Fast algorithm for low-rank matrix recovery in Poisson noise

Yao Xie
Joint work with Yang Cao

H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology
Presented at GlobalSIP 2014
Static feature-specific imaging (SFSI)

Object prior knowledge: high/low diversity training sets $A^1$

Calculate $R_z^1$ from $A^1$

Find $M$ eigenvectors of $R_z^1$ to form projection matrix $F (M \times N)$

Parallel architecture FSI

Object $x$

Feature measurement $y = Fx$

Object estimate $\hat{x} = F^r y$
Formulation

- Goal: recover a matrix $M^* \in \mathbb{R}^{m_1 \times m_2}_{+}$
- Observation: $N$ linear measurements with Poisson noise

\[ y_i \sim \text{Poisson}(\langle AM^* \rangle_i), \quad i = 1, \ldots, N, \]

- linear operator $A : \mathbb{R}^{m_1 \times m_2}_{+} \rightarrow \mathbb{R}^N$ models the measuring process of physical devices

\[ [AM]_i = \langle A, M \rangle \triangleq \text{tr}(A^\top M), \]

$A_i \in \mathbb{R}^{m_1 \times m_2}$
Connection with earlier work

- matrix completion:
  - noiseless $y_i = [AM]_i$ [Candes, Recht 2009],
  - with Gaussian measurement noise $y_i = [AM]_i + w_i$ [Candes, Yaniv 2009] [Cai, Candes, Shen 2010]
- compressed sensing with Poisson noise [Raginsky et.al. 2010]

$$y_i \sim \text{Poisson}([Ax]_i)$$

- feasibility study and performance bound for matrix recovery with Poisson noise [Xie, Chi, Calderbank 2013]
- low-rank signal recovery with Poisson noise (directly observe entries) [Soni, Haupt 2014]

$$Y_{ij} \sim \text{Poisson}([M]_{ij}), \quad \{M \text{ is low-rank}\}$$

- this work: efficient algorithm for matrix recovery with Poisson noise
Regularized maximum-likelihood estimator

\[ \hat{M} \triangleq \arg \min_{M \in \Gamma} \left[ -\log p(y|AM) + \lambda \rho(M) \right] \]

\[ = \arg \min_{M \in \Gamma} \left[ -\sum_{j=1}^{N} y_j \log [AM]_j - [AM]_j + \lambda \rho(M) \right] \]

- \( \rho(M) > 0 \): regularization function
- \( \lambda > 0 \): regularization parameter
- \( \Gamma \): set of feasible estimators
Assumptions

- Total intensity of $M^*$ is known a priori
  \[ I \triangleq \| M^* \|_{1,1}, \]
  where \( \| X \|_{1,1} = \sum_i \sum_j [X]_{ij} \)

- Positivity-preserving of $A$: $[M]_{ij} \geq 0$ for all $i, j \Rightarrow [AM]_i \geq 0$, for all $i$

- Flux-preserving of $A$: $\sum_{i=1}^N [AM]_i \leq \| M \|_{1,1}$
Linear sensing operator $A$, 

$[A_{i}]_{jk} = \begin{cases} 
0, & \text{with probability } p; \\
1/N, & \text{with probability } 1 - p. 
\end{cases}$

satisfies earlier assumptions, and the restrictive isometry property (RIP) 

[Raginsky, Willett, Harmaney, Marcia 2010], [Xie, Chi, Calderbank 2013]
Optimization problem

- \( \rho(M) = \| M \|_* \)
- Optimization problem

\[
\min_{M \in \Gamma_0} f(M) + \lambda \| M \|_* ,
\]

where \( f(M) = -\log p(y|AM) \)

\[
\Gamma_0 \triangleq \{ M \in \mathbb{R}^{m_1 \times m_2}_+ : \| M \|_{1,1} = I \}.
\]

- Convex optimization: Semidefinite program (SDP)
- More efficient algorithm: eigenvalue thresholding
Taylor expansion and approximation

