ELEG 867 - Compressive Sensing and Sparse Signal Representations

Introduction to Matrix Completion and Robust PCA

Gonzalo Garateguy
Depart. of Electrical and Computer Engineering
University of Delaware

Fall 2011
Matrix Completion Problems - Motivation

Recomender Systems

<table>
<thead>
<tr>
<th>Items</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>User 1</td>
<td>x</td>
<td>x</td>
<td>?</td>
<td>?</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>User 2</td>
<td>?</td>
<td>?</td>
<td>x</td>
<td>x</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>

Collaborative filtering (Amazon, last.fm)
Content based (Pandora, www.nanocrowd.com)
Netflix prize competition boosted interest in the area

http://sahd.pratt.duke.edu/Videos/keynote.html
Matrix Completion Problems - Motivation

Sensor location estimation in Wireless Sensor Networks

Distance matrix

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>$d_{1,2}$</td>
<td>$d_{1,3}$</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>2</td>
<td>$d_{2,1}$</td>
<td>0</td>
<td>?</td>
<td>$d_{2,4}$</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>3</td>
<td>?</td>
<td>$d_{3,1}$</td>
<td>0</td>
<td>$d_{3,4}$</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>4</td>
<td>?</td>
<td>$d_{4,2}$</td>
<td>$d_{4,3}$</td>
<td>0</td>
<td>$d_{4,5}$</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>5</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>$d_{5,4}$</td>
<td>0</td>
<td>$d_{5,6}$</td>
<td>$d_{5,7}$</td>
</tr>
<tr>
<td>6</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>$d_{6,5}$</td>
<td>0</td>
<td>$d_{6,7}$</td>
</tr>
<tr>
<td>7</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>$d_{7,5}$</td>
<td>$d_{7,6}$</td>
<td>0</td>
</tr>
</tbody>
</table>

- The problem is to find the positions of the sensors in $\mathbb{R}^2$ given the partial information about relative distances.
- A distance matrix like this has rank 2 in $\mathbb{R}^2$.
- For certain types of graphs the problem can be solved if we know the whole distance matrix.
Matrix Completion Problems - Motivation

Image reconstruction from incomplete data

Reconstructed image

Incomplete image 50% of the pixels
Robust PCA - Motivation

Foreground identification for surveillance applications

E.J. Candes, X. Li, Y. Ma, and Wright, J. “Robust principal component analysis?” http://arxiv.org/abs/0912.3599
Robust PCA - Motivation

Image alignment and texture recognition

(a) Input \(r = 35\)  (b) Input \(r = 15\)  (c) Input \(r = 53\)  (d) Input \(r = 13\)

(e) Output \(r = 14\)  (f) Output \(r = 8\)  (g) Output \(r = 19\)  (h) Output \(r = 6\)

Z. Zhang, X. Liang, A. Ganesh, and Y. Ma, “TILT: transform invariant low-rank textures” Computer Vision–ACCV 2010
Robust PCA - Motivation

Camera calibration with radial distortion

Motivation

Many other applications

- System Identification in control theory
- Covariance matrix estimation
- Machine Learning
- Computer Vision

Videos to watch

Matrix Completion via Convex Optimization: Theory and Algorithms by Emmanuel Candes
http://videolectures.net/mlss09us_candes_mccota/

Low Dimensional Structures in Images or Data by Yi Ma, Workshop in Signal Processing with Adaptive Sparse Structured Representations (June 2011)
http://ecos.maths.ed.ac.uk/SPARS11/YiMa.wmv
Problem Formulation

Matrix completion

\[
\text{minimize} \quad \text{rank}(A) \\
\text{subject to} \quad A_{ij} = D_{ij} \quad \forall (i,j) \in \Omega
\]

Robust PCA

\[
\text{minimize} \quad \text{rank}(A) + \lambda ||E||_0 \\
\text{subject to} \quad A_{ij} + E_{ij} = D_{ij} \quad \forall (i,j) \in \Omega
\]

- Very hard to solve in general without any assumptions, some times NP hard.
- Even if we can solve them, are the solutions always what we expect?
- Under which conditions we can have exact recovery of the real matrices?
Outline

- Convex Optimization concepts

- Matrix Completion
  - Exact Recovery from incomplete data by convex relaxation
  - ALM method for Nuclear Norm Minimization

- Robust PCA
  - Exact Recovery from incomplete data and corrupted data by convex relaxation
  - ALM method for Low rank and Sparse separation
Convex sets and Convex functions

