The University of Edinburgh

Compressive Independent Component Analysis

Michael P. Sheehan

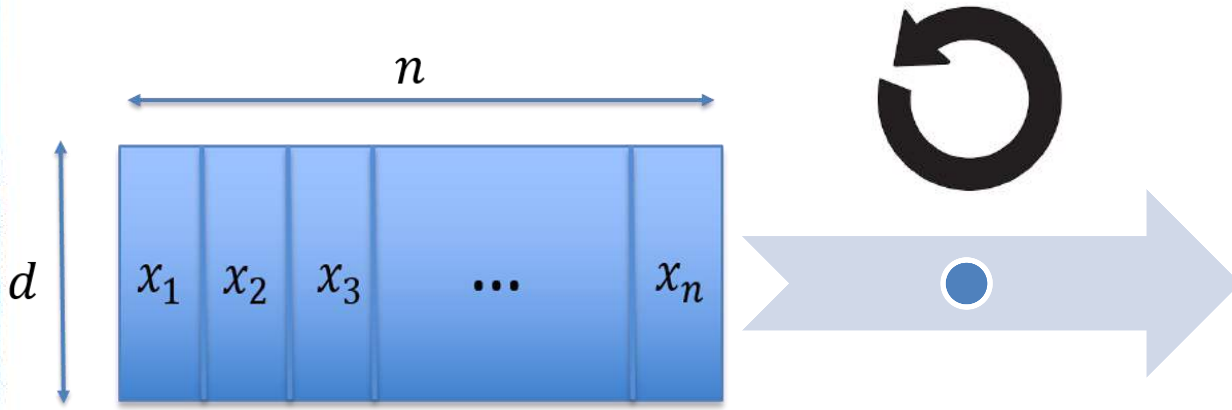
IDCOM, University of Edinburgh

Joint work with Mike E. Davies and Madeleine Kotzagiannidis

Outline

1. Compressive Learning
2. Compressive ICA
   i. Independent Component Analysis
   ii. Motivation
   iii. Theory
   iv. Inverse Problem
3. Results
   i. Phase Transition
   ii. Toy Example
4. Outlook

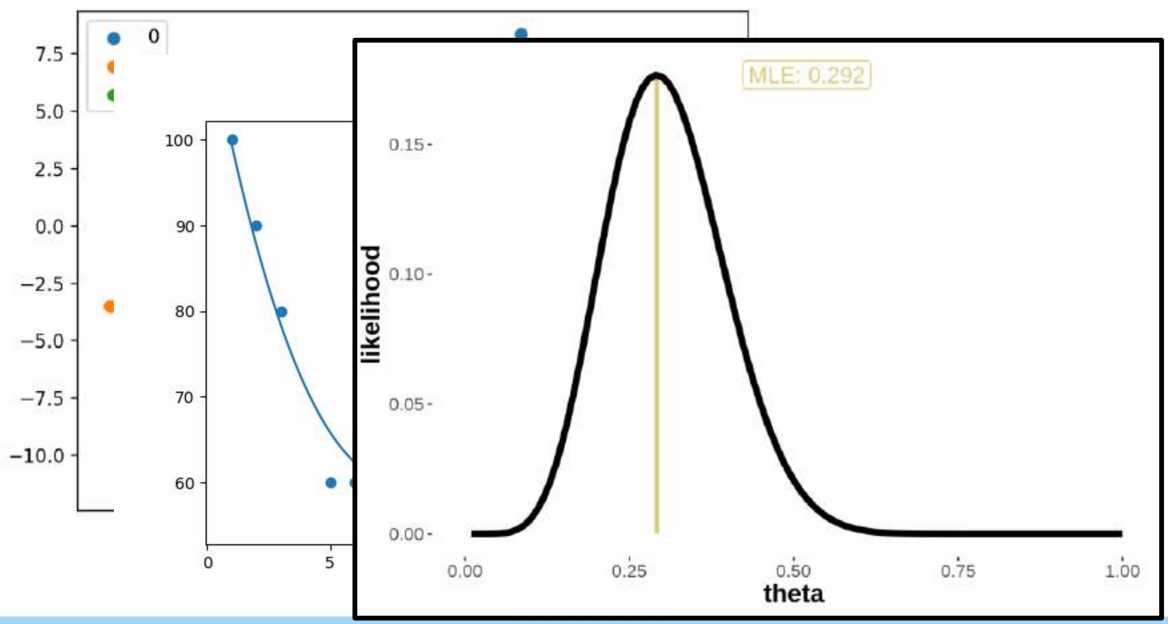
1. Compressive Learning

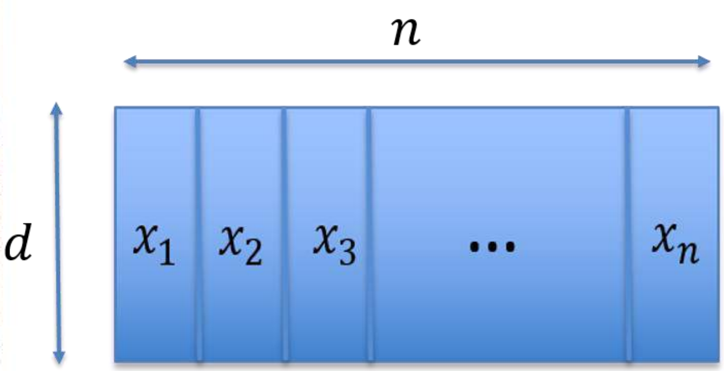


Recourse

Learning

- Classification
- Regression
- Parameter estimation



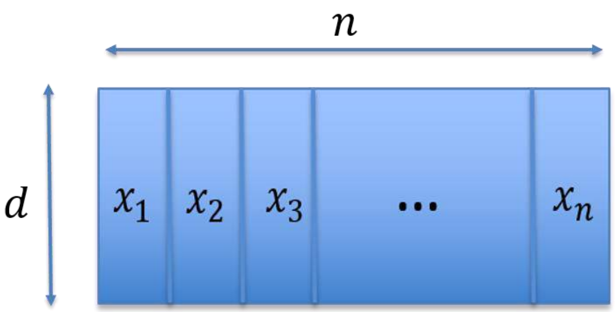


Challenges

1. Dataset has to be stored in memory
2. Computation complexity may scale with the dataset dimensions
3. Amenable to online/distributed learning ?

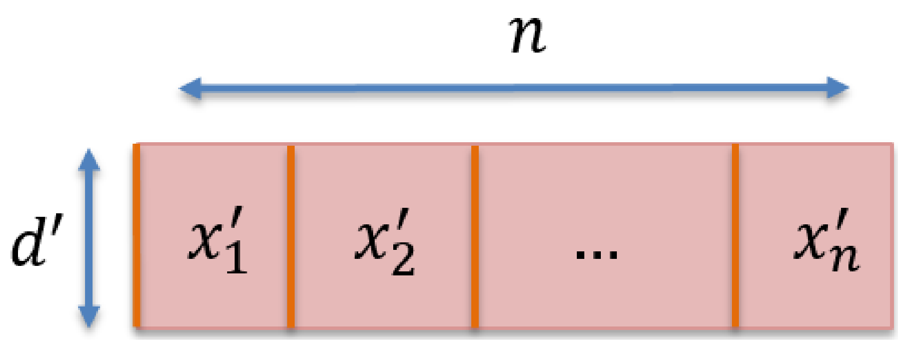
Large Scale Learning

What if d or n is large?



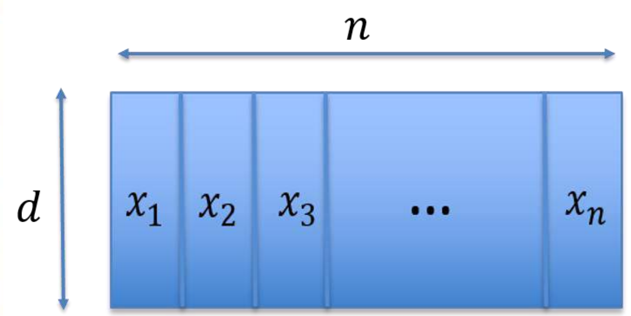
a) Original

Compression Schemes

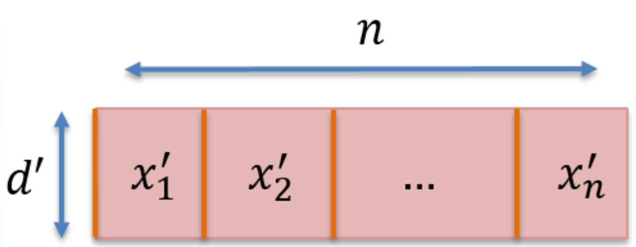


- **e.g. Feature Selection, Random Projection**
- **Does not compress the number of data points n**

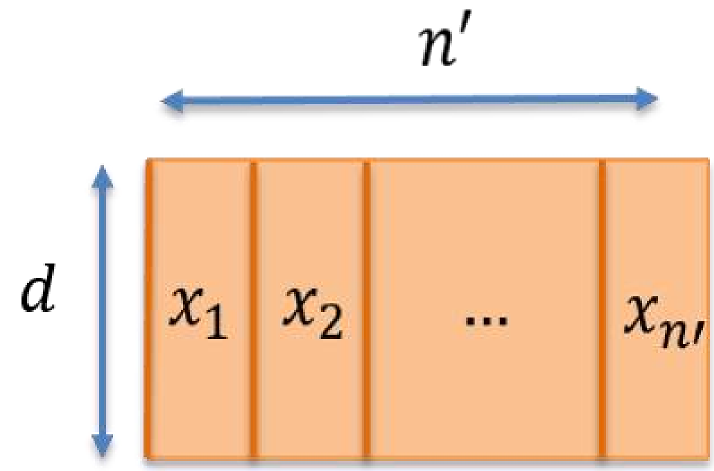
Compression Schemes



a) Original

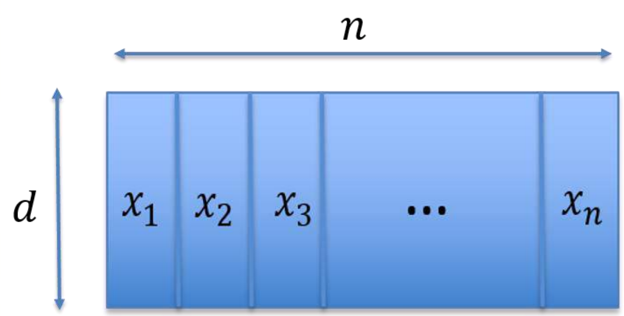


b) Dimensionality Reduction

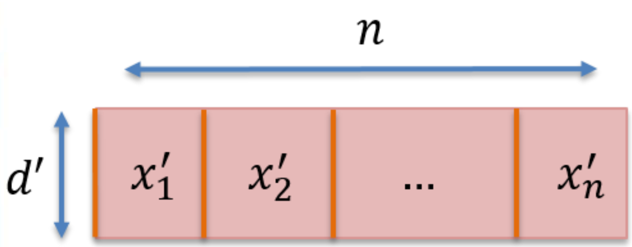


- **e.g. sampling, Nystrom, coresets**
- **Does not compress the feature space**
- **Could potentially discard important items**

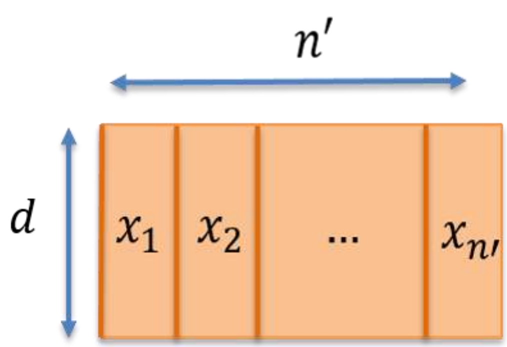
Linear Sketches



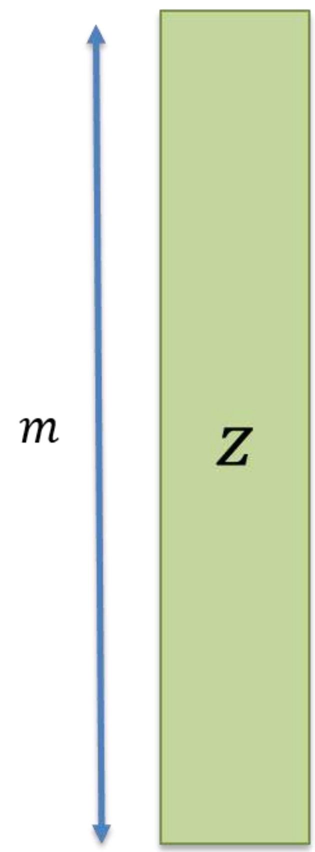
a) Original



b) Dimensionality Reduction

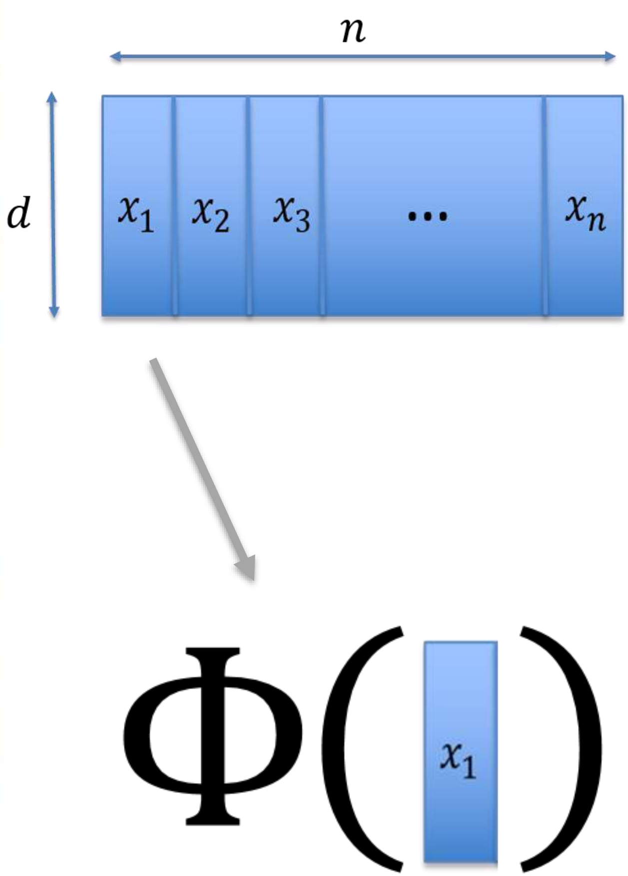


c) Subsampling



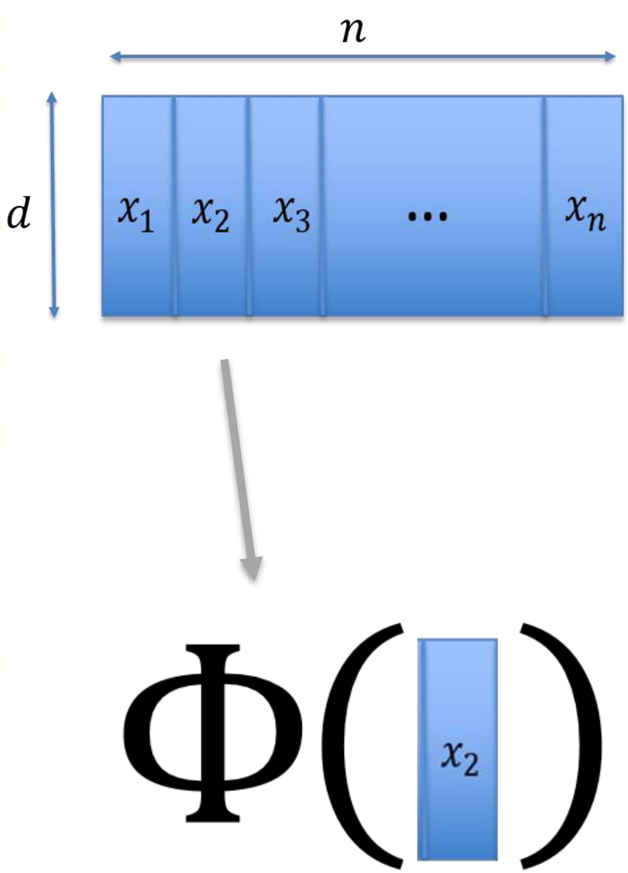
- **The sketch has dimension $m \ll n\, d$**
- **The size m typically scales independent of n and d**
- **Amenable to online learning**

Compressive learning: How do we form the sketch ?



