On Subspaces and Missing Data

Daniel L. Pimentel-Alarcón

University of Wisconsin-Madison

Joint work: Nigel Boston and Robert Nowak

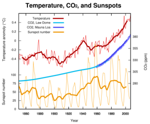
Applied Algebra Seminar, UC-Berkeley

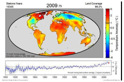
Introduction

We have lots of data

Introduction

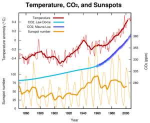
We have lots of data

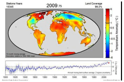




Introduction

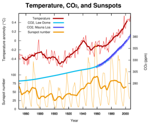
We have lots of data

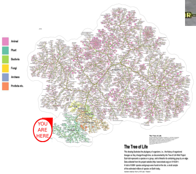


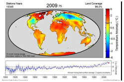


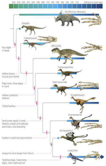
Introduction

We have lots of data



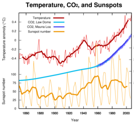


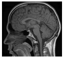


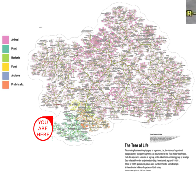


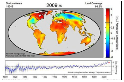
Introduction

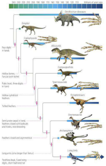
We have lots of data





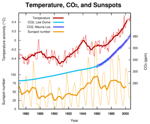


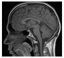


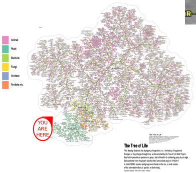


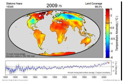
Introduction

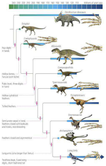
We have lots of data





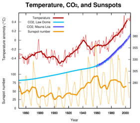


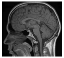


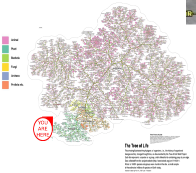


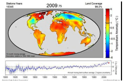
Introduction

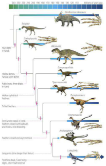
We have lots of data











Introduction

We have lots of data

And we want to analyze it.

Introduction

That's all very nice, but... often data is missing!

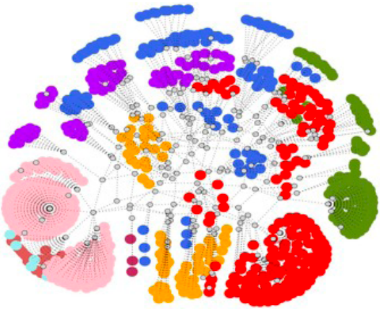
- Example: Vision.

Image: Hopkins 155 Dataset

Introduction

Often data is missing!

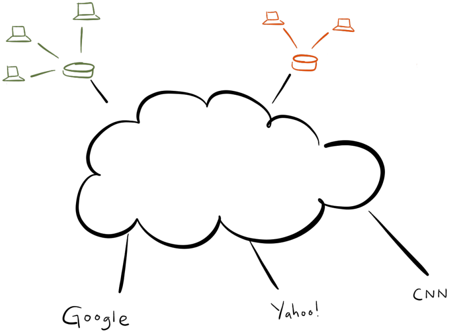
- Other example: Network topology estimation



Introduction

Often data is missing!

- Other example: Network topology estimation



Introduction

Often data is missing!

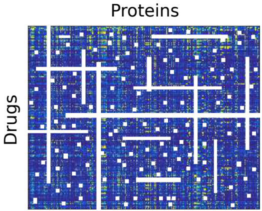
- Other example: Network topology estimation

$$\text{monitors}\left\{\vphantom{\begin{bmatrix}1\\2\\3\\4\\5\\6\end{bmatrix}}\right. \underbrace{\begin{bmatrix} 1 & \cdot & \cdot & 3 & \cdot & 3 & \cdot & 1 & 2 & \cdot \\ 2 & \cdot & 2 & \cdot & \cdot & 6 & \cdot & \cdot & 4 & \cdot \\ \cdot & \cdot & 3 & \cdot & \cdot & 9 & \cdot & 3 & 6 & \cdot \\ 1 & \cdot & 1 & 3 & 6 & \cdot & 4 & 1 & 2 & 2 \\ \cdot & 8 & \cdot & \cdot & 6 & \cdot & 4 & \cdot & \cdot & \cdot \\ \cdot & 8 & \cdot & \cdot & \cdot & \cdot & 4 & \cdot & \cdot & 2 \end{bmatrix}}_{\text{IP's}}$$

Introduction

Often data is missing!

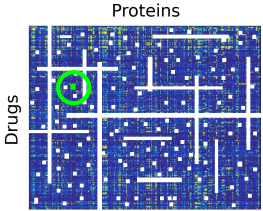
- Other example: Drug-target interactions



Introduction

Often data is missing!

- Other example: Drug-target interactions



Introduction

Introduction

What am I telling you?

Goal: Analyze
Incomplete Data

What am I telling you?

Goal: Analyze Incomplete Data

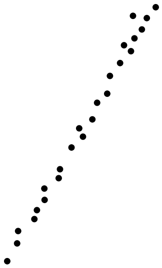
How? Finding Subspaces

How?

Data is often well-modeled by linear subspaces.

- Linear Algebra is one of our favorite tools.

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



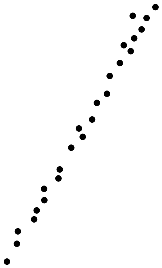
How?

Data is often well-modeled by linear subspaces.

- Linear Algebra is one of our favorite tools.
- We want to extend linear algebra to incomplete datasets.

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

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



What am I telling you?

Low-Rank Matrix Completion

Low-Rank Matrix Completion:

- Given a subset of entries in a rank-r matrix, exactly recover *all* of the missing entries.

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

Low-Rank Matrix Completion

Low-Rank Matrix Completion:

- Given a subset of entries in a rank-r matrix, exactly recover *all* of the missing entries.

