

1-Bit Matrix Completion

Mark A. Davenport

School of Electrical and Computer Engineering
Georgia Institute of Technology

Yaniv Plan



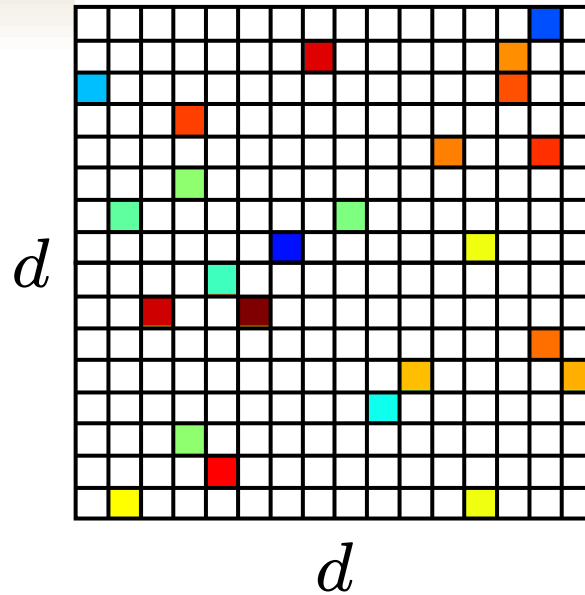
Mary Wootters



Ewout van den Berg

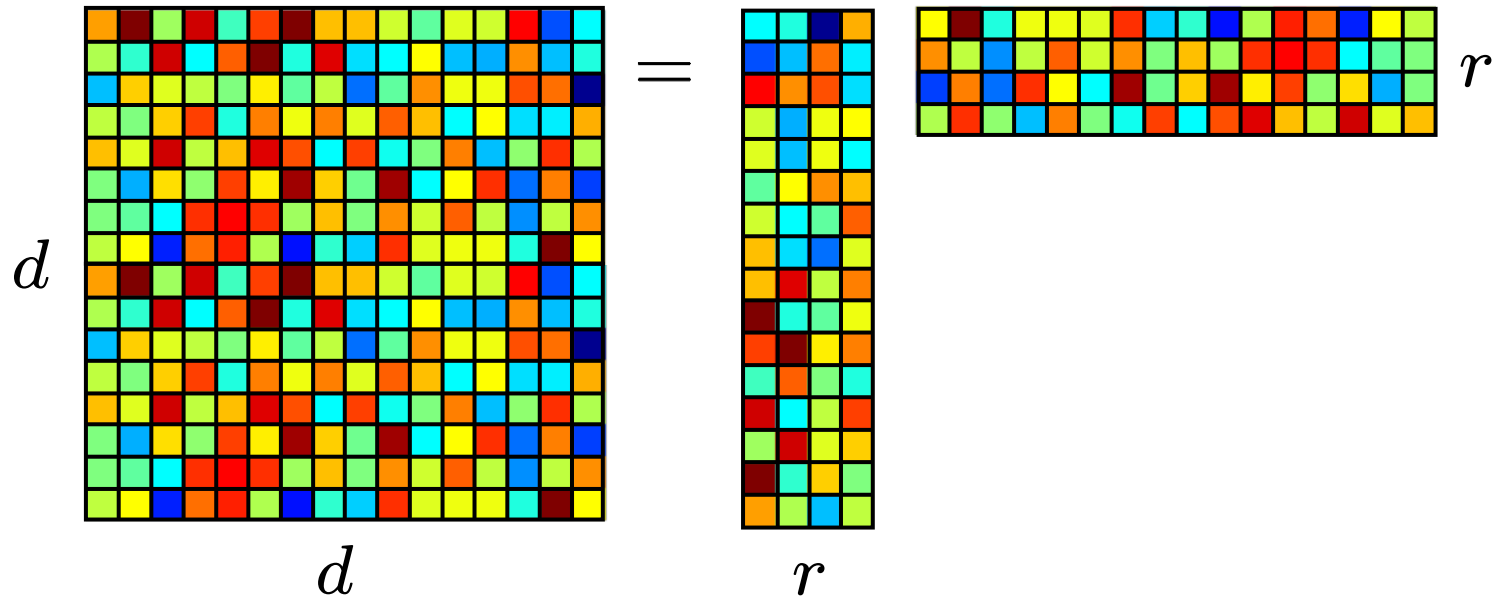


Matrix Completion



- 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:

$$M = U\Sigma V^*$$



$$\approx dr \ll d^2$$

degrees of freedom

Collaborative Filtering

The “Netflix Problem”

