Closed-Form Solutions in Low-Rank Subspace Recovery Models and Their Implications

Zhouchen Lin (林宙辰)
北京大学
Nov. 7, 2015
Outline

• Sparsity vs. Low-rankness
• Closed-Form Solutions of Low-rank Models
• Applications of Closed-Form Solutions
• Conclusions
Challenges of High-dim Data

Images
> 1M dim

Videos
> 1B dim

Web data
> 10B+ dim?

Courtesy of Y. Ma.
Sparsity vs. Low-rankness
**Sparse Models**

- **Sparse Representation**

\[
\begin{align*}
\min & \quad ||z||_0, \\
\text{s.t.} & \quad x = Az. \\
\end{align*}
\] (1)

- **Sparse Subspace Clustering**

\[
\begin{align*}
\min & \quad ||z_i||_0, \\
\text{s.t.} & \quad x_i = X_\hat{i} z_i, \quad \forall i. \\
\end{align*}
\] (2)

where \( X_\hat{i} = [x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_n]\).

\[
\begin{align*}
\min & \quad ||Z||_0, \\
\text{s.t.} & \quad X = X Z, \text{diag}(Z) = 0. \\
\end{align*}
\] (3)

\[
\begin{align*}
\min & \quad ||Z||_1, \\
\text{s.t.} & \quad X = X Z, \text{diag}(Z) = 0. \\
\end{align*}
\] (4)

Low-rank Models

• Matrix Completion (MC)
  \[
  \min \text{rank}(A), \quad \text{s.t.} \quad D = \pi(A).
  \]

• Robust PCA
  \[
  \min \text{rank}(A) + \lambda \|E\|_l_0, \quad \text{s.t.} \quad D = A + E.
  \]

• Low-rank Representation (LRR)
  \[
  \min \text{rank}(Z) + \lambda \|E\|_{2,0}, \quad \text{s.t.} \quad X = XZ + E.
  \]

\[
\|E\|_{l_0} = \#\{E_{ij} | E_{ij} \neq 0\} \quad \|E\|_{2,0} = \#\{i | \|E_{i,:}\|_2 \neq 0\}.
\]

NP Hard!
Convex Program Formulation

- Matrix Completion (MC)
  \[
  \min \|A\|_*, \quad s.t. \quad D = \pi(A).
  \]

- Robust PCA
  \[
  \min \|A\|_* + \lambda\|E\|_{l_1}, \quad s.t. \quad D = A + E.
  \]

- Low-rank Representation (LRR)
  \[
  \min \|Z\|_* + \lambda\|E\|_{2,1}, \quad s.t. \quad X = XZ + E.
  \]

  \[
  \|A\|_* = \sum \sigma_i(A),
  \]

  \[
  \|E\|_{l_1} = \sum_{i,j} |E_{ij}|.
  \]

  \[
  \|E\|_{2,1} = \sum_i \|E_{:,i}\|_2.
  \]
Applications of Low-rank Models

- Background modeling
- Robust Alignment
- Image Rectification
- Motion Segmentation
- Image Segmentation
- Saliency Detection
- Image Tag Refinement
- Partial Duplicate Image Search
- ......

林宙辰、马毅，信号与数据处理中的低秩模型，中国计算机学会通讯，2015年第4期。
Closed-form Solution of LRR

• Closed form solution at noiseless case

\[ \min_{Z} \|Z\|_*, \]
\[ s.t. \quad X = X Z, \]

has a unique closed-form optimal solution: \( Z^* = V_r V_r^T \), where \( U_r \Sigma_r V_r^T \) is the skinny SVD of \( X \).

– Shape Interaction Matrix
– when \( X \) is sampled from independent subspaces, \( Z^* \) is block diagonal, each block corresponding to a subspace

\[ \min_{X=XZ} \|Z\|_* = \text{rank}(X). \]

Closed-form Solution of LRR

- Closed form solution at general case

\[ \min_Z \|Z\|_*, \quad s.t. \quad X = AZ, \]

has a unique closed-form optimal solution: \( Z^* = A^\dagger X \).

