# ECE 18-898G: Special Topics in Signal Processing: Sparsity, Structure, and Inference

## Introduction

Yuejie Chi

Department of Electrical and Computer Engineering

Carnegie Mellon University

Spring 2018*

*Slides adapted from Yuxin Chen@Princeton.

# What is sparsity?

A signal is said to be sparse when most of its components vanish.

- Formally, $\boldsymbol{x} \in \mathbb{R}^p$ is said to be $k$-sparse if it has at most $k$ nonzero entries

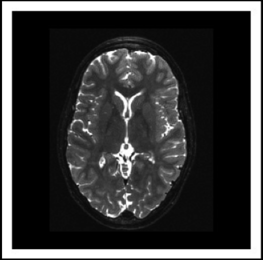
# What is sparsity?

A signal is said to be sparse when most of its components vanish.

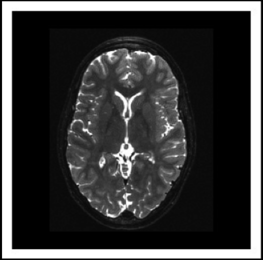
- Formally, $\boldsymbol{x} \in \mathbb{R}^p$ is said to be $k$-sparse if it has at most $k$ nonzero entries
- Think of a $k$-sparse signal as having $k$ degrees of freedom

*Engineers wish to describe / approximate data in the most parsimonious terms!*

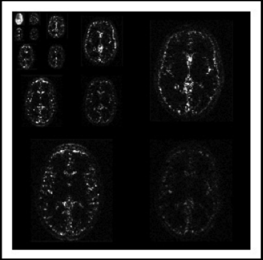
# Only a small number of parameters matter



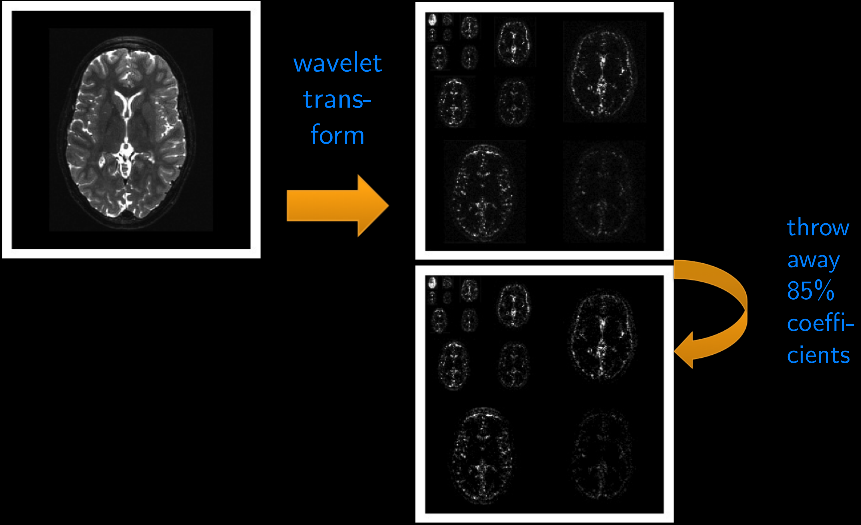
# Only a small number of parameters matter



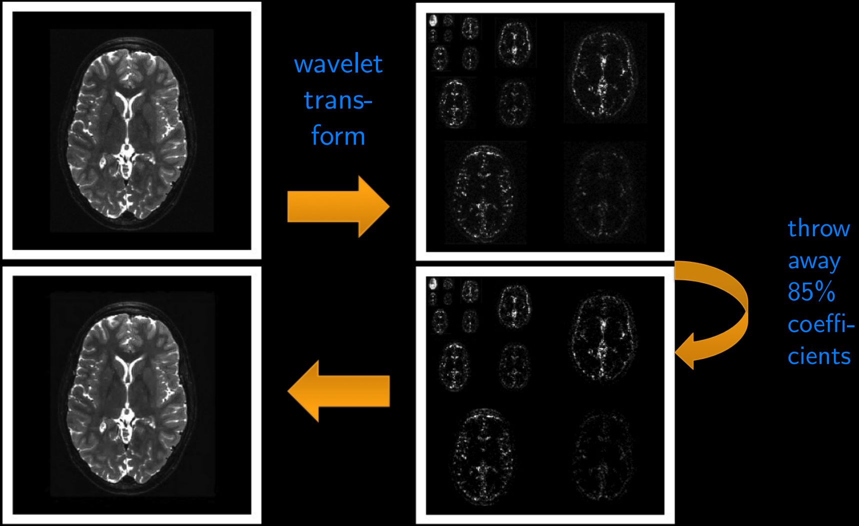
wavelet
trans-
form



# Only a small number of parameters matter



# Only a small number of parameters matter




Signal is very sparse in some transform domain (e.g. wavelet)

# Only a small number of parameters matter

- Compute $10^6$ wavelet coeffients
- Keep only the $25K$ largest coefficients
- Inverse wavelet transform

