The Compressed Sensing Paradigm

MA4291 Presentation

Ang Yan Sheng
A0144836Y

6 Nov 2017

Contents

The problem with cameras

When you take a picture:

- Make $\sim 10^7$ measurements
- Change basis (discrete cosine, wavelet, etc.)
- Take largest $\sim 10^5$ components

What if measurement is expensive?

Infrared cameras	sensor cost
Astronomy	data bandwidth, power limit
Medical imaging	exposure, time

Data:	k-sparse vector	$x \in \Sigma_k \subseteq \mathbb{R}^n$
Measurement:	dot product	$\langle a_i, x \rangle$
Set of measurements:	matrix product	Ax, $A \in \mathbb{R}^{m \times n}$

Main Problem

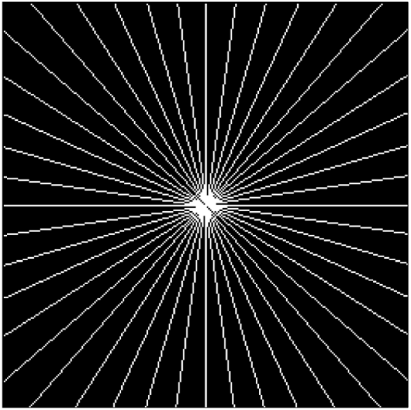
Given $k \ll n$, find measurement matrix $A \in \mathbb{R}^{m \times n}$ ($m \ll n$) such that any $x \in \Sigma_k$ can be efficiently recovered from Ax.

Need A injective on Σ_k, ie.

$$\begin{aligned} \{0\} &= \ker(A) \cap (\Sigma_k - \Sigma_k) \\ &= \ker(A) \cap \Sigma_{2k} \end{aligned}$$

Example (Candès-Romberg-Tao 2006)

$x =$ Logan-Shepp phantom



$A = 50\times$ undersampling
(frequency domain)

Example (Candès-Romberg-Tao 2006)

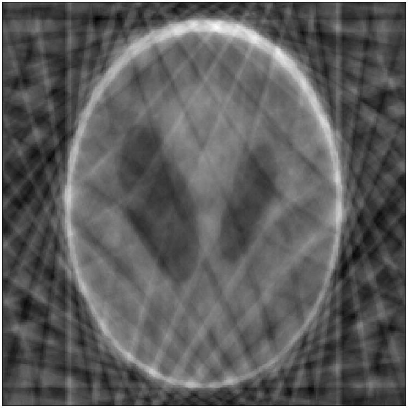
$x =$ Logan-Shepp phantom

$z = \ell_0$-minimisation

$$\min_z \|z\|_0 \qquad \text{with} \quad Az = Ax \qquad (P_0)$$

Example (Candès-Romberg-Tao 2006)

$x =$ Logan-Shepp phantom



$z = \ell_2$-minimisation

$$\min_z \|z\|_2 \qquad \text{with} \quad Az = Ax \qquad\qquad (P_2)$$

Example (Candès-Romberg-Tao 2006)

$x =$ Logan-Shepp phantom

$z = \ell_1$-minimisation

$$\min_z \|z\|_1 \qquad \text{with} \quad Az = Ax \qquad (P_1)$$

Null Space Property

For $T \subseteq [n]$, let v_T be v projected onto coordinates in T.

Definition

$A \in \mathbb{R}^{m \times n}$ *has NSP$_k$* if for all $v \in \ker A \backslash \{0\}$, and $T \subseteq [n]$ of size k,

$$\|v_T\|_1 < \|v_{T^c}\|_1.$$

(ie. largest k coordinates of v contains less than half the mass)

Theorem

A has NSP$_k$ $\iff$ (P_1) has unique solution $z = x$ for all $x \in \Sigma_k$.