Valid for any unitary invariant norm!

\[ \|X\|_{UI} = \|UXV^T\|_{UI}, \quad \forall \ U^TU = I, \ V^TV = I. \]

Liu et al., Robust Recovery of Subspace Structures by Low-Rank Representation, TPAMI 2013.
Yao-Liang Yu and Dale Schuurmans, Rank/Norm Regularization with Closed-Form Solutions: Application to Subspace Clustering, UAI2011.
Closed-form Solution of LRR

- Closed form solution of the original LRR

\[
\min_{Z} \text{rank}(Z), \quad \text{s.t.} \quad A = XZ. \tag{1}
\]

**Theorem:** Suppose \( U_X \Sigma_X V_X^T \) and \( U_A \Sigma_A V_A^T \) are the skinny SVD of \( X \) and \( A \), respectively. The complete solutions to feasible generalized LRR problem (1) are given by

\[
Z^* = X^+ A + SV_A^T, \tag{2}
\]

where \( S \) is any matrix such that \( V_X^T S = 0 \).

Latent LRR

• Small sample issue

\[
\begin{align*}
\min & \|Z\|_*, \\
\text{s.t.} & \ X = XZ.
\end{align*}
\]

• Consider unobserved samples

\[
\begin{align*}
\min & \|Z\|_*, \\
\text{s.t.} & \ X_O = [X_O, X_H]Z.
\end{align*}
\]

Latent LRR

\[
\text{min } \|Z\|_*, \\
s.t. \quad X_O = [X_O, X_H]Z.
\]

**Theorem:** Suppose \( Z_{O,H}^* = [Z_{O|H}^*; Z_{H|O}^*] \). Then

\[
Z_{O|H}^* = V_O V_O^T, \quad Z_{H|O}^* = V_H V_H^T,
\]

where \( V_O \) and \( V_H \) are calculated as follows. Compute the skinny SVD: \([X_O, X_H] = U \Sigma V^T\) and partition \( V \) as \( V = [V_O; V_H] \).

**Proof:** That \([X_O, X_H] = U \Sigma [V_O; V_H]^T\) implies

\[
X_O = U \Sigma V_O^T, \quad X_H = U \Sigma V_H^T.
\]

So \( X_O = [X_O, X_H]Z \) reduces to:

\[
V_O^T = V^T Z.
\]

So \( Z_{O,H}^* = V V_O^T = [V_O V_O^T; V_H V_O^T] \).

Latent LRR

\[
\begin{align*}
\min & \|Z\|_*, \\
\text{s.t.} & \quad X_O = [X_O, X_H]Z. \\
Z^*_{O|H} & = V_OV_O^T, \quad Z^*_{H|O} = V_HV_O^T. \\
X_O & = [X_O, X_H]Z^*_{O,H} \\
& = X_OZ^*_{O|H} + X_HZ^*_{H|O} \\
& = X_OZ^*_{O|H} + X_HV_HV_O^T \\
& = X_OZ^*_{O|H} + U\Sigma V_H^TV_HV_O^T \\
& = X_OZ^*_{O|H} + U\Sigma V_H^TV_H\Sigma^{-1}U^TX_O \\
& \equiv X_OZ^*_{O|H} + L^*_{H|O}X_O. 
\end{align*}
\]

min rank(Z) + rank(L), \
\text{s.t.} \quad X = XZ + LX. 

\[
\begin{align*}
\min & \|Z\|_* + \|L\|_*, \\
\text{s.t.} & \quad X = XZ + LX. 
\end{align*}
\]

Latent LRR

\[
\min \|Z\|_* + \|L\|_* + \lambda \|E\|_1,
\]
\[
s.t. \quad X = XZ + LX + E.
\]
Analysis on LatLRR

• Noiseless LatLRR has non-unique closed form solutions!

Theorem: The complete solutions to

$$\min_{Z,L} \text{rank}(Z) + \text{rank}(L), \quad s.t. \quad X = XZ + LX$$

are as follows

$$Z^* = V_X \tilde{W} V_X^T + S_1 \tilde{W} V_X^T$$
$$L^* = U_X \Sigma_X (I - \tilde{W}) \Sigma_X^{-1} U_X^T + U_X \Sigma_X (I - \tilde{W}) S_2,$$

where $\tilde{W}$ is any idempotent matrix and $S_1$ and $S_2$ are any matrices satisfying:

