Low-Rank Matrix Completion
Geometric Approach and Performance Guarantees

Wei Dai, Ely Kerman, and Olgica Milenkovic

Imperial College London (IC)
University of Illinois at Urbana-Champaign (UIUC)

ICASSP 2011

Outline

Low-rank matrix completion (LRMC)

Introduction

What is missing for LRMC?

Problem

The natural approach does not work.

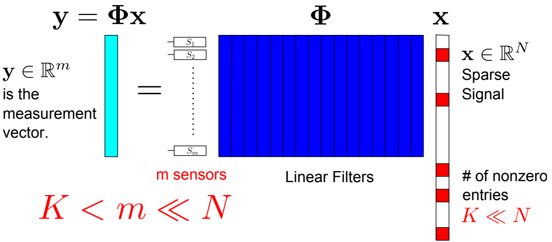
Solution

A new “norm”!

Results

Strong performance guarantees for two scenarios.

Compressive Sensing



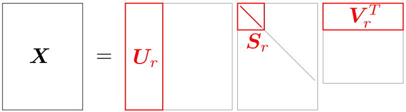
- ℓ_0-search.
- ℓ_1-minimization.

$$\min_{\boldsymbol{x}} \ \|\boldsymbol{x}\|_1 \text{ s.t. } \boldsymbol{y} = \boldsymbol{\Phi x}.$$

- Greedy algorithms.
  - OMP, Subspace pursuit (SP), IHT $\cdots$

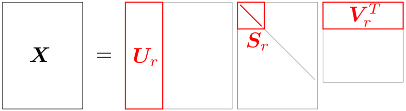
LRMC

- Rank $r \ll \min(m, n)$



LRMC

- Rank $r \ll \min(m, n)$



LRMC

Given

- $\Omega \subset [m] \times [n]$: the index set of the observed entries
- $\boldsymbol{X}_\Omega$: partial observations

$$\boldsymbol{X}_\Omega = \begin{bmatrix} 1 & 5 & ? & 9 & ? & \cdots \\ ? & 6 & ? & 2 & 7 & \cdots \\ 2 & ? & 4 & ? & 7 & \cdots \\ ? & 3 & ? & 8 & 1 & \cdots \\ 4 & 10 & ? & ? & 5 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

Find an $\hat{\boldsymbol{X}}$ so that rank $\left(\hat{\boldsymbol{X}}\right) \leq r$ and $\left(\hat{\boldsymbol{X}}\right)_\Omega = \boldsymbol{X}_\Omega$.

Methods to Solve LRMC

- ℓ_1-minimization

$$\min_{\boldsymbol{X}'} \ \left\|\boldsymbol{X}'\right\|_* \text{ s.t. } (\boldsymbol{X}')_\Omega = \boldsymbol{X}_\Omega.$$

- Greedy algorithms
  - SP⇒ADMiRA.
  - IHT⇒SVP.
- How to do ℓ_0-search?

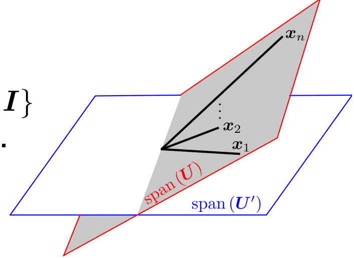
Methods to Solve LRMC

- ℓ_1-minimization

$$\min_{\boldsymbol{X}'} \ \|\boldsymbol{X}'\|_* \text{ s.t. } (\boldsymbol{X}')_\Omega = \boldsymbol{X}_\Omega.$$

- Greedy algorithms
  - SP⇒ADMiRA.
  - IHT⇒SVP.
- How to do ℓ_0-search?

The natural formulation



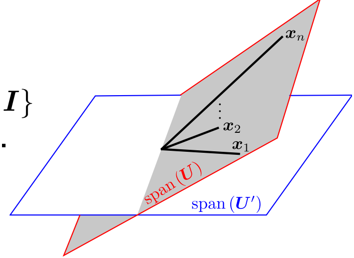
Definition: $\mathcal{U}_{m,r} = \{ \boldsymbol{U} \in \mathbb{R}^{m \times r} : \boldsymbol{U}^* \boldsymbol{U} = \boldsymbol{I} \}$

$\boldsymbol{U} \in \mathcal{U}_{m,r} \Rightarrow$ an r-d subspace.