$M_{i,j}$ = how much user i likes movie j

Rank 1 model: u_i = how much user i likes romantic movies

v_j = amount of romance in movie j

$$M_{i,j} = u_i v_j$$

Rank 2 model: w_i = how much user i likes zombie movies

x_j = amount of zombies in movie j

$$M_{i,j} = u_i v_j + w_i x_j$$

LOVE
MEANS
NEVER
HAVING
TO SAY
YOU'RE
UNDEAD



WARM BODIES

PG-13 PARENTS STRONGLY CAUTIONED
SOME MATERIAL MAY BE INAPPROPRIATE FOR CHILDREN UNDER 13
ZOMBIE VIOLENCE AND SOME LANGUAGE

FEBRUARY 1 #WARMBODIES

Quebec



Beyond Netflix

- Recovery of incomplete survey data
- Analysis of voting data
- Sensor localization
- Student response data
- Quantum state tomography
- ...

Low-Rank Matrix Recovery

Given:

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

How can we recover M ?

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

Can we replace this with something computationally feasible?

Nuclear Norm Minimization

Convex relaxation!

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

$$\widehat{M} = \arg \inf_{X: X = Y} \|X\|_*$$

If $\|Y\|_* = O(r d \log d)$, under certain assumptions, this procedure can recover M !

Matrix Completion in Practice

- Noise

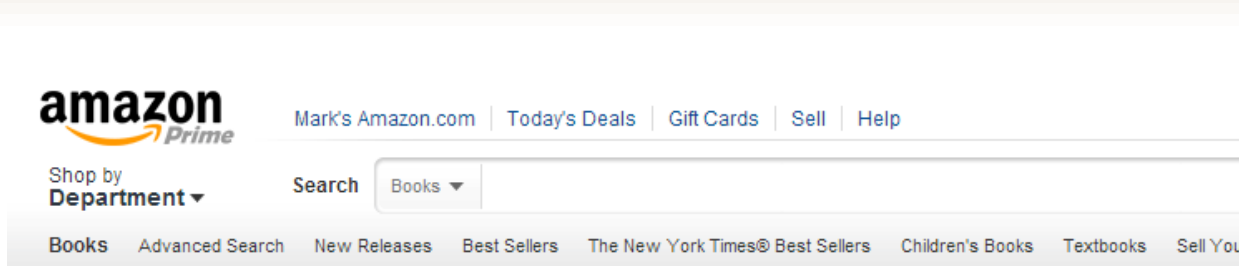
$$Y = (M + Z)$$

- ***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*

What's the Problem?



amazon Prime

Mark's Amazon.com | Today's Deals | Gift Cards | Sell | Help

Shop by Department

Search Books

Books | Advanced Search | New Releases | Best Sellers | The New York Times® Best Sellers | Children's Books | Textbooks | Sell Your Books

Customer Reviews

Linear Operator Theory in Engineering and Science (Applied Mathematical Sciences)

4 Reviews

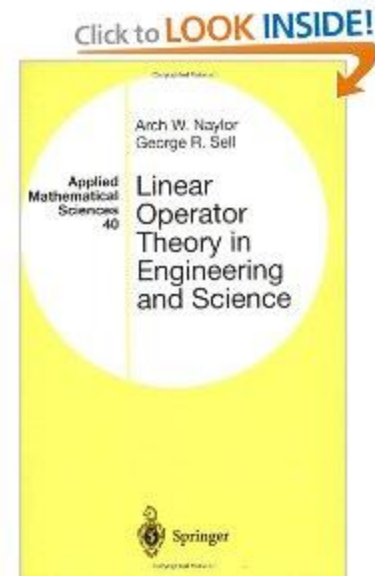
- 5 star: (4)
- 4 star: (0)
- 3 star: (0)
- 2 star: (0)
- 1 star: (0)

Average Customer Review

★★★★★ (4 customer reviews)

Share your thoughts with other customers

Create your own review



★★★★★ i would give 6 stars!, February 26, 2012

By [Chris](#) See all my reviews

Amazon Verified Purchase (What's this?)

