Poisson Matrix Completion

Yang Cao and Yao Xie†

† H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology

Background

- **Matrix completion** recover a matrix $M \in \mathbb{R}^{d_1 \times d_2}$ from incomplete observations.
- **Low-rank matrix:**
  (i) exact low-rank matrix: rank$(M) \leq r$;
  (ii) approximate low-rank matrix: $\|M\|_\infty \leq \alpha$ and rank$(M) \leq r \Rightarrow \|M\|_* \leq \alpha\sqrt{r d_1 d_2}$.
- **Quantized matrix completion:** observations and noise are 'quantized' and classical techniques can not guarantee the recovery performance. [1] gives the first result for a quantized case in which the observations are -1 or 1.
- **Applications:** Netflix problem, network traffic, counting data.
- **Recovery algorithms:** solving a penalized nuclear norm regularization optimization problem (convex optimization).
  $\Longrightarrow$ **Can we guarantee the recovery performance for Poisson observations?**

Formulation

- Suppose the matrix $M \in \mathbb{R}_+^{d_1 \times d_2}$ consists of underlying parameters for the Poisson observations is (approximate) low-rank.
- Assume a subset $\Omega \subseteq \{(i,j) : 1 \leq i \leq d_1, 1 \leq j \leq d_2\}$ and we observe
$$Y_{ij} \sim \text{Poisson}(M_{ij}), \quad \forall (i,j) \in \Omega.$$
- **Assumptions:**
  (i) M is not too spiky: $M_{ij} \leq \alpha$ for some $\alpha > 0$;
  (ii) M is approximate low-rank: $\|M\|_* \leq \alpha\sqrt{r d_1 d_2}$;
  (iii) **for Poisson case only:** $M_{ij} \geq \beta$ for some $\beta > 0$. β can be interpreted by the minimum Signal-to-Noise Ratio (SNR).
- **Goal:** Recover M based on the incomplete observations $Y_{ij}, (i,j) \in \Omega$.

Regularized maximum-likelihood estimator

- **Method:** Maximizing the log-likelihood function of the optimization variable X given our observations subject to a set of convex constraints.
- Log-likelihood function for Poisson matrix completion:
$$F_{\Omega,Y}(X) = \sum_{(i,j)\in\Omega} Y_{ij} \log X_{ij} - X_{ij}.$$
- Candidate set:
$$\mathcal{S} \triangleq \{X \in \mathbb{R}_+^{d_1 \times d_2} : \|X\|_* \leq \alpha\sqrt{r d_1 d_2}, \; \beta \leq X_{ij} \leq \alpha, \forall (i,j) \in [d_1] \times [d_2]\}.$$
- Estimator:
$$\widehat{M} = \arg\max_{X \in \mathcal{S}} F_{\Omega,Y}(X). \tag{1}$$

Performance bounds

Performance metric: square error
$$R(\widehat{M}, M) = \|\widehat{M} - M\|_F^2.$$

Theorem (Upper bound): Assume $M \in \mathcal{S}$, Ω is chosen at random following our Bernoulli sampling model with $\mathbb{E}[|\Omega|] = m$, and $\widehat{M}$ is the solution to (1). Then with a probability exceeding $(1 - C/(d_1 d_2))$, we have
$$\frac{1}{d_1 d_2} R(M, \widehat{M}) \leq C' \left(\frac{8\alpha T}{1 - e^{-T}}\right) \cdot \left(\frac{\alpha\sqrt{r}}{\beta}\right) \cdot \left(\alpha(e^2 - 2) + 3\log(d_1 d_2)\right) \cdot \left(\frac{d_1 + d_2}{m}\right)^{1/2} \cdot \left[1 + \frac{(d_1 + d_2)\log(d_1 d_2)}{m}\right]^{1/2}. \tag{2}$$

If $m \geq (d_1 + d_2)\log(d_1 d_2)$, then (2) simplifies to
$$\frac{1}{d_1 d_2} R(M, \widehat{M}) \leq \sqrt{2} C' \left(\frac{8\alpha T}{1 - e^{-T}}\right) \cdot \left(\frac{\alpha\sqrt{r}}{\beta}\right) \cdot \left(\alpha(e^2 - 2) + 3\log(d_1 d_2)\right) \cdot \left(\frac{d_1 + d_2}{m}\right)^{1/2}. \tag{3}$$

- C', C are absolute constants and T depends only on α and β.
- The expectation and probability are with respect to the random Poisson observations and Bernoulli sampling model.

Proof sketch:

- Relate likelihood function deviate from Kullback-Leibler (K-L) divergence
- Relate Hellinger distance to K-L divergence
- Lower bound for Hellinger distance

Remark: If the number of observations is on the order of $(d_1 + d_2)\log^\delta(d_1 d_2)$ with $\delta > 2$, mean squared error
$$\frac{1}{d_1 d_2} R(\widehat{M}, M) \to 0 \quad \text{as} \quad (d_1 + d_2) \to \infty.$$

Theorem (Lower bound): Fix α, β, r, d_1, and d_2 to be such that $\alpha \geq 1$, $\alpha \geq 2\beta$, $r \geq 4$, and $\alpha^2 r \max\{d_1, d_2\} \geq C_0$. Fix Ω_0 be an arbitrary subset of $[d_1] \times [d_2]$ with cardinality m. Consider any algorithm which, for any $M \in \mathcal{S}$, returns an estimator $\widehat{M}$. Then there exists $M \in \mathcal{S}$ such that with probability at least $3/4$,
$$\frac{1}{d_1 d_2} R(M, \widehat{M}) \geq \min\left\{\frac{1}{256}, C_2 \alpha^{3/2} \left[\frac{r \max\{d_1, d_2\}}{m}\right]^{1/2}\right\}, \tag{4}$$
as long as the right-hand side of (4) exceeds $C_1 r \alpha^2 / \min\{d_1, d_2\}$, where C_0, C_1 and C_2 are absolute constants. Here the probability is with respect to the random Poisson observations only.

