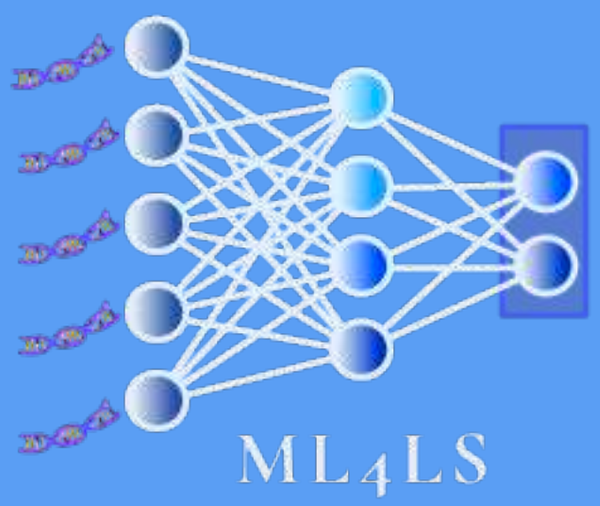
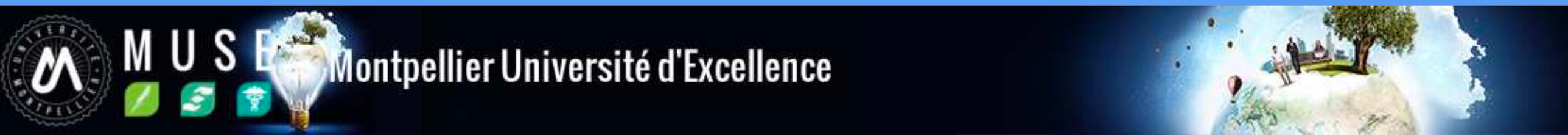
Applied Math
UNIVERSITY OF COLORADO **BOULDER**

Introduction to Compressed Sensing and Matrix Completion

Stephen Becker
University of Colorado Boulder, USA, Dept. of Applied Math
visitor at: **Institut Montpelliérain Alexander Grothendieck**



Machine Learning for Life Sciences
15-17 Nov 2022 Montpellier
KIM Data & Life Sciences



Outline

Goal of the talk

- Introduce some new tools (compressed sensing ...)
- A few examples of applications to life science
- Inspire you to use them in more applications

Outline

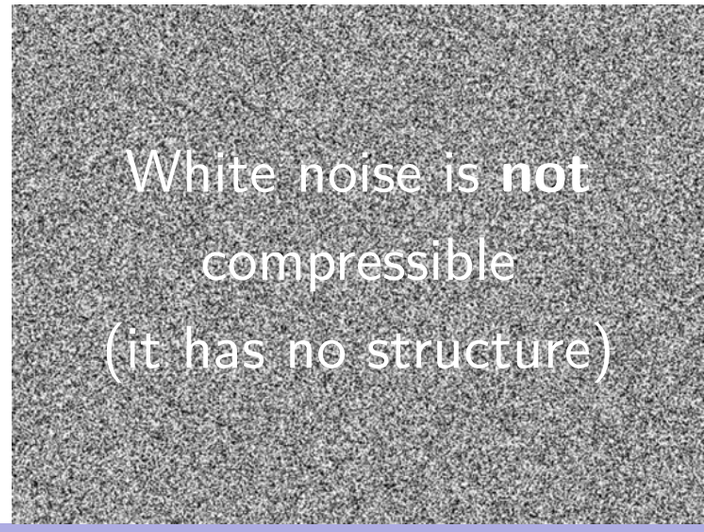
- Compressed sensing
  - Motivating sparsity
  - Theorems
  - Examples
  - Smashed Filter
- Matrix completion
  - Examples
  - Robust PCA
    - Examples
- Where are we today?
  - Tensors
  - Sketching
  - Computational paradigm

Compression

- Which picture is the JPEG (8 kB), which is the PNG (60 kB)?

- Compression is everywhere
  - images, audio, video, documents
  - **interesting** signals have <u>structure</u>

Structure: frequency, wavelets, curvelets…

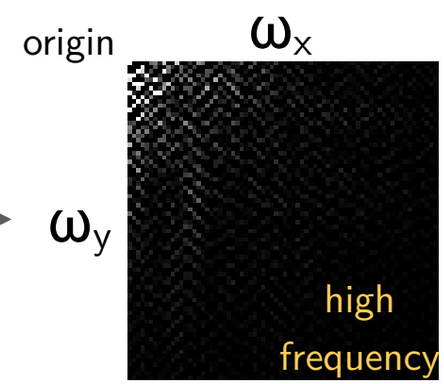


Compression = sparsity

Image

2D DCT

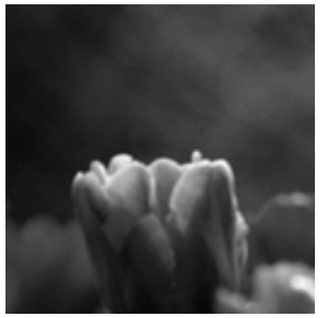
Fourier domain



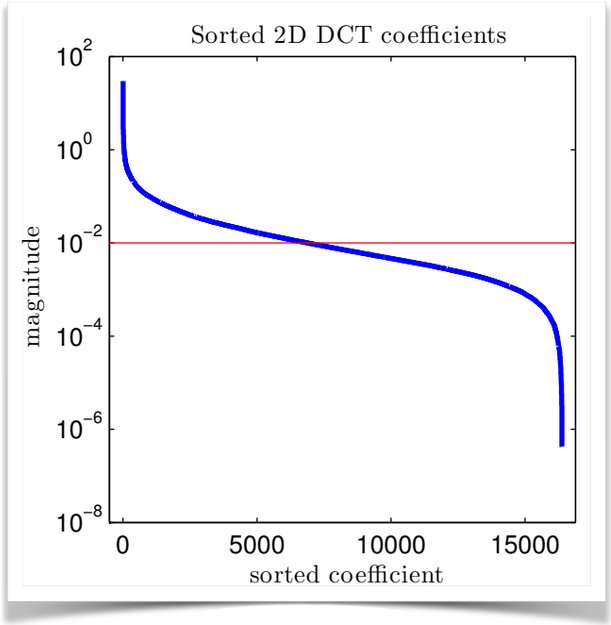
Truncation,

inverse DCT

Reconstruction



Sparsity



(Actual JPEG does DCT on 8x8 patches, in HSV coordinates, and quantizes)

Exploiting sparsity

- Our picture had less **information** than **data**.

  We used this fact to **compress** it

  Could we just collect less data in the first place?

Exploiting sparsity

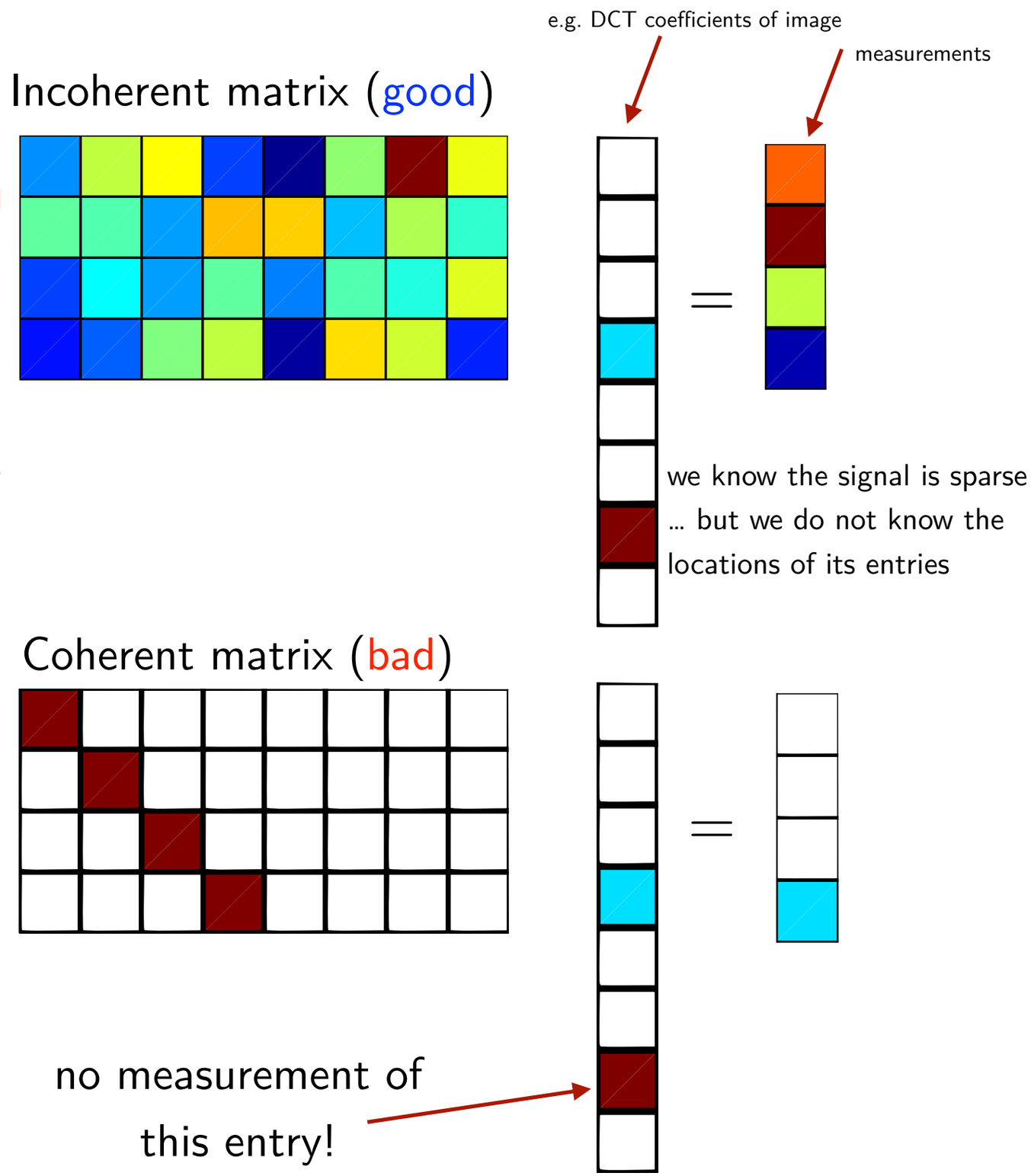
- Our picture had less **information** than **data**.

  We used this fact to **compress** it

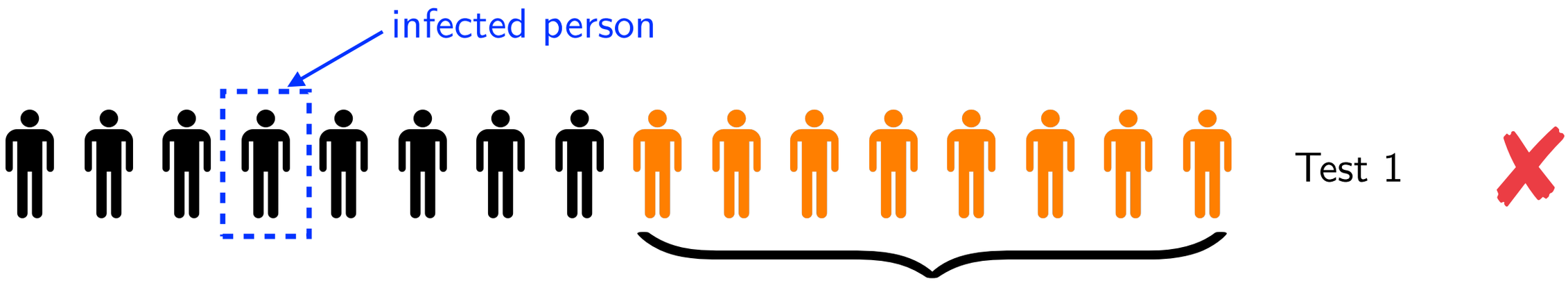
  Could we just collect less data in the first place?

- **Yes**! But must be careful!

- Require **incoherence**. "Random" measurements are good.



Example of "incoherence": group testing



pool all these blood samples together,
apply expensive test just once

means person not included in test

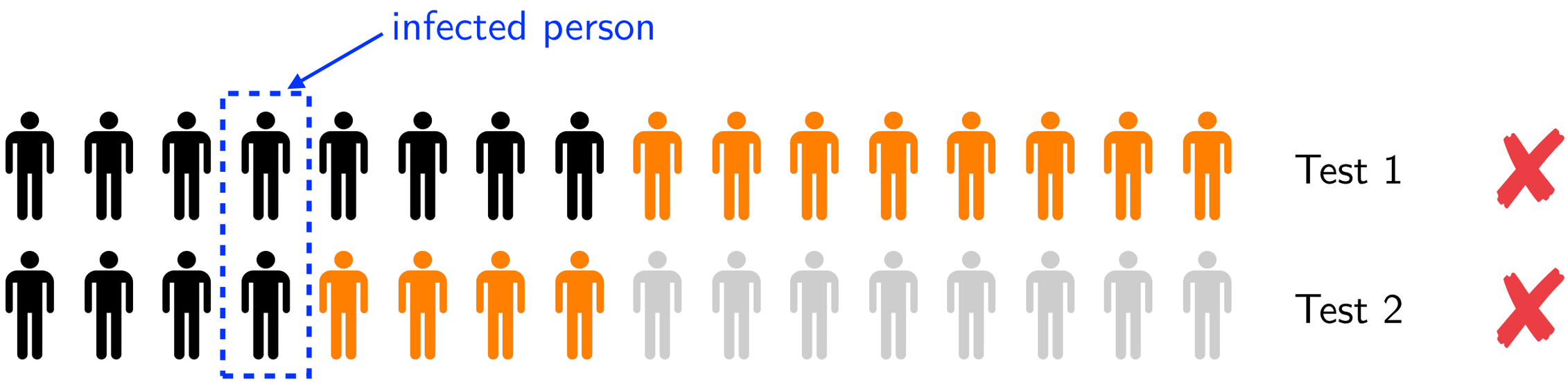
means person included in test

Introduced by Dorfman, 1943 for testing solders for syphilis

Pooling blood samples is a good idea if only a few soldiers are infected
(and we assume test is highly sensitive)

e.g., if we know that exactly 1 soldier out of n is infected,
need only $\log_2(n)$ tests instead of n