This review is from: **Linear Operator Theory in Engineering and Science (Applied Mathematical Sciences) (Paperback)**

I'm doing a PhD in econometrics and I need to apply operator theories in constructing a linear or nonlinear operator to help explain individual economic behaviour. This book contains numerous useful ideas and applications with exercises thoroughly designed; one of the questions in the exercise gave me an idea of creating a matrix for describing a nonlinear operator. That question asks for a matrix that describes a second order differential operator and that gave me an idea that Taylor series approximation can be used to linearise a nonlinear operator and hence a nonlinear operator may also be described by a matrix.

Help other customers find the most helpful reviews

Was this review helpful to you?

[Report abuse](#) | [Permalink](#)

1-Bit Matrix Completion

Extreme case

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

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!

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

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

Fraction of positive/negative observations tells us something about λ

Example of the power of *dithering*

Observation Model

For $(i, j) \in \mathcal{I}$ 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 $Z_{i,j}$ is drawn uniformly at random

Examples

- Logistic regression / Logistic noise

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

$Z_{i,j} \sim$ 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:

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

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

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 \hat{M} is chosen at random with $\mathbb{E}|\hat{M}_{ij}| = m > d \log d$, then with high probability

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

where

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

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 \mathcal{S} is chosen at random with $|\mathcal{S}| = m > d \log d$, 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)

There exist M satisfying the assumptions above such that for any set \mathcal{S} with $|\mathcal{S}| = m$, we have (under mild technical assumptions) that

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

Logistic Model

$$L_\alpha = 1 \quad \beta_\alpha \approx e^\alpha$$

Theorem (Upper bound achieved by convex ML estimator)

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

Theorem (Lower bound on any estimator)

$$\inf_{\widehat{M}} \mathbb{E} \left[\frac{1}{d^2} \|\widehat{M} - M\|_F^2 \right] \geq c\alpha e^{\frac{3}{8}\alpha} \sqrt{\frac{rd}{m}}$$

Probit Model

$$L_\alpha \approx \frac{\frac{\alpha}{\sigma} + 1}{\sigma} \quad \beta_\alpha \approx \sigma^2 e^{\alpha^2/2\sigma^2}$$

Theorem (Upper bound achieved by convex ML estimator)

$$\frac{1}{d^2} \|\widehat{M} - M\|_F^2 \leq C \left(\frac{\alpha}{\sigma} + 1 \right) e^{\alpha^2/2\sigma^2} \sigma \alpha \sqrt{\frac{rd}{m}}$$

Two regimes

- High signal-to-noise ratio: $\sigma \leq \alpha$
- Low signal-to-noise ratio: $\sigma \geq \alpha$

Compare to how well we can estimate M from unquantized, noisy measurements

Probit Model (High SNR)

Theorem (Upper bound achieved by convex ML estimator)

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

Theorem (Lower bound on any estimator with unquantized measurements)

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

Probit Model (Low SNR)

Theorem (Upper bound achieved by convex ML estimator)

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

Theorem (Lower bound on any estimator with unquantized measurements)

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

More noise can lead to *improved* performance!