Remark: The gap between the upper bound in (3) and the lower bound in (4) is asymptotical on the order of $\log(d_1 d_2)$. This gap is due to: (i) locally sub-gaussian property of Poisson distribution; (ii) unbounded range of Poisson random variables. The same order of gap can also be found in a recent similar work [2].

Penalized maximum likelihood singular value threshold(PMLSVT)

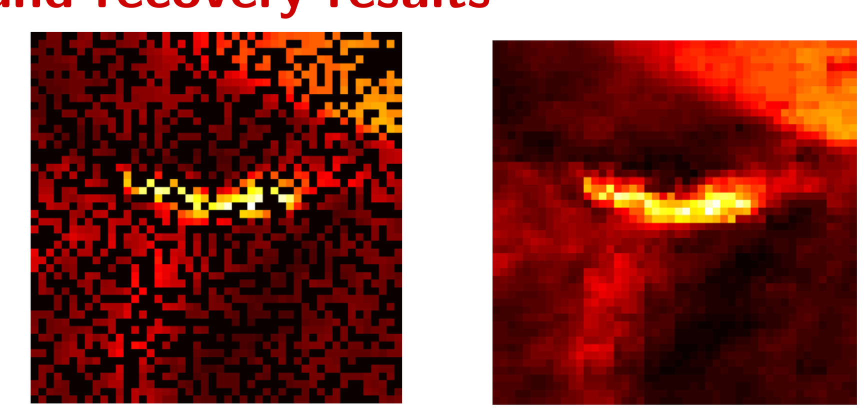
PMLSVT algorithm for Poisson matrix completion

- Input: Y, Ω.
- Output: An approximate solution to (1).
- Define $f(X) \triangleq -F_{\Omega,Y}(X)$ be the negative log-likelihood function. Define Π_{Γ_1} as the projection of a matrix onto $\Gamma_1 \triangleq \{X \in \mathbb{R}^{d_1 \times d_2} : \beta \leq X_{ij} \leq \alpha\}$.

Algorithm 1 PMLSVT for Poisson matrix completion

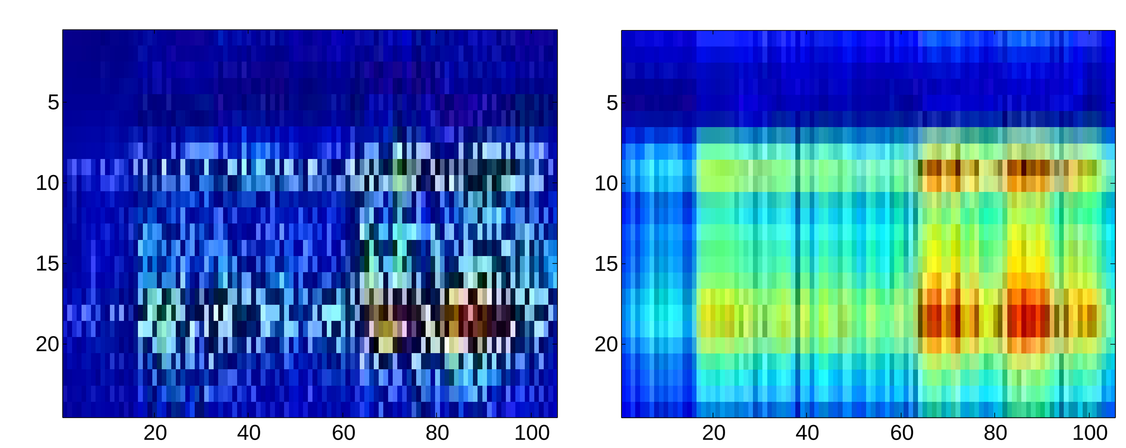
1: Initialize: The maximum number of iterations K, parameters η, and t. $[M_0]_{ij} = Y_{ij}$ for $(i,j) \in \Omega$ and is $(\alpha + \beta)/2$ otherwise.
2: **for** $k = 1, 2, \ldots K$ **do**
3: $\quad C = M_{k-1} - (1/t)\nabla f(M_{k-1})$
4: $\quad C = UDV^T$ {singular value decomposition}
5: $\quad D_{\text{new}} = \text{diag}((\text{diag}(D) - \lambda/t)_+)$
6: $\quad M_k = \Pi_{\Gamma_1}(U D_{\text{new}} V^T)$
7: $\quad$ If $f(M_k) > f(M_{k-1})$ then $t = \eta t$, go to 4.
8: $\quad$ If $|f(M_k) - f(M_{k-1})| < 0.5/K$ then exit;
9: **end for**

Examples and recovery results



(a) $p = 0.5$. (b) $\lambda = 0.1, K = 2000$.

Figure: Matrix completion from partial observations: (a): 50% of entries observed (dark spots represent missing entries); (b): images formed by complete matrix with $\lambda = 0.1$ and no more than 2000 iterations, and the run time of the PMLSVT algorithm is 1.1 seconds.



(a) original data, $p = 0.5$. (b) $\lambda = 100, K = 4000$.

Figure: Bike sharing count data: (a): observed matrix M with 50% missing entries; (b): recovered matrix with $\lambda = 100$ and 4000 iterations, with an elapsed time of 3.1 seconds.

[1] 1-Bit matrix completion. Information and Inference: A Journal of the IMA *(Mark A. Davenport/Yaniv Plan/Ewout van den Berg/Mary Wootters,2014)*
[2] Low Rank Matrix Completion with Exponential Family Noise *(Jean Lafond,2015)*