The Information-Theoretic Requirements of Subspace Clustering with Missing Data

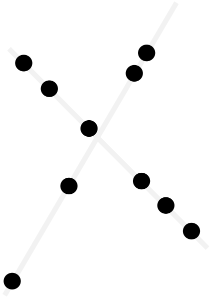
Daniel L. Pimentel-Alarcón
Robert Nowak
University of Wisconsin-Madison

ICML 2016

Subspace Clustering

- **We are given:** Columns in a Union of Subspaces.

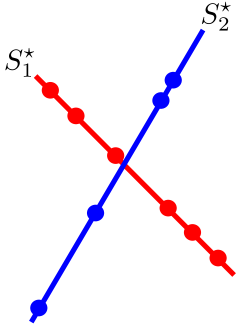
$$\begin{bmatrix} 1 & 4 & 1 & 3 & 3 & 1 & 2 & 1 & 2 & 1 \\ 2 & 4 & 2 & 6 & 3 & 2 & 2 & 2 & 4 & 1 \\ 3 & 4 & 3 & 9 & 3 & 3 & 2 & 3 & 6 & 1 \\ 1 & 8 & 1 & 3 & 6 & 1 & 4 & 1 & 2 & 2 \\ 2 & 8 & 2 & 6 & 6 & 2 & 4 & 2 & 4 & 2 \\ 3 & 8 & 3 & 9 & 6 & 3 & 4 & 3 & 6 & 2 \end{bmatrix}$$



Subspace Clustering

- **We are given:** Columns in a Union of Subspaces.
- **Goal:** Cluster the columns or find the subspaces.

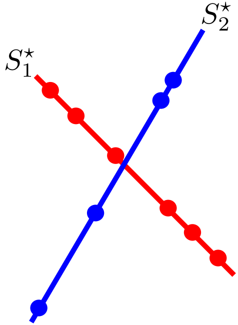
$$\begin{bmatrix} 1 & 4 & 1 & 3 & 3 & 1 & 2 & 1 & 2 & 1 \\ 2 & 4 & 2 & 6 & 3 & 2 & 2 & 2 & 4 & 1 \\ 3 & 4 & 3 & 9 & 3 & 3 & 2 & 3 & 6 & 1 \\ 1 & 8 & 1 & 3 & 6 & 1 & 4 & 1 & 2 & 2 \\ 2 & 8 & 2 & 6 & 6 & 2 & 4 & 2 & 4 & 2 \\ 3 & 8 & 3 & 9 & 6 & 3 & 4 & 3 & 6 & 2 \end{bmatrix}$$



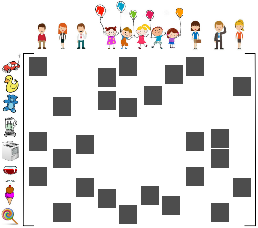
Subspace Clustering with Missing Data (SCMD)

- **We are given:** Incomplete columns in a Union of Subspaces.
- **Goal:** Cluster the columns or find the subspaces.

$$\begin{bmatrix} 1 & \cdot & \cdot & 3 & \cdot & 3 & \cdot & 1 & 2 & \cdot \\ 2 & \cdot & 2 & \cdot & \cdot & 6 & \cdot & \cdot & 4 & \cdot \\ \cdot & \cdot & 3 & \cdot & \cdot & 9 & \cdot & 3 & 6 & \cdot \\ 1 & \cdot & 1 & 3 & 6 & \cdot & 4 & 1 & 2 & 2 \\ \cdot & 8 & \cdot & \cdot & 6 & \cdot & 4 & \cdot & \cdot & \cdot \\ \cdot & 8 & \cdot & \cdot & \cdot & \cdot & 4 & \cdot & \cdot & 2 \end{bmatrix}$$



This arises in many Applications



First thing we need to ask:

In principle, what do we need to succeed?

First thing we need to ask:

In principle, what do we need to succeed?

To find out, let us look back at the full-data case.