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

- $\sim$ Identifying the subspace spanned by the columns, $S^\star$. Here

$$S^\star = \operatorname{span} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}.$$

Low-Rank Matrix Completion

Notation

- $\mathbf{\Omega}$ will indicate the observed entries:

$$\mathbf{X}_{\mathbf{\Omega}} = \begin{bmatrix} 1 & \cdot & 3 & \cdot \\ 1 & 2 & \cdot & \cdot \\ \cdot & 2 & 3 & \cdot \\ \cdot & \cdot & \cdot & 4 \\ \cdot & \cdot & \cdot & 4 \end{bmatrix} \qquad \mathbf{\Omega} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Low-Rank Matrix Completion

Existing theory (e.g. Candès and Recht, '09):

- Under some conditions on $\mathbf{\Omega}$ (e.g., uniform sampling):

If the columns of $\mathbf{X}$ lie in an r-dimensional subspace $S^\star$

$\Downarrow$

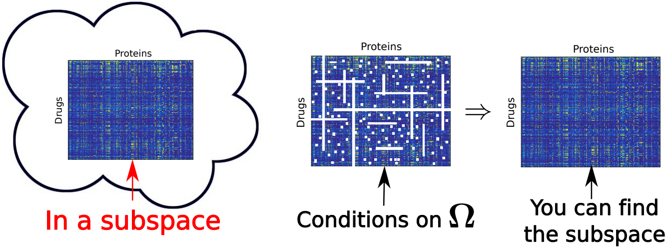
$S^\star$ is the only r-dimensional subspace that agrees with $\mathbf{X}_{\mathbf{\Omega}}$.

Low-Rank Matrix Completion

Existing theory (e.g. Candès and Recht, '09):

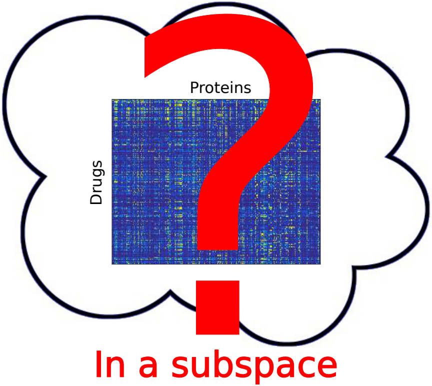
- Under some conditions on $\mathbf{\Omega}$ (e.g., uniform sampling):

> If the columns of $\mathbf{X}$ lie in an r-dimensional subspace $S^\star$
>
> $\Downarrow$
>
> $S^\star$ is the only r-dimensional subspace that agrees with $\mathbf{X}_{\mathbf{\Omega}}$.



Low-Rank Matrix Completion

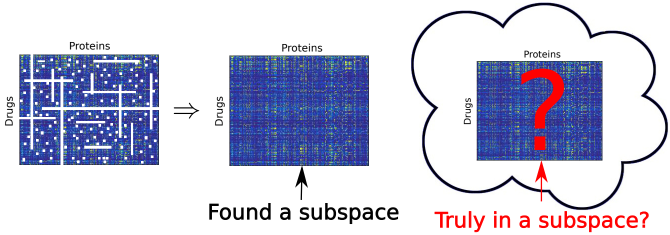
In practice, we hardly ever know whether our matrix lies in a subspace.



Low-Rank Matrix Completion

We need to turn things around:

- Say I have an incomplete matrix $\mathbf{X}_{\mathbf{\Omega}}$.
- Say I find an r-dimensional subspace S that agrees with $\mathbf{X}_{\mathbf{\Omega}}$.
- Is $\mathbf{X}$ truly in S?



Low-Rank Matrix Completion

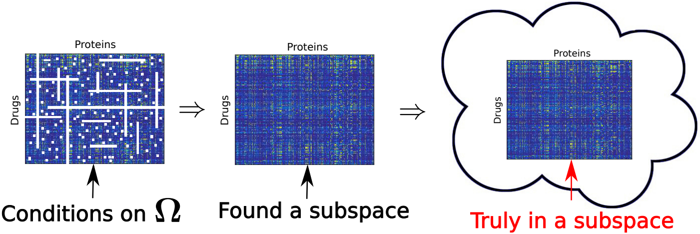
We need a converse to LRMC:

- Under some conditions on $\mathbf{\Omega}$:

If there is an r-dimensional subspace S that agrees with $\mathbf{X}_{\mathbf{\Omega}}$

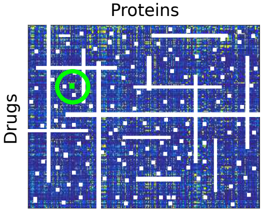
$\Downarrow$

The columns of $\mathbf{X}$ truly lie in S.

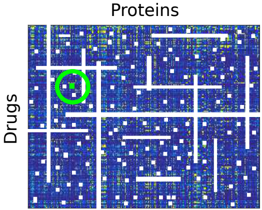


Why worry about this?

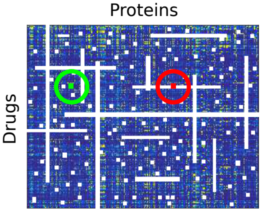
Why worry about this?



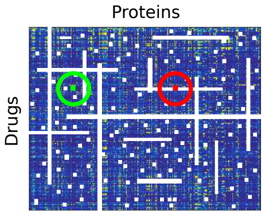
Why worry about this?

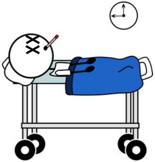


Why worry about this?



Why worry about this?



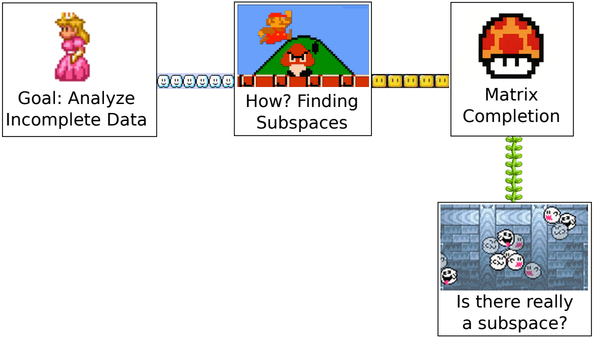


- (Pretty bad secondary effect!)

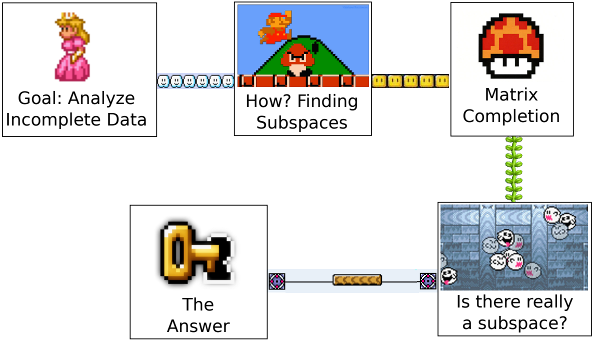
So, is there really a subspace?

