Necessary and Sufficient Conditions for Sketched Subspace Clustering

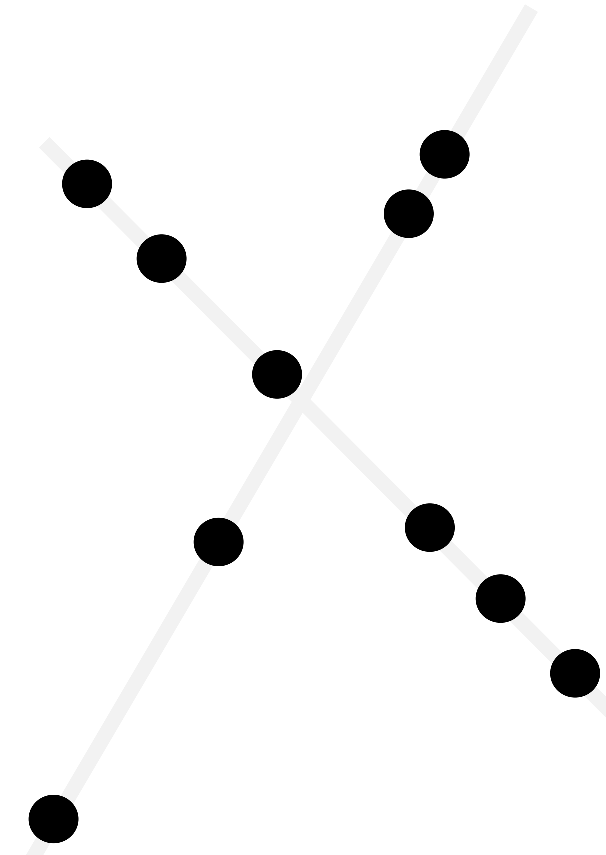
Daniel Pimentel-Alarcón, Laura Balzano & Robert Nowak
UNIVERSITY *of* WISCONSIN-MADISON

Allerton 2016

Subspace Clustering

Subspace Clustering

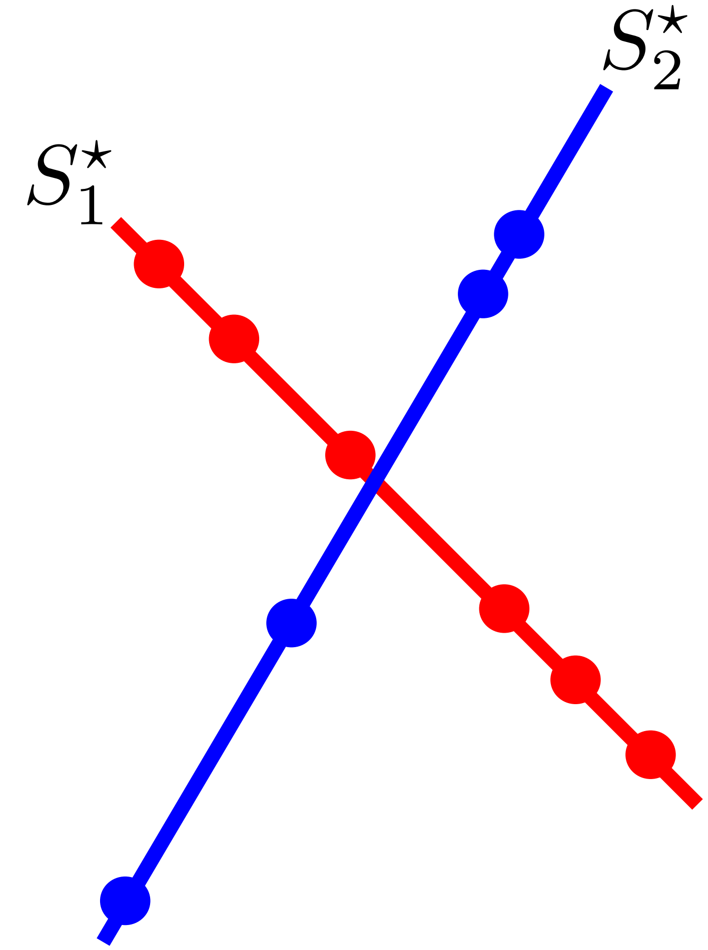
$$\begin{bmatrix} 1 & 4 & 1 & 3 & 3 & 1 & 2 & 1 & 2 & 1 \\ 2 & 4 & 2 & 6 & 3 & 2 & 2 & 2 & 4 & 1 \\ 3 & 4 & 3 & 9 & 3 & 3 & 2 & 3 & 6 & 1 \\ 1 & 8 & 1 & 3 & 6 & 1 & 4 & 1 & 2 & 2 \\ 2 & 8 & 2 & 6 & 6 & 2 & 4 & 2 & 4 & 2 \\ 3 & 8 & 3 & 9 & 6 & 3 & 4 & 3 & 6 & 2 \end{bmatrix}$$



Subspace Clustering

We are given: Columns in a union of subspaces.

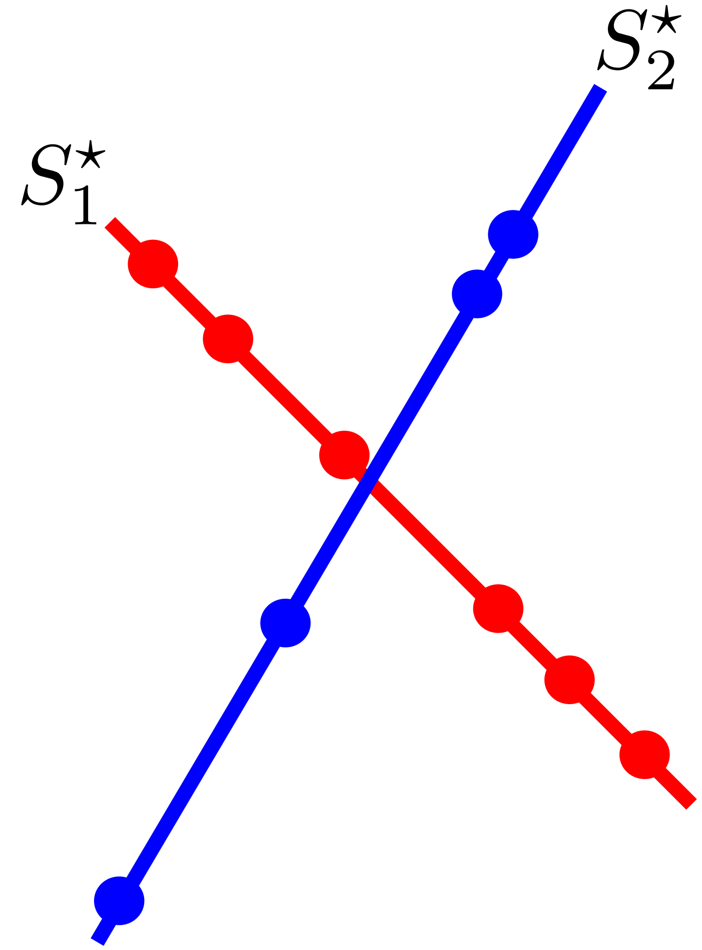
$$\begin{bmatrix} 1 & 4 & 1 & 3 & 3 & 1 & 2 & 1 & 2 & 1 \\ 2 & 4 & 2 & 6 & 3 & 2 & 2 & 2 & 4 & 1 \\ 3 & 4 & 3 & 9 & 3 & 3 & 2 & 3 & 6 & 1 \\ 1 & 8 & 1 & 3 & 6 & 1 & 4 & 1 & 2 & 2 \\ 2 & 8 & 2 & 6 & 6 & 2 & 4 & 2 & 4 & 2 \\ 3 & 8 & 3 & 9 & 6 & 3 & 4 & 3 & 6 & 2 \end{bmatrix}$$



Subspace Clustering

We are given: Columns in a union of subspaces.
Goal: Cluster the columns, or find the subspaces.

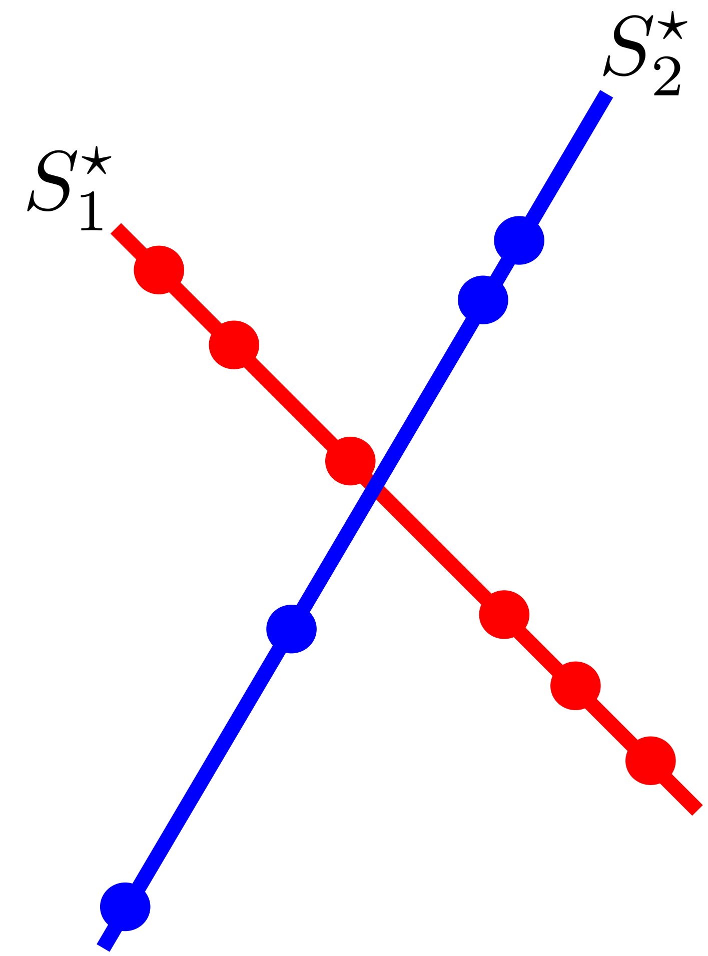
$$\begin{bmatrix} 1 & 4 & 1 & 3 & 3 & 1 & 2 & 1 & 2 & 1 \\ 2 & 4 & 2 & 6 & 3 & 2 & 2 & 2 & 4 & 1 \\ 3 & 4 & 3 & 9 & 3 & 3 & 2 & 3 & 6 & 1 \\ 1 & 8 & 1 & 3 & 6 & 1 & 4 & 1 & 2 & 2 \\ 2 & 8 & 2 & 6 & 6 & 2 & 4 & 2 & 4 & 2 \\ 3 & 8 & 3 & 9 & 6 & 3 & 4 & 3 & 6 & 2 \end{bmatrix}$$



Sketched
Subspace Clustering

We are given: Projections of Columns in a union of subspaces.
Goal: Cluster the columns, or find the subspaces.