Full-data case

- Suppose I have unlimited computational power
- Say I want to identify r-dimensional subspaces from **complete** columns:

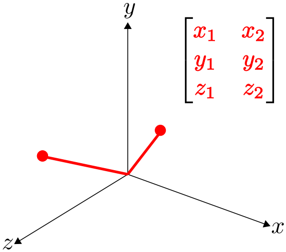
$$\mathbf{X} = \begin{bmatrix} 3 & 1 & 3 & 2 & 4 & 5 & 7 & 1 & 8 & 5 \\ 3 & 3 & 1 & 2 & 4 & 5 & 5 & 1 & 8 & 7 \\ 1 & 3 & 2 & 3 & 5 & 4 & 7 & 5 & 5 & 8 \\ 2 & 1 & 2 & 3 & 3 & 5 & 5 & 4 & 7 & 4 \end{bmatrix}$$

*

Full-data case

Key Idea:

- r columns define a *candidate* subspace.



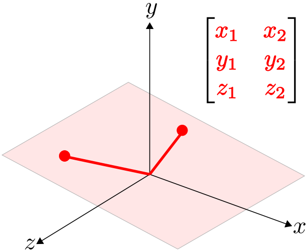
$$\begin{bmatrix} x_1 & x_2 \\ y_1 & y_2 \\ z_1 & z_2 \end{bmatrix}$$

*

Full-data case

Key Idea:

- r columns define a *candidate* subspace.

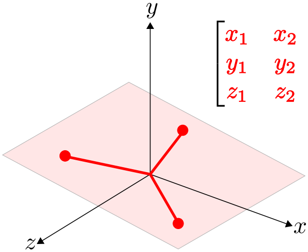


*

Full-data case

Key Idea:

- r columns define a *candidate* subspace.
- I can certify this subspace with an $(r+1)^{\text{th}}$ column.



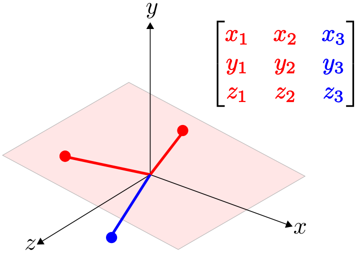
$$\begin{bmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{bmatrix}$$

*

Full-data case

Key Idea:

- r columns define a *candidate* subspace.
- I can certify this subspace with an $(r+1)^{\text{th}}$ column.



*

Full-data case

I can try all combinations of $r+1$ columns (here $r=2$).

$$\begin{bmatrix} 3 & 1 & 3 \\ 3 & 3 & 1 \\ 1 & 3 & 2 \\ 2 & 1 & 2 \end{bmatrix}$$

Linearly independent

$\Updownarrow$

Columns come from different subspaces.

$$\begin{bmatrix} 3 & 2 & 5 \\ 3 & 2 & 5 \\ 1 & 3 & 4 \\ 2 & 3 & 5 \end{bmatrix}$$

Linearly dependent

$\Updownarrow$

Columns come from the same subspace.

*

Full-data case

We can try combinations of $r+1$ columns until we identify all subspaces

$$S_1^\star = \text{span}\begin{bmatrix}1 & 3\\3 & 1\\3 & 2\\1 & 2\end{bmatrix} \qquad S_2^\star = \text{span}\begin{bmatrix}3 & 2\\3 & 2\\1 & 3\\2 & 3\end{bmatrix}$$

Then we can trivially cluster the columns.

$$\begin{bmatrix}3 & 1 & 3 & 2 & 4 & 5 & 7 & 1 & 8 & 5\\3 & 3 & 1 & 2 & 4 & 5 & 5 & 1 & 8 & 7\\1 & 3 & 2 & 3 & 5 & 4 & 7 & 5 & 5 & 8\\2 & 1 & 2 & 3 & 3 & 5 & 5 & 4 & 7 & 4\end{bmatrix}$$

