta + \sigma\xi_i,\ \ i = 1, \ldots, n$$

$\theta \in \mathbb{R}^d,\ \ \|\theta\| = 1$ Parameter of interest

$\xi_i \sim \mathcal{N}_d(0, I_d)$ i.i.d.

$R_\ell\theta$ Cyclic shift of θ by ℓ coordinates:

$$\big[R_\ell\theta\big]_j = \theta_{j+\ell \pmod d}$$

e.g. $R_1 \begin{pmatrix}1\\2\\3\end{pmatrix} = \begin{pmatrix}2\\3\\1\end{pmatrix}$

Multi-reference alignment

[BCSZ'14]

$$Y_i = R_{\ell_i}\theta + \sigma\xi_i,\ i = 1, \ldots, n$$

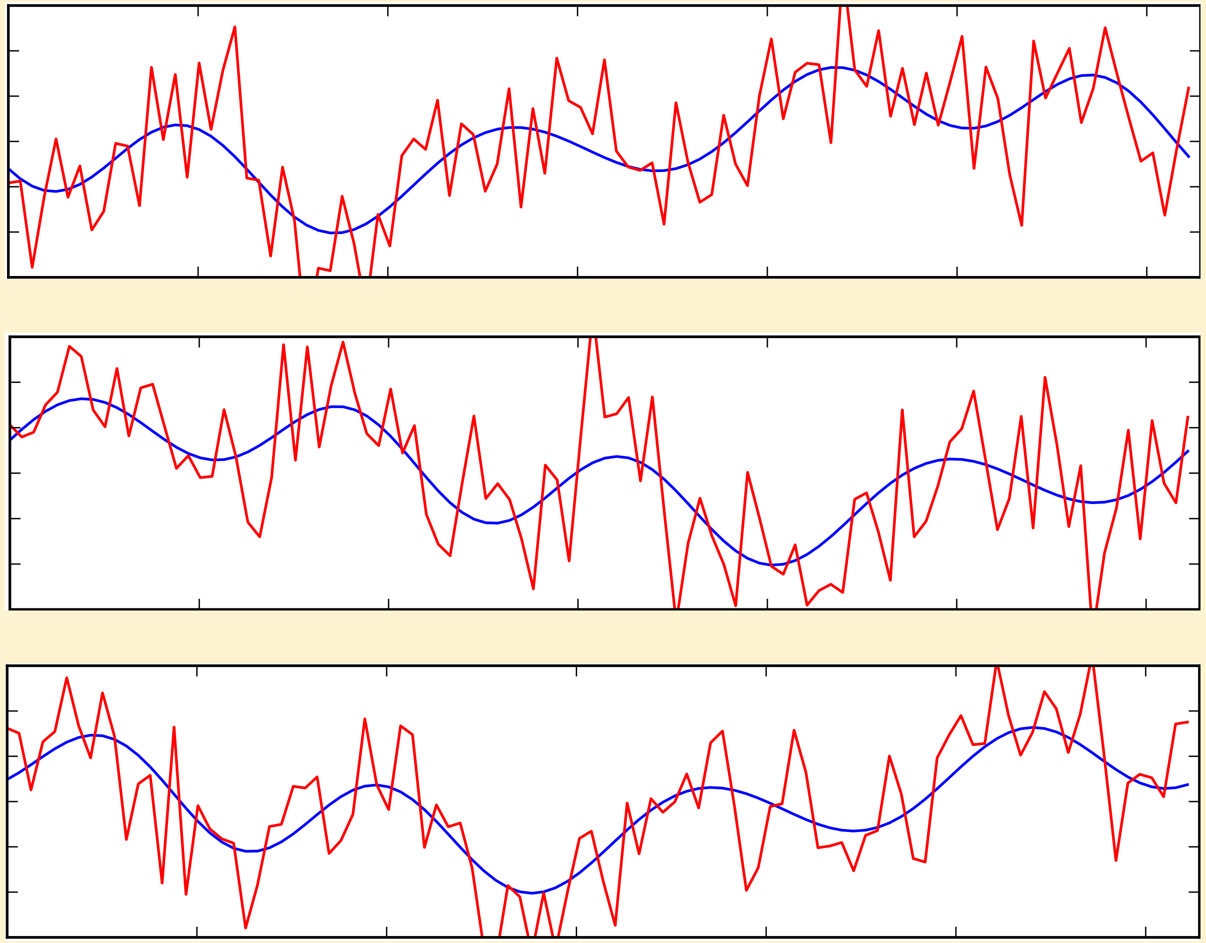
- How many **samples** are needed to estimate signal?
- Are there **efficient algorithms** to recover the signal?
- How does rate of estimation depend on **group structure**?