- We needed to know when will a set $\mathbf{X}_{\boldsymbol{\Omega}}$ of incomplete vectors define a subspace.
- What are the conditions on $\boldsymbol{\Omega}$?

What am I telling you?



What am I telling you?



The Answer

Notation

- For any matrix $\mathbf{\Omega}'$ formed with a subset of the columns in $\mathbf{\Omega}$:

$$\mathbf{\Omega}' = \underbrace{\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}}_{n(\mathbf{\Omega}') \,:=\, \#\text{columns}} \left.\vphantom{\begin{matrix} 1 \\ 1 \\ 0 \end{matrix}}\right\} m(\mathbf{\Omega}') := \#\text{nonzero rows}$$

- Assume without loss of generality:
  - $\mathbf{\Omega}$ has $r+1$ nonzero entries per column.
  - $\mathbf{\Omega}$ has $r(d-r)$ columns.

The Answer

Technical detail (so there are no secrets between us):

$$\mathbf{X} = \underbrace{\mathbf{U}^\star}_{d\times k}\underbrace{\boldsymbol{\Theta}^\star}_{k\times N}.$$

- $\nu_{\mathrm{G}} =$ Uniform measure on $\mathrm{Gr}(k, \mathbb{R}^d)$.
- $\nu_{\Theta} =$ Lebesgue measure on $\mathbb{R}^{k\times N}$.
- Our results hold almost surely w.r.t. product measure $\nu_{\mathrm{G}} \times \nu_{\Theta}$.

The Answer

Theorem (P.-A., Boston, Nowak (Allerton '15))

For almost every $\mathbf{X}$, at most finitely many r-dimensional subspaces can agree with $\mathbf{X}_{\mathbf{\Omega}}$ if and only if every matrix $\mathbf{\Omega}'$ formed with a subset of the columns in $\mathbf{\Omega}$ satisfies

$$m(\mathbf{\Omega}') \; \geq \; n(\mathbf{\Omega}')/r + r.$$

The Answer

Theorem (P.-A., Boston, Nowak (Allerton '15))

For almost every $\mathbf{X}$, at most finitely many r-dimensional subspaces can agree with $\mathbf{X}_{\mathbf{\Omega}}$ if and only if every matrix $\mathbf{\Omega}'$ formed with a subset of the columns in $\mathbf{\Omega}$ satisfies

$$m(\mathbf{\Omega}') \;\geq\; n(\mathbf{\Omega}')/r + r.$$

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

The Answer

Theorem (P.-A., Boston, Nowak (Allerton '15))

For almost every $\mathbf{X}$, at most finitely many r-dimensional subspaces can agree with $\mathbf{X}_{\mathbf{\Omega}}$ if and only if every matrix $\mathbf{\Omega}'$ formed with a subset of the columns in $\mathbf{\Omega}$ satisfies

$$m(\mathbf{\Omega}') \ \geq \ n(\mathbf{\Omega}')/r + r.$$

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

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

The Answer

For almost every $\mathbf{X}$, at most finitely many r-dimensional subspaces can agree with $\mathbf{X}_{\mathbf{\Omega}}$ if and only if every matrix $\mathbf{\Omega}'$ formed with a subset of the columns in $\mathbf{\Omega}$ satisfies

$$m(\mathbf{\Omega}') \;\geq\; n(\mathbf{\Omega}')/r + r.$$

There is a set of measure zero of bad matrices for which this theorem does not apply.

$$\mathbf{X} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \qquad\qquad \mathbf{X}_{\mathbf{\Omega}} = \begin{bmatrix} 0 & \cdot & \cdot \\ 0 & 0 & \cdot \\ \cdot & 0 & 0 \\ \cdot & \cdot & 0 \end{bmatrix}$$

The Answer

For almost every $\mathbf{X}$, at most finitely many r-dimensional subspaces can agree with $\mathbf{X}_{\boldsymbol{\Omega}}$ if and only if every matrix $\boldsymbol{\Omega}'$ formed with a subset of the columns in $\boldsymbol{\Omega}$ satisfies

$$m(\boldsymbol{\Omega}') \ \geq \ n(\boldsymbol{\Omega}')/r + r.$$

This is the answer!

Every subset of n columns of $\boldsymbol{\Omega}$ has at least $n/r + r$ nonzero rows.

$$\boldsymbol{\Omega} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad \Rightarrow \ \text{Check:} \ \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

The answer

Now we know when there are at most finitely many completions.

- Then what?
- Just a few additional entries give us the converse we were looking for.

The Answer

Theorem (P.-A., Boston, Nowak (Allerton '15))

Suppose $\mathbf{X}_{\mathbf{\Omega}}$ has an additional $(d-r)$ columns observed on $\hat{\mathbf{\Omega}}$, such that every matrix $\mathbf{\Omega}'$ formed with a subset of the columns in $\hat{\mathbf{\Omega}}$ satisfies

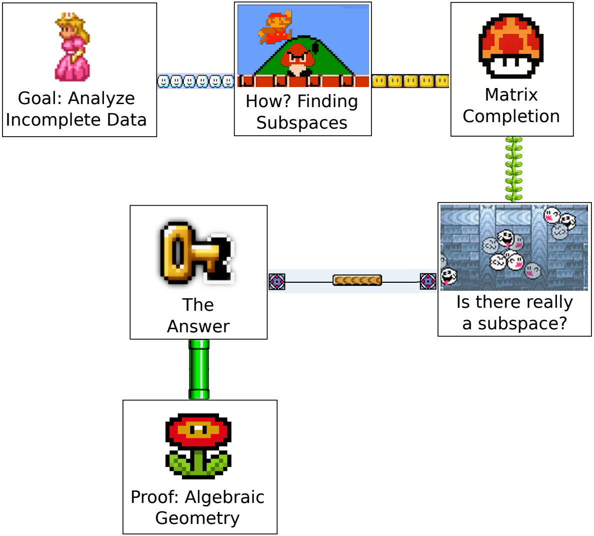
$$m(\mathbf{\Omega}') \;\geq\; n(\mathbf{\Omega}') + r.$$

If there is an r-dimensional subspace S that agrees with $\mathbf{X}_{\mathbf{\Omega}}$

$$\Downarrow$$

The columns of $\mathbf{X}$ truly lie in S.