1. $V_X^T S_1 = 0$ and $S_2 U_X = 0$; and
2. $\text{rank}(S_1) \cdot \text{rank}(\tilde{W})$ and $\text{rank}(S_2) \cdot \text{rank}(I - \tilde{W}).$

$$A^2 = A$$

Zhang et al, A Counterexample for the Validity of Using Nuclear Norm as a Convex Surrogate of Rank, ECML/PKDD 2013.
Analysis on LatLRR

**Theorem:** The complete solutions to

$$\min_{Z,L} \|Z\|_* + \|L\|_*, \quad s.t. \quad X = XZ + LX$$

are as follows

$$Z^* = V_X \hat{W} V_X^T \quad \text{and} \quad L^* = U_X (I - \hat{W}) U_X^T,$$

where $\hat{W}$ is any block diagonal matrix satisfying:

1. its blocks are compatible with $\Sigma_X$, i.e., if $[\Sigma_X]_{ii} \neq [\Sigma_X]_{jj}$ then $[\hat{W}]_{ij} = 0$; and

2. both $\hat{W}$ and $I - \hat{W}$ are positive semi-definite.

$$\Sigma_X = \begin{bmatrix} 10 & 0 & 0 & 0 & 0 & 0 \\ 0 & 8 & 0 & 0 & 0 & 0 \\ 0 & 0 & 8 & 0 & 0 & 0 \\ 0 & 0 & 0 & 8 & 0 & 0 \\ 0 & 0 & 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3 \end{bmatrix} \quad \Rightarrow \quad \hat{W} = \begin{bmatrix} 0.8 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.5 & 0.1 & 0.3 & 0 & 0 \\ 0 & 0.1 & 0.6 & 0.2 & 0 & 0 \\ 0 & 0.3 & 0.2 & 0.4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Positive Semi-definite

Robust LatLRR

$$\min_{Z, W} \|Z\|_1, \text{ s.t. } Z = V_X W V_X^T, \text{ } W \text{ is diagonal,}$$

$$0 \leq \text{diag}(W) \leq 1, \text{ } \text{tr}(W) = 1,$$

Comparison on the synthetic data as the percentage of corruptions increases.

Robust LatLRR

Table 1: Segmentation errors (%) on the Hopkins 155 data set. For robust LatLRR, the parameter $\lambda$ is set as $0.806/\sqrt{n}$. The parameters of other methods are also tuned to be the best.

<table>
<thead>
<tr>
<th></th>
<th>SSC</th>
<th>LRR</th>
<th>RSI</th>
<th>LRSC</th>
<th>LatLRR</th>
<th>Robust LatLRR</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAX</td>
<td>46.75</td>
<td>49.88</td>
<td>47.06</td>
<td>40.55</td>
<td>42.03</td>
<td>35.06</td>
</tr>
<tr>
<td>MEAN</td>
<td>2.72</td>
<td>5.64</td>
<td>6.54</td>
<td>4.28</td>
<td>4.17</td>
<td>3.74</td>
</tr>
<tr>
<td>STD</td>
<td>8.20</td>
<td>10.35</td>
<td>9.84</td>
<td>8.55</td>
<td>9.14</td>
<td>7.02</td>
</tr>
</tbody>
</table>

Table 2: Segmentation accuracy (%) on the Extended Yale B data set, with different number of persons. For robust LatLRR, the parameter $\lambda$ is set as 0.014, 0.013, and 0.0135, respectively. The parameters of other methods are also tuned to be the best.

<table>
<thead>
<tr>
<th>Persons</th>
<th>SSC</th>
<th>LRR</th>
<th>RSI</th>
<th>LRSC</th>
<th>LatLRR</th>
<th>Robust LatLRR</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>90.31</td>
<td>91.25</td>
<td>88.44</td>
<td>75.94</td>
<td>69.06</td>
<td>95.00</td>
</tr>
<tr>
<td>7</td>
<td>87.05</td>
<td>72.77</td>
<td>89.73</td>
<td>65.63</td>
<td>42.86</td>
<td>92.86</td>
</tr>
<tr>
<td>9</td>
<td>71.35</td>
<td>60.42</td>
<td>87.67</td>
<td>51.04</td>
<td>36.11</td>
<td>91.49</td>
</tr>
</tbody>
</table>
Relationship Between LR Models

(Original RPCA) \( \min_{A,E} \text{rank}(A) + \lambda \|E\|_\ell, \)
\[
\text{s.t. } X = A + E.
\]