$$Y_i = R_{\ell_i} + \sigma \xi_i$$

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

R_{ℓ_i} : shift

Synchronization

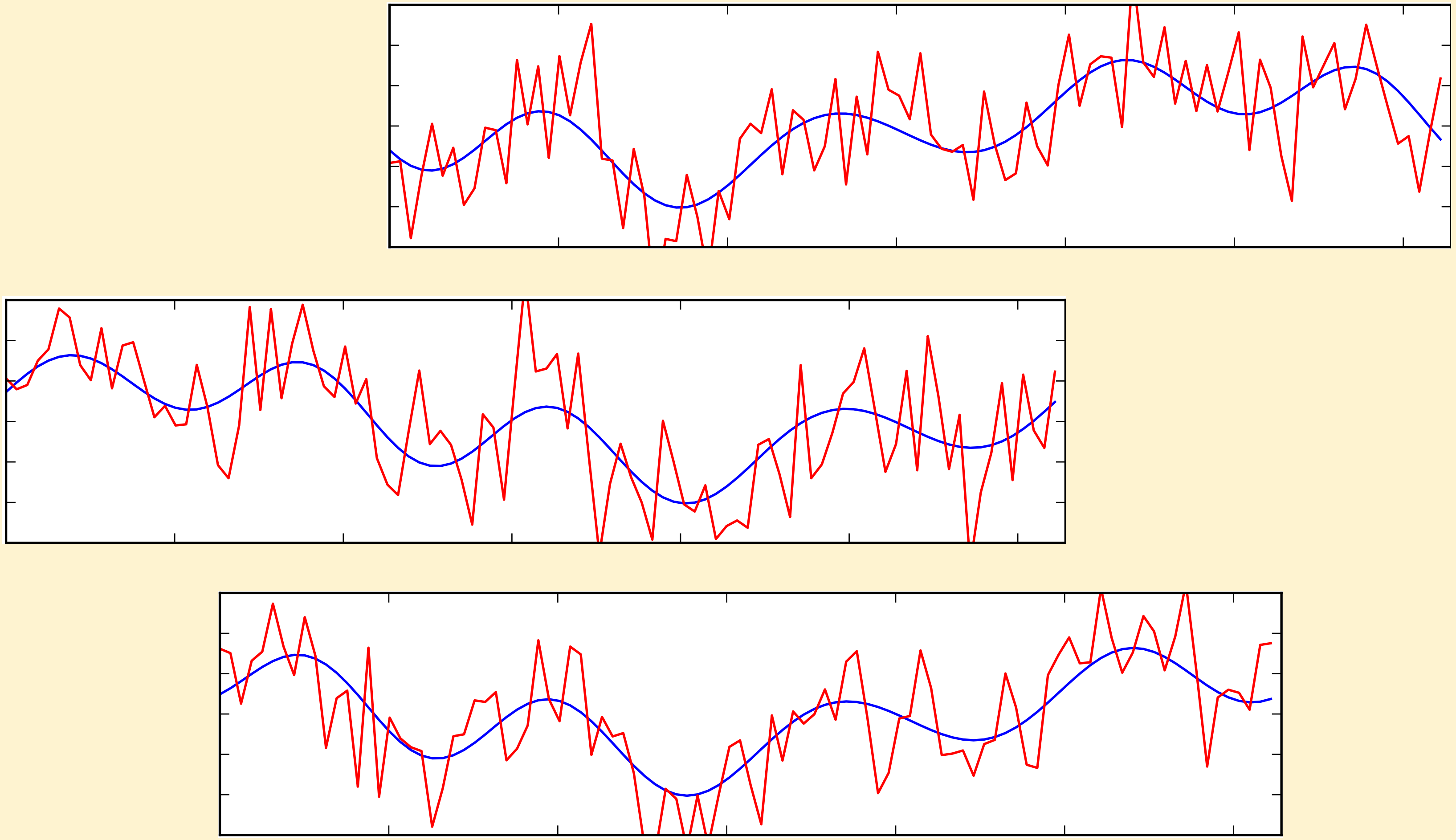


$$Y_i = R_{\ell_i} + \sigma \xi_i$$

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

R_{ℓ_i} : shift

Synchronization

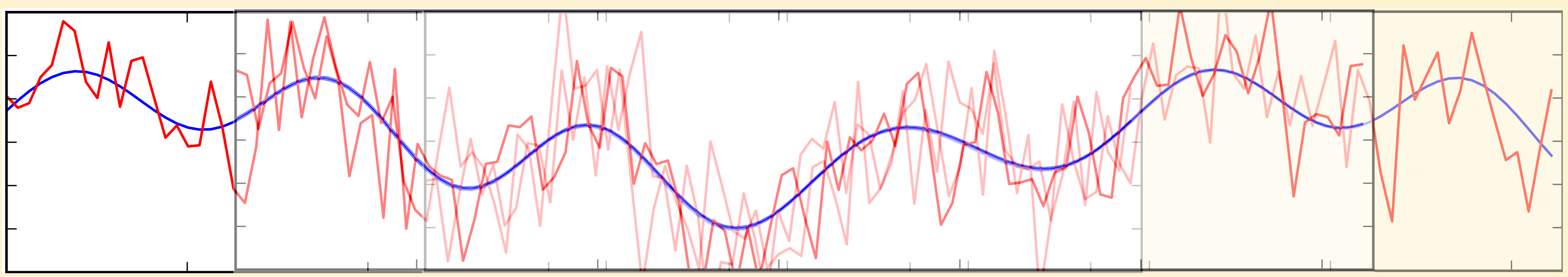


$$Y_i = R_{\ell_i} + \sigma \xi_i$$

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

R_{ℓ_i} : shift

Synchronization



[BCSZ'14]: "we focus on the problem of estimating the shifts"

$$Y_i = R_{\ell_i} + \sigma \xi_i$$

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

R_{ℓ_i} : shift

Synchronization

[BCSZ'14]: "we focus on the problem of estimating the shifts"