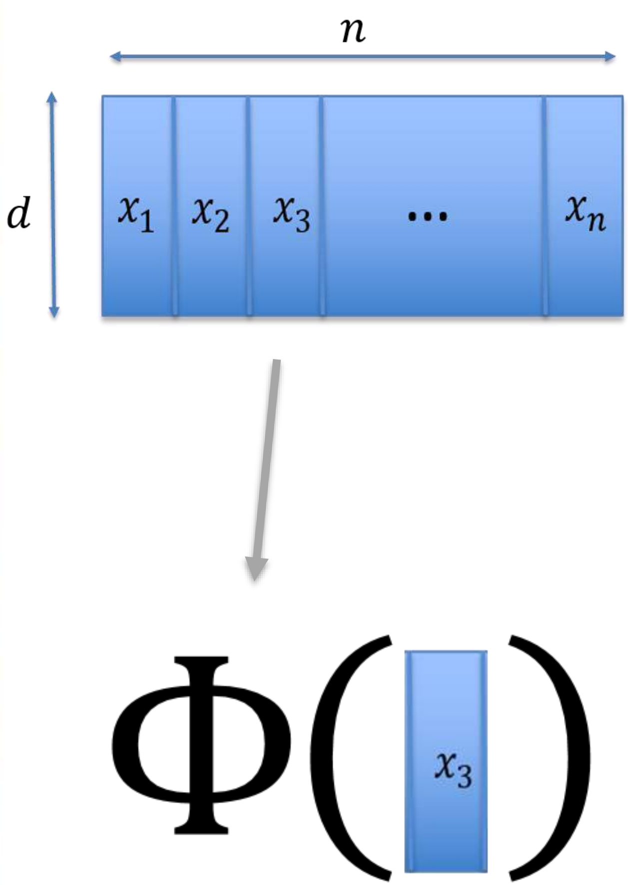
- **We pass each data point $\boldsymbol{x_i}$ through a feature function $\boldsymbol{\Phi}: \mathbb{R}^{\boldsymbol{d}} \rightarrow \mathbb{C}^{\boldsymbol{m}}$**

Compressive learning: How do we form the sketch ?

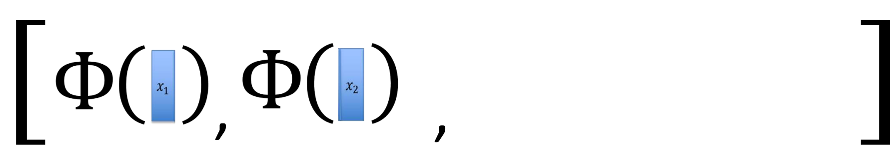


- **We pass each data point $\boldsymbol{x_i}$ through a feature function $\boldsymbol{\Phi}$: $\mathbb{R}^{\boldsymbol{d}} \rightarrow \mathbb{C}^{\boldsymbol{m}}$**

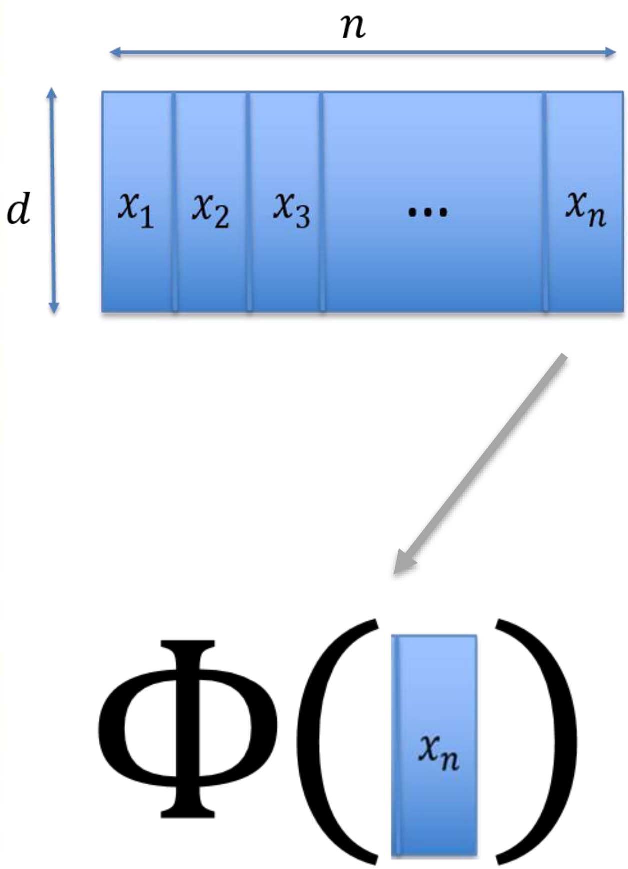
Compressive learning: How do we form the sketch ?



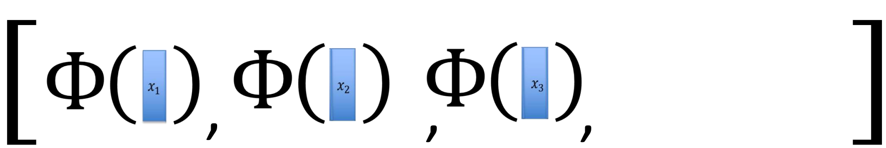
- **We pass each data point $\boldsymbol{x_i}$ through a feature function $\boldsymbol{\Phi}$: $\mathbb{R}^{\boldsymbol{d}} \rightarrow \mathbb{C}^{\boldsymbol{m}}$**



Compressive learning: How do we form the sketch ?



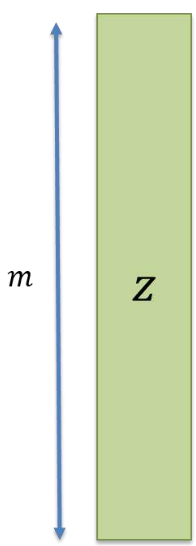
- **We pass each data point $\boldsymbol{x_i}$ through a feature function $\boldsymbol{\Phi}$: $\mathbb{R}^{\boldsymbol{d}} \to \mathbb{C}^{\boldsymbol{m}}$**



Compressive learning: How do we form the sketch ?

$$\left[\Phi(x_1), \Phi(x_2), \Phi(x_3), \ldots, \Phi(x_n)\right]$$

- We then pool and average the feature function of each data point to form the sketch



$$= \frac{1}{n}\sum_{i=1}^{n}\Phi(x_i)$$

Compressive learning: How do we form the sketch ?

$$\underbrace{z}_{m} = \frac{1}{n}\sum_{i=1}^{n}\Phi(x_i)$$

Advantages

- Only the sketch of size m has to be stored in memory
- Typically, $m \ll nd$
- Easily amenable to online and distributed learning

How do we learn from a sketch?

- We want to learn the parameters θ of the learning model $\pi_\theta \in \mathcal{P}$ where $x \sim \pi_\theta$
- Similar to moment matching, we match the sketch with it's expectation

$$\min_{\theta \in \Theta} \left\| z - \mathbb{E}_{x \sim \pi_\theta} \Phi(x) \right\|$$

Sketch (empirical moment)

True moment

Compressive Learning: How is it possible?

- Sketch reformulation: $\mathcal{A}(\pi_\theta) := \mathbb{E}_{x\sim\pi_\theta}\Phi(x)$
- $\mathcal{A}$: $\mathcal{P}\rightarrow\mathbb{R}^m$ equivalently a linear operator acting the model

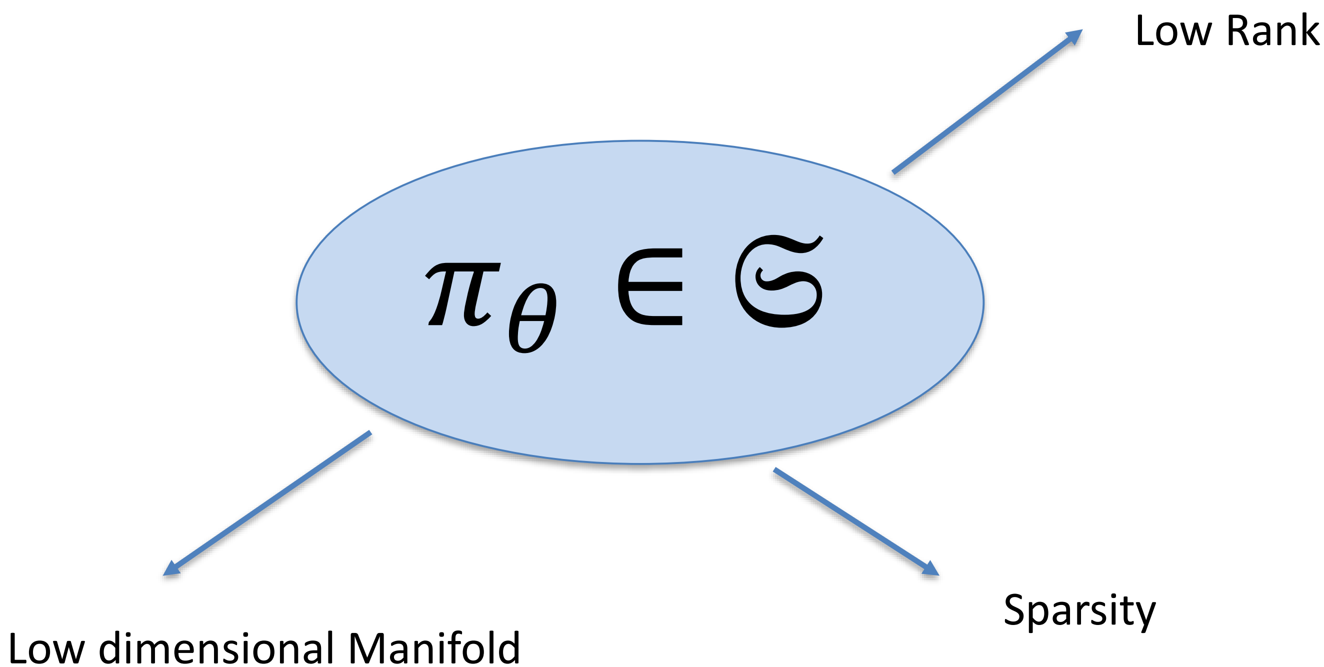
Reformulated

$$\min_{\pi_\theta \in \mathfrak{S}} \|z - \mathcal{A}(\pi_\theta)\|$$

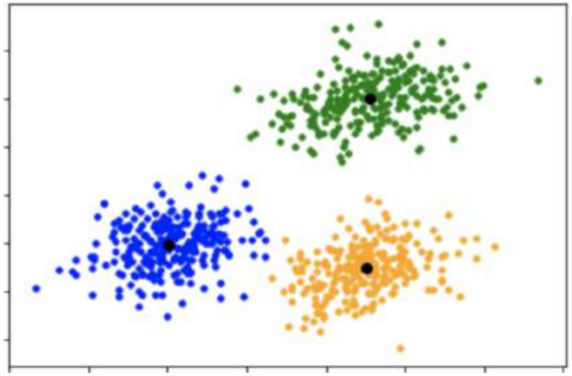
Constrained model set

Compressive Learning: How is it possible?

$$\min_{\pi_\theta \in \mathfrak{S}} \|z - \mathcal{A}(\pi_\theta)\|$$



What has compressive learning achieved so far?

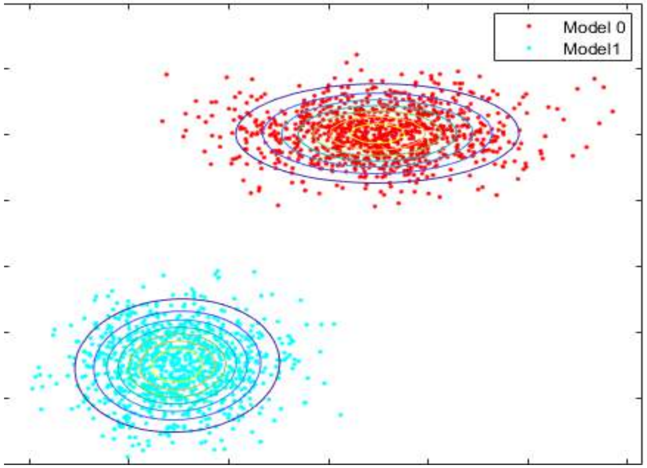


1. Compressive k means

- **Model set** $\mathfrak{S}$ $\pi_\theta : k$ centres $c_1, c_2, \dots, c_k$
- **Feature Function** $\Phi(x) = \left(\frac{e^{i\omega_j^T x}}{w\ \omega_j} \right)_{\mathrm{j=1}}^{\mathrm{m}}$
- **Sketch Size** $m \approx \mathcal{O}(\mathrm{kd})$

"Compressive Statistical Learning with Random Feature Moments" Gribonval et. al. 2020

What CL has achieved so far?



1. Compressive k means
2. Compressive GMM

- **Model set** $\mathfrak{S}$ π_θ: mixture of k Gaussians
- **Feature Function** $\Phi(x) = \left(e^{i\omega_j^T x}\right)_{\mathrm{j=1}}^{\mathrm{m}}$
- **Sketch Size** $m \approx \mathcal{O}(\mathrm{kd})$

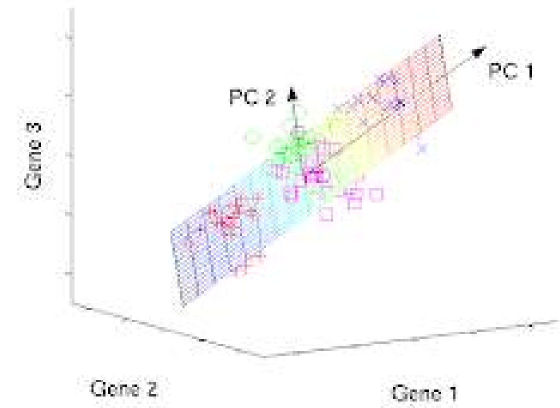
"Compressive Statistical Learning with Random Feature Moments" Gribonval et. al. 2020

What CL has achieved so far?