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



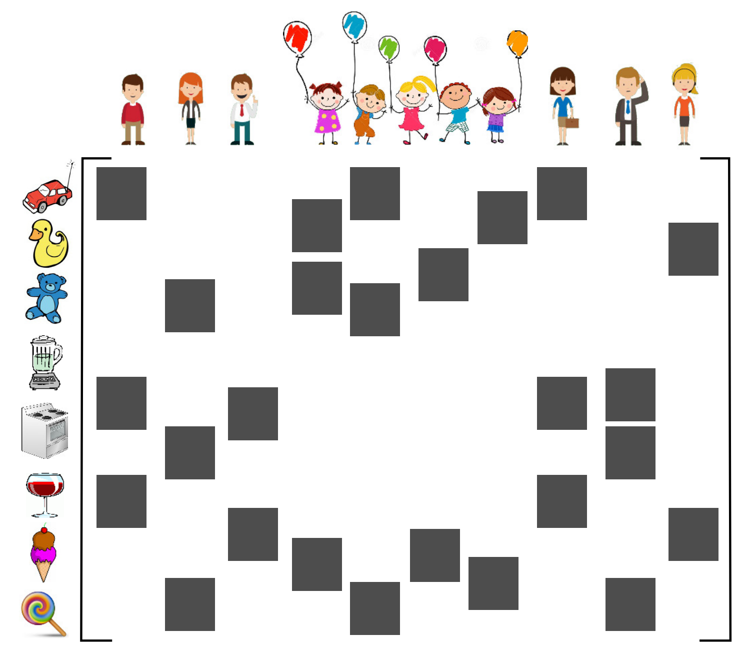
Sketched Subspace Clustering

We are given: Projections of Columns in a union of subspaces.

Goal: Cluster the columns, or find the subspaces.

Applications

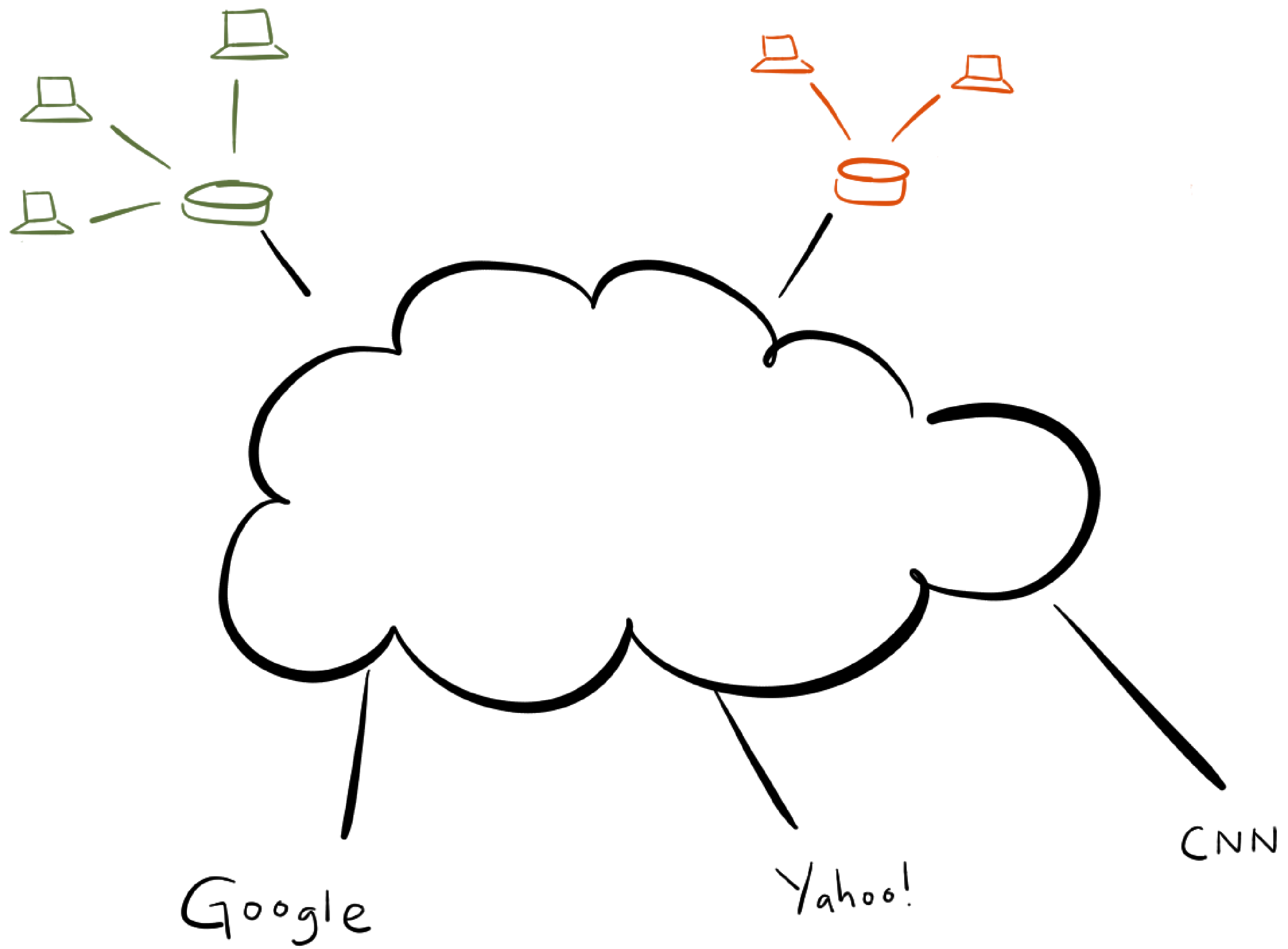
Applications



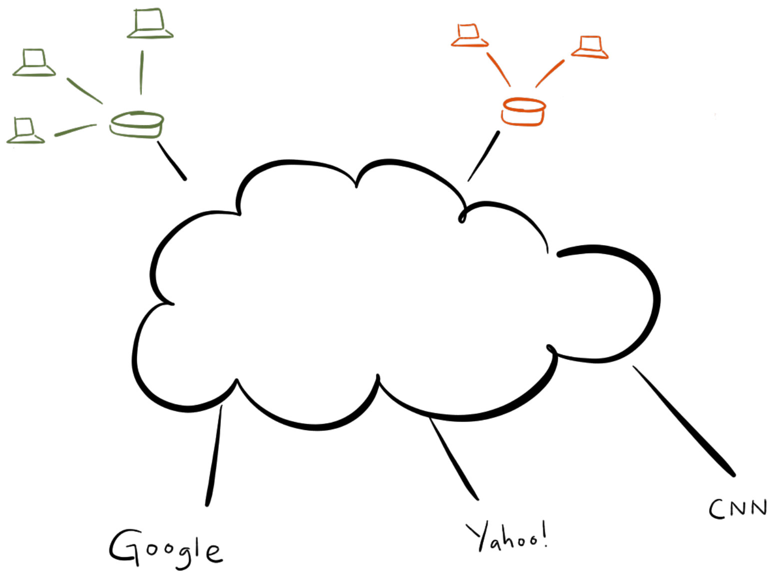
Applications

Example: Network Topology Estimation

Example: Network Topology Estimation



Example: Network Topology Estimation



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

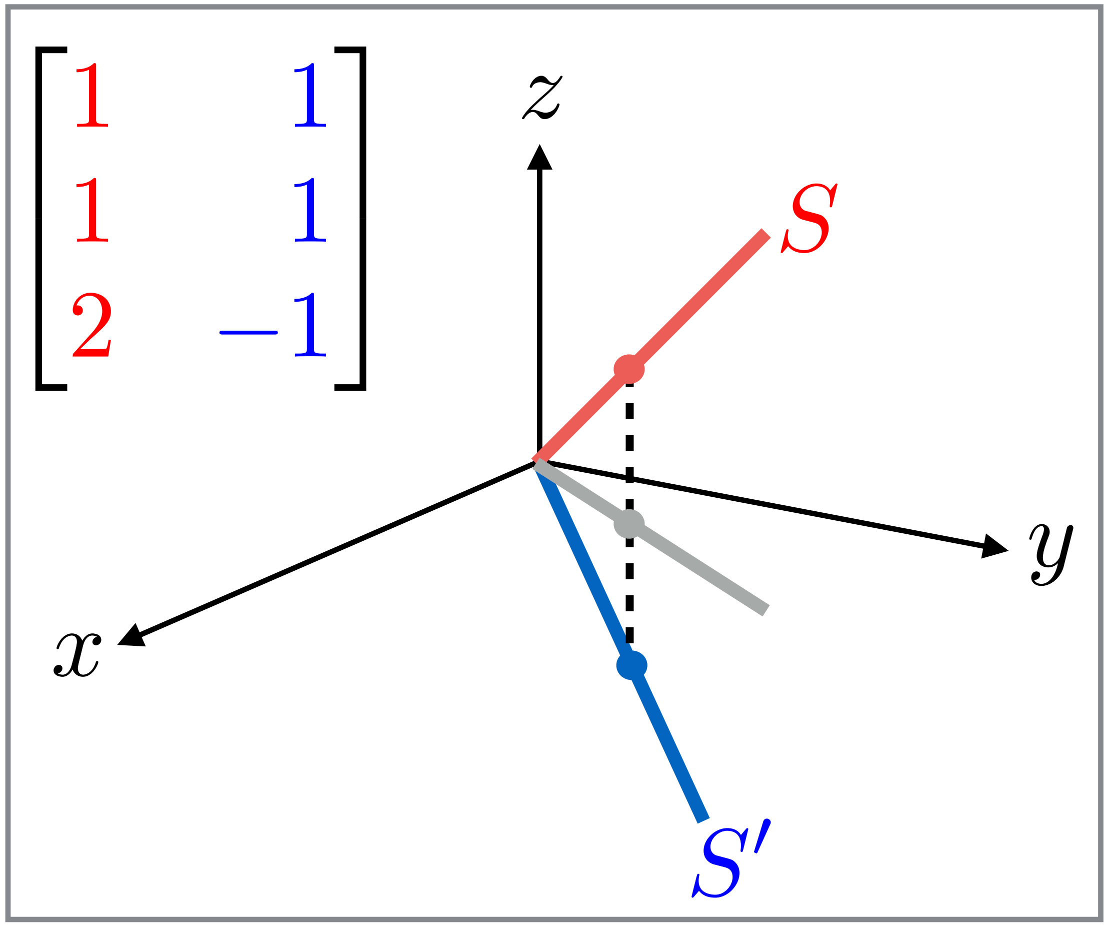
Example: Network Topology Estimation

Complication

Two subspaces (even orthogonal) can appear identical if they are only observed on a subset of coordinates

