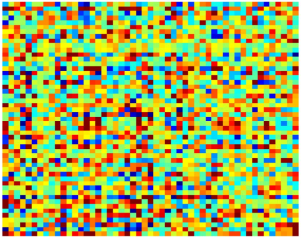
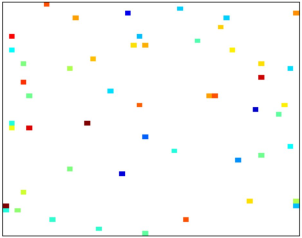
Robust Matrix Completion

Olga Klopp

CREST - Université Paris Ouest

Matrix Completion





Problem

Infer missing entries

Motivation

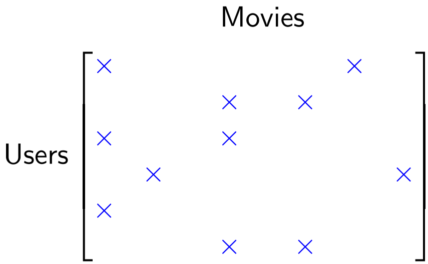
The Netflix problem

Example from the St Flour's Lectures by Emmanuel Candès

- Netflix database
  - About half a million users
  - About 18,000 movies
- People rate movies
- Sparsely sampled entries

The Netflix problem

- Netflix database
  - About half a million users
  - About 18,000 movies
- People rate movies
- Sparsely sampled entries



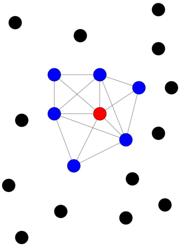
Problem

Complete the "Netflix matrix"

Global positioning from local distances

Example from the St Flour's Lectures by Emmanuel Candès

- Points $\{x_j\}_{1\leq j\leq n} \in \mathbb{R}^d$
- Partial information about distances $M_{ij} = \|x_i - x_j\|$



Example (Singer, Biswas et al.)

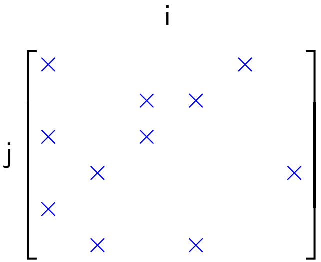
- Low-powered wirelessly networked sensors
- Each sensor can construct a distance estimate from nearest neighbor

Global positioning from local distances

- Points $\{x_j\}_{1 \le j \le n} \in \mathbb{R}^d$
- Partial information about distances
  $M_{ij} = \|x_i - x_j\|$

Example (Singer, Biswas et al.)

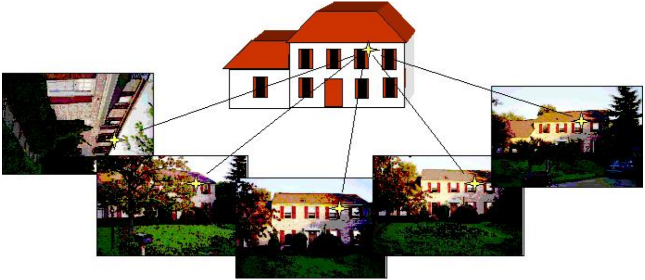
- Low-powered wirelessly networked sensors
- Each sensor can construct a distance estimate from nearest neighbor



Problem

Locate the sensors

Structure-from-motion problem



Problem

Recover 3D shape from 2D images

Structure-from-motion problem

- P features over F frames
- (x_{fp}, y_{fp}) = position of feature p at frame f
- $2F \times P$ measurement matrix

$$\begin{bmatrix} x_{11} & \cdots & x_{1P} \\ & \cdots & \\ x_{F1} & \cdots & x_{FP} \\ y_{11} & \cdots & y_{1P} \\ & \cdots & \\ y_{F1} & \cdots & y_{FP} \end{bmatrix}$$

Structure-from-motion problem

- P features over F frames
- (x_{fp}, y_{fp}) = position of feature p at frame f
- W a $2F \times P$ measurement matrix
- Occlusions $\rightarrow W$ partially filled in