Example of "incoherence": group testing

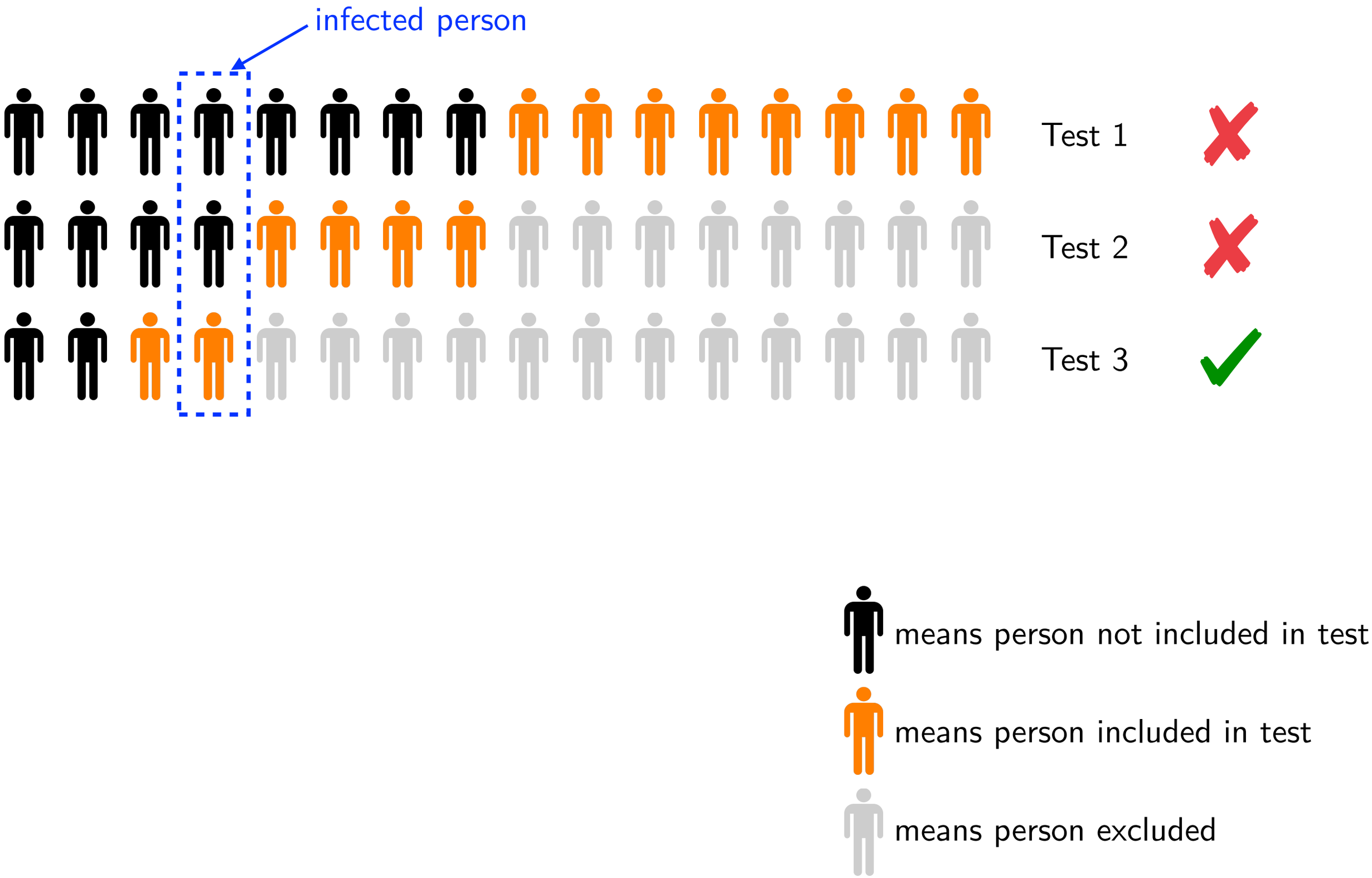


means person not included in test

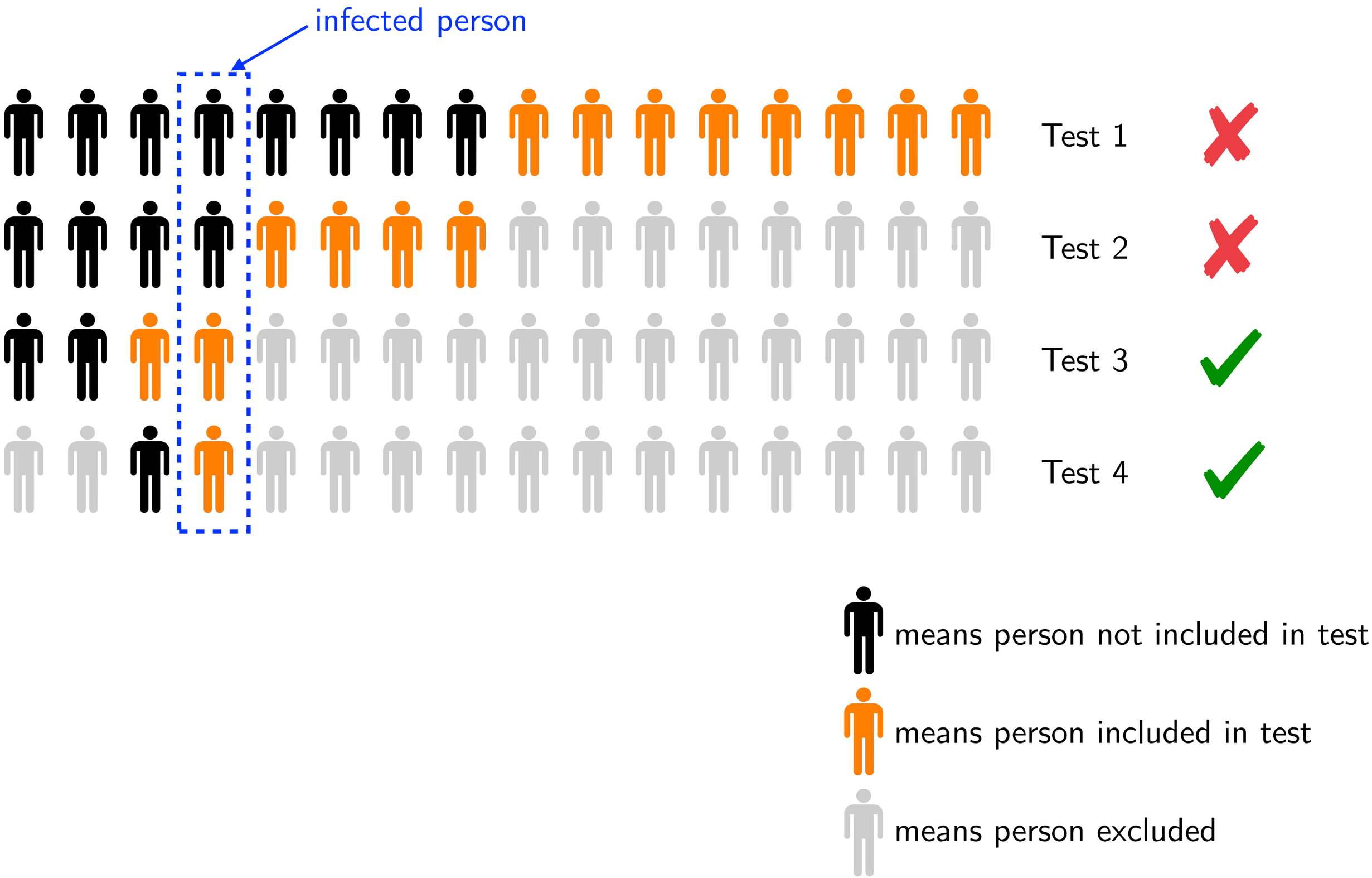
means person included in test

means person excluded

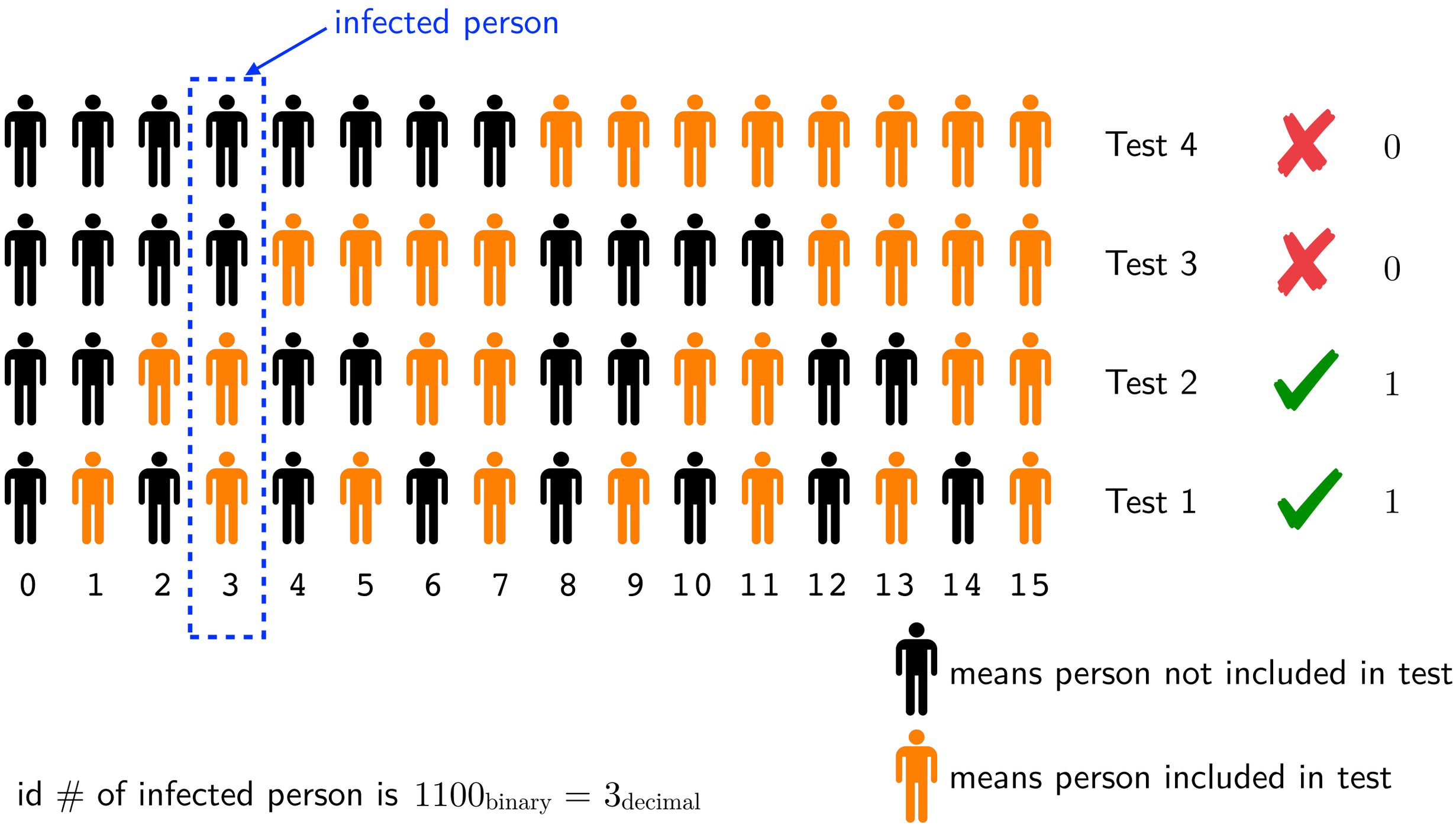
Example of "incoherence": group testing



Example of "incoherence": group testing



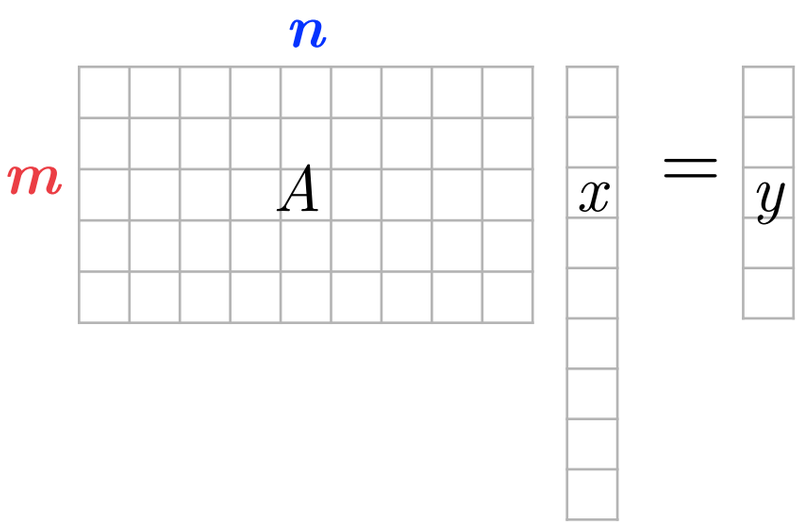
Group testing can even be non-adaptive



id # of infected person is $1100_{\text{binary}} = 3_{\text{decimal}}$

(we're exploiting the fact that we know exactly one person is infected)

How many measurements do we need?



$y = Ax_{\text{true}} + z$

or just

$y = Ax_{\text{true}}$ for now

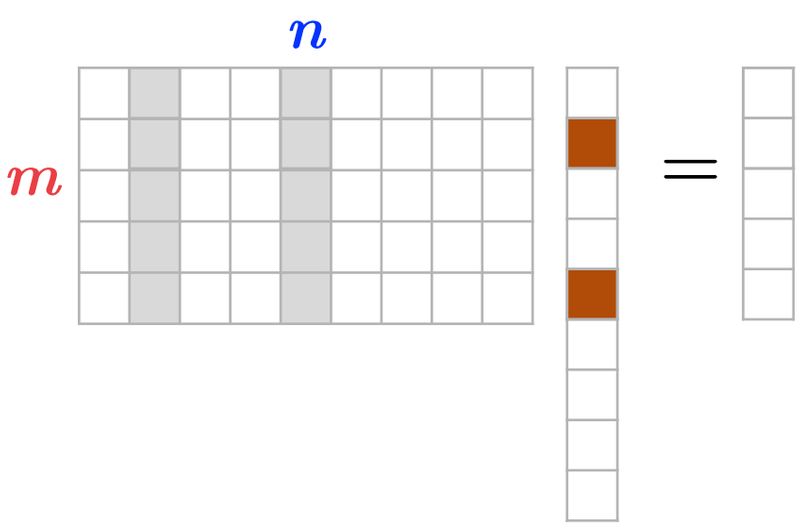
With no assumptions on x, we need $m \geq n$

Why? There could be two different inputs that give the same output, so from looking at output alone, we can never determine which input was used

Mathematically, we need the measurement operator to be injective (aka one-to-one).

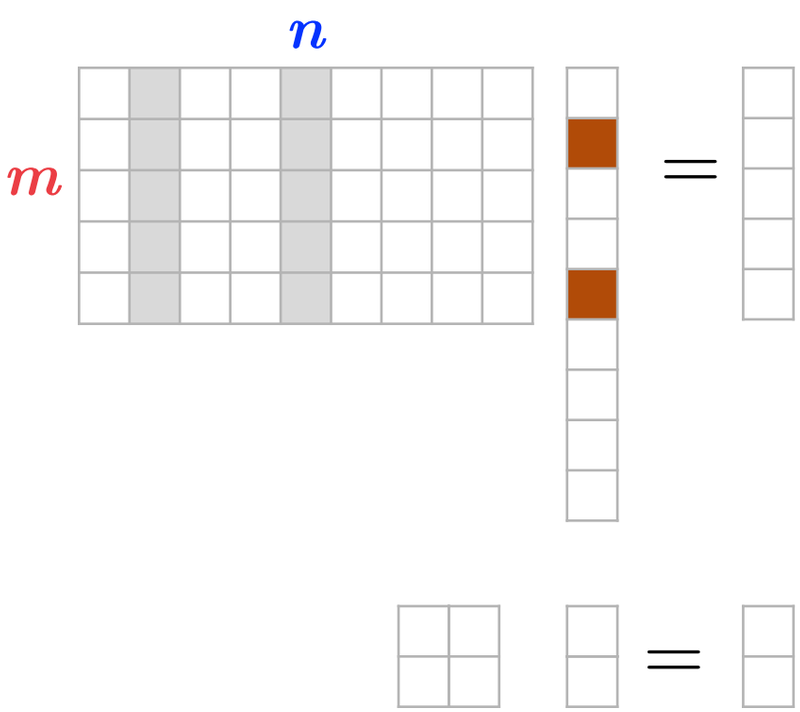
Since we have linear measurements, this reduces to requiring the matrix to be full column rank

How many measurements do we need?



Now assume x has only k (of n possible) nonzero entries

How many measurements do we need?



Now assume x has only k (of n possible) nonzero entries

If we **know the positions** of those nonzero entries...

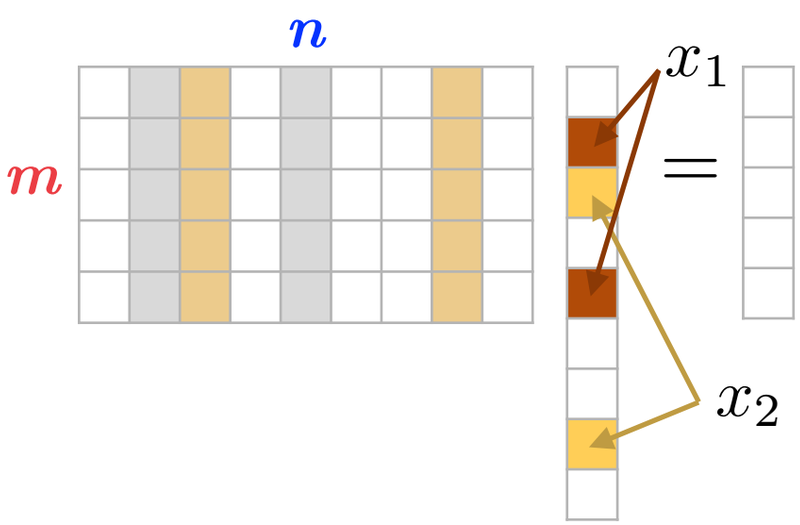
...need just $m \geq k$

(since a square matrix can be nonsingular)

Aside: can **all** possible square sub-matrices be nonsingular?

- **Yes.** For example, a Gaussian, since columns are in general position
- ... but explicitly constructing/checking is not easy

How many measurements do we need?



Now assume x has only k (of n possible) nonzero entries

If we **don't know** the positions of those nonzero entries...

Suppose there are *two* k sparse signals, x_1 and x_2, that give the same measurement

$$Ax_1 = y$$
$$Ax_2 = y$$

If $x_1 \neq x_2$, then we cannot recover (which one would we choose?)

Defining $x = x_1 - x_2$, $Ax = 0$ and x has up to $2k$ nonzero entries

So if the m by $2k$ submatrix of A is invertible, $x = 0$ is the only solution, thus $x_1 = x_2$

...hence we need $m \geq 2k$

For some matrices (e.g., Gaussian, or anything with "spark" greater than $2k$) this is also <u>sufficient</u>

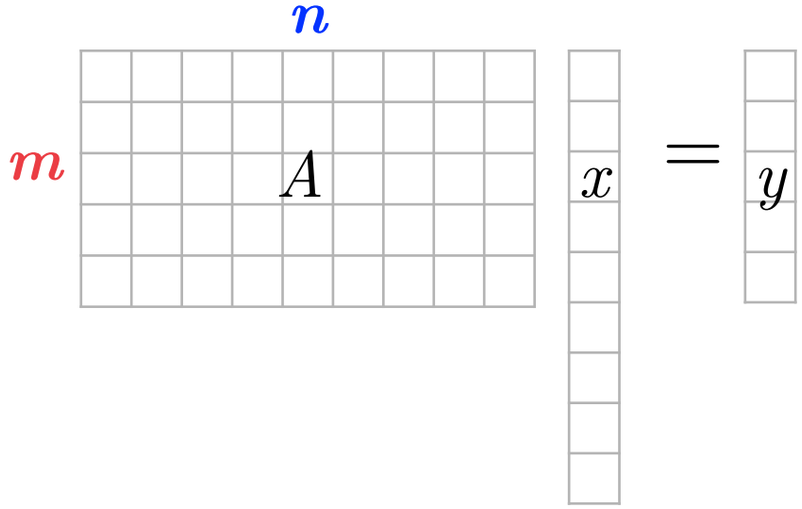
Suggests the algorithm: **check all possible submatrices with k columns**

Using optimization to find a signal

measurement signal

$$y = Ax$$

If A is underdetermined, *which* solution to pick?



Using optimization to find a signal

measurement signal

$$y = Ax$$

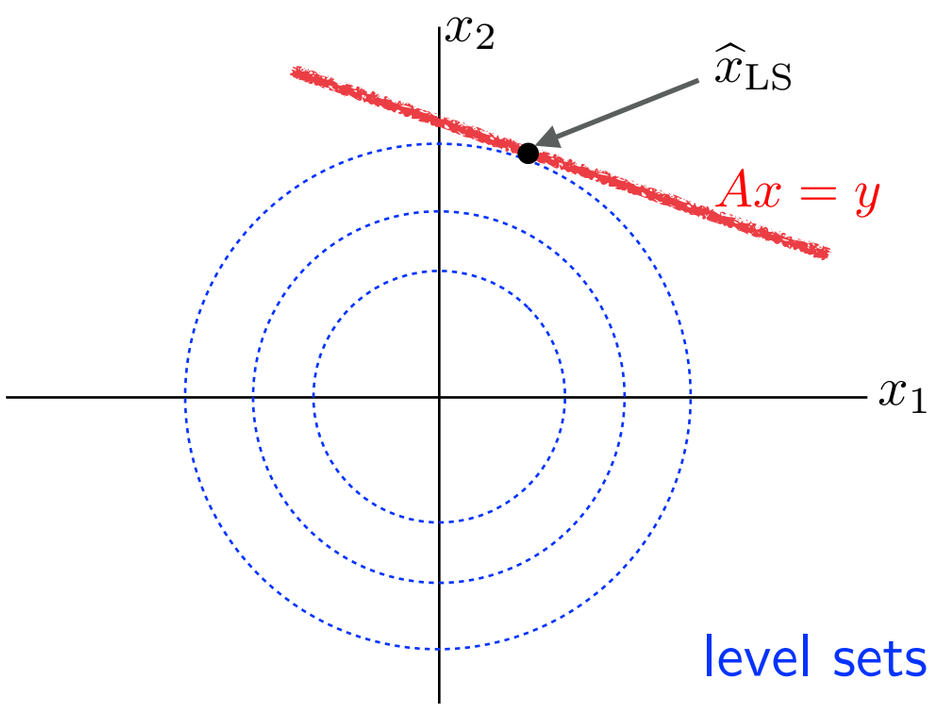
If A is underdetermined, *which* solution to pick?

$$\|x\|_2 = \sqrt{\sum_{i=1}^{n} |x_i|^2}$$

Conventional choice:

$$\widehat{x}_{\mathrm{LS}} = \underset{x}{\operatorname{argmin}} \ \|x\|_2 \quad \text{such that} \quad Ax = y$$

"least-squares"



level sets of the norm in blue

Using optimization to find a signal

measurement signal

$$y = Ax$$

If A is underdetermined, *which* solution to pick?

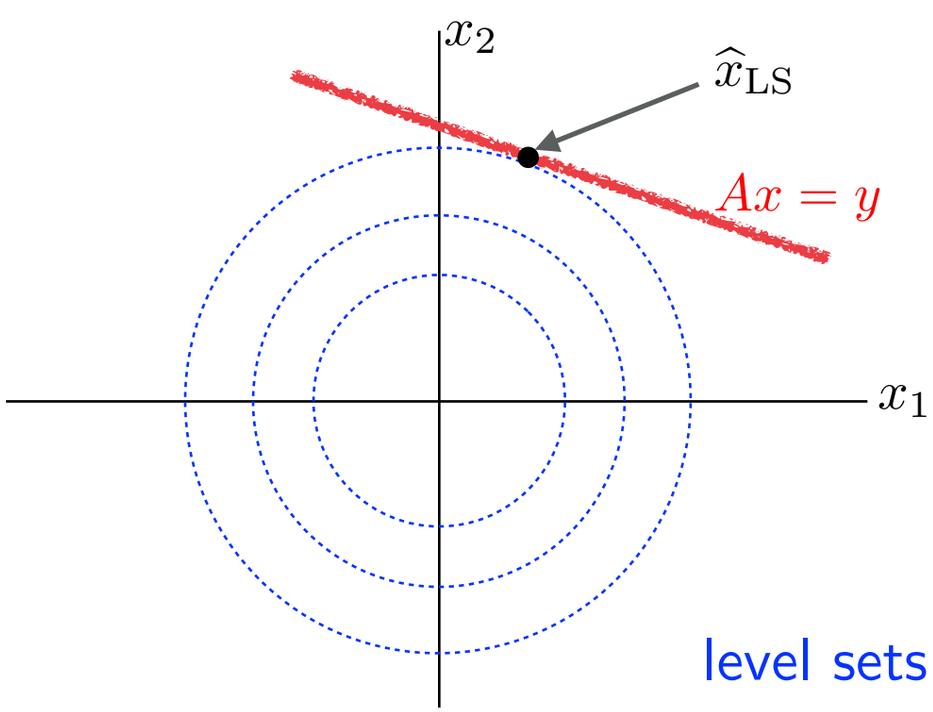
"ℓ_1 -norm promotes sparsity"

$$\|x\|_2 = \sqrt{\sum_{i=1}^{n} |x_i|^2}$$

Conventional choice:

$$\widehat{x}_{\mathrm{LS}} = \underset{x}{\operatorname{argmin}} \ \|x\|_2 \quad \text{such that} \quad Ax = y$$

"least-squares"

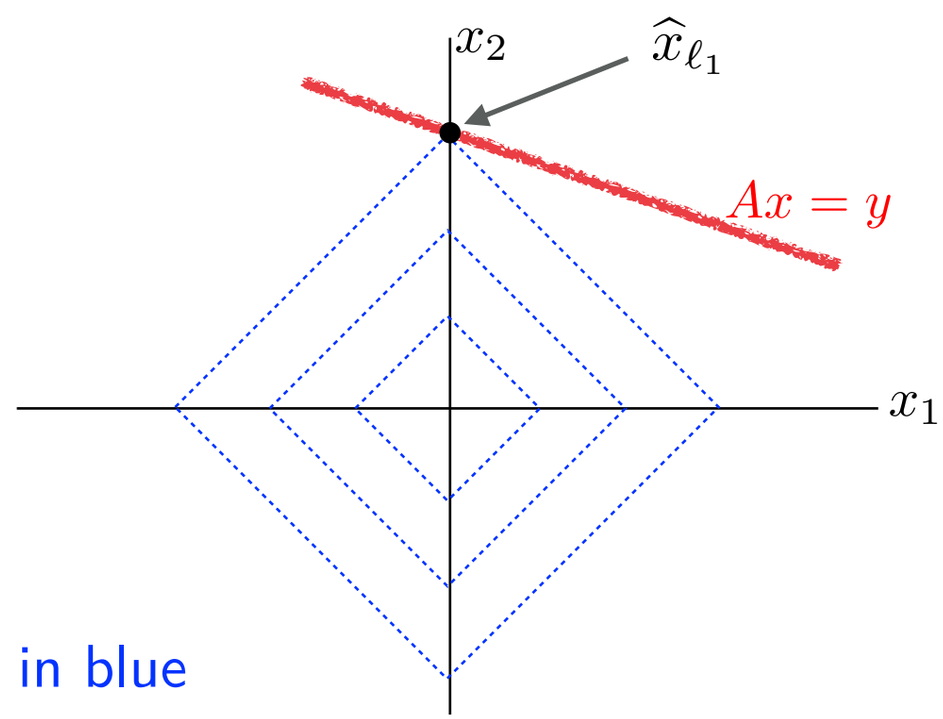


$$\|x\|_1 = \sum_{i=1}^{n} |x_i|$$

But if we want sparsity, try

$$\widehat{x}_{\ell_1} = \underset{x}{\operatorname{argmin}} \ \|x\|_1 \quad \text{such that} \quad Ax = y$$

"basis-pursuit"



level sets of the norm in blue

Compressed Sensing

So for $m = 2k$, we can check *every subset of size* k to find correct one

Complexity is thus $\binom{n}{k} = \frac{n!}{k!(n-k)!} \approx n^k$ (theoretically possible, but *computationally intractable*)

* Aside: if you can bound the magnitude of the coefficients, then this brute-force algorithm can be greatly improved by using integer programming techniques.