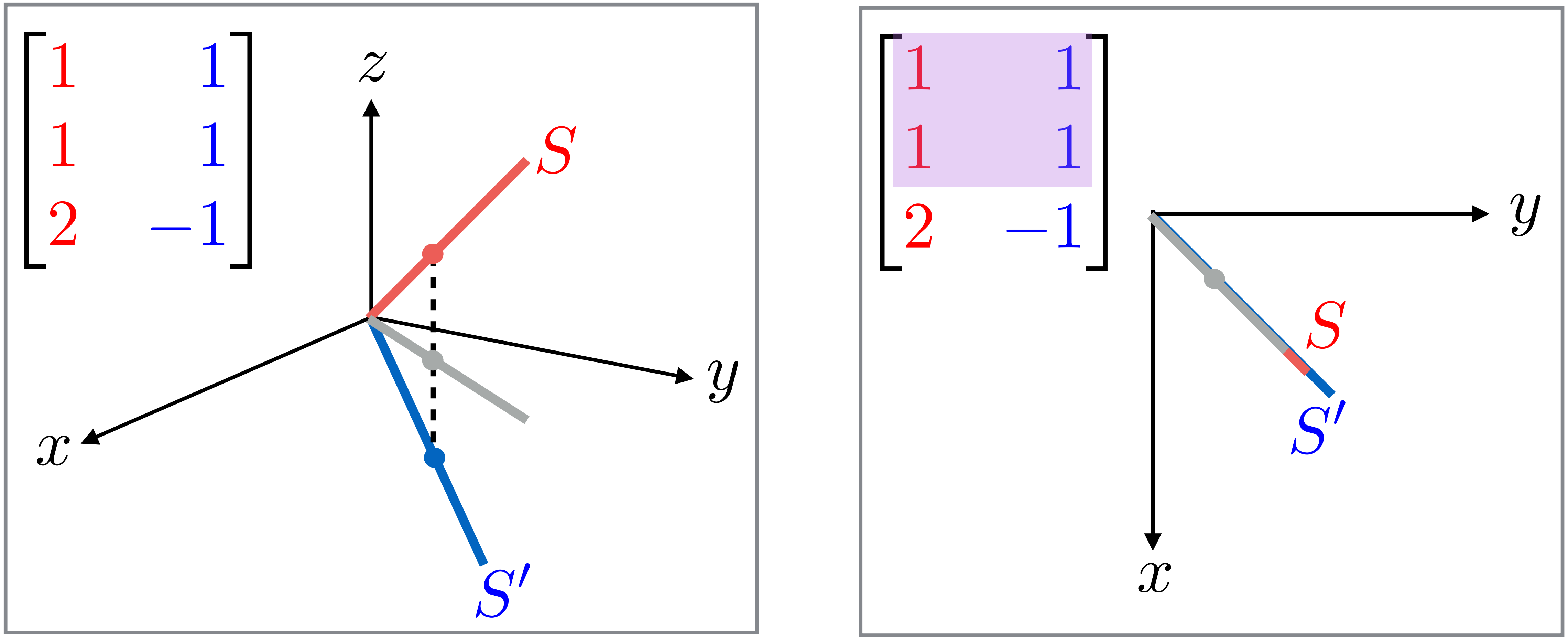
Complication

Two subspaces (even orthogonal) can appear identical if they are only observed on a subset of coordinates



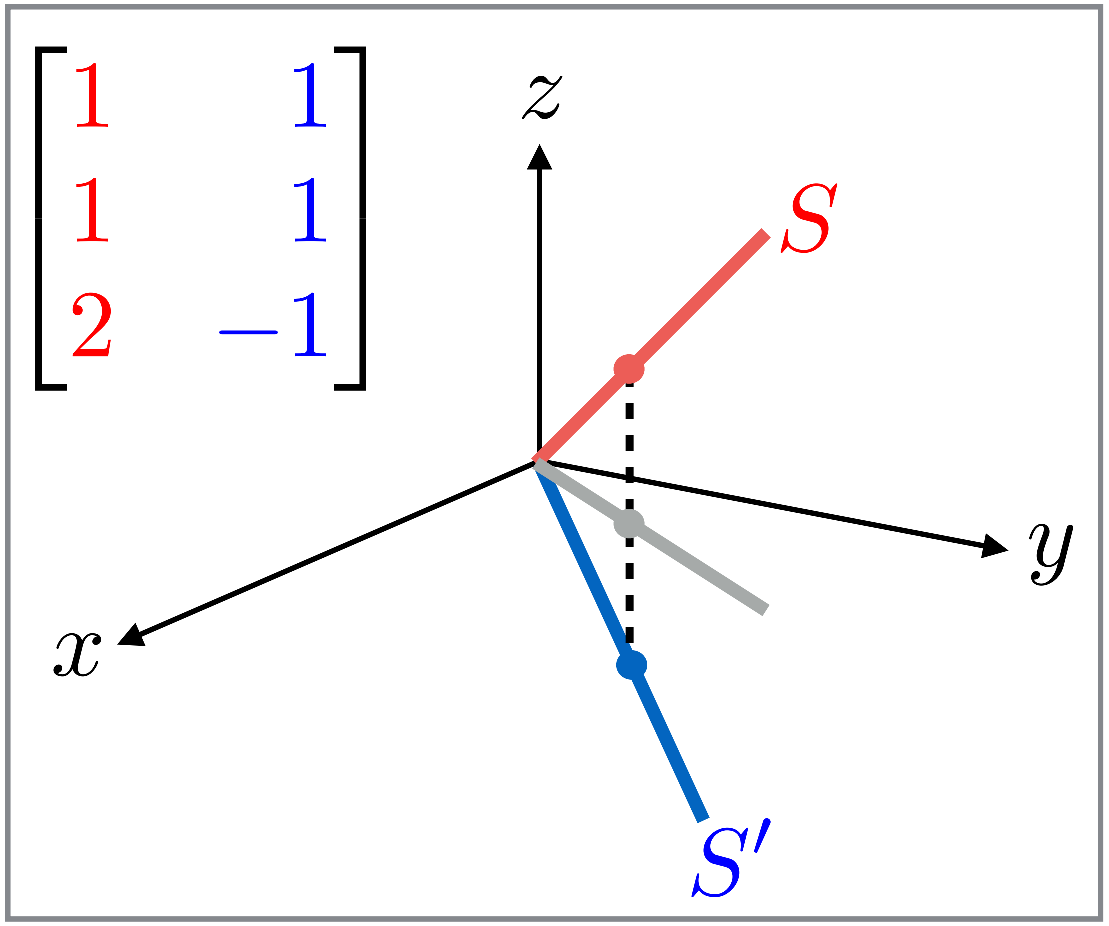
Complication

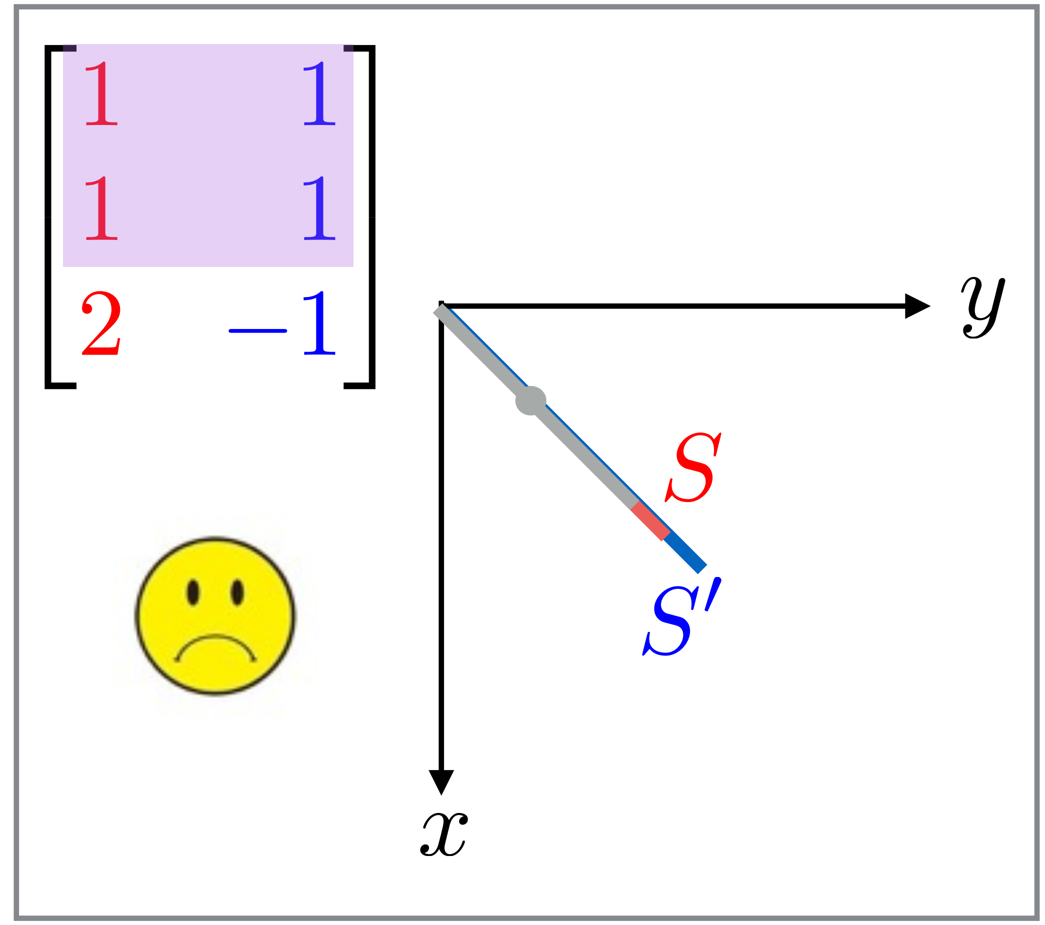
Two subspaces (even orthogonal) can appear identical if they are only observed on a subset of coordinates