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

Problem

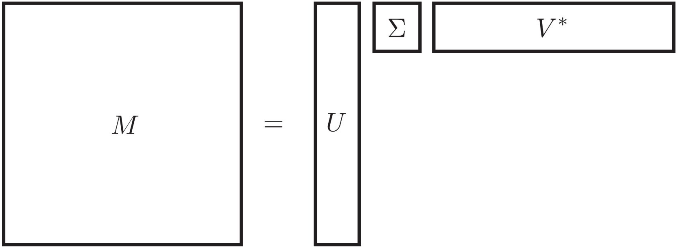
Recover the missing measurements

Low-dimensional structure

Engineering/scientific applications: unknown matrix has often (approx.) low rank

- Netflix matrix
- Sensor-net matrix: $\|x_i - x_j\|^2$, $\{x_i\} \in \mathbb{R}^d$
  - rank 2 if $d = 2$
  - rank 3 if $d = 3$
  - ...
- Structure-from-motion problem: $\mathrm{rank} \leq 4$
- Many others (e.g. machine learning, quantum tomography ...)

Dimension reduction



$M \in \mathbb{R}^{m_1 \times m_2}$ of rank r depends upon $(m_1 + m_2 - r)r$ free parameters

- $r \ll \min(m_1, m_2) \Rightarrow (m_1 + m_2 - r)r \ll m_1 m_2$
- Completion impossible if $n < (m_1 + m_2 - r)r$

Rank minimization

$$\begin{aligned} \textbf{minimize} \quad & \operatorname{rank}(A) \\ \textbf{subject to} \quad & A_{ij} = M_{ij} \\ & (i,j) \in E \end{aligned}$$

- (Usually) NP-hard

Trace - norm heuristics

Rank minimization

$$\begin{aligned} \textbf{minimize} \quad & \operatorname{rank}(A) \\ \textbf{subject to} \quad & A_{ij} = M_{ij} \\ & (i,j) \in E \end{aligned}$$

- (Usually) NP-hard

Trace-norm minimization

$$\begin{aligned} \textbf{minimize} \quad & \|A\|_* \\ \textbf{subject to} \quad & A_{ij} = M_{ij} \\ & (i,j) \in E \end{aligned}$$

- Convex relaxation (Fazel (2002))
- Trace norm:

$$\|A\|_* = \Sigma\, \sigma_i(A).$$

- Semidefinite program (SDP)

Trace Regression Model

$$Y_i = \mathrm{tr}(X_i^T M) + \xi_i, \quad i = 1, \dots n$$

- (X_i, Y_i), $i = 1 \dots n$ observations, $X_i \in \mathbb{R}^{m_1 \times m_2}$;
- $M \in \mathbb{R}^{m_1 \times m_2}$ unknown matrix of interest;
- ξ_i i.i.d. random errors: $\mathbb{E}\,\xi_i = 0$, $\mathbb{E}\,\xi_i^2 = \sigma^2$.

Trace Regression Model

$$Y_i = \mathrm{tr}(X_i^T M) + \xi_i, \quad i = 1, \dots n$$

- (X_i, Y_i), $i = 1 \dots n$ observations, $X_i \in \mathbb{R}^{m_1 \times m_2}$;
- $M \in \mathbb{R}^{m_1 \times m_2}$ unknown matrix of interest;
- ξ_i i.i.d. random errors: $\mathbb{E}\, \xi_i = 0$, $\mathbb{E}\, \xi_i^2 = \sigma^2$.

Problem

Recover M from (X_i, Y_i) when $m_1 m_2 \gg n$

Matrix Completion

$$X_i = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ & & \cdots & \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$

The design matrices X_i are **i.i.d** copies of a random matrix X having distribution Π on the set $\mathcal{X}$.

$$\mathcal{X} = \left\{ e_j(m_1)e_k^T(m_2), 1 \leq j \leq m_1, 1 \leq k \leq m_2 \right\}$$

$e_l(m)$ are the canonical basis vectors in $\mathbb{R}^m$.

If $X_i = e_k(m_1)e_l^T(m_2)$

$$\mathrm{tr}(X_i^T M) = M_{kl}$$

M_{kl} is (k, l)th entry of M.

Matrix Completion: Equivalent formulation

- A subset of indexes

$$E \in \{1, \dots m_1\} \times \{1, \dots m_2\}, \quad \mathrm{Card}\,(E) = n.$$

- We observe the noisy entries of M:

$$y_{kl} = M_{kl} + \xi_{kl}, \quad (k, l) \in E$$

M_{kl} is (k, l)th entry of M

- Difference: an entry appears **at most once**.