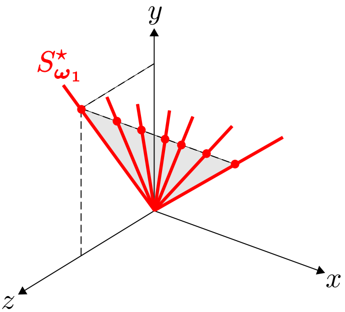
What am I telling you?



Idea of the proof

A column with $r+1$ samples imposes one restriction on what $S^\star$ may be.

$$\mathbf{X}_{\mathbf{\Omega}} \quad = \quad \begin{array}{c} \mathbf{x}_{\boldsymbol{\omega}_1} \\ \left[\begin{array}{c} \cdot \\ 1 \\ 1 \end{array}\right] \end{array}$$



- A subspace S *agrees* with $\mathbf{x}_{\boldsymbol{\omega}_1}$ $\iff$ $\underbrace{f_1(S) = 0}_{\text{degree-}r\text{ polynomial}}$.

Idea of the proof

More precisely:

- Take a basis of S:

$$S = \operatorname{span} \underbrace{\begin{bmatrix} \\ \mathbf{U} \\ \\ \end{bmatrix}}_{r} \Bigg\} d.$$

- Then $\mathbf{x}_{\boldsymbol{\omega}_i} \in S$ is equivalent to:

$$r+1 \left\{ \begin{bmatrix} \mathbf{x}_{\boldsymbol{\omega}_i} \end{bmatrix} \right. = \begin{bmatrix} \mathbf{U}_{\boldsymbol{\omega}_i} \end{bmatrix} \boldsymbol{\theta}_{\boldsymbol{i}}.$$

Idea of the proof

- We can split this as:

$$\begin{array}{c} r\left\{\begin{array}{c} \\ \\ \\ \end{array}\right. \\ 1\left\{\begin{array}{c} \\ \end{array}\right. \end{array} \begin{bmatrix} \mathbf{x}_{\boldsymbol{\Delta}_i} \\ \hline \mathbf{x}_{\boldsymbol{\nabla}_i} \end{bmatrix} = \begin{bmatrix} \mathbf{U}_{\boldsymbol{\Delta}_i} \\ \hline \mathbf{U}_{\boldsymbol{\nabla}_i} \end{bmatrix} \boldsymbol{\theta}_i.$$

- We can use the top block to solve for $\boldsymbol{\theta}_i$:

$$\boldsymbol{\theta}_i = \mathbf{U}_{\boldsymbol{\Delta}_i}^{-1} \mathbf{x}_{\boldsymbol{\Delta}_i}.$$

- Plug this in the last row:

$$\mathbf{x}_{\boldsymbol{\nabla}_i} = \mathbf{U}_{\boldsymbol{\nabla}_i} \mathbf{U}_{\boldsymbol{\Delta}_i}^{-1} \mathbf{x}_{\boldsymbol{\Delta}_i}.$$

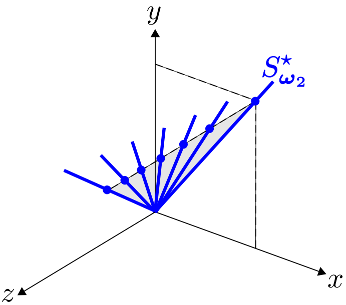
- Or equivalently

$$\underbrace{\mathbf{x}_{\boldsymbol{\nabla}_i} - \mathbf{U}_{\boldsymbol{\nabla}_i} \mathbf{U}_{\boldsymbol{\Delta}_i}^{-1} \mathbf{x}_{\boldsymbol{\Delta}_i}}_{f_i(\mathbf{U}_{\boldsymbol{\omega}_i} | \mathbf{x}_{\boldsymbol{\omega}_i})} = 0.$$

Idea of the proof

An other column with $r+1$ samples imposes an other restriction.

$$\mathbf{X}_{\boldsymbol{\Omega}} = \begin{array}{c} \mathbf{x}_{\boldsymbol{\omega}_2} \\ \left[\begin{array}{c} 2 \\ 2 \\ \cdot \end{array}\right] \end{array}$$

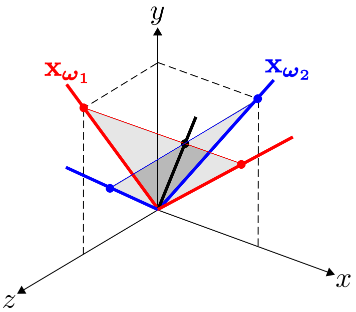


- A subspace S *agrees* with $\mathbf{x}_{\boldsymbol{\omega}_2} \iff f_2(\mathbf{U}_{\boldsymbol{\omega}_2}|\mathbf{x}_{\boldsymbol{\omega}_2}) = 0$.

Idea of the proof

Each column with $r+1$ samples imposes one restriction.

$$\mathbf{X}_{\boldsymbol{\Omega}} \;=\; \begin{array}{c} \begin{array}{cc} \mathbf{x}_{\omega_1} & \mathbf{x}_{\omega_2} \end{array} \\ \left[\begin{array}{cc} \cdot & 2 \\ 1 & 2 \\ 1 & \cdot \end{array}\right] \end{array}$$



- A subspace S *agrees* with $\mathbf{X}_{\boldsymbol{\Omega}}$ $\iff \begin{cases} f_1(\mathbf{U}_{\boldsymbol{\omega}_1}|\mathbf{x}_{\boldsymbol{\omega}_1}) = 0 \\ f_2(\mathbf{U}_{\boldsymbol{\omega}_2}|\mathbf{x}_{\boldsymbol{\omega}_2}) = 0 \end{cases}$.

Idea of the proof

- We thus obtain a set of polynomials:

$$f_1, f_2, \ldots, f_N.$$

- $\mathbf{U}$ has $r(d-r)$ degrees of freedom:

$$\mathbf{U} = \begin{bmatrix} \mathbf{I} \\ \hline \mathbf{V} \end{bmatrix} \begin{matrix} \}\, r \\ \}\, d-r. \end{matrix}$$

- We want $r(d-r)$ algebraically independent polynomials.