1 megapixel image

25k term approximation

# Only a small number of parameters matter

Raw: 15MB

# Only a small number of parameters matter

Raw: 15MB

JPEG: 150KB

There is (almost) no loss in quality between the raw image and its JPEG compressed form

# General philosophy

We are drowning in information and starving for knowledge

Rutherford Roger

- Massive data acquisition
- Most data is redundant and can be thrown away

# General philosophy

We are drowning in data and starving for information

Rutherford Roger

- Massive data acquisition
- Most data is redundant and can be thrown away

# General philosophy

We are drowning in data and starving for information

Rutherford Roger

- Massive data acquisition
- Most data is redundant and can be thrown away

Will such "information sparsity" be useful in data acquisition, statistical inference and information recovery?

# Advantages of sparsity

- Interpretability of our estimate / fitted model
  - particularly important when sample size $\ll$ # unknowns

# Advantages of sparsity

- Interpretability of our estimate / fitted model
  - particularly important when sample size $\ll$ # unknowns
- Computational convenience
  - in many cases we have scalable procedures to promote sparsity

# Advantages of sparsity

- Interpretability of our estimate / fitted model
  - particularly important when sample size $\ll$ # unknowns
- Computational convenience
  - in many cases we have scalable procedures to promote sparsity
- "Bet on sparsity" principle
  - *use a procedure that does well in sparse problems, since no procedure does well in dense problems*
  - "less is more": sparse model might be easier to estimate than dense models
  - Occam's razor

# Example: compressed sensing

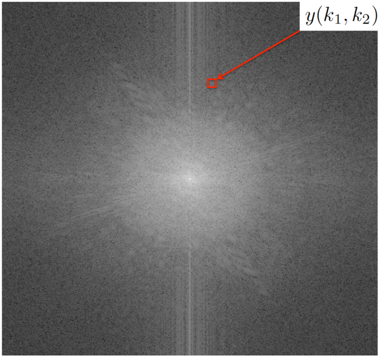
# Magnetic Resonance Imaging (MRI)

MR scanner

MR image

*K. Pauly, G. Gold, RAD220*

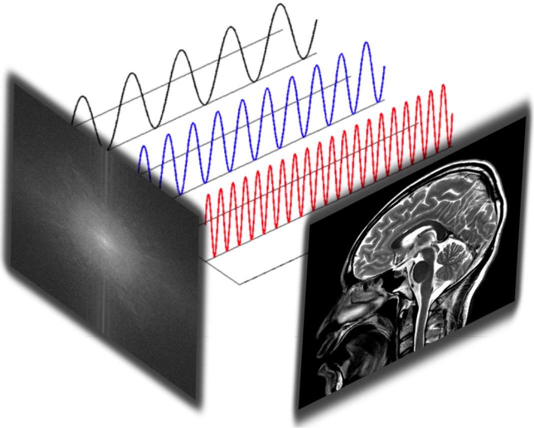
# What an MRI machine sees




Measured data $y(k_1, k_2)$ $\longleftarrow$ Fourier transform of image $f(x_1, x_2)$

# Fourier transform

$$y(k_1, k_2) \approx \sum_{x_1} \sum_{x_2} f(x_1, x_2) \mathrm{e}^{-i2\pi(k_1 x_1 + k_2 x_2)}$$



# How do we form an image?

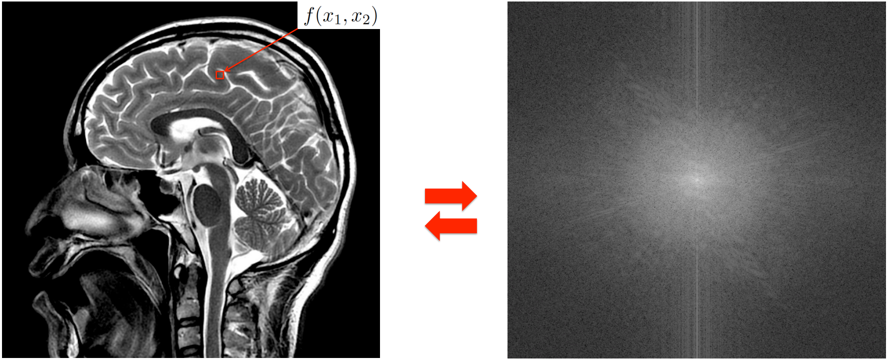
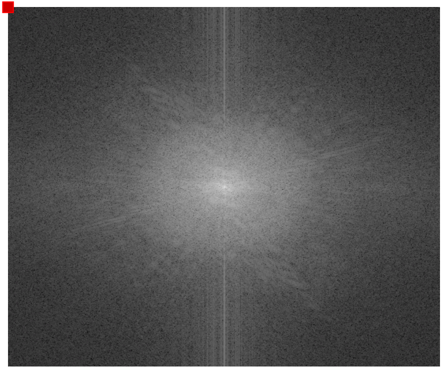