Complication

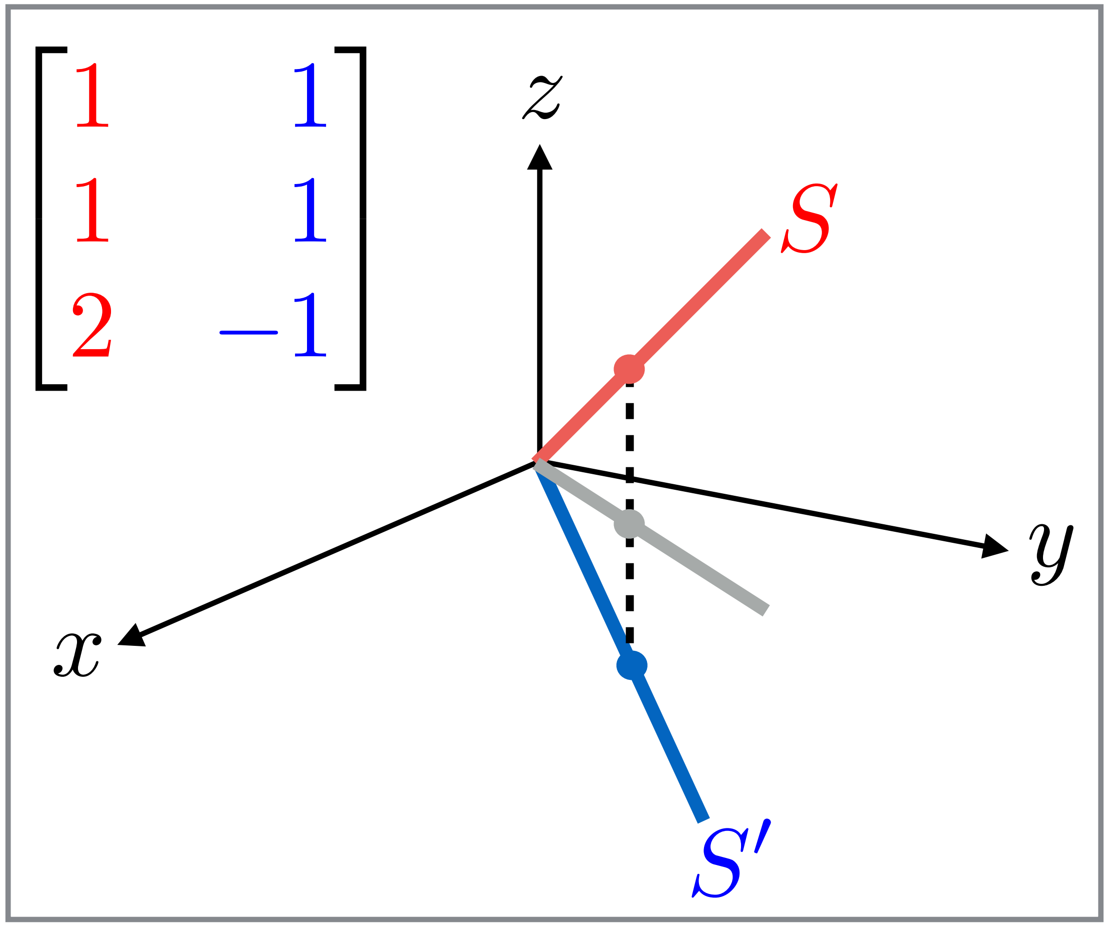
Two subspaces (even orthogonal) can appear identical if they are only observed on a subset of coordinates

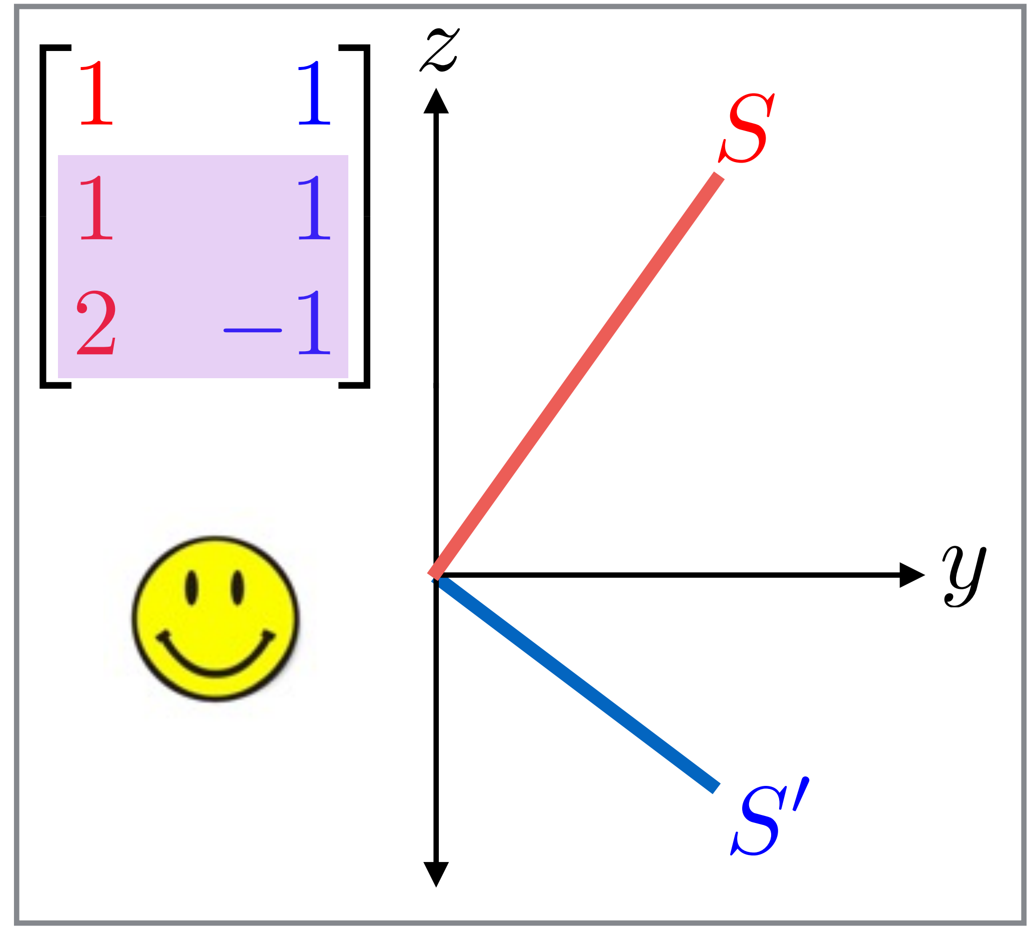




Complication

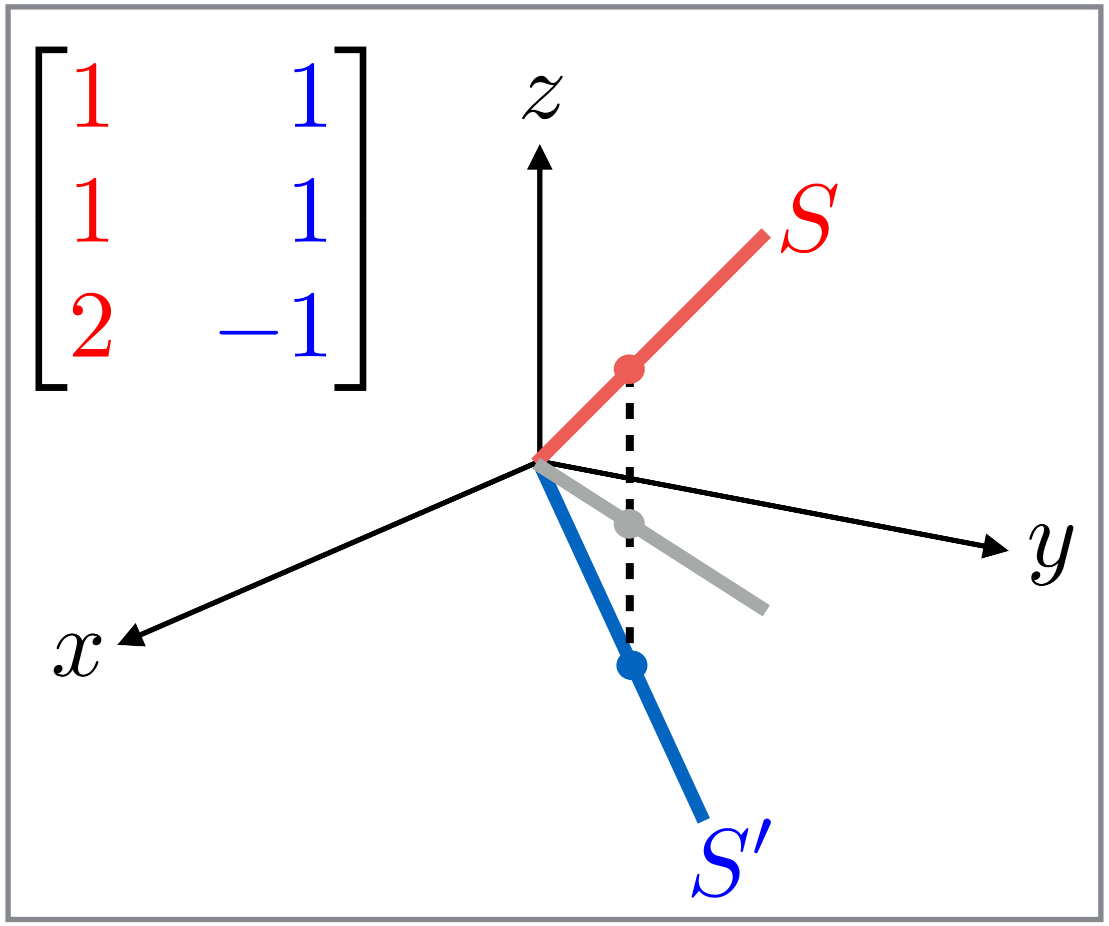
Two subspaces (even orthogonal) can appear identical if they are only observed on a subset of coordinates