- This talk: how to find U^*.
  Will not talk about the uniqueness.
- Why ℓ_0? Optimization on a smooth manifold.

The natural formulation

Definition: $\mathcal{U}_{m,r} = \{\boldsymbol{U} \in \mathbb{R}^{m \times r} : \boldsymbol{U}^* \boldsymbol{U} = \boldsymbol{I}\}$
$\boldsymbol{U} \in \mathcal{U}_{m,r} \Rightarrow$ an r-d subspace.



ℓ_0-Search

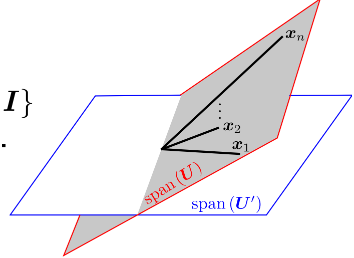
$\forall \boldsymbol{U} \in \mathcal{U}_{m,r}$, define $f_F(\boldsymbol{U}) = \min_{\boldsymbol{W} \in \mathbb{R}^{r \times n}} \|\boldsymbol{X}_\Omega - (\boldsymbol{U}\boldsymbol{W})_\Omega\|_F^2$.

$$\text{LRMC} \equiv \text{find a } \boldsymbol{U}^* \in \mathcal{U}_{m,r} \text{ s.t. } f_F(\boldsymbol{U}^*) = 0.$$

Then $\hat{\boldsymbol{X}} = \boldsymbol{U}^* \boldsymbol{W}_{\boldsymbol{U}^*}$.

- This talk: how to find U^*.
  Will not talk about the uniqueness.
- Why ℓ_0? Optimization on a smooth manifold.

The natural formulation



Definition: $\mathcal{U}_{m,r} = \{ U \in \mathbb{R}^{m \times r} : U^* U = I \}$

$U \in \mathcal{U}_{m,r} \Rightarrow$ an r-d subspace.

ℓ_0-Search

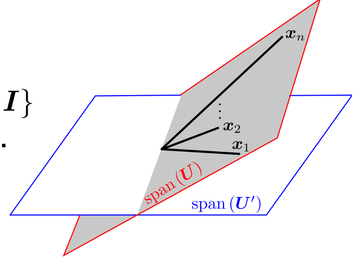
$\forall U \in \mathcal{U}_{m,r}$, define $f_F(U) = \min_{W \in \mathbb{R}^{r \times n}} \| X_\Omega - (UW)_\Omega \|_F^2$.

$$\text{LRMC} \equiv \text{find a } U^* \in \mathcal{U}_{m,r} \text{ s.t. } f_F(U^*) = 0.$$

Then $\hat{X} = U^* W_{U^*}$.

- This talk: how to find U^*.
  Will not talk about the uniqueness.
- Why ℓ_0? Optimization on a smooth manifold.

The natural formulation



Definition: $\mathcal{U}_{m,r} = \{\boldsymbol{U} \in \mathbb{R}^{m \times r} : \boldsymbol{U}^* \boldsymbol{U} = \boldsymbol{I}\}$

$\boldsymbol{U} \in \mathcal{U}_{m,r} \Rightarrow$ an r-d subspace.