1. Compressive k means

2. Compressive GMM

3. Compressive PCA



- **Model set $\mathfrak{S}$** π_θ: k dimensional subspace - (rank$\left(\Sigma_{\pi_\theta}\right) \leq k$)
- **Feature Function** $\Phi(x) = \left\langle a_j, xx^T \right\rangle_{j=1}^{m}$
- **Sketch Size** $m \approx \mathcal{O}$(kd)

“Compressive Statistical Learning with Random Feature Moments” Gribonval et. al. 2020

2. Compressive ICA

Independent Component Analysis

- ICA is a method to identify latent variables that are mutually independent to one another.
- Applications: Blind source separation, EEG recordings, financial modelling, telecommunications
- Given a dataset $\boldsymbol{X} = (\boldsymbol{x_1}, \boldsymbol{x_2}, \dots, \boldsymbol{x_n}) \in \mathbb{R}^{n\times d}$
- $\boldsymbol{S} = (\boldsymbol{s_1}, \boldsymbol{s_2}, \dots, \boldsymbol{s_n}) \in \mathbb{R}^{n\times d}$
- Mixing matrix $\boldsymbol{Q} \in \mathbb{R}^{d\times d}$

<u>Goal</u>

Estimate **Q** from $\boldsymbol{X}$

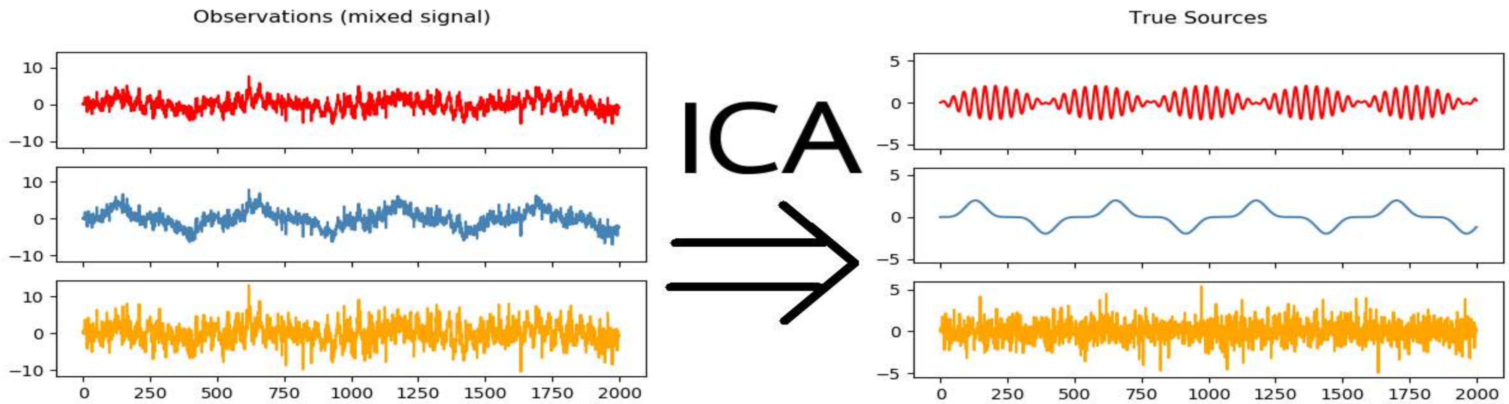
$$\boldsymbol{X} = \boldsymbol{QS}$$

Dataset

Mixing Matrix

Independent Components

Independent Component Analysis



ICA assumptions

- Each source signal $\boldsymbol{s} = (s_1, s_2, \ldots, s_d)$ in time has components that are mutually independent:

$$\pi(\boldsymbol{s}) = \prod_{i=1}^{d} \pi_i(s_i)$$

- The individual distributions are assumed non-Gaussian but left unspecified (Semi-parametric)

How do we estimate $\boldsymbol{Q}$?

Optimise $\widehat{\boldsymbol{Q}}$

Maximize the independence $\hat{\boldsymbol{s}}$

$$\hat{\boldsymbol{s}} = \widehat{\boldsymbol{Q}}^T \boldsymbol{x}$$

<u>Measures of independence</u>

- Mutual information
- KL divergence
- Kurtosis (cumulant based methods)

Cumulant Based ICA (Kurtosis)

$$\hat{\boldsymbol{s}} = \widehat{\boldsymbol{Q}}^T \boldsymbol{x}$$

- The 4^{th} order cumulant (kurtosis) of a $\boldsymbol{x}$ is defined as

$$(\mathcal{X}_{\pi_\theta})_{ijkl} = \mathbb{E}_{\pi_\theta}[x_i x_j x_k x_l] - \mathbb{E}_{\pi_\theta}[x_i x_j]\mathbb{E}_{\pi_\theta}[x_k x_l] - \mathbb{E}_{\pi_\theta}[x_i x_k]\mathbb{E}_{\pi_\theta}[x_j x_l] - \mathbb{E}_{\pi_\theta}[x_i x_l]\mathbb{E}_{\pi_\theta}[x_j x_k]$$

- In our multivariate setting, the 4th order cumulant gives rise to a 4th order tensor $\mathcal{X}_{\pi_\theta} \in \mathbb{R}^{d\times d\times d\times d}$
- <u>Note</u>

1. The diagonal entries $(\mathcal{X}_{\pi_\theta})_{iiii}$ auto-cumulants
2. The off-diagonal entries $(\mathcal{X}_{\pi_\theta})_{ijkl}$ are the cross-cumulants

Cumulant Based ICA (Kurtosis)

$\hat{\boldsymbol{s}} = \widehat{\boldsymbol{Q}}^T \boldsymbol{x}$

- The 4^{th} order cumulant (kurtosis) of a $\boldsymbol{x}$ is defined as

$$(\mathcal{X}_{\pi_\theta})_{ijkl} = \mathbb{E}_{\pi_\theta}[x_i x_j x_k x_l] - \mathbb{E}_{\pi_\theta}[x_i x_j]\mathbb{E}_{\pi_\theta}[x_k x_l] - \mathbb{E}_{\pi_\theta}[x_i x_k]\mathbb{E}_{\pi_\theta}[x_j x_l] - \mathbb{E}_{\pi_\theta}[x_i x_l]\mathbb{E}_{\pi_\theta}[x_j x_k]$$

- In our multivariate setting, the 4^{th} order cumulant gives rise to a 4^{th} order tensor $\mathcal{X}_{\pi_\theta} \in \mathbb{R}^{d\times d\times d\times d}$
- Note

1. The diagonal entries $(\mathcal{X}_{\pi_\theta})_{iiii}$ auto-cumulants
2. The off-diagonal entries $(\mathcal{X}_{\pi_\theta})_{ijkl}$ are the cross-cumulants

Zero when independent

Cumulant Based ICA (Kurtosis)

$\hat{\boldsymbol{s}} = \widehat{\boldsymbol{Q}}^T \boldsymbol{x}$

GOAL

Optimize $\widehat{\boldsymbol{Q}}$ such that

$$\hat{\mathcal{S}} = \mathcal{X}_{\pi_\theta} \times_1 \widehat{\boldsymbol{Q}}^T \times_2 \widehat{\boldsymbol{Q}}^T \times_3 \widehat{\boldsymbol{Q}}^T \times_4 \widehat{\boldsymbol{Q}}^T$$

is **diagonal**

Compressive ICA

- The geometry of the cumulant tenors gives rise to a natural model set:

$$\mathfrak{S} = \{\pi_\theta \mid \mathcal{X}_{\pi_\theta} = \mathcal{S} \times_1 \boldsymbol{Q} \times_2 \boldsymbol{Q} \times_3 \boldsymbol{Q} \times_4 \boldsymbol{Q}\}$$

Compressive ICA

- The geometry of the cumulant tenors gives rise to a natural model set:

$$\mathfrak{S} = \{\pi_\theta \mid \mathcal{X}_{\pi_\theta} = \mathcal{S} \times_1 \boldsymbol{Q} \times_2 \boldsymbol{Q} \times_3 \boldsymbol{Q} \times_4 \boldsymbol{Q}\}$$

Diagonal Tensor with
d degrees of freedom

Compressive ICA

- The geometry of the cumulant tenors gives rise to a natural model set:

$$\mathfrak{S} = \{\pi_\theta \mid \mathcal{X}_{\pi_\theta} = \mathcal{S} \times_1 \boldsymbol{Q} \times_2 \boldsymbol{Q} \times_3 \boldsymbol{Q} \times_4 \boldsymbol{Q}\}$$

Diagonal Tensor with d degrees of freedom

Orthogonal matrix with $\frac{d(d-1)}{2}$ degrees of freedom

Compressive ICA

- The geometry of the cumulant tenors gives rise to a natural model set:

$$\mathfrak{S} = \{\pi_\theta \mid \mathcal{X}_{\pi_\theta} = \mathcal{S} \times_1 \boldsymbol{Q} \times_2 \boldsymbol{Q} \times_3 \boldsymbol{Q} \times_4 \boldsymbol{Q}\}$$

Total of $\frac{d(d+1)}{2}$ degrees of freedom

Diagonal Tensor with d degrees of freedom

Orthogonal matrix with $\frac{d(d-1)}{2}$ degrees of freedom

Compressive ICA

- The geometry of the cumulant tenors gives rises to a natural model set:

$$\mathfrak{S} = \{\pi_\theta \mid \mathcal{X}_{\pi_\theta} = \mathcal{S} \times_1 \boldsymbol{Q} \times_2 \boldsymbol{Q} \times_3 \boldsymbol{Q} \times_4 \boldsymbol{Q}\}$$

Total of d$(d+1)$ degrees of freedom

- $\mathfrak{S}$ is a low dimension model set residing in $\mathbb{R}^{d\times d\times d\times d}$

Compressive ICA

- The geometry of the cumulant tenors gives rises to a natural model set:

$$\mathfrak{S} = \{\pi_\theta \mid \mathcal{X}_{\pi_\theta} = \mathcal{S} \times_1 \boldsymbol{Q} \times_2 \boldsymbol{Q} \times_3 \boldsymbol{Q} \times_4 \boldsymbol{Q}\}$$

Total of d$(d+1)$ degrees of freedom

- $\mathfrak{S}$ is a low dimension model set residing in $\mathbb{R}^{d\times d\times d\times d}$
- Can we exploit the structure/sparsity of $\mathcal{X}_{\pi_\theta}$ to from a sketch?

Compressive ICA: Forming the Sketch

- Feature function: $\Phi(\boldsymbol{x}) = \langle \boldsymbol{a_j}, \boldsymbol{x} \otimes \boldsymbol{x} \otimes \boldsymbol{x} \otimes \boldsymbol{x} \rangle_{j=1}^{m}$

- Equivalent Sketching Operator: $\mathcal{A}(\mathcal{X}_{\pi_\theta}) = \boldsymbol{A}\mathrm{vec}(\mathcal{X}_{\pi_\theta})$

  where $\boldsymbol{A} \in \mathbb{R}^{m \times d^4}$ is a random (sub) Gaussian matrix

c.f. compressive sensing

Compressive ICA: Forming the Sketch

- Feature function: $\Phi(\boldsymbol{x}) = \langle \boldsymbol{a_j}, \boldsymbol{x} \otimes \boldsymbol{x} \otimes \boldsymbol{x} \otimes \boldsymbol{x} \rangle_{j=1}^{m}$

- Equivalent Sketching Operator: $\mathcal{A}\left(\mathcal{X}_{\pi_\theta}\right) = \boldsymbol{A}\mathrm{vec}(\mathcal{X}_{\pi_\theta})$

  where $\boldsymbol{A} \in \mathbb{R}^{m \times d^4}$ is a random (sub) Gaussian matrix

Recall: $\boldsymbol{z} = \frac{1}{n}\sum_{i=1}^{n} \Phi(\boldsymbol{x}_i)$

Compressive ICA: Forming the Sketch

- Feature function: $\Phi(\boldsymbol{x}) = \langle \boldsymbol{a_j}, \boldsymbol{x} \otimes \boldsymbol{x} \otimes \boldsymbol{x} \otimes \boldsymbol{x} \rangle_{j=1}^{m}$

- Equivalent Sketching Operator: $\mathcal{A}(\mathcal{X}_{\pi_\theta}) = \boldsymbol{A}\mathrm{vec}(\mathcal{X}_{\pi_\theta})$

  where $\boldsymbol{A} \in \mathbb{R}^{m \times d^4}$ is a random (sub) Gaussian matrix

Recall: $\boldsymbol{z} = \frac{1}{n}\sum_{i=1}^{n} \Phi(\boldsymbol{x}_i)$

Inverse Problem

$$\min_{\mathcal{X}_{\pi_\theta} \in \mathfrak{S}} \|\boldsymbol{z} - \mathcal{A}(\mathcal{X}_{\pi_\theta})\|$$

Compressive ICA Restricted Isometry Property

Let $A_{ij} \sim N(0, \sqrt{m^{-1}})$, then for any $\xi, \delta \in (0,1)$ and $\mathcal{X}_{\pi_\theta} \in \mathfrak{S}$, we have

$$(1-\delta)\left\|\mathcal{X}_{\pi_{\theta_1}} - \mathcal{X}_{\pi_{\theta_2}}\right\|_F^2 \le \left\|\mathcal{A}\left(\mathcal{X}_{\pi_{\theta_1}} - \mathcal{X}_{\pi_{\theta_2}}\right)\right\|_2^2$$

Compressive ICA Restricted Isometry Property

Let $A_{ij} \sim N(0, \sqrt{m^{-1}})$, then for any $\xi, \delta \in (0,1)$ and $\mathcal{X}_{\pi_\theta} \in \mathfrak{S}$, we have

$$(1-\delta)\left\|\mathcal{X}_{\pi_{\theta_1}} - \mathcal{X}_{\pi_{\theta_2}}\right\|_F^2 \leq \left\|\mathcal{A}\left(\mathcal{X}_{\pi_{\theta_1}} - \mathcal{X}_{\pi_{\theta_2}}\right)\right\|_2^2$$

with prob. $1-\xi$ provided that the sketch size

$$m \geq \frac{C}{\delta^2} \max\left\{4d(d+1)\log(6)\,, \log(\frac{6}{\xi})\right\}$$

Compressive ICA Restricted Isometry Property

$$m \approx \mathcal{O}\big(d(d+1)\big)$$

$$m \geq \frac{C}{\delta^2} \max\left\{4d(d+1)\log(6)\,, \log(\frac{6}{\xi})\right\}$$

Constrained Optimization

$$\min_{\mathcal{X}_{\pi_\theta} \in \mathfrak{S}} \left\| \boldsymbol{z} - \mathcal{A}(\mathcal{X}_{\pi_\theta}) \right\|$$

- We propose an iterative projection gradient descent scheme to solve the OP

Constrained Optimization

$$\min_{\mathcal{X}_{\pi_\theta} \in \mathfrak{S}} \left\| \boldsymbol{z} - \mathcal{A}\left(\mathcal{X}_{\pi_\theta}\right) \right\|$$

- We propose an iterative projection gradient descent scheme to solve the OP
- After each gradient step, we project onto the model set $\mathfrak{S}$

