Harmonic Mean Iteratively Reweighted Least Squares for Low-Rank Matrix Recovery

Christian Kümmerle
joint work with
Juliane Sigl (TUM)

Technische Universität München,
Department of Mathematics

SampTA 2017, July 7, 2017
Problem: Low-Rank Matrix Recovery

Harmonic Mean Iteratively Reweighted Least Squares

Numerical Experiments

Conclusions
## Matrix completion

<table>
<thead>
<tr>
<th>User</th>
<th>movie I</th>
<th>movie II</th>
<th>movie III</th>
<th>movie IV</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>User A</td>
<td>1</td>
<td>?</td>
<td>5</td>
<td>4</td>
<td>...</td>
</tr>
<tr>
<td>User B</td>
<td>?</td>
<td>2</td>
<td>3</td>
<td>?</td>
<td>...</td>
</tr>
<tr>
<td>User C</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>?</td>
<td>...</td>
</tr>
<tr>
<td>User D</td>
<td>?</td>
<td>5</td>
<td>1</td>
<td>3</td>
<td>...</td>
</tr>
<tr>
<td>User E</td>
<td>1</td>
<td>2</td>
<td>?</td>
<td>?</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

**Training Data:**
- 480K users, 18K movies, 100M ratings
- Ratings 1-5 (99% ratings missing)

**Goal:**
- $1M prize for 10% reduction in RMSE over Cinematch

BellKor’s Pragmatic Chaos declared winners on 9/21/2009 used ensemble of models, an important ingredient being low-rank factorization

---

**Data and Mathematics**

**Recommender systems**

- Goal: Reconstruct missing entries
- Highly ill-posed problem
- Netflix prize problem
- Observation: Only few different “types of users”
- Low-rank matrix completion

---

**Problem:**

\[
\begin{align*}
\text{min} & \quad \| P \|_F \\
\text{subject to} & \quad R_1, R_2 \text{ of rank } r
\end{align*}
\]

\[
\| P \|_F = \sqrt{\sum_{i,j} P_{ij}^2}
\]

\[r \leq \min(p, d)\]

\[
\| X \|_2 \leq \epsilon
\]

**Note:**

\[
\text{df} = r \leq \min(p, d)
\]

\[
\text{cannot work, count degrees of freedom of SVD}
\]
Matrix completion

Problem: Recover $X_0 \in \mathbb{R}^{d_1 \times d_2}$ of rank $r \ll \min(d_1, d_2)$ from

$$((X_0)_{i_\ell j_\ell})_{\ell=1}^m = \langle e_{i_\ell} e_{j_\ell}^T, X_0 \rangle_{\ell=1}^m$$

and $m$ small.

- Note: $m < df := r(d_1 + d_2 - r)$ cannot work, count degrees of freedom of SVD
Low-rank matrix recovery problem

Assume rank-$r$ matrix $X_0 \in \mathbb{R}^{d_1 \times d_2}$, measured by given $m$ matrices $(A_\ell)_{\ell=1}^m \subset \mathbb{R}^{d_1 \times d_2}$, with $m \ll d_1 \cdot d_2$.

Find $X \in \mathbb{R}^{d_1 \times d_2}$

s.t. $\langle A_\ell, X \rangle = \langle A_\ell, X_0 \rangle$ for all $\ell \in [m]$, 

$\text{rank}(X) = r$
Low-rank matrix recovery problem

Assume rank-$r$ matrix $X_0 \in \mathbb{R}^{d_1 \times d_2}$, measured by given $m$ matrices $(A_\ell)_{\ell=1}^m \subset \mathbb{R}^{d_1 \times d_2}$, with $m \ll d_1 \cdot d_2$.

Find $X \in \mathbb{R}^{d_1 \times d_2}$

s.t. $\langle A_\ell, X \rangle = \langle A_\ell, X_0 \rangle$ for all $\ell \in [m]$,

$\text{rank}(X) = r$

General framework for:

- matrix completion
- phase retrieval
- blind deconvolution
- quantum state tomography
- ...