Idea of the proof

- Recall:

$$\boldsymbol{\Omega}' = \underbrace{\left.\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}\right.}_{n(\boldsymbol{\Omega}') \ := \ \#\text{columns}} \quad \left.\begin{matrix} \\ \\ \\ \end{matrix}\right\} m(\boldsymbol{\Omega}') := \#\text{nonzero rows}$$

- $f_i(\mathbf{U}_{\boldsymbol{\omega}_i} | \mathbf{x}_{\boldsymbol{\omega}_i})$ only involves the variables corresponding to the nonzero rows of $\boldsymbol{\omega}_i$.
- $\mathcal{F}'(\mathbf{U}_{\boldsymbol{\Omega}'} | \mathbf{X}_{\boldsymbol{\Omega}'})$ = subset of polynomials corresponding to $\boldsymbol{\Omega}'$:
  - $n(\boldsymbol{\Omega}')$ polynomials.
  - $r(m(\boldsymbol{\Omega}') - r)$ variables.

$$\mathbf{U} = \begin{bmatrix} \mathbf{I} \\ \hline \mathbf{V} \end{bmatrix} \begin{matrix} \}\, r \\ \}\, m(\boldsymbol{\Omega}') - r \end{matrix}$$

Idea of the proof

There is a subspace that agrees with $\mathbf{X}_{\boldsymbol{\Omega}}$

$$\Downarrow$$

$\mathcal{F}'(\mathbf{U}_{\boldsymbol{\Omega}'}|\mathbf{X}_{\boldsymbol{\Omega}'}) = 0$ has at least one solution.

- If

$$\underbrace{n(\boldsymbol{\Omega}')}_{equations} > \underbrace{r(m(\boldsymbol{\Omega}') - r)}_{unknowns}$$

$\Rightarrow$ Polynomials are dependent.

- (That is the easy direction)

Idea of the proof

- Our results hold a.s. w.r.t.

$$\underbrace{\nu_{\mathrm{G}}}_{\text{Uniform on } \mathrm{Gr}(k,\mathbb{R}^d)} \times \underbrace{\nu_{\Theta}}_{\text{Lebesgue on } \mathbb{R}^{k\times N}} .$$

- $\exists$ Bijection between dense open subset of $\mathrm{Gr}(k,\mathbb{R}^d)$ and $\mathbb{R}^{(d-k)\times k}$ via

$$S = \operatorname{span} \begin{bmatrix} \mathbf{I} \\ \mathbf{V} \end{bmatrix} \begin{matrix} \}k \\ \}d-k \end{matrix}$$

- Our results hold a.s. w.r.t.

$$\underbrace{\nu_{\mathbf{V}}}_{\text{Lebesgue on } \mathbb{R}^{(d-k)\times k}} \times \underbrace{\nu_{\Theta}}_{\text{Lebesgue on } \mathbb{R}^{k\times N}} .$$

Idea of the proof

Recall:

$$\mathbf{X} = \mathbf{U}^\star \boldsymbol{\Theta}^\star = \underbrace{\begin{bmatrix} \mathbf{I} \\ \mathbf{V}^\star \end{bmatrix}}_{d \times k} \underbrace{\boldsymbol{\Theta}^\star}_{k \times N}$$

$\Rightarrow \mathcal{F}'(\mathbf{U}_{\boldsymbol{\Omega}'} | \mathbf{X}_{\boldsymbol{\Omega}'}) = \mathcal{F}'(\mathbf{U}_{\boldsymbol{\Omega}'} | \mathbf{V}^\star_{\boldsymbol{\Omega}'}, \boldsymbol{\Theta}^\star)$.

- The elements of $\mathbf{V}^\star_{\boldsymbol{\Omega}'}, \boldsymbol{\Theta}^\star$ are *generic* real numbers.

Idea of the proof

- For a.e. $(\mathbf{V}^\star, \mathbf{\Theta})$, if

$$\underbrace{n(\mathbf{\Omega}')}_{equations} \leq \underbrace{r(m(\mathbf{\Omega}') - r)}_{unknowns} \qquad \forall\ \mathbf{\Omega}' \subset \mathbf{\Omega}$$

$\Rightarrow$ Polynomials are a algebraically independent.

Idea of the proof

- For a.e. $(\mathbf{V}^\star, \mathbf{\Theta})$, if

$$\underbrace{n(\mathbf{\Omega}')}_{equations} \leq \underbrace{r(m(\mathbf{\Omega}') - r)}_{unknowns} \qquad \forall\, \mathbf{\Omega}' \subset \mathbf{\Omega}$$

$\Rightarrow$ Polynomials are a algebraically independent.

After this, deep algebraic geometry results do the heavy lifting:

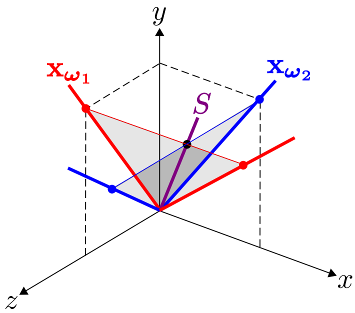
$\Rightarrow$ Polynomials are a regular sequence.

$\Rightarrow$ Polynomials define a zero-dimensional variety.

$\Rightarrow$ At most finitely many solutions (subspaces) will agree with $\mathbf{X}_{\mathbf{\Omega}}$.

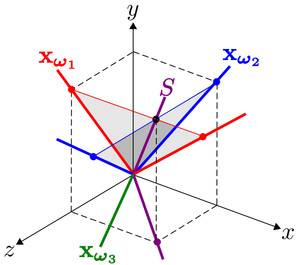
Idea of the proof

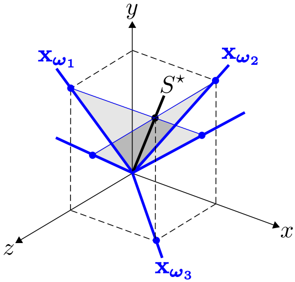
- How do we know $\mathbf{X}$ truly lies in S?



Idea of the proof

We need a few additional *checksum* polynomials (consistency check):



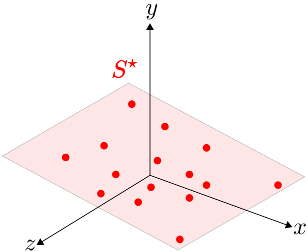


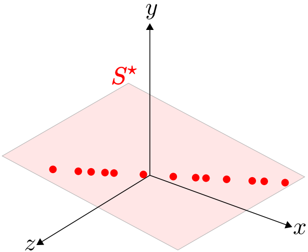
Idea of the proof

Full-data case:

- How can I know if $\mathbf{X}$ truly lies in S?
- With one *generic* column (consistency check):

What do I mean generic?





