Compressive Sensing, Lecture 3

Yesterday

- Proof of RIP property for Gaussian matrices
- Proof of signal recovery for RIP and RIPless sensing
- Convex optimization

Today

- Matrix rank minimization and nuclear norm
- Low rank plus sparse decomposition
- Transform invariant low-rank textures

Linear matrix equations

Problem

Recover matrix $\bar{X} \in \mathbb{R}^{n \times n}$ from $m \ll n^2$ linear measurements

$$b_k = \langle A_k, \bar{X} \rangle, \; k = 1, \ldots, m \rightsquigarrow b = \mathcal{A}(\bar{X}).$$

- In general this is impossible.
- Suppose we know rank$(\bar{X}) = r \ll n$.
  Could we get by with fewer than n^2 measurements?

Possible approach for low-rank $\bar{X}$

Take $m \ll n^2$ measurements $b = \mathcal{A}(\bar{X})$ and then solve

$$\begin{array}{ll} \min & \text{rank}(X) \\ & \mathcal{A}(X) = b. \end{array}$$

Applications

Matrix completion ("netflix" problem)

- Preference matrix M.
- We only observe a small portion of its entries M_{ij}.
- Fill in missing entries of M.

Sensor location

- n locations in $\mathbb{R}^d$.
- We only measure a subset of pairwise distances.
- Find the locations.

Linear system identification

- Dynamical linear system $\begin{cases} x(t+1) = Ax(t) + Bu(t) \\ y(t) = Cx(t) + Du(t) \end{cases}$
- Find A, B, C, D from observations of input $u(t)$ and output $y(t)$ for $t = 0, 1 \ldots, N$.

Norms

In compressed sensing there are three key norms: $\ell_1, \ell_2, \ell_\infty$.

Matrix norms

- Operator norm: $\|M\| := \max_{\|x\|_2 = 1} \|Mx\|_2 = \|\sigma(M)\|_\infty$
- Frobenius norm: $\|M\|_F := \sqrt{\sum_{i=1}^n \sum_{j=1}^n |M_{ij}|^2} = \|\sigma(M)\|_2$
- Nuclear norm $\|M\|_* := \|\sigma(M)\|_1$

Endow $\mathbb{R}^{n \times n}$ with the inner product: $\langle M, N \rangle = \text{trace}(M^\mathsf{T} N)$.

With this inner product:

- The Frobenius norm is the Hilbert space norm.
- The nuclear norm is the dual of the operator norm.

Rank minimization and convex relaxation

Nuclear norm heuristic for rank minimization problem (Fazel)

$$\begin{array}{ll} \min & \text{rank}(X) \\ & \mathcal{A}(X) = b \end{array} \quad \rightsquigarrow \quad \begin{array}{ll} \min & \|X\|_* \\ & \mathcal{A}(X) = b \end{array}$$

Nuclear norm

$$\|X\|_* := \sum_{i=1}^{n} \sigma_i(X).$$

Theorem (Fazel)

The nuclear norm is the convex envelope of the rank function on $\{M : \|M\| \leq 1\}$.

Fact

Nuclear norm minimization can be cast as a semidefinite program.

Dictionary vectors/matrices

Analogy between compressed sensing and low-rank recovery:

Parsimony concept	cardinality	rank
Hilbert space norm	Euclidean	Frobenius
Relaxed norm	ℓ_1	nuclear
Dual norm	ℓ_∞	operator
Convex relaxation	linear programming	semidefinite programming

The above can be seen in terms of the singular value map

$$X \mapsto \sigma(X).$$

Deterministic approach: restricted isometry

Restricted isometry property (RIP)

Given $\mathcal{A} : \mathbb{R}^{n \times n} \to \mathbb{R}^m$ and $k \in \{1, \ldots, m\}$, the k-isometry constant δ_k is the smallest $\delta \geq 0$ such that

$$(1-\delta)\|X\|_F^2 \leq \|\mathcal{A}(X)\|_2^2 \leq (1+\delta)\|X\|_F^2$$

for all $X \in \mathbb{R}^{n \times n}$ with rank$(X) \leq k$.

If $\delta_k < 1$, we say that A satisfies the RIP with constant δ_k.

Low-rank recovery and RIP property (Recht-Parrilo-Fazel)

Theorem

Assume $\bar{X} \in \mathbb{R}^{n \times n}$ *satisfies* rank$(\bar{X}) \leq r$.