Constrained Optimization

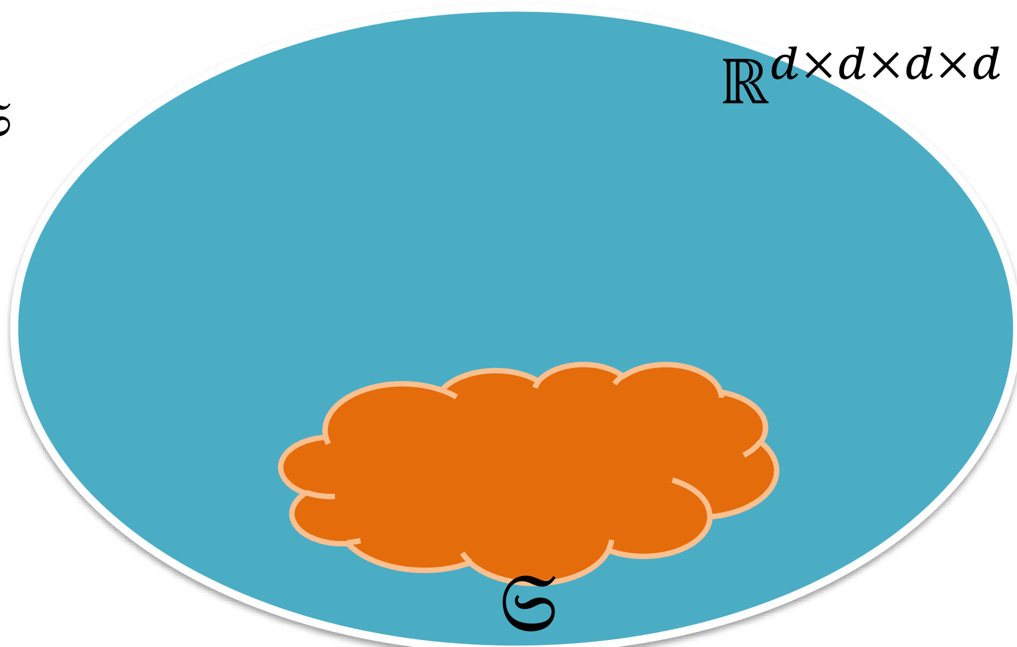
$$\min_{\mathcal{X}_{\pi_\theta} \in \mathfrak{S}} \left\| \boldsymbol{z} - \mathcal{A}\left(\mathcal{X}_{\pi_\theta}\right) \right\|$$

- We propose an iterative projection gradient descent scheme to solve the OP
- After each gradient step, we project onto the model set $\mathfrak{S}$
- Akin to Compressive Sensing (hard thresholding onto the sparse set)

Constrained Optimization

$$\min_{\mathcal{X}_{\pi_\theta} \in \mathfrak{S}} \left\| \boldsymbol{z} - \mathcal{A}(\mathcal{X}_{\pi_\theta}) \right\|$$

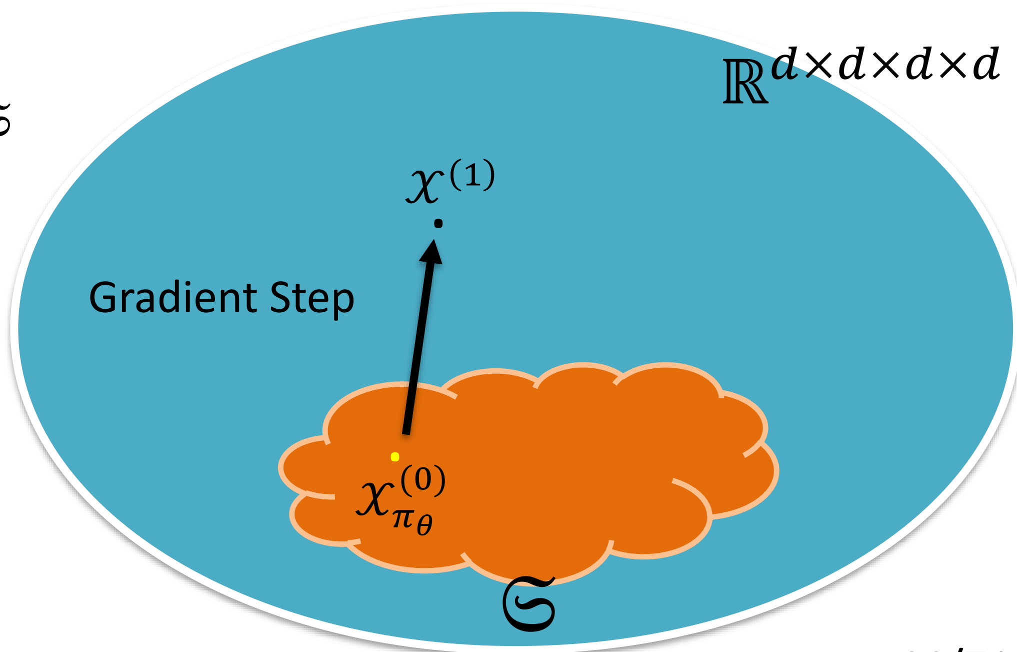
- We propose an iterative projection gradient descent scheme to solve the OP
- After each gradient step, we project onto the model set $\mathfrak{S}$
- Akin to Compressive Sensing (hard thresholding onto the sparse set)



Constrained Optimization

$$\min_{\mathcal{X}_{\pi_\theta}\in\mathfrak{S}} \left\|\boldsymbol{z} - \mathcal{A}(\mathcal{X}_{\pi_\theta})\right\|$$

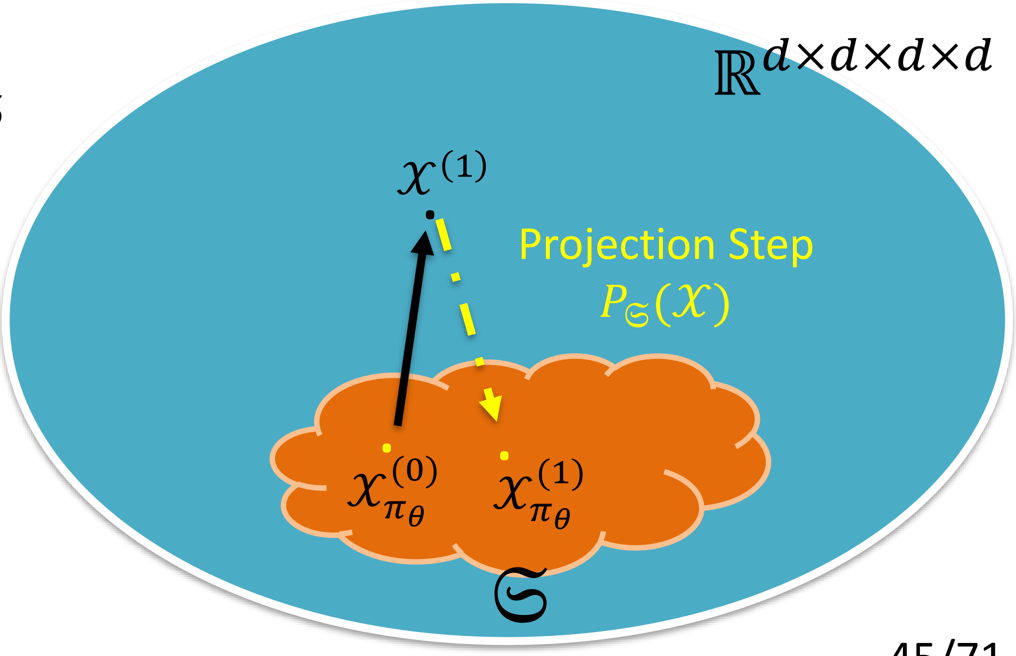
- We propose an iterative projection gradient descent scheme to solve the OP
- After each gradient step, we project onto the model set $\mathfrak{S}$
- Akin to Compressive Sensing (hard thresholding onto the sparse set)



Constrained Optimization

$$\min_{\mathcal{X}_{\pi_\theta} \in \mathfrak{S}} \left\| \boldsymbol{z} - \mathcal{A}(\mathcal{X}_{\pi_\theta}) \right\|$$

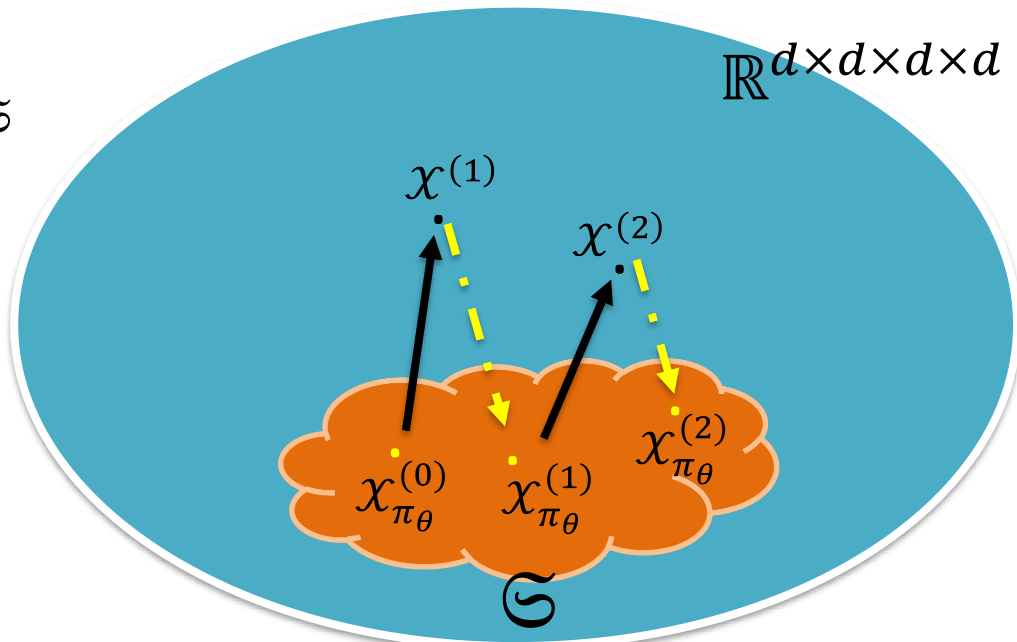
- We propose an iterative projection gradient descent scheme to solve the OP
- After each gradient step, we project onto the model set $\mathfrak{S}$
- Akin to Compressive Sensing (hard thresholding onto the sparse set)



Constrained Optimization

$$\min_{\mathcal{X}_{\pi_\theta} \in \mathfrak{S}} \left\| \boldsymbol{z} - \mathcal{A}\left(\mathcal{X}_{\pi_\theta}\right) \right\|$$

- We propose an iterative projection gradient descent scheme to solve the OP
- After each gradient step, we project onto the model set $\mathfrak{S}$
- Akin to Compressive Sensing (hard thresholding onto the sparse set)



3. Results

Compressive ICA RIP in Practice

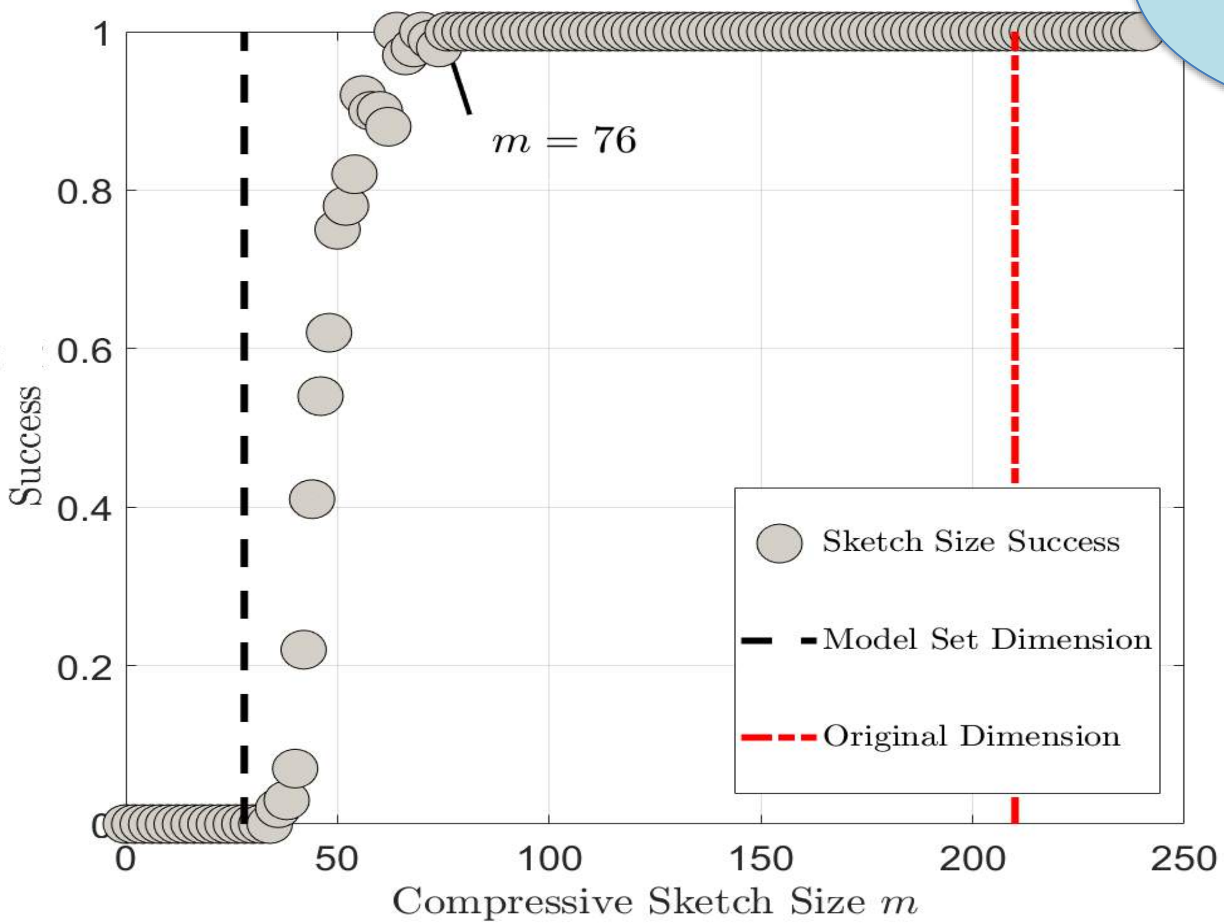
Compressive ICA RIP:
$m \approx \mathcal{O}\big(d(d+1)\big)$

Experiment 1

- We mix the cumulant tensor of $d = 7$ sources signals by a known mixing matrix $\boldsymbol{Q}$ to construct $\mathcal{X}_{\pi_\theta}$.
- For varying sketch sizes m, we compute the sketch and obtain the estimate $\widehat{\boldsymbol{Q}}$ using the IPG algorithm.
- Consider the estimation successful if $\mathrm{D}\big(\boldsymbol{Q}, \widehat{\boldsymbol{Q}}\big) \leq 10^{-7}$, where D is the Amari error/distance.