- Taylor expansion of log-likelihood function at \((k - 1)\)th solution

\[
Q_{t_k}(M, M_{k-1}) \triangleq f(M_{k-1}) + \langle M - M_{k-1}, \nabla f(M_{k-1}) \rangle \\
+ \frac{t_k}{2} \| M - M_{k-1} \|_F^2,
\]

\[
\propto \frac{t_k}{2} \left\| M - \left( M_{k-1} - \frac{1}{t_k} \nabla f(M_{k-1}) \right) \right\|_F^2
\]

where \(t_k\) is the step size at \(k\)th iteration
Solution for the next iteration

\[ M_k = \arg \min_M \left[ \frac{1}{2} \left\| M - \left( M_{k-1} - \frac{1}{t_k} \nabla f(M_{k-1}) \right) \right\|^2_F + \frac{\lambda}{t_k} \| M \|_* \right] \]

Theorem (Cai, Candes, Shen 2010)

For each \( \tau \geq 0 \), and \( X \in \mathbb{R}^{n_1 \times n_2} \):

\[ D_\tau(X) = \arg \min_{Y \in \mathbb{R}^{n_1 \times n_2}} \left\{ \frac{1}{2} \| Y - X \|_F^2 + \tau \| Y \|_* \right\} \]  \hspace{1cm} (1)

\[ D_\tau(X) \triangleq U D_\tau(\Sigma) V^T, \]

\[ X = U \Sigma V^T \]
Solution of the approximate optimization problem

Solution given by **Singular Value Thresholding (SVT)**

\[ M_k = D_{\lambda/t_k} \left( M_{k-1} - \frac{1}{t_k} \nabla f(M_{k-1}) \right). \]
Poisson noise Maximal Likelihood Singular Value thresholding (PMLSV) algorithm

Algorithm 1 PMLSV

1: Initialize: $M_0 = \mathcal{P}(\sum_{i=1}^n y_i A_i)$, parameter $\gamma$, step size $L$
2: for $k = 1, 2, \ldots NOI$ do
3: \[ G(M_{k-1}) := \nabla \left[ - \log p(y | AM_{k-1}) \right] \]
4: \[ C := M_{k-1} - \frac{1}{L} G(M_{k-1}) \]
5: singular value decomposition: $C := UDV^T$
6: \[ D_{\text{new}} := \text{diag}((\text{diag}(D) - \frac{\lambda}{L})_+) \]
7: \[ W_k := \mathcal{P}(UD_{\text{new}}V^T). \]
8: If $F(M_k) < F(M_{k-1})$, then $k = k + 1$; else $L = \gamma L$, go to 6.
9: If $|F(M_k) - F(M_{k-1})| < 0.5/NOI$, then $k = k - 1$, exit;
10: end for
Numerical examples

- normalized risk:

\[ R(M^*, M) \triangleq \frac{1}{I^2} \| M^* - M \|_F^2. \]

- Parameters are as follows: \( I = 2.37 \times 10^7 \), \( L = 10^{-5} \), \( \gamma = 1.1 \), and \( NOI = 2500 \).

(a) original solar flare image    (b) solar flare image with rank 10
Solution quality compared with SDP

**Fig. 2**: Risk vs the number of measurements when $\alpha = 4$ and $\lambda = 0.002$. Points from left to right correspond to risk when $N = 500, 750, 1000, 1250, 1500$, respectively, and solved by CVX and PMLSV.

**Table 1**: CPU time (in seconds) of solving SDP by using CVX and our PMLSV algorithm when fixing $\alpha = 4$ and $\lambda = 0.002$ with $500, 750, 1000, 1250$ and $1500$ measurements, respectively.

<table>
<thead>
<tr>
<th>$N$</th>
<th>SDP</th>
<th>PMLSV</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>725s</td>
<td>172s</td>
</tr>
<tr>
<td>750</td>
<td>1146s</td>
<td>232s</td>
</tr>
<tr>
<td>1000</td>
<td>1510s</td>
<td>378s</td>
</tr>
<tr>
<td>1250</td>
<td>2059s</td>
<td>490s</td>
</tr>
<tr>
<td>1500</td>
<td>2769s</td>
<td>642s</td>
</tr>
</tbody>
</table>

Larger $\alpha$ (hence the higher the SNR) we have, the lower the risk as demonstrated in Fig. 4.

**Fig. 3**: Recovery results when fixing $N = 1000, \lambda = 0.002$ with different values of $\alpha$.

**Fig. 4**: Risk vs $\alpha$ when fixing $N = 1000, \lambda = 0.002$. Points from left to right mean the risk with $\alpha = 1$ to $\alpha = 9$.