If signals can be synchronized perfectly, then

$$\mathbb{E}[\|\tilde{\theta} - \theta\|] \lesssim \sigma \sqrt{\frac{d}{n}}$$

↓

$$\mathbb{E}\left[\min_{R \in \mathbb{Z}_d} \|\tilde{\theta} - R\theta\|\right] =: \mathbb{E}[\rho(\tilde{\theta}, \theta)]$$

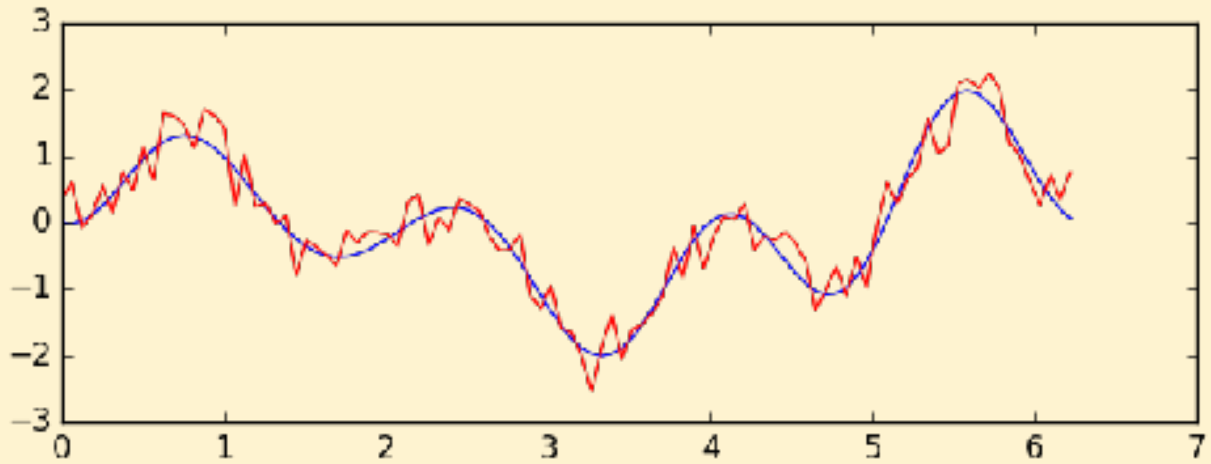
$$Y_i = R_{\ell_i} + \sigma \xi_i$$

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

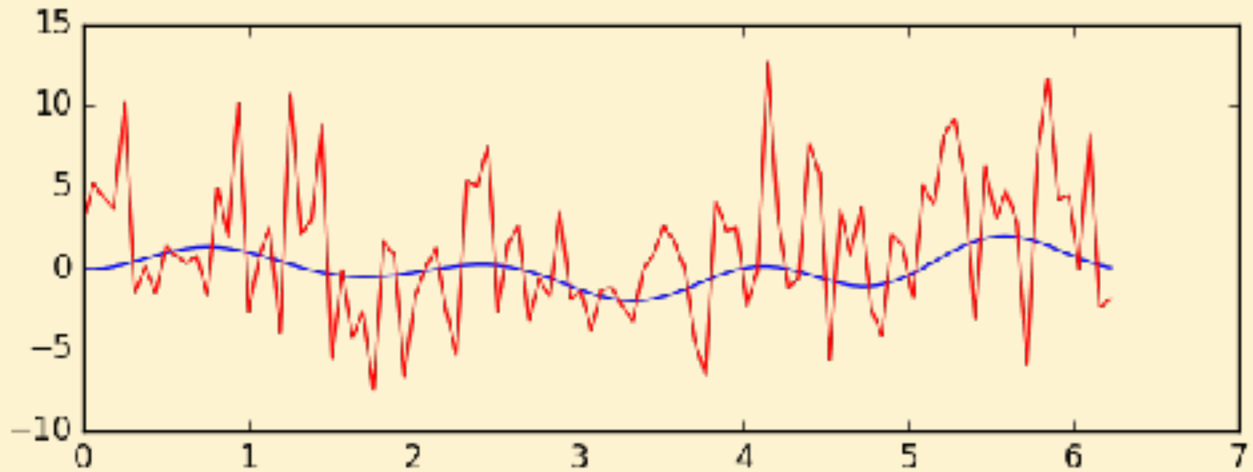
R_{ℓ_i} : shift

Synchronization

High SNR



Low SNR



Threshold effect [ADBS'16]: critical SNR below which no synchronization possible, even in infinite-sample limit

e.g. $d = 2, \quad \theta_1 = 0, \ \theta_2 = 1$

"Fundamental limits in multi-image alignment," Augerrebere, Delbracio, Bartesaghi, Sapiro. 2016.