- *If* $\delta_{2r}(\mathcal{A}) < 1$ *then* $\bar{X}$ *can be recovered (via, e.g., rank minimization) from* $b = \mathcal{A}(\bar{X})$.
- *If* $\delta_{4r}(\mathcal{A}) < \sqrt{2} - 1$ *then the nuclear norm solution recovers* $\bar{X}$.

RIP for randomly generated matrices:

Theorem

If $\mathcal{A}$ *is Gaussian, then* $\mathcal{A}$ *satisfies* $\delta_{4r} < \sqrt{2} - 1$ *if* $m \geq Crn$ *for some suitable constant* C.

The proofs of the above are extensions of the analogous proofs for RIP approach to compressed sensing.

A probabilistic approach

Probabilistic approach:

- Fix $\bar{X} \in \mathbb{R}^{n\times n}$ with rank$(\bar{X}) = r \ll n$.
- Pick random Gaussian $\mathcal{A} : \mathbb{R}^{n\times n} \to \mathbb{R}^m$ and put $b = \mathcal{A}(\bar{X})$.
- Let $\hat{X} := \text{argmin}_X \{\|X\|_* : \mathcal{A}(X) = b\}$.

Theorem (Candès and Recht)

If $m \geq r(6n - 5r)$ for $\beta > 1$, then recovery is exact with probability at least $1 - 2e^{(1-\beta)n/8}$.

Matrix completion

Problem

Assume M low rank and observe a subset of entries. Recover M.

- This is certainly an undetermined system of matrix equations.
- Unfortunately RIP fails in most interesting cases.

Model

Ω uniform random subset of $\{1, \ldots, n\} \times \{1, \ldots, n\}$.

When is matrix completion possible?

Bad cases

Observe that if

$$M = e_1 v^*$$

then recovery is not possible from a random small set of entries.

Likewise if

$$M = uv^*$$

where u, v are sparse vectors.

Coherence

- In the above cases the rows and/or columns of M are aligned with the basis vectors.
- Coherence is a measure of this kind of alignment.

Incoherence

Assume M has singular value decomposition

$$M = U\Sigma V^*,$$

and let $r = \text{rank}(M)$.

Coherence parameter

Smallest $\mu > 0$ such that for $i = 1, \ldots, n$

$$\|U^* e_i\|_2^2 \leq \frac{\mu r}{n}, \quad \|V^* e_i\|_2^2 \leq \frac{\mu r}{n}$$

and

$$|UV^*|_{ij} \leq \frac{\mu r}{n^2}$$

Incoherence and matrix completion

Consider the nuclear norm heuristic

$$\begin{array}{ll}\min & \|X\|_* \\ & X_{ij} = M_{ij}, \ \ (i,j) \in \Omega.\end{array}$$

Theorem (Candès & Recht)

Assume rank$(M) = r$ and Ω is a random set of size m. If

$$m \geq C\mu r(1+\beta)\log^2 n$$

then the solution to the nuclear norm heuristic is exact with probability at least

$$1 - n^{-\beta}.$$

Theoretical limits of matrix completion

The above result is nearly optimal:

Theorem (Candès & Tao)

No method can ensure recovery with high probability if

$$m \lesssim \mu \cdot nr \cdot \log n.$$

Neat connection with random graph theory

- For successful matrix recovery, the adjacency graph defined by entries in Ω must be connected.
- Given a bipartite graph, how many random edges should we pick to get a single component?

More general result

Assume $\{A_1, \dots, A_{n^2}\}$ orthonormal basis of $\mathbb{R}^{n\times n}$ and $M = U\Sigma V^*$.

Coherence

M has coherence μ with respect to $\{A_1, \dots, A_{n^2}\}$ if either

$$\max_k \|A_k\|^2 \leq \frac{\mu}{n}$$

or

$$\max_k \|P_U A_k\|^2 \leq \frac{\mu r}{n}, \ \max_k \|A_k P_V\|^2 \leq \frac{\mu r}{n}, \ \max_k |\langle A_k, UV^* \rangle \leq \frac{\mu r}{n^2}.$$

Theorem (Gross)

Exact recovery with probability at least $1 - n^{-\beta}$ if

$$m \geq C\mu r(1+\beta)\log^2 n.$$

Low rank + sparse decomposition

Separation problem

Suppose $M = L_0 + S_0$ where L_0 low rank and S_0 sparse.
If we observe M, could we recover L_0 and S_0?