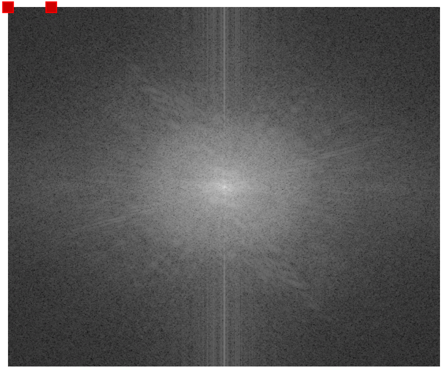

image $f(x_1, x_2)$ $\longleftarrow$ inverse Fourier transform of measurements

$$f(x_1, x_2) \;\approx\; \sum\sum y(k_1, k_2)\mathrm{e}^{i2\pi(k_1 x_1 + k_2 x_2)}$$

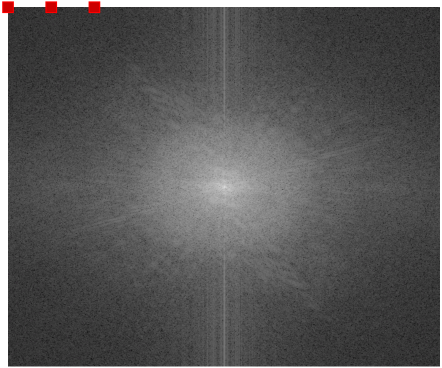
# MRI data collection is inherently slow



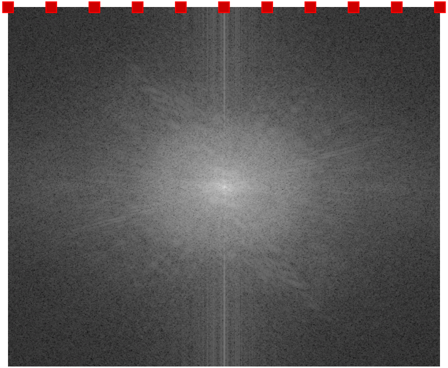
# MRI data collection is inherently slow



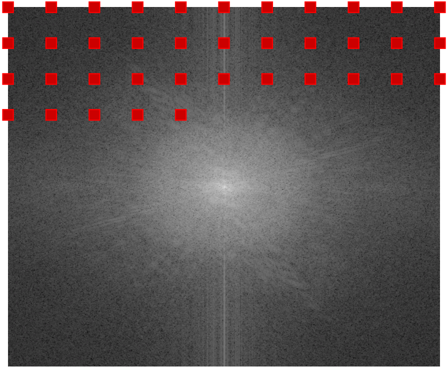
# MRI data collection is inherently slow



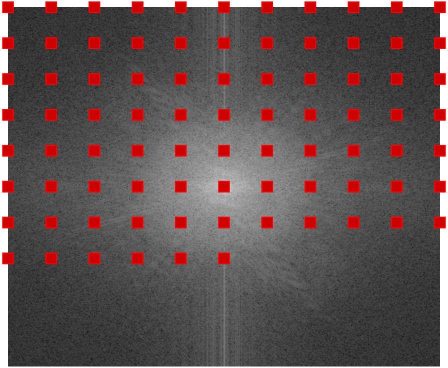
# MRI data collection is inherently slow



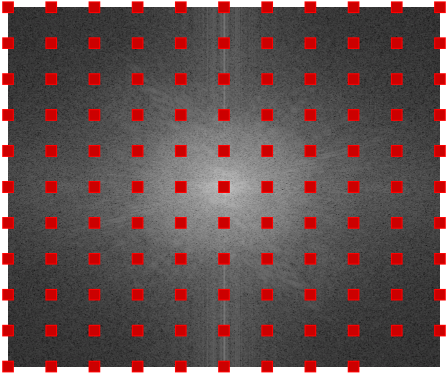
# MRI data collection is inherently slow



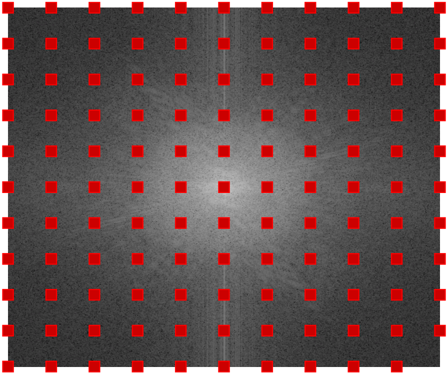
# MRI data collection is inherently slow



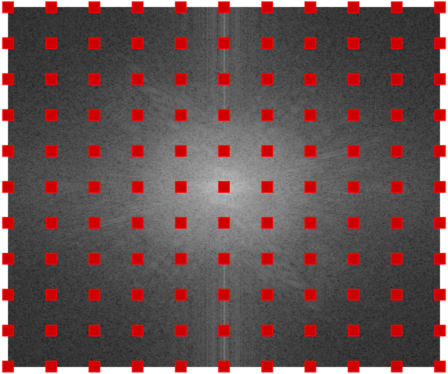
# MRI data collection is inherently slow



