Lower Bounds for Quantized Matrix Completion

Mark A. Davenport

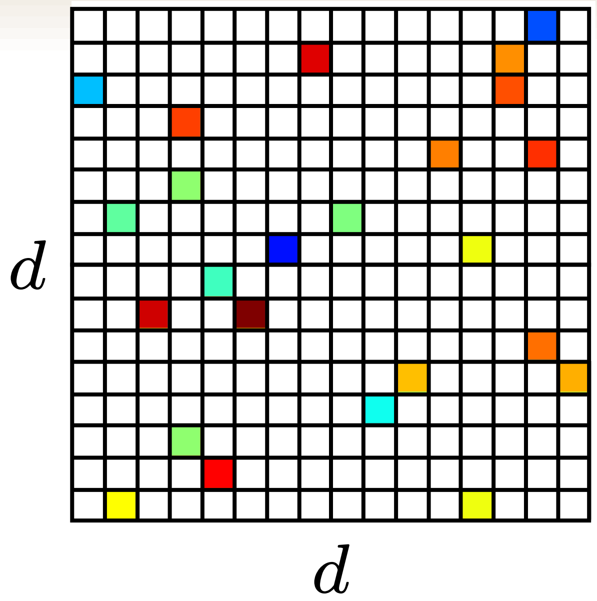
School of Electrical and Computer Engineering
Georgia Institute of Technology

Mary Wootters

Yaniv Plan

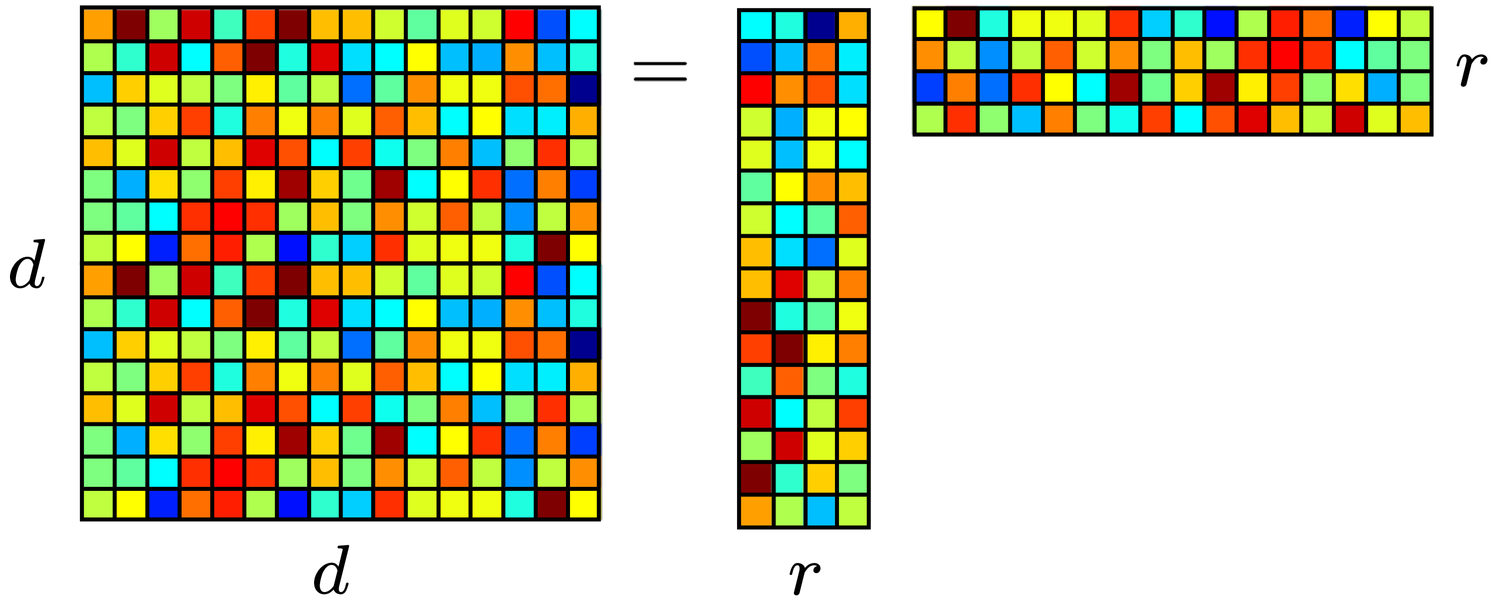
Ewout van den Berg

Matrix Completion



- When is it possible to recover the original matrix?
- How can we do this efficiently?
- How many samples will we need?

Low-Rank Matrices



Singular value decomposition:

$$M = U\Sigma V^* \quad \Rightarrow \quad \approx dr \ll d^2$$

degrees of freedom

Low-Rank Matrix Recovery

Given:

- a $d \times d$ matrix M of rank r
- samples of M on the set : $Y = M$

How can we recover M?

$$\widehat{M} = \arg\inf_{X : X \ \ = Y} \operatorname{rank}(X)$$

Can we replace this with something computationally feasible?

Nuclear Norm Minimization

Convex relaxation!

Replace $\mathrm{rank}(X)$ with $\|X\|_* = \sum_{j=1}^{d} |\sigma_j|$

$$\widehat{M} = \underset{X : X \quad = Y}{\arg\inf} \|X\|_*$$

If $| \quad | = O(r\, d \log d)$, this procedure can recover M !