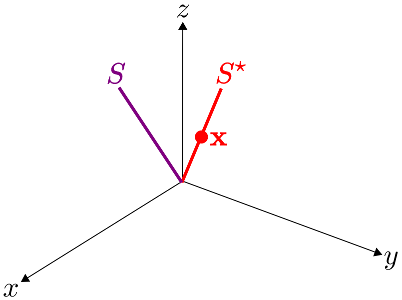
Idea of the proof

Full-data case:

- How can I know if $\mathbf{X}$ truly lies in S?
- With one *generic* column (consistency check):

$$\begin{array}{c} \mathbf{x} \\ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \end{array}$$

- $\mathbf{x} \in S \iff S = S^{\star}$.



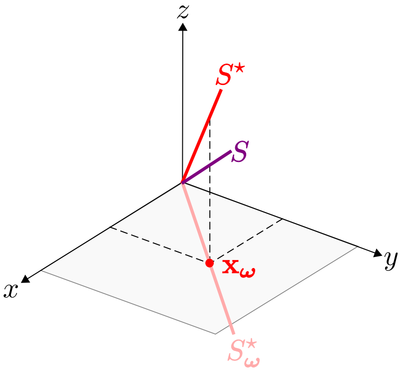
Idea of the proof

Missing-data case:

- Something similar:

$$\begin{array}{c} \color{red}{\mathbf{x}_{\boldsymbol{\omega}}} \\ \left[\begin{array}{c} \color{red}{1} \\ \color{red}{1} \\ \cdot \end{array}\right] \end{array}$$

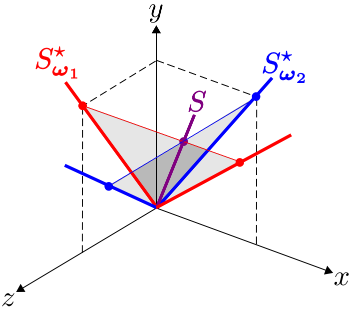
- $\mathbf{x}_{\boldsymbol{\omega}} \in S \iff S_{\boldsymbol{\omega}} = S^\star_{\boldsymbol{\omega}}$.



Idea of the proof

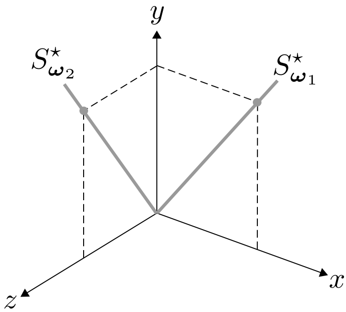
So the question is:

- If $S_{\boldsymbol{\omega}_i} = S^\star_{\boldsymbol{\omega}_i}$ for every i...
- Can we guarantee that $S = S^\star$?



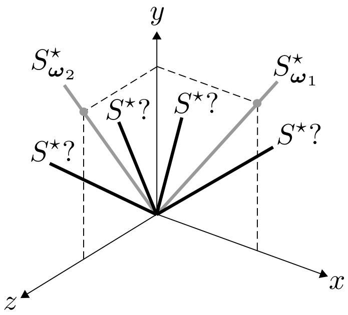
Idea of the proof

Suppose I don't tell you $S^\star$... but I give you a set of canonical projections of $S^\star$.



Idea of the proof

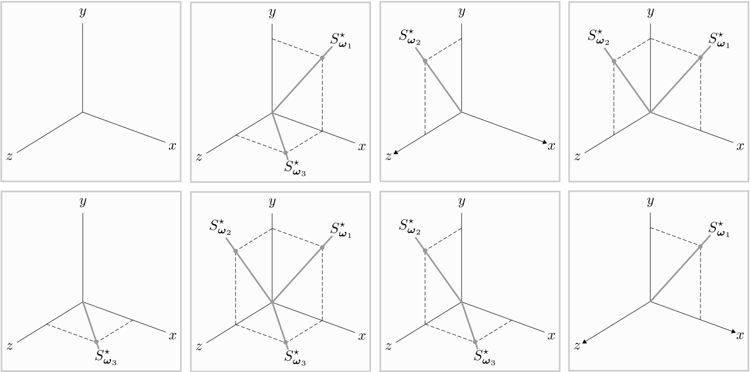
Suppose I don't tell you $S^\star$...but I give you a set of canonical projections of $S^\star$.



Can you uniquely determine $S^\star$ from this set of projections?

Idea of the proof

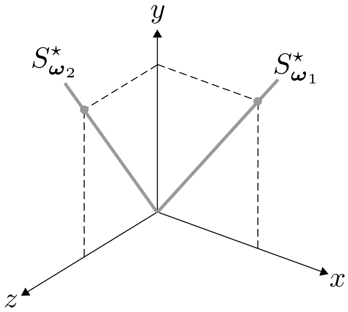
Well... sometimes you can, sometimes you can't.



We characterized when you can, and when you can't.

Idea of the proof

The columns of $\hat{\mathbf{\Omega}}$ will index the given projections.



$$\hat{\mathbf{\Omega}} = \begin{array}{c} \begin{array}{cc} \boldsymbol{\omega}_1 & \boldsymbol{\omega}_2 \end{array} \\ \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} \end{array}$$

- Assume without loss of generality:
  - $\hat{\mathbf{\Omega}}$ has $r+1$ nonzero entries per column.
  - $\hat{\mathbf{\Omega}}$ has $d-r+1$ columns.

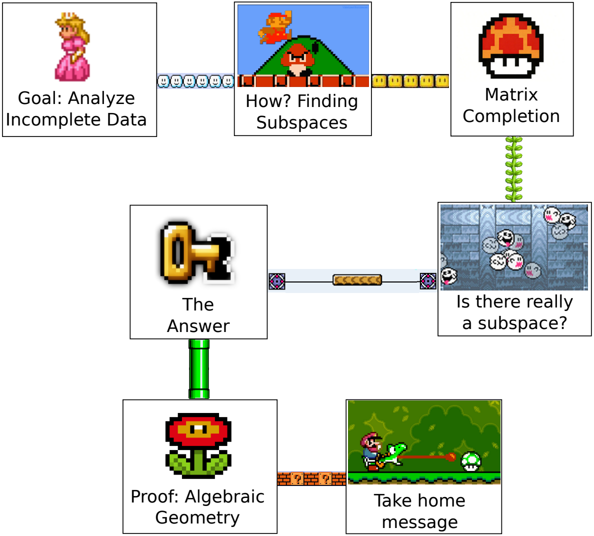
Idea of the proof

Theorem (P.-A., Boston, Nowak, ISIT '15)

For almost every $S^\star$, $S^\star$ is the only subspace that agrees with the given projections if every matrix $\mathbf{\Omega}'$ formed with a proper *subset of the columns in $\hat{\mathbf{\Omega}}$ satisfies*

$$m(\mathbf{\Omega}') \;\geq\; n(\mathbf{\Omega}') + r.$$

What am I telling you?