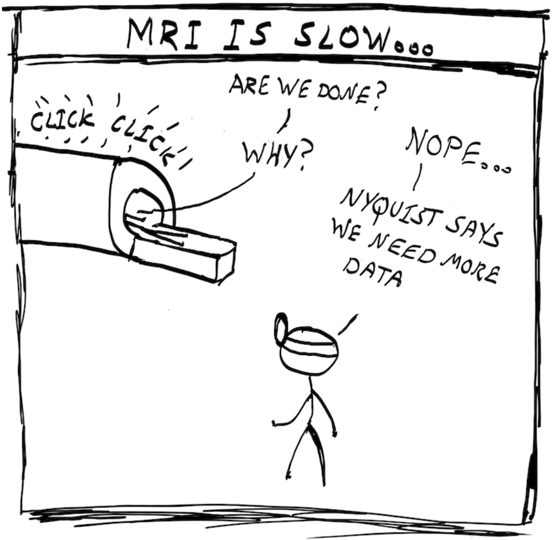
# MRI data collection is inherently slow



# MRI data collection is inherently slow



Done!


MRI IS SLOW...
CLICK CLICK CLICK
ARE WE DONE?
WHY?
NOPE...
NYQUIST SAYS WE NEED MORE DATA

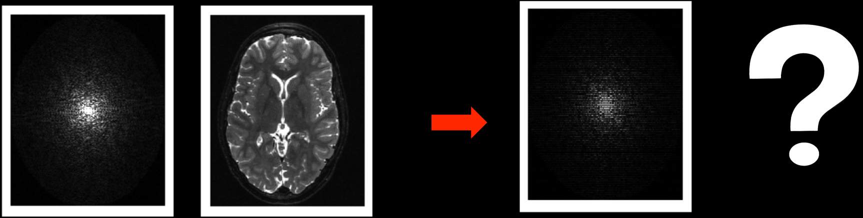
*M. Lustig*

# Fact: impact of MRI on children health is limited


Children cannot stay still or breathhold!

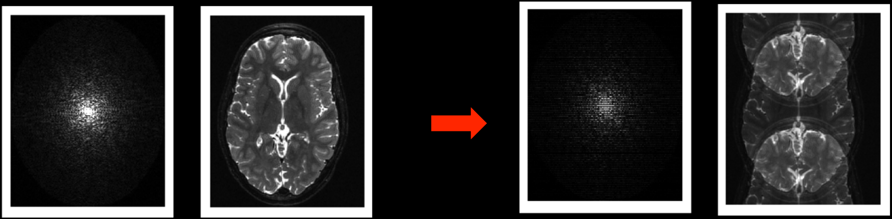
- (deep) anesthesia required
- respiration suspension

# Is it possible to take fewer samples to reduce scan time?



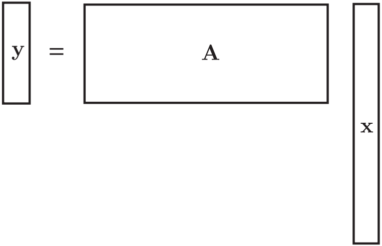
uniform undersampling by a factor of 2

# Is it possible to take fewer samples to reduce scan time?



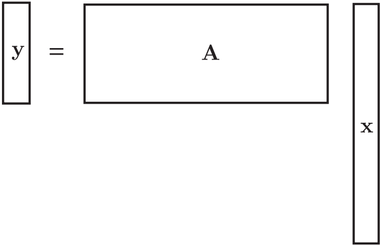
uniform undersampling by a factor of 2

# Fewer equations than unknowns!




How can we possibly solve an underdetermined system?

# Fewer equations than unknowns!




How can we possibly solve an underdetermined system?

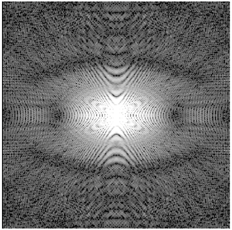
Carl Friedrich Gauss

We need at least as many equations as unknowns!