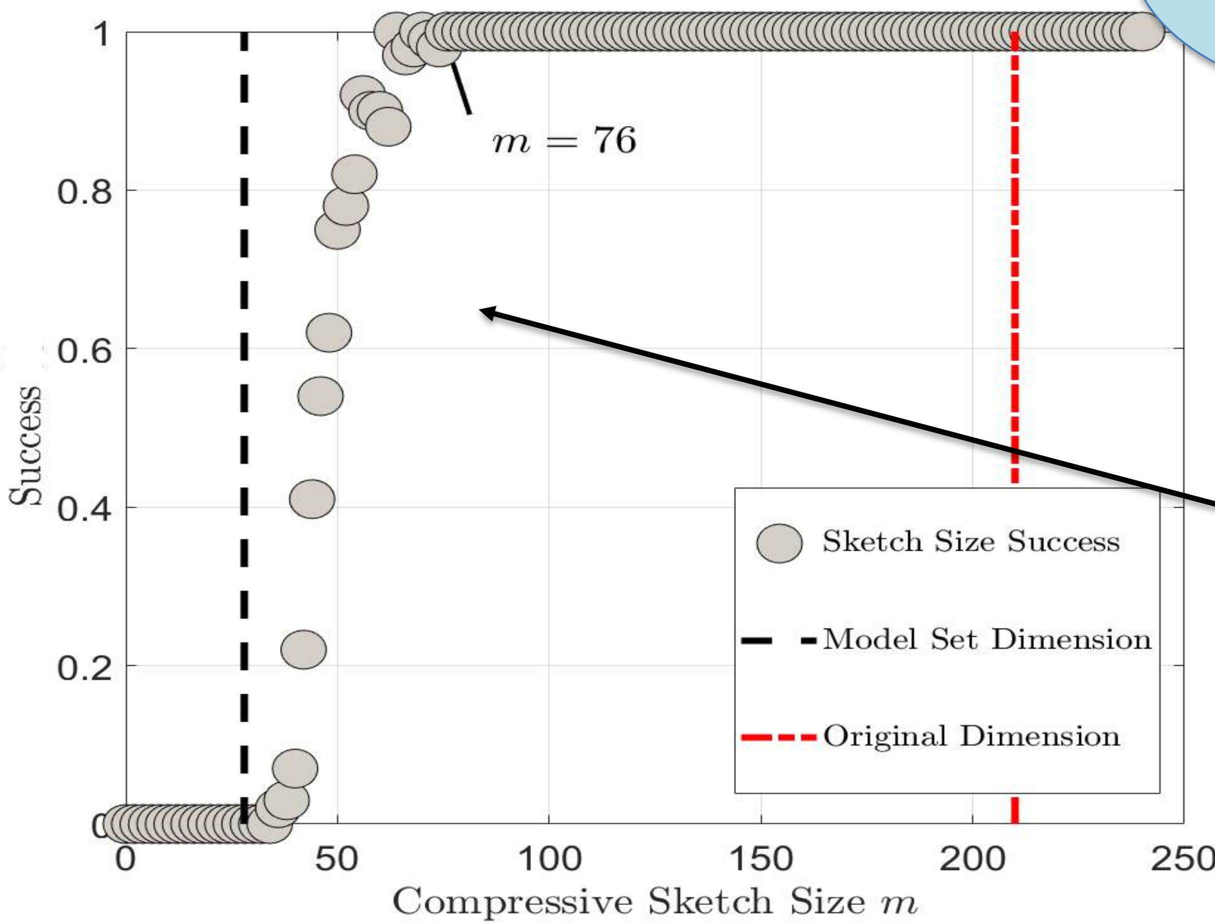
Compressive ICA RIP in Practice

Compressive ICA RIP:
$m \approx \mathcal{O}\big(d(d+1)\big)$



Compressive ICA RIP in Practice

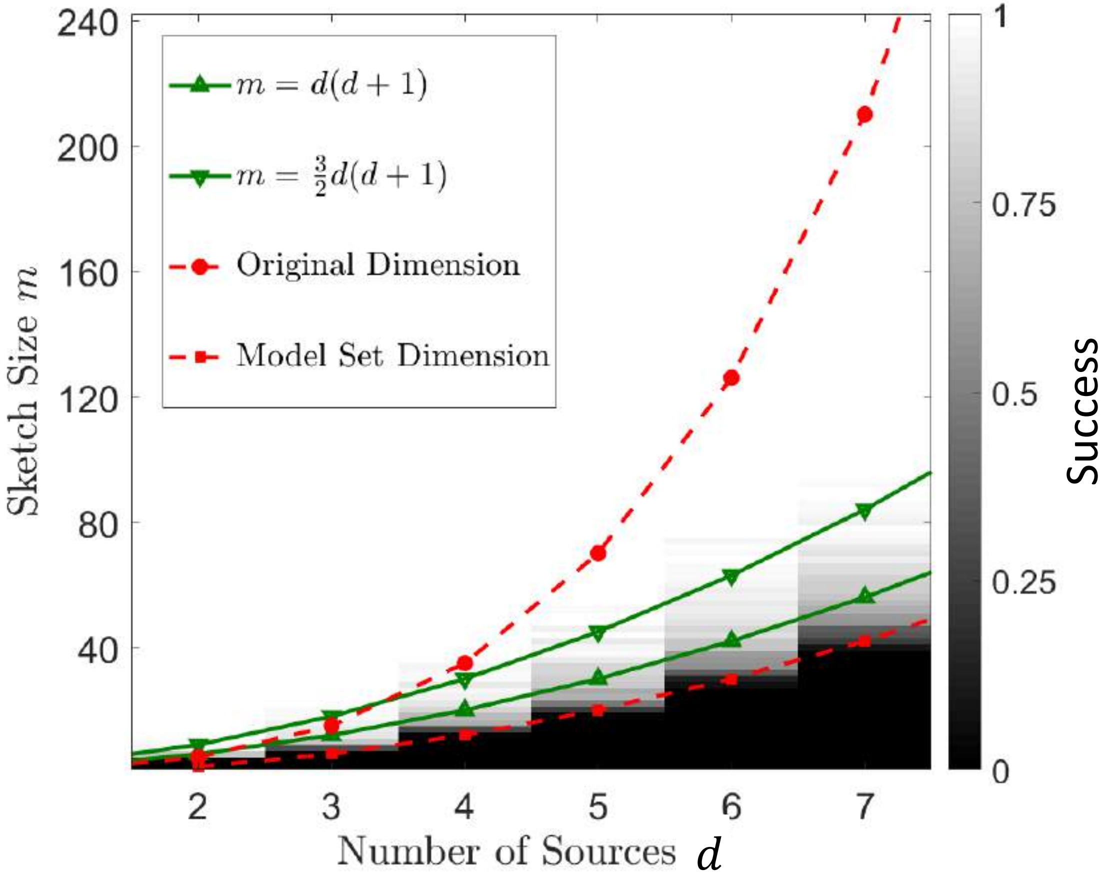
Compressive ICA RIP:
$m \approx \mathcal{O}\big(d(d+1)\big)$



Phase transition

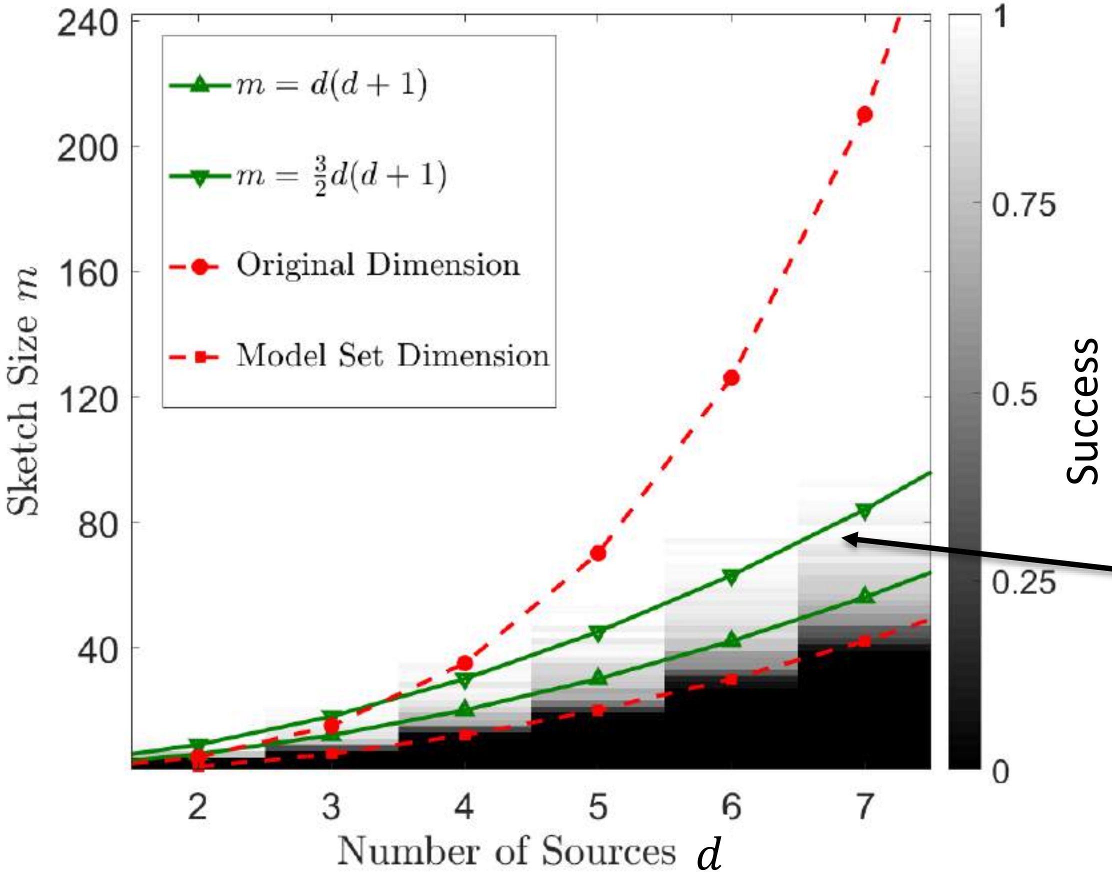
Compressive ICA RIP in Practice

Compressive ICA RIP:
$m \approx \mathcal{O}\big(d(d+1)\big)$



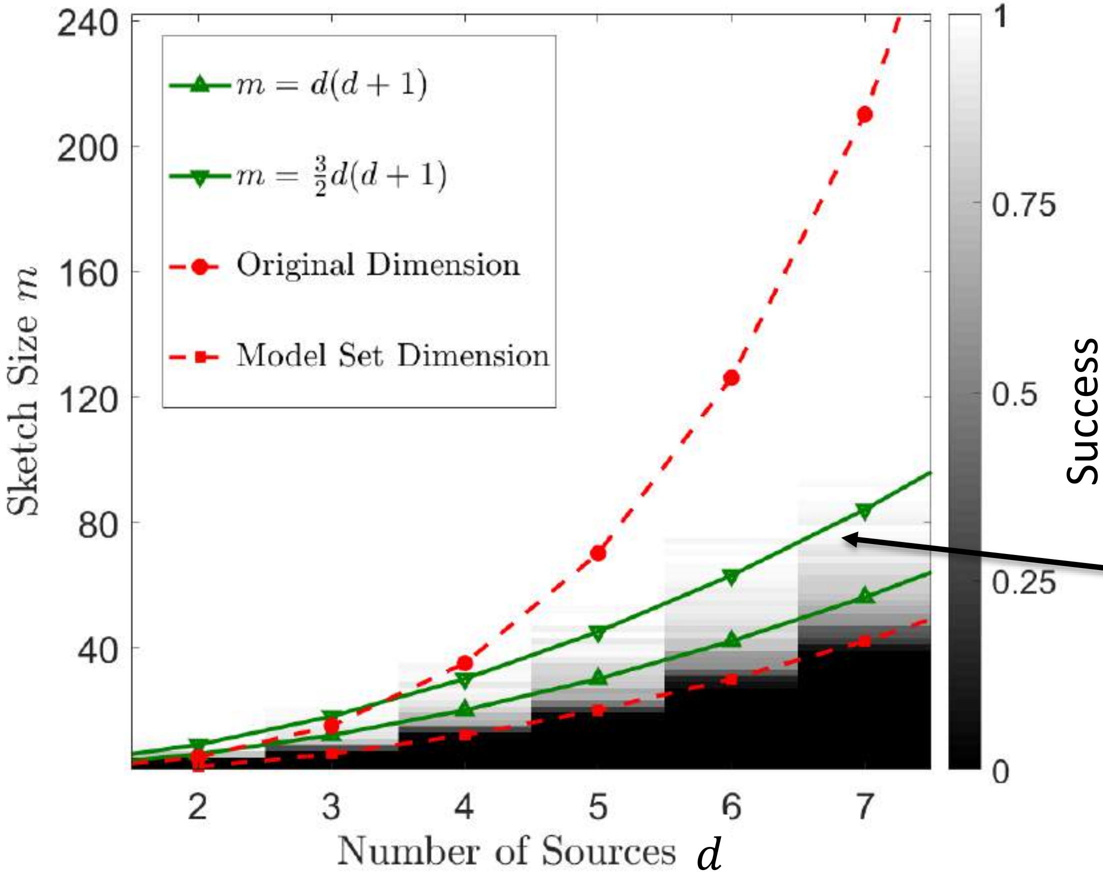
Compressive ICA RIP in Practice

Compressive ICA RIP:
$m \approx \mathcal{O}\big(d(d+1)\big)$



Phase transition occurs around $m = 2d(d+1)$

Compressive ICA RIP in Practice



Compressive ICA RIP:
$m \approx \mathcal{O}\big(d(d+1)\big)$

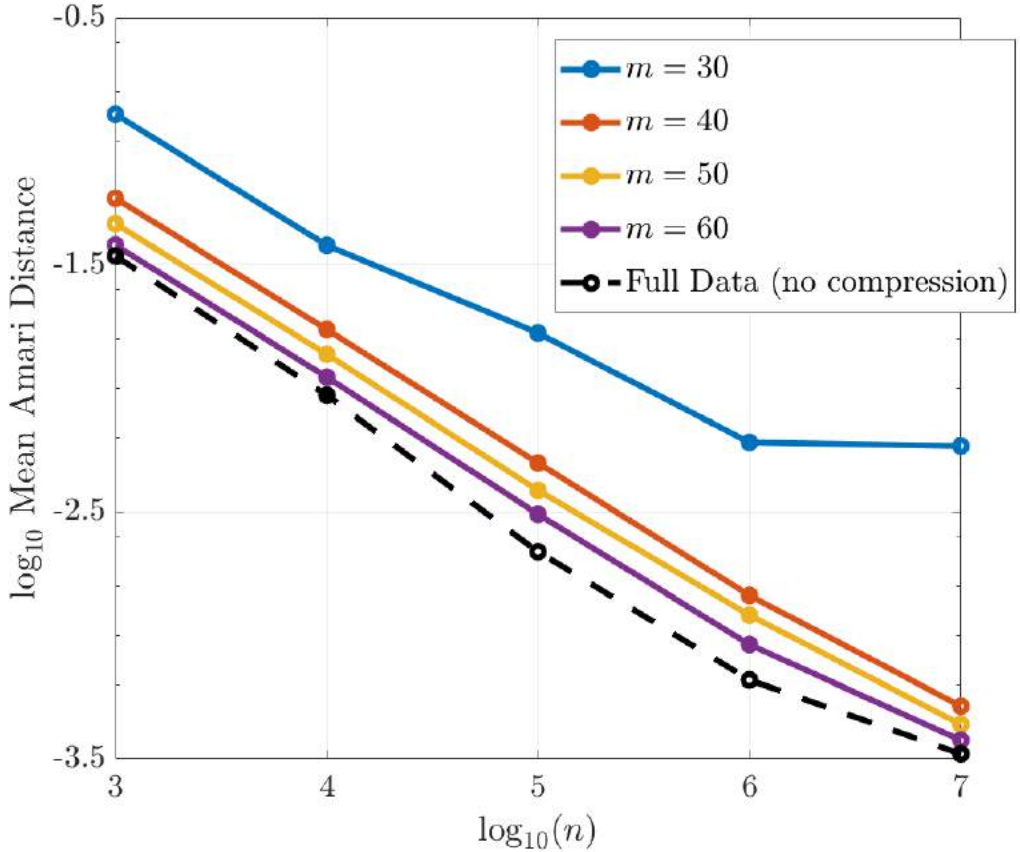
In practice the order is ≈ 2.5

Phase transition occurs around $m = 2d(d+1)$

How efficient are sketches?