Convex set

A set $C$ is convex if the line segment between any two points in $C$ lies in $C$. For any $x_1, x_2 \in C$ and any $\theta$ with $0 \leq \theta \leq 1$ we have

$$\theta x_1 + (1 - \theta)x_2 \in C.$$
Convex sets and Convex functions

Convex combination
A convex combination of $k$ points $x_1, \ldots, x_k$ is defined as

$$\theta_1 x_1 + \ldots + \theta_k x_k$$

where $\theta_i \geq 0$ and $\theta_1 + \ldots + \theta_k = 1$

Convex hull
The convex hull of $C$ is the set of all convex combinations of points in $C$

$$\text{conv } C = \{ \theta_1 x_1 + \ldots + \theta_k x_k | x_i \in C, \theta_i \geq 0, i = 1, \ldots, k, \theta_1 + \ldots + \theta_k = 1 \}$$
Convex sets and Convex functions

Operations that preserve convexity

Intersection

If $S_1$ and $S_2$ are convex, then $S_1 \cap S_2$ is convex.
In general if $S_\alpha$ is convex for every $\alpha \in \mathcal{A}$, then $\bigcap_{\alpha \in \mathcal{A}} S_\alpha$ is convex.

Subspaces, affine sets and convex cones are therefore closed under arbitrary intersections.

Affine functions

Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be affine, $f(x) = Ax + b$, where $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. If $S \subseteq \mathbb{R}^n$ is convex, then the image of $S$ under $f$

$$f(S) = \{f(x) \mid x \in S\}$$

is convex
Convex sets and Convex functions

Convex functions

A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if $\text{dom} f$ is a convex set and if for all $x, y \in \text{dom} f$, and $\theta$ with $0 \leq \theta \leq 1$, we have

$$f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta f(y))$$

we say that $f$ is strictly convex if the strict inequality holds whenever $x \neq y$ and $0 < \theta < 1$.
Operations that preserve convexity

Composition with an affine mapping

Suppose \( f : \mathbb{R}^n \rightarrow \mathbb{R}, A \in \mathbb{R}^{n \times m} \) and \( b \in \mathbb{R}^n \). Define \( g : \mathbb{R}^m \rightarrow \mathbb{R} \) by

\[
g(x) = f(Ax + b)
\]

with \( \text{dom} g = \{ x | Ax + b \in \text{dom} f \} \). Then if \( f \) is convex, so is \( g \).

Pointwise maximum

If \( f_1 \) and \( f_2 \) are convex functions then their pointwise maximum \( f \) defined by

\[
f(x) = \max\{f_1(x), f_2(x)\}
\]

with \( \text{dom} f = \text{dom} f_1 \cap \text{dom} f_2 \) is also convex. This also extend to the case where \( f_1, ..., f_m \) are convex, then

\[
f(x) = \max\{f_1(x), ..., f_m(x)\}, \quad \text{is also convex}
\]
Pointwise maximum of convex functions

\[ f(x) = \max \{ f_1(x), f_2(x) \} \]
Convex sets and Convex functions

Convex differentiable functions

If $f$ is differentiable (i.e. its gradient $\nabla f$ exist at each point in $\text{dom} f$). Then $f$ is convex if and only if $\text{dom} f$ is convex and

$$f(y) \geq f(x) + \nabla f(x)^T (y - x)$$

holds for all $x, y \in \text{dom} f$. 
Second order conditions

If $f$ is twice differentiable, i.e. its Hessian $\nabla^2 f$ exist at each point in $\text{dom} f$. Then $f$ is convex if and only if $\text{dom} f$ is convex and its Hessian is positive semidefinite for all $x \in \text{dom} f$

$$\nabla^2 f(x) \succeq 0$$
Convex non-differentiable functions

The concept of gradient can be extended to non-differentiable functions introducing the subgradient

Subgradient of a function

A vector $g \in \mathbb{R}^n$ is a subgradient of $f : \mathbb{R}^n \rightarrow \mathbb{R}$ at $x \in \text{dom} f$ if for all $z \in \text{dom} f$

$$f(z) \geq f(x) + g^T(z - x)$$
Subgradients

Observations

- If $f$ is convex and differentiable, then its gradient at $x$, $\nabla f(x)$ is its only subgradient

Subdifferentiable functions

A function $f$ is called subdifferentiable at $x$ if there exist at least one subgradient at $x$

Subdifferential at a point

The set of subgradients of $f$ at the point $x$ is called the subdifferential of $f$ at $x$, and is denoted $\partial f(x)$

Subdifferentiability of a function

A function $f$ is called subdifferentiable if it is subdifferentiable at all $x \in \text{dom} f$
Existence of the subgradient of a convex function

If $f$ is convex and $x \in \text{int dom} f$, then $\partial f(x)$ is nonempty and bounded.