# A surprising experiment

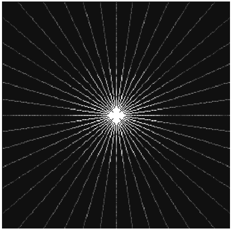
# A surprising experiment

Fourier transform



# A surprising experiment

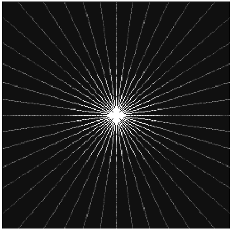
Fourier transform



highly subsampled

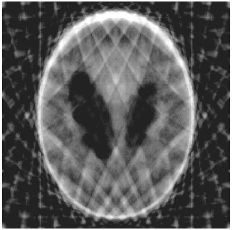
# A surprising experiment

Fourier transform



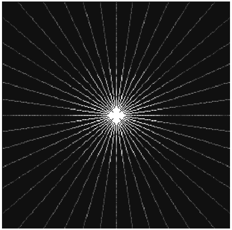
highly subsampled

classical reconstruction



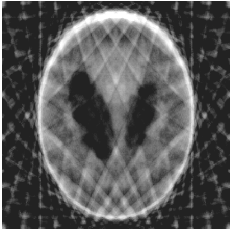
# A surprising experiment

Fourier transform



highly subsampled

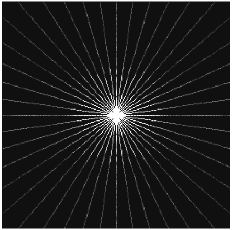
classical reconstruction



compressed sensing reconstruction

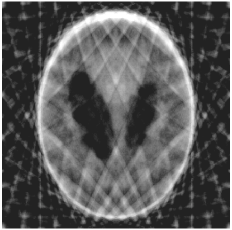
# A surprising experiment

Fourier transform



highly subsampled

classical reconstruction

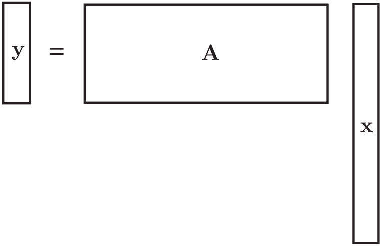


compressed sensing reconstruction

CS algorithm:

$$\min \quad \sum_x ||\nabla f(x)||_1 \text{ subj. to data constraints}$$

# Structured solutions

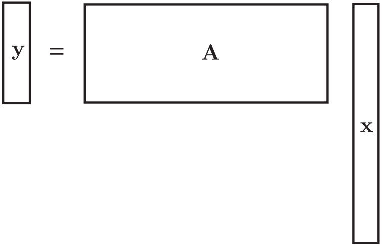



How can we possibly solve?

Need some structure

$\boldsymbol{x}$ is $k$-sparse $\rightarrow$ at most $k$ degrees of freedom

# Ingredients for success




- Exploit signal structure: sparsity
- Recovery via efficient algorithms (e.g. convex optimization)
- Incoherent sensing mechanism

# Translation to practice...

*Rather than taking nearly six minutes with multiple breath-holds, a Cardiac Cine scan can now be done within 25 seconds – in free-breathing.*

News | February 21, 2017

## FDA Clears Compressed Sensing MRI Acceleration Technology From Siemens Healthineers

- *New technology employs iterative reconstruction to produce high-quality MR images at a rapid rate with zero diagnostic information loss*
- *Compressed Sensing Cardiac Cine – the technology's first application – enables diagnostic cardiac imaging of patients with arrhythmias or respiratory problems*

Siemens Healthineers has announced that the Food and Drug Administration (FDA) has cleared the company's revolutionary Compressed Sensing technology, which slashes the long acquisition times

# Going beyond sparsity

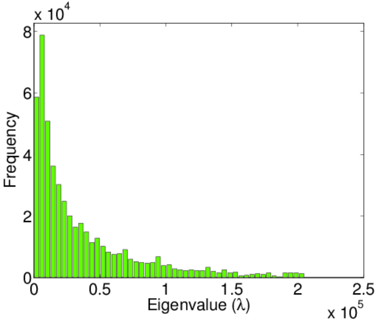
# Netflix challenge: predict unseen ratings

# Can we infer the missing entries?

$$\begin{bmatrix} \checkmark & ? & ? & ? & \checkmark & ? \\ ? & ? & \checkmark & \checkmark & ? & ? \\ \checkmark & ? & ? & \checkmark & ? & ? \\ ? & ? & \checkmark & ? & ? & \checkmark \\ \checkmark & ? & ? & ? & ? & ? \\ ? & \checkmark & ? & ? & \checkmark & ? \\ ? & ? & \checkmark & \checkmark & ? & ? \end{bmatrix}$$

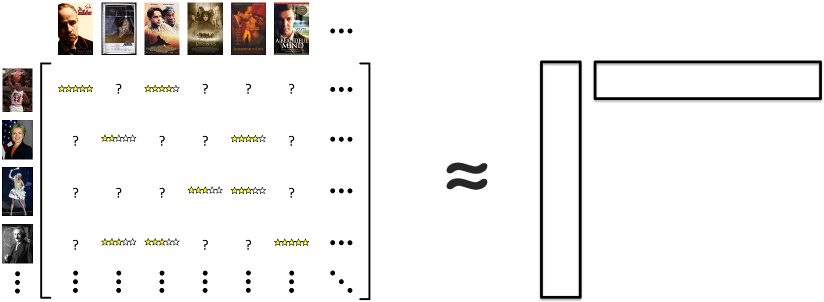
- Underdetermined system (more unknowns than revealed entries)
- Seems hopeless

# What if unknown matrix has structure?




A few factors explain most of the data

# What if unknown matrix has structure?

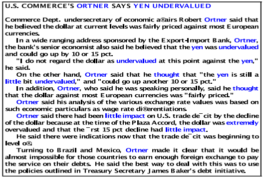


A few factors explain most of the data $\longrightarrow$ low-rank approximation

# Big data

images


videos



U.S. COMMERCE'S ORTNER SAYS YEN UNDERVALUED

Commerce Dept. undersecretary of economic a®airs Robert Ortner said that he believed the dollar at current levels was fairly priced against most European currencies.