Compressed Sensing

So for $m = 2k$, we can check *every subset of size* k to find correct one

Complexity is thus $\binom{n}{k} = \frac{n!}{k!(n-k)!} \approx n^k$ (theoretically possible, but *computationally intractable*)

Contrast this to **basis pursuit** (this is a linear program, i.e., *tractable*)

$$\min_x \|x\|_1 \quad \text{subject to} \quad Ax = b$$

Compressed sensing (Candès, Romberg, Tao, Donoho, 2004)

Suppose x_0 is k-sparse and $z{=}0$, then...

Theorem

Let $A \in \mathbb{R}^{m \times n}$ have i.i.d. scaled Gaussian/Bernoulli entries and

$$m \geq C \cdot k \cdot \log(n/k)$$

for some constant C independent of n. Then with overwhelming probability, **basis pursuit** gives an estimator $\hat{x}$ such that $\hat{x} = x_0$.

and if we have noise, solve this variant:

$$\min_x \|x\|_1 \quad \text{subject to} \quad \|Ax - b\|_2 \leq \epsilon$$

Summary so far

For sparse recovery, there are two kinds of limits:

Information limits ⟶ $x + y + z = 3 \quad x - y = 4z = 2 \quad z = ?$

and

Computation limits ⟶ What are the prime factors of 5,569,145,346,157,157,894,612,234,343 ?

Before compressed sensing, the limits didn't match, e.g.,

we were limited by computational power

Ingredients for a compressed sensing problem

(1) Signal must be **sparse**

(may need to change the basis)

(2) Take **incoherent** measurements

(Gaussian measurements often impractical;

measurements must be incoherent with the chosen basis)

... plus nearly 2 decades of improvements & variations & solvers

Outline

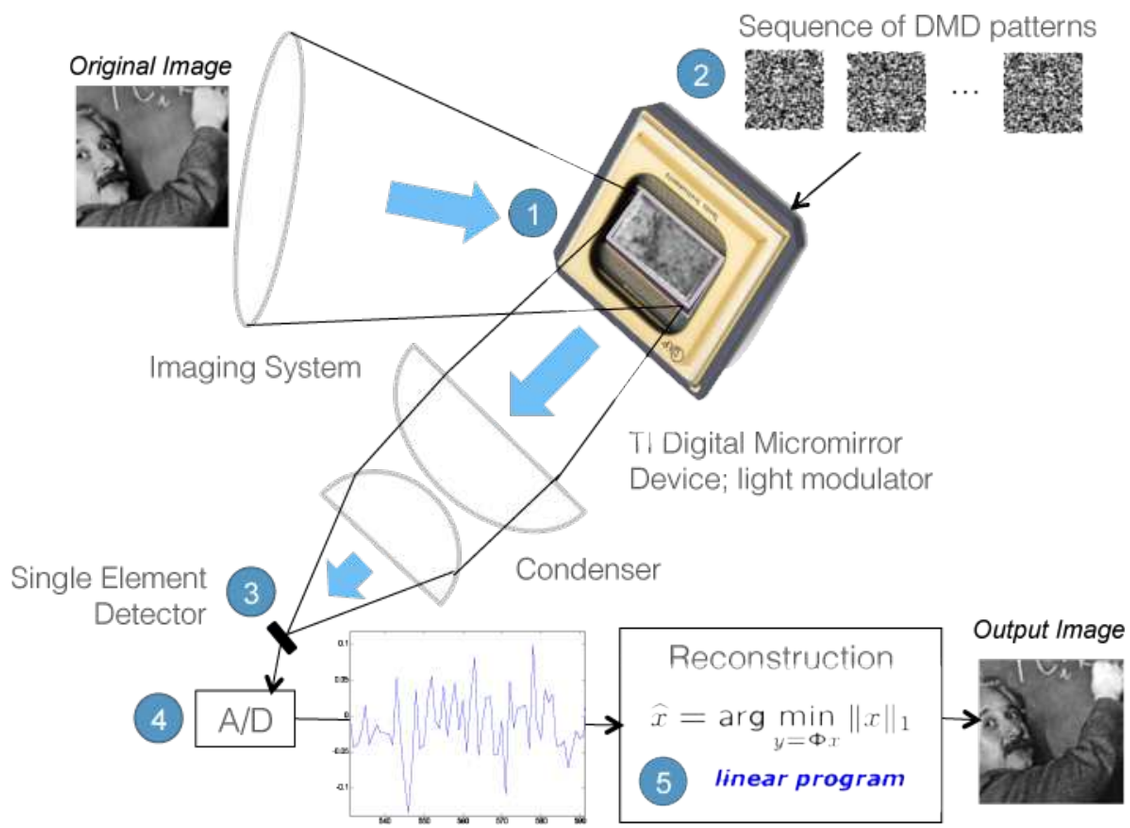
Examples of **compressed sensing**

- Compressed sensing
  - Motivating sparsity
  - Theorems
  - Examples
  - Smashed Filter
- Matrix completion
  - Examples
  - Robust PCA
    - Examples
- Where are we today?
  - Tensors
  - Sketching
  - Computational paradigm

CS hardware: single pixel camera

one of the top 10 emerging technologies by MIT Technology Review 2007

Baraniuk's group at Rice, 2006; InView Corp.



Like coded aperture but not invertible

Idea: trade-off longer acquisition time for fewer pixels. Crazy?

Application: wavelengths where single pixel is expensive

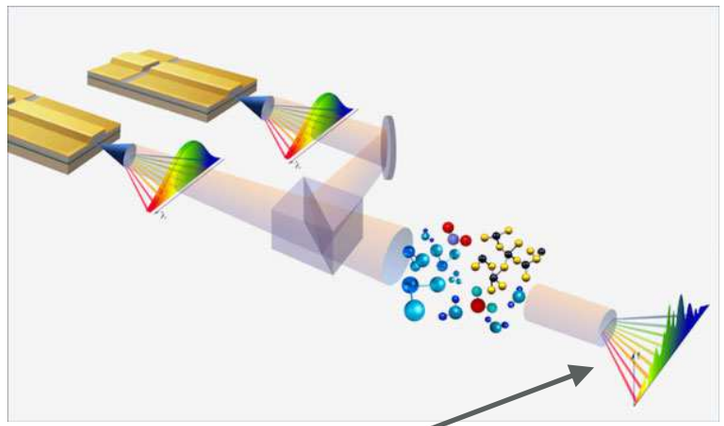
Issues: noise (non-negative measurements), calibration

Other applications of compressed sensing: MRI, DNA microarray, hyperspectral imaging, radar, seismic imaging, image processing, error coding

CS hardware: single pixel camera

one of the top 10 emerging technologies by MIT Technology Review 2007

Example: dual-comb spectroscopy



Used for, e.g., finding heat profile of substance, and many others

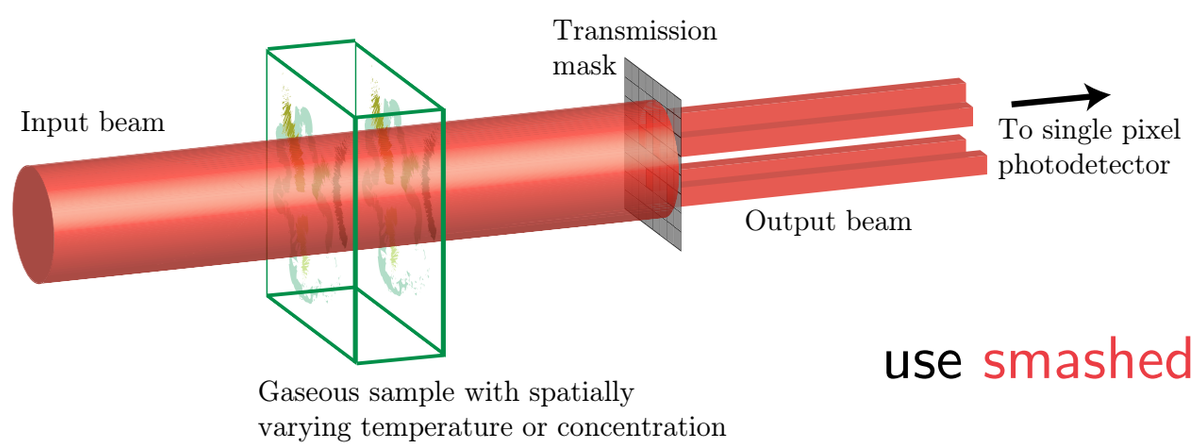
University of Colorado Boulder www.colorado.edu/lab/rieker/research

Precision Laser Diagnostics for Energy and the Environment

Research | People | Publications | Collaborators | Opportunities

image: https://microdevices.jpl.nasa.gov/capabilities/semiconductor-lasers/dual-comb-spectroscopy/

single near-IR photon sensor... perfect single pixel camera



use smashed filter ideas

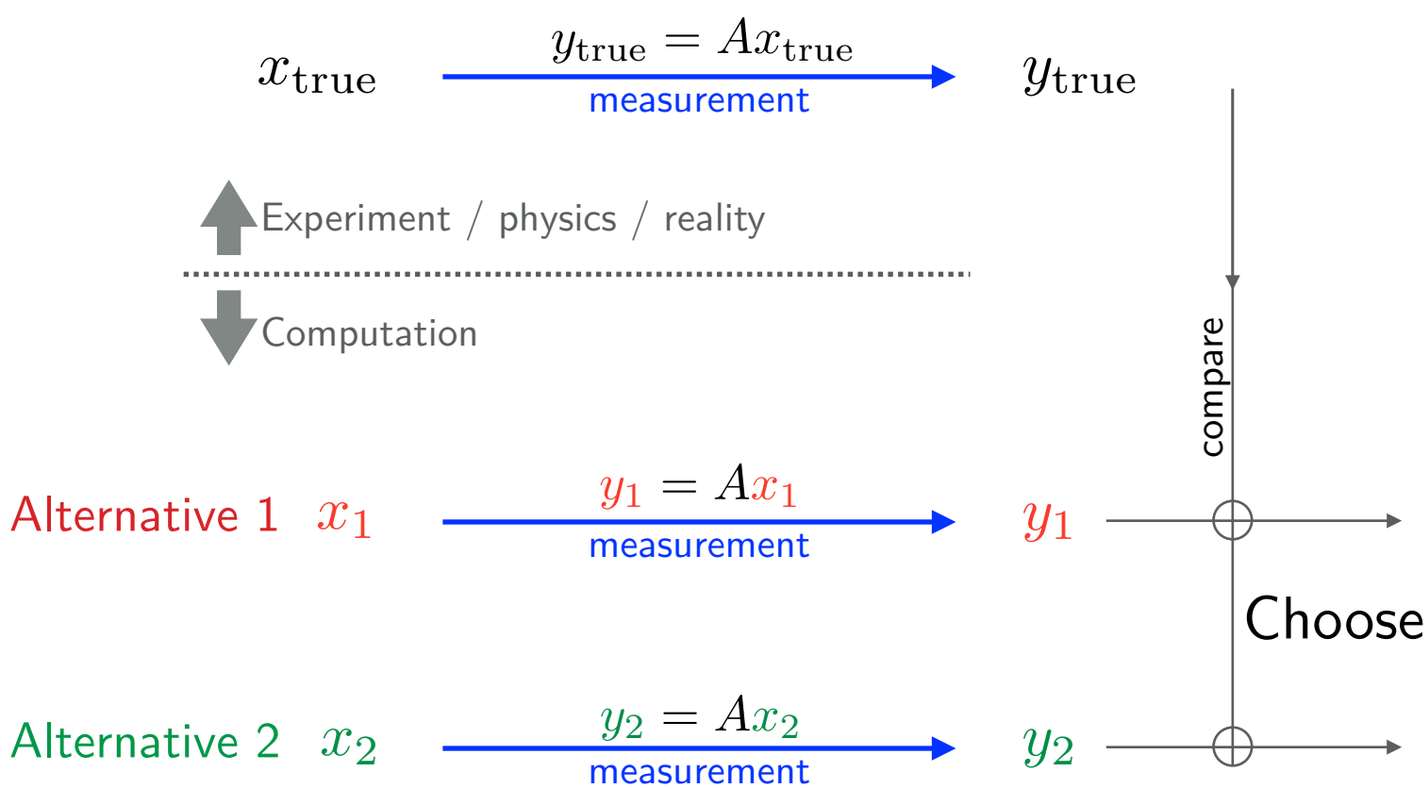
Smashed filter idea

Not compressed sensing (though still benefits from incoherent measurements)

- no sparsity needed
- no ℓ_1 minimization

Useful with a **parametric** model

For simplicity, let there be just *two alternatives*



Choose whichever better matches true measurement

Think of as combining **hashing** with a **likelihood ratio** test

Comparison often via *nonlinear least-squares*

* name from Davenport, Duarte, Wakin, Laska, Takhar, Kelly, Baraniuk 2007

CS hardware: MRI

Exploit compressibility of medical images (e.g., in a wavelet domain)

1 year old female with liver lesions, 8x acceleration AND better reconstruction

- Lustig, Pauly, Vasanawala et al. (Stanford – Berkeley)
- Trzasko, Manduca, Borisch (Mayo Clinic)
- Axel et al. (New York University)
- Mistretta (U. of Wisconsin)

MRI data collection is inherently slow

Case study: body **pediatric MRI**

Currently, MRI does not help children much:

- Children cannot keep still
- Children cannot hold breath
- low tolerance of long exams

<u>Previous solutions</u>

- sedation
- CT/CAT scans, involves radiation

Parallel imaging (PI)

Compressed sensing + PI

Lesions are barely seen with linear reconstruction

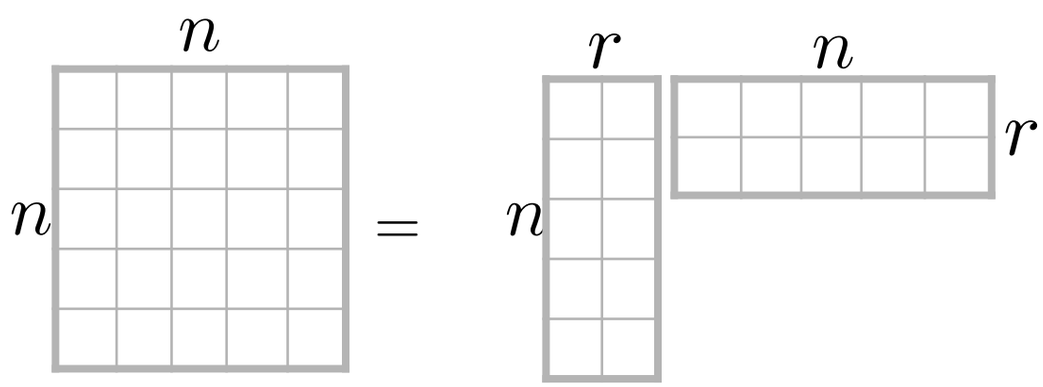
Outline

from vectors to matrices: **Matrix Completion**

- Compressed sensing
  - Motivating sparsity
  - Theorems
  - Examples
  - Smashed Filter
- Matrix completion
  - Examples
  - Robust PCA
    - Examples
- Where are we today?
  - Tensors
  - Sketching
  - Computational paradigm

Matrix completion

Assumption: low-rank



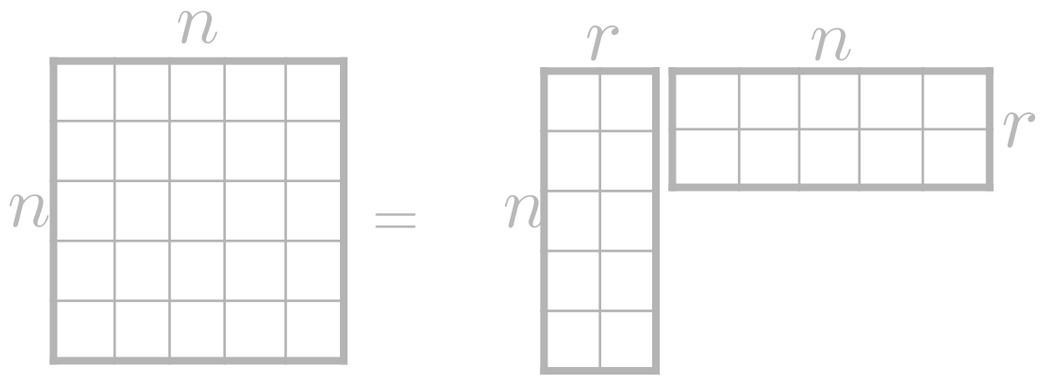
$\approx 2nr$ degrees of freedom

Measurements:

$$\mathcal{A}(X) = \begin{bmatrix} \times & ? & ? & \times & ? \\ ? & \times & ? & \times & \times \\ \times & \times & ? & ? & ? \\ ? & \times & \times & \times & ? \\ ? & ? & \times & ? & \times \end{bmatrix}$$

Matrix completion

Assumption: low-rank



$\approx 2nr$ degrees of freedom

Measurements:

$$\mathcal{A}(X) = \begin{bmatrix} \times & ? & ? & \times & ? \\ ? & \times & ? & \times & \times \\ \times & \times & ? & ? & ? \\ ? & \times & \times & \times & ? \\ ? & ? & \times & ? & \times \end{bmatrix}$$

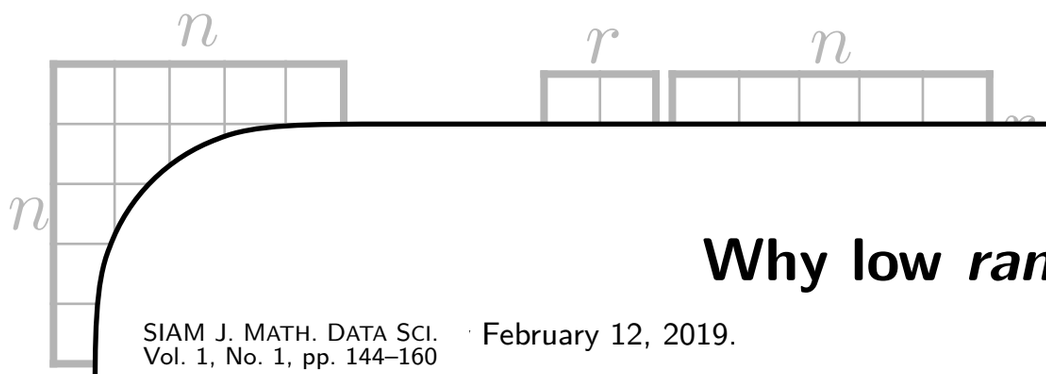
Matrix completion is like an extension of compressed sensing to matrices, but

- measurements are of a different nature (missing data), not Gaussians
- assumption is that we are **low-rank, rather than sparse**

(There *is* an analogous theory for Gaussian measurements, which is actually simpler than the vector case. Related to **sketching** and **randomized linear algebra**)

Matrix completion

Assumption: low-rank



Measurements:

Why low *rank* matrices?

SIAM J. MATH. DATA SCI. February 12, 2019.
Vol. 1, No. 1, pp. 144–160

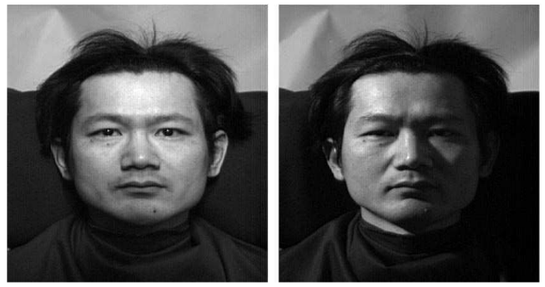
Why Are Big Data Matrices Approximately Low Rank?*

Madeleine Udell† and Alex Townsend‡

Example: motion of rigid object tracked in a video is rank 4

e.g., Tron and Vidal (2007)

Example: set of all **illuminations** of convex Lambertian object is ∼ rank 9



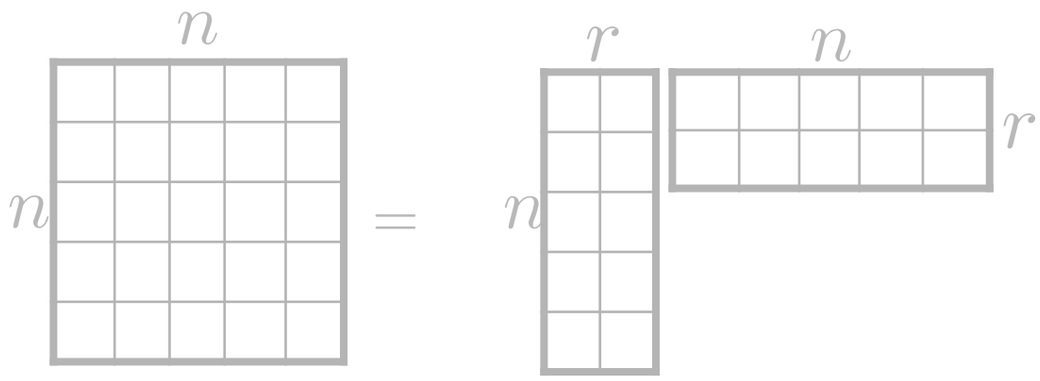
Basri and Jacobs (2003)

Other examples (some coming later in the talk)

- PCA, MDS
- quantum states
- minimum order system identification
- combinatorial relaxations
- phase recovery
- distance matrices

Matrix completion

Assumption: low-rank



$\approx 2nr$ degrees of freedom

Measurements:

$$\mathcal{A}(X) = \begin{bmatrix} \times & ? & ? & \times & ? \\ ? & \times & ? & \times & \times \\ \times & \times & ? & ? & ? \\ ? & \times & \times & \times & ? \\ ? & ? & \times & ? & \times \end{bmatrix}$$

Matrix completion is like an extension of compressed sensing to matrices, but

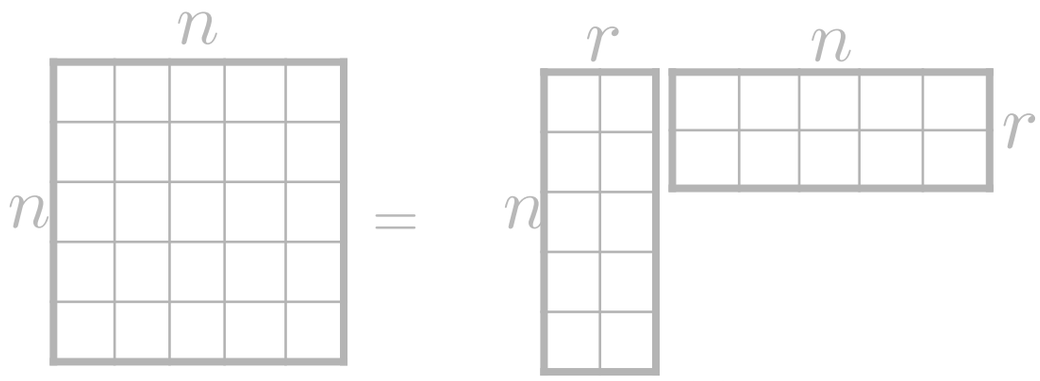
- measurements are of a different nature (missing data), not Gaussians
- assumption is that we are **low-rank, rather than sparse**
- **solve via optimization**:

$$\min_X \operatorname{rank}(X) \quad \text{subject to} \quad \mathcal{A}(X) = b,\ X \in \mathbb{R}^{n_1 \times n_2}$$

this is intractable!