Classical approach: Convex Relaxation

**Nuclear Norm Minimization** [Fazel ’02, Recht and Candès ’08]:

\[
\min \|X\|_{S_1} \quad (\|X\|_{S_1} = \|\sigma(X)\|_{\ell_1})
\]

s.t. \( \Phi(X)_\ell := \langle A_\ell, X \rangle = \langle A_\ell, X_0 \rangle \) for all \( \ell \in [m] \).
Classical approach: Convex Relaxation

**Nuclear Norm Minimization** [Fazel ’02, Recht and Candès ’08]:
\[
\min \|X\|_{S_1} \quad (\|X\|_{S_1} = \|\sigma(X)\|_{\ell_1})
\]
\[
s.t. \Phi(X) = \langle A_\ell, X \rangle = \langle A_\ell, X_0 \rangle \quad \text{for all } \ell \in [m].
\]

**Theorem (Recht, Candès ’08, Recht ’11, Candès, Plan ’11)**

Let \(X_0 \in \mathbb{R}^{d_1 \times d_2}\) be “generic” rank-\(r\), \(d_f = r(d_1 + d_2 - r)\)

1. **There exists** \(C_1 \geq 1\) **such that if** \((A_\ell)_{\ell=1}^m = (e_i e_j^T)_{\ell=1}^m\) **are** \(m\) **uniformly distributed matrix completion measurements and if**

\[
m \geq C_1 d_f \log^2(\max(d_1, d_2)),
\]

**NNM succeeds to recover** \(X_0\) **with high probability.**

2. **There exists** \(C_2 \geq 1\) **such that if** \((A_\ell)_{\ell=1}^m\) **are i.i.d. Gaussians and if**

\[
m \geq C_2 d_f,
\]

**NNM succeeds to recover** \(X_0\) **with high probability.**
Beyond convex relaxations

- [Tropp et al. ’14]: $C_2 = 3$ for NNM. But: $3 \times$ oversampling not desirable!
  
  Q: Are there algorithms that solve the problem already for
  
  \[ m \approx C d_f = d_f, \]
  
  i.e., with a $C \approx 1$ and smaller than the one of NNM?

- Need for fast algorithms with low storage requirements:
  Few and fast iterations, maybe $O(d_f)$ variables instead of $O(d_1 d_2)$
Beyond convex relaxations

[Tropp et al. '14]: $C_2 = 3$ for NNM. But: $3 \times$ oversampling not desirable!

Q: Are there algorithms that solve the problem already for

$$m \approx C d_f = d_f,$$

i.e., with a $C \approx 1$ and smaller than the one of NNM?

Need for fast algorithms with low storage requirements:
Few and fast iterations, maybe $O(d_f)$ variables instead of $O(d_1 d_2)$

→ Wide variety of non-convex algorithms:
Optimization of matrix factors $U, V$ such that $X = UV^T$
("alternating minimization"), iterative hard thresholding, manifold optimization, belief propagation...
Beyond convex relaxations

- [Tropp et al. ’14]: $C_2 = 3$ for NNM. But: $3 \times$ oversampling not desirable!

Q: Are there algorithms that solve the problem already for

\[ m \approx d_f, \]

i.e., with a $C \approx 1$ and smaller than the one of NNM?

- Need for fast algorithms with low storage requirements:
  Few and fast iterations, maybe $O(d_f)$ variables instead of $O(d_1 d_2)$

→ Wide variety of non-convex algorithms:
  Optimization of matrix factors $U$, $V$ such that $X = UV^T$
  (“alternating minimization”), iterative hard thresholding, manifold optimization, belief propagation...