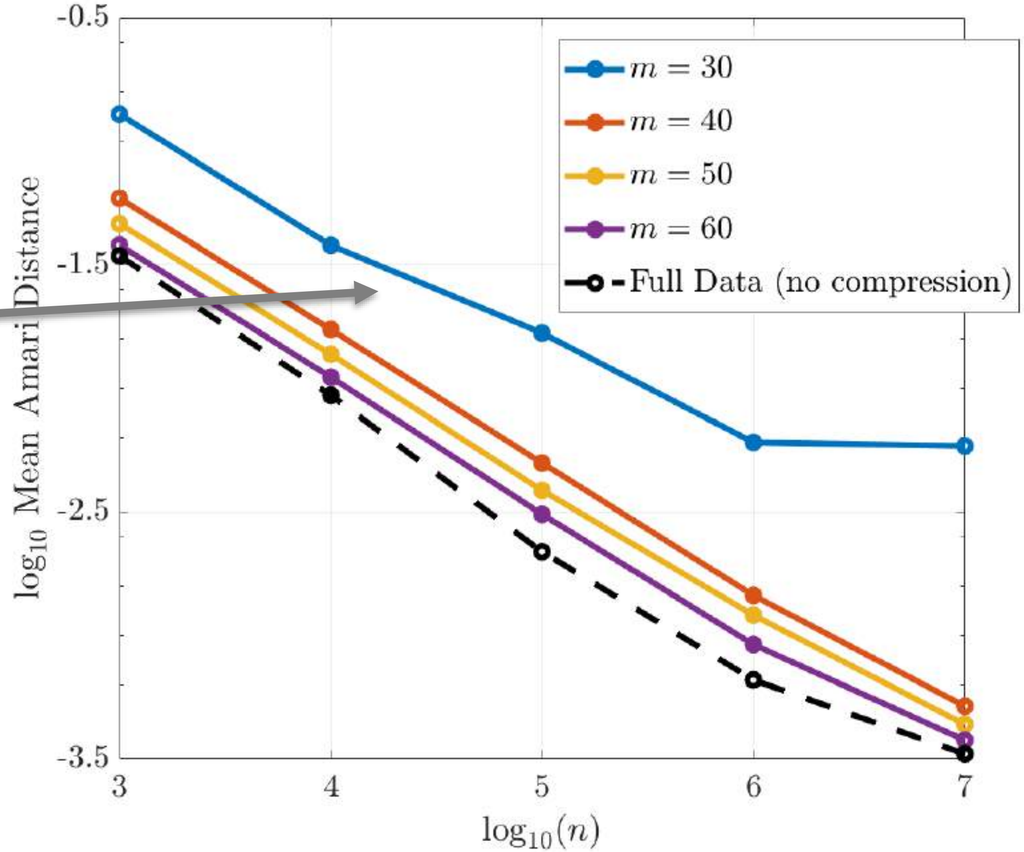
<u>Experiment 2</u>

- $d = 7$ source signals mixed by $\boldsymbol{Q}$
- Measure the average Amari error $\mathrm{D}(\boldsymbol{Q}, \widehat{\boldsymbol{Q}})$ over 1000 trials
- Plotted as function of the number of data points n



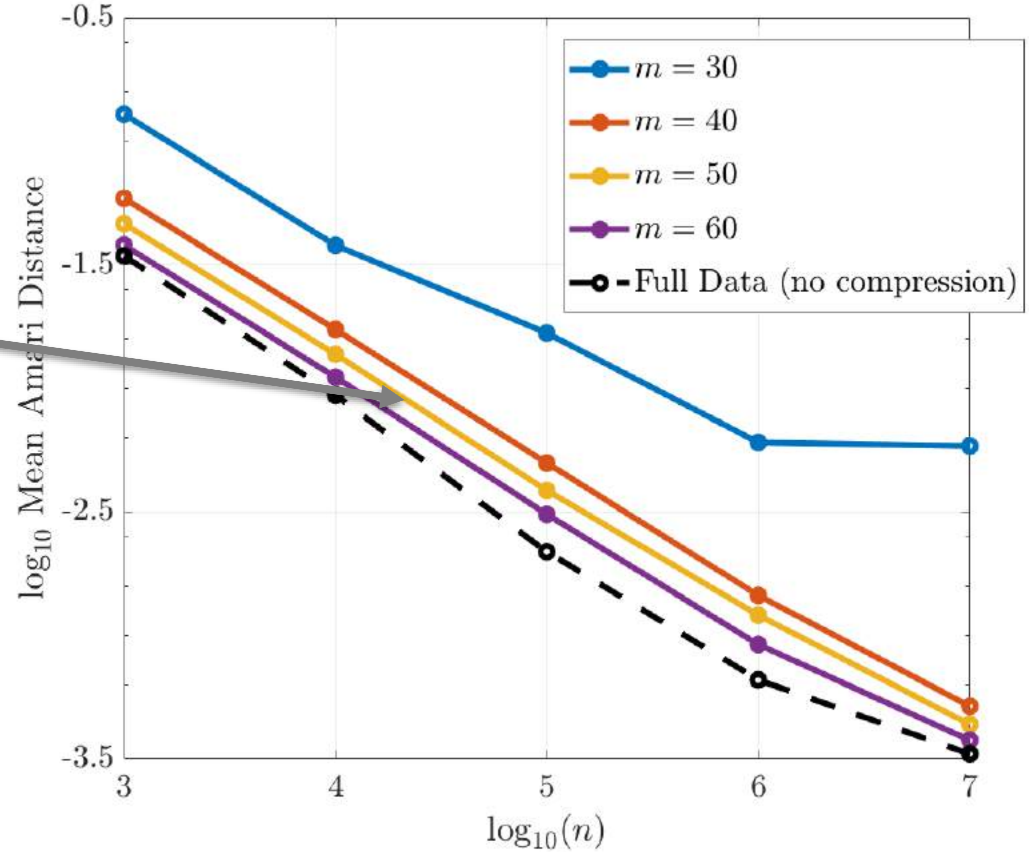
How efficient are sketches?

Sketch size is not sufficient so fails.

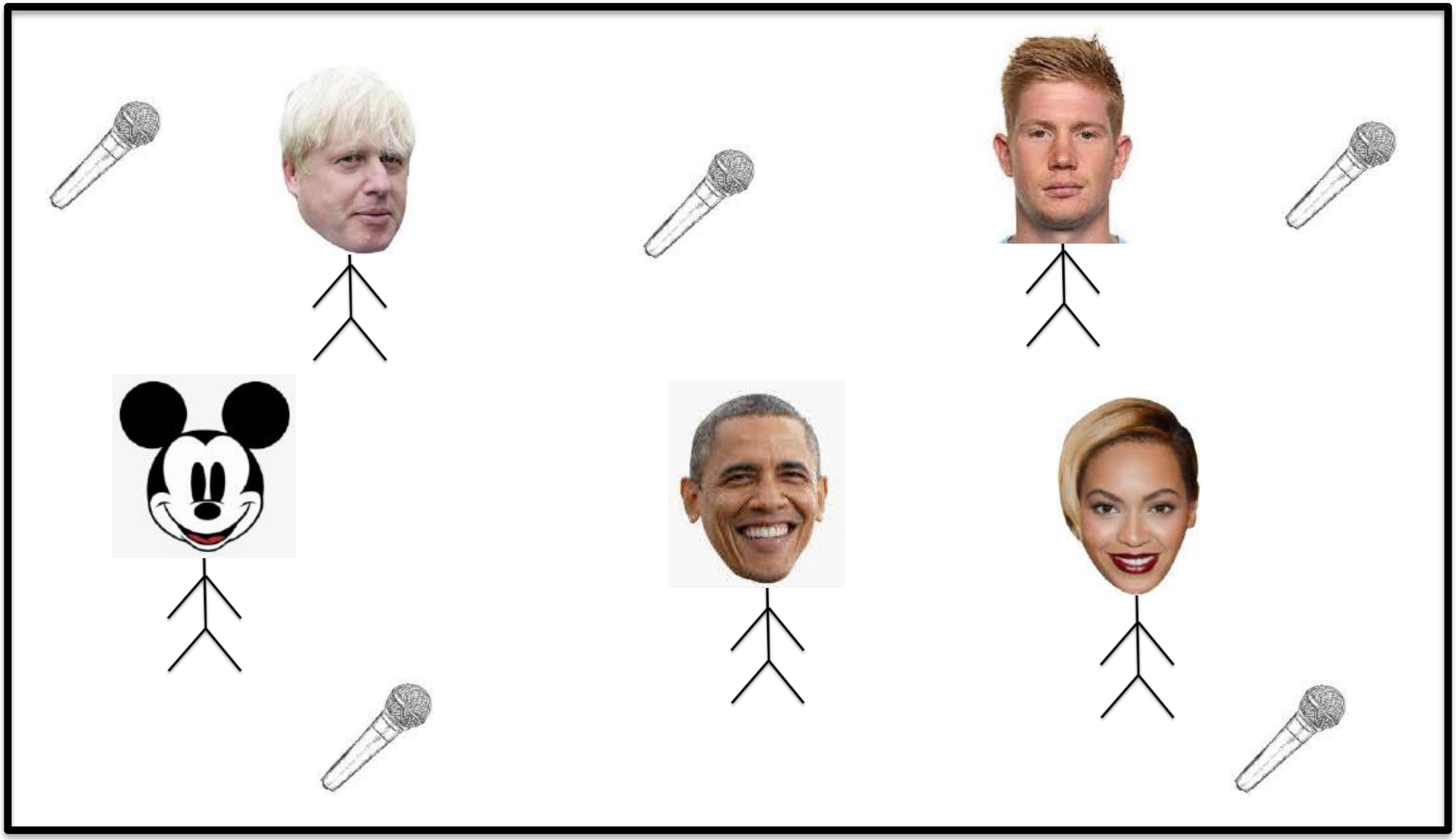


How efficient are sketches?

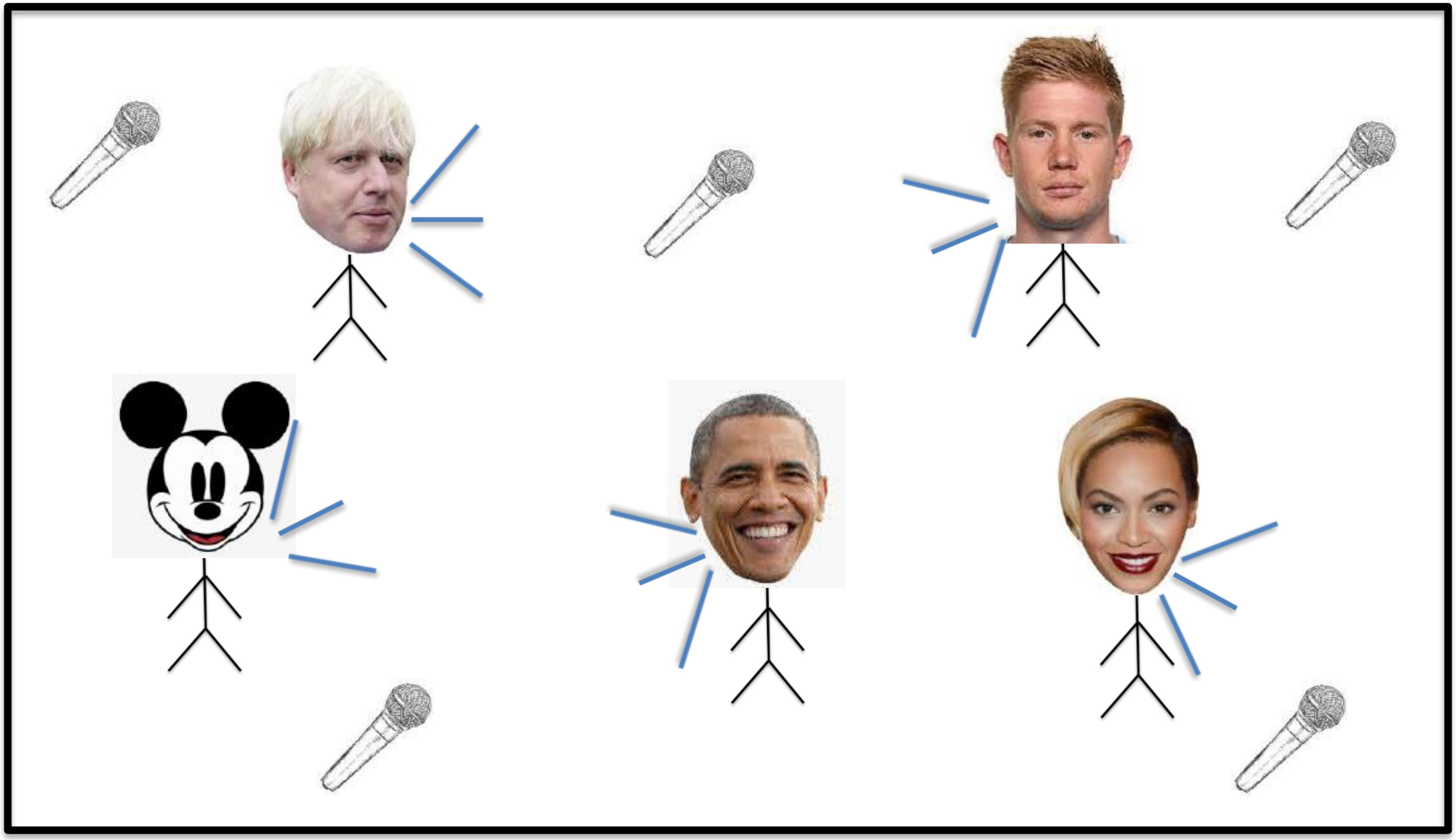
The mean Amari error of the sketch converges toward the full data error.



