Low-rank matrix completion and denoising under Poisson noise

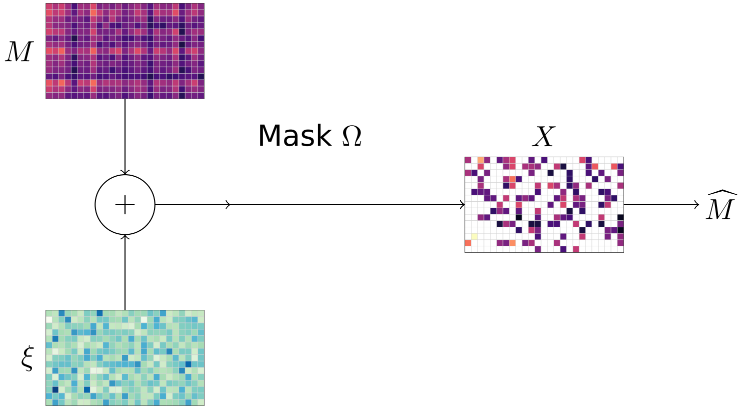
Andrew D. McRae
SPARS 2019

Work supervised by Mark Davenport

Georgia Institute of Technology
School of Electrical and Computer Engineering

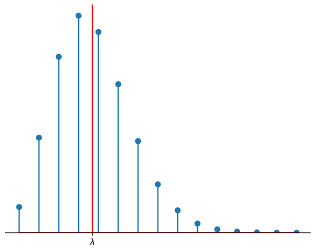
Matrix completion and denoising

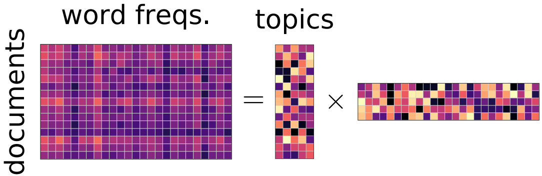
- Denoising: exploit structure to reduce noise contamination
- Completion: recover matrix from only a few noisy entries!



The Poisson noise case

- Often an ideal model for *count* data
- Example: photons arriving at an imaging sensor
- Example: topic modeling





Prior work: generic noise

- A lot of literature (Candes and Plan 2010; Keshavan, Montanari, and Oh 2010; Negahban and Wainwright 2012, etc.)
- Typical estimator:

$$\widehat{M} = \operatorname*{argmin}_{M'} \sum_{(i,j)\in\Omega} (X_{ij} - M'_{ij})^2 + \alpha \|M'\|_*$$

- Typical theoretical bound (see, e.g., Klopp 2015):

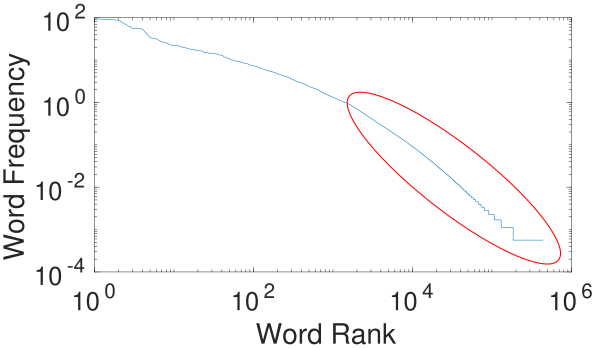
$$\|\widehat{M} - M\|_F^2 \lesssim \frac{\overset{\overset{\text{rank}(M)}{\downarrow}}{r}(m+n)}{\underset{\underset{\text{fraction of observed entries}}{\uparrow}}{p}} \left(\max_{i,j} |M_{ij}|^2 + \max_{i,j} \operatorname{var}(\xi_{ij}) \right)$$

- Requires uniform upper bound on matrix entries and noise variance
- Best results don't apply to Poisson noise

Prior work: Poisson noise

- Typically MLE with structural constraints/penalties (e.g., Gunasekar, Ravikumar, and Ghosh 2014; Lafond 2015; Soni et al. 2016; Cao and Xie 2016)
- Similar dependence on dimension as generic methods
- Poorly conditioned with low rates: unreasonable assumption

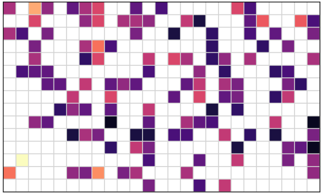
Word frequencies from ~2000 Project Gutenberg texts



Our framework

- True low-rank rate matrix $M \in [0, \infty)^{m \times n}$
- Bernoulli sampling model: $(i, j) \in \Omega$ independently with probability p
- Sampling operator $\mathcal{A}_\Omega \colon \mathbf{R}^{m \times n} \to \mathbf{R}^\Omega$
- Poisson observations: $X \mid \Omega \sim \mathrm{Poisson}(\mathcal{A}_\Omega(M))$
- Consider matrix "Dantzig selector"-type estimator:

$$\widehat{M}^\delta = \operatorname*{argmin}_{M' \in [0,\infty)^{m \times n}} \|M'\|_* \text{ s.t. } \|\mathcal{A}_\Omega^*(X) - pM'\| \leq \delta$$

$\mathcal{A}_\Omega^*(X) =$ 

Upper bound on error

Theorem

If δ is chosen properly, we have

$$\|\widehat{M}^{\delta} - M\|_F \lesssim \sqrt{\frac{r}{p}}\widetilde{\sigma}(M) + \log \text{term},$$

with high probability, where

$$\widetilde{\sigma}(M) = \max_i \sqrt{\sum_{j=1}^{n} M_{ij} + (1-p)M_{ij}^2} + \max_j \sqrt{\sum_{i=1}^{m} M_{ij} + (1-p)M_{ij}^2}.$$

In many common situations, the logarithmic term is negligible.