In a wide ranging address sponsored by the Export-Import Bank, Ortner, the bank's senior economist also said he believed that the yen was undervalued and could go up by 10 or 15 pct.

"I do not regard the dollar as undervalued at this point against the yen," he said.

On the other hand, Ortner said that he thought that "the yen is still a little bit undervalued," and "could go up another 10 or 15 pct."

In addition, Ortner, who said he was speaking personally, said he thought that the dollar against most European currencies was "fairly priced."

Ortner said his analysis of the various exchange rate values was based on such economic particulars as wage rate di®erentiations.

Ortner said there had been little impact on U.S. trade de¯cit by the decline of the dollar because at the time of the Plaza Accord, the dollar was extremely overvalued and that the ¯rst 15 pct decline had little impact.

He said there were indications now that the trade de¯cit was beginning to level o®.

Turning to Brazil and Mexico, Ortner made it clear that it would be almost impossible for those countries to earn enough foreign exchange to pay the service on their debts. He said the best way to deal with this was to use the policies outlined in Treasury Secretary James Baker's debt initiative.

text


web data

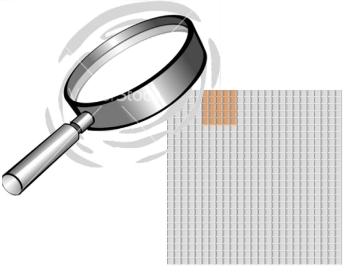
Huge data sizes
but often
low-dimensional structure

Engineering applications: unknown matrix is often (approx.) low rank

# Low-rank matrix completion?

| | | | | |
|---|---|---|---|---|
| 3 | 2 | 4 | 2 | 1 |
| 4 | 2 | 6 | 4 | 2 |
| 3 | 1 | 5 | 4 | 2 |
| 3 | 1 | 4 | 3 | 1 |
| 1 | 0 | 3 | 3 | 2 |

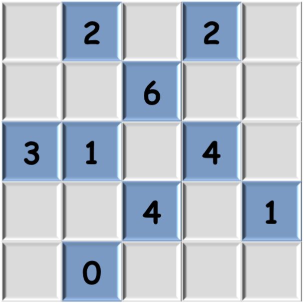
Ground truth



$50 \times 50$ low-rank

matrix

# Another surprising experiment




Observed samples

# Another surprising experiment

| | | | | |
|---|---|---|---|---|
| | 2 | | 2 | |
| | | 6 | | |
| 3 | 1 | | 4 | |
| | | 4 | | 1 |
| | 0 | | | |

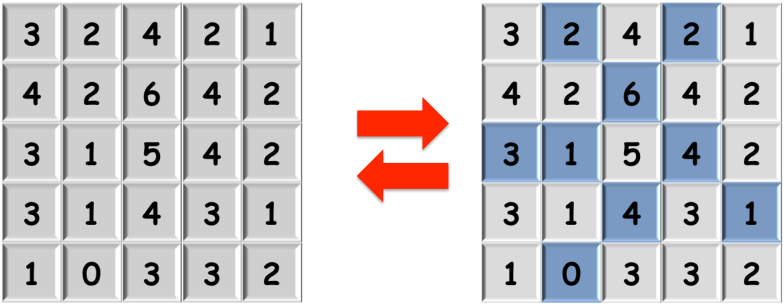
Observed samples

| | | | | |
|---|---|---|---|---|
| 3 | 2 | 4 | 2 | 1 |
| 4 | 2 | 6 | 4 | 2 |
| 3 | 1 | 5 | 4 | 2 |
| 3 | 1 | 4 | 3 | 1 |
| 1 | 0 | 3 | 3 | 2 |

Estimate via nuclear norm min

minimize $\underbrace{\text{sum-of-singular-values}}_{\text{nuclear norm}}$ subj. to data constraints

# Another surprising experiment

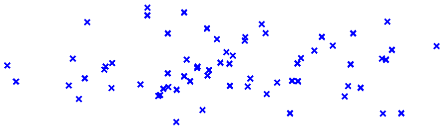



Ground truth

Estimate via nuclear norm min

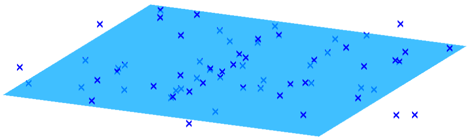
minimize $\underbrace{\text{sum-of-singular-values}}_{\text{nuclear norm}}$ subj. to data constraints