ℓ_0-Search

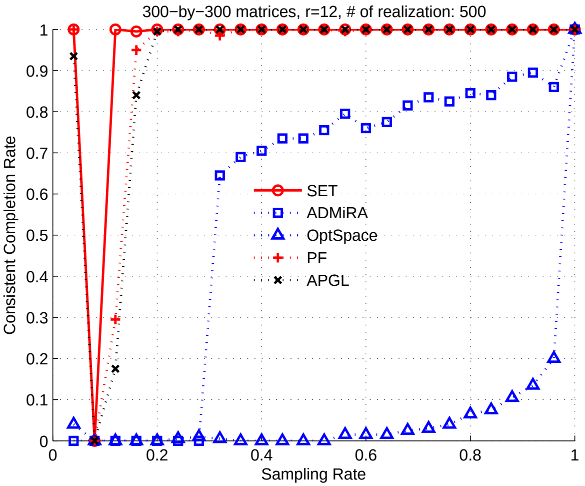
$\forall \boldsymbol{U} \in \mathcal{U}_{m,r}$, define $f_F(\boldsymbol{U}) = \min_{\boldsymbol{W} \in \mathbb{R}^{r \times n}} \| \boldsymbol{X}_\Omega - (\boldsymbol{U}\boldsymbol{W})_\Omega \|_F^2$.

LRMC $\equiv$ find a $\boldsymbol{U}^* \in \mathcal{U}_{m,r}$ s.t. $f_F(\boldsymbol{U}^*) = 0$.

Then $\hat{\boldsymbol{X}} = \boldsymbol{U}^* \boldsymbol{W}_{\boldsymbol{U}^*}$.

- This talk: how to find $\boldsymbol{U}^*$.
  Will not talk about the uniqueness.
- Why ℓ_0? Optimization on a smooth manifold.

A “modified” ℓ_0-search



W. Dai, O. Milenkovic, and E. Kerman, “Subspace evolution and transfer (SET) for low-rank matrix completion,” IEEE Trans. Signal Processing, accepted, 2011.

The natural approach doesn't work

$$f_F(\boldsymbol{U}) = \min_{\boldsymbol{W}\in\mathbb{R}^{r\times n}} \left\| \boldsymbol{X}_\Omega - (\boldsymbol{U}\boldsymbol{W})_\Omega \right\|_F^2 = \sum_{i=1}^n \underbrace{\min_{\boldsymbol{w}\in\mathbb{R}} \left\| (\boldsymbol{X}_\Omega)_{:,i} - (\boldsymbol{U}\boldsymbol{w})_{\Omega_i} \right\|_F^2}_{f_{F,i}(\boldsymbol{U})}$$

$$\boldsymbol{X}_\Omega = \begin{bmatrix} ? \\ 1 \\ 1 \end{bmatrix}, \quad \boldsymbol{U}(t) = \begin{bmatrix} \sqrt{1-2t^2} \\ t \\ t \end{bmatrix}.$$

$$f_F(\boldsymbol{U}(t)) = \min_{w\in\mathbb{R}} \left\| \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \left(\begin{bmatrix} t \\ t \end{bmatrix} w \right) \right\|_F^2 = \begin{cases} 0 & \text{if } t \neq 0, \\ 2 & \text{if } t = 0. \end{cases}$$

The natural approach doesn't work

$$f_F(\boldsymbol{U}) = \min_{\boldsymbol{W}\in\mathbb{R}^{r\times n}} \left\| \boldsymbol{X}_\Omega - (\boldsymbol{U}\boldsymbol{W})_\Omega \right\|_F^2 = \sum_{i=1}^n \underbrace{\min_{\boldsymbol{w}\in\mathbb{R}} \left\| (\boldsymbol{X}_\Omega)_{:,i} - (\boldsymbol{U}\boldsymbol{w})_{\Omega_i} \right\|_F^2}_{f_{F,i}(\boldsymbol{U})}$$

$$\boldsymbol{X}_\Omega = \begin{bmatrix} ? \\ 1 \\ 1 \end{bmatrix}, \quad \boldsymbol{U}(t) = \begin{bmatrix} \sqrt{1-2t^2} \\ t \\ t \end{bmatrix}.$$

$$f_F(\boldsymbol{U}(t)) = \min_{w\in\mathbb{R}} \left\| \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \left(\begin{bmatrix} t \\ t \end{bmatrix} w \right) \right\|_F^2 = \begin{cases} 0 & \text{if } t \neq 0, \\ 2 & \text{if } t = 0. \end{cases}$$

$f_F(\boldsymbol{U})$ is not continuous.