Mixture of Gaussians

R_{ℓ_i} latent variables $\longrightarrow$ R uniform from $\mathbb{Z}_d$ (WLOG)

$$Y_i = R_{\ell_i}\theta + \sigma\xi_i \quad \longrightarrow \quad R_{U_i}Y_i = R_{U_i+\ell_i}\theta + \sigma R_{U_i}\xi_i$$

Equivalent to samples from **Gaussian mixture model**:

$$Y_i \sim \frac{1}{d}\sum_{R\in\mathbb{Z}_d}\mathcal{N}(R\theta, \sigma^2 I) =: \mathrm{P}_\theta$$

Rates of estimation

Curse of dimensionality

Gaussian mixtures with d centers [C'95, HK'15]

$$\mathbb{E}\big[\rho(\tilde{\theta},\theta)\big] \asymp C(\sigma,d)n^{-\frac{1}{2d}}$$

Parametric rate

We assume that Fourier transform of θ satisfies:

$$|\hat{\theta}_j| = 0 \quad \text{or} \quad |\hat{\theta}_j| > C \quad \Longrightarrow \mathbb{E}\big[\rho(\tilde{\theta},\theta)\big] \asymp C(\sigma,d)n^{-\frac{1}{2}}$$

Dependence in σ ?

"Optimal Rate of Convergence for Finite Mixture Models", Chen. 1995
"Optimal rates for finite mixture estimation", Heinrich, Kahn. 2015

Mixture of Gaussians

What can be estimated from Y_i ? Moments

$\mathbb{E}[Y_i] = \mathbb{E}[R\theta + Z] = \mathbb{E}[R\theta] = \bar{\theta}\mathbb{1}$ (mean)

$\mathbb{E}[Y_i^{\otimes 2}] - \sigma^2 I = \mathbb{E}[(R\theta)^{\otimes 2}]$ (autocovariance)

$\mathbb{E}[(Y_i - \bar{\theta}\mathbb{1})^{\otimes 3}] = \mathbb{E}[(R\theta)^{\otimes 3}]$ (triple covariance)

$\vdots$ $\vdots$

Mixture of Gaussians

What can be estimated from Y_i ? Moments

$$\mathbb{E}[Y_i] = \mathbb{E}[R\theta + Z] = \mathbb{E}[R\theta] = \bar{\theta}\mathbb{1}$$ (mean)

$$\mathbb{E}[Y_i^{\otimes 2}] - \sigma^2 I = \mathbb{E}[(R\theta)^{\otimes 2}]$$ (autocovariance)

$$\mathbb{E}[(Y_i - \bar{\theta}\mathbb{1})^{\otimes 3}] = \mathbb{E}[(R\theta)^{\otimes 3}]$$ (triple covariance)

$$\vdots \qquad \vdots$$

(mean) → $\hat{\theta}_0$

(autocovariance) → $\{|\hat{\theta}_j|^2\}$

(triple covariance) → $\{\hat{\theta}_i \hat{\theta}_j \hat{\theta}_{-i-j}\}$

Reconstruction from bispectrum [G'89, SG'92]

$$\hat{\theta} = (\hat{\theta}_0, \ldots, \hat{\theta}_{d-1})$$

First moment:	$\hat{\theta}_0$	(DC)
Second moment:	$\{\lvert\hat{\theta}_j\rvert^2\}$	(magnitude)
Third moment:	$\{\hat{\theta}_i \hat{\theta}_j \hat{\theta}_{-i-j}\}$	(phase)

Reconstruction possible if $\hat{\theta}_j \neq 0 \quad \forall j$

"Signal reconstruction from multiple correlations: frequency and time domain approaches," Giannaki. 1989
"Shift- and rotation-invariant object reconstruction using the bispectrum," Sadler, Giannakis. 1992

Reconstruction from bispectrum [G'89, SG'92]

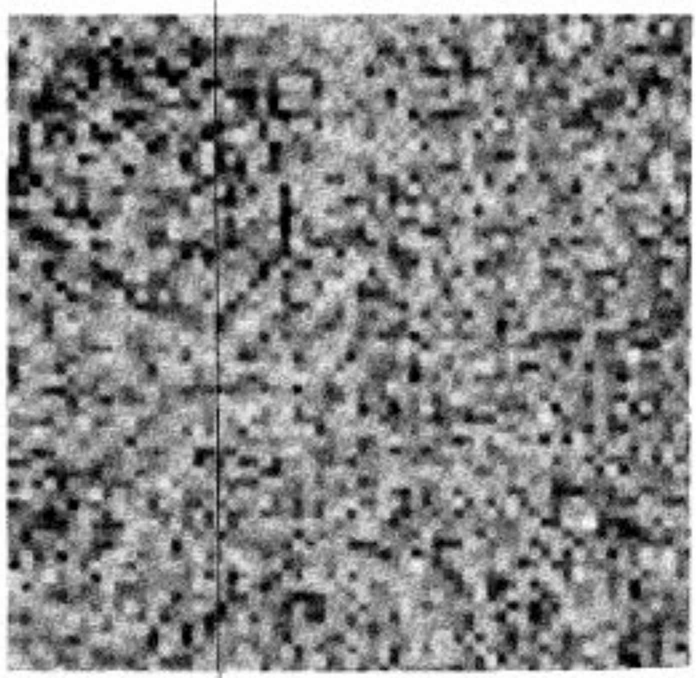


Figure 3
Object "Calvin" in white Gaussian noise.
SNR = $-10dB$.

