

Tail bounds for Matrix De-noising application

Kevin Kim¹, Dong Hu¹, Georgios Mavroudeas¹

¹Department of Computer Science
Rensselaer Polytechnic Institute

December, 2019

Table of Contents

- 1 Motivation & Background
- 2 Problem & Modeling
- 3 From Random Variables to Random Matrices
- 4 Simulation & Results
- 5 Conclusion

Table of Contents

- 1 Motivation & Background
- 2 Problem & Modeling
- 3 From Random Variables to Random Matrices
- 4 Simulation & Results
- 5 Conclusion

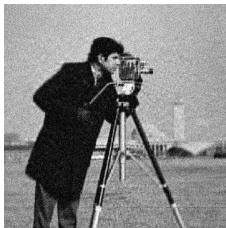
- **Overview:** Matrix de-noising is one of the popular topics in Machine Learning. Many existing papers have focused on the assumption that the data matrix has low rank. In this presentation we want to focus on one specific de-noising algorithm on low rank matrices and analyze it from the probability perspective.

Motivation II

One important application of matrix de-noising is image de-noising



(a) Ground Truth



(b) Noisy Input($\sigma = 30$)



(c) De-noised Image

Figure: Example of image de-noising

Table of Contents

- 1 Motivation & Background
- 2 Problem & Modeling
- 3 From Random Variables to Random Matrices
- 4 Simulation & Results
- 5 Conclusion

Problem Formulation I

Oracle: We can sample any matrix with noise where the noise has distribution $\mathcal{N}(0, \sigma^2)$ for some σ .

Problem: Given a low-rank matrix $\mathbf{A}^{m \times n}$ with $rk(\mathbf{A}) = r \ll \min\{m, n\}$, we infect it with Gaussian noise and we get a noisy observation of the matrix.

$$\mathbf{C} = \mathbf{A} + \mathbf{G}$$

where $[\mathbf{G}^{m \times n}]_{i,j} \sim \mathcal{N}(0, \sigma^2)$.

Our goal is to recover our initial matrix \mathbf{A} from the noisy counterpart \mathbf{C} .

Problem Formulation II

We pick linear operator to recover the matrix, and we can get our estimator $\hat{\mathbf{A}} = \mathbf{C}\mathbf{X}$ for some linear operator \mathbf{X} . Then finding the best estimator involves solving the optimization problem

$$\begin{aligned} \min_{\mathbf{X}} \quad & \mathbb{E} \left\| \hat{\mathbf{A}} - \mathbf{A} \right\|_F^2 \\ \text{s.t.} \quad & \hat{\mathbf{A}} = \mathbf{C}\mathbf{X} \\ & rk(\hat{\mathbf{A}}) \leq r \end{aligned}$$

Problem Solution

It can be shown that

$$\mathbf{X}^* = \arg \min_{\mathbf{X}} \mathbb{E} \left\| \hat{\mathbf{A}} - \mathbf{A} \right\|_F^2 \quad (1)$$

$$= (\mathbf{A}^\top \mathbf{A} + \sigma^2 m \mathbf{I})^{-\frac{1}{2}} \left[(\mathbf{A}^\top \mathbf{A} + \sigma^2 m \mathbf{I})^{-\frac{1}{2}} \mathbf{A}^\top \mathbf{A} \right]_r \quad (2)$$

where $[\mathbf{M}]_r = \mathbf{U}_r \mathbf{\Sigma}_r \mathbf{V}_r^\top$ is the rank- r truncation of \mathbf{M} if $\mathbf{M} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^\top$ is its SVD.

Optimal Solution

Can We Really Use it?

Notice that the optimal matrix \mathbf{X}^* , involves \mathbf{A} in the formula.

$$\mathbf{X}^* = (\mathbf{A}^\top \mathbf{A} + \sigma^2 m \mathbf{I})^{-\frac{1}{2}} \left[(\mathbf{A}^\top \mathbf{A} + \sigma^2 m \mathbf{I})^{-\frac{1}{2}} \mathbf{A}^\top \mathbf{A} \right]_r$$

WARNING: We are not allowed to use \mathbf{A} in our expression!!!

Proposed Solution

Observation:

$$\mathbb{E} \left[\mathbf{C}^\top \mathbf{C} \right] = \mathbf{A}^\top \mathbf{A} + \sigma^2 m \mathbf{I}$$

Proposed linear operator:

$$\tilde{\mathbf{X}} = (\mathbf{C}^\top \mathbf{C})^{-\frac{1}{2}} \left[(\mathbf{C}^\top \mathbf{C})^{-\frac{1}{2}} (\mathbf{C}^\top \mathbf{C} - \sigma^2 m \mathbf{I}) \right]_r$$

We use $\tilde{\mathbf{A}} = \mathbf{C} \tilde{\mathbf{X}}$ to approximate $\hat{\mathbf{A}} = \mathbf{C} \mathbf{X}^*$.

Challenge: Is this estimation accurate?

Probability to the Rescue

Bounding our Solution

- The main challenge is to find probability tail bounds that shows our solution concentrates around and is unlikely to deviate far from the optimal solution.
- Namely, we want to show

$$\mathbb{E} \left\| \mathbf{C} \tilde{\mathbf{X}} - \mathbf{C} \mathbf{X}^* \right\|_2 < \varepsilon$$

w.h.p for some small ε .

- Notice

$$\mathbb{E} \left\| \mathbf{C} \tilde{\mathbf{X}} - \mathbf{C} \hat{\mathbf{X}} \right\|_2 \leq \mathbb{E} \left[\left\| \mathbf{C} \right\|_2 \left\| \tilde{\mathbf{X}} - \mathbf{X}^* \right\|_2 \right]$$

Table of Contents

- 1 Motivation & Background
- 2 Problem & Modeling
- 3 From Random Variables to Random Matrices**
- 4 Simulation & Results
- 5 Conclusion

Main Theory I

Probability Tools Overview

Here some of the existing probability tools will be presented for scalar random variables.

- Sub-Gaussian Bounds:

- A random variable X is sub-Gaussian if it satisfies: $\forall \lambda \in \mathbb{R}$

$$\mathbb{E} \left[e^{\lambda(X-\mu)} \right] \leq e^{\frac{\sigma^2 \lambda^2}{2}}$$

and the concentration bound is:

$$\mathbb{P} [|X - \mu| \geq t] \leq 2e^{-\frac{t^2}{2\sigma^2}}$$

- Sub-exponential Bounds:

- A random variable X is sub-exponential if it satisfies: $\exists v, \alpha > 0$, such that $\forall |\lambda| < \frac{1}{\alpha}$

$$\mathbb{E} \left[e^{\lambda(X-\mu)} \right] \leq e^{\frac{v^2 \lambda^2}{2}}$$

and it's a generalization of the sub-Gaussian random variables. The sub-exponential tail bound can be formulated as

$$\mathbb{P}[X - \mu \geq t] \leq \begin{cases} e^{-\frac{t^2}{2v^2}} & \text{if } 0 \leq t \leq \frac{v^2}{\alpha} \\ e^{-\frac{t}{2\alpha}} & \text{for } t \geq \frac{v^2}{\alpha} \end{cases}$$

- Bernstein Bounds:

- Can be applied to sum of series of centered bounded i.i.d random variables. Let S_1, \dots, S_n be random variables with $\mathbb{E}[S_i] = 0$ and $|S_i| \leq L$ and let

$$Z = \sum_{i=1}^n S_i, \quad \text