A Geometric Performance Metric

Replace $f_{F,i}(\boldsymbol{U})$ with $f_{G,i}(\boldsymbol{U})$:

A Geometric Performance Metric

Replace $f_{F,i}(\boldsymbol{U})$ with $f_{G,i}(\boldsymbol{U})$:

- $\begin{bmatrix} ? \\ 1 \\ 1 \end{bmatrix} \quad \Rightarrow \quad \mathcal{B} = \text{span}\left(\left\{\begin{bmatrix} x_1 \\ 1 \\ 1 \end{bmatrix} : x_1 \in \mathbb{R}\right\}\right).$

A Geometric Performance Metric

Replace $f_{F,i}(\boldsymbol{U})$ with $f_{G,i}(\boldsymbol{U})$:

- $\begin{bmatrix} ? \\ 1 \\ 1 \end{bmatrix} \quad \Rightarrow \quad \mathcal{B} = \text{span}\left(\left\{\begin{bmatrix} x_1 \\ 1 \\ 1 \end{bmatrix} : x_1 \in \mathbb{R}\right\}\right).$
- Let $\theta_{\min}$ be the minimum principal angle between $\mathcal{B}$ and span$(\boldsymbol{U})$.

$$\theta_{\min} = 0 \quad \Leftrightarrow \quad \text{span}(\boldsymbol{U}) \bigcap \mathcal{B} \neq \{\mathbf{0}\}.$$

A Geometric Performance Metric

Replace $f_{F,i}(\boldsymbol{U})$ with $f_{G,i}(\boldsymbol{U})$:

- $\begin{bmatrix} ? \\ 1 \\ 1 \end{bmatrix} \quad \Rightarrow \quad \mathcal{B} = \text{span}\left(\left\{\begin{bmatrix} x_1 \\ 1 \\ 1 \end{bmatrix} : x_1 \in \mathbb{R}\right\}\right).$

- Let $\theta_{\min}$ be the minimum principal angle between $\mathcal{B}$ and span $(\boldsymbol{U})$.

$$\theta_{\min} = 0 \quad \Leftrightarrow \quad \text{span}(\boldsymbol{U}) \bigcap \mathcal{B} \neq \{\mathbf{0}\}.$$

- $f_{G,i}(\boldsymbol{U}) = \sin^2 \theta_{\min,i}$.
  $f_G(\boldsymbol{U}) = \sum_i f_{G,i}(\boldsymbol{U})$.

A Geometric Performance Metric

Replace $f_{F,i}(\boldsymbol{U})$ with $f_{G,i}(\boldsymbol{U})$:

- $\begin{bmatrix} ? \\ 1 \\ 1 \end{bmatrix} \Rightarrow \mathcal{B} = \text{span}\left(\left\{\begin{bmatrix} x_1 \\ 1 \\ 1 \end{bmatrix} : x_1 \in \mathbb{R}\right\}\right).$
- Let $\theta_{\min}$ be the minimum principal angle between $\mathcal{B}$ and span$(\boldsymbol{U})$.

$$\theta_{\min} = 0 \quad \Leftrightarrow \quad \text{span}(\boldsymbol{U}) \bigcap \mathcal{B} \neq \{\mathbf{0}\}.$$

- $f_{G,i}(\boldsymbol{U}) = \sin^2 \theta_{\min,i}$.
  $f_G(\boldsymbol{U}) = \sum_i f_{G,i}(\boldsymbol{U})$.

ℓ_0-search:

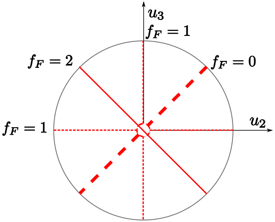
$$\text{Find } \boldsymbol{U} \in \mathcal{U}_{m,r} \text{ s.t. } f_G(\boldsymbol{U}) = 0.$$

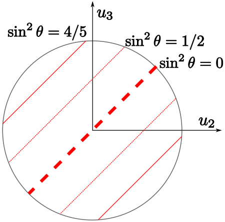
Continuity!