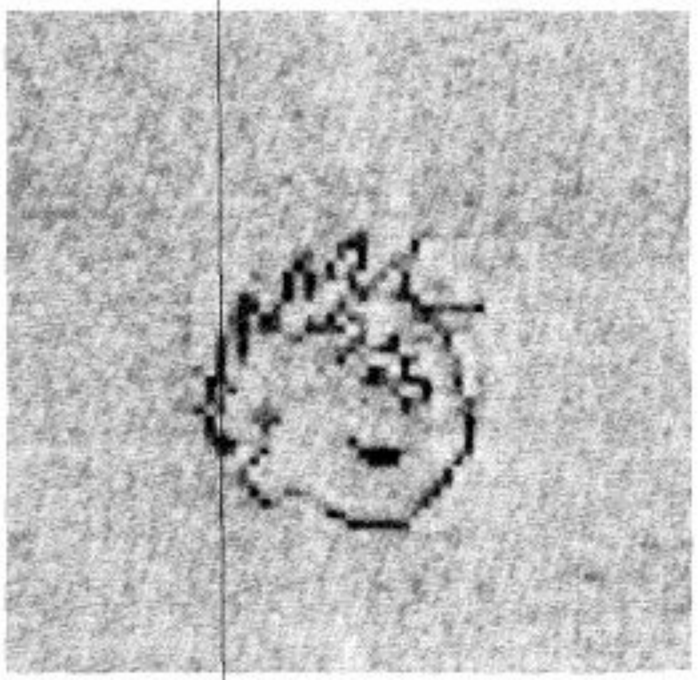


Figure 4
Reconstruction from 10 frames with random translation.

Reconstruction from bispectrum [G'89, SG'92]

$$\hat{\theta} = (\hat{\theta}_0, \ldots, \hat{\theta}_{d-1})$$

First moment:	$\hat{\theta}_0$	(DC)
Second moment:	$\{\lvert\hat{\theta}_j\rvert^2\}$	(magnitude)
Third moment:	$\{\hat{\theta}_i \hat{\theta}_j \hat{\theta}_{-i-j}\}$	(phase)

$$\frac{1}{n}\sum_{i=1}^{n} Y_i^{\otimes 3} \to \mathbb{E}[Y_i^{\otimes 3}]$$ at rate $\sigma^3/\sqrt{n}$

$$Y_i = R_{\ell_i} + \sigma \xi_i$$
$\xi_i \sim \mathcal{N}(0, I)$
R_{ℓ_i} : shift

Orbit recovery

Theorem [BRW'17]:

- Optimal rate of estimation for **worst case** signals:

$$\mathbb{E}[\rho(\tilde{\theta}, \theta)] \asymp \frac{\sigma^{d-2}}{\sqrt{n}}$$

$$\rho(\tilde{\theta}, \theta) := \min_{R \in \mathbb{Z}_d} \|\tilde{\theta} - R\theta\|$$

- Optimal rate of estimation for **typical** signals:

$$\mathbb{E}\big[\rho(\tilde{\theta}, \theta)\big] \asymp \frac{\sigma^{3}}{\sqrt{n}}$$

- Can interpolate between them: for $2 \leq s \leq d/2$, there exists a class of signals on which optimal rate is:

$$\mathbb{E}\big[\rho(\tilde{\theta}, \theta)\big] \asymp \frac{\sigma^{2s-1}}{\sqrt{n}}$$

Moments are enough

Main technical Theorem [BRW'17]:

Let $\theta, \tau \in \mathbb{R}^d$, $\rho(\theta, \tau) = \varepsilon$. $\Delta_m = \mathbb{E}[(R\theta)^{\otimes m}] - \mathbb{E}[(R\tau)^{\otimes m}]$.

If there exists a k such that

$$\|\Delta_m\| = o(\varepsilon) \text{ for } m = 1, \ldots, k-1,$$

$$\|\Delta_k\| = \Omega(\varepsilon)$$

as $\varepsilon \to 0$,

then

$$D(\mathrm{P}_\theta \,\|\, \mathrm{P}_\tau) \asymp \sigma^{-2k} \varepsilon^2$$

"If you can match $k-1$ moments, divergence is $\Theta(\sigma^{-2k})$"

Note: holds for *any* subgroup of orthogonal group.

Moments = Rates

For optimal estimator,

$$\mathbb{E}\big[\rho(\tilde{\theta},\theta)\big] \asymp \frac{\sigma^k}{\sqrt{n}}$$

where $k-1$ is the maximum number of matchable moments.