The subdifferential $\partial f(x)$ is always a closed convex set, even if $f$ is not convex. This follows from the fact that it is the intersection of an infinite set of halfspaces

$$\partial f(x) = \bigcap_{z \in \text{dom} f} \{g | f(z) \geq f(x) + g^T(z - x)\}.$$
Basic properties

Nonnegative scaling
For $\alpha \geq 0$, $\partial(\alpha f)(x) = \alpha \partial f(x)$

Subgradient of the sum
Given $f = f_1 + ... + f_m$, where $f_1, ..., f_m$ are convex functions, the subgradient of $f$ at $x$ is given by $\partial f(x) = \partial f_1(x) + ... + \partial f_m(x)$

Affine transformations of domain
Suppose $f$ is convex, and let $h(x) = f(Ax + b)$. Then $\partial h(x) = A^T \partial f(Ax + b)$.

Pointwise maximum
Suppose $f$ is the pointwise maximum of convex functions $f_1, ..., f_m$, $f(x) = \max_{i=1,...,m} f_i(x)$, then $\partial f(x) = \text{Co} \cup \{\partial f_i(x) | f_i(x) = f(x)\}$
Subgradient of the pointwise maximum of two convex functions

\[ f(x) = \max\{f_1(x), f_2(x)\} \]

\[ f_1(x) + f'_1(x_0)(x - x_0) \]

\[ f_2(x) + f'_2(x_0)(x - x_0) \]
Subgradient of the pointwise maximum of two convex functions

\[ f(x) = \max\{f_1(x), f_2(x)\} \]

\[ f(x) + g(x - x_0) \]

\[ f_1(x_0) + f'_1(x_0)(x - x_0) \]

\[ f_2(x_0) + f'_2(x_0)(x - x_0) \]
Subgradient of the pointwise maximum of two convex functions

\[ g \in [f_2'(x_0), f_1'(x_0)] \rightarrow g = \theta f_2'(x_0) + (1 - \theta) f_2'(x_0) \text{ with } 0 \leq \theta \leq 1 \]
Examples

Consider the function \( f(x) = |x| \). At \( x_0 = 0 \), the subdifferential is defined by the inequality

\[
f(z) \geq f(x_0) + g(z - x_0), \quad \forall z \in \text{dom} \ f \\
|z| \geq gz, \quad \forall z \in R \\
\partial f(0) = \{ g \mid g \in [-1, 1] \}
\]

then for all \( x \)

\[
\partial f(x) = \begin{cases} 
-1 & \text{for } x < 0 \\
1 & \text{for } x > 0 \\
\{ g \mid g \in [-1, 1] \} & \text{for } x = 0 
\end{cases}
\]
Example: $\ell_1$ norm

Consider $f(x) = \|x\|_1 = |x_1| + \cdots + |x_n|$, and note that $f$ can be expressed as the maximum of $2^n$ linear functions

$$\|x\|_1 = \max\{ f_1(x), \ldots, f_{2^n}(x) \}$$

$$\|x\|_1 = \max\{ s_1^T x, \ldots, s_{2^n}^T x \mid s_i \in \{-1, 1\}^n \}$$

The active functions $f_i(x)$ at $x$ are the ones for which $s_i^T x = \|x\|_1$. Then denoting

$$s_i = [s_{i,1}, \ldots, s_{i,n}]^T, \; s_{i,j} \in \{-1, 1\}$$

the set of indices of the active functions at $x$ is

$$\mathcal{A}_x = \left\{ i \mid \begin{array}{ll} s_{i,j} = -1 & \text{for } x_j < 0 \\ s_{i,j} = 1 & \text{for } x_j > 0 \\ s_{i,j} = -1 \text{ or } 1 & \text{for } x_j = 0 \end{array} \right\}$$
subgradient of the $\ell_1$ norm