Matrix completion

Assumption: low-rank



$\approx 2nr$ degrees of freedom

Measurements:

$$\mathcal{A}(X) = \begin{bmatrix} \times & ? & ? & \times & ? \\ ? & \times & ? & \times & \times \\ \times & \times & ? & ? & ? \\ ? & \times & \times & \times & ? \\ ? & ? & \times & ? & \times \end{bmatrix}$$

Matrix completion is like an extension of compressed sensing to matrices, but

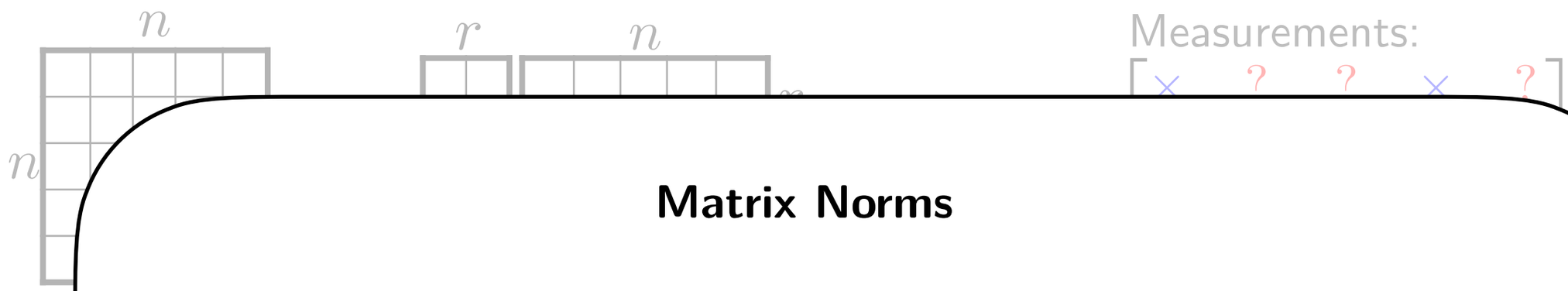
- measurements are of a different nature (missing data), not Gaussians
- assumption is that we are **low-rank, rather than sparse**
- using **nuclear norm** gives a tractable problem:

$$\min_X \|X\|_* \quad \text{subject to} \quad \mathcal{A}(X) = b, \; X \in \mathbb{R}^{n_1 \times n_2}$$

Optimization problem is a semi-definite program (**SDP**), so difficult to solve via IPM when large.
For first-order methods, difficulty due to non-smoothness and constraints

Matrix completion

Assumption: low-rank



Matrix Norms

Singular Value Decomposition (SVD)

$X = U\Sigma V^T, \quad \Sigma = \mathrm{diag}(\sigma),\ \sigma_i \geq 0$

- $\|X\| = \max_i\ \sigma_i = \|\sigma\|_\infty$ spectral
- $\|X\|_F = \sqrt{\sum_i \sigma_i^2} = \|\sigma\|_2$ Frobenius
- $\|X\|_* = \sum_i \sigma_i = \|\sigma\|_1 = \mathrm{tr}\left(\sqrt{X^T X}\right)$ nuclear (or Schatten-1, or trace)

matrix analog of l1 norm for vectors

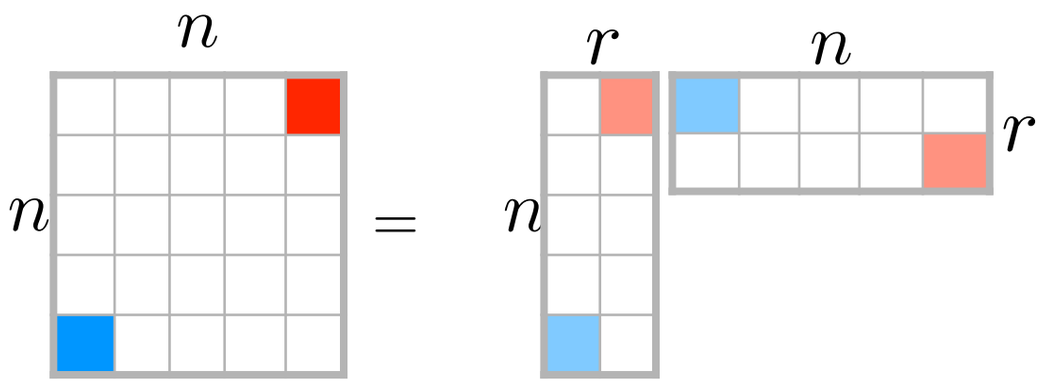
σ are called "singular values"

Optimiza M when large.

For first-order methods, difficulty due to non-smoothness and constraints

What about "incoherent measurements"?

Assumption: low-rank



Measurements:

$$\mathcal{A}(X) = \begin{bmatrix} \times & ? & ? & \times & ? \\ ? & \times & ? & \times & \times \\ \times & \times & ? & ? & ? \\ ? & \times & \times & \times & ? \\ ? & ? & \times & ? & \times \end{bmatrix}$$

This matrix is **low-rank** and **sparse**.

Unlikely for *uniform entry-wise sampling* to find non-zero entries

$\Longrightarrow$ it's indistinguishable from the zero matrix, and cannot recover

We need to assume the left and right singular vectors are **incoherent** (i.e., not sparse)

Or, if they are not incoherent, need to do to non-uniform sampling

Outline

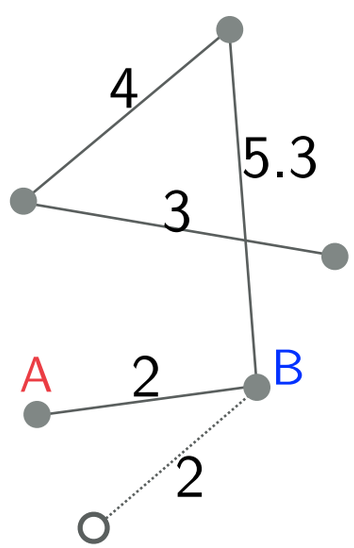
Examples of matrix completion

- Compressed sensing
  - Motivating sparsity
  - Theorems
  - Examples
  - Smashed Filter
- Matrix completion
  - Examples
  - Robust PCA
    - Examples
- Where are we today?
  - Tensors
  - Sketching
  - Computational paradigm

Euclidean Distance Matrices

	Denver	Chicago	Atlanta	Dallas	Seattle
Denver	0	?	591	398	?
Chicago	?	0	218	?	982
Atlanta	591	218	0	634	?
Dallas	398	?	634	0	521
Seattle	?	982	?	521	0

$$x_i \in \mathbb{R}^3,\ i = 1, \ldots, p, \quad \text{so } x \in \mathbb{R}^{p \times 3}$$



i.e., **nodes** = cities, **edges** = Euclidean distance

A classic problem in graph theory

Results on conditions for *existence*, *uniqueness*, etc.

Our focus: how to actually **compute** it!

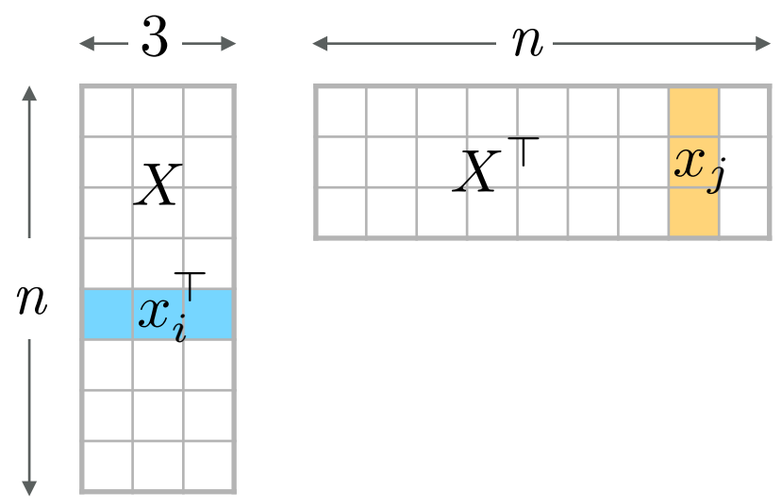
(... up to translations and rotations)

Euclidean Distance Matrices

Locations of cities

$x_i \in \mathbb{R}^3, i = 1, \ldots, n$

$$
\begin{aligned}
\mathrm{EDM}_{i,j} &\stackrel{\text{def}}{=} \|x_i - x_j\|^2 \\
&= \|x_i\|^2 - 2x_i^\top x_j + \|x_j\|^2 \\
&= \left(\mathbb{1} \cdot \mathrm{diag}(XX^\top)^\top - 2XX^\top + \mathrm{diag}(XX^\top) \cdot \mathbb{1}^\top\right)_{i,j} \\
&= \left(\mathcal{K}(XX^\top)\right)_{i,j}
\end{aligned}
$$



Euclidean Distance Matrices

Locations of cities

$x_i \in \mathbb{R}^3, i = 1, \ldots, n$

$$\begin{aligned}
\mathrm{EDM}_{i,j} &\stackrel{\text{def}}{=} \|x_i - x_j\|^2 \\
&= \|x_i\|^2 - 2x_i^\top x_j + \|x_j\|^2 \\
&= \left(\mathbb{1} \cdot \mathrm{diag}(XX^\top)^\top - 2XX^\top + \mathrm{diag}(XX^\top) \cdot \mathbb{1}^\top\right)_{i,j} \\
&= \left(\mathcal{K}(XX^\top)\right)_{i,j}
\end{aligned}$$

$$\text{Find } X \text{ such that } \left(\mathcal{K}(XX^\top)\right)_{i,j} \approx \mathrm{EDM}_{i,j} \text{ if } (i,j) \in \Omega$$

Euclidean Distance Matrices

Locations of cities

$x_i \in \mathbb{R}^3, i = 1, \ldots, n$

$$\begin{aligned}
\mathrm{EDM}_{i,j} &\stackrel{\text{def}}{=} \|x_i - x_j\|^2 \\
&= \|x_i\|^2 - 2x_i^\top x_j + \|x_j\|^2 \\
&= \left(\mathbb{1} \cdot \mathrm{diag}(XX^\top)^\top - 2XX^\top + \mathrm{diag}(XX^\top) \cdot \mathbb{1}^\top\right)_{i,j} \\
&= \left(\mathcal{K}(XX^\top)\right)_{i,j}
\end{aligned}$$

Find X such that $\left(\mathcal{K}(XX^\top)\right)_{i,j} \approx \mathrm{EDM}_{i,j}$ if $(i,j) \in \Omega$

$$Z = XX^\top$$

Find Z such that $\left(\mathcal{K}(Z)\right)_{i,j} \approx \mathrm{EDM}_{i,j}$ if $(i,j) \in \Omega, \quad Z \succeq 0, \ \mathrm{rank}(Z) = 3$

means **symmetric** and **positive semi-definite**

prevents things like $\begin{aligned} Z &= XY^\top \\ Z &= -XX^\top \end{aligned}$

Euclidean Distance Matrices

Locations of cities

$x_i \in \mathbb{R}^3, i = 1, \ldots, n$

$$\begin{aligned}
\mathrm{EDM}_{i,j} &\stackrel{\text{def}}{=} \|x_i - x_j\|^2 \\
&= \|x_i\|^2 - 2x_i^\top x_j + \|x_j\|^2 \\
&= \left(\mathbb{1} \cdot \mathrm{diag}(XX^\top)^\top - 2XX^\top + \mathrm{diag}(XX^\top) \cdot \mathbb{1}^\top\right)_{i,j} \\
&= \left(\mathcal{K}(XX^\top)\right)_{i,j}
\end{aligned}$$

Find X such that $\left(\mathcal{K}(XX^\top)\right)_{i,j} \approx \mathrm{EDM}_{i,j}$ if $(i,j) \in \Omega$

$Z = XX^\top$

Find Z such that $\left(\mathcal{K}(Z)\right)_{i,j} \approx \mathrm{EDM}_{i,j}$ if $(i,j) \in \Omega$, $\quad Z \succeq 0$, $\mathrm{rank}(Z) = 3$

Minimize $\mathrm{rank}(Z)$ such that $\left(\mathcal{K}(Z)\right)_{i,j} \approx \mathrm{EDM}_{i,j}$ if $(i,j) \in \Omega$, $\quad Z \succeq 0$

Euclidean Distance Matrices

Locations of cities

$x_i \in \mathbb{R}^3, i = 1, \ldots, n$

$$\begin{aligned}
\mathrm{EDM}_{i,j} &\stackrel{\text{def}}{=} \|x_i - x_j\|^2 \\
&= \|x_i\|^2 - 2x_i^\top x_j + \|x_j\|^2 \\
&= \left(\mathbb{1} \cdot \mathrm{diag}(XX^\top)^\top - 2XX^\top + \mathrm{diag}(XX^\top) \cdot \mathbb{1}^\top\right)_{i,j} \\
&= \left(\mathcal{K}(XX^\top)\right)_{i,j}
\end{aligned}$$

Find X such that $\left(\mathcal{K}(XX^\top)\right)_{i,j} \approx \mathrm{EDM}_{i,j}$ if $(i,j) \in \Omega$

$Z = XX^\top$

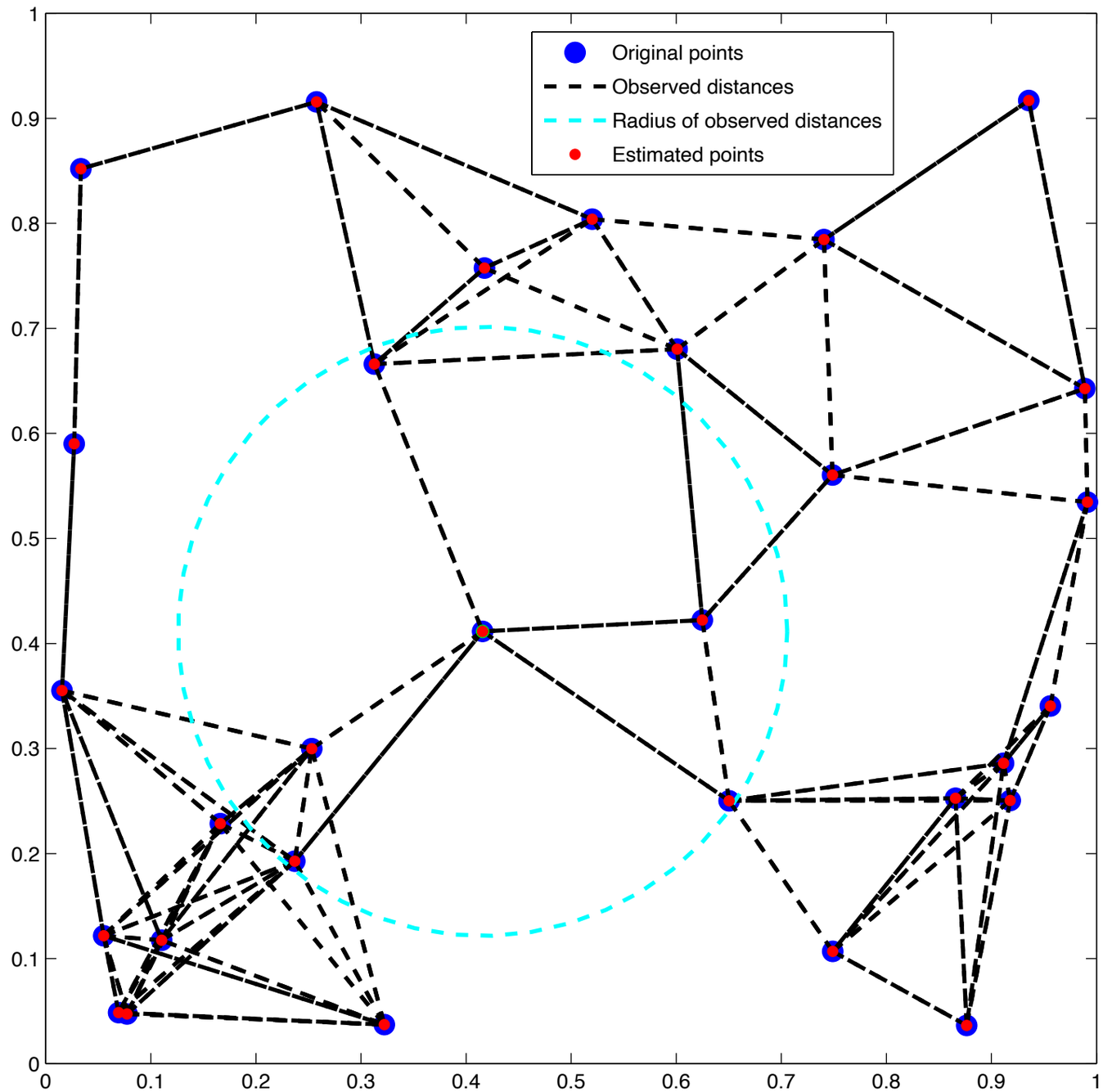
Find Z such that $\left(\mathcal{K}(Z)\right)_{i,j} \approx \mathrm{EDM}_{i,j}$ if $(i,j) \in \Omega$, $\quad Z \succeq 0$, $\mathrm{rank}(Z) = 3$

Minimize $\mathrm{rank}(Z)$ such that $\left(\mathcal{K}(Z)\right)_{i,j} \approx \mathrm{EDM}_{i,j}$ if $(i,j) \in \Omega$, $\quad Z \succeq 0$

Minimize $\|Z\|_*$ such that $\left(\mathcal{K}(Z)\right)_{i,j} \approx \mathrm{EDM}_{i,j}$ if $(i,j) \in \Omega$, $\quad Z \succeq 0$

This is now a **tractable** semi-definite program (SDP)

Example (synthetic data)



$n = 30$ points

Observe 190 of

$$\binom{n}{2} = \frac{30 \cdot 29}{2} = 435$$

total possible pairwise distances

(i.e., 43.7%)

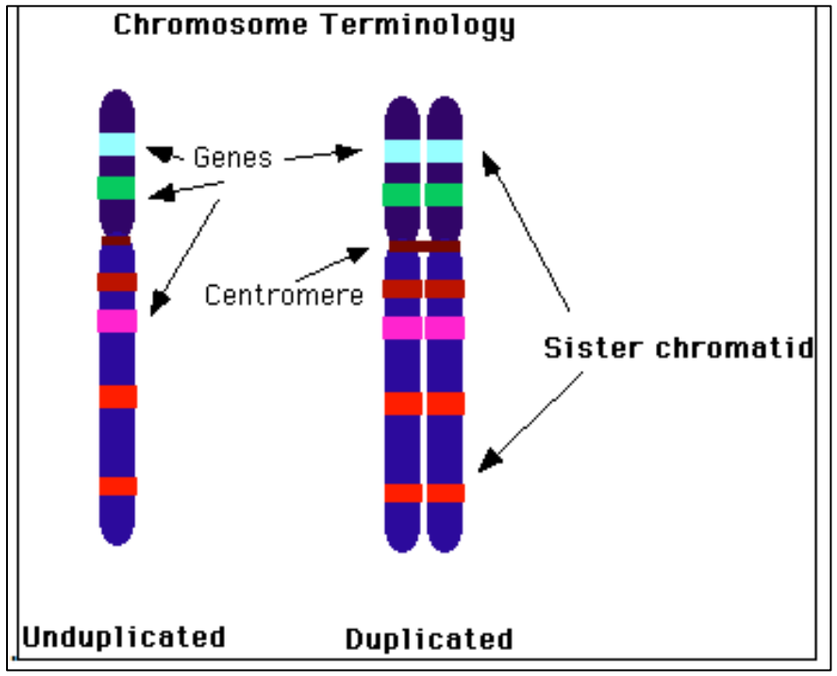
After correcting for rotation and translation, error is $5.6 \cdot 10^{-16}$ using SDP approach.

