Compressed Sensing
for Sparse and Low-Rank Models

David Gross
April 2017
Outline

- What this is NOT about: Deep Learning
- Three Examples of Compressed Sensing
What this talk is NOT about:

Deep neural networks
Deep Learning

These days: ML associated with *deep learning*:

- **Supervised learning**: neural nets trained on labeled data.

**Advantage:**

- Can treat highly complex situations for which no explicit model is known.

**Problem:**

- We don’t understand when and why it works.
Deep Learning

Successful *impersonation attacks:*
- Ability to treat complex tasks $\Rightarrow$ hard to understand

[Sharif et al., 2016]
What this talk IS about:

Compressed Sensing
Compressed Sensing

High-level principle:

Recover data with low-dimensional description from reduced set of measurements.

- Unsupervised: No learning phase with labeled data
- Requires explicit statistical model
- Often: complete and rigorous understanding
- Emphasize on provably efficient algorithms

Three examples:

4th example: Sparse PCA by Frank Vallentin
Sparsity

Learning data with sparse description:
Example from digital photography

- Typical JPG-compressed picture only 10% of raw data size
- ...in some sense 90% of data wasn’t actually necessary

Could one design process that directly records data in “compressed form”?
JPG based on *sparsity*:

- Divide image into blocks of $8 \times 8$ pixels
- Expand each block in *discrete cosine basis*:
- Turns out: natural images approx. sparse in this basis. (Example above: 7%).
- Keep only large coefficients.
Q: Can one recover picture perfectly from 7% of pixels?
**Sparsity: Geometry**

- Candidate image compatible with observed points form affine space.
- Which point in this high-dimensional space has sparsest description?
- ...solved by Donoho, Candes, Tao *et al.* around 2004 to great effect.
Compressed Sensing

**Thm.** [Candes, Tao, Donoho, et al. (2004)]

- Any signal with up to $r$ non-vanishing Fourier coefficients can be recovered exactly from $r (\log n)$ randomly chosen measurements.
- The signal minimizes the $\ell_1$-norm over the affine plane.

- Efficient *convex optimization*
- Handle tens of thousands of variables.

- Signals that can be synthesized out of few harmonics can be learned with few measurements and low computational effort.
Netflix.com

- Netflix is a US online movie platform
- In 2006, offered $1m for numerical solution of this problem:

\[
\begin{bmatrix}
\text{User} & 1 & 2 & 3 & 4 & 5 \\
\text{Video} & X & X & X & X & X
\end{bmatrix}
\]

- Users can rate videos
Netflix.com

- Netflix is a US online movie platform
- In 2006, offered $1m for numerical solution of this problem:

\[ \text{Users can rate videos} \]
\[ \text{Any given user will only rate small fraction} \]
\[ \text{Can one predict the missing entries?} \]
Prediction only possible if there are patterns.

Model:
- Taste vectors are linear combination of few prototypes
- Prototypes not assumed to be known!

In any case:
- Full matrix has low (column) rank.
Can one reconstruct a low-rank matrix from few measurements?

From singular-value decomposition:

\[ X = \sum_{i=1}^{r} s_i (\phi_i \otimes \psi_i^*), \quad s_i \in \mathbb{R}, \phi_i, \psi_i \in \mathbb{C}^n \]

...can hope that \( O(r n) \ll n^2 \) measurements are sufficient.
Low rank

Theorem [DG ’10; building on Candes and Tao ’10]. Choose a basis in the space of \((n \times n)\) matrices.

- Any rank-\(r\) matrix can be recovered exactly from \(O(rn \log^2 n)\) randomly chosen expansion coefficients w.r.t. the basis.
- The matrix minimizes the nuclear norm over the compatible affine space.

ARTICLE
Received 23 Aug 2016 | Accepted 20 Mar 2017 | Published xx xxx 2017

Experimental quantum compressed sensing for a seven-qubit system

C.A. Riofrío\(^1\), D. Gross\(^2\), S.T. Flammia\(^3\), T. Monz\(^4\), D. Nigg\(^4\), R. Blatt\(^4\) & J. Eisert\(^1\)
Blind source separation

Problem:
- Locate illegal broadcasters (black)...
- ...using antennas in known positions (red).

- Each antenna records mix of signals:
  \[ y_i(t) = \sum_u \frac{1}{d_{i,u}^2} x_u(t), \]
- ...attenuated by (unknown) squared distance.
- Task: Find the distances, from \( y_i(t) \)'s.
Blind source separation

Look at third-order *cumulant* of time series:

\[ T_{i,j,k} = \frac{1}{T} \sum_t y_i(t)y_j(t)y_k(t) \]

⇒ Unique tensor low-rank decomposition:

\[ T = \sum_{u=1}^{r} s_u (\delta_u \otimes \delta_u \otimes \delta_u) \]

in terms of distance vectors \( \delta_u \)

\[ (\delta_u)_i = \frac{1}{d_{u,i}} \]

Compare to matrix case:

\[ X_{i,j} = \sum_{i=1}^{r} s_i (\phi_i \otimes \psi_i^*) \]
Blind source separation

Theorem [2018].

...
Geometry of $\ell_1$-minimization

$\mathbb{R}^3$

$\circ = \pm e_i$

$\circ$ = 1-sparse

$\longrightarrow$ = 2-sparse

$\triangle$ = 3-sparse
Summary

Compressed sensing...

- ...allows for efficient learning of data with low-dim description
- ...comes with efficient algorithms based on convex optimization
- ...often amenable to mathematically rigorous treatment.

Thanks for your attention!