$(\Leftarrow)$: $Av_T = A(-v_{T^c})$, so $\|v_T\|_1 < \|-v_{T^c}\|_1$.
$(\Rightarrow)$: $x \in \Sigma_k$ supported on T, with $Ax = Az$:

$$\begin{aligned} \|x\|_1 &\leq \|x - z_T\|_1 + \|z_T\|_1 \\ &< \|(x - z)_{T^c}\|_1 + \|z_T\|_1 = \|z\|_1. \end{aligned}$$

Restricted Isometry Property

Definition

Restricted isometry constant $\delta_k(A)$: minimal such that

$$\big|\|Ax\|_2^2 - \|x\|_2^2\big| \le \delta_k \|x\|_2^2 \qquad \text{for all } x \in \Sigma_k.$$

Note that $0 \le \delta_1(A) \le \delta_2(A) \le \cdots \le \delta_k(A)$.

Proposition

$$|\langle Ax, Az\rangle - \langle x, z\rangle| \le \delta_{2k}\|x\|_2\|z\|_2 \qquad \textit{for all } x, z \in \Sigma_k.$$

Assume that $\|x\|_2 = \|z\|_2 = 1$. Then

$$\text{LHS} = \left|\frac{\|A(x+z)\|_2^2 - \|A(x-z)\|_2^2}{4} - \frac{\|x+z\|_2^2 - \|x-z\|_2^2}{4}\right|$$

$$\le \delta_{2k}\frac{\|x+z\|_2^2 + \|x-z\|_2^2}{4} = \delta_{2k}\frac{\|x\|_2^2 + \|z\|_2^2}{2} = \text{RHS}.$$

Restricted Isometry Property

Theorem

If $\delta_{2k}(A) < 1/3$, then A has NSP_k.

Fix $v \in \ker A$, $T_0 \subseteq [n]$ the k largest entries in v,
$T_1 \subseteq [n]$ the k largest entries in $v_{T_0^c}$,
$T_2 \subseteq [n]$ the k largest entries in $v_{(T_0 \cup T_1)^c}$, ...

Note that $0 = Av = Av_{T_0} + Av_{T_1} + \cdots$, so

$$(1-\delta_k)\|v_{T_0}\|_2^2 \leq \|Av_{T_0}\|_2^2 = \langle Av_{T_0}, -(Av_{T_1} + Av_{T_2} + \cdots)\rangle$$

$$\leq \sum_{j\geq 1} |\langle Av_{T_0}, Av_{T_j}\rangle| \leq \delta_{2k}\|v_{T_0}\|_2 \sum_{j\geq 1} \|v_{T_j}\|_2$$

$$\sqrt{k}\sum_{j\geq 1}\|v_{T_j}\|_2 \leq k\sum_{j\geq 1}\max_{l\in T_j}|v_l| \leq \sum_{j\geq 1}\sum_{l\in T_{j-1}}|v_l| = \|v\|_1$$

$$\|v_{T_0}\|_1 \leq \sqrt{k}\|v_{T_0}\|_2 \leq \frac{\delta_{2k}}{1-\delta_k}\|v\|_1 < \frac{\|v\|_1}{2}.$$

Strategy

Let $\omega_{ij} \sim N(0,1)$, and

$$A = \frac{1}{\sqrt{m}} \begin{pmatrix} \omega_{11} & \cdots & \omega_{1n} \\ \vdots & \ddots & \vdots \\ \omega_{m1} & \cdots & \omega_{mn} \end{pmatrix}.$$

We want $\delta_{2k}(A) < 1/3$ w.h.p.:

- Show $\|Ax\|_2 \approx \|x\|_2$ w.v.h.p. for fixed $x \in \mathbb{R}^n$
- For each k-dimensional subspace in Σ_k:
  - Find points x_i "covering" $\mathbb{S}^{k-1}$
  - Show $\|Ax_i\|_2 \approx \|x_i\|_2 \implies \|Az\|_2 \approx \|z\|_2$ on $\mathbb{S}^{k-1}$
- Finish by union bound

Step 1

- Show $\|Ax\|_2 \approx \|x\|_2$ w.v.h.p. for fixed $x \in \mathbb{R}^n$

Assume $\|x\|_2 = 1$. Then for $0 < \delta < 1$,

$$\begin{aligned}
&\mathbb{P}\left(\left|\|Ax\|_2^2 - 1\right| \geq \delta\right) \\
&= \mathbb{P}\left(\left|(\omega_{11}x_1 + \cdots)^2 + \cdots + (\omega_{m1}x_1 + \cdots)^2 - m\right| \geq \delta m\right) \\
&= \mathbb{P}\left(\left|\omega_1^2 + \cdots + \omega_m^2 - m\right| \geq \delta m\right) \\
&\overset{?}{\leq} 2\exp(-C\delta^2 m)
\end{aligned}$$