EDM application: DNA structure

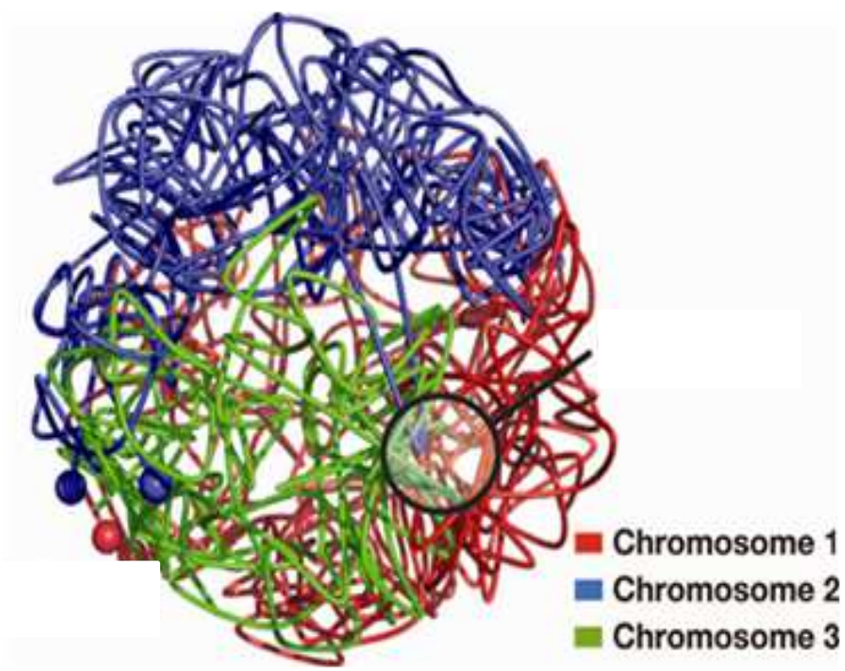
Study structural/spatial properties of chromosomes

(this is *not* already known! we know "flattened" chromosomes)

This structure affects regulation of gene expression, replication, repair, ...



"Flattened" Chromosome



"Actual" 3-d geometry

DNA example: measurements

Measurements come via *chromosome conformation capture* (3C) and variants

3C: cf. Dekker et al, *Science* (2002)

roughly speaking: we see pairwise distances between *loci*,
which we collect in the "contact" map C

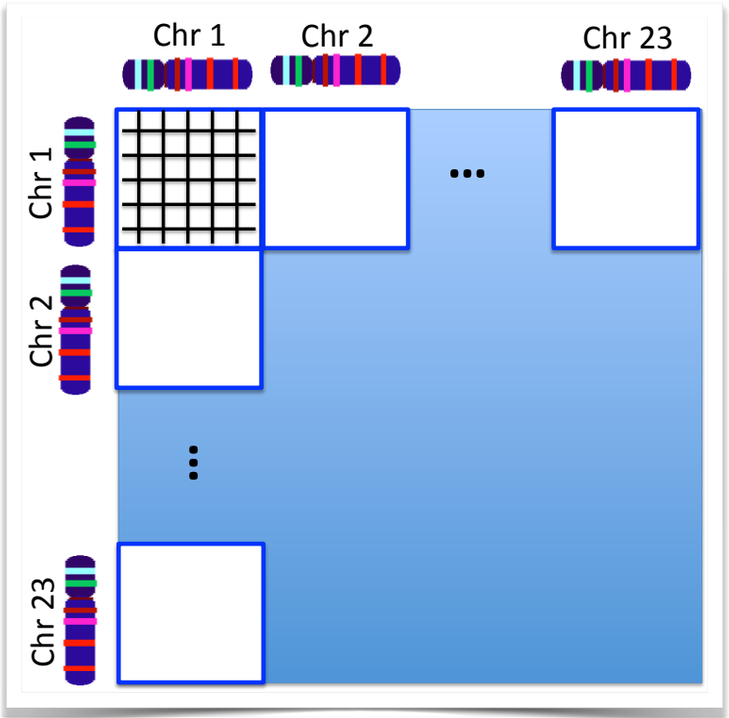
By itself, even with perfect measurements, this is a hard problem

To make things worse...

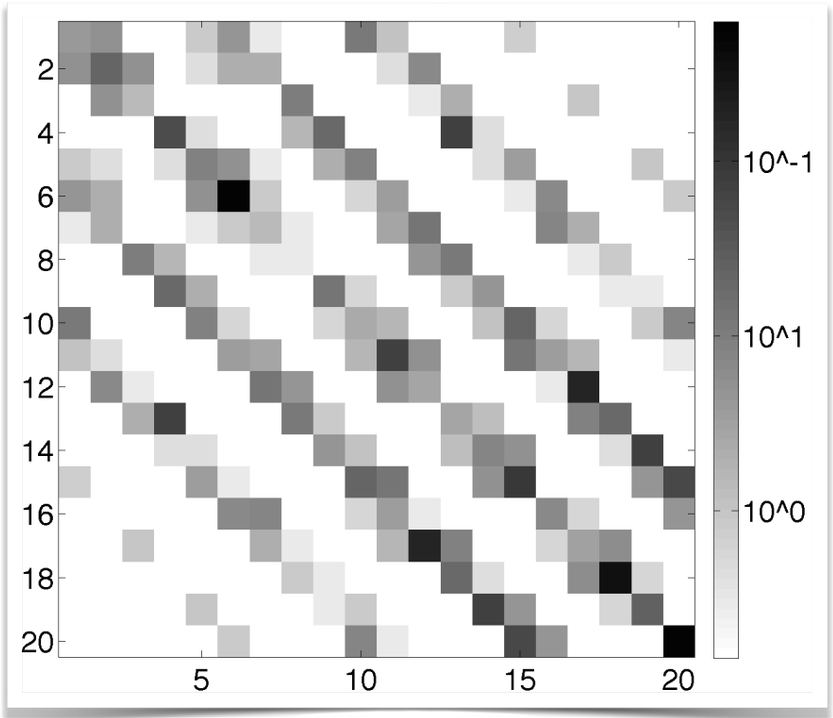
our measurements are indirect: random variables, affected by pairwise distance *and* other covariates ... and hence hard to model, and very noisy

structure of the $p \times p$ contact map

C =

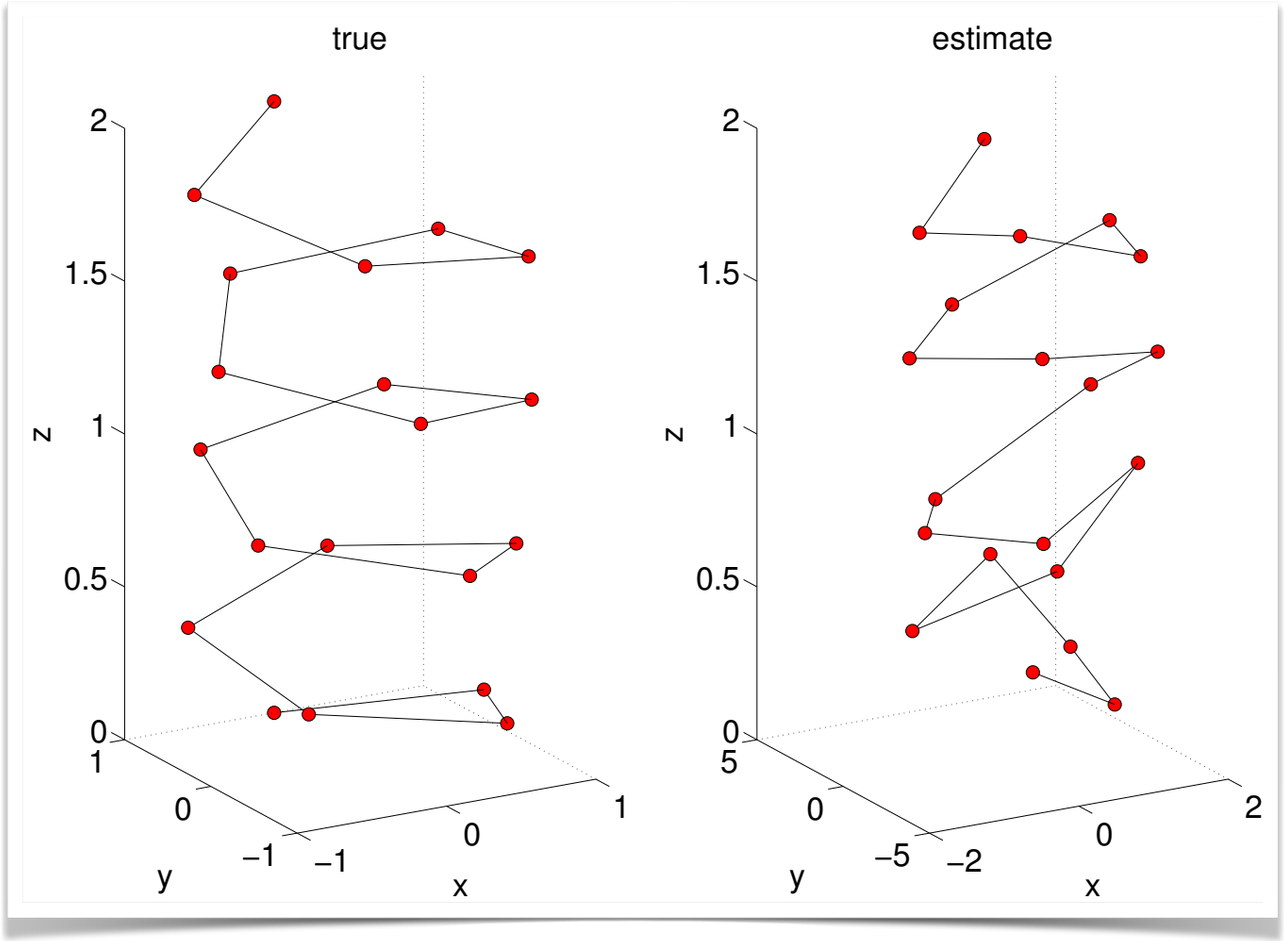


actual contact map (for simulated data)

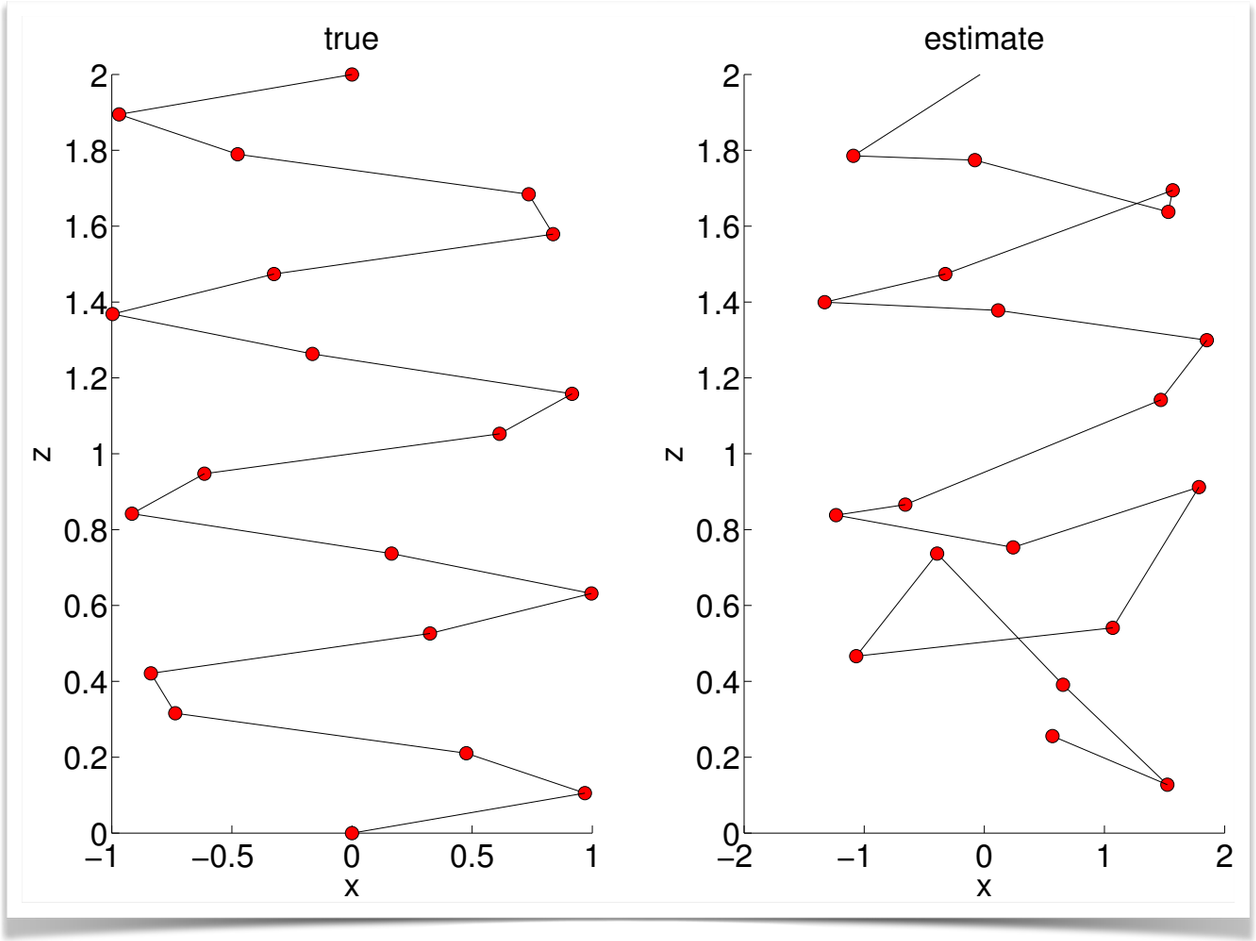


DNA example: preliminary results

3D view of true/recovered points



2D projection of same points



Simulated data, according to our Poisson model

We recover model coefficients and locations

Aravkin, Becker, Drusvyatskiy, Lozano (*Allerton '16*)

Other challenges: data comes from a mix of many different cells. There is no "ground-truth"

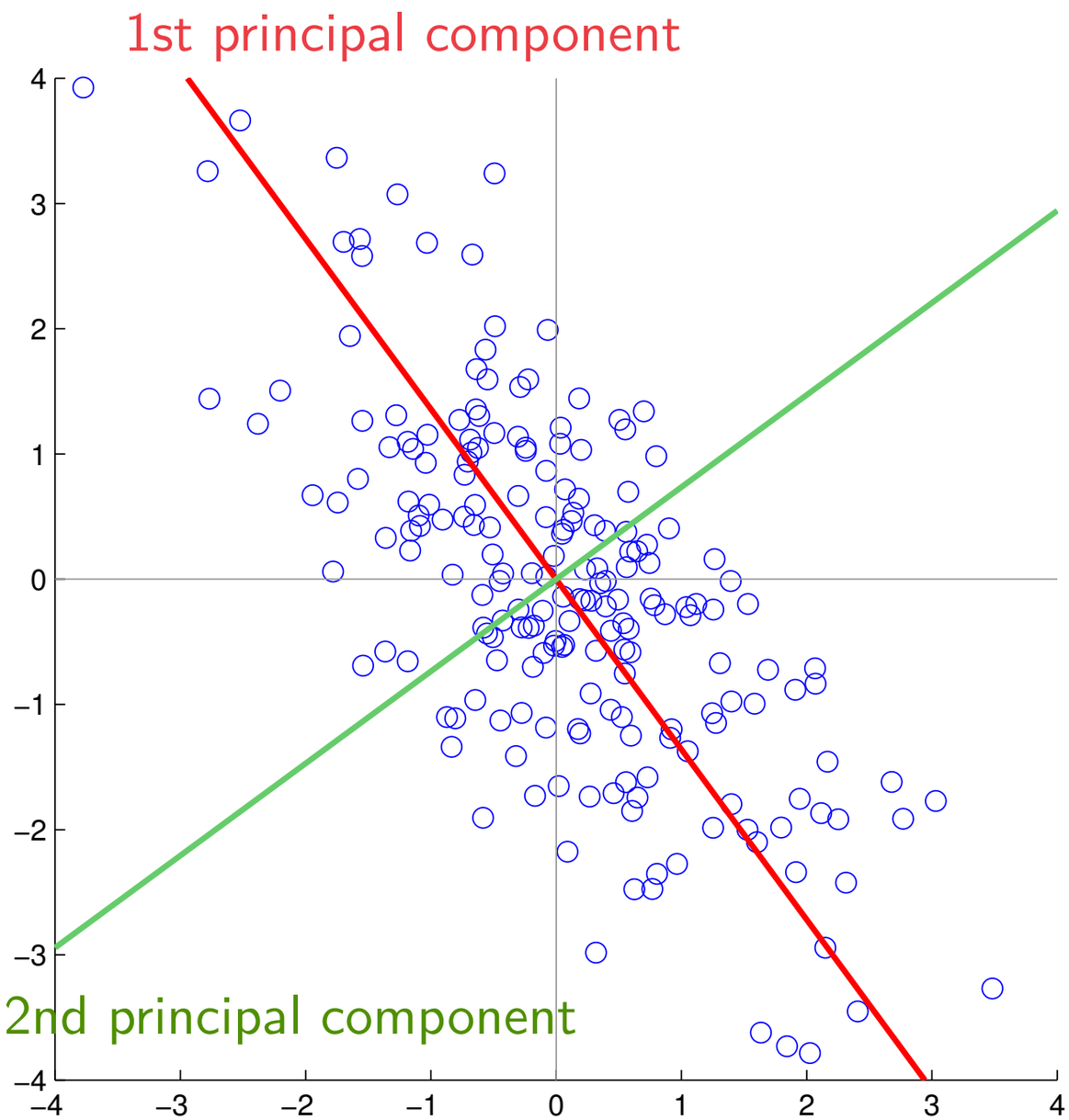
Technique: combine EDM with Poisson regression

Outline

Robust PCA

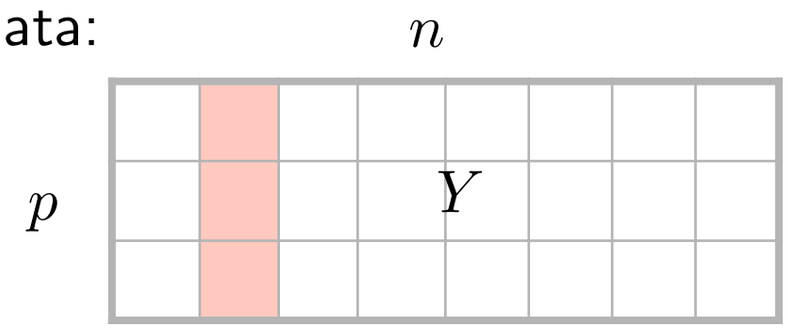
- Compressed sensing
  - Motivating sparsity
  - Theorems
  - Examples
  - Smashed Filter
- Matrix completion
  - Examples
  - Robust PCA
    - Examples
- Where are we today?
  - Tensors
  - Sketching
  - Computational paradigm

Principal Component Analysis (PCA)



note: typically Y is *centered* in PCA

Data:



each <u>column</u> of Y is a new observation

PCA finds the most significant *left* singular vectors of Y (via SVD)

$$Y = UDV^T$$

Equivalently, find most significant *eigenvectors* of sample covariance matrix

$$\widehat{\Sigma} = \frac{1}{n-1} YY^T, \quad \widehat{\Sigma} = UDU^T$$

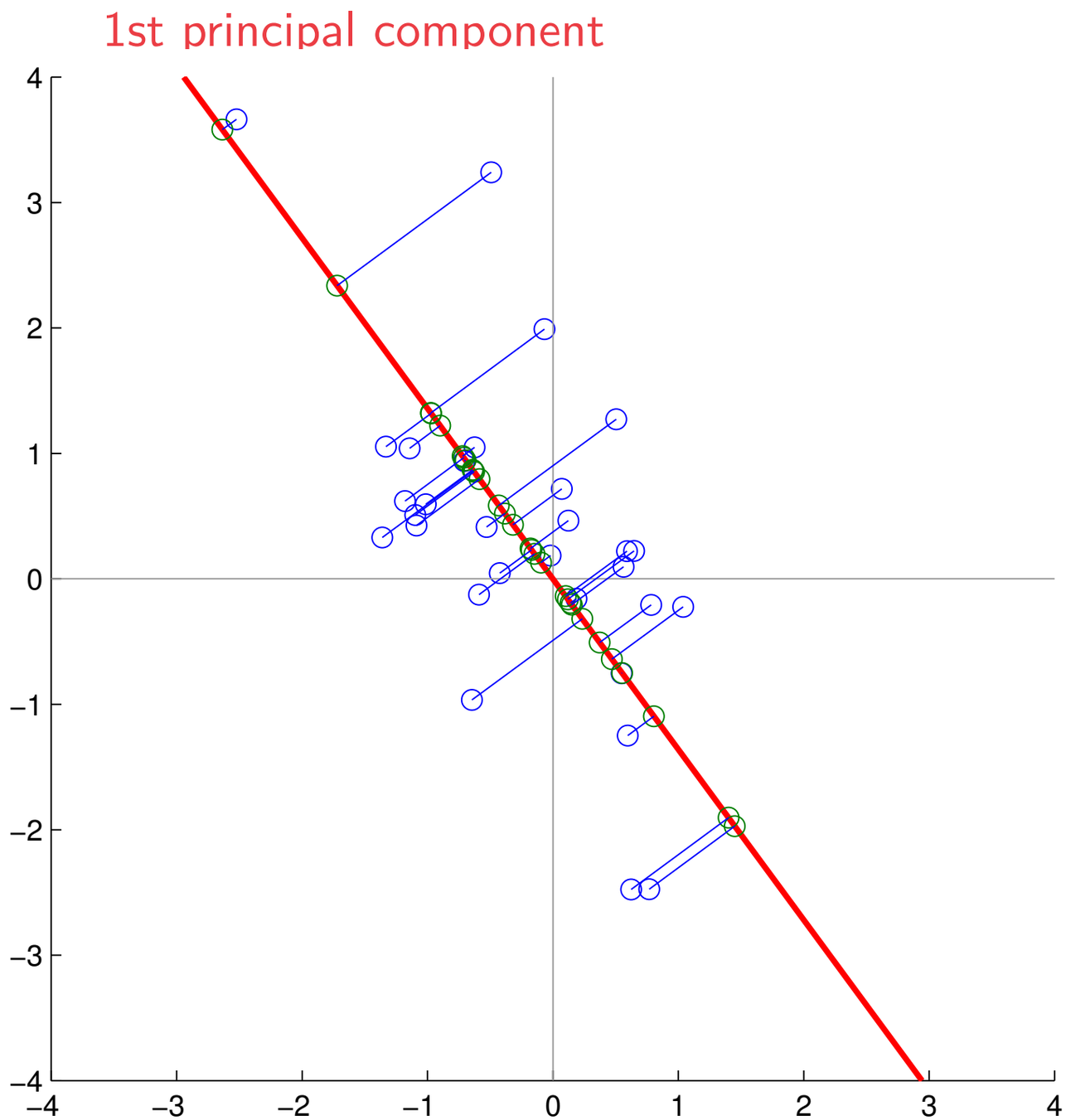
The principal components create a **new coordinate system**:

- The first coordinate is the most crucial one
- The second coordinate is the next most useful
- etc.