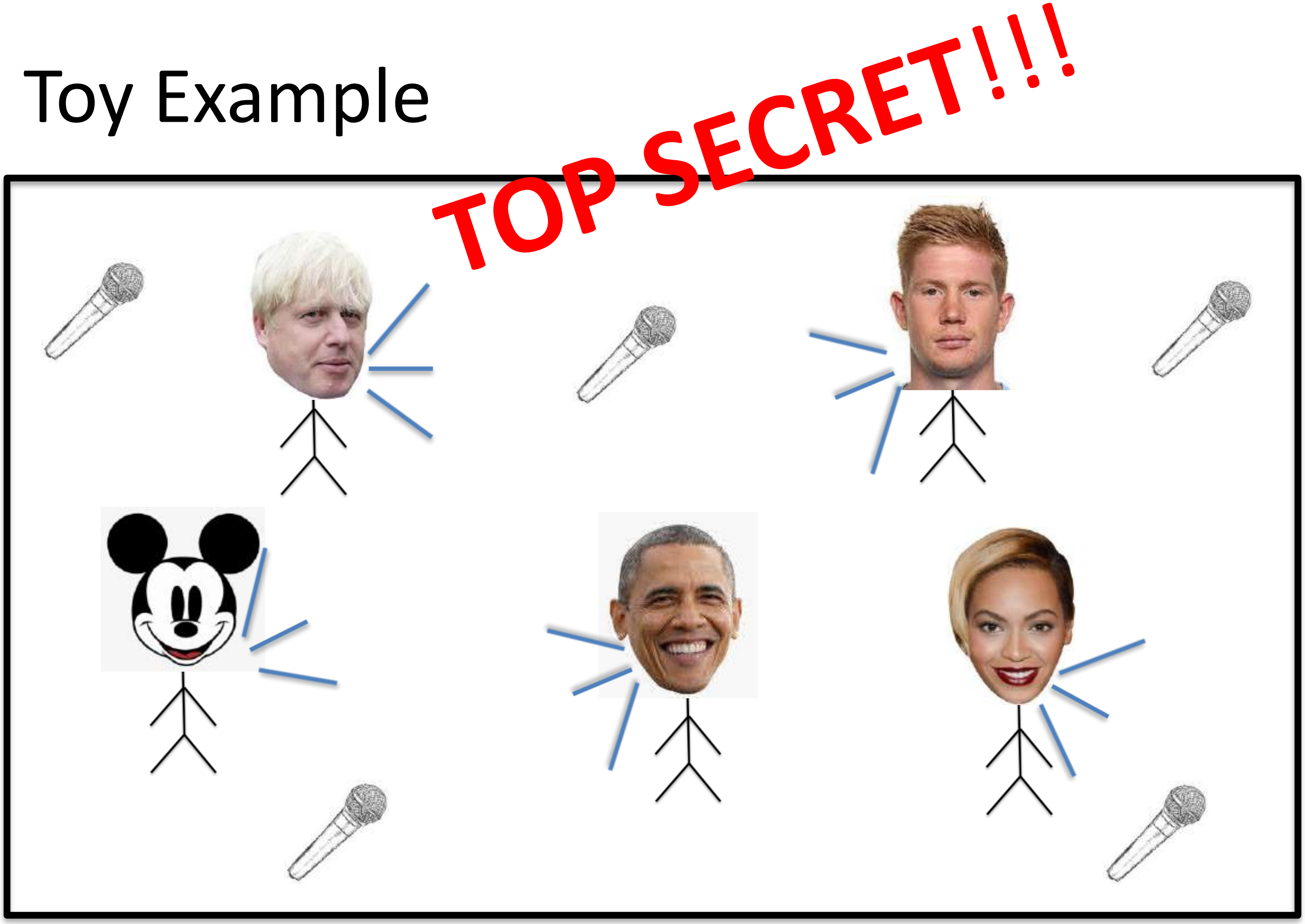
Toy Example



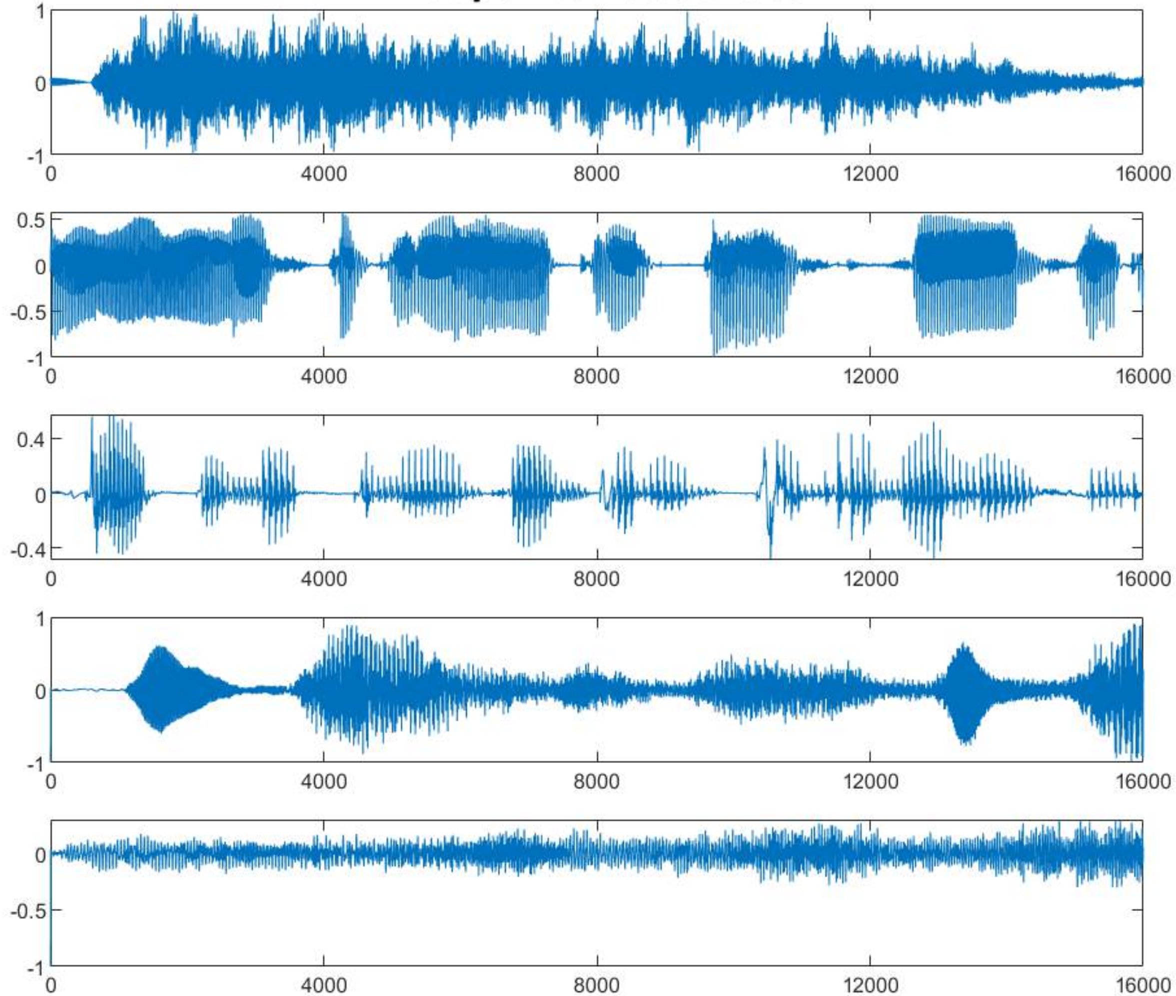
Toy Example



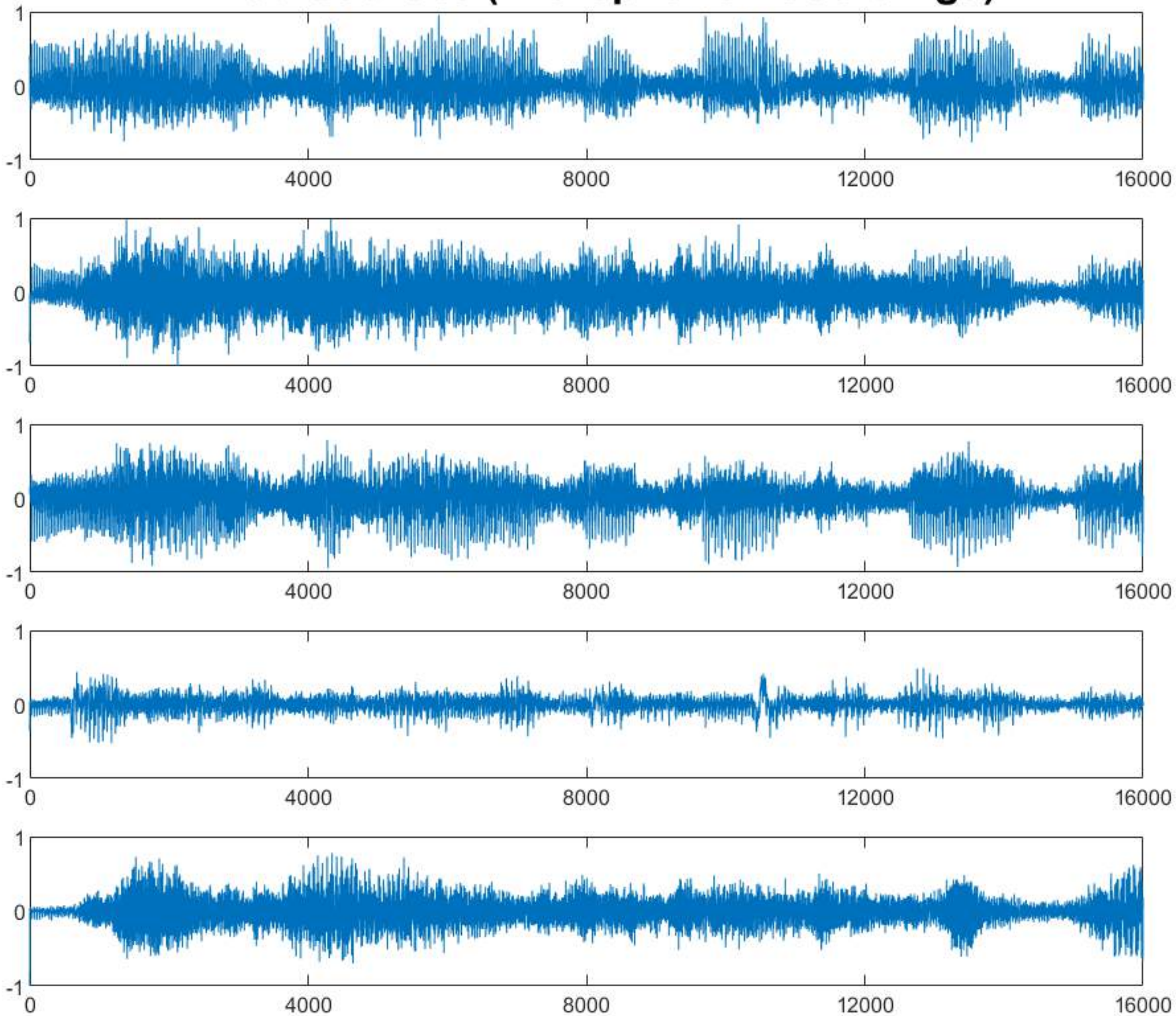
Toy Example

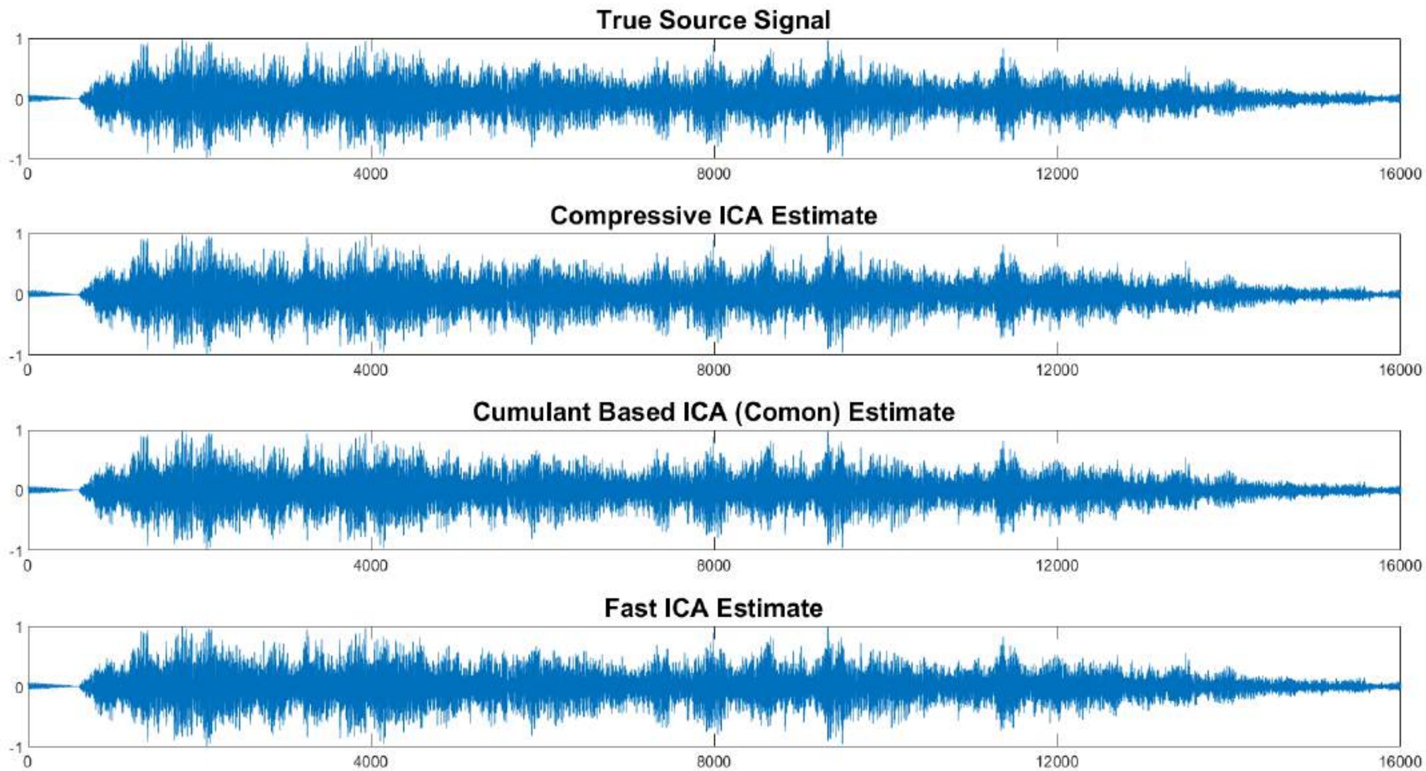


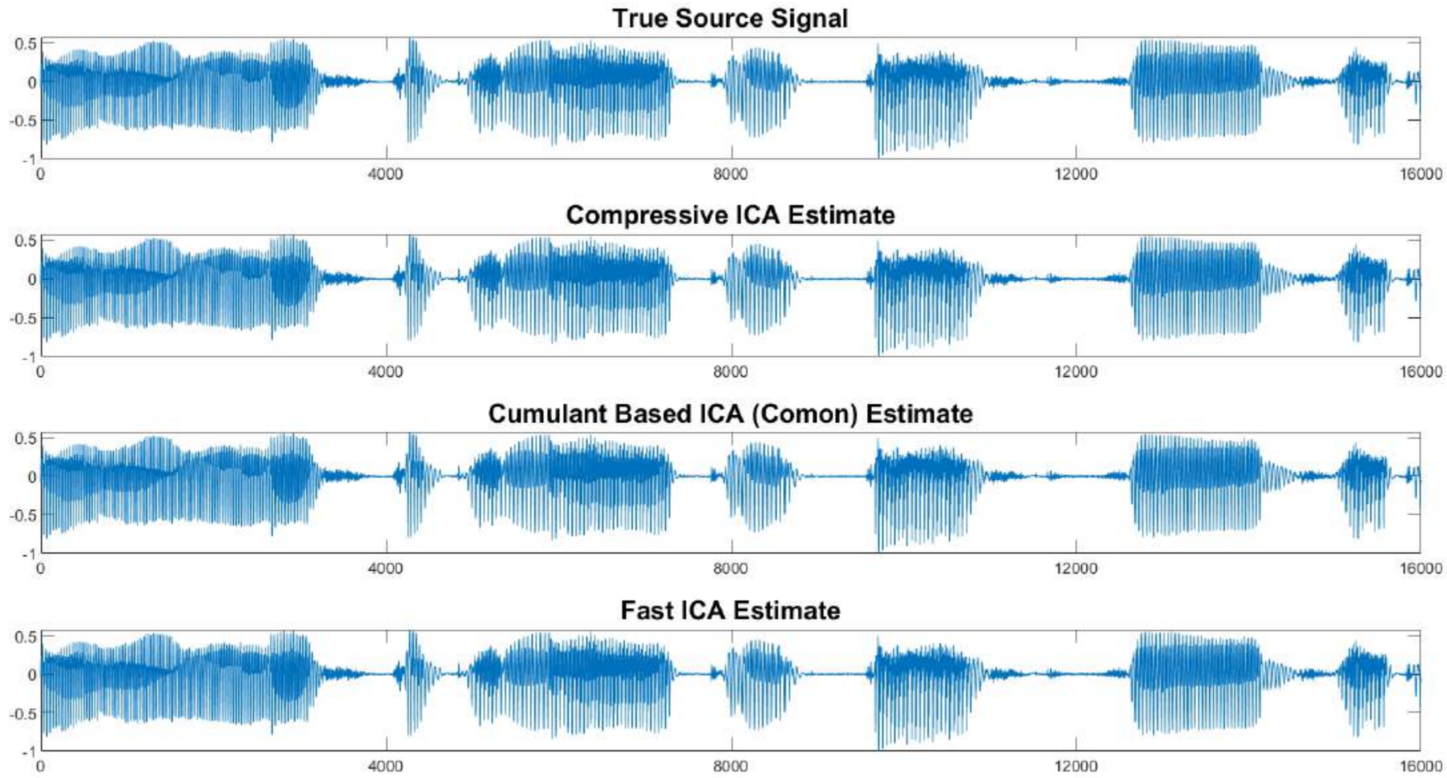
Independent Sources

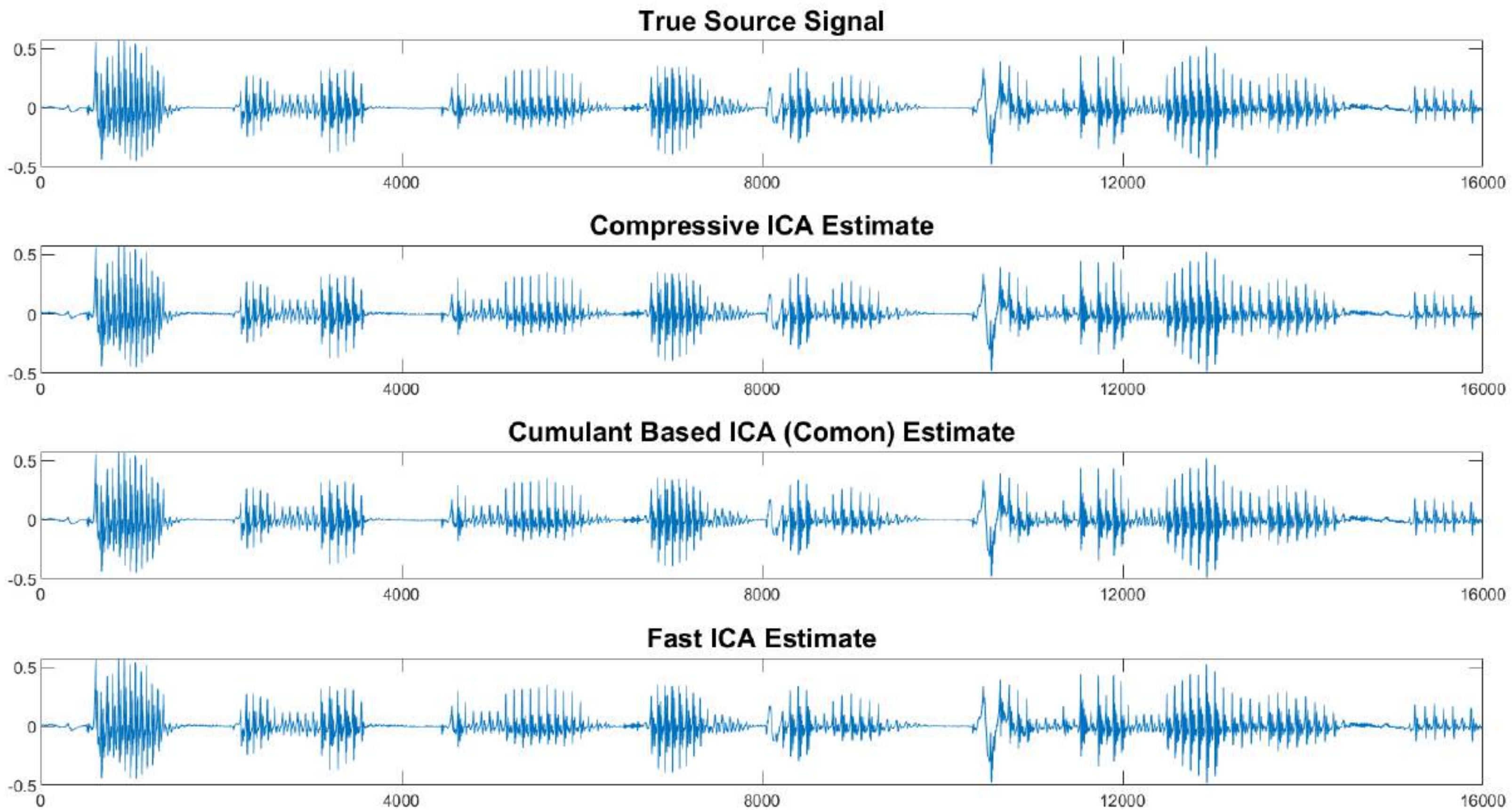


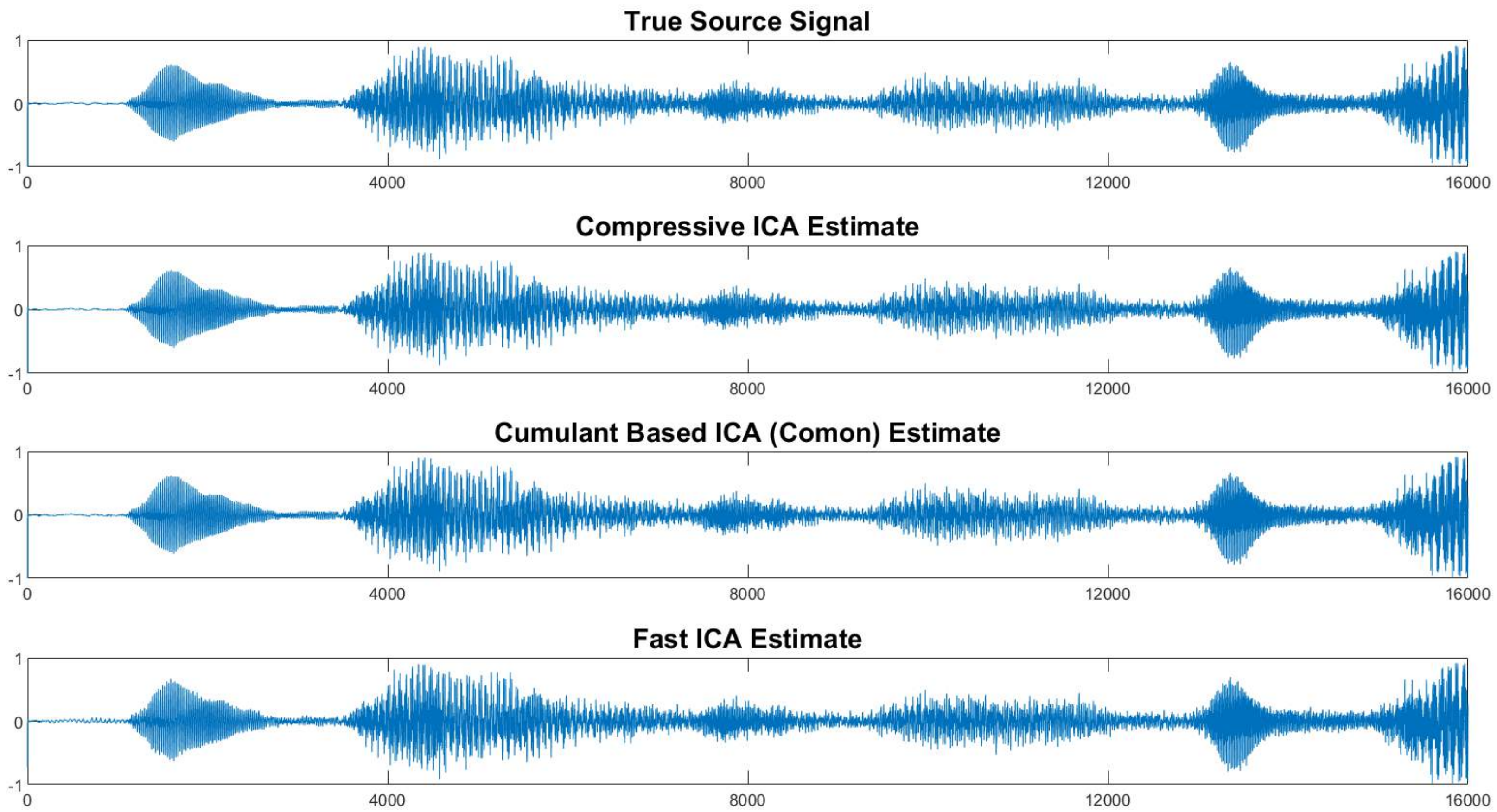
Mixed Sources (Microphone Recordings)

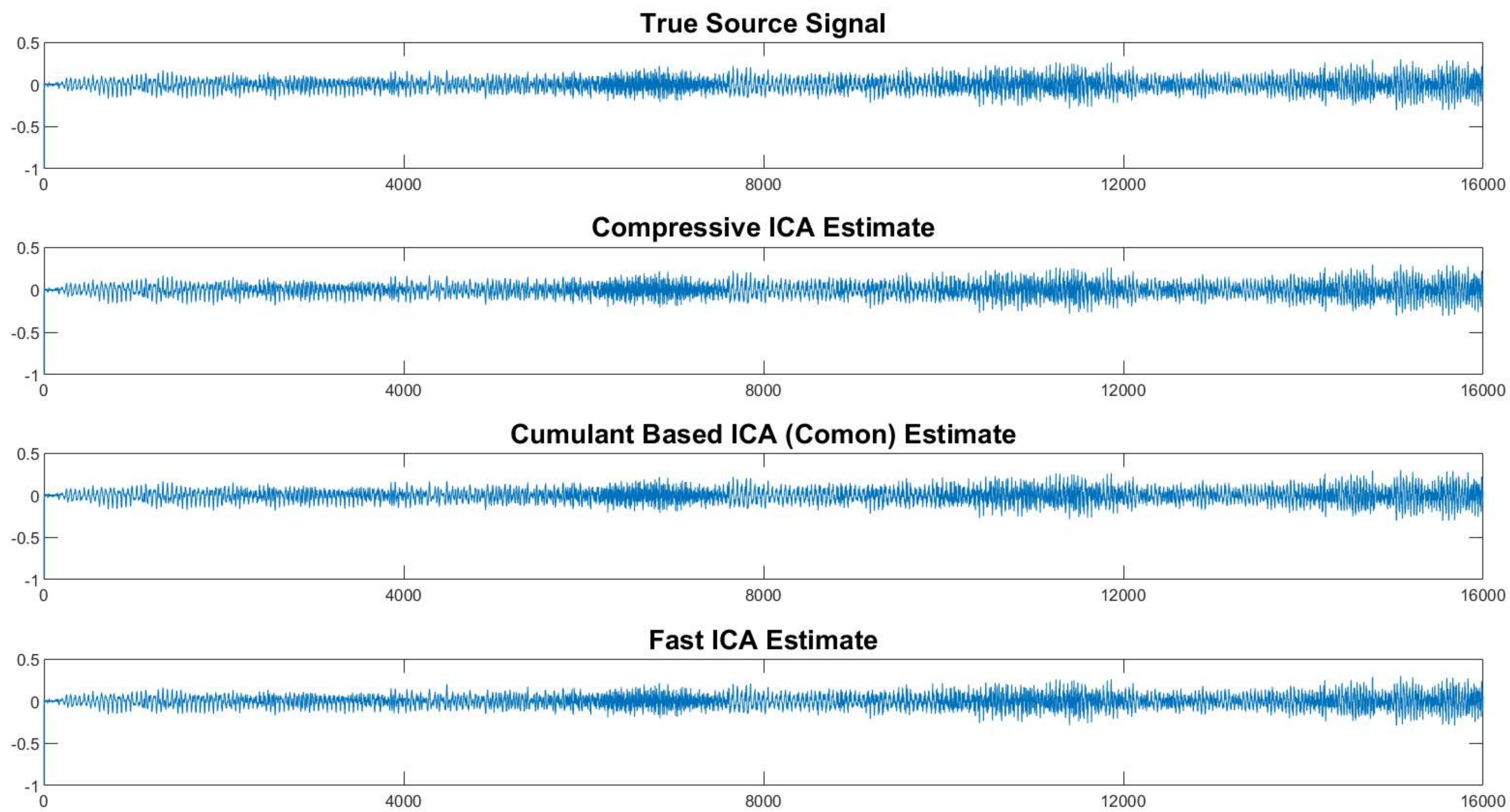


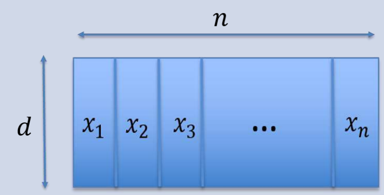
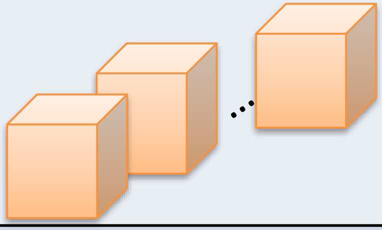
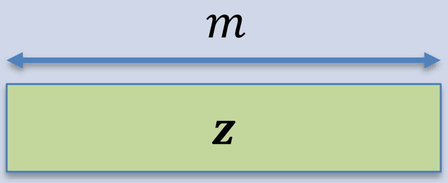

True Source Signal
1
0
-1
0
4000
8000
12000
16000
Compressive ICA Estimate
1
0
-1
0
4000
8000
12000
16000
Cumulant Based ICA (Comon) Estimate
1
0
-1
0
4000
8000
12000
16000
Fast ICA Estimate
1
0
-1
0
4000
8000
12000
16000


True Source Signal
Compressive ICA Estimate
Cumulant Based ICA (Comon) Estimate
Fast ICA Estimate
0.5
0
-0.5
-1
0
4000
8000
12000
16000


True Source Signal
Compressive ICA Estimate
Cumulant Based ICA (Comon) Estimate
Fast ICA Estimate
0.5
0
-0.5
0
4000
8000
12000
16000


True Source Signal
1
0
-1
0
4000
8000
12000
16000
Compressive ICA Estimate
1
0
-1
0
4000
8000
12000
16000
Cumulant Based ICA (Comon) Estimate
1
0
-1
0
4000
8000
12000
16000
Fast ICA Estimate
1
0
-1
0
4000
8000
12000
16000


True Source Signal
Compressive ICA Estimate
Cumulant Based ICA (Comon) Estimate