The subgradient of $\|x\|_1$ at a generic point $x$ is defined by

\[
\partial \|x\|_1 = \text{co} \cup \{ \partial f_i(x) \mid i \in A_x \}
\]
\[
\partial \|x\|_1 = \text{co} \{ \nabla f_i(x) \mid i \in A_x \}
\]
\[
\partial \|x\|_1 = \text{co} \{ s_i \mid i \in A_x \}
\]
\[
\partial \|x\|_1 = \{ g \mid g = \sum_{i \in A_x} \theta_i s_i \, , \, \theta_i \geq 0 \, , \, \sum_i \theta_i = 1 \}
\]

or equivalently

\[
\partial \|x\|_1 = \begin{cases} 
    g & g_j = -1 \quad \text{for } x_j < 0 \\
    g & g_j = 1 \quad \text{for } x_j > 0 \\
    g_j = \zeta \in [-1, 1] \quad \text{for } x_j = 0
\end{cases}
\]
$\ell_1$ norm on $R^2$

in $R^2$ the set of subgradients are

$$s_1 = [-1, 1]^T$$
$$s_2 = [-1, -1]^T$$
$$s_3 = [1, 1]^T$$
$$s_4 = [1, -1]^T$$
An optimization problem is convex if its objective is a convex function, the inequality constraints $f_j$ are convex and the equality constraints $h_j$ are affine.

$$\begin{align*}
\text{minimize} & \quad f_0(x) \quad \text{(Convex function)} \\
\text{s.t.} & \quad f_i(x) \leq 0 \quad \text{(Convex sets)} \\
& \quad h_j(x) = 0 \quad \text{(Affine)}
\end{align*}$$

or equivalently

$$\begin{align*}
\text{minimize} & \quad f_0(x) \quad \text{(Convex function)} \\
\text{s.t.} & \quad x \in C \quad C \text{ is a convex set} \\
& \quad h_j(x) = 0 \quad \text{(Affine)}
\end{align*}$$
Theorem
If \( x^* \) is a local minimizer of a convex optimization problem, it is a global minimizer.

Optimality conditions
A point \( x^* \) is a minimizer of a convex function \( f \) if and only if \( f \) is subdifferentiable at \( x^* \) and
\[
0 \in \partial f(x^*)
\]
Convex optimization problems

Given the convex problem

\[
\begin{align*}
\text{minimize} & \quad f_0(x) \\
\text{s.t.} & \quad f_i(x) \leq 0, \quad i = \{1, \ldots, k\} \\
& \quad h_j(x) = 0, \quad j = \{1, \ldots, l\}
\end{align*}
\]

its Lagrangian function is defined as

\[
\mathcal{L}(x, \lambda, \nu) = f_0(x) + \sum_{j=1}^{l} \lambda_j h_j(x) + \sum_{i=1}^{k} \nu_i f_i(x)
\]

where \( \nu_i \geq 0, \lambda_i \in R \)
Augmented Lagrangian Method

Considering the problem

$$\begin{align*}
\text{minimize} & \quad f(x) \\
\text{s.t.} & \quad x \in C \\
& \quad h(x) = 0
\end{align*}$$

The augmented lagrangian is defined as

$$\mathcal{L}(x, \lambda, c) = f(x) + \lambda^T h(x) + \frac{\mu}{2} \| h(x) \|^2_2$$

where $\mu$ is a penalty parameter and $\lambda$ is the multiplier vector.
The augmented lagrangian method consist of solving a sequence of problems of the form

$$\min_x \mathcal{L}(x, \lambda_k, \mu_k) = f(x) + \lambda_k^T h(x) + \frac{\mu_k}{2} \|h(x)\|_2^2$$

s.t. $x \in C$

where $\{\lambda_k\}$ is a bounded sequence in $\mathbb{R}^l$ and $\{\mu_k\}$ is a penalty parameter sequence satisfying

$$0 < \mu_k < \mu_{k+1} \quad \forall k, \mu_k \to \infty$$
Augmented Lagrangian Method

The exact solution to problem (3) can be found using the following iterative algorithm

set $\rho > 1$

while not converged do

solve $x_{k+1} = \text{argmin}_{x \in C} \mathcal{L}(x, \lambda_k, \mu_k)$

$\lambda_{k+1} = \lambda_k + \mu_k h(x_{k+1})$

$\mu_k = \rho \mu_k$

end while

Output $x_k$
Matrix completion

Optimization problem

\[
\begin{align*}
\text{minimize} & \quad \text{rank}(A) \\
\text{subject to} & \quad A_{ij} = D_{ij} \quad \forall (i,j) \in \Omega
\end{align*}
\] (4)

- We look for the simplest explanation for the observed data
- Given enough number of samples, the likelihood of the solution to be unique should be high
Matrix completion

\[ \begin{align*}
\text{minimize} & \quad \text{rank}(A) \\
\text{subject to} & \quad A_{ij} = D_{ij} \quad \forall (i, j) \in \Omega
\end{align*} \]

- The minimization of the rank(·) function is a combinatorial problem, with exponential complexity in the size of the matrix!