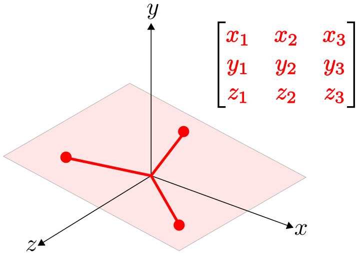
Full-data case

Key idea:

$r+1$ **complete** columns *fit* in an r-dimensional subspace

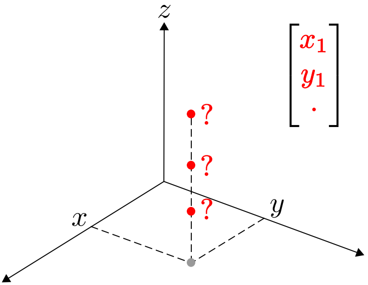
$\Updownarrow$

Columns come from the same subspace.



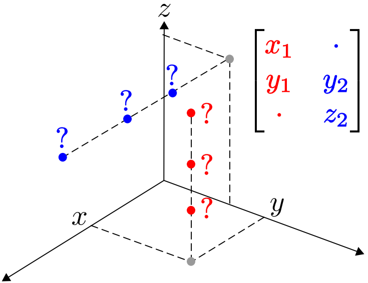
What changes with missing data?

- We don't know where points really are!
- Say I give you a point *without* the z coordinate.



What changes with missing data?

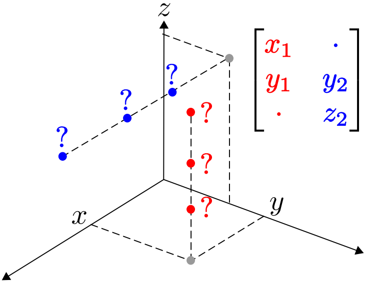
- We don't know where points really are!
- Say I give you a point *without* the z coordinate.
- Say I give you an other point *without* the x coordinate.



$$\begin{bmatrix} x_1 & \cdot \\ y_1 & y_2 \\ \cdot & z_2 \end{bmatrix}$$

What changes with missing data?

- We don't know where points really are!
- Say I give you a point *without* the z coordinate.
- Say I give you an other point *without* the x coordinate.



$$\begin{bmatrix} x_1 & \cdot \\ y_1 & y_2 \\ \cdot & z_2 \end{bmatrix}$$

- Is there only one subspace that agrees with these columns?

What changes with missing data?

- $\exists$ **False** subspaces that can fit arbitrarily many **incomplete** columns from different subspaces.

$$\begin{bmatrix} 1 & \cdot & \cdot & 3 & \cdot & 3 & \cdot & 1 & 2 & \cdot \\ 2 & \cdot & 2 & \cdot & \cdot & 6 & \cdot & \cdot & 4 & \cdot \\ \cdot & \cdot & 3 & \cdot & \cdot & 9 & \cdot & 3 & 6 & \cdot \\ 1 & \cdot & 1 & 3 & 6 & \cdot & 4 & 1 & 2 & 2 \\ \cdot & 8 & \cdot & \cdot & 6 & \cdot & 4 & \cdot & \cdot & \cdot \\ \cdot & 8 & \cdot & \cdot & \cdot & \cdot & 4 & \cdot & \cdot & 2 \end{bmatrix} \subset \text{span} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

What changes with missing data?

- So even with **unlimited** computational power:
  - We could run into **false** subspaces!
  - And get a wrong clustering!
- How can we guarantee that this won't happen?
- We need to make sure that columns are observed in the right places.

What do I mean observed in the right places?

We say $\mathbf{X}_{\mathbf{\Omega}}$ is observed in the right places if every matrix $\mathbf{X}'_{\mathbf{\Omega}'}$ formed with a *proper* subset of the columns in $\mathbf{X}_{\mathbf{\Omega}}$ satisfies

$$\#RowsWithObservations(\mathbf{X}'_{\mathbf{\Omega}'}) \;\geq\; \#Columns(\mathbf{X}'_{\mathbf{\Omega}'}) + r.$$

What do I mean observed in the right places?