Non-noisy case

- Candès/Recht (2008), Candès/Tao (2009)
- Gross (2009), Recht (2009)
- Different approach Keshavan et al (2009) (OPTSPACE)

Recht (2009)

Exact reconstruction with high probability if

$$n > C \log^2(m)\,(m_1 + m_2)\mathrm{rank}(M)$$

$m = \min\{m_1, m_2\}$.

Exact reconstruction: conditions

- **Sampling uniformly at random**
- **"Incoherence" condition**:

$A \in \mathbb{R}^{m_1 \times m_2} = U\, DV^T$, $\operatorname{rank}(A) = r$, $\nu = O(1)$ and $d = \max(m_1, m_2)$

$$\left\|U^T e_i\right\|^2 \leq \frac{\nu r}{d}, \quad \left\|V^T e_i\right\|^2 \leq \frac{\nu r}{d}$$

and

$$\left|U\, V^T\right|_{ij}^2 \leq \frac{\nu r}{d^2}$$

(intuition: column and row spaces cannot be aligned with basis vectors)

Constrained Matrix LASSO

$$\widehat{M} = \underset{\|A\|_\infty \leq \gamma}{\text{argmin}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \left(Y_i - \langle X_i, A \rangle\right)^2 + \lambda \|A\|_* \right\}$$

- $\lambda > 0$ is a regularization parameter.
- γ is an upper bound on $\|M\|_\infty = \max_{i,j} \mid M_{ij} \mid$.

  Often known in applications! (e.g. NETFLIX maximal rating)

- $\gamma \rightarrow$ the ball over which we are minimizing.
- Optimal choice

$$\lambda = C^* \sigma \sqrt{\frac{\log(m_1 + m_2)}{\min(m_1, m_2) n}}.$$

Matrix LASSO: bounds on estimation error

Theorem (K., 2012)

With a good choice of λ, with high probability

$$\frac{\|\widehat{M} - M\|_2^2}{m_1 m_2} \leq C \max(\sigma^2, \gamma^2) \log(m_1 + m_2) \frac{\max(m_1, m_2)\,\mathrm{rank}(M)}{n}.$$

$$n > C \log(m_1 + m_2)\, \max(m_1, m_2)\mathrm{rank}(M)$$

- Low rank matrix M: $\mathbf{n \ll m_1 m_2}$.
- n close to the number of degrees of freedom of a rank r matrix

$$(m_1 + m_2)r - r^2$$

.

- Minimax optimality: Koltchinskii et al (2011)

Assumptions on the sampling scheme

We consider a **general** (**unknown**) weighted sampling scheme:

- π_{jk} = probability to observe the (j,k)-th entry;
- $C_k = \sum_{j=1}^{m_1} \pi_{jk}$ the probability to observe an element from the k-th column;
- $R_j = \sum_{k=1}^{m_2} \pi_{jk}$ the probability to observe an element from the j-th row.

Assumption 1

There exists a positive constant $\mu \leq 1$ such that

$$\pi_{jk} \geq \frac{\mu}{m_1 m_2}$$

Assumptions on the sampling scheme

The nuclear-norm penalization fails when some columns or rows are sampled with very high probability (Salakhudinov et al (2010))

Assumption 2

There exists a positive constant $\nu \geq 1$ such that

$$\max_{i,j} (C_i, R_j) \leq \frac{\nu}{\min(m_1, m_2)}.$$

- Uniform sampling: $\nu = \mu = 1.$

Robust Matrix Completion

joint work with K. Lounici and A. Tsybakov

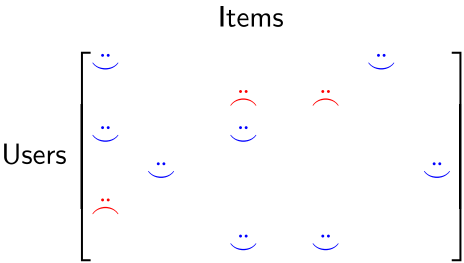
Motivation

Gross errors frequently occur in many applications

- Web data analysis
- Occlusions
- Malicious tampering
- Image processing
- ...