Applications

- Robust PCA
- Latent variable detection
- Video surveillance

Nuclear & ℓ_1 heuristic (PCP)

$$\begin{array}{ll}\min & \|L\|_* + \lambda \cdot \|S\|_1 \\ & L + S = M.\end{array}$$

Recovery theorems

Low rank plus sparse separation:

- It is not possible in certain cases, e.g, if a matrix is both low rank and sparse.
- Recovery statements depend on matrix coherence.

Theorem

Suppose $M = L_0 + S_0 \in \mathbb{R}^{n\times n}$ *where*

$$\text{rank}(L_0) \le \frac{\rho_r n}{\mu}, \quad \|S_0\|_0 \le \rho_s n^2.$$

Then PCP succeeds with probability $1 - \frac{C}{n^{10}}$ *for* $\lambda = 1/\sqrt{n}$.

Matrix completion with corrupted data

Suppose entries of M may be both missing and corrupted. Can we recover M?

Extended PCP

$$\begin{aligned} \min \quad & \|L\|_* + \lambda \cdot \|S\|_1 \\ & L_{ij} + S_{ij} = M_{ij}, \quad (i,j) \in \Omega. \end{aligned}$$

Theorem

- $L_0 \in \mathbb{R}^{n\times n}$, *rank*$(L_0) \le \frac{\rho_r n}{\mu \log^2 n}$
- Ω *random set of size* cn^2, *where* $c \in (0,1)$.
- *Each observed entry is corrupted with probability* $\tau \le \tau_s$.

Then PCP succeeds with probability $1 - \frac{C}{n^{10}}$ *for* $\lambda = \frac{1}{\sqrt{cn}}$.

Relationship with matrix completion

Suppose we have no corruption.

MC: recovery via

$$\begin{aligned} \min \quad & \|L\|_* \\ & L_{ij} = M_{ij}, \quad (i,j) \in \Omega. \end{aligned}$$

PCP: recovery via

$$\begin{aligned} \min \quad & \|L\|_* + \lambda \cdot \|S\|_1 \\ & L_{ij} + S_{ij} = M_{ij}, \quad (i,j) \in \Omega. \end{aligned}$$

Under suitable conditions both yield the same answer.

The second one is a robust version of the first one.

Transform invariant low-rank textures (TILT)

Zhang-Ganesh-Liang-Ma

Suppose a low rank matrix is both corrupted and misaligned. Can we recover it?

Model

$$M \circ \tau = L_0 + S_0$$

L_0: low rank, S_0: sparse, τ: parametric deformation.

Problem

Given M, can we find the above decomposition?

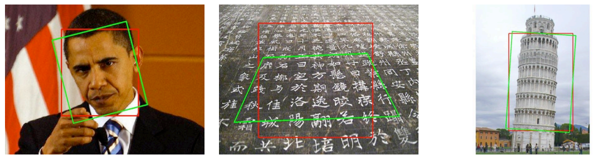
Approach

Find L, S, τ that solves

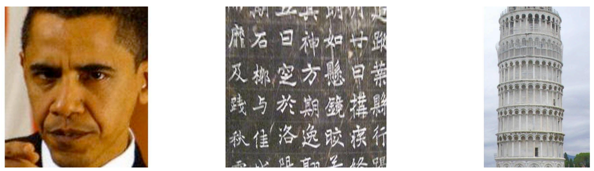
$$\begin{aligned} \min \quad & \|L\|_* + \lambda \cdot \|S\|_1 \\ & L + S = M \circ \tau. \end{aligned}$$

Examples

red windows indicate the original input



green windows indicate texture found



Main references for today's material

- Slides for this minicourse:
  `http://andrew.cmu.edu/user/jfp/UNencuentro`

- B. Recht, M. Fazel, P. Parrilo, "Guaranteed Minimum-Rank Solutions of Linear Matrix Equations via Nuclear Norm Minimization," *SIAM Review,* vol 52, no 3, pp. 471–501, 2010.
- V. Chandrasekaran, S. Sanghavi, P.A. Parrilo and A. Willsky , "Rank-Sparsity Incoherence for Matrix Decomposition," *SIAM Journal on Optimization,* vol. 21, issue 2, pp. 572–596, 2011
- E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust Principal Component Analysis?," *Journal of ACM,* vol 58, no 3, article no 11, 2011.
- Z. Zhang, A. Ganesh, X. Liang, and Y. Ma, "TILT: Transform Invariant Low-rank Textures," To Appear in International Journal of Computer Vision.