(Original Robust LRR) \( \min_{Z,E} \text{rank}(Z) + \lambda \|E\|_\ell, \)
\[
\text{s.t. } X - E = (X - E)Z.
\]

(Original Robust LatLRR) \( \min_{Z,L,E} \text{rank}(Z) + \text{rank}(L) + \lambda \|E\|_\ell, \)
\[
\text{s.t. } X - E = (X - E)Z + L(X - E).
\]
Relationship Between LR Models

(Heuristic RPCA) $\min_{A,E} \|A\|_* + \lambda \|E\|_\ell,$

s.t. $X = A + E.$

(Heuristic Robust LRR) $\min_{Z,E} \|Z\|_* + \lambda \|E\|_\ell,$

s.t. $X - E = (X - E)Z.$

(Heuristic Robust LatLRR) $\min_{Z,L,E} \|Z\|_* + \|L\|_* + \lambda \|E\|_\ell,$

s.t. $X - E = (X - E)Z + L(X - E).$

Relationship Between LR Models

Original Robust LRR

\[
\min_{Z,E} \text{rank}(Z) + \lambda \|E\|_{\ell}, \\
\text{s.t. } X - E = (X - E)Z.
\]

Heuristic Robust LRR

\[
\min_{Z,E} \|Z\|_* + \lambda \|E\|_{\ell}, \\
\text{s.t. } X - E = (X - E)Z.
\]

Original RPCA

\[
\min_{A,E} \text{rank}(A) + \lambda \|E\|_{\ell}, \\
\text{s.t. } X = A + E.
\]

Original Robust LatLRR

\[
\min_{Z,L,E} \text{rank}(Z) + \text{rank}(L) + \lambda \|E\|_{\ell}, \\
\text{s.t. } X - E = (X - E)Z + L(X - E).
\]

Heuristic Robust LatLRR

\[
\min_{Z,L,E} \|Z\|_* + \|L\|_* + \lambda \|E\|_{\ell}, \\
\text{s.t. } X - E = (X - E)Z + L(X - E).
\]

Implications

• We could obtain a *globally optimal* solution to other low rank models.
• We could have *much faster* algorithms for other low rank models.
Implications

• Comparison of optimality

Comparison of accuracies of solutions to relaxed R-LRR computed by REDU-EXPR and partial ADM, where the parameter is adopted as $1/\sqrt{\log n}$ and n is the input size. The program is run by 10 times and the average accuracies are reported.

Implications

• Comparison of speed

<table>
<thead>
<tr>
<th>Model</th>
<th>Method</th>
<th>Accuracy</th>
<th>CPU Time (h)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LRR</td>
<td>ADM</td>
<td>-</td>
<td>&gt;10</td>
</tr>
<tr>
<td>R-LRR</td>
<td>ADM</td>
<td>-</td>
<td>did not converge</td>
</tr>
<tr>
<td>R-LRR</td>
<td>partial ADM</td>
<td>-</td>
<td>&gt;10</td>
</tr>
<tr>
<td>R-LRR</td>
<td>REDU-EXPR</td>
<td>61.6365%</td>
<td>0.4603</td>
</tr>
</tbody>
</table>

Table 1: Unsupervised face image clustering results on the Extended YaleB database. REDU-EXPR means reducing to RPCA first and then express the solution as that of RPCA.

Implications

• Comparison of optimality and speed

Table 4: Comparison of robustness and speed between partial ADM (LRSC) (Favaro et al., 2011) and REDU-EXPR (RSI) (Wei and Lin, 2010) methods for solving R-LRR when the percentage of corruptions increases. All the experiments are run ten times and the $\lambda$ is set to be the same: $\lambda = 1/\sqrt{\log n}$, where $n$ is the data size.