Ranking and Collaborative Filtering

- People rate items
- Some entries have been tampered with



Problem

Make approach robust vis-à-vis corruptions

Model

- Observations (Y_i, X_i) satisfying the *trace regression model*

$$Y_i = \mathrm{tr}(X_i^T M_0) + \xi_i, \quad i = 1, \dots N$$

- We observe noisy entries of $M_0 = L_0 + S_0$
  - $L_0 \in \mathbb{R}^{m_1 \times m_2}$ is low rank
  - $S_0 \in \mathbb{R}^{m_1 \times m_2}$ gross/malicious corruptions
- We do not know which entries are corrupt!

Model

- Observations (Y_i, X_i) satisfying the *trace regression model*

$$Y_i = \mathrm{tr}(X_i^T M_0) + \xi_i, \quad i = 1, \dots N$$

- We observe noisy entries of $M_0 = L_0 + S_0$
  - $L_0 \in \mathbb{R}^{m_1 \times m_2}$ is low rank
  - $S_0 \in \mathbb{R}^{m_1 \times m_2}$ gross/malicious corruptions
- We do not know which entries are corrupt!

Goal

Recover (L_0, S_0) from (X_i, Y_i) when $m_1 m_2 \gg N$

Matrix Decomposition Problem

We observe ALL entries of $M_0 = L_0 + S_0$

- Chandrasekaran et al (2011) : S_0 is element-wise sparse;
- Hsu et al (2011): milder conditions for recovery;
- Xu et al (2012) : S_0 is column-wise sparse;
- Agarwal et al (2012) : "spikiness condition":
  - element-wise sparsity: $\|L\|_\infty \leq \frac{\alpha}{\sqrt{m_1 m_2}}$
  - column-wise sparsity: $\|L\|_{2,1} \leq \frac{\alpha}{\sqrt{m_2}}$

Robust Matrix Completion: non-noisy case

We observe a small fraction of entries of $M_0 = L_0 + S_0$

- Candès et al (2009):
  - S_0 is element-wise sparse;
  - random positions of corruptions;
  - $N = 0.1 m_1 m_2$.
- Chen et al (2011):
  - S_0 is column-wise sparse;
  - random positions of corruptions;
  - Sparse/low-rank incoherent condition.
- Chen et al (2013) and Li (2013):
  - S_0 is element-wise sparse;
  - L_0 is incoherent;
  - assumptions on the number of observations.

.

Convex relaxation for robust matrix completion

(X_i, Y_i), $i = 1 \dots N$ observations

$$Y_i = \mathrm{tr}(X_i^T M_0) + \xi_i, \quad \text{and} \quad M_0 = L_0 + S_0$$

$$(\widehat{L}, \widehat{S}) \in \underset{\substack{\|L\|_\infty \leq \mathbf{a} \\ \|S\|_\infty \leq \mathbf{a}}}{\mathrm{argmin}} \left\{ \frac{1}{N} \sum_{i=1}^{N} (Y_i - \langle X_i, L + S \rangle)^2 + \lambda_1 \|L\|_* + \lambda_2 \mathcal{R}(S) \right\}$$

- λ_1, λ_2 are regularization parameters.
- $\mathbf{a}$ is an upper bound on $\|L_0\|_\infty$ and $\|S_0\|_\infty$.
- $\mathcal{R}(S)$ norm-based penalty $\rightarrow$ corruptions

Sparsity structure

- Column-wise sparsity: small number $s < m_2$ of non-zero columns

$$\mathcal{R}(S) = \|S\|_{2,1} = \sum_{k=1}^{m_2} \|S^k\|_2 \qquad \begin{bmatrix} 0 & \times & 0 & \dots & 0 & \times & 0 \\ 0 & \times & 0 & \dots & 0 & \times & 0 \\ 0 & \times & 0 & \dots & 0 & \times & 0 \\ 0 & \times & 0 & \dots & 0 & \times & 0 \\ 0 & \times & 0 & \dots & 0 & \times & 0 \\ 0 & \times & 0 & \dots & 0 & \times & 0 \end{bmatrix}$$

- Element-wise sparsity: small number $s << m_1 m_2$ of non-zero entries:

$$\mathcal{R}(S) = \|S\|_1 = \sum_{ij} |S_{ij}| \qquad \begin{bmatrix} 0 & \times & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & \times & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & \times & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & \times \end{bmatrix}$$

Assumptions $\mathcal{R}$

- $\mathcal{R}$ is *decomposable* with respect to a properly chosen set of indices I.

$$\mathcal{R}(A) = \mathcal{R}(A_I) + \mathcal{R}(A_{\bar{I}})$$

  - $(2,1)-$norm is decomposable with respect to any set I such that

$$I = \{1, \dots, m_1\} \times C$$

    where $C \subset \{1, \dots, m_2\}$.