- Continuous

$$\boldsymbol{X}_\Omega = \begin{bmatrix} ? \\ 1 \\ 1 \end{bmatrix}$$

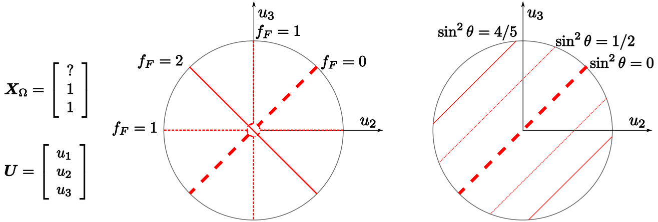
$$\boldsymbol{U} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}$$





Continuity!

- Continuous



- Theorem:

$$\{U \in \mathcal{U}_{m,r} : \ f_G(U) = 0\} = \overline{\{U \in \mathcal{U}_{m,r} : \ f_F(U) = 0\}}.$$

Strong Performance Guarantees

Theorem

For the cases:

- rank-one matrices with arbitrary sampling pattern
- full-sampling matrices with arbitrary ranks

A gradient-descent finds a consistent solution with prob. one.

Dai, Kerman, and Milenkovic, "A Geometric Approach to Low-Rank Matrix Completion," IT, submitted.

Strong Performance Guarantees

Theorem

For the cases:

- rank-one matrices with arbitrary sampling pattern
- full-sampling matrices with arbitrary ranks

A gradient-descent finds a consistent solution with prob. one.

- No local minimum / saddle points.

Dai, Kerman, and Milenkovic, "A Geometric Approach to Low-Rank Matrix Completion," IT, submitted.

Strong Performance Guarantees

Theorem

For the cases:

- rank-one matrices with arbitrary sampling pattern
- full-sampling matrices with arbitrary ranks

A gradient-descent finds a consistent solution with prob. one.

- No local minimum / saddle points.
- Do not require incoherence conditions.
- Holds for arbitrary matrix size.

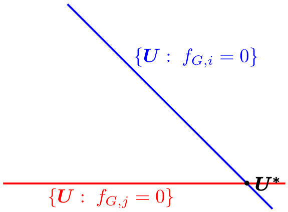
Dai, Kerman, and Milenkovic, “A Geometric Approach to Low-Rank Matrix Completion,” IT, submitted.

Idea of Proof

1. Let $\boldsymbol{U}^*$ be a global minimizer: $f_G = \mathbf{0}$.
   $\boldsymbol{U}^* \in \bigcap_i \{\boldsymbol{U} : f_{G,i}(\boldsymbol{U}) = 0\}$.
2. For a random $\boldsymbol{U}_0$, compute $-\nabla f_{G,i}$.
3. $\forall i$, Proj $(-\nabla f_{G,i}, \boldsymbol{U}^* - \boldsymbol{U}_0) \geq 0$, where "=" iff $f_{G,i}(\boldsymbol{U}_0) = 0$.
4. $\nabla f_G = \sum_i \nabla f_{G,i} \Rightarrow$
   Proj $(-\nabla f_G, \boldsymbol{U}^* - \boldsymbol{U}_0) \geq 0$, where "=" iff $f_G(\boldsymbol{U}_0) = 0$.

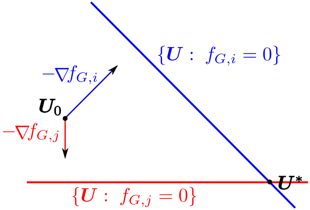
Idea of Proof

1. Let $\boldsymbol{U}^*$ be a global minimizer: $f_G = \mathbf{0}$.
   $\boldsymbol{U}^* \in \bigcap_i \{\boldsymbol{U} : f_{G,i}(\boldsymbol{U}) = 0\}$.
2. For a random $\boldsymbol{U}_0$, compute $-\nabla f_{G,i}$.
3. $\forall i$, Proj $(-\nabla f_{G,i}, \boldsymbol{U}^* - \boldsymbol{U}_0) \geq 0$, where "=" iff $f_{G,i}(\boldsymbol{U}_0) = 0$.
4. $\nabla f_G = \sum_i \nabla f_{G,i} \Rightarrow$
   Proj $(-\nabla f_G, \boldsymbol{U}^* - \boldsymbol{U}_0) \geq 0$, where "=" iff $f_G(\boldsymbol{U}_0) = 0$.