<table>
<thead>
<tr>
<th>Noise Percentage (%)</th>
<th>0</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rank($Z$) (partial ADM)</td>
<td>20</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>Rank($Z$) (REDU-EXPR)</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>$|E|_2,0$ (partial ADM)</td>
<td>0</td>
<td>99</td>
<td>200</td>
<td>300</td>
<td>400</td>
<td>500</td>
</tr>
<tr>
<td>$|E|_2,0$ (REDU-EXPR)</td>
<td>0</td>
<td>100</td>
<td>200</td>
<td>300</td>
<td>400</td>
<td>500</td>
</tr>
<tr>
<td>Objective (partial ADM)</td>
<td><strong>20.00</strong></td>
<td>67.67</td>
<td>106.10</td>
<td>144.14</td>
<td>182.19</td>
<td>220.24</td>
</tr>
<tr>
<td>Objective (REDU-EXPR)</td>
<td><strong>20.00</strong></td>
<td><strong>58.05</strong></td>
<td><strong>96.10</strong></td>
<td><strong>134.14</strong></td>
<td><strong>172.19</strong></td>
<td><strong>210.24</strong></td>
</tr>
<tr>
<td>Time (s, partial ADM)</td>
<td><strong>4.89</strong></td>
<td>124.33</td>
<td>126.34</td>
<td>119.12</td>
<td>115.20</td>
<td>113.94</td>
</tr>
<tr>
<td>Time (s, REDU-EXPR)</td>
<td>10.67</td>
<td><strong>9.60</strong></td>
<td><strong>8.34</strong></td>
<td><strong>8.60</strong></td>
<td><strong>9.00</strong></td>
<td><strong>12.86</strong></td>
</tr>
</tbody>
</table>

$O(n^1)$ RPCA by $l_1$-Filtering

- Assumption: $\text{rank}(A) = o(n)$.
\(O(n^1)\) RPCA by \(l_1\)-Filtering

- First, randomly sample \(D_{sub}\).

\[
\begin{align*}
K &= cr \\
D_{sub} \\
D \\
\min \|A\|_* + \lambda \|E\|_{l_1}, \\
s.t. \quad D = A + E.
\end{align*}
\]

$O(n^1)$ RPCA by $l_1$-Filtering

- Second, solve RPCA for $D_{sub}$: $D_{sub} = A_{sub} + E_{sub}$.

$$\min \|A_{sub}\|_* + \lambda \|E_{sub}\|_1 \quad \text{s.t.} \quad D_{sub} = A_{sub} + E_{sub}.$$
$O(n^1)$ RPCA by $l_1$-Filtering

- Third, find the full rank submatrix $B$ of $A_{sub}$.

\[ \begin{align*}
K &= cr \\
r &= o(n) \\
\min \|A\|_* + \lambda \|E\|_{l_1}, \\
\text{s.t.} \quad D &= A + E.
\end{align*} \]

$O(n^1)$ RPCA by $l_1$-Filtering

• Fourth, correct the $Kx(n-K)$ submatrix of $D$ by $C_{sub}$.

$$\min_{p_c} \|d_c - C_{sub}p_c\|_1$$

$O(n)$ complexity! Can be done in parallel!!

$$\min \|A\|_* + \lambda \|E\|_{l_1}, \quad s.t. \quad D = A + E.$$
$O(n^1)$ RPCA by $l_1$-Filtering

- Fifth, correct the $(n-K)\times K$ submatrix of $D$ by $R_{sub}$.

\[
\min_{q_r} \| d_r - q_r R_{sub} \|_1
\]

$O(n)$ complexity! Can be done in parallel!!

Rows and columns can also be processed in parallel!!

\[
\min \| A \|_* + \lambda \| E \|_{l_1}, \quad s.t. \quad D = A + E.
\]

$O(n^1)$ RPCA by $l_1$-Filtering

- Finally, the rest part of $A$ is $A^c = Q_rBP_c$.

\[ P_c = (p_1, p_2, \cdots, p_{n-K}) \]

\[ Q_r = \begin{pmatrix} q_1 \\ q_2 \\ \cdots \\ q_{n-K} \end{pmatrix} \]

\[ K = cr \]

\[ r = o(n) \]

A compact representation of $A$!

\[ \min \|A\|_* + \lambda\|E\|_{l_1}, \text{ s.t. } D = A + E. \]

$O(n^1)$ RPCA by $l_1$-Filtering

- What if the rank is unknown?