, our algorithm: HM-IRLS
Outline

Problem: Low-Rank Matrix Recovery

Harmonic Mean Iteratively Reweighted Least Squares

Numerical Experiments

Conclusions
Iteratively Reweighted Least Squares for Matrix Recovery

If $\sigma(X)$ vector of singular values of $X \in \mathbb{R}^{d_1 \times d_2}$,

$$\|X\|_{S_p}^p := \sum_{i=1}^{\min(d_1,d_2)} \sigma_i(X)^p \xrightarrow{p \to 0} \text{rank}(X).$$

Strategy of IRLS: For some $0 < p < 1$, mimic

$$\min \|X\|_{S_p}^p \quad \text{s.t. } \Phi(X) = \Phi(X_0)$$

(Schatten-$p$ minimization)

[Oymak, Mohan, Fazel, Hassibi ’11, Fornasier, Rauhut, Ward ’11]
by a sequence of (re-)weighted least squares problems.
Iteratively Reweighted Least Squares for Matrix Recovery

If $\sigma(X)$ vector of singular values of $X \in \mathbb{R}^{d_1 \times d_2}$,

$$\|X\|_{S_p}^p := \sum_{i=1}^{\min(d_1,d_2)} \sigma_i(X)^p \xrightarrow{p \to 0} \text{rank}(X).$$

- **Strategy of IRLS:** For some $0 < p < 1$, mimic

  $$\min \|X\|_{S_p}^p \quad \text{s.t. } \Phi(X) = \Phi(X_0)$$

  (Schatten-$p$ minimization)


- **Schatten-$p$ min** is at least as successful as Schatten-$q$ min if $1 \leq p \leq q \leq 1$

- **non-convex** if $p < 1 \Rightarrow$ optimization method crucial

- Analogy to $\ell_p$-minimization $\leftrightarrow \ell_1$-minimization in compressed sensing
Let $X \in \mathbb{R}^{d \times d}$ be of full rank [Mohan, Fazel '11, Fornasier, Rauhut, Ward '11]. Then:

$$
\|X\|_{S_p}^p = \text{tr}[{(XX^T)^{p/2}}] = \text{tr}[{(XX^T)^{p-2}}XX^T] = \text{tr}(W_L XX^T) =: \|X\|_{F(W_L)}^2,
$$

(column space weighting).

Also:

$$
\|X\|_{S_p}^p = \text{tr}[{(XX^T)^{p/2}}] = \text{tr}[{(XX^T)^{p-2}}XX^T] = \text{tr}(W_R XX^T) =: \|X\|_{F(W_R)}^2,
$$

(row space weighting).
Schatten-$p$ quasinorms and weighted Frobenius norms

Let $X \in \mathbb{R}^{d \times d}$ be of full rank [Mohan, Fazel '11, Fornasier, Rauhut, Ward '11]. Then:

$$\|X\|_{S_p}^p = \text{tr}[(XX^T)^{p/2}] = \text{tr}[\left(XX^T\right)^{\frac{p-2}{2}}XX^T] = \text{tr}(W_LXX^T) =: \|X\|^2_{F(W_L)},$$

(column space weighting). Also:

$$\|X\|_{S_p}^p = \text{tr}[(X^TX)^{p/2}] = \text{tr}[X^TX\left(X^TX\right)^{\frac{p-2}{2}}] = \text{tr}(X^TXW_R) =: \|X^T\|^2_{F(W_R)},$$

(row space weighting).
Derive iterative procedure, e.g.:

- Given $X^{(n)}$, solve

$$X^{(n+1)} := \arg \min_{\Phi(X)=Y} \|X\|_{\ell_2(W_{L}^{(n)})}^2$$

with $W_{L}^{(n)}=(X^{(n)}X^{(n)T} + \epsilon(n)^2 \text{Id}_{d_1})^{\frac{p-2}{2}}$ with $\epsilon(n) = \sigma_{r+1}(X^{(n)})$.