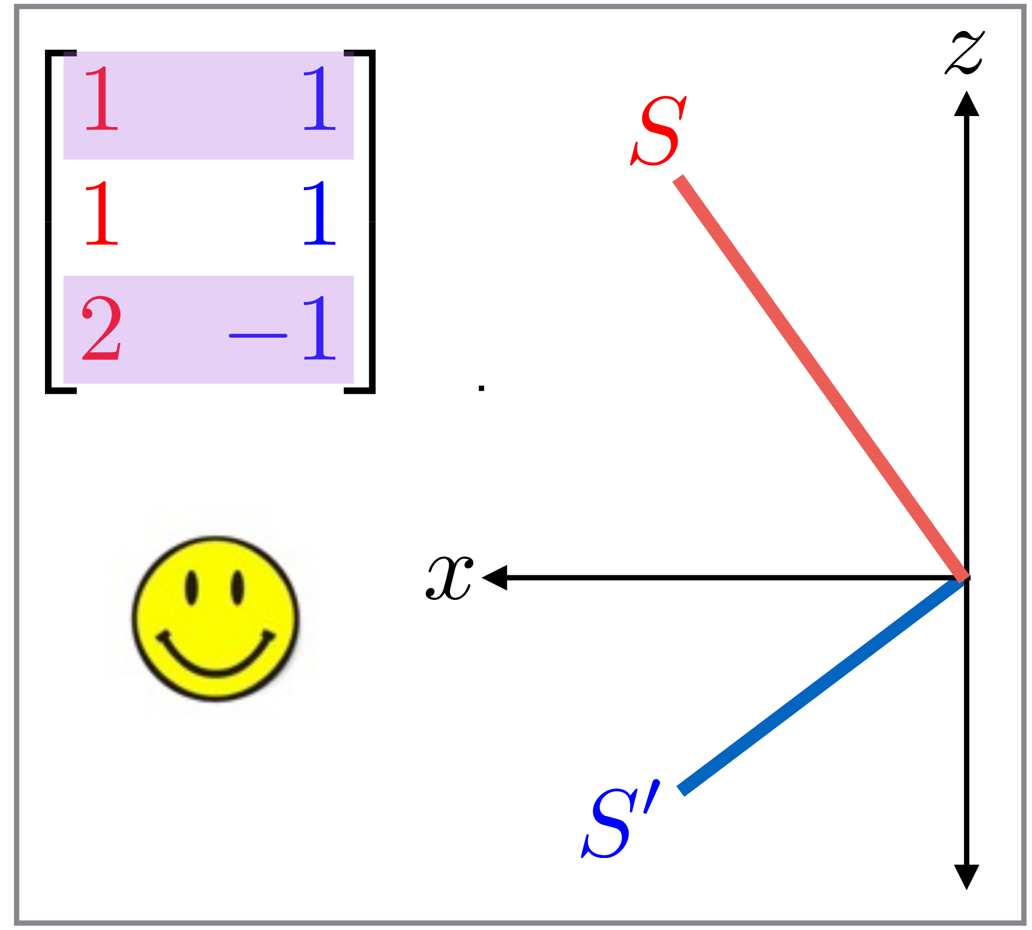




Fortunately

Not all subsets of coordinates are bad



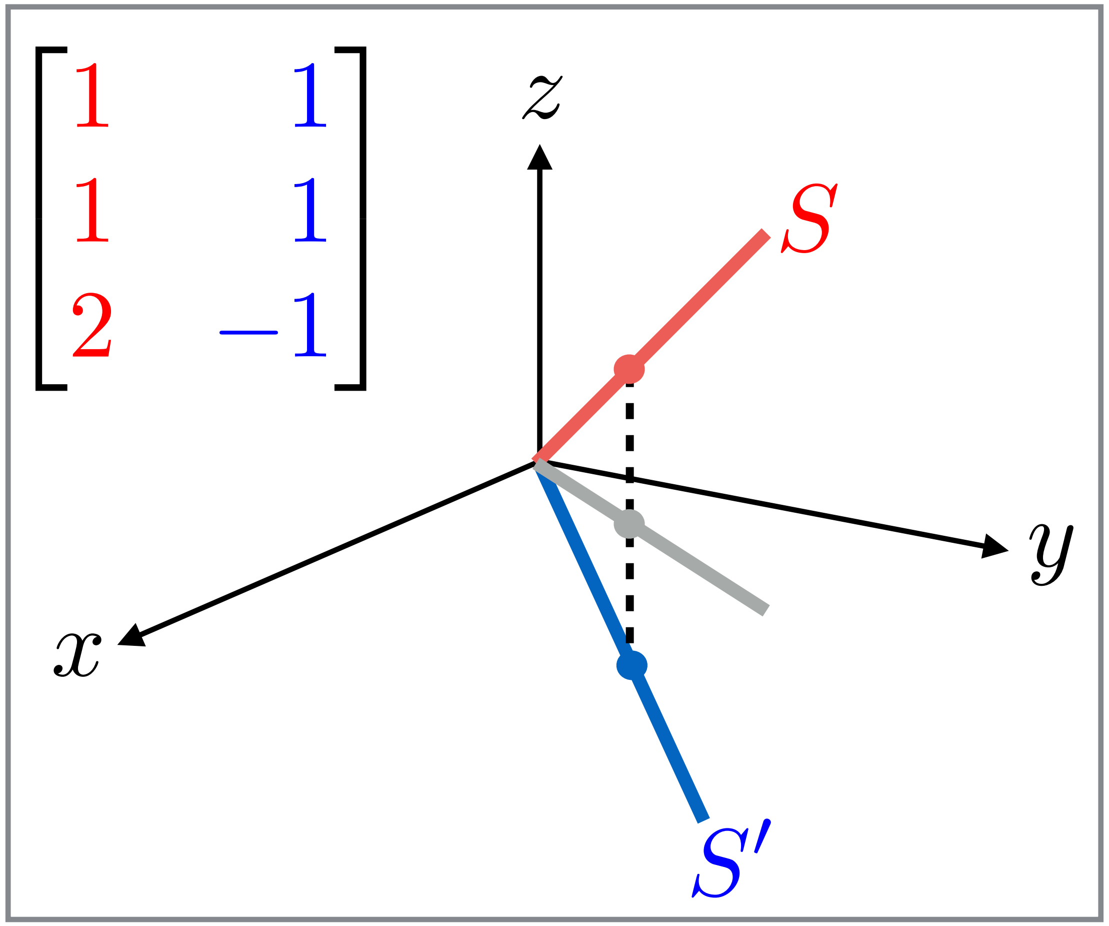


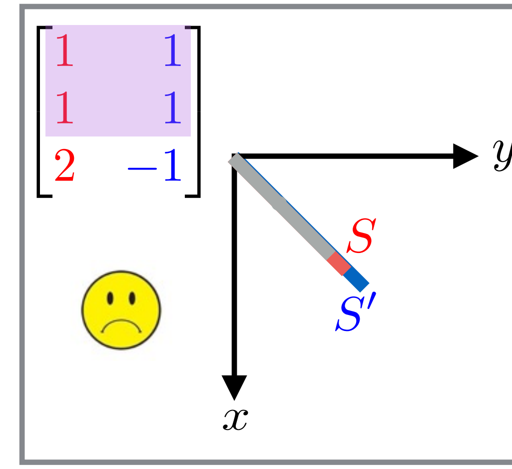
Fortunately

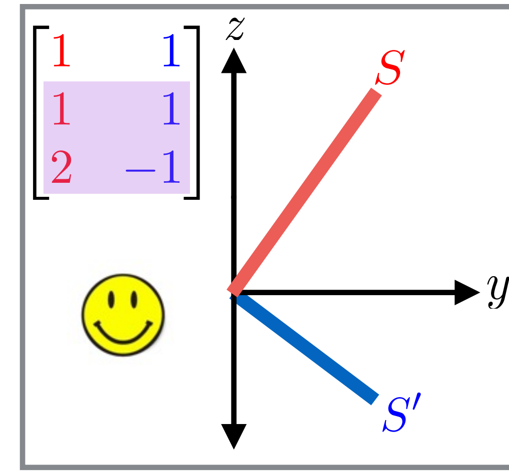
Not all subsets of coordinates are bad

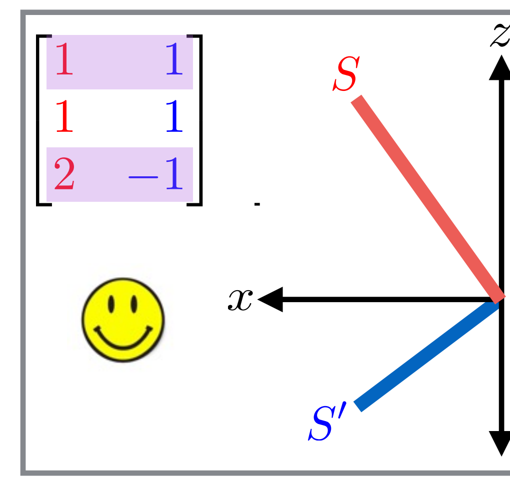
If we pick *the right* subsets of coordinates, we will be fine.

The catch: How do we know which are *the right* subsets?