Applications

- Collaborative Filtering (aka the "Netflix Problem")
- Recovery of incomplete survey data
- Analysis of voting data
- Sensor localization
- Quantum state tomography
- …

Matrix Completion in Practice

- Noise

$$Y = (M + Z)$$

- ***Quantization***
  - Netflix: Ratings are integers between 1 and 5
  - Survey responses: True/False, Yes/No, Agree/Disagree
  - Voting data: Yea/Nay
  - Quantum state tomography: Binary outcomes

Extreme quantization *destroys low-rank structure*

1-Bit Matrix Completion

Extreme case

$$Y = \text{sign}(M \quad)$$

Claim: Recovering M from Y is impossible!

$$M = \begin{bmatrix} \lambda & \lambda & \lambda & \lambda \\ \lambda & \lambda & \lambda & \lambda \\ \lambda & \lambda & \lambda & \lambda \\ \lambda & \lambda & \lambda & \lambda \end{bmatrix}$$

No matter how many samples we obtain, all we can learn is whether $\lambda > 0$ or $\lambda < 0$

Is There Any Hope?

If we consider a noisy version of the problem, recovery becomes feasible!

$$Y = \mathrm{sign}(M \quad + Z \quad)$$

$$M + Z = \begin{bmatrix} \lambda + Z_{1,1} & \lambda + Z_{1,2} & \lambda + Z_{1,3} & \lambda + Z_{1,4} \\ \lambda + Z_{2,1} & \lambda + Z_{2,2} & \lambda + Z_{2,3} & \lambda + Z_{2,4} \\ \lambda + Z_{3,1} & \lambda + Z_{3,2} & \lambda + Z_{3,3} & \lambda + Z_{3,4} \\ \lambda + Z_{4,1} & \lambda + Z_{4,2} & \lambda + Z_{4,3} & \lambda + Z_{4,4} \end{bmatrix}$$

Fraction of positive/negative observations tells us something about λ

Example of the power of *dithering*

Observation Model

For $(i, j) \in$ we observe

$$Y_{i,j} = \begin{cases} +1 & \text{with probability } f(M_{i,j}) \\ -1 & \text{with probability } 1 - f(M_{i,j}) \end{cases}$$

If f behaves like a CDF, then this is equivalent to

$$Y_{i,j} = \text{sign}(M_{i,j} + Z_{i,j})$$

where $Z_{i,j}$ is drawn according to a suitable distribution

We will assume that is drawn uniformly at random

Examples

- Logistic regression / Logistic noise

$$f(x) = \frac{e^x}{1+e^x}$$

$$Z_{i,j} \sim \text{logistic distribution}$$

- Probit regression / Gaussian noise

$$f(x) = \Phi(x/\sigma)$$

$$Z_{i,j} \sim \mathcal{N}(0, \sigma^2)$$

Maximum Likelihood Estimation

Log-likelihood function:

$$F(X) = \sum_{(i,j)\in \;_{+}} \log(f(X_{i,j})) + \sum_{(i,j)\in \;_{-}} \log(1 - f(X_{i,j}))$$

$$\widehat{M} = \arg\max_{X} F(X)$$
$$\text{s.t. } \frac{1}{d\alpha}\|X\|_* \leq \sqrt{r}$$
$$\|X\|_\infty \leq \alpha$$

Recovery of the Matrix

***Theorem** (Upper bound achieved by convex ML estimator)*

Assume that $\frac{1}{d\alpha}\|M\|_* \leq \sqrt{r}$ and $\|M\|_\infty \leq \alpha$. If is chosen at random with $\mathbb{E}|\quad| = m > d \log d$, then with high probability

$$\frac{1}{d^2}\|\widehat{M} - M\|_F^2 \leq C\alpha L_\alpha \beta_\alpha \sqrt{\frac{rd}{m}}$$

where

$$L_\alpha := \sup_{|x|\leq\alpha} \frac{|f'(x)|}{f(x)(1-f(x))} \qquad \beta_\alpha := \sup_{|x|\leq\alpha} \frac{f(x)(1-f(x))}{(f'(x))^2}$$

Is this bound tight?

Recovery of the Matrix

***Theorem** (Upper bound achieved by convex ML estimator)*

Assume that $\frac{1}{d\alpha}\|M\|_* \leq \sqrt{r}$ and $\|M\|_\infty \leq \alpha$. If is chosen at random with $\mathbb{E}| \quad | = m > d \log d$, then with high probability

$$\frac{1}{d^2}\|\widehat{M} - M\|_F^2 \leq C\alpha L_\alpha \beta_\alpha \sqrt{\frac{rd}{m}}$$

***Theorem** (Lower bound on any estimator)*

For any recovery algorithm $\widehat{M}$ there exist M satisfying the assumptions above such that for any set with $| \quad | = m$, we have (under mild technical assumptions) that

$$\mathbb{E}\left[\frac{1}{d^2}\|\widehat{M} - M\|_F^2\right] \geq c\alpha\sqrt{\beta_{\frac{3}{4}\alpha}}\sqrt{\frac{rd}{m}}$$

Proof Outline

- Construct a set $\mathcal{X}$ of low-rank, bounded matrices with the properties that
  - $|\mathcal{X}|$ is large
  - For any $X_i, X_j \in \mathcal{X}$, $\|X_i - X_j\|_F$ is relatively large
- Apply Fano's inequality to show that given observations of a particular X_i, there is a lower bound on how well we can correctly identify the chosen X_i
- If we cannot identify the chosen X_i, then we cannot estimate it very accurately either
- Randomized construction of $\mathcal{X}$

Conclusions

- Lower bounds can also be stated for
  - How well can we recover low-rank matrices in the presence of Gaussian noise?
  - How well can we recover the *distribution* $f(M)$
- Quantized (especially 1-bit) matrix completion is difficult
  - Naïve approaches don't work well
  - We have algorithms that are near-optimal
  - Seems to work well in practice

Thank You!