Step 1

Proposition

Let $\omega_1, \ldots, \omega_m \sim N(0,1)$ *and* $0 < \delta < 1$. *Then*

$$\left.\begin{aligned} \mathbb{P}\left(\omega_1^2 + \cdots + \omega_m^2 \geq (1+\delta)m\right) \\ \mathbb{P}\left(\omega_1^2 + \cdots + \omega_m^2 \leq (1-\delta)m\right) \end{aligned}\right\} \leq \exp\left(-\frac{m}{2}\left(\frac{\delta^2}{2} - \frac{\delta^3}{3}\right)\right).$$

$$\mathbb{E}\exp(\lambda\omega^2) = 1/\sqrt{1-2\lambda}$$

$$\begin{aligned} \mathrm{LHS}_+ &= \mathbb{P}\left(\exp(\lambda(\omega_1^2 + \cdots + \omega_m^2 - (1+\delta)m)) \geq 1\right) \\ &\leq \mathbb{E}\exp(\lambda(\omega_1^2 + \cdots + \omega_m^2 - (1+\delta)m)) \\ &= (1-2\lambda)^{-m/2}\exp(-\lambda(1+\delta)m) \\ &= (1+\delta)^{m/2}\exp(-\frac{\delta}{2}m) \qquad \left(\lambda = \frac{\delta}{2(1+\delta)}\right) \\ &= \exp\left(-\frac{m}{2}(\delta - \ln(1+\delta))\right) \leq \mathrm{RHS}. \end{aligned}$$

Step 1

- Show $\|Ax\|_2 \approx \|x\|_2$ w.v.h.p. for fixed $x \in \mathbb{R}^n$

Assume $\|x\|_2 = 1$. Then for $0 < \varepsilon < 1$,

$$\begin{aligned} &\mathbb{P}\left(\left|\|Ax\|_2^2 - 1\right| \geq \delta\right) \\ &= \mathbb{P}\left(\left|(\omega_{11}x_1 + \cdots)^2 + \cdots + (\omega_{m1}x_1 + \cdots)^2 - m\right| \geq \delta m\right) \\ &= \mathbb{P}\left(\left|\omega_1^2 + \cdots + \omega_m^2 - m\right| \geq \delta m\right) \\ &\leq 2\exp(-C\delta^2 m), \end{aligned}$$

where we may take $C = 1/12$.

Step 2a

- Find points x_i "covering" $\mathbb{S}^{k-1} \subseteq \mathbb{R}^k$

Proposition

There is a set $S \subseteq \mathbb{S}^{k-1}$ of 9^k points such that

$$\min_{x_i \in S} \|z - x_i\|_2 \leq \frac{1}{4} \qquad \text{for all } z \in \mathbb{S}^{k-1}. \qquad (*)$$

Pick maximal $x_1, x_2, \ldots x_N \in \mathbb{S}^{k-1}$ with

$$\|x_i - x_j\|_2 > 1/4 \quad \text{for all } j < i.$$

Then $S = \{x_1, \ldots, x_N\}$ satisfies $(*)$.
$B(x_i, 1/8)$ pairwise disjoint and all contained in $B(0, 9/8)$:

$$N \operatorname{Vol}(B(0,1/8)) \leq \operatorname{Vol}(B(0,9/8)) \implies N \leq 9^k.$$

Step 2b

- Show $\|Ax_i\|_2 \approx \|x_i\|_2 \implies \|Az\|_2 \approx \|z\|_2$ on $\mathbb{S}^{k-1}$

Proposition

Take $S \subseteq \mathbb{S}^{k-1}$ as before, and let

$$\sup_{x_i \in S} \left| \|Ax_i\|_2^2 - 1 \right| = \delta/2, \qquad \sup_{z \in \mathbb{S}^{k-1}} \left| \|Az\|_2^2 - 1 \right| = \gamma.$$

Then $\gamma \le \delta$.