One of the reasons PCA is useful is because many datasets can be well-explained by just a few top principal components

... so, success of PCA is because data have a low-rank representation

Interpretations of PCA



First principal component is the axis that ...

- most greatly reduces the variance

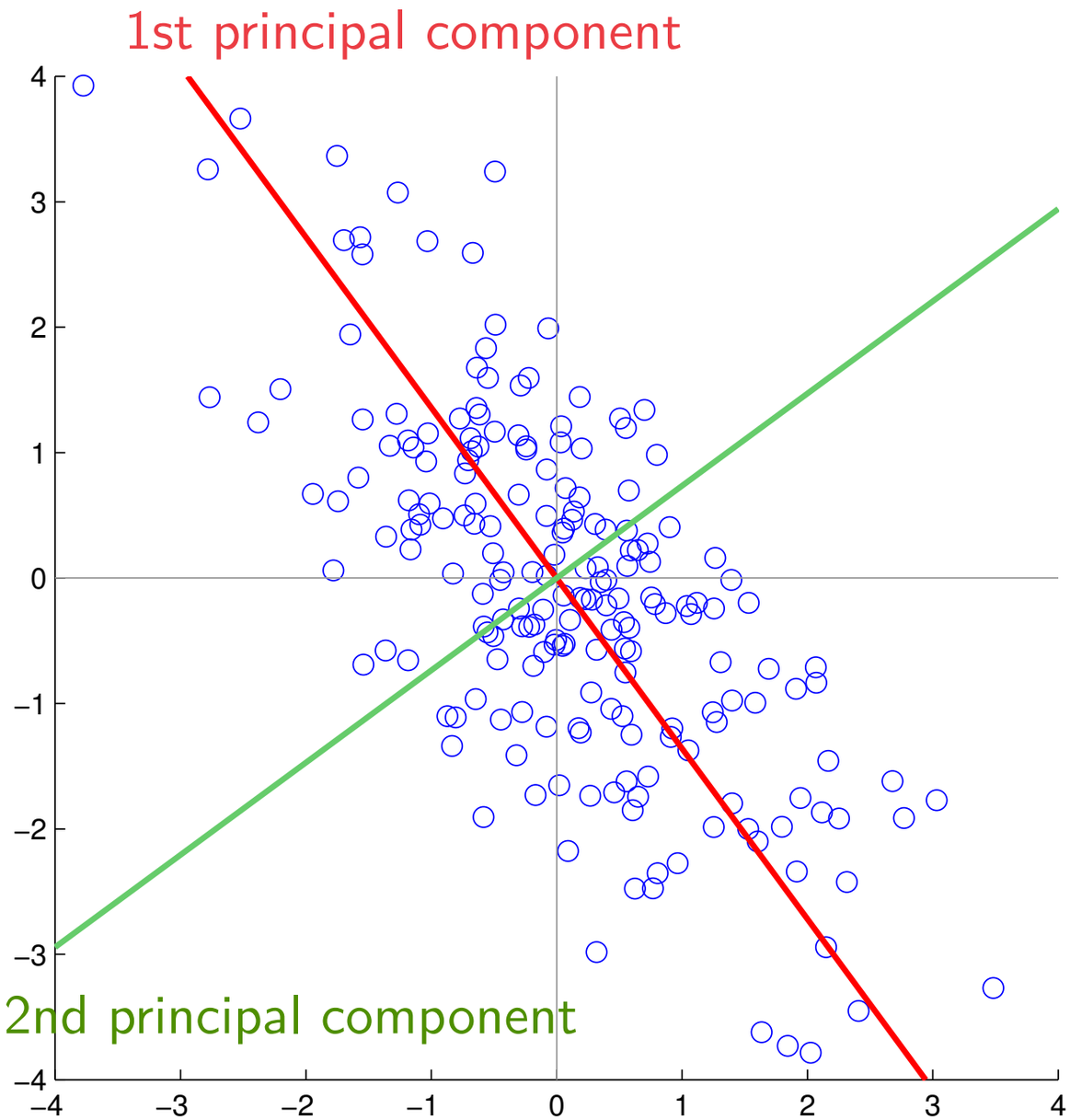
or (equivalently)...

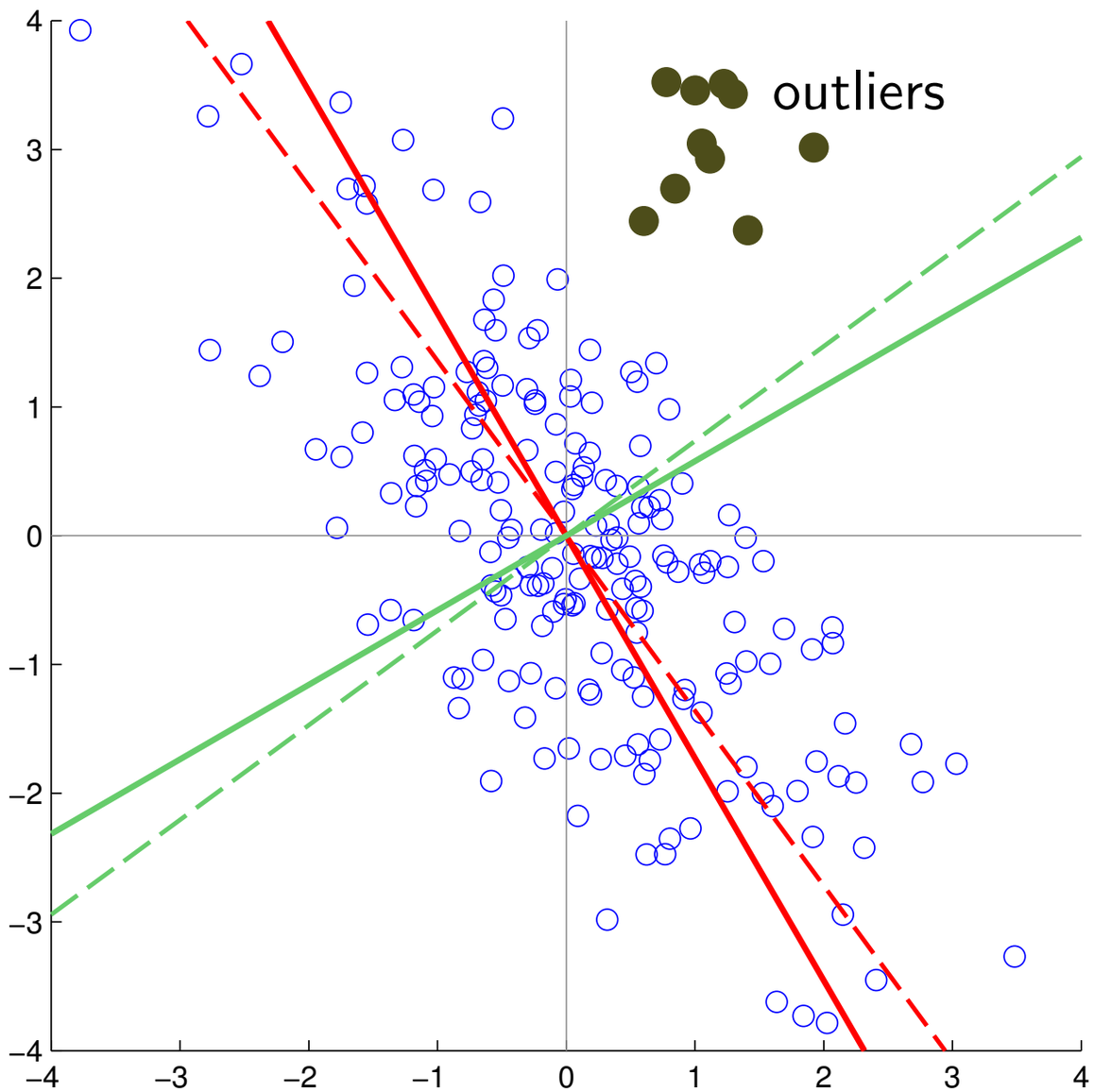
- minimizes the sum-of-squared distance of the points to this axis

Squared distance is not robust

Outliers corrupt

...and absolute outliers corrupt absolutely





Outliers corrupt

$$Y = \begin{array}{|c|c|c|c|c|}\hline 6 & 5 & 4 & 5 & 5 \\ \hline 12 & 12 & 9 & 9 & 9 \\ \hline 10 & 11 & 8 & 7 & 7 \\ \hline 6 & 5 & 4 & 5 & 5 \\ \hline 12 & 12 & 9 & 9 & 9 \\ \hline \end{array} \qquad \mathrm{rank}(Y) = 2$$

$$S = \begin{array}{|c|c|c|c|c|}\hline & & & & \\ \hline -1 & & & & \\ \hline & & & 1 & \\ \hline & -1 & & & \\ \hline & & & & \\ \hline \end{array} \qquad \mathrm{rank}(Y + S) = 5$$

Full rank!

Prelude to RPCA: demixing problems

Suppose we observe $y = x + z$ and want to "unmix" x and z

This is the cocktail party problem, and is generally <u>impossible!</u>

e.g., $y = (x + 5) + (z - 5)$

$= (x + 7) + (z - 7)$

But if x is sparse in one domain, and z is sparse in another domain, and these domains are **incoherent**, then there is hope

Some early ideas: Zibulevsky, Pearlmutter '01;
"morphological diversity" of Bobin, Moudden, Starck, Elad '06

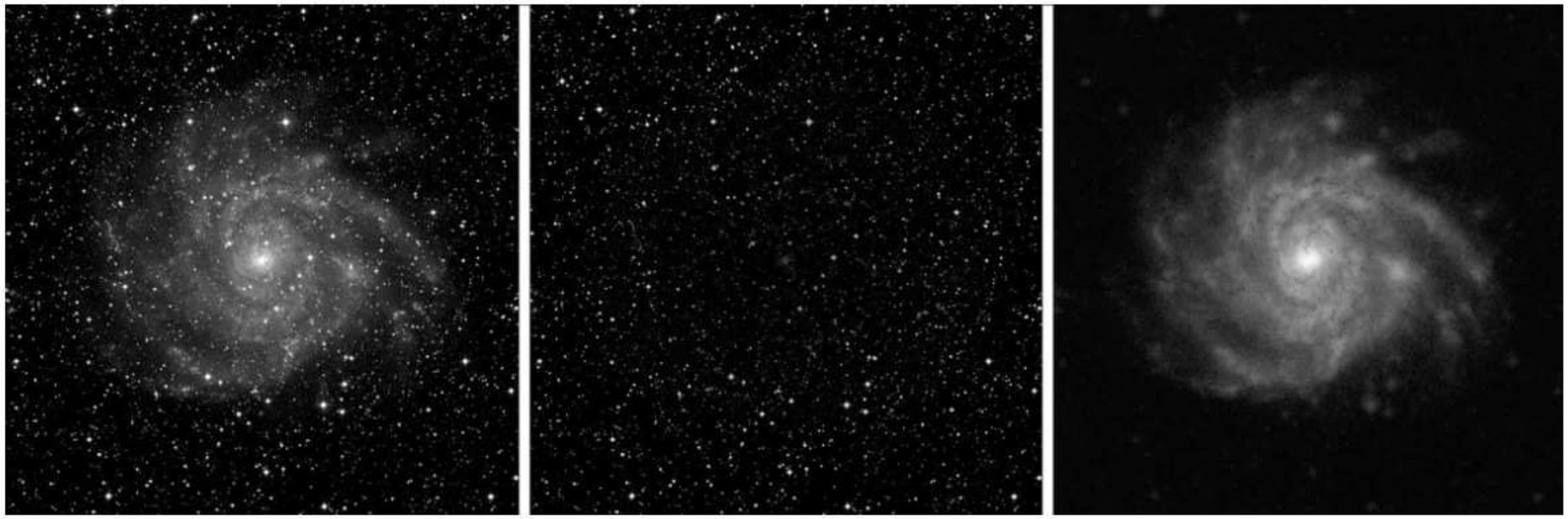


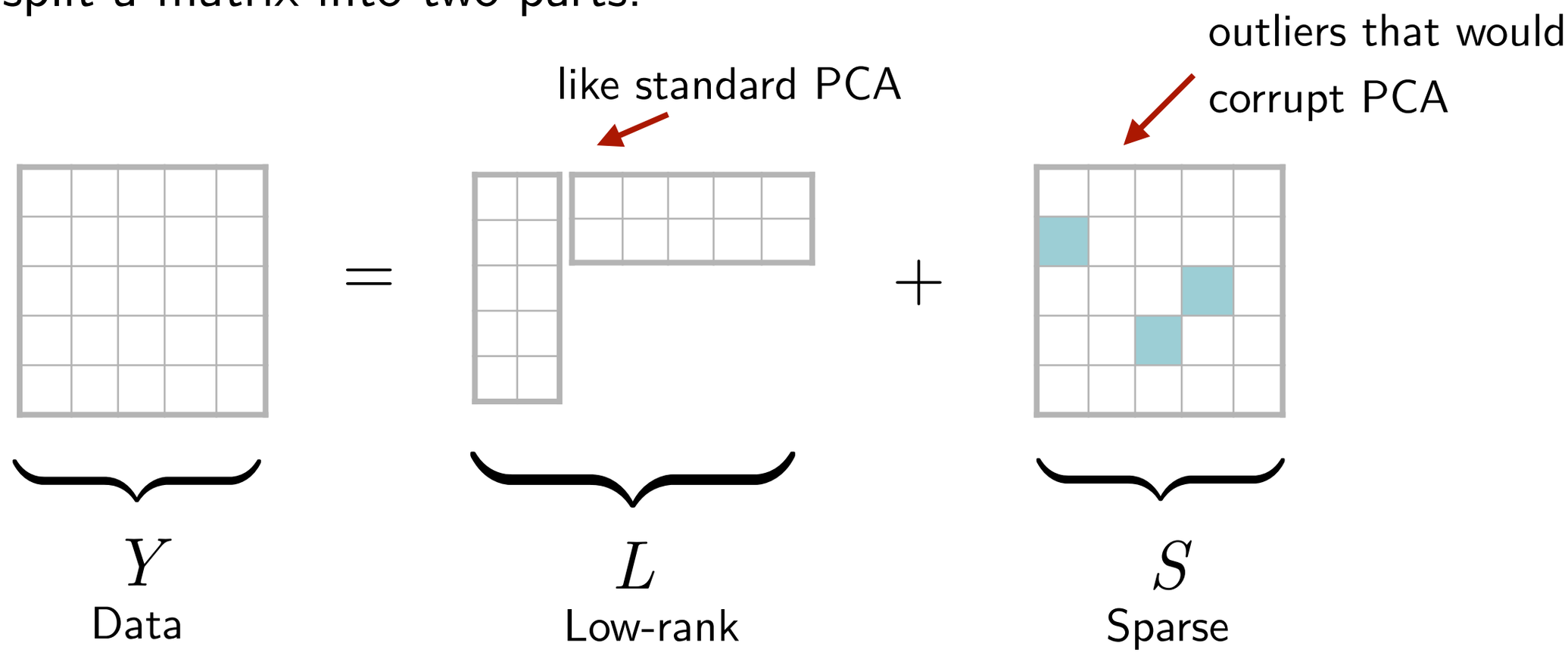
Image credit: NASA

Observation z_0 Sparse component x_0 DCT-sparse component y_0

from McCoy et al., *Convexity in source separation: Models, geometry and algorithms*, 2014

Robust PCA ("RPCA"): demixing

Idea is to split a matrix into two parts:



$$\begin{array}{ll} \text{minimize}_{L,S} & \text{rank}(L) + (\#\text{ nonzeros})(S) \\ \text{subject to} & L + S = Y \end{array}$$

Formulation of demixing problem
(but intractable)

$$\begin{array}{ll} \text{minimize}_{L,S} & \|L\|_* + \lambda \|S\|_1 \\ \text{subject to} & L + S = Y \end{array}$$

Convex relaxation
(tractable)

Noise-aware variations of this formulation exist

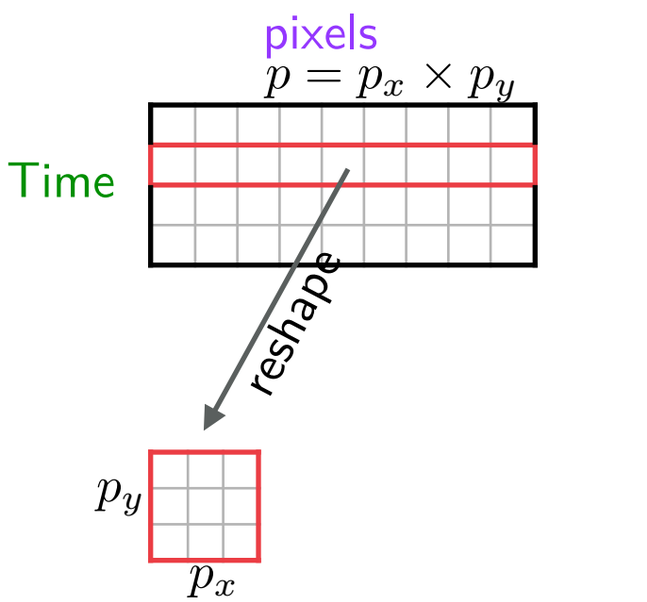
Outline

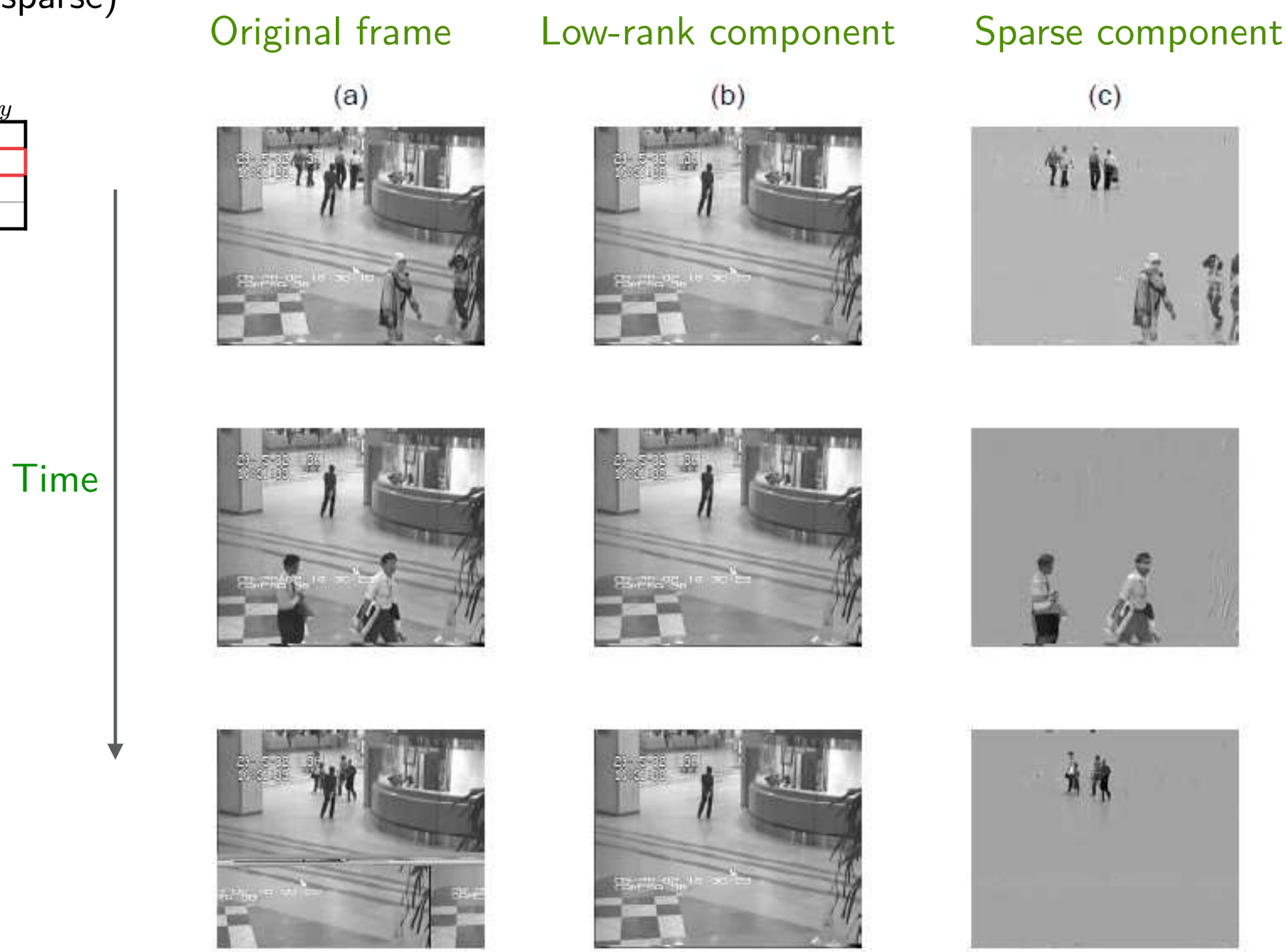
Examples of RPCA and related techniques

- Compressed sensing
  - Motivating sparsity
  - Theorems
  - Examples
  - Smashed Filter
- Matrix completion
  - Examples
  - Robust PCA
    - Examples
- Where are we today?
  - Tensors
  - Sketching
  - Computational paradigm