  - l_1-norm is decomposable with respect to any subspace of indices I.

- $\mathcal{R}$ is *absolute*:

$$\mathcal{R}(A) = \mathcal{R}(|A|).$$

  - l_p and $\|\cdot\|_{2,1}$ norms are absolute.

Sampling scheme

Set of observations $= \Omega \cup \tilde{\Omega}$

- Ω and $\tilde{\Omega}$ unknown
- $\Omega \cap \tilde{\Omega} = \emptyset$ and $|\Omega| + |\tilde{\Omega}| = N$
- Ω "non-corrupted" observations $\rightarrow$ noisy entries of L_0
- $\tilde{\Omega}$ "corrupted" observations $\rightarrow$ noisy entries of S_0
- $|\Omega|$ and $|\tilde{\Omega}|$ non-random and unknown
- On Ω usual matrix completion sampling.

No assumptions on $\tilde{\Omega}$!

Bounds on estimation error: column-wise sparsity

$$(\widehat{L}, \widehat{S}) \in \operatorname*{argmin}_{\substack{\|L\|_\infty \leq \mathbf{a} \\ \|S\|_\infty \leq \mathbf{a}}} \left\{ \frac{1}{N} \sum_{i=1}^{N} \left(Y_i - \langle X_i, L + S \rangle\right)^2 + \lambda_1 \|L\|_* + \lambda_2 \|S\|_{2,1} \right\}.$$

Theorem (K., Lounici and Tsybakov 2014)

With high probability

$$\frac{\|L_0 - \widehat{L}\|_2^2}{m_1 m_2} \leq \frac{r\,M}{N} + \frac{|\tilde{\Omega}|}{N} + \frac{\mathbf{a}^2 s}{m_2} \quad \textit{and} \quad \frac{\|\widehat{S}_{\mathcal{I}}\|_2^2}{|\mathcal{I}|} \leq \frac{|\tilde{\Omega}|}{N} + \frac{\mathbf{a}^2 s}{m_2}.$$

- $r = \operatorname{rank} L_0$, $M = \max(m_1, m_2)$,
- $|\tilde{\Omega}|$ number of corrupt observations, s number of corrupt columns and $\mathcal{I}$ the set of the non-corrupt columns.

Bounds on estimation error: element-wise sparsity

$$(\widehat{L},\widehat{S}) \in \underset{\substack{\|L\|_\infty \le \mathbf{a} \\ \|S\|_\infty \le \mathbf{a}}}{\operatorname{argmin}} \left\{ \frac{1}{N}\sum_{i=1}^{N} \left(Y_i - \langle X_i, L+S\rangle\right)^2 + \lambda_1 \|L\|_* + \lambda_2 \|S\|_1 \right\}.$$

Theorem (K.,Lounici and Tsybakov 2014)

With high probability

$$\frac{\|L_0-\widehat{L}\|_2^2}{m_1 m_2} \le \frac{r\,M}{N} + \frac{|\tilde{\Omega}|}{N} + \frac{\mathbf{a}^2 s}{m_1 m_2} \quad \textit{and} \quad \frac{\|\widehat{S}_{\mathcal{I}}\|_2^2}{|\mathcal{I}|} \le \frac{|\tilde{\Omega}|}{N} + \frac{\mathbf{a}^2 s}{m_1 m_2}.$$

- $r = \operatorname{rank} L_0$, $M = \max(m_1, m_2)$,
- $|\tilde{\Omega}|$ number of corrupt observations, s number of corrupt entries and $\mathcal{I}$ the set of the non-corrupt entries.

Bounds on estimation error

- Minimax optimality (up to a logarithmic factor).
- All entries are observed ($N = m_1 m_2$) $\rightarrow$ matrix decomposition.
- Small number of corruptions $\rightarrow$ recovery of L_0 from a nearly minimal number of observations.
- Does not require strong assumption on the unknown matrix.
- Adaptive $\rightarrow$ does not require knowledge of $\operatorname{rank} L_0$ and sparsity level of S_0.

Thank you!