…a growing theme in statistics [LNS'99, CL'11, WY'16]

…actually common in Gaussian mixtures [L'89, HK'15]

"On estimation of the L_r norm of a regression function," Lepski, Nemirovski, Spokoiny. 1999

"Testing composite hypotheses, Hermite polynomials and optimal estimation of a nonsmooth functional," Cai, Low. 2011

"Minimax rates of entropy estimation on large alphabets via best polynomial approximation," Wu, Yang. 2016

"Moment matrices: applications in mixtures", Lindsay. 1989

"Optimal rates for finite mixture estimation", Heinrich, Kahn. 2015

Typical signals

$$\hat{\theta} = (\hat{\theta}_0, \ldots, \hat{\theta}_{d-1})$$

First moment:	$\hat{\theta}_0$	(DC)
Second moment:	$\{\lvert\hat{\theta}_j\rvert^2\}$	(magnitude)
Third moment:	$\{\hat{\theta}_i\hat{\theta}_j\hat{\theta}_{-i-j}\}$	(phase)

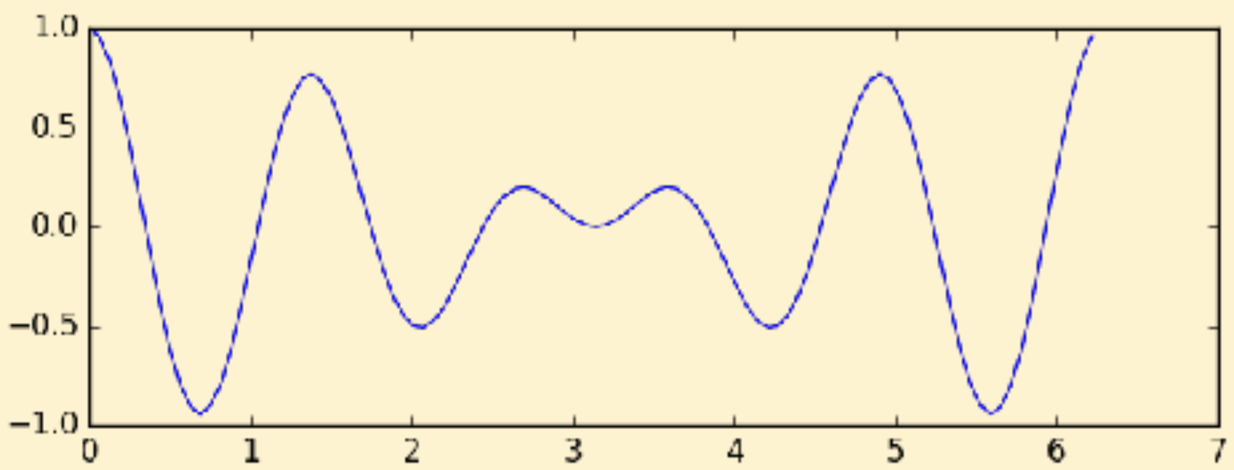
$\frac{1}{n}\sum_{i=1}^{n} Y_i^{\otimes 3} \to \mathbb{E}[Y_i^{\otimes 3}]$ at rate $\sigma^3/\sqrt{n}$ **actually optimal**

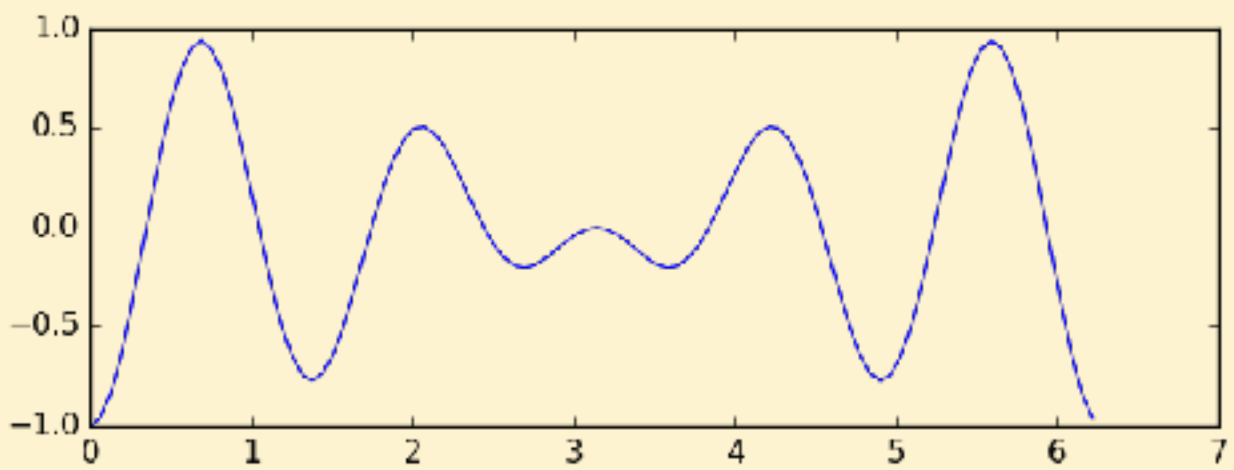
Worst-case signals

$$\hat{\theta}_4 = \hat{\theta}_{-4} = \frac{1}{2} \qquad \hat{\tau}_4 = \hat{\tau}_{-4} = -\frac{1}{2}$$