- Need for a convex relaxation

\[ \begin{align*}
\text{rank}(A) &= \| \text{diag}(\Sigma) \|_0 \\
A &= U\Sigma V^T \\
\downarrow \\
\| A \|_* &= \| \text{diag}(\Sigma) \|_1
\end{align*} \]

Convex relaxation

\[ \begin{align*}
\text{minimize} & \quad \| A \|_* \\
\text{subject to} & \quad A_{ij} = D_{ij} \quad \forall (i, j) \in \Omega
\end{align*} \]
Matrix Completion

Nuclear Norm

The nuclear norm of a matrix $A \in \mathbb{R}^{m \times n}$ is defined as $\|A\|_* = \sum_{i=1}^{r} \sigma_i(A)$, where $\{\sigma_i(A)\}_{i=1}^{r}$ are the elements of the diagonal matrix $\Sigma$ from the SVD decomposition of $A = U\Sigma V^T$.

Observations

- $r = \text{rank}(A)$ can be $r < m, n$. If this is the case we say that the matrix is low rank.
- The singular values $\sigma_i(A) = \sqrt{\lambda_i(A^T A)}$ are obtained as the square root of the eigenvalues of $A^T A$ and are always $\sigma_i \geq 0$.
- The left singular vectors $U$ are the eigenvectors of $AA^T$.
- The right singular vectors $V$ are the eigenvectors of $A^T A$. 
Matrix Completion

Spectral Norm

The spectral norm of a matrix $A \in \mathbb{R}^{m \times n}$ is defined as $\|A\|_2 = \sigma_{\text{max}}(A)$, where

$$\sigma_{\text{max}} = \max \{ \sigma_i(A) \}^r_{i=1}$$

Dual Norm

Given an arbitrary norm $\| \cdot \|_\diamond$ in $\mathbb{R}^n$, its dual norm $\| \cdot \|_\dagger$ is defined as

$$\|z\|_\dagger = \sup \{ z^T x \mid \|x\|_\diamond \leq 1 \}$$

Observations

- The nuclear norm is the dual norm of the spectral norm

$$\|A\|_* = \sup \{ \text{tr}(A^T X) \mid \|X\|_2 \leq 1 \}$$
Matrix Completion

Convex relaxation of the rank

Convex envelope of a function

Let $f : C \rightarrow R$ where $C \subseteq R^n$. The convex envelope of $f$ (on $C$) is defined as the largest convex function $g$ such that $g(x) \leq f(x)$ for all $x \in C$

Theorem

The convex envelope of the function $\phi(X) = \text{rank}(X)$ on $C = \{X \in R^{m \times n} \|X\|_2 \leq 1\}$, is $\phi_{env}(X) = \|X\|_\ast$.

Observations

- The convex envelope of rank($X$) on a the set $\{X \|X\|_2 \leq M\}$ is given by $\frac{1}{M} \|X\|_\ast$
- By solving the heuristic problem we obtain a lower bound on the optimal value of the original problem (provided we can identify a bound $M$ on the feasible set).

Convex relaxation

\[
\text{minimize} \quad \|A\|_* \\
\text{subject to} \quad A_{ij} = D_{ij} \quad \forall (i, j) \in \Omega
\]  

- The original problem is now a problem with a non-smooth but convex function as the objective.
- The remaining problem is the number of measurements and in which positions have to be taken in order to guarantee that the solution is equal to the matrix $D$?
Matrix completion

Which types of matrices can be completed exactly?
Consider the matrix

\[ M = e_1.e_n^T = \begin{pmatrix}
0 & 0 & \cdots & 0 & 1 \\
0 & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \cdots & 0 & 0 \\
0 & 0 & \cdots & 0 & 0
\end{pmatrix} \]

- Can it be recovered from 90% of its samples?
- Is the sampling set important?
- Which sampling sets work and which ones doesn’t?
Matrix completion

Sampling set \( \Omega \)

The sampling set \( \Omega \) is defined as \( \Omega = \{(i,j) \mid D_{ij} \text{ is observed} \} \)

Consider

\[
D = xy^T \quad x \in \mathbb{R}^m, \ y \in \mathbb{R}^n
\]

\[
D_{ij} = x_i y_j
\]

- If the sampling set avoids row \( i \), then \( x_i \) can not be recovered by any method whatsoever

Observation

- No columns or rows from \( D \) can be avoided in the sampling set
- There is a need for a characterization of the sampling operator with respect to the set of matrices that we want to recover
Matrix completion

- To recover a low rank matrix, this matrix cannot be in the null space of the sampling operator.
- If the singular vectors of $D = USV^T$ are highly concentrated, then $D$ is more likely to be in the null space of a given sampling operator.
Matrix completion

Intuition
- the singular vectors need to be sufficiently spread, i.e. uncorrelated with the standard basis in order to minimize the number of observations needed to recover a low rank matrix.