First thing to ask

How many subsets of coordinates are good?
(This depends on the subspaces)

New measure of similarity

of subsets of $r+1$ coordinates where two subspaces differ

New measure of similarity

of subsets of $r+1$ coordinates where two subspaces differ

Definition 1. *Given* $S, S' \in \mathrm{Gr}(r, \mathbb{R}^d)$, *define the* partial coordinate discrepancy *between* S *and* S' *as:*

$$\delta(S, S') \;:=\; \frac{1}{\binom{d}{r+1}} \sum_{\boldsymbol{\omega} \in [d]^{r+1}} \mathbb{1}_{\{S_{\boldsymbol{\omega}} \neq S'_{\boldsymbol{\omega}}\}}.$$

New measure of similarity

\# of subsets of *r+1* coordinates where two subspaces differ

Dimension of subspaces

Definition 1. *Given* $S, S' \in \mathrm{Gr}(r, \mathbb{R}^d)$, *define the* partial coordinate discrepancy *between* S *and* S' *as:*

$$\delta(S, S') \;:=\; \frac{1}{\binom{d}{r+1}} \sum_{\boldsymbol{\omega} \in [d]^{r+1}} \mathbb{1}_{\{S_{\boldsymbol{\omega}} \neq S'_{\boldsymbol{\omega}}\}}.$$

New measure of similarity

\# of subsets of *r+1* coordinates where two subspaces differ

Dimension of subspaces

Ambient dimension

Definition 1. *Given* $S, S' \in \mathrm{Gr}(r, \mathbb{R}^d)$, *define the* partial coordinate discrepancy *between* S *and* S' *as:*

$$\delta(S, S') := \frac{1}{\binom{d}{r+1}} \sum_{\omega \in [d]^{r+1}} \mathbb{1}_{\{S_\omega \neq S'_\omega\}}.$$

New measure of similarity

\# of subsets of *r+1* coordinates where two subspaces differ

Dimension of subspaces

Ambient dimension

Definition 1. *Given* $S, S' \in \mathrm{Gr}(r, \mathbb{R}^d)$, *define the* partial coordinate discrepancy *between* S *and* S' *as:*

$$\delta(S, S') := \frac{1}{\binom{d}{r+1}} \sum_{\omega \in [d]^{r+1}} \mathbb{1}_{\{S_\omega \neq S'_\omega\}}.$$

Subsets of $\{1,\ldots,d\}$ with exactly $r+1$ elements

New measure of similarity

of subsets of $r+1$ coordinates where two subspaces differ

Dimension of subspaces

Ambient dimension

Definition 1. *Given* $S, S' \in \mathrm{Gr}(r, \mathbb{R}^d)$, *define the* partial coordinate discrepancy *between* S *and* S' *as:*

$$\delta(S, S') := \frac{1}{\binom{d}{r+1}} \sum_{\omega \in [d]^{r+1}} \mathbb{1}_{\{S_\omega \neq S'_\omega\}}.$$

Subsets of {*1*,…,*d*} with exactly *r+1* elements

Are subspaces equal on these *r+1* coordinates?

New measure of similarity

of subsets of *r+1* coordinates where two subspaces differ

Example

$$\begin{array}{cc} \color{red}{S} & \color{blue}{S'} \\ \begin{bmatrix} \color{red}{1} \\ \color{red}{1} \\ \color{red}{1} \\ \color{red}{1} \end{bmatrix} & \begin{bmatrix} \color{blue}{1} \\ \color{blue}{1} \\ \color{blue}{-1} \\ \color{blue}{-1} \end{bmatrix} \end{array}$$

Example

$$\begin{array}{cc} \color{red}{S} & \color{blue}{S'} \\ \begin{bmatrix} \color{red}{1} \\ \color{red}{1} \\ \color{red}{1} \\ \color{red}{1} \end{bmatrix} & \begin{bmatrix} \color{blue}{1} \\ \color{blue}{1} \\ \color{blue}{-1} \\ \color{blue}{-1} \end{bmatrix} \end{array}$$

$$\boxed{\delta(\color{red}{S}, \color{blue}{S'}) = \frac{4}{6}}$$

Example

$$\begin{array}{cc} S & S' \\ \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} & \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix} \end{array}$$

$$\boxed{\delta(S, S') = \frac{4}{6}}$$

# of good combinations of *r+1* coordinates

Example

$$\begin{array}{cc} S & S' \\ \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} & \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix} \end{array}$$

# of good combinations
of *r+1* coordinates

$$\boxed{\delta(S, S') = \frac{4}{6}}$$

Example

$$\begin{array}{cc} S & S' \\ \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} & \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix} \end{array}$$

# of good combinations
of *r+1* coordinates

$$\boxed{\delta(S, S') = \frac{4}{6}}$$

Example

$$\begin{array}{cc} S & S' \\ \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} & \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix} \end{array}$$

$$\boxed{\delta(S, S') = \frac{4}{6}}$$

# of good combinations of *r+1* coordinates

# of total combinations of *r+1* coordinates

Example

$$\begin{array}{cc} \color{red}{S} & \color{blue}{S'} \\ \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} & \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix} \end{array}$$

# of good combinations of *r+1* coordinates

$$\delta(S, S') = \frac{4}{6}$$

# of total combinations of *r+1* coordinates

Probability that 2 subspaces are different on *r+1* coordinates chosen randomly

What does this mean?

Depending on the subspaces,

there may be way too many bad subsets!

What does this mean?

Depending on the subspaces,
there may be way too many bad subsets!

Lucky break!

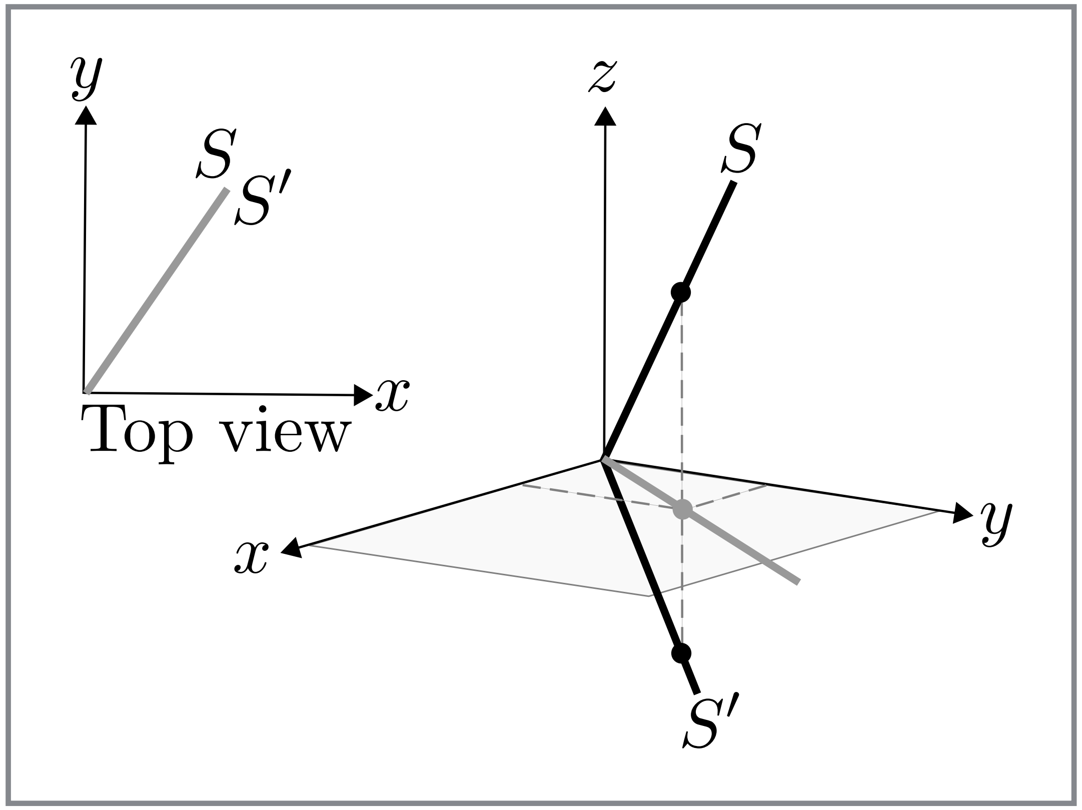
If we rotate subspaces randomly,
all subsets will be good!

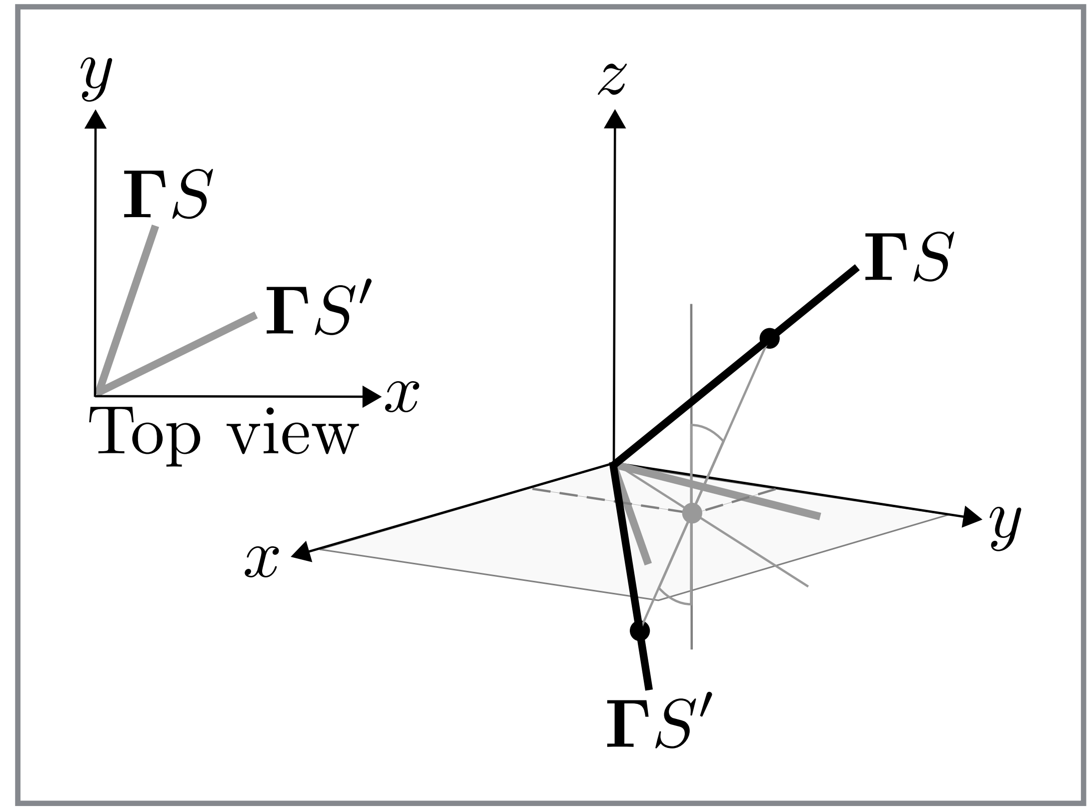
What does this mean?