$\left| \|Au\|_2^2 - \|u\|_2^2 \right| \le \gamma \|u\|_2^2$ for all $u \in \mathbb{R}^k$, so

$$|\langle Au, Av\rangle - \langle u, v\rangle| \le \gamma \|u\|_2 \|v\|_2 \qquad \text{for all } u, v \in \mathbb{R}^k.$$

For any $z \in \mathbb{S}^{k-1}$, take $x_i \in S$ with $\|z - x_i\|_2 \le 1/4$:

$$\begin{aligned} \left| \|Az\|_2^2 - 1 \right| &= \left| \|Ax_i\|_2^2 - 1 + \langle A(z + x_i), A(z - x_i)\rangle - \langle z + x, z - x\rangle \right| \\ &\le \delta/2 + \gamma \|z + x\|_2 \|z - x\|_2 \le \delta/2 + \gamma/2. \end{aligned}$$

Step 3

- Finish by union bound.

We win if $\|Ax_i\|_2 \approx \|x_i\|_2$ for all $x_i \in S$ for all subspaces in Σ_k.

$$\mathbb{P}(\delta_k(A) > \delta) \leq \binom{n}{k}(9^k)(2\exp(-C(\delta/2)^2 m))$$

$$\leq \exp\left(k\ln(\frac{en}{k}) + k\ln 9 + \ln 2 - \frac{C\delta^2}{4}m\right).$$

Theorem

There exists an absolute $C' > 0$ such that if

$$m \geq C'\delta^{-2}(k\ln(en/k) + \ln(2/\varepsilon)),$$

then with $A = \frac{1}{\sqrt{m}}(\omega_{ij})_{i,j=1}^{m,n}$,

$$\mathbb{P}(\delta_k(A) \leq \delta) \geq 1 - \varepsilon.$$

Stability and robustness

Is ℓ_1-minimisation:

- *stable*, when x is only *compressible* (close to Σ_k)?
- *robust*, when $y = Ax + e$ with noise $\|e\|_2 \leq \eta$?

Replace (P_1) with

$$\min_z \|z\|_1 \qquad \text{with} \quad \|Az - y\|_2 \leq \eta \qquad (P_{1,\eta})$$

Definition

$A \in \mathbb{R}^{m \times n}$ *has* $\ell_2 RNSP_k(\rho, \tau)$ if for all $T \subseteq [n]$ of size k,

$$\|v_T\|_2 \leq \frac{\rho \|v_{T^c}\|_1}{\sqrt{k}} + \tau \|Av\|_2.$$

Stability and robustness

Proposition

A has $\ell_2 RNSP_k(\rho, \tau) \implies (P_{1,\eta})$ *has solution* z *with*

$$\|z - x\|_2 \leq \frac{C}{\sqrt{k}} \inf_{\hat{x} \in \Sigma_k} \|\hat{x} - x\|_1 + D\eta$$

with $C, D > 0$ *depending only on* ρ, τ.

Proposition

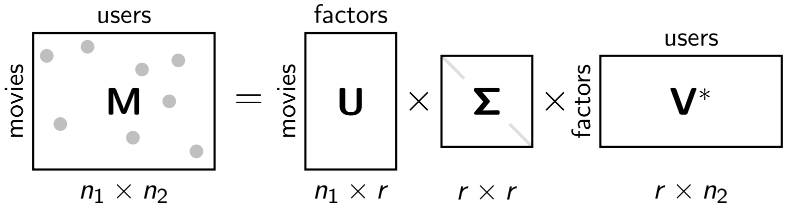
If $\delta_{2k}(A) < 1/3$, *then* A *has* $\ell_2 RNSP_k(\rho, \tau)$, *with* ρ, τ *depending only on* $\delta_{2k}(A)$.