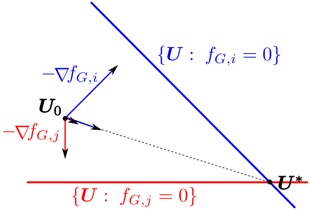
Idea of Proof

1. Let $\boldsymbol{U}^*$ be a global minimizer: $f_G = \mathbf{0}$.
   $\boldsymbol{U}^* \in \bigcap_i \{\boldsymbol{U} : f_{G,i}(\boldsymbol{U}) = 0\}$.
2. For a random $\boldsymbol{U}_0$, compute $-\nabla f_{G,i}$.
3. $\forall i$, Proj $(-\nabla f_{G,i}, \boldsymbol{U}^* - \boldsymbol{U}_0) \geq 0$, where "=" iff $f_{G,i}(\boldsymbol{U}_0) = 0$.
4. $\nabla f_G = \sum_i \nabla f_{G,i} \Rightarrow$
   Proj $(-\nabla f_G, \boldsymbol{U}^* - \boldsymbol{U}_0) \geq 0$, where "=" iff $f_G(\boldsymbol{U}_0) = 0$.



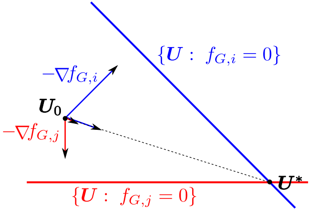
Idea of Proof

1. Let $\boldsymbol{U}^*$ be a global minimizer: $f_G = \mathbf{0}$.
   $\boldsymbol{U}^* \in \bigcap_i \{\boldsymbol{U} : f_{G,i}(\boldsymbol{U}) = 0\}$.
2. For a random $\boldsymbol{U}_0$, compute $-\nabla f_{G,i}$.
3. $\forall i$, Proj $(-\nabla f_{G,i}, \boldsymbol{U}^* - \boldsymbol{U}_0) \geq 0$, where "=" iff $f_{G,i}(\boldsymbol{U}_0) = 0$.
4. $\nabla f_G = \sum_i \nabla f_{G,i} \Rightarrow$
   Proj $(-\nabla f_G, \boldsymbol{U}^* - \boldsymbol{U}_0) \geq 0$, where "=" iff $f_G(\boldsymbol{U}_0) = 0$.



Idea of Proof

1. Let $\boldsymbol{U}^*$ be a global minimizer: $f_G = \mathbf{0}$.
   $\boldsymbol{U}^* \in \bigcap_i \{\boldsymbol{U} : f_{G,i}(\boldsymbol{U}) = 0\}$.
2. For a random $\boldsymbol{U}_0$, compute $-\nabla f_{G,i}$.
3. $\forall i$, Proj $(-\nabla f_{G,i}, \boldsymbol{U}^* - \boldsymbol{U}_0) \geq 0$, where "=" iff $f_{G,i}(\boldsymbol{U}_0) = 0$.
4. $\nabla f_G = \sum_i \nabla f_{G,i} \Rightarrow$
   Proj $(-\nabla f_G, \boldsymbol{U}^* - \boldsymbol{U}_0) \geq 0$, where "=" iff $f_G(\boldsymbol{U}_0) = 0$.



Summary

- ℓ_0-search for LRMC.
- Geometric objective function.
- Strong performance guarantees for two cases.
  - Do not require incoherence.
  - Holds with probility one.
  - Valid for arbitrary matrix size.
- Future work
  - Analysis for the general case.