Depending on the subspaces,
there may be way too many bad subsets!

Lucky break!

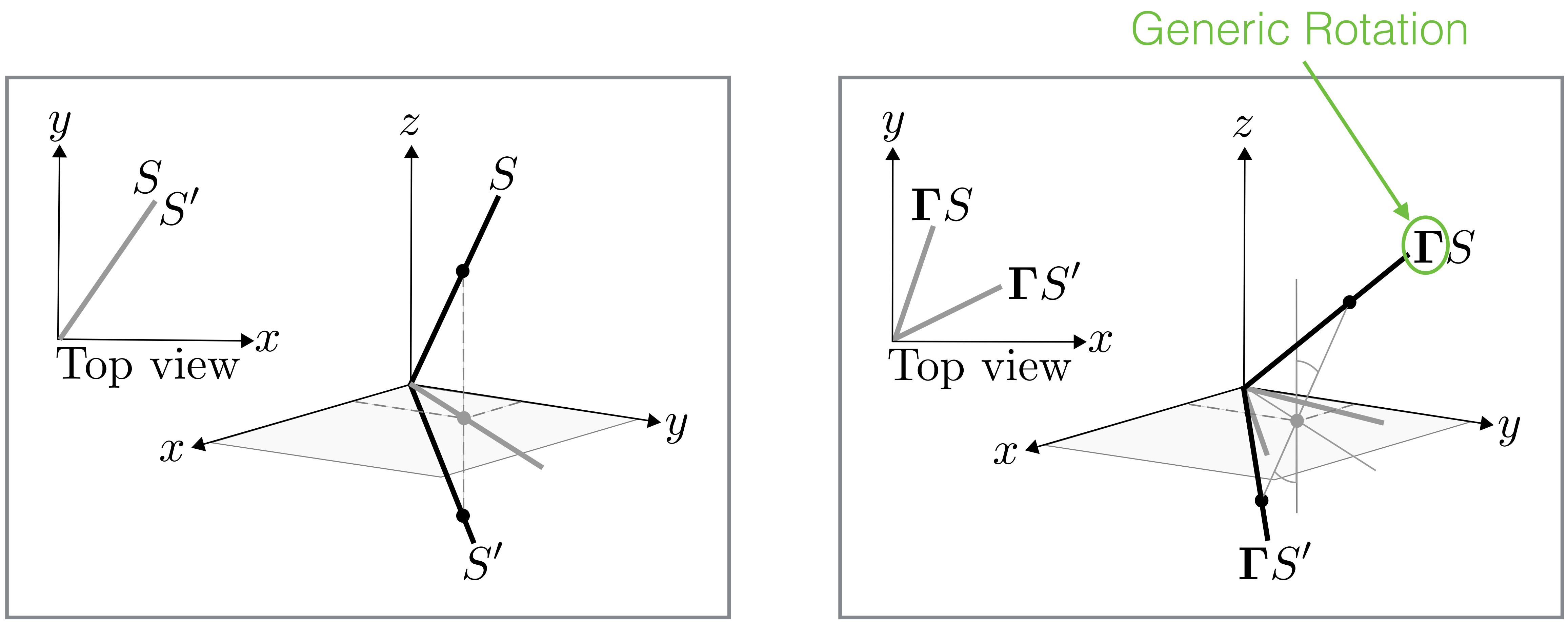
If we rotate subspaces randomly,
all subsets will be good!





Lucky break!

If we rotate subspaces randomly,
all subsets will be good!



Lucky break!

If we rotate subspaces randomly,
all subsets will be good!

What does this mean?

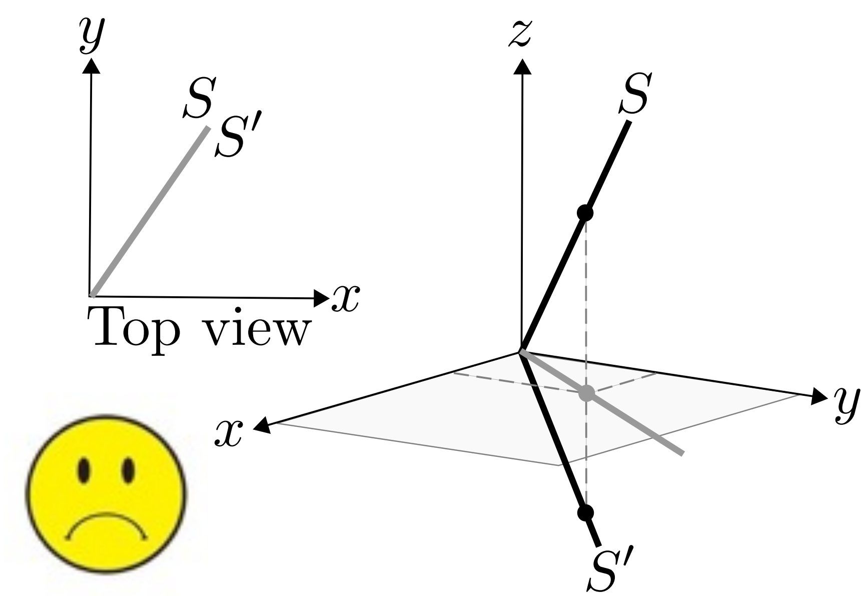
We can rotate, subsample and cluster

What does this mean?

We can rotate, subsample and cluster

Original Data

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

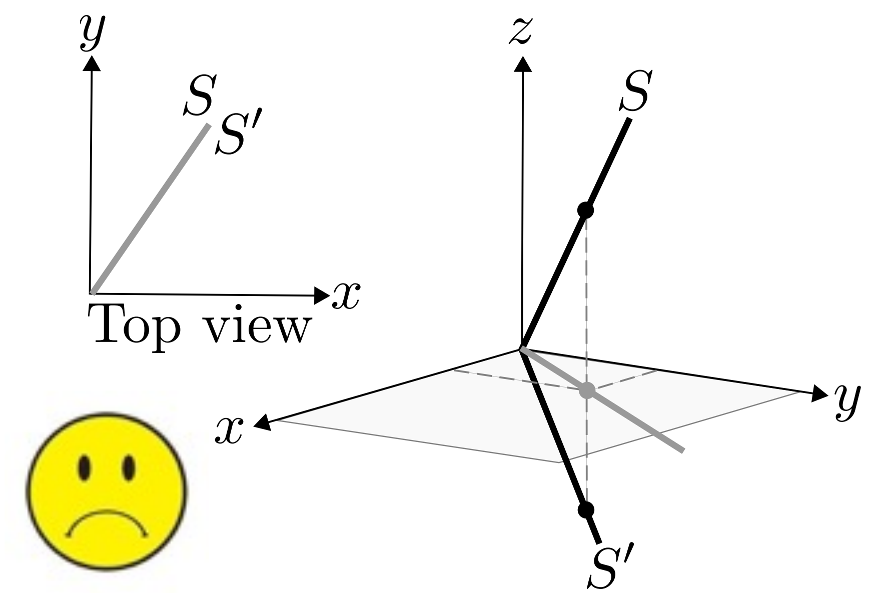


What does this mean?

We can rotate, subsample and cluster

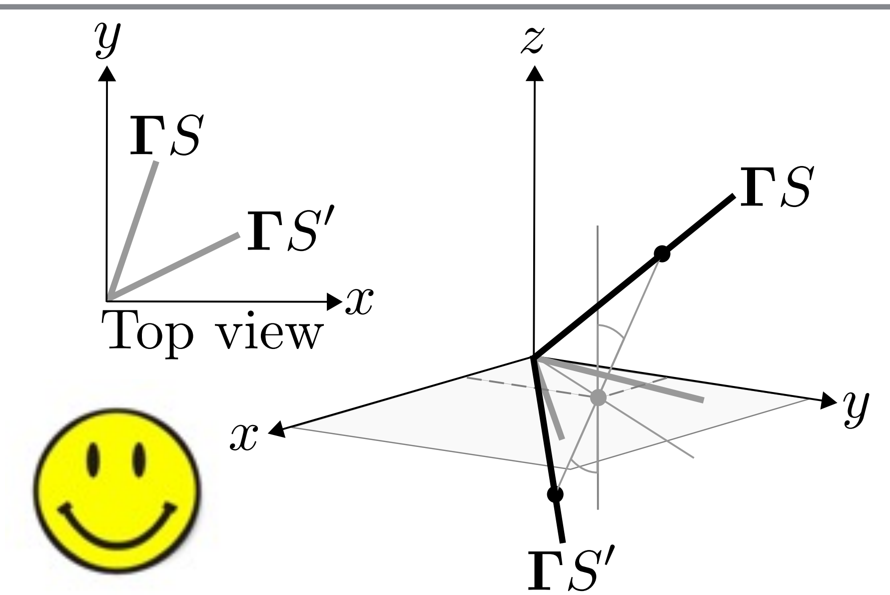
Original Data

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



Rotated Data

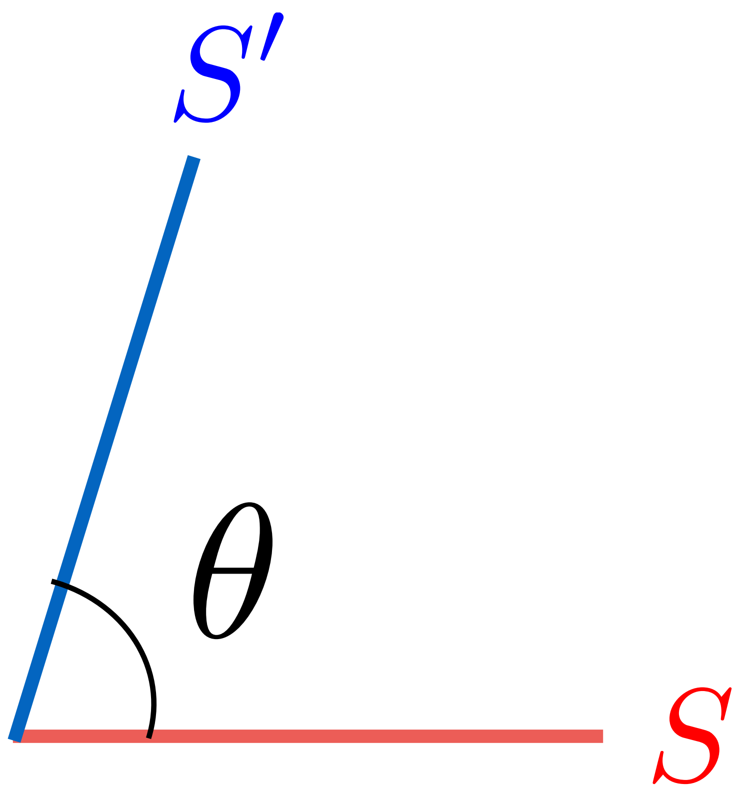
$$\mathbf{\Gamma}\begin{bmatrix} 1 & 4 & 1 & 3 & 3 & 1 & 2 & 1 & 2 & 1 \\ 2 & 4 & 2 & 6 & 3 & 2 & 2 & 2 & 4 & 1 \\ 3 & 4 & 3 & 9 & 3 & 3 & 2 & 3 & 6 & 1 \\ 1 & 8 & 1 & 3 & 6 & 1 & 4 & 1 & 2 & 2 \\ 2 & 8 & 2 & 6 & 6 & 2 & 4 & 2 & 4 & 2 \\ 3 & 8 & 3 & 9 & 6 & 3 & 4 & 3 & 6 & 2 \end{bmatrix}$$