1. If $\text{rank}(A_{sub}) > K/c$, then increase $K$ to $c \text{rank}(A_{sub})$.
2. Otherwise, resample another $D_{sub}$ for cross validation.

$$\min \|A\|_* + \lambda \|E\|_{l_1}, \quad s.t. \quad D = A + E.$$
# Experiments

## Synthetic Data

<table>
<thead>
<tr>
<th>Size</th>
<th>Method</th>
<th>$\frac{|L_0 - L^*|_F}{|L_0|_F}$</th>
<th>rank($L_0$)</th>
<th>$|L^<em>|_</em>$</th>
<th>$|S^*<em>0|</em>{l_0}$</th>
<th>$|S^*<em>1|</em>{l_1}$</th>
<th>Time(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>S-ADM</td>
<td>$1.46 \times 10^{-8}$</td>
<td>20</td>
<td>39546</td>
<td>39998</td>
<td>998105</td>
<td>84.73</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>L-ADM</td>
<td>$4.72 \times 10^{-7}$</td>
<td>20</td>
<td>39546</td>
<td>40229</td>
<td>998105</td>
<td>27.41</td>
</tr>
<tr>
<td></td>
<td>$l_1$</td>
<td>$1.66 \times 10^{-8}$</td>
<td>20</td>
<td>39546</td>
<td>40000</td>
<td>998105</td>
<td>5.56</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\frac{|L_0 - L^*|_F}{|L_0|_F}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>rank($L_0$) = 50, $|L_0|<em>* = 249432$, $|S_0|</em>{l_0} = 250000$, $|S_0|_{l_1} = 6246093$</td>
<td>1093.96</td>
<td>195.79</td>
<td>42.34 = 19.66 + 22.68</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5000</td>
<td>S-ADM</td>
<td>$7.13 \times 10^{-9}$</td>
<td>50</td>
<td>249432</td>
<td>249995</td>
<td>6246093</td>
<td>1093.96</td>
</tr>
<tr>
<td></td>
<td>L-ADM</td>
<td>$4.28 \times 10^{-7}$</td>
<td>50</td>
<td>249432</td>
<td>250636</td>
<td>6246158</td>
<td>195.79</td>
</tr>
<tr>
<td></td>
<td>$l_1$</td>
<td>$5.07 \times 10^{-9}$</td>
<td>50</td>
<td>249432</td>
<td>250000</td>
<td>6246093</td>
<td>42.34 = 19.66 + 22.68</td>
</tr>
<tr>
<td>10000</td>
<td>S-ADM</td>
<td>$1.23 \times 10^{-8}$</td>
<td>100</td>
<td>997153</td>
<td>1000146</td>
<td>25004071</td>
<td>11258.51</td>
</tr>
<tr>
<td></td>
<td>L-ADM</td>
<td>$4.26 \times 10^{-7}$</td>
<td>100</td>
<td>997153</td>
<td>1000744</td>
<td>25005109</td>
<td>1301.83</td>
</tr>
<tr>
<td></td>
<td>$l_1$</td>
<td>$2.90 \times 10^{-10}$</td>
<td>100</td>
<td>997153</td>
<td>1000023</td>
<td>25004071</td>
<td>276.54 = 144.38 + 132.16</td>
</tr>
</tbody>
</table>

## Structure form Motion Data

<table>
<thead>
<tr>
<th>Size</th>
<th>Method</th>
<th>rank($L_0$)</th>
<th>$|L_0|<em>* = 31160$, $|S_0|</em>{l_0} = 900850$, $|S_0|_{l_1} = 3603146$</th>
<th>$|L^<em>|_</em>$</th>
<th>$|S^*<em>0|</em>{l_0}$</th>
<th>$|S^*<em>1|</em>{l_1}$</th>
<th>Time(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4002 x 2251</td>
<td>S-ADM</td>
<td>$5.38 \times 10^{-8}$</td>
<td>4</td>
<td>31160</td>
<td>900726</td>
<td>3603146</td>
<td>925.28</td>
</tr>
<tr>
<td></td>
<td>L-ADM</td>
<td>$3.25 \times 10^{-7}$</td>
<td>4</td>
<td>31160</td>
<td>1698387</td>
<td>3603193</td>
<td>62.08</td>
</tr>
<tr>
<td></td>
<td>$l_1$</td>
<td>$1.20 \times 10^{-8}$</td>
<td>4</td>
<td>31160</td>
<td>900906</td>
<td>3603146</td>
<td>5.29 = 3.51 + 1.78</td>
</tr>
</tbody>
</table>

Experiments

Conclusions

- Low-rank models have much richer mathematical properties than sparse models.
- Closed-form solutions to low-rank models are useful in both theory and applications.
Thanks!

• zlin@pku.edu.cn
• http://www.cis.pku.edu.cn/faculty/vision/zlin/zlin.htm