- Iterate for $n = 1, 2, \ldots$
IRLS with column space reweighting

Proposed in: [Fornasier, Rauhut, Ward ’11, Mohan, Fazel ’11]

Algorithm: **IRLS-FRW**

**Input:** linear $\Phi : \mathbb{R}^{d_1 \times d_2} \rightarrow \mathbb{R}^m$, $\Phi(X_0) \in \mathbb{R}^m$, rank parameter $R$, $0 < p \leq 1$

**Output:** $(X^{(n)})_{n=1}^{n_0}$

Initialize $n = 0$, $\epsilon^{(0)} = 1$, $W_L^{(1)} = I_{d_1}$.

for $n = 1, 2, \ldots$ do

1. Solve linearly constrained QP $\overset{\hat{\cdot}}{\text{Solve linear equation:}}$

   $$X^{(n)} = \arg\min_{\Phi(X) = \Phi(X_0)} \|X\|^2_F(W_L^{(n)})$$
IRLS with column space reweighting

Proposed in: [Fornasier, Rauhut, Ward ’11, Mohan, Fazel ’11]

Algorithm: IRLS-FRW

Input: linear $\Phi : \mathbb{R}^{d_1 \times d_2} \to \mathbb{R}^m$, $\Phi(X_0) \in \mathbb{R}^m$, rank parameter $R$, $0 < p \leq 1$

Output: $(X^{(n)})_{n=1}^{n_0}$

Initialize $n = 0$, $\epsilon^{(0)} = 1$, $W_L^{(1)} = I_{d_1}$.

for $n = 1, 2, \ldots$ do

1. Solve linearly constrained QP $\hat{X}$ Solve linear equation:

   $$X^{(n)} = \arg \min_{\Phi(X)=\Phi(X_0)} \|X\|_F^2(W_L^{(n)})$$

2. Perform a singular value decomposition: $X^{(n)} = U^{(n)}\Sigma^{(n)}V^{(n)T}$,

   $$\epsilon^{(n)} = \min \left( \epsilon^{(n-1)}, \Sigma^{(n)}_{R+1,R+1} \right),$$

3. Define $W_L^{(n+1)} = (X^{(n)}X^{(n)T} + \epsilon^{(n)2}I_{d_1})^{\frac{p-2}{2}}$
Discussion of prior IRLS algorithms for Schatten-$p$ minimization

- Little gain in empirical success rate of IRLS-FRW for $p < 1$ compared to convex approaches.
Discussion of prior IRLS algorithms for Schatten-$p$ minimization

- Little gain in empirical success rate of IRLS-FRW for $p < 1$ compared to convex approaches
- Related IRLS algorithm for $\ell_p$-minimization in compressed sensing [Daubechies, DeVore, Fornasier, Güntürk ’10] exhibits local convergence rate of order $2 - p$, but this is not observed for IRLS-FRW in low-rank matrix recovery.
Little gain in empirical success rate of \textit{IRLS-FRW} for \( p < 1 \) compared to convex approaches

Related IRLS algorithm for \( \ell_p \)-minimization in compressed sensing [Daubechies, DeVore, Fornasier, Güntürk ’10] exhibits local convergence rate of order \( 2 - p \), but this is not observed for IRLS-FRW in low-rank matrix recovery.

Prior art performs \textit{either row space reweighting or column space reweighting}. Is the structure of the problem used in an optimal way in the IRLS algorithms?
Algorithm: **IRLS-FRW** [Fornasier, Rauhut, Ward ’11]

**Input:** linear $\Phi : \mathbb{R}^{d_1 \times d_2} \to \mathbb{R}^m$, $\Phi(X_0) \in \mathbb{R}^m$, rank parameter $R$, $0 < p \leq 1$

**Output:** $(X^{(n)})_{n=1}^{n_0}$

Initialize $n = 0$, $\epsilon^{(0)} = 1$, $W_L^{(1)} = I_{d_1}$.