# Another problem: principal component analysis



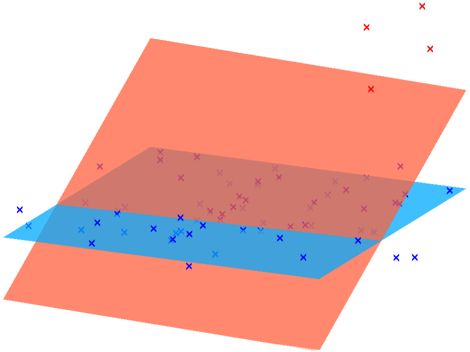
$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{x}_1 & \dots & \boldsymbol{x}_n \end{bmatrix}$$

# Another problem: principal component analysis



minimize $\|\boldsymbol{X} - \boldsymbol{L}\|$ subject to $\operatorname{rank}(\boldsymbol{L}) \leq k$

# Another problem: principal component analysis



minimize $\|\boldsymbol{X} - \boldsymbol{L}\|$ subject to $\operatorname{rank}(\boldsymbol{L}) \le k$

# Robust principal component analysis

Recover low-dimensional structure from corrupted data

$$\boldsymbol{Y} = \boldsymbol{L} + \boldsymbol{S}$$

- $\boldsymbol{Y}$: data matrix (observed)
- $\boldsymbol{L}$: low-rank component (unobserved)
- $\boldsymbol{S}$: sparse outliers (unobserved)

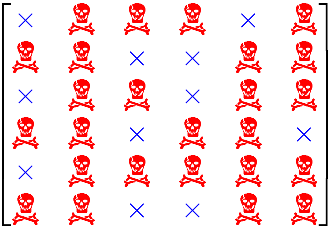
$$\begin{bmatrix} \times & \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times & \times \end{bmatrix}$$

# Robust principal component analysis

Recover low-dimensional structure from corrupted data

$$\boldsymbol{Y} = \boldsymbol{L} + \boldsymbol{S}$$

- $\boldsymbol{Y}$: data matrix (observed)
- $\boldsymbol{L}$: low-rank component (unobserved)
- $\boldsymbol{S}$: sparse outliers (unobserved)

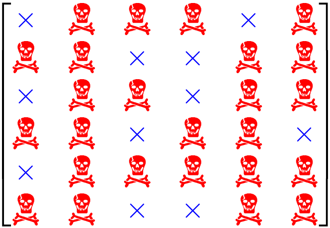


# Robust principal component analysis

Recover low-dimensional structure from corrupted data

$$\boldsymbol{Y} = \boldsymbol{L} + \boldsymbol{S}$$

- $\boldsymbol{Y}$: data matrix (observed)
- $\boldsymbol{L}$: low-rank component (unobserved)
- $\boldsymbol{S}$: sparse outliers (unobserved)

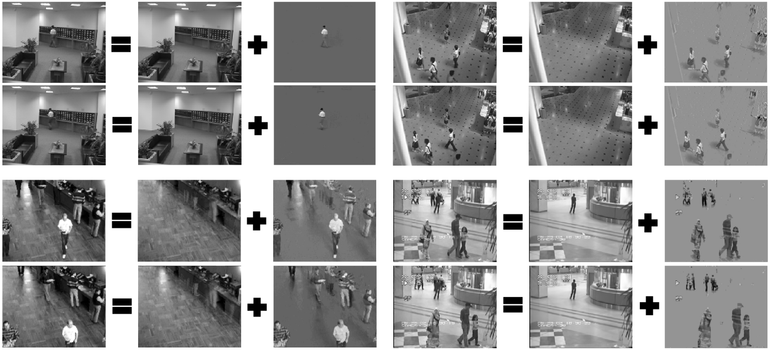


Can we separate $\boldsymbol{L}$ and $\boldsymbol{S}$?

# De-mixing by (non)convex programming

Spoiler: convex relaxation often enables perfect separation; nonconvex ones might work even better!

**Example:** separation of background (low-rank) and foreground (sparse) in videos



# Keywords of this course

- Low-dimensional structure (e.g. sparsity, low rank)

# Keywords of this course

- Low-dimensional structure (e.g. sparsity, low rank)
- Statistical models of data collection (incoherent sensing mechanism, often has some“randomness”)

# Keywords of this course

- Low-dimensional structure (e.g. sparsity, low rank)
- Statistical models of data collection (incoherent sensing mechanism, often has some“randomness”)
- Efficient algorithms (convex optimization, numerical methods, gradient descent, etc.)

# Keywords of this course

- Low-dimensional structure (e.g. sparsity, low rank)
- Statistical models of data collection (incoherent sensing mechanism, often has some“randomness”)
- Efficient algorithms (convex optimization, numerical methods, gradient descent, etc.)