The Answer in Words

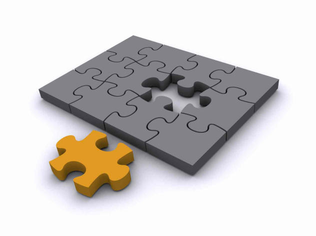
If a matrix does not satisfy our sampling conditions, then you cannot find its subspace.

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

The Answer in Words

If a matrix satisfies our sampling conditions, then you can find its subspace up to finite choice.

$$\mathbf{X}_{\boldsymbol{\Omega}} = \begin{bmatrix} 1 & 1 & 3 & \cdot \\ 1 & 2 & \cdot & 1 \\ 3 & \cdot & 5 & 4 \\ \cdot & 7 & 6 & 5 \end{bmatrix}$$



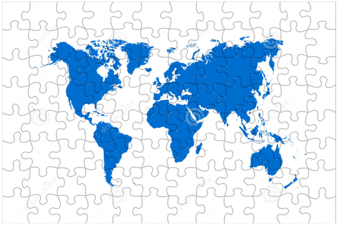
Sometimes finite choice = unique choice (e.g., rank$= 1$), but sometimes not.

The Answer in Words

With just a few additional samples we can make sure that

- $\mathbf{X}$ really is in an subspace.
- You found the right subspace.

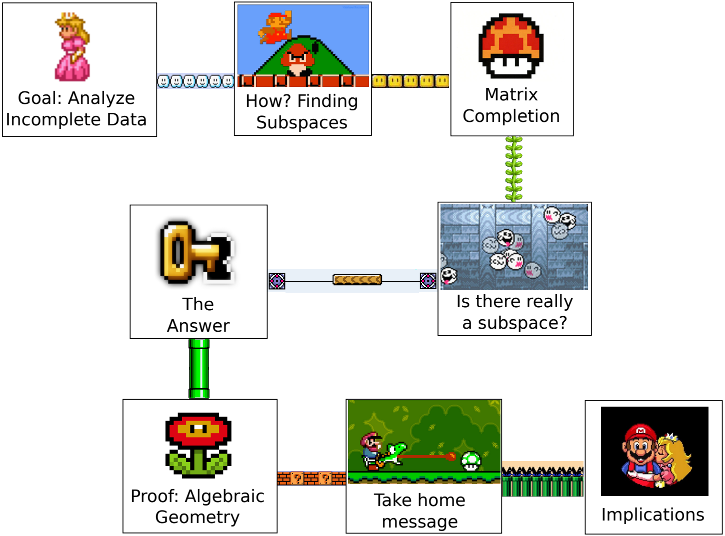
$$\mathbf{X}_{\boldsymbol{\Omega}} = \begin{bmatrix} 1 & 1 & 3 & \cdot & -1 & 1 \\ 1 & 2 & \cdot & 1 & \cdot & -1 \\ 3 & \cdot & 5 & 4 & 3 & \cdot \\ \cdot & 7 & 6 & 5 & 5 & -2 \end{bmatrix}$$



The Big Picture

	Uniquely define a subspace	Verify there is truly a subspace
Full data	r	$r + 1$
Missing data	$(r+1)(d-r)$	$(r+1)(d-r+1)$

What am I telling you?

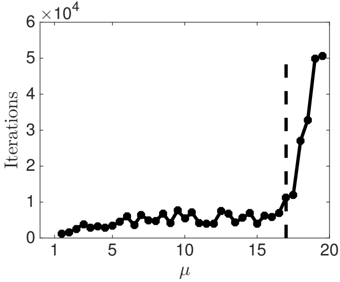


Implications on LRMC

- P.-A., Boston, Nowak (Allerton '15):
  - For almost every matrix, $\mathcal{O}(\max\{r, \log d\})$ uniform random entries per column are sufficient for completion.
- Regardless of coherence! (at least theoretically)

Implications on LRMC

- P.-A., Boston, Nowak (Allerton '15):
  - For almost every matrix, $\mathcal{O}(\max\{r, \log d\})$ uniform random entries per column are sufficient for completion.
- Regardless of coherence! (at least theoretically)
- But coherence seems to come at a price in practice



Implications on LRMC

Validation criteria:

- Suppose you observe the right entries.
- Try to complete the matrix using any method.
- If you find a rank-r completion, then it is the right completion.
- In lieu of coherence assumptions.
- In lieu of uniform sampling assumptions.
- With probability 1 (as opposed to *with high probability*).

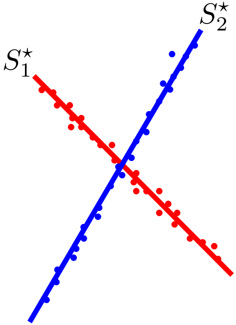
Implications

- Our results tell us exactly which entries to observe.
  - We can now design Adaptive LRMC Algorithms.

Implications

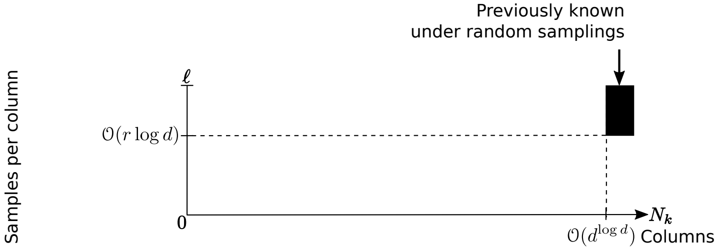
Help answer an important open question:

- The Sample Complexity of Subspace Clustering with Missing Data.