Implementation

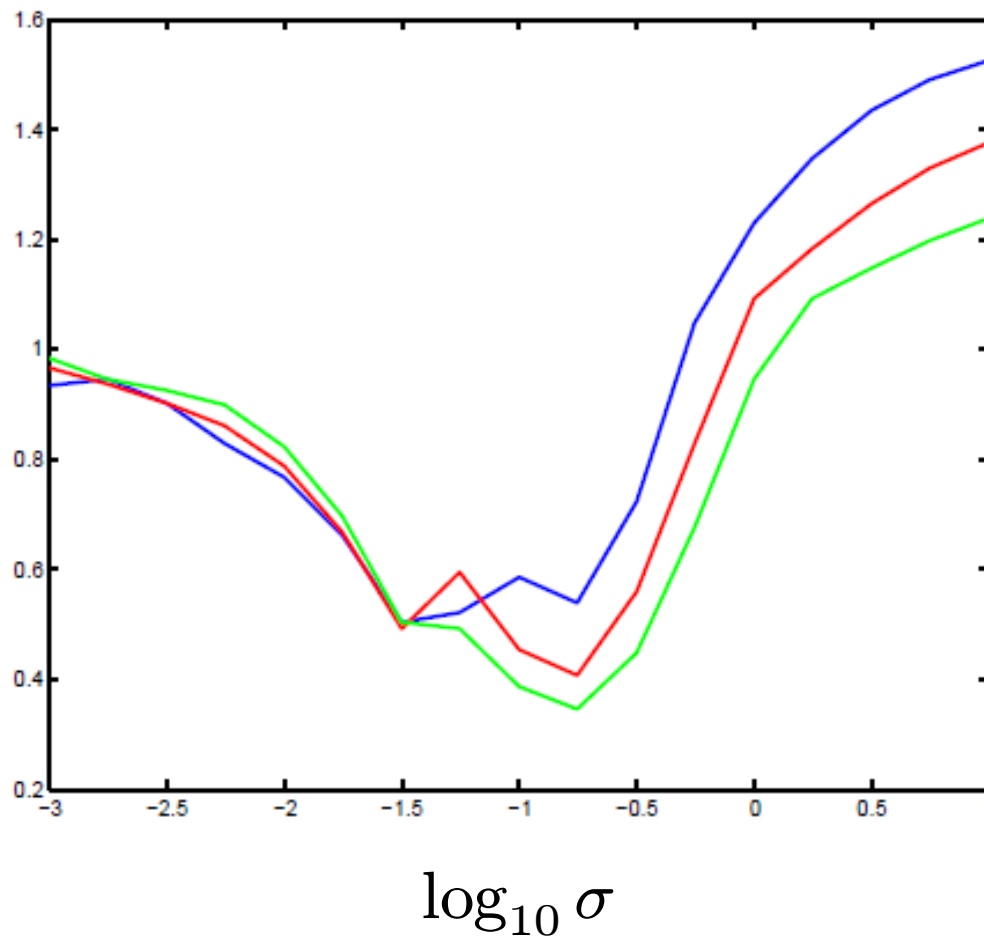
$$\underset{x}{\text{minimize}} \quad f(x) \quad \text{subject to} \quad x \in \mathcal{C}$$

- $f(x)$ is smooth, convex
- \mathcal{C} is a closed, convex set
- Nonmonotone spectral projected-gradient (SPG) algorithm
- Iterative algorithm, each iteration requires computation of
 - $f(x)$
 - $\nabla f(x)$

Synthetic Simulations

$$d = 500 \quad m = .15d^2$$

$$\frac{\|\widehat{M} - M\|_F}{\|M\|_F}$$



$$r = 5$$

$$r = 3$$

$$r = 2$$

MovieLens Data Set

- 100,000 movie ratings on a scale from 1 to 5
- Convert to binary outcomes by comparing each rating to the average rating in the data set
- Evaluate by checking if we predict the correct sign
- Training on 95,000 ratings and testing on remainder
 - “standard” matrix completion: 60% accuracy

1: 64% 2: 56% 3: 44% 4: 65% 5: 74%

- 1-bit matrix completion: 73% accuracy

1: 79% 2: 73% 3: 58% 4: 75% 5: 89%

Conclusions

- 1-bit matrix completion is hard!
- What did you really expect?
- Sometimes 1-bit is all we can get...
- We have algorithms that are near optimal
- Open questions
 - Are there simpler/better/faster/stronger algorithms?
 - What about 2.32-bit matrix completion?

Thank You!