*We can recover many low-dimensional structures of interest from highly incomplete data by efficient algorithms*

# Logistics

# Basic information

- Mon/Wed: 4:30 – 6:00 pm
- Instructor's office hours: Thursday 2-3:30pm, PH B25
- TA's office hours: Rohan Varma, Monday 10-12, PH B44

# Why you **should** consider taking this course

# Why you **should** consider taking this course

- There will be quite a few THEOREMS and PROOFS ...

# Why you **should** consider taking this course

- There will be quite a few THEOREMS and PROOFS ...
    - Promote deeper understanding of scientific results

# Why you should consider taking this course

- There will be quite a few THEOREMS and PROOFS ...
  - Promote deeper understanding of scientific results
- Nonrigorous / heuristic from time to time

# Why you should consider taking this course

- There will be quite a few THEOREMS and PROOFS ...
  - Promote deeper understanding of scientific results
- Nonrigorous / heuristic from time to time
  - "Nonrigorous" but grounded in rigorous theory
  - Help develop intuition

# Why you should consider taking this course

- There will be quite a few THEOREMS and PROOFS ...
    - Promote deeper understanding of scientific results
- Nonrigorous / heuristic from time to time
    - "Nonrigorous" but grounded in rigorous theory
    - Help develop intuition
- No exams!

# Tentative topics

First half: Fundamentals:

- Sparse representation
- Sparse linear regression and model selection
- Sparsity in graphical models
- Compressed sensing and sparse recovery
- Low-rank matrix recovery and matrix completion

# Tentative topics

First half: Fundamentals:

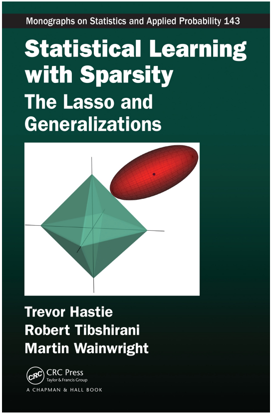
- Sparse representation
- Sparse linear regression and model selection
- Sparsity in graphical models
- Compressed sensing and sparse recovery
- Low-rank matrix recovery and matrix completion

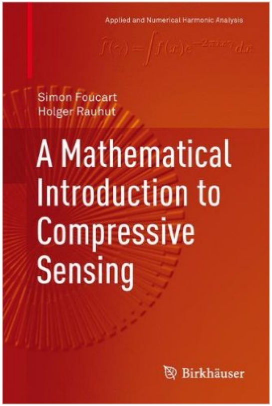
Second half: Special topics:

- phase retrieval / solving systems of quadratic equations
- Super-resolution and spectral estimation
- dictionary learning
- Neural networks
- implicit regularization: how optimization interacts with statistical inference

# Textbooks

We recommend these two books, but will not follow them closely ...



# Other useful references

- *Mathematics of sparsity (and a few other things)*, Emmanuel Candes, International Congress of Mathematicians, 2014.
- *Sparse and redundant representations: from theory to applications in signal and image processing*, Michael Elad, Springer, 2010.
- *Graphical models, exponential families, and variational inference*, Martin Wainwright, and Michael Jordan, Foundations and Trends in Machine Learning, 2008.
- *Introduction to the non-asymptotic analysis of random matrices*, Roman Vershynin, Compressed Sensing: Theory and Applications, 2010.
- *Convex optimization*, Stephen Boyd, and Lieven Vandenberghe, Cambridge University Press, 2004.
- *Topics in random matrix theory*, Terence Tao, American Mathematical Society, 2012.

More references will be provided at each lecture.

# Prerequisites

- linear algebra
- probability
- a programming language (e.g. Matlab, Python, ...)
- *knowledge in basic convex optimization is a plus*

# Prerequisites

- linear algebra
- probability
- a programming language (e.g. Matlab, Python, ...)
- *knowledge in basic convex optimization is a plus*
- *Concentration inequalities and non-asymptotic random matrix theory*

# Grading

- Homeworks (30%): $\sim$4 problem sets
- Midterm Paper Presentations (20%)
  - An in-class presentation on a selected paper from a given pool is arranged in lieu of the midterm.
  - About 20 min each, highlight at least one key result
- Term project (50%)

# Term project

Two forms

- literature review on a research **topic** (individual)
- original research (can be individual or a group of two)
  - *You are strongly encouraged to combine it with your own research*

# Term project

Two forms

- literature review on a research **topic** (individual)
- original research (can be individual or a group of two)
  - *You are strongly encouraged to combine it with your own research*

Three milestones

- Proposal (March 28): up to 2 pages (NIPS format). Plan early!
- Presentation (last week of class)
- Report (May 14): up to 4 pages with unlimited appendix

# Reference


[1] "*Mathematics of sparsity (and a few other things)*," E. Candes, *International Congress of Mathematicians*, 2014.

[2] "*Statistical learning with sparsity: the Lasso and generalizations*," T. Hastie, R. Tibshirani, and M. Wainwright, 2015.