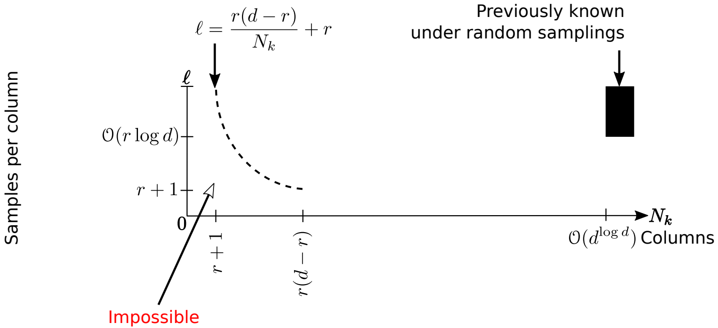


$$\begin{bmatrix} 1 & \cdot & \cdot & 3 & \cdot & 3 & \cdot & 1 & 2 & \cdot \\ 2 & \cdot & 2 & \cdot & \cdot & 6 & \cdot & \cdot & 4 & \cdot \\ \cdot & \cdot & 3 & \cdot & \cdot & 9 & \cdot & 3 & 6 & \cdot \\ 1 & \cdot & 1 & 3 & 6 & \cdot & 4 & 1 & 2 & 2 \\ \cdot & 8 & \cdot & \cdot & 6 & \cdot & 4 &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

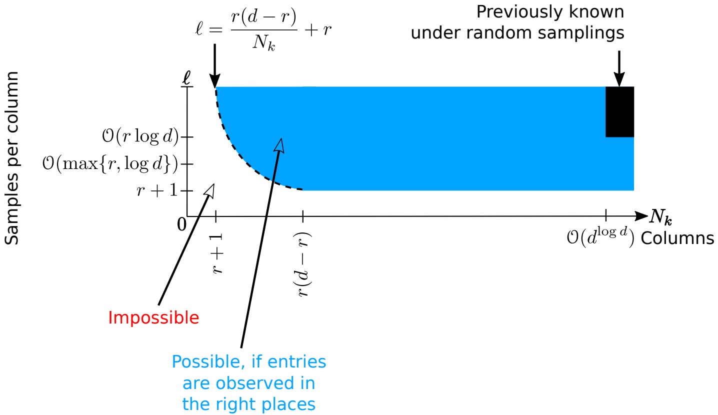
Sample Complexity of SCMD



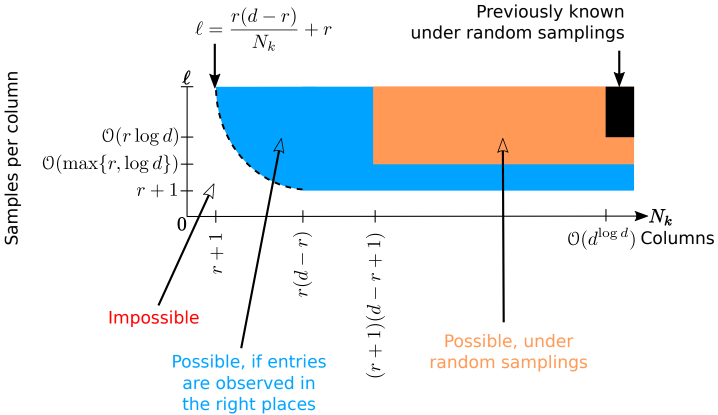
Sample Complexity of SCMD



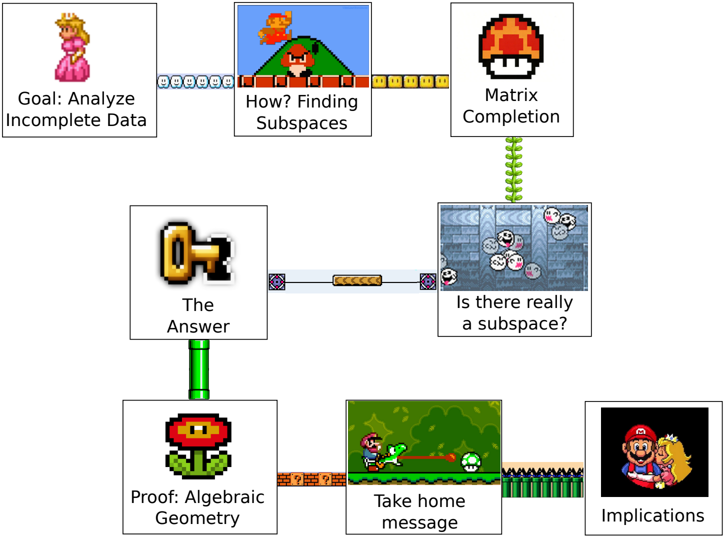
Sample Complexity of SCMD

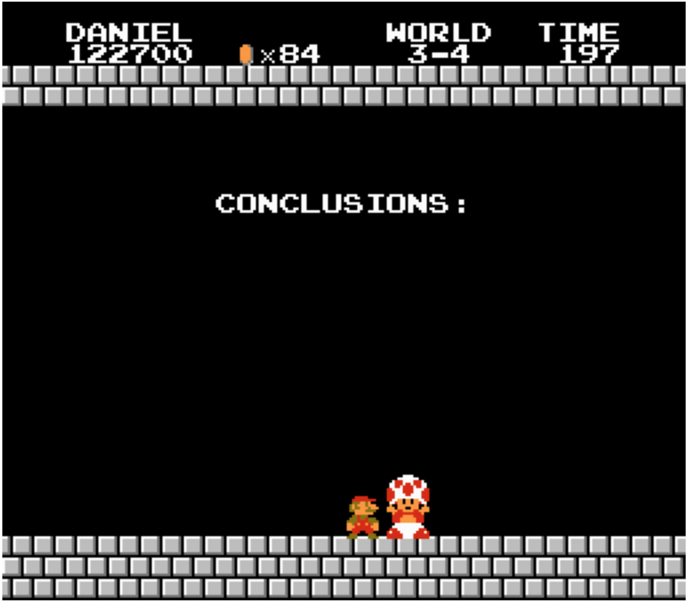


Sample Complexity of SCMD



Long Story Short




DANIEL
122700
×84
WORLD
3-4
TIME
197
CONCLUSIONS:

Conclusions

Now we know:

	Uniquely define a subspace	Verify there is truly a subspace
Full data	r	$r+1$
Missing data	$(r+1)(d-r)$	$(r+1)(d-r+1)$

This has important implications on:

- LRMC.
- SCMD.
- Adaptive strategies
- Related problems.


DANIEL
122700
×84
WORLD
3-4
TIME
197
THANK YOU BERKELEY!