Contents

The Netflix Prize (2006–2009)

- $n_1 = 1.8 \times 10^4$ movies
- $n_2 = 5 \times 10^5$ users
- 10^8 ratings (1%)
- Challenge: predict 3×10^6 unknown ratings

Low rank assumption, singular value decomposition:



$\Omega \subseteq [n_1] \times [n_2]$: sampling operator $\mathcal{R}_\Omega : \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}^{n_1 \times n_2}$
Recommender systems, global positioning, ...

Coherence

Coupon collector: $|\Omega|$ needs to be $O(n_2 \ln n_2)$

$$M = \begin{pmatrix} x_1 & x_2 & \cdots & x_{n_2} \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}: \ |\Omega| \text{ needs to be } O(n_1 n_2)!$$

Coherence

Definition

For a subspace $U \subseteq \mathbb{R}^n$ of dimension r, the *coherence of U* is

$$\frac{n}{r} \max_i \|\mathbf{P}_U \mathbf{e}_i\|_2^2.$$

eg. If $\mathbf{e}_i \in U$, then coherence $= n/r$.
eg. If U spanned by $(\pm 1/\sqrt{n}, \dots, \pm 1/\sqrt{n})$, then coherence $= 1$.

Definition (Incoherence assumptions)

A0: RS(**M**) and CS(**M**) have coherences $\leq \mu_0$.
A1: The entries of $\mathbf{UV}^*$ are $\leq \mu_1 \sqrt{r/n_1 n_2}$ in absolute value.

Review: matrix analysis

Let $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{n_1 \times n_2}$.

- Singular values $\sigma_1(\mathbf{X}) \geq \sigma_2(\mathbf{X}) \geq \cdots \geq 0$ (eigenvalues of $\sqrt{\mathbf{X}^*\mathbf{X}}$)
- Inner product $\langle \mathbf{X}, \mathbf{Y} \rangle = \mathrm{Tr}(\mathbf{X}^*\mathbf{Y})$
- Frobenius norm $\|\mathbf{X}\|_F = \langle \mathbf{X}, \mathbf{X} \rangle^{1/2}$
- Spectral norm $\|\mathbf{X}\| = \sigma_1(\mathbf{X}) = \max\{\|\mathbf{X}\mathbf{x}\|_2 : \|\mathbf{x}\|_2 = 1\}$
- Nuclear norm $\|\mathbf{X}\|_* = \sum_k \sigma_k(\mathbf{X})$

The main problem

Assume that $|\Omega| = m$ is chosen uniformly.

$$\min_{\mathbf{X}} \operatorname{rank}(\mathbf{X}) \qquad \text{with} \quad \mathcal{R}_\Omega \mathbf{X} = \mathcal{R}_\Omega \mathbf{M}$$

$$\min_{\mathbf{X}} \|\mathbf{X}\|_* \qquad \text{with} \quad \mathcal{R}_\Omega \mathbf{X} = \mathcal{R}_\Omega \mathbf{M} \qquad (P_*)$$

Theorem (Recht 2011)

If **M** *satisfies* **A0** *and* **A1**, $\beta > 1$, *and*

$$m \geq 32 \max(\mu_1^2, \mu_0) r (n_1 + n_2) \beta \ln^2(2n_2),$$

then (P_*) *has unique solution* $\mathbf{X} = \mathbf{M}$ *with probability at least*

$$1 - 6(n_1 + n_2)^{2-2\beta} \ln(n_2) - n_2^{2-2\sqrt{\beta}}.$$

Concentration bounds

Proposition (Bernstein inequality)

Let $X_1, \ldots, X_L$ be independent zero-mean random variables, with $|X_k| \leq M$. Let $\rho_k^2 = \mathbb{E}(X_k^2)$. Then for any $t > 0$,

$$\mathbb{P}\left(\left|\sum_{k=1}^{L} X_i\right| > t\right) \leq 2\exp\left(\frac{-t^2/2}{\sum_{k=1}^{L}\rho_k^2 + Mt/3}\right).$$

Proposition (Noncommutative Bernstein inequality)