Background subtraction via RPCA

split an object into "predictable" (low-rank)
and "innovation" (sparse)





from Goldfarb, Ma, Scheinberg (2010)

Background subtraction via RPCA

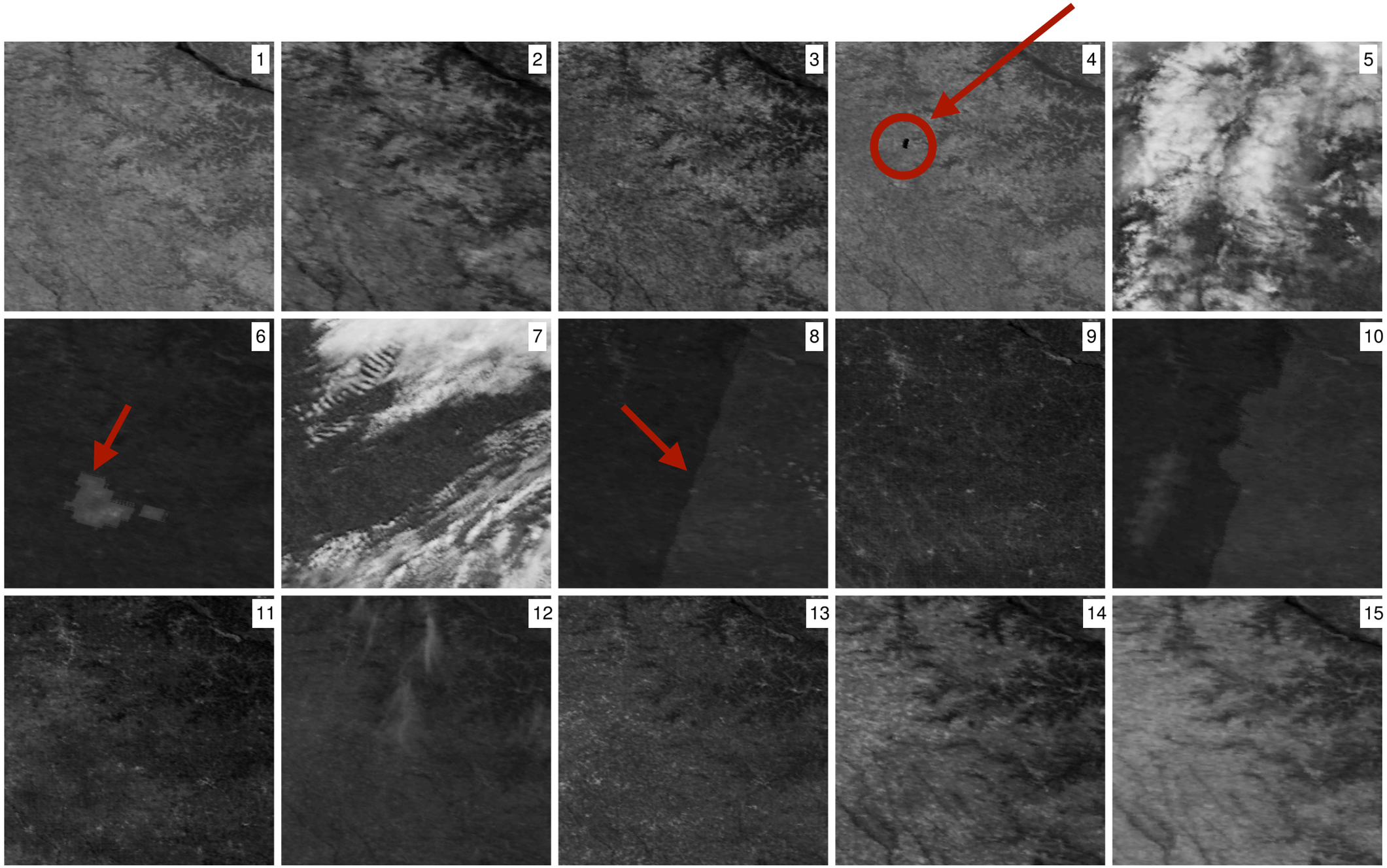
Not a compelling technique for easy scenarios,

but true advantage is to remove predictable, but non-stationary, objects

Original Low-rank Sparse

from demo of TFOCS software (Becker, Candès, Grant 2010)

Removing clouds from satellite data



Satellite photos of same location, different days.

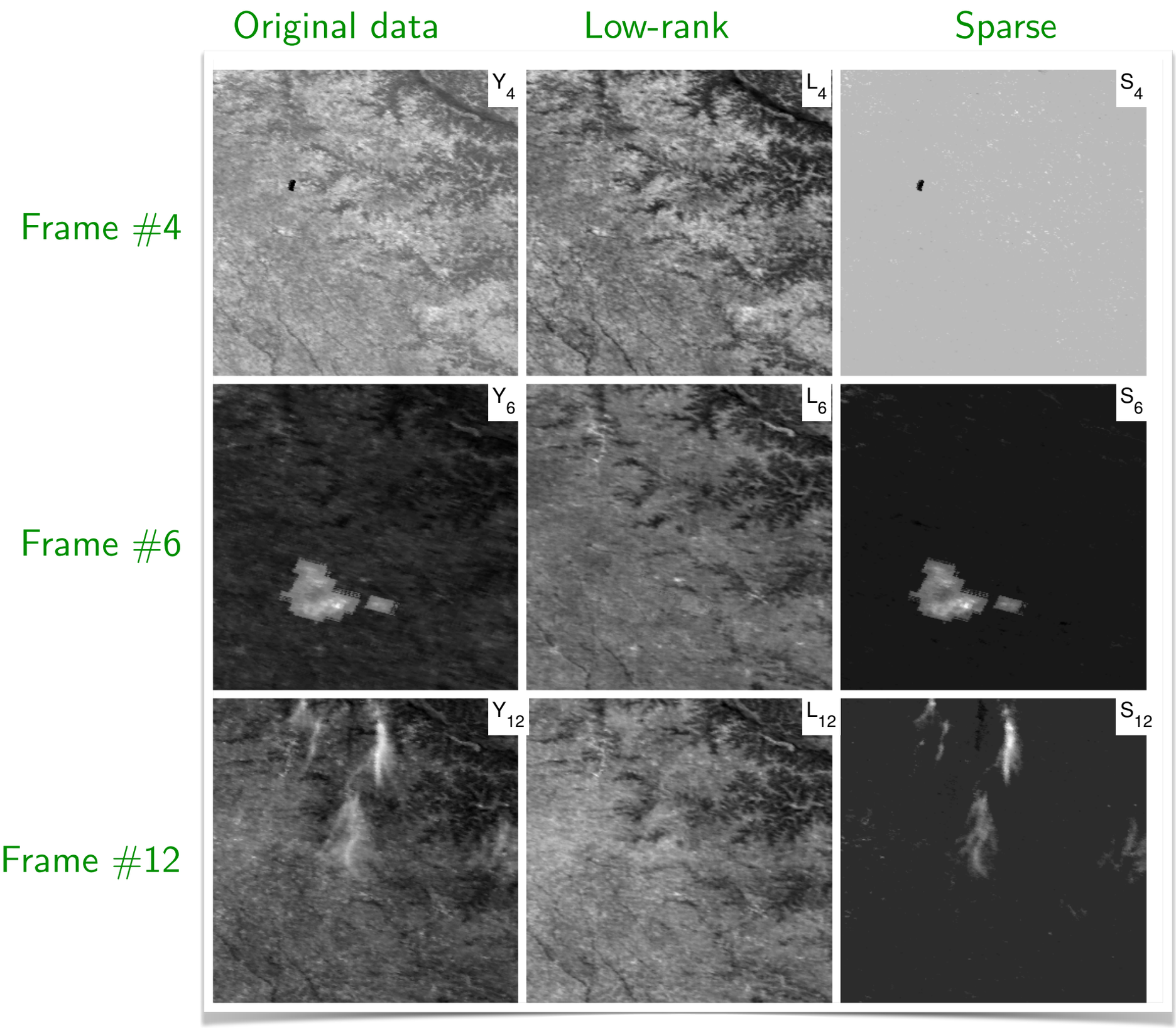
From MODIS satellite data (300 x 300 px)

Results

Parameters chosen via cross-validation

Takes 5 seconds

Images are modified to enhance contrast



from Aravkin, Becker, Cevher, Olsen 2014

Improved dataset

- 300 x 300 px, 3 (of 7) spectral bands used, from MODIS satellite
- 1 picture/day, 180 days, so 270,000 pixels total
- lots of clouds
- MODIS provides crude cloud mask
- Solve with modified non-convex version of matrix completion
- Good initialization is important

Weights based on cloud mask estimate

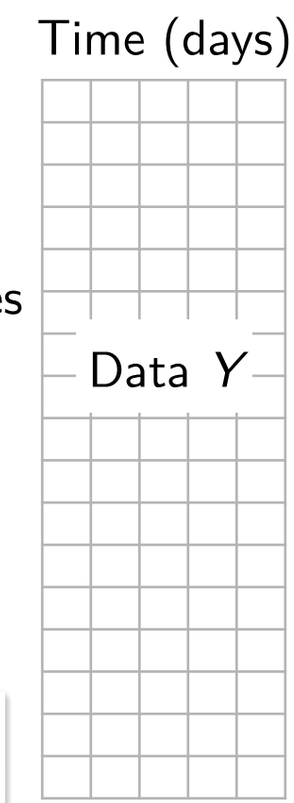
$$\underset{U,V}{\text{minimize}} \ \frac{1}{2}\|A \cdot \left(UV^T - Y\right)\|_F^2 + \gamma \mathrm{TV}(V)$$

$$U \in \mathbb{R}^{n_1 \times r}, \ V \in \mathbb{R}^{n_2 \times r}$$

Penalize rapid changes in the time component

Pixels x frequencies

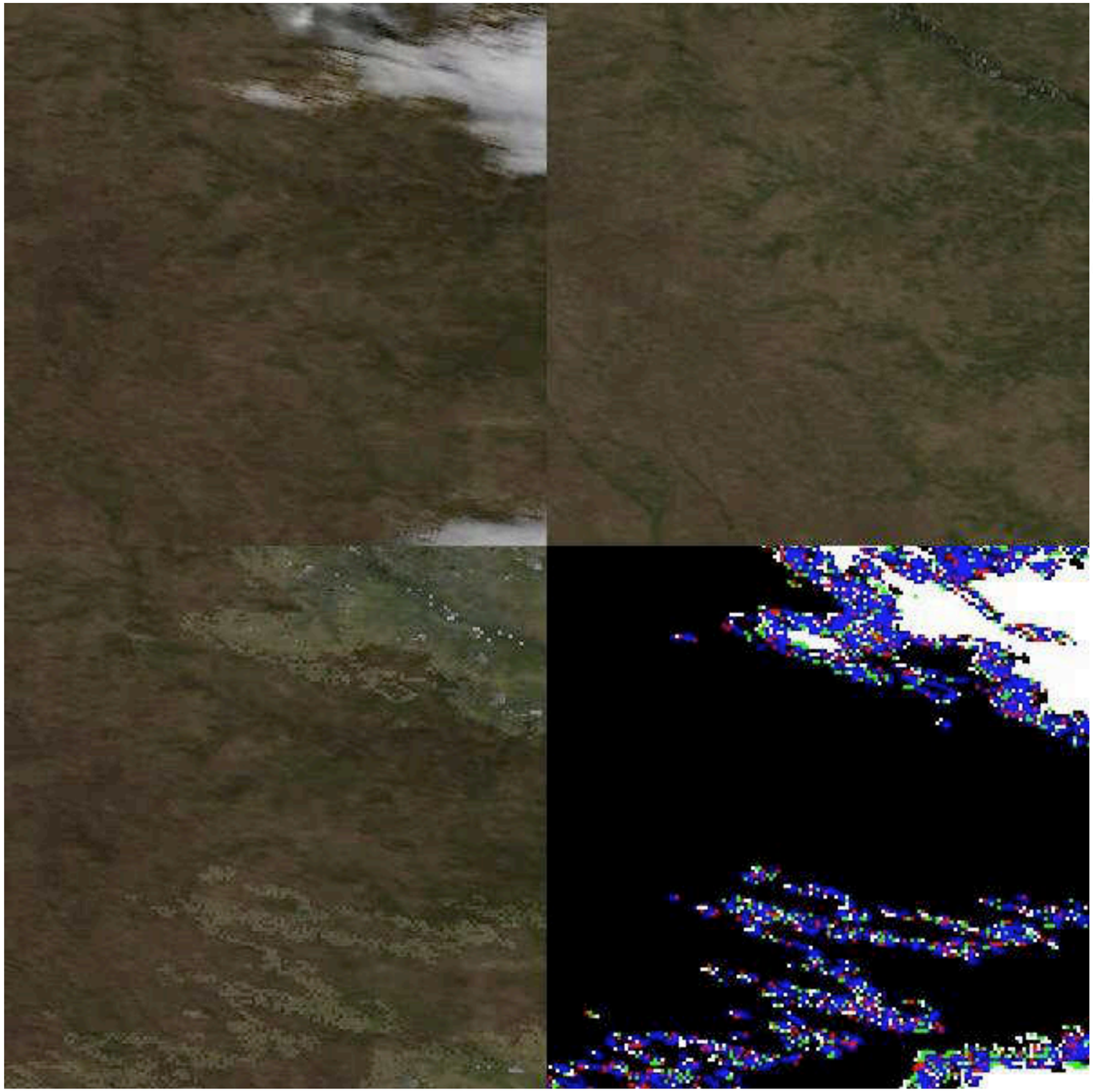
300^2 x 3



Punchline: good results relied on custom tweaks to model and solver

Cloud removal in action

Raw data



Our cloud-free estimate

Estimate based on using median for missing pixels

(Inaccurate) cloud mask provided by data source

joint work with
Marek Petrik and Peder Olsen (IBM Research)

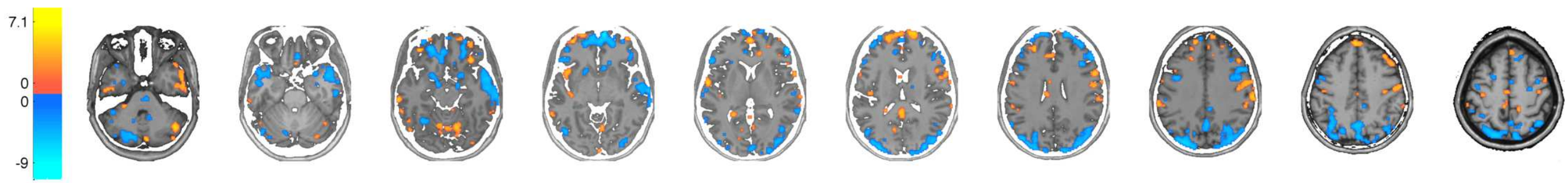
fMRI neuroscience data

Response to pain

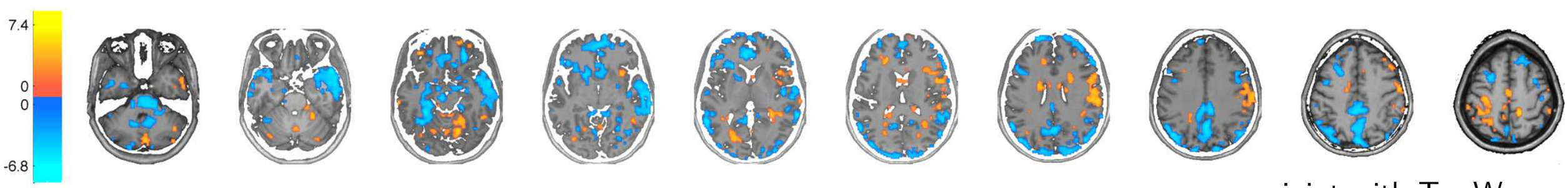
fMRI of a single subject (4 runs, each has 460 timepoints x 300,000 voxels)

Can RPCA **reduce artifacts** that occur due to head motion, respiration, scanner, etc.?