We say $\mathbf{X}_{\mathbf{\Omega}}$ is observed in the right places if every matrix $\mathbf{X}'_{\mathbf{\Omega}'}$ formed with a *proper* subset of the columns in $\mathbf{X}_{\mathbf{\Omega}}$ satisfies

$$\#RowsWithObservations(\mathbf{X}'_{\mathbf{\Omega}'}) \;\geq\; \#Columns(\mathbf{X}'_{\mathbf{\Omega}'}) + r.$$

$$\mathbf{X}_{\mathbf{\Omega}} = \begin{bmatrix} 1 & \cdot & 3 & \cdot \\ 1 & 2 & \cdot & \cdot \\ \cdot & 2 & 3 & \cdot \\ \cdot & \cdot & \cdot & 4 \\ \cdot & \cdot & \cdot & 4 \end{bmatrix} \qquad \mathbf{X}_{\mathbf{\Omega}} = \begin{bmatrix} 1 & \cdot & \cdot & \cdot \\ 1 & 2 & \cdot & \cdot \\ \cdot & 2 & 3 & \cdot \\ \cdot & \cdot & 3 & 4 \\ \cdot & \cdot & \cdot & 4 \end{bmatrix}$$

$$\underbrace{m(\mathbf{\Omega}')}_{3} \ngeq \underbrace{n(\mathbf{\Omega}')/r + r}_{4} \qquad \underbrace{m(\mathbf{\Omega}')}_{4} \geq \underbrace{n(\mathbf{\Omega}')/r + r}_{4}$$

How do we know if columns come from same subspace?

Key intuition:

> $d - r + 1$ **incomplete** columns (observed in the right places) *behave* as **one complete** column.

$$\underbrace{\begin{bmatrix} 4 & 1 & \cdot \\ 4 & 3 & 1 \\ \cdot & 3 & 2 \\ 3 & \cdot & 2 \end{bmatrix}}_{d-r+1} \quad \sim \quad \underbrace{\begin{bmatrix} 1 \\ 3 \\ 3 \\ 1 \end{bmatrix}}_{1}$$

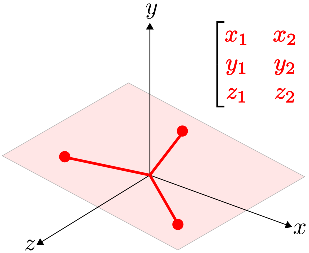
How do we know if columns come from same subspace?

Recall:

$r+1$ **complete** columns *fit* in an r-dimensional subspace

$\Updownarrow$

Columns come from the same subspace.



$$\begin{bmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{bmatrix}$$

How do we know if columns come from same subspace?

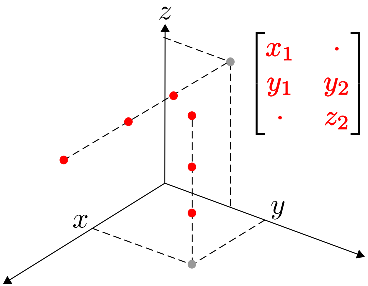
Analogously:

Theorem (P.-A., Nowak, ICML '16)

$r+1$ *sets of* $d-r+1$ *incomplete columns (observed in the right places)* fit *in an* r*-dimensional subspace*

$$\Updownarrow$$

Columns come from the same subspace.