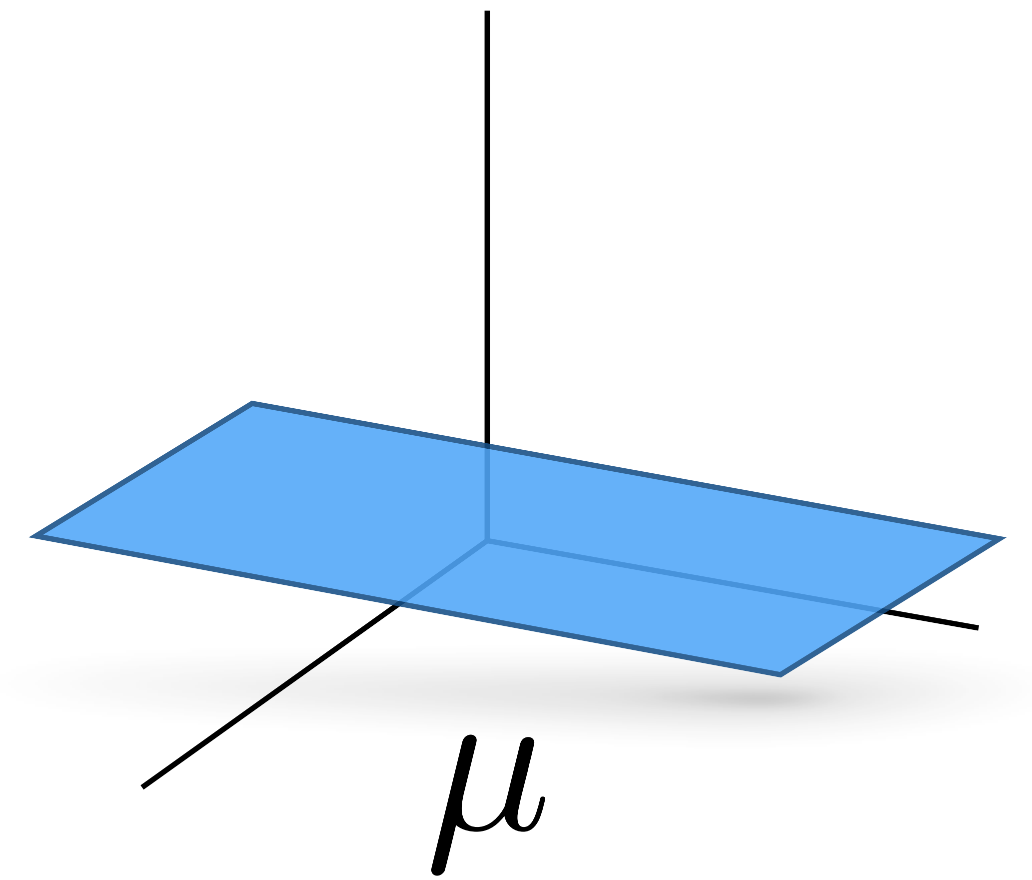
What does this mean?

We can rotate, subsample and cluster

A little more about δ



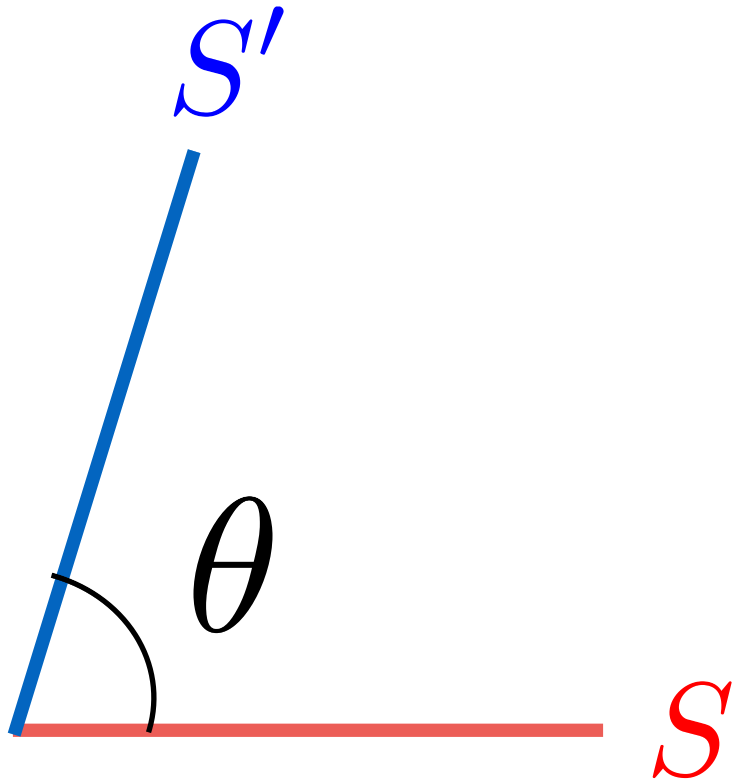
Principal Angle
Distance



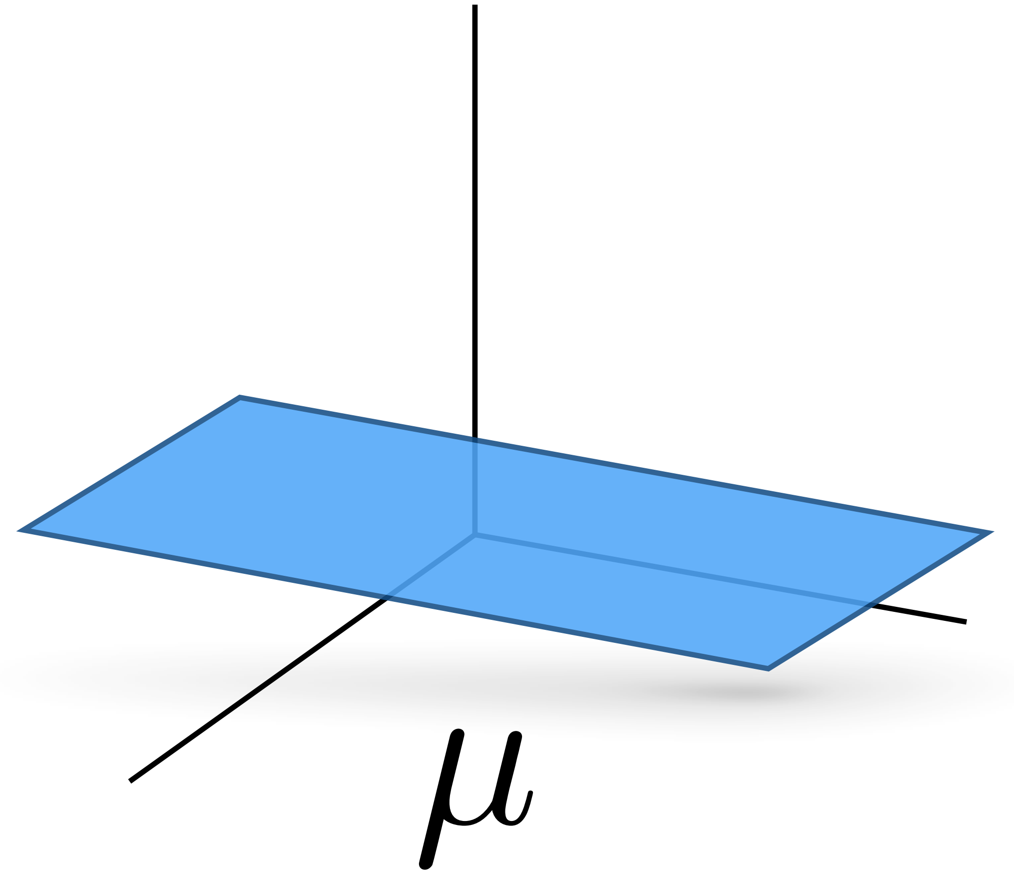
Coherence

A little more about δ

Long story short: none implies the other.



Principal Angle
Distance



Coherence

A little more about δ

Long story short: none implies the other.

What’s next?

- δ is an *all or nothing* metric.

What's next?

- δ is an *all or nothing* metric.

$$\delta(S, S') := \frac{1}{\binom{d}{r+1}} \sum_{\boldsymbol{\omega} \in [d]^{r+1}} \mathbb{1}_{\{S_{\boldsymbol{\omega}} \neq S'_{\boldsymbol{\omega}}\}}.$$

kind of an ℓ_0 norm.

What's next?

- δ is an *all or nothing* metric.

$$\delta(S, S') := \frac{1}{\binom{d}{r+1}} \sum_{\boldsymbol{\omega} \in [d]^{r+1}} \mathbb{1}_{\{S_{\boldsymbol{\omega}} \neq S'_{\boldsymbol{\omega}}\}}.$$

kind of an ℓ_0 norm.

- Can we come up with more practical metrics?

kind of an ℓ_1 norm.

What's next?

Thank you.