We demonstrate its accuracy and efficiency compared with the semi-definite program (SDP) and tested on real data examples of solar flare images. Future work includes analyzing the convergence property of the algorithm, and extension to the related matrix completion problem with Poisson noise.

**Fig. 5**: Recovery results with different values of $\lambda$ when fixing $N = 1000$ and $\alpha = 4$.

**Fig. 6**: Risk vs $\lambda$ when $N = 1000, \alpha = 4$. Points from left to right means the risk when $\lambda = 0.0007$ to $\lambda = 0.0039$ with step size 0.0004.
Computation time

Table 1: CPU time (in seconds) of solving SDP by using CVX and our PMLSV algorithm when fixing $\alpha = 4$ and $\lambda = 0.002$ with 500, 750, 1000, 1250 and 1500 measurements, respectively.

<table>
<thead>
<tr>
<th>$N$</th>
<th>500</th>
<th>750</th>
<th>1000</th>
<th>1250</th>
<th>1500</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDP</td>
<td>725s</td>
<td>1146s</td>
<td>1510s</td>
<td>2059s</td>
<td>2769s</td>
</tr>
<tr>
<td>PMLSV</td>
<td>172s</td>
<td>232s</td>
<td>378s</td>
<td>490s</td>
<td>642s</td>
</tr>
</tbody>
</table>
Changing SNR

- to change SNR of the image, we scale the image intensity by \( \alpha \geq 1 \).
- recovery results when \( N = 1000 \) and \( \lambda = 0.002 \).
Changing SNR

N vs Number of Observations when $\lambda = 0.002$.

Table 1: CPU time (in seconds) of solving SDP by using CVX and our PMLSV algorithm when fixing $\alpha = 4$ and $\lambda = 0.002$ with 500, 750, 1000, 1250 and 1500 measurements, respectively.

---

500 600 700 800 900 1000 1100 1200 1300 1400 1500
5.2
5.4
5.6
5.8
6
6.2
6.4
6.6
6.8
7
7.2 x 10^{-5}

---

Fig. 2: Risk vs the number of measurements when $\alpha = 4$ and $\lambda = 0.002$. Points from left to right correspond to risk when $N = 500$, 750, 1000, 1250, and 1500, respectively, and solved by CVX and PMLSV.

---

Fig. 3: Recovery results when fixing $N = 1000$, $\lambda = 0.002$ with different value of $\alpha$.

---

Third, we run our algorithm with different values of $\lambda$ when fixing $N = 1000$ and $\alpha = 4$. The results are shown in Fig. 5. From Fig. 6, we can see that there is an optimal value for $\lambda$ which leads to the smallest risk.

---

5. CONCLUSION AND FUTURE WORK

We have presented a new algorithm for low-rank matrix recovery with linear measurements contaminated with Poisson noise: the Poisson noise Maximal Likelihood Singular Value Thresholding (PMLSV) algorithm, based on solving a regularized maximum likelihood problem with nuclear norm as the regularizer.

We demonstrate its accuracy and efficiency compared with the semi-definite program (SDP) and tested on real data examples of solar flare images. Future work include analyzing the convergence property of the algorithm, and extension to the related matrix completion problem with Poisson noise.
Choice of penalty parameter $\lambda$

Recovery results when $N = 1000$ and $\alpha = 4$,

- $\lambda = 0.0007$
- $\lambda = 0.0011$
- $\lambda = 0.0015$
- $\lambda = 0.0019$
- $\lambda = 0.0023$
- $\lambda = 0.0027$
- $\lambda = 0.0031$
- $\lambda = 0.0035$
- $\lambda = 0.0039$

Fig. 5: Recovery results with different value of $\lambda$ when fixing $N = 1000$ and $\alpha = 4$.

Fig. 6: Risk vs $\lambda$ when $N = 1000$, $\alpha = 4$. Points from left to right means the risk when $\lambda = 0.0007$ to $\lambda = 0.0039$ with step size 0.0004.
Choice of penalty parameter $\lambda$

### Fig. 2

Risk vs the number of measurements when $\alpha = 4$ and $\lambda = 0.002$. Points from left to right correspond to risk when $N = 500, 750, 1000, 1250,$ and $1500$, respectively, and solved by CVX and PMLSV.