for $n = 1, 2, \ldots$ do

1. Solve linearly constrained QP $\overset{\wedge}{\text{Solve linear equation}}$:

   \[
   X^{(n)} = \arg \min_{\Phi(X) = \Phi(X_0)} \|X\|^2_F (W_L^{(n)})
   \]

2. Perform a singular value decomposition: $X^{(n)} = U^{(n)} \Sigma^{(n)} V^{(n)*}$,

   $\epsilon^{(n)} = \min \left( \epsilon^{(n-1)}, \Sigma^{(n)}_{R+1,R+1} \right)$,

3. Define $W_L^{(n+1)} = (X^{(n)} X^{(n)*} + \epsilon^{(n)^2} I_{d_1})^{\frac{p-2}{2}}$
IRLS with harmonic row and column space reweighting

Algorithm: IRLS-FRW [Fornasier, Rauhut, Ward '11]

Input: linear $\Phi : \mathbb{R}^{d_1 \times d_2} \to \mathbb{R}^m$, $\Phi(X_0) \in \mathbb{R}^m$, rank parameter $R$, $0 < p \leq 1$

Output: $(X^{(n)})_{n=1}^{n_0}$

Initialize $n = 0, \epsilon^{(0)} = 1, W_L^{(1)} = I_{d_1}$. 

for $n = 1, 2, \ldots$ do

1. Solve linearly constrained QP $\triangleright$ Solve linear equation:

\[
X^{(n)} = \arg \min_{\Phi(X) = \Phi(X_0)} \left\| X_{\text{vec}} \right\|_{\ell_2}^2 (\mathbf{I}_{d_2} \otimes W_L^{(n)})
\]

2. Perform a singular value decomposition: $X^{(n)} = U^{(n)} \Sigma^{(n)} V^{(n)\ast}$,

$\epsilon^{(n)} = \min \left( \epsilon^{(n-1)}, \Sigma_{R+1,R+1}^{(n)} \right)$,

3. Define $W_L^{(n+1)} = \left( X^{(n)} X^{(n)\ast} + \epsilon^{(n)2} \mathbf{I}_{d_1} \right)^{\frac{p-2}{2}}$
IRLS with *harmonic* row and column space reweighting

**Algorithm:** HM-IRLS [K., Sigl ’17]

**Input:** linear $\Phi : \mathbb{R}^{d_1 \times d_2} \to \mathbb{R}^m$, $\Phi(X_0) \in \mathbb{R}^m$, rank parameter $R$, $0 < p \leq 1$

**Output:** $(X^{(n)})_{n=1}^{n_0}$

Initialize $n = 0$, $\epsilon(0) = 1$, $W_L^{(1)} = I_{d_1}$.

for $n = 1, 2, \ldots$ do

1. Solve linearly constrained QP $\tilde{\Phi}$ = Solve linear equation:

   $$X^{(n)} = \arg\min_{\Phi(X)=\Phi(X_0)} \|X_{vec}\|_2^2 \ell_2(\tilde{W}^{(n)})$$

2. Perform a singular value decomposition: $X^{(n)} = U^{(n)} \Sigma^{(n)} V^{(n)*}$,

   $$\epsilon^{(n)} = \min \left( \epsilon^{(n-1)}, \Sigma_{R+1,R+1}^{(n)} \right)$$

3. Def. $W_L^{(n+1)} = (X^{(n)} X^{(n)T} + \epsilon^{(n)} 2 \text{Id}_{d_1})^{\frac{p-2}{2}}$, $W_R^{(n+1)} = (X^{(n)} X^{(n)} + \epsilon^{(n)} 2 \text{Id}_{d_2})^{\frac{p-2}{2}}$,

   $$\tilde{W}^{(n+1)} := 2 \left( \text{Id}_{d_2} \otimes (W_L^{(n+1)})^{-1} + (W_R^{(n+1)})^{-1} \otimes \text{Id}_{d_1} \right)^{-1} \in \mathbb{R}^{d_1 d_2 \times d_1 d_2}$$
Theory: Harmonic Mean Iteratively Reweighted Least Squares