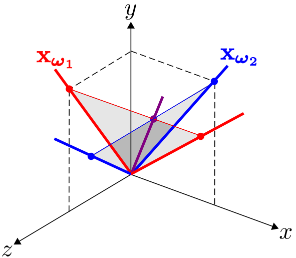


$$\begin{bmatrix} x_1 & \cdot \\ y_1 & y_2 \\ \cdot & z_2 \end{bmatrix}$$

Main Idea of the Proof

Each column with $r+1$ samples imposes one polynomial constraint on the subspaces that can *agree* with it.

$$\mathbf{X}_{\mathbf{\Omega}} = \begin{array}{c} \\ \end{array}\overset{\begin{array}{cc}\mathbf{x}_{\omega_1} & \mathbf{x}_{\omega_2}\end{array}}{\begin{bmatrix} \cdot & 2 \\ 1 & 2 \\ 1 & \cdot \end{bmatrix}}$$



- A subspace S *agrees* with $\mathbf{X}_{\mathbf{\Omega}}$ $\iff \begin{cases} f_1(S_{\omega_1}|\mathbf{x}_{\omega_1}) = 0 \\ f_2(S_{\omega_2}|\mathbf{x}_{\omega_2}) = 0 \end{cases}$.
- If our columns are observed in the right places, only the true subspaces will be consistent with the constraints.

Extend Linear Algebra Results to Missing Data

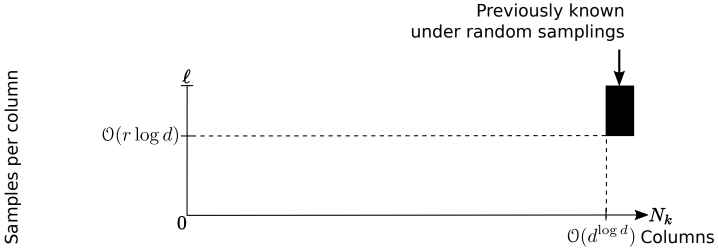
	Uniquely define a subspace	Certify/Discard subspaces
Full data	r	$r+1$
Missing data		

Extend Linear Algebra Results to Missing Data

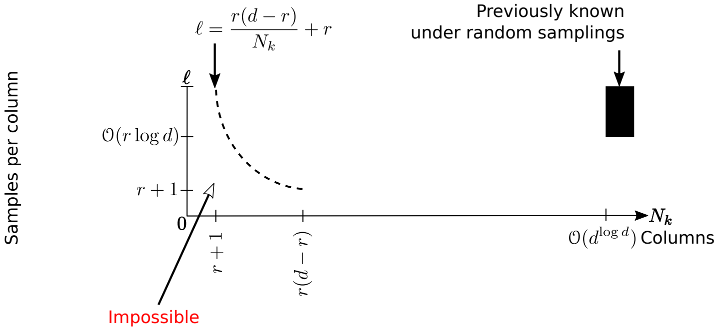
	Uniquely define a subspace	Certify/Discard subspaces
Full data	r	$r+1$
Missing data	$(r+1)(d-r)$ *	$(r+1)(d-r+1)$ *

* Observed in the right places.