Discussion

$$\|\widehat{M}^{\delta} - M\|_F \lesssim \underbrace{\sqrt{\frac{r}{p}}}_{\text{Same as before}} \left(\max_i \sqrt{\sum_{j=1}^{n} \underbrace{M_{ij}}_{\text{Noise variance}} + (1-p)M_{ij}^2} + \max_j \sqrt{\sum_{i=1}^{m} M_{ij} + \underbrace{(1-p)M_{ij}^2}_{\text{From sampling variance}}} \right)$$

— Works with low rates!

— No uniform upper bound; compare to previous:

$$\|\widehat{M} - M\|_F \lesssim \sqrt{\frac{r(m+n)}{p}} \sqrt{\max_{i,j} M_{i,j}^2 + \max_{i,j} M_{i,j}}$$

— More refined bound and works with Poisson

Example: denoising case ($p = 1$)

— MLE: $\widehat{M} = X$

$$\mathbf{E}\|\widehat{M} - M\|_F^2 = \sum_{i,j} M_{ij}$$

— If rows and columns have comparable energy, we get

$$\|\widehat{M}^\delta - M\|_F^2 \lesssim \frac{r}{m \wedge n} \mathbf{E}\|\widehat{M} - M\|_F^2$$

— Error reduction $\approx$ reduction in degrees of freedom

Proof outline

- Deterministic: if $\|\mathcal{A}_\Omega^*(X) - pM\| \leq \delta$, then $\|\widehat{M}^\delta - M\|_F$ is small (standard techniques)
- Use matrix concentration to show event holds w.h.p. (Bandeira and R. van Handel 2016; Latała, Ramon van Handel, and Youssef 2018)
- These results assume bounded or Gaussian noise...
- Need truncation argument to apply to Poisson noise
- Avoiding likelihood function avoids poor conditioning at low rates

Minimax lower bound

Theorem

If p is not too small, then, for all sufficiently large $\sigma > 0$,

$$\inf_{\widehat{M}} \sup_{\substack{M \in [0,\infty)^{m \times n} \\ \operatorname{rank}(M) \leq r \\ \widetilde{\sigma}(M) \leq \sigma}} \mathbf{E} \|\widehat{M} - M\|_F^2 \gtrsim \frac{r}{p}\sigma^2,$$

where the infimum is over all estimators

- Matches upper bound within a constant; this is the best we can do!
- Proof by standard arguments from information theory and hypothesis testing

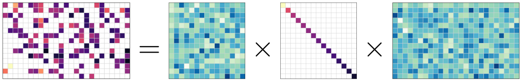
Computation

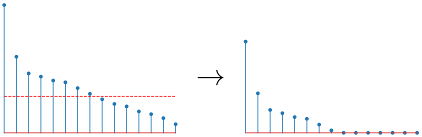
$$\widehat{M}^{\delta} = \underset{M' \in [0,\infty)^{m \times n}}{\operatorname{argmin}} \|M'\|_* \text{ s.t. } \|\mathcal{A}_\Omega^*(X) - pM'\| \leq \delta$$

— Estimator can be computed with semidefinite program

— Without ≥ 0 constraint, solvable with singular value thresholding

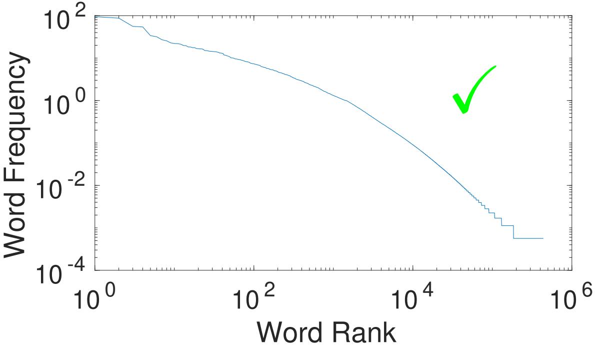
— Same theoretical guarantees hold without constraint





Poisson recap

- Our approach works for Poisson, unlike the best general analysis
- Good performance at low rates!



- Minimax optimal performance
- Amenable to more efficient algorithm than previous work

Generalizations and implications

- All of our methods generalize to many other noise models
- In a minimax sense, simple singular value thresholding is the best we can do!
- Perhaps more sophisticated algorithms can still do better in more restricted settings
- Reminder of gaps in noisy MC theory

Conclusions

- First truly minimax optimal result in Poisson case
- Generalizations refine and extend current state-of-the-art MC results
- Surprising implication: SVD-based algorithms are minimax optimal!