- Theory of IRLS-FRW extended to HM-IRLS, covering also non-convex case \( p < 1 \):

**Theorem ([K., Sigl ’17])**

Let \( \Phi : \mathbb{R}^{d_1 \times d_2} \to \mathbb{R}^m \) linear, \( X_0 \in \mathbb{R}^{d_1 \times d_2} \) be of rank \( r \). Let \( (X^{(n)}, \epsilon^{(n)})_{n \geq 1} \) be output of HM-IRLS for input \( \Phi, \Phi(X_0), r \) and \( 0 < p \leq 1 \), let \( \epsilon := \lim_{n \to \infty} \epsilon^{(n)} \).

(i) If \( \epsilon \geq 0 \), then each accumulation point \( \bar{X}_\epsilon \) of \( (X^{(n)})_{n \geq 1} \) is a stationary point of the \( \epsilon \)-perturbed Schatten-\( p \) objective

\[
g_{\epsilon,p}(X) = \sum_{i=1}^{d_1 \wedge d_2} (\sigma_i(X)^2 + \epsilon^2)^{\frac{p}{2}}
\]

under the linear constraint \( \Phi(X) = \Phi(X_0) \).
Main result: Local Convergence with Superlinear Rate

Theorem ([K., Sigl ’17])

In the above setting, if Φ fulfills a Schatten-$p$ null space property of order $2r$ with constant $\gamma_{2r} < 1$, the iterates $(X^{(n)})_{n \geq 1}$ of HM-IRLS converge locally with a rate of $2 - p$ to $X_0$ in a neighborhood of the rank-$r$ matrix $X_0$, i.e., there exists a constant $\mu > 1$ (that depends on $X_0$) such that if $\|X^{(n_0)} - X_0\|_{2 \to 2}$ is small enough, then

$$\|X^{(n)} - X_0\|_{2 \to 2} \leq \mu \|X^{(n-1)} - X_0\|_{2 \to 2}^{2-p}$$

for all $n \geq n_0$. 

Schatten-$p$ null space property of order $2r$ is fulfilled with high probability e.g. for Φ with i.i.d. Gaussian entries if $m \geq Cd\|\cdot\|_{\infty}^2$ [Oymak et al. 11, Chavez-Dominguez et al. ’15].

This theorem does not (yet) cover the matrix completion case.
Main result: Local Convergence with Superlinear Rate

Theorem ([K., Sigl ’17])

In the above setting, if \( \Phi \) fulfills a Schatten-\( p \) null space property of order \( 2r \) with constant \( \gamma_{2r} < 1 \), the iterates \( (X^{(n)})_{n \geq 1} \) of HM-IRLS converge locally with a rate of \( 2 - p \) to \( X_0 \) in a neighborhood of the rank-\( r \) matrix \( X_0 \), i.e., there exists a constant \( \mu > 1 \) (that depends on \( X_0 \)) such that if \( \|X^{(n_0)} - X_0\|_{2 \rightarrow 2} \) is small enough, then

\[
\|X^{(n)} - X_0\|_{2 \rightarrow 2} \leq \mu \|X^{(n-1)} - X_0\|_{2 \rightarrow 2}^{2-p}
\]

for all \( n \geq n_0 \).

- Schatten-\( p \) null space property of order \( 2r \) is fulfilled with high probability e.g. for \( \Phi \) with i.i.d. Gaussian entries if \( m \geq Cd_f \) [Oymak et al. 11, Chavez-Dominguez et al. ’15].
- This theorem does not (yet) cover the matrix completion case.
Outline

Problem: Low-Rank Matrix Recovery

Harmonic Mean Iteratively Reweighted Least Squares

Numerical Experiments

Conclusions
Convergence speed experiments: Matrix completion

Figure: Log. error plot for HM-IRLS and IRLS-FRW. Problem parameters $d_1 = d_2 = 40$, $r = 10$, for problems with easy ($\rho = \frac{m}{d_f} = 2$) and harder ($\rho = \frac{m}{d_f} = 1.2$) recovery problems

- Easy recovery problems
- Hard recovery problems
Experiments confirm theory: local convergence rate of $2 - p$ for $p < 1 \Rightarrow$ superlinear convergence of HM-IRLS.
Experiments confirm theory: **local convergence rate of $2 - p$ for $p < 1 \Rightarrow$ superlinear convergence of HM-IRLS.**

This rate cannot be observed for IRLS algorithms as IRLS-FRW.