$$\hat{\theta}_5 = \hat{\theta}_{-5} = \frac{1}{2} \qquad \hat{\tau}_5 = \hat{\tau}_{-5} = -\frac{1}{2}$$

all other entries zero





Worst-case signals

$$\hat{\theta}_4 = \hat{\theta}_{-4} = \frac{1}{2} \qquad \hat{\tau}_4 = \hat{\tau}_{-4} = -\frac{1}{2}$$

$$\hat{\theta}_5 = \hat{\theta}_{-5} = \frac{1}{2} \qquad \hat{\tau}_5 = \hat{\tau}_{-5} = -\frac{1}{2}$$

all other entries zero

First moment:	$\hat{\theta}_0$	(DC)	**same**
Second moment:	$\{\lvert\hat{\theta}_j\rvert^2\}$	(magnitude)	**same**
Third moment:	$\{\hat{\theta}_i \hat{\theta}_j \hat{\theta}_{-i-j}\}$	(phase)	**vanishes**

First **eight** moments match!

Worst-case signals

$$\hat{\theta}_4 = \hat{\theta}_{-4} = \frac{1}{2} \qquad \hat{\tau}_4 = \hat{\tau}_{-4} = -\frac{1}{2}$$

$$\hat{\theta}_5 = \hat{\theta}_{-5} = \frac{1}{2} \qquad \hat{\tau}_5 = \hat{\tau}_{-5} = -\frac{1}{2}$$

First **eight** moments match!

$$D(\mathrm{P}_\theta \parallel \mathrm{P}_\tau) \asymp \sigma^{-18}$$

Takeaways

Divergence between Gaussian mixtures usually hard to compute, but reduces to **moment matching**

Maximum likelihood estimator gives **optimal** noise dependence

Algorithms

[PWBRS'17]: New **tensor-based** algorithms for MRA

Achieves optimal σ^3 dependence on SNR for generic signals

Modified algorithm can handle **heterogenous** mixtures with σ^5 dependence

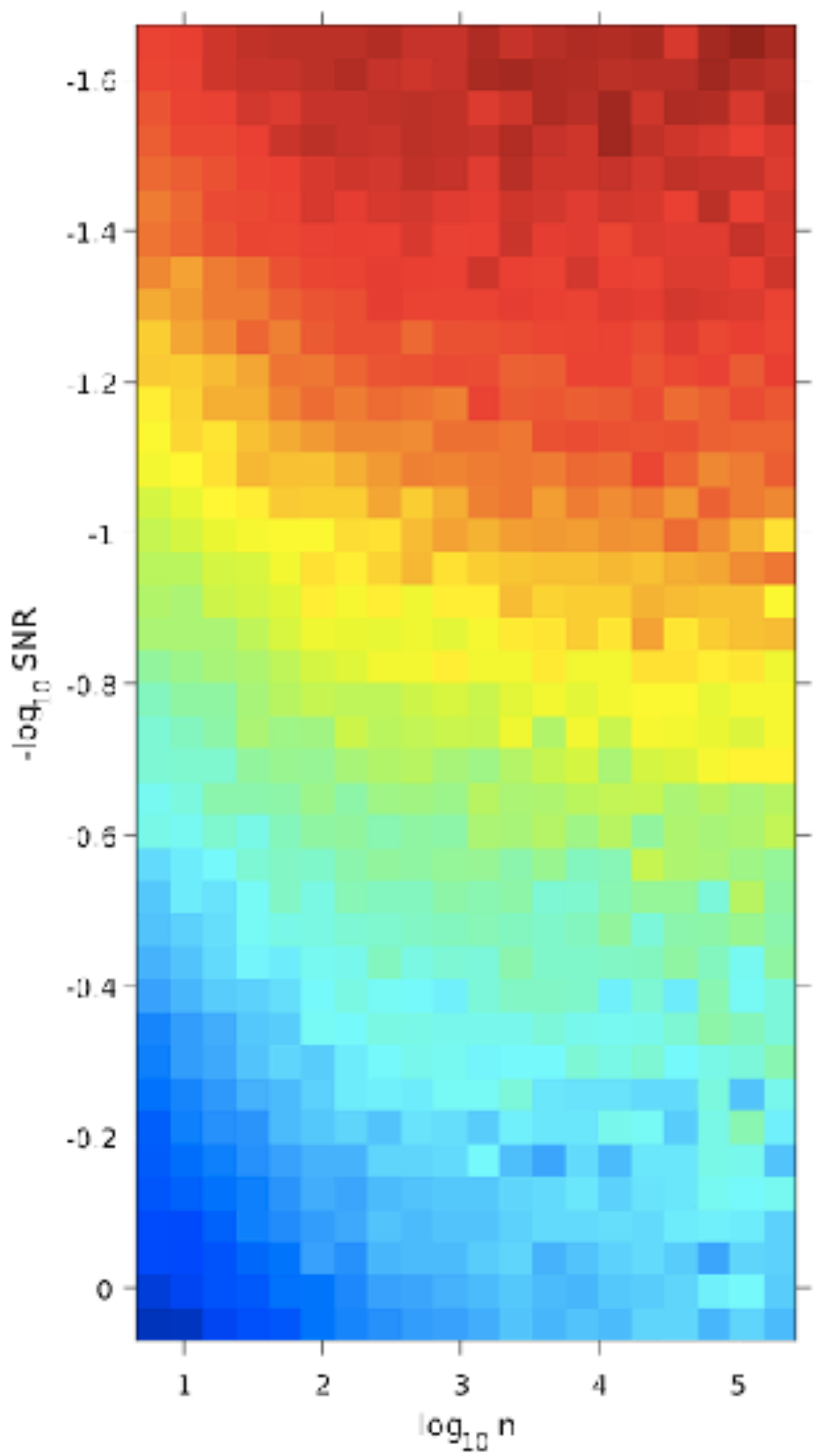
$$Y_i = R_i \theta^{(i)} + \sigma \xi_i$$