Thanks!

References I

- Bandeira, A. and R. van Handel (July 2016). "Sharp nonasymptotic bounds on the norm of random matrices with independent entries". In: *Ann. Probab.* 44.4, pp. 2479–2506.
- Candes, E. and Y. Plan (June 2010). "Matrix Completion With Noise". In: *Proc. IEEE* 98.6, pp. 925–936.
- Cao, Y. and Y. Xie (Mar. 2016). "Poisson Matrix Recovery and Completion". In: *IEEE Trans. Signal Process.* 64.6, pp. 1609–1620.
- Gunasekar, S., P. Ravikumar, and J. Ghosh (June 2014). "Exponential Family Matrix Completion under Structural Constraints". In: *Proc. Int. Conf. Mach. Learn. (ICML)*. Beijing, China.
- Keshavan, R., A. Montanari, and S. Oh (2010). "Matrix Completion from Noisy Entries". In: *J. Mach. Learn. Res.* 11, pp. 2057–2078.
- Klopp, O. (2015). "Matrix Completion by Singular Value Thresholding: Sharp Bounds". In: *Electron. J. Stat.* 9, pp. 2348–2369.

References II

- Lafond, Jean (July 2015). "Low Rank Matrix Completion with Exponential Family Noise". In: *Proc. Conf. Learn. Theory (COLT)*. Paris, France.
- Latała, Rafał, Ramon van Handel, and Pierre Youssef (Dec. 2018). "The dimension-free structure of nonhomogeneous random matrices". In: *Invent. Math.* 214 (3), pp. 1031–1080.
- Negahban, Sahand and Martin Wainwright (May 2012). "Restricted Strong Convexity and Weighted Matrix Completion: Optimal Bounds with Noise". In: *J. Mach. Learn. Res.* 13, pp. 1665–1697.
- Soni, A. et al. (June 2016). "Noisy Matrix Completion Under Sparse Factor Models". In: *IEEE Trans. Inf. Theory* 62.6, pp. 3636–3661.

Formal Statements: Upper Bound

Theorem

Let M be a non-negative $m \times n$ matrix with rank r. Let $\lambda_{\max} = \max_{i,j} M_{ij}$. Suppose Ω is chosen according to a Bernoulli sampling model with sampling probability p, and suppose $X \sim \mathrm{Poisson}(\mathcal{A}_\Omega(M))$ conditioned on Ω. Set $\epsilon \in (0, 1/2)$, and choose δ such that

$$\delta \geq 2\sqrt{p}\widetilde{\sigma}(M) + \frac{8\epsilon}{\sqrt{mn}} + C \max\left\{\lambda_{\max}, 4\log\frac{2mn}{\epsilon}\right\}\sqrt{\log\frac{m \vee n}{\epsilon}},$$

where C is a universal constant. Then, with probability at least $1 - 2\epsilon$,

$$\|\widehat{M}^\delta - M\|_F \leq \frac{4\sqrt{2r}\delta}{p}.$$

Formal Statements: Lower Bound I

Theorem

Let r, k*, and* ℓ *be positive integers, with* $k \geq \ell$*, and take* $m = rk, n = r\ell$*. Let* $p \in (0, 1]$*,* $\lambda_{\max} \geq 1/8\ell p$*, and set* $\sigma_1^2 = k\lambda_{\max}$*. Define*

$$S_1 = \left\{ M \in [0, \lambda_{\max}]^{m \times n} : \operatorname{rank}(M) \leq r, \right.$$
$$\left. \sqrt{\max_i \sum_j M_{ij}} + \sqrt{\max_j \sum_i M_{ij}} \leq 2\sigma_1 \right\}.$$

Then, under a Bernoulli sampling model with sampling probability p*,*

$$\inf_{\widehat{M}} \sup_{M \in S_1} \mathbf{P}\left(\|\widehat{M} - M\|_F \geq \frac{\sqrt{r}\sigma_1}{8\sqrt{2p}} \right) \geq \frac{1}{2} - \frac{8 \log 2}{m}.$$

Format Statements: Lower Bound II

Theorem

Again, take $m = rk, n = r\ell$ *with* $m \geq n$*. Set* $\sigma_2^2 = k\lambda_{\max}^2$*. Let*

$$S_2 = \left\{ M \in [0, \lambda_{\max}]^{m\times n} : \operatorname{rank}(M) \leq r, \right.$$
$$\left. \sqrt{\max_i \sum_j M_{ij}^2} + \sqrt{\max_j \sum_i M_{ij}^2} \leq 2\sigma_2 \right\}.$$

Suppose $p \geq \frac{r}{2n}$*. Then*

$$\inf_{\widehat{M}} \sup_{M\in S_2} \mathbf{E}\|\widehat{M} - M\|_F^2 \geq \frac{r\sigma_2^2}{8} \max\left\{ \frac{1}{2}\left\lfloor \frac{1}{2p} \right\rfloor, 1 - p \right\}$$
$$\geq \frac{1}{64}\frac{1-p}{p} r\sigma_2^2.$$