**HM-IRLS** one of very few iterative algorithms for low-rank recovery with superlinear rate
Sample complexity experiments: Matrix Completion

![Graph showing the empirical recovery success rate with varying oversampling factor $\rho$ for various MC algorithms, including HM-IRLS for $p = 0.01$ and $p = 0.1$. Problem parameters $d_1 = d_2 = 100$, $r = 8$, random $X_0$, modified uniform sampling.](image)

**Figure:** Empirical recovery success rate with varying oversampling factor $\rho := m/d_f$ for various MC algorithms, including HM-IRLS for $p = 0.01$ and $p = 0.1$. Problem parameters $d_1 = d_2 = 100$, $r = 8$, random $X_0$, modified uniform sampling.
Sample complexity experiments: Matrix Completion

- **HM-IRLS** needs fewer measurements ($\approx$ theoretical lower bound) for successful recovery than any other comparison algorithm.
Observation: Breakdown of global convergence to $x_0$ for $\tau < 0.5$
$\ell_\tau$ minimization for sparse recovery, $\tau < 1$ by IRLS

[Daubechies, DeVore, Fornasier, Güntürk '10, Lai, Xu, Yin '13]

- Observation: Breakdown of global convergence to $x_0$ for $\tau < 0.5$
- Not the case for HM-IRLS for low-rank recovery!
  → A property of HM-IRLS that is fundamentally more favorable than in the case of [DDFG '10]
Why does the harmonic mean weight matrix work?

\[ W_L = (XX^T + \epsilon^2 \text{Id}_{d_1})^{\frac{p-2}{2}}, \quad W_R = (X^TX + \epsilon^2 \text{Id}_{d_2})^{\frac{p-2}{2}} \]

Recall harmonic mean weight matrix:

\[ \hat{W} := 2 \left( \text{Id}_{d_2} \otimes W_L^{-1} + W_R^{-1} \otimes \text{Id}_{d_1} \right)^{-1} \in \mathbb{R}^{d_1d_2 \times d_1d_2} \]

- If rank(\(X\)) \(\approx r\) and \(\epsilon \to 0\): \(\hat{W}\) has \(d_f = r(d_1 + d_2 - r)\) large singular values and "only" \(d_1d_2 - d_f\) small ones.
Why does the harmonic mean weight matrix work?

\[ W_L = (XX^T + \epsilon^2 \text{Id}_{d_1})^{\frac{p-2}{2}}, \quad W_R = (X^TX + \epsilon^2 \text{Id}_{d_2})^{\frac{p-2}{2}} \]

Recall harmonic mean weight matrix:

\[ \tilde{W} := 2 \left( \text{Id}_{d_2} \otimes W_L^{-1} + W_R^{-1} \otimes \text{Id}_{d_1} \right)^{-1} \in \mathbb{R}^{d_1 d_2 \times d_1 d_2} \]

- If rank(\(X\)) \(\approx r\) and \(\epsilon \to 0\): \(\tilde{W}\) has \(d_f = r(d_1 + d_2 - r)\) large singular values and ”only“ \(d_1 d_2 - d_f\) small ones.
- Not the case for a arithmetic mean weight matrix

\[ \tilde{W}^{\text{AM}} := \frac{1}{2} \left( \text{Id}_{d_2} \otimes W_L + W_R \otimes \text{Id}_{d_1} \right) \]

since only \(r^2 \ll d_f\) large ones.
Why does the harmonic mean weight matrix work?