Fast ICA Estimate
0.5
0
-0.5
-1
0
4000
8000
12000
16000

Method	Amari Error	Space Complexity
Fast ICA (recourse to data)	**0.4087**	5×16000 Data matrix 
Cumulant based ICA (no Compression)	0.4129	$5\times 5 \times 5 \times 5$ cumulant Tensor (70 DoF) 
Compressive ICA	0.4156	$m = 38$ size sketch 

Limitations

- In general it is difficult to find closed form projections onto model sets

- Here we use a proxy projection, where we first *partially* diagonalise the cumulant tensor using existing techniques and then threshold the cross cumulants to zero.

- As a result, the computational complexity is equivalent to other cumulant based method

Summary

- We have shown that a low dimensional model set exists in the space of cumulant tensors for the ICA problem
- As a result, we can form sketches that are of the order of the model set to estimate the parameters of the ICA model
- The memory complexity is reduced from $\mathcal{O}(d^4)$ to $\mathcal{O}(d(d+1))$

Outlook

- Seek a cheaper projection operator or proxy that exhibits a computational complexity that scales with m
- Quantify theoretically the controlled loss of information/efficiency of taking a sketch of size m
- Can we leverage other sufficient statistics to produce sketches from when the distribution, like ICA, is left unspecified?

Thank you for your attention!

Any questions?