R_i i.i.d uniform cyclic shifts

$\theta^{(i)}$ i.i.d from finite mixture of linearly indep. signals

Heterogeneity

3 components



relative error
(red is good)

Takeaway

Polynomial-time **tensor algorithms** are (much) stronger than synchronization approaches

More information about bispectrum-based algorithms in next talk!

Group Structure

Main technical Theorem [BRW'17]:

Let $\theta, \tau \in \mathbb{R}^d$, $\rho(\theta,\tau) = \varepsilon.$ $\quad\Delta_m = \mathbb{E}[(R\theta)^{\otimes m}] - \mathbb{E}[(R\tau)^{\otimes m}].$

If there exists a k such that

$$\|\Delta_m\| = o(\varepsilon) \text{ for } m = 1, \ldots, k-1,$$
$$\|\Delta_k\| = \Omega(\varepsilon)$$

as $\varepsilon \to 0,$

then

$$D(\mathrm{P}_\theta \parallel \mathrm{P}_\tau) \asymp \sigma^{-2k}\varepsilon^2$$

"If you can match $k-1$ moments, divergence is $\Theta(\sigma^{-2k})$"

Note: holds for *any* subgroup of orthogonal group.

Group Structure

$$\Delta_m = \mathbb{E}[(R\theta)^{\otimes m}] - \mathbb{E}[(R\tau)^{\otimes m}]$$

$\downarrow$ $\swarrow$

depend on group structure

entries of $\mathbb{E}[(R\theta)^m]$ are degree m polynomials in entries of θ

invariant theory

Invariant Theory

Classical question in group theory: describe a

ring of polynomials in d variables which are

invariant under the action of a group $\mathcal{G}$ on $\mathbb{R}^d$

$$P(R\theta) = P(\theta) = \sum_{\mathbf{k}} \alpha_{\mathbf{k}} \theta^{\mathbf{k}}$$

$$\mathbf{k} = (k_1, \ldots, k_d), \;\; \theta^{\mathbf{k}} = \theta_1^{k_1} \theta_2^{k_2} \cdots \theta_d^{k_d}$$

Invariant Theory

"Theorem": If the algebra $R^{\mathcal{G}}$ of $\mathcal{G}$-invariant polynomials is generated as an algebra by polynomials of degree at most g, then orbit recovery problem can be solved at rate $\sigma^g/\sqrt{n}$.

Proof:

Generated by polynomials of degree $\leq g$. $\longrightarrow$ Orbit of θ determined by $\mathbb{E}[(R\theta)^{\otimes m}], m \leq g$ $\longrightarrow$ $D(\mathrm{P}_\theta \parallel \mathrm{P}_\tau) \gtrsim \sigma^{-2g}$ for $\theta \neq \tau$.

Example

$\mathcal{G} = SO(d)$ group of all rotations

$Y_i = R_{\ell_i}\theta + \sigma\xi_i$ "norm recovery"

algebra of invariants = generated by squared 2 norm

expect optimal
rate of estimation
$\sigma^2/\sqrt{n}$

Example

$\mathcal{G} = \mathfrak{S}_d$ group of all permutations

$Y_i = R_{\ell_i}\theta + \sigma\xi_i$ "bag of values"

algebra of invariants = symmetric polynomials

expect optimal
rate of estimation
$\sigma^d/\sqrt{n}$

Takeaway

Rates of estimation correspond to properties of **algebra of invariant polynomials**—a well-studied object in group theory

Multi-reference alignment

$$Y_i = R_{\ell_i}\theta + \sigma Z_i$$

- **Optimal** rates of estimation
- **Efficient algorithms** via tensor decomposition
- Link between rates and **polynomial invariants**

- **Dimension dependence?**
- **Projection step?** Observe ΠY_i

Sources

- A. S. Bandeira, P. Rigollet, J. Weed.
  Optimal rates of estimation for multi-reference alignment.
  2017.
  arXiv:1702.08546 [math.ST]

- A. Perry, J. Weed, A. S. Bandeira, P. Rigollet, A. Singer.
  The sample complexity of multi-reference alignment.
  2017.
  Manuscript.