Let $\mathbf{X}_1, \ldots, \mathbf{X}_L$ be independent zero-mean random $d_1 \times d_2$ matrices, with $\|\mathbf{X}_k\| \leq M$. Let $\rho_k^2 = \max(\|\mathbb{E}\mathbf{X}_k\mathbf{X}_k^*\|, \|\mathbb{E}\mathbf{X}_k^*\mathbf{X}_k\|)$. Then for any $t > 0$,

$$\mathbb{P}\left(\left\|\sum_{k=1}^{L} \mathbf{X}_i\right\| > t\right) \leq (d_1 + d_2)\exp\left(\frac{-t^2/2}{\sum_{k=1}^{L}\rho_k^2 + Mt/3}\right).$$

The dual problem

The *subdifferential* of $\|\cdot\|_*$ at $\mathbf{M}$ is

$$\partial\|\mathbf{M}\|_* = \{\mathbf{Y} : \|\mathbf{M}\|_* = \langle \mathbf{Y}, \mathbf{M}\rangle \text{ and } \|\mathbf{Y}\| \leq 1\}.$$

Proposition

If there exists $\mathbf{Y} \in \mathrm{Ran}(\mathcal{R}_\Omega) \cap \partial\|\mathbf{M}\|_*$, *then* (P_*) *has unique solution* $\mathbf{X} = \mathbf{M}$.

For all $\mathbf{Z} \in \ker(\mathcal{R}_\Omega)$, we have $\|\mathbf{M}+\mathbf{Z}\|_* \geq \langle \mathbf{M}+\mathbf{Z}, \mathbf{Y}\rangle = \|\mathbf{M}\|_*$.

Proposition

$\mathbf{Y} \in \partial\|\mathbf{M}\|_* \iff \mathbf{Y} = \mathbf{U}\mathbf{V}^* + \mathbf{W}$, *where* $\|\mathbf{W}\| \leq 1$, $\mathrm{CS}(\mathbf{W}) \perp \mathrm{CS}(\mathbf{U})$, *and* $\mathrm{RS}(\mathbf{W}) \perp \mathrm{RS}(\mathbf{V}^*)$.

Hence for some $\mathbb{R}^{n_1 \times n_2} = T \oplus T^\perp$,

$$\mathcal{P}_T(\mathbf{Y}) = \mathbf{U}\mathbf{V}^*, \quad \|\mathcal{P}_{T^\perp}(\mathbf{Y})\| \leq 1.$$

Proof sketch

We need the following to hold w.h.p.:

$$\frac{n_1 n_2}{m}\left\|\mathcal{P}_T\mathcal{R}_\Omega\mathcal{P}_T - \frac{m}{n_1 n_2}\mathcal{P}_T\right\| \le \frac{1}{2}, \qquad \|\mathcal{R}_\Omega\| \le \frac{8}{3}\sqrt{\beta}\ln(n_2),$$

and there exists $\mathbf{Y} \in \operatorname{Ran}(\mathcal{R}_\Omega)$ with

$$\|\mathcal{P}_T(\mathbf{Y}) - \mathbf{UV}^*\|_F \le \sqrt{\frac{r}{2n_2}}, \qquad \|\mathcal{P}_{T^\perp}(\mathbf{Y})\| < \frac{1}{2}.$$

Candès-Recht 2009: Decoupling on first 4 terms of infinite series
Candès-Tao 2009: Intensive combinatorial analysis, $O(\ln n_2)$ terms

Final step (Recht 2011)

Sample Ω with replacement, take $\mathcal{R}_\Omega$ with multiplicity.

References

- H. Boche, et al.
  "A Survey of Compressed Sensing."
  Compressed Sensing and its Applications, edited by H. Boche, et al.
  Birkhäuser Basel, 2015, pp. 1–37.
- E. Candès, B. Recht.
  "Exact Matrix Completion via Convex Optimization."
  Found. of Comput. Math., **9**(6), 2009, pp. 717–772.
- E. Candès, J. Romberg, T. Tao.
  "Robust Uncertainty Principles: Exact Signal Reconstruction From Highly Incomplete Frequency Information."
  IEEE Trans. Inform. Theory, **52**, 2006, pp. 489–509.
- B. Recht.
  "A Simpler Approach to Matrix Completion."
  J. Mach. Learn. Res., **12**, 2011, pp. 3413–3430.