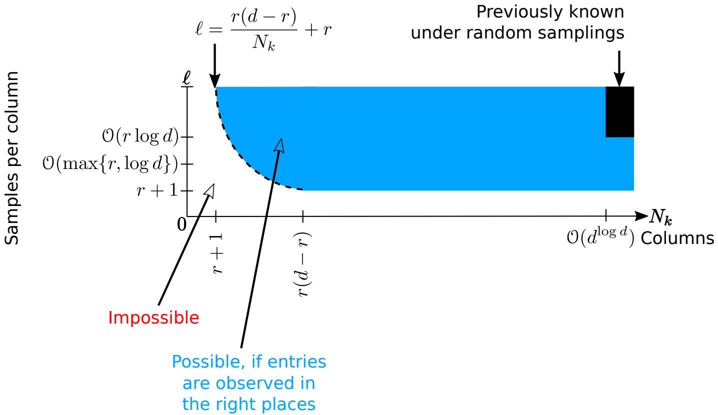
The Big Picture



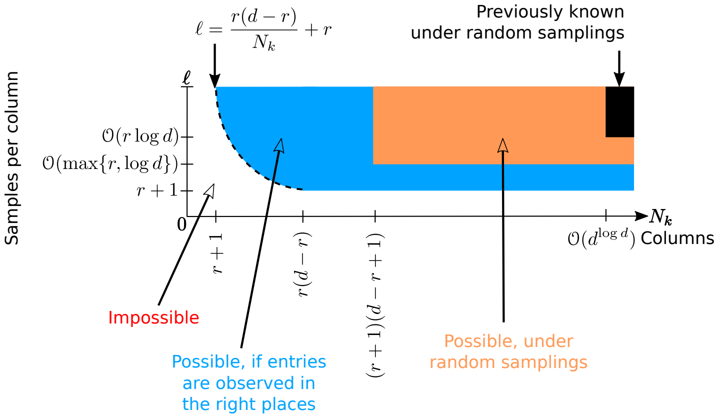
The Big Picture



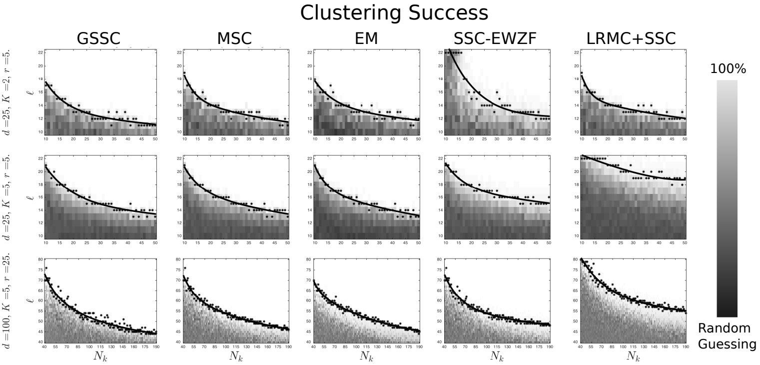
The Big Picture



The Big Picture



The Big Picture



How am I on time?

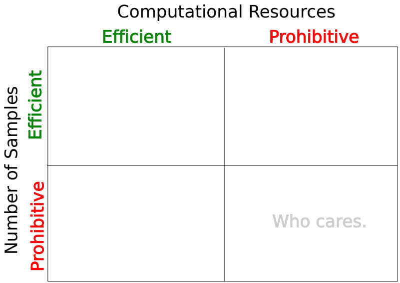
*

Now we know when we
should be able to succeed.

Now we know when we
should be able to succeed.

Now the question is: How?

Algorithms?


Computational Resources
Efficient
Prohibitive
Number of Samples
Efficient
Prohibitive
Who cares.

Algorithms?

Number of Samples \ Computational Resources	Efficient	Prohibitive
Efficient		
Prohibitive	HRMC	Who cares.

Algorithms?

Number of Samples \ Computational Resources	Efficient	Prohibitive
Efficient		$\mathcal{F} = \mathbf{0}$
Prohibitive	HRMC	Who cares.

Algorithms?

Computational Resources

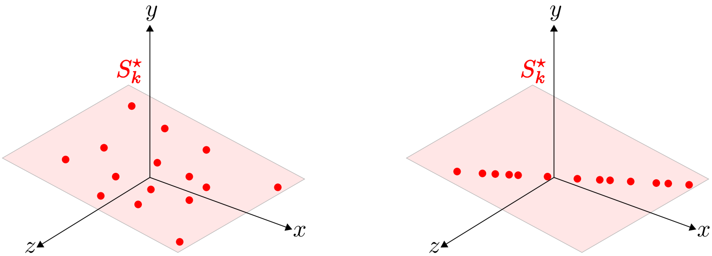
Number of Samples	Efficient	Prohibitive
Efficient	EM k-GROUSE SSC-EWZF LRMC+SSC	$\mathcal{F} = \mathbf{0}$
Prohibitive	HRMC	Who cares.

Algorithms?

Number of Samples		Computational Resources: Efficient	Computational Resources: Prohibitive
	Efficient	EM k-GROUSE SSC-EWZF LRMC+SSC ?	$\mathcal{F} = \mathbf{0}$
	Prohibitive	HRMC	Who cares.

Thanks.

What do I mean generic?



*

**