\[ W_L = \left( XX^T + \epsilon^2 I_{d_1} \right)^{\frac{p-2}{2}}, \quad W_R = \left( X^T X + \epsilon^2 I_{d_2} \right)^{\frac{p-2}{2}} \]

Recall harmonic mean weight matrix:

\[ \tilde{W} := 2 \left( I_{d_2} \otimes W_L^{-1} + W_R^{-1} \otimes I_{d_1} \right)^{-1} \in \mathbb{R}^{d_1 d_2 \times d_1 d_2} \]

- Action of \( \tilde{W} \) on tangent space
  \( T_X \mathcal{M}_r = \{ Z \in \mathbb{R}^{d_1 d_2} : Z = X R^T + L X^T \text{ s.t. } L \in \mathbb{R}^{d_1 \times r}, R \in \mathbb{R}^{d_2 \times r} \} \)
  at \( X \) of rank-\( r \) manifold is particularly well-behaved.

- Note: Computationally, huge matrix \( \tilde{W} \) does never have to be computed explicitly!
Outline

Problem: Low-Rank Matrix Recovery

Harmonic Mean Iteratively Reweighted Least Squares

Numerical Experiments

Conclusions
Conclusions

- We propose and analyze a new, quickly convergent algorithm for non-convex Schatten-$p$ minimization based on IRLS.

Future Work:
- Improve computational efficiency to tackle large-scale problems.
- Derive a global convergence analysis.
- Extension to related problems: Matrix completion with outliers, robust PCA, imaging problems...
Conclusions

- We propose and analyze a new, quickly convergent algorithm for non-convex Schatten-$p$ minimization based on IRLS.
- Promising empirical behavior for the low-rank matrix recovery problem, especially in terms of sample complexity needed for recovery, compared to a variety of other approaches.

Future Work:
- Improve computational efficiency to tackle large-scale problems.
- Derive a global convergence analysis.
- Extension to related problems: Matrix completion with outliers, robust PCA, imaging problems...
Conclusions

- We propose and analyze a new, quickly convergent algorithm for non-convex Schatten-$p$ minimization based on IRLS.
- Promising empirical behavior for the low-rank matrix recovery problem, especially in terms of sample complexity needed for recovery, compared to a variety of other approaches.
- Extension to noisy setting is straightforward.

Future Work:
- Improve computational efficiency to tackle large-scale problems.
- Derive a global convergence analysis.
- Extension to related problems: Matrix completion with outliers, robust PCA, imaging problems...
Conclusions

- We propose and analyze a new, quickly convergent algorithm for non-convex Schatten-$p$ minimization based on IRLS.
- Promising empirical behavior for the low-rank matrix recovery problem, especially in terms of sample complexity needed for recovery, compared to a variety of other approaches.
- Extension to noisy setting is straightforward.

Future Work:
- Improve computational efficiency to tackle large-scale problems.
Conclusions

- We propose and analyze a new, quickly convergent algorithm for non-convex Schatten-$p$ minimization based on IRLS.
- Promising empirical behavior for the low-rank matrix recovery problem, especially in terms of sample complexity needed for recovery, compared to a variety of other approaches.
- Extension to noisy setting is straightforward.

Future Work:

- Improve computational efficiency to tackle large-scale problems.
- Derive a global convergence analysis.
- Extension to related problems: Matrix completion with outliers, robust PCA, imaging problems...
Thank you!

Code available at:

http://www-m15.ma.tum.de/Allgemeines/ChristianKuemmerle

Harmonic Mean Iteratively Reweighted Least Squares for Low-Rank Matrix Recovery

Iteratively Reweighted Least Squares Minimization for Sparse Recovery

Low-rank Matrix Recovery via Iteratively Reweighted Least Squares Minimization

Iterative reweighted algorithms for matrix rank minimization