Standard

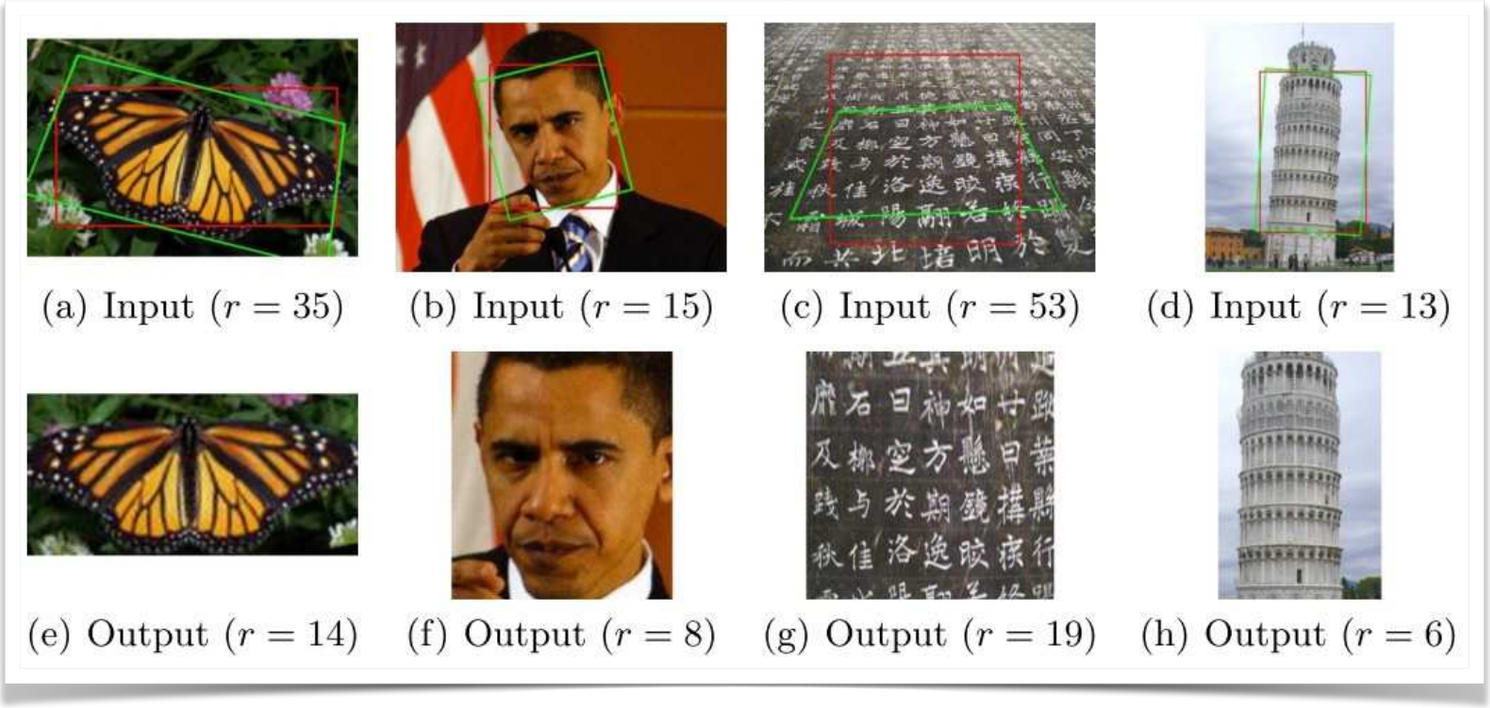


After RPCA processing



joint with Tor Wager

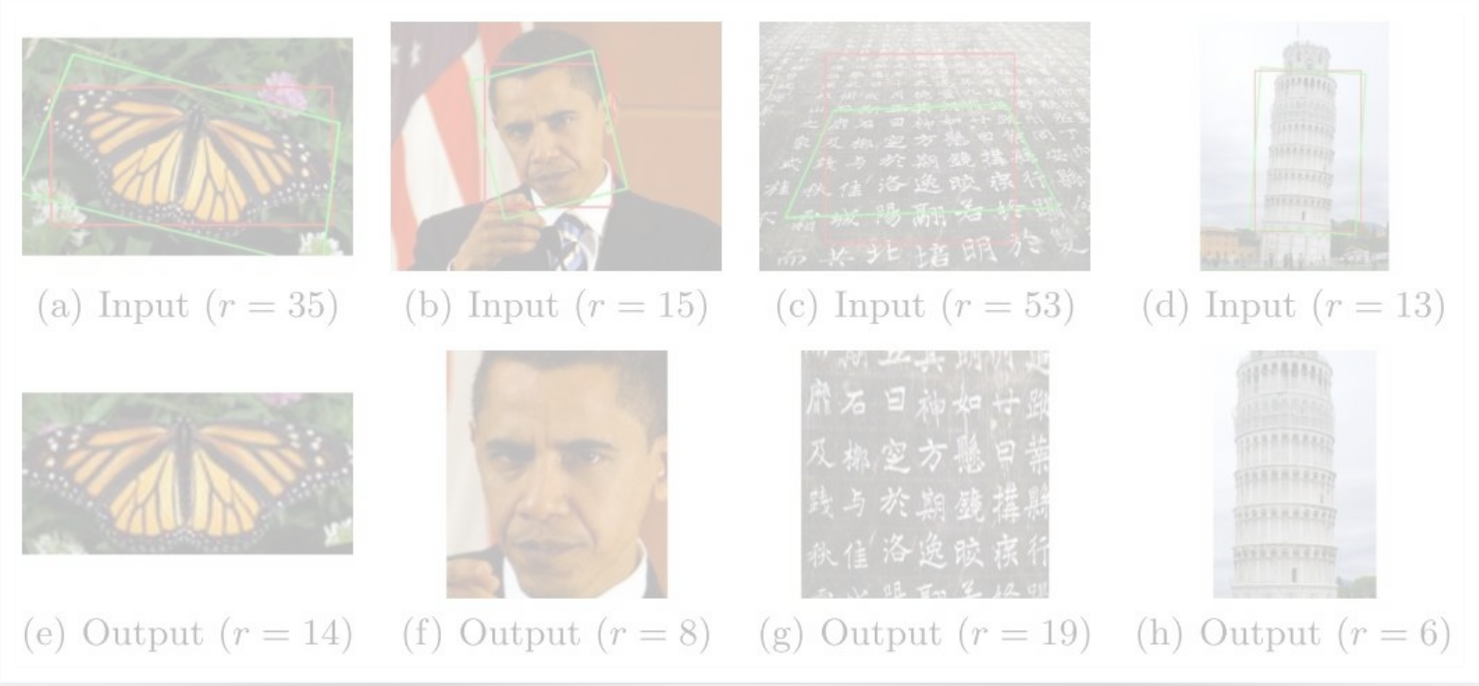
Interesting extensions


(a) Input ($r = 35$) (b) Input ($r = 15$) (c) Input ($r = 53$) (d) Input ($r = 13$)
(e) Output ($r = 14$) (f) Output ($r = 8$) (g) Output ($r = 19$) (h) Output ($r = 6$)

RPCA for alignment

TILT: Transform Invariant Low-Rank Textures, Zhang et al., Int. J. Computer Vision (2012)

Interesting extensions


(a) Input ($r = 35$) (b) Input ($r = 15$) (c) Input ($r = 53$) (d) Input ($r = 13$)
(e) Output ($r = 14$) (f) Output ($r = 8$) (g) Output ($r = 19$) (h) Output ($r = 6$)

RPCA for alignment

TILT: Transform Invariant Low-Rank Textures, Zhang et al., Int. J. Computer Vision (2012)

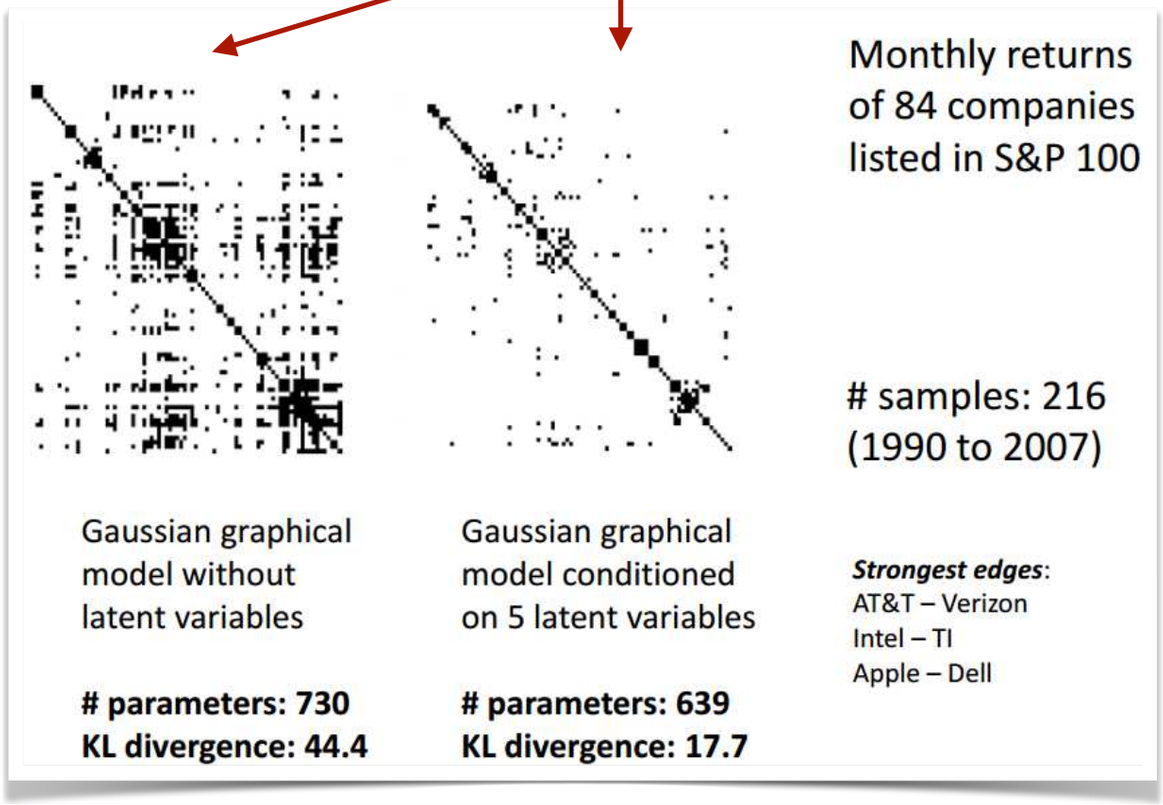
Covariance selection, graphical lasso
(for graphical models)

Results in same kind of difficult optimization problem

Efficient solvers:
QUIC & DIRTY, Hsieh, Dhillon, Ravikumar, Becker, Olsen, NIPS (2014)

Ex.: account for latent (hidden) variables
Latent Variable Graphical Model Selection via Convex Optimization, Chandrasekaran, Parrilo, Willsky, Annals Stat. (2010)

precision matrix w/o and w/ latent variables



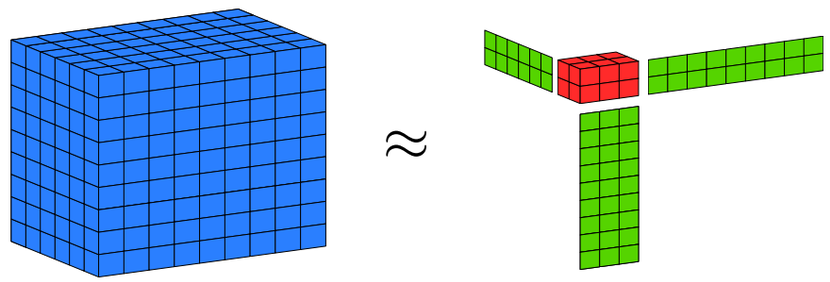
Outline

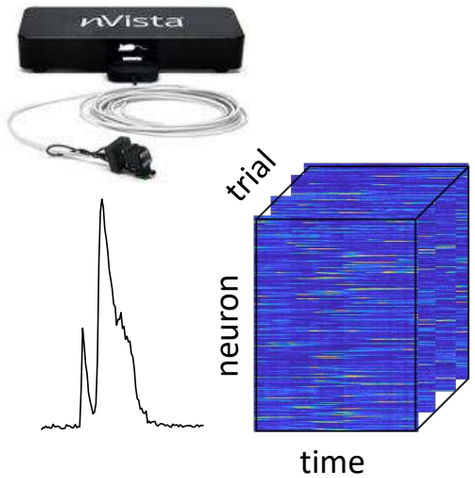
Where are we today?

- from matrices to tensors
- "sketches" for many purposes

- Compressed sensing
  - Motivating sparsity
  - Theorems
  - Examples
  - Smashed Filter
- Matrix completion
  - Examples
- Robust PCA
  - Examples
- Where are we today?
  - Tensors
  - Sketching
  - Computational paradigm

From Matrices to Tensors





Neuroscience (neuron x time x trial/ stimulus, for calcium imaging)

Find / exploit **low-rank tensors**

- … but notion of "rank" now has many ways to be generalized (CP, Tucker, Tensor Train, Hierarchical Tucker, Tensor Ring, etc.)
- Used for **recovery**, **analysis**, **compression**, etc.

Many of the matrix concepts (matrix completion, RCPA) extend though generally not as well, and algorithms are slower

Active area of research

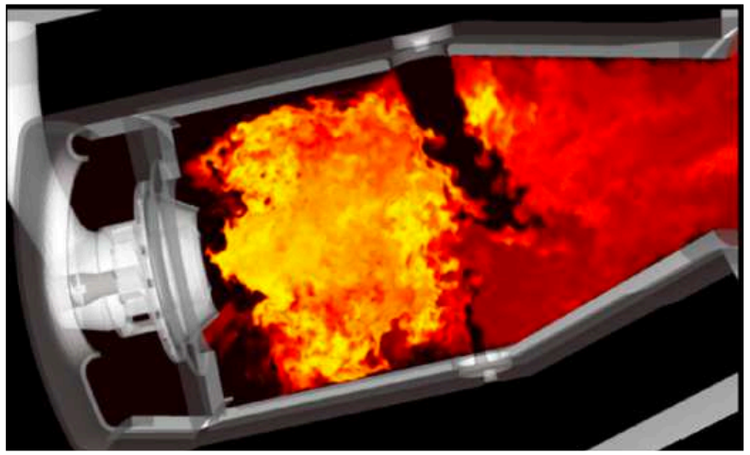
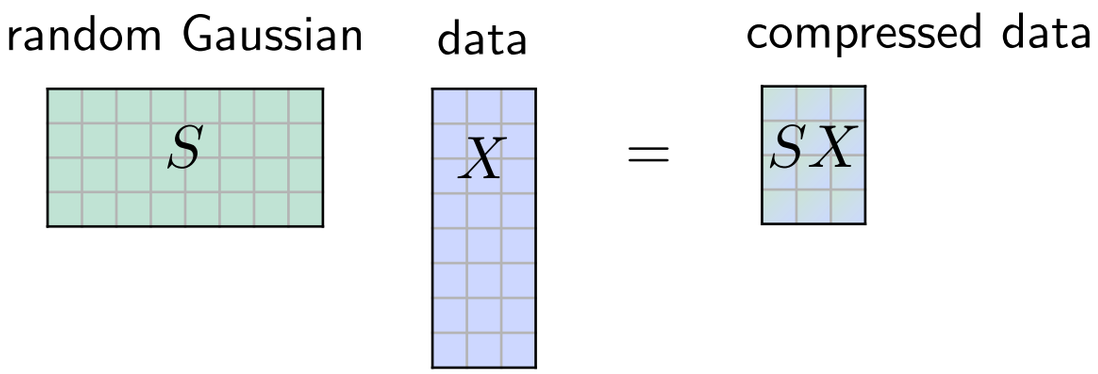


Image: Center for Turbulence Research (Stanford)

Sketching and randomized linear algebra



Example from Paschout et al. (2010)

650k SNPs. Predict *ancestry*

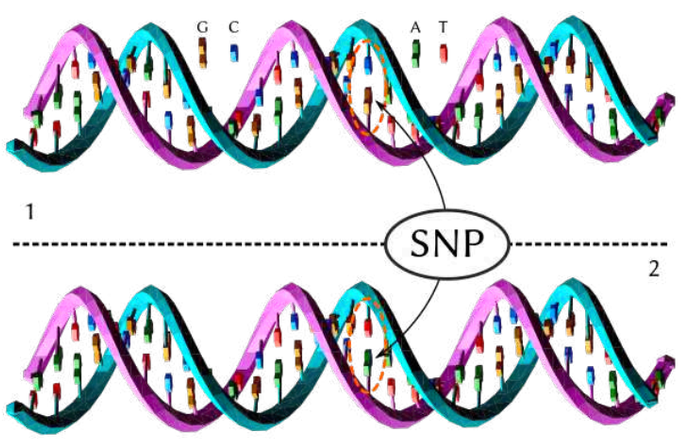
Use **interpolative decomposition**

(like PCA but more interpretable)

One use is for working with giant out-of-core datasets

Ex: UK Biobank data, 500k x 500k

Single-nucleotide polymorphism (SNP)



The upper DNA molecule differs from the lower DNA molecule at a single base-pair location (a G/A polymorphism). Credit: wikipedia

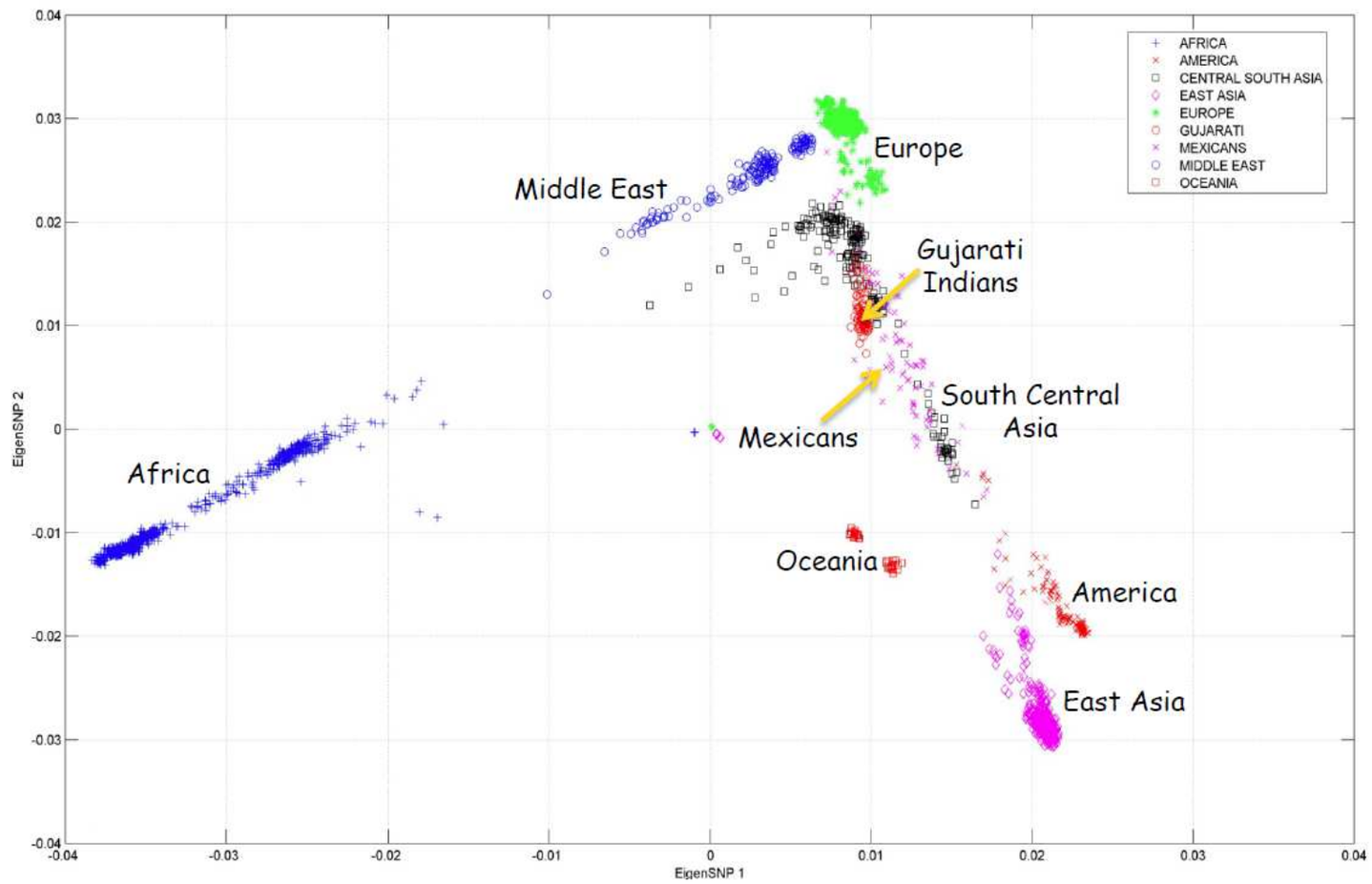


Image from Drineas (2019).

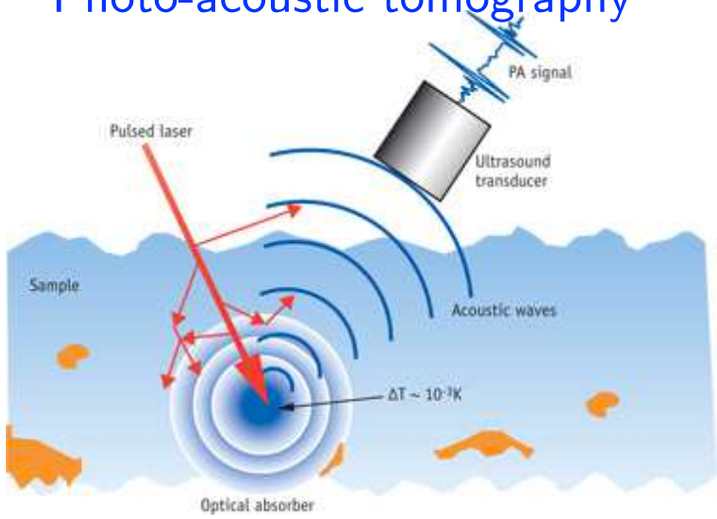
Computational Paradigm

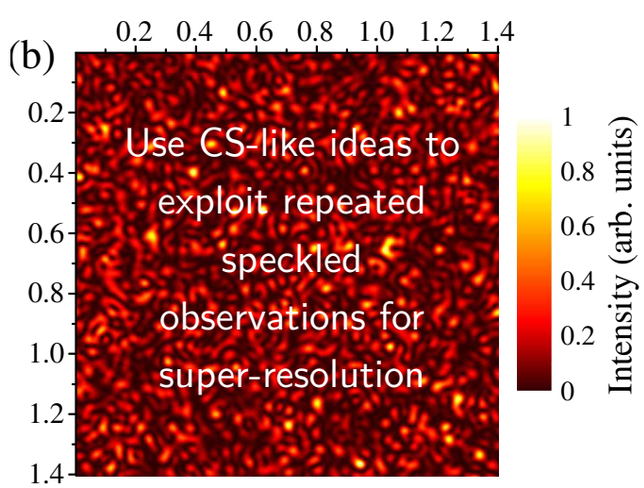
Biggest effect of compressed sensing may be helping to spread the idea of computational processing (as compared to traditional *linear processing*)

$$\min_x \underbrace{\mathcal{L}(x; \text{data})}_{\text{loss function}} + \lambda \cdot \text{regularizer}(x)$$

(not a new idea of course!)

Photo-acoustic tomography





Powerful because you only need to specify the **forward model**

i.e., how the data were collected

Our research group has successfully applied this to:

- Photo-acoustic tomography
- Fluorescent microscopy imaging
- Ultrasound
- MEG (Magnetoencephalography)

Thanks for your attention

Generic Compressed Sensing References

Surveys, general:

- *An Introduction To Compressive Sampling,* Wakin and Candes, IEEE Sig. Proc. Mag, 2008
- *Compressive Sensing Fundamentals* (book chapter), Wakin, 2013

Surveys/monographs/books with proofs:

- *A Mathematical Introduction to Compressive Sensing*, Foucart and Rauhut, 2013 (Birkhauser)
  - Exhaustive in detail, many interesting/tangential topics
- *Compressive Sensing and Structured Random Matrices*, Rauhut, 2010, monograph/habilitation
  - The basis for his book with Foucart but not as overwhelming. Recommended!
  - http://www.mathc.rwth-aachen.de/~rauhut/files/LinzRauhut.pdf

Original papers:

- Donoho 2004, Candes, Tao, Romberg 2004
- Extends prior work by Donoho, Huo, Elad, Bruckstein, Tropp, Gribonval, Nielsen, Temlyakov, etc.
- See Rauhut or Intro of my PhD thesis for details (http://resolver.caltech.edu/CaltechTHESIS:06022011-152525054)