Coherence of a subspace
Let $U$ be a subspace of $R^n$ of dimension $r$ and $P_U$ be the orthogonal projection onto $U$. Then the coherence of $U$ is defined to be

$$\mu(U) = \frac{n}{r} \max_{1 \leq i \leq n} \|P_U e_i\|^2$$

Observations
- The minimum value that $\mu(U)$ can achieve is 1 for example if $U$ is spanned by vectors whose entries all have magnitude $1/\sqrt{n}$
- The largest possible value for $\mu(U)$ is $n/r$ corresponding to a subspace that contains a standard basis element.
Matrix completion

\( \mu_0 \) coherence

A matrix \( D = \sum_{1 \leq k \leq r} \sigma_k u_k v_k^T \) is \( \mu_0 \) coherent if for some positive \( \mu_0 \)
\[ \max(\mu(U), \mu(V)) \leq \mu_0 \]

\( \mu_1 \) coherence

A matrix \( D = \sum_{1 \leq k \leq r} \sigma_k u_k v_k^T \) has \( \mu_1 \) coherence if
\[ \| UV^T \|_\infty \leq \mu_1 \sqrt{r/mn} \]
for some \( \mu_1 > 0 \)

Observation

- If \( D \) is \( \mu_0 \) coherent then it is \( \mu_1 \) coherent for \( \mu_1 = \mu_0 \sqrt{r} \)
Matrix completion

Theorem

Let $D \in \mathbb{R}^{m \times n}$ of rank $r$ be $(\mu_0, \mu_1)$-coherent and let $N = \max(m, n)$. If we observe $M$ entries of $D$ with locations sampled uniformly at random. Then there exist constants $C$ and $c$ such that if

$$M \geq C \max(\mu_1^2, \mu_0^{1/2} \mu_1, \mu_0 N^{1/4})Nr(\beta \log N)$$

for some $\beta > 2$, then the minimizer of (6) is unique and equal to $D$ with probability at least $1 - cn^{-\beta}$. If in addition $r \leq \mu_0^{-1} N^{1/5}$ then the number of observations can be improved to

$$M \geq C\mu_0 N^{6/5}r(\beta \log N)$$

Candès, E.J. and Recht, B. “Exact matrix completion via convex optimization”, Foundations of Computational Mathematics 2009
Matrix completion

Recovery performance

Figure: The $x$ axis corresponds to $\text{rank}(A)/\min\{m,n\}$ and the $y$ axis to $\rho_s = 1 - M/mn$ (probability that an entry is omitted from the observations)

Emmanuel J. Candes, Xiaodong Li, Yi Ma, John Wright “Robust Principal Component Analysis?”

http://arxiv.org/abs/0912.3599
Matrix completion

Other bounds on number of measurements and sampling operators

- Emmanuel J. Candes, Xiaodong Li, Yi Ma, John Wright “Rmdbust Principal Component Analysis?”
  http://arxiv.org/abs/0912.3599
- Venkat Chandrasekaran, Sujay Sanghavi, Pablo A. Parrilo, Alan S. Willsky “Rank-Sparsity Incoherence for Matrix Decomposition”
  http://arxiv.org/abs/0906.2220
- Zihan Zhou, Xiaodong Li, John Wright, Emmanuel Candes, Yi Ma “Stable Principal Component Pursuit”
- Raghunandan H. Keshavan, Andrea Montanari, Sewoong Oh “Matrix Completion from a Few Entries”
  http://arxiv.org/abs/0901.3150
- Sahand Negahban, Martin J. Wainwright “Restricted strong convexity and weighted matrix completion: Optimal bounds with noise”
  http://arxiv.org/abs/1009.2118v2
- Yonina C. Eldar, Deanna Needell, Yaniv Plan “Unicity conditions for low-rank matrix recovery”
  http://arxiv.org/abs/1103.5479