### Table 1

<table>
<thead>
<tr>
<th>$N$</th>
<th>SDP (in seconds)</th>
<th>PMLSV (in seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>725</td>
<td>172</td>
</tr>
<tr>
<td>750</td>
<td>1146</td>
<td>232</td>
</tr>
<tr>
<td>1000</td>
<td>1510</td>
<td>378</td>
</tr>
<tr>
<td>1250</td>
<td>2059</td>
<td>490</td>
</tr>
<tr>
<td>1500</td>
<td>2769</td>
<td>642</td>
</tr>
</tbody>
</table>

### Fig. 3

Recovery results when fixing $N = 1000$, $\lambda = 0.002$ with different value of $\alpha$.

### Fig. 4

Risk vs $\alpha$ when fixing $N = 1000$, $\lambda = 0.002$. Points from left to right means the risk with $\alpha = 1$ to $\alpha = 9$.

### Fig. 5

Recovery results with different value of $\lambda$ when fixing $N = 1000$ and $\alpha = 4$.

### Fig. 6

Risk vs $\lambda$ when $N = 1000$, $\alpha = 4$. Points from left to right means the risk when $\lambda = 0.0007$ to $\lambda = 0.0039$ with step size $0.0004$.

We have presented a new algorithm for low-rank matrix recovery with linear measurements contaminated with Poisson noise: the Poisson noise Maximal Likelihood Singular Value Thresholding (PMLSV) algorithm, based on solving a regularized maximum likelihood problem with nuclear norm as the regularizer.

We demonstrate its accuracy and efficiency compared with the semi-definite program (SDP) and tested on real data examples of solar flare images. Future work include analyzing the convergence property of the algorithm, and extension to the related matrix completion problem with Poisson noise.
Extension to matrix completion

\[ F_{\Omega,Y}(X) = \sum_{(i,j) \in \Omega} Y_{i,j} \log X_{i,j} - X_{i,j}, \]

matrix recovery by the following method

\[
\hat{M} = \arg \max_{X \in \mathbb{R}^{d_1 \times d_2}} F_{\Omega,Y}(X),
\]

\[ s.t. \quad \|X\|_* \leq \alpha \sqrt{d_1 d_2 r}, \quad \gamma_1 \leq \|X\|_\infty \leq \gamma_2 \]
Performance guarantee

**Theorem**

Assume that $\|M\|_* \leq \alpha \sqrt{d_1 d_2 r}$ and $\gamma_1 \leq \|M\|_\infty \leq \gamma_2$. $\Omega$ is chosen at random binomial with $\mathbb{E}|\Omega| = m$. $\widehat{M}$ is defined in (2). Then with probability at least $(1 - C/(d_1 + d_2))$, we have

$$\frac{1}{d_1 d_2} \|M - \widehat{M}\|^2_F \leq C^* \left( \frac{8\gamma_2 T}{1 - e^{-T}} \right) \left( \alpha \sqrt{r} + \frac{1}{\sqrt{d_1 d_2}} \right) \sqrt{\frac{d_1 + d_2}{m}} \sqrt{1 + \frac{(d_1 + d_2) \log(d_1 d_2)}{m}},$$

(3)

If $m \geq (d_1 + d_2) \log(d_1 d_2)$ then this simplifies to

$$\frac{1}{d_1 d_2} \|M - \widehat{M}\|^2_F \leq \sqrt{2} C^* \left( \frac{8\gamma_2 T}{1 - e^{-T}} \right) \left( \alpha \sqrt{r} + \frac{1}{\sqrt{d_1 d_2}} \right) \sqrt{\frac{d_1 + d_2}{m}}.$$

$T, C^*, C$: constants.
Fast algorithm for matrix completion

- algorithm similar to PMLSV can be used except that we modify the likelihood function, which only affect the gradient
- solution again based on singular value thresholding

(a) $p = 0.5$  
(b) $\alpha = 1000, \gamma_1 = 1, \gamma_2 = 12000$

**Figure**: Image with missing data and recovery result
Summary

- Fast algorithm for low-rank matrix recovery and matrix completion
- Key idea: approximating the log likelihood function by sequential Taylor expansion
- Known exact solution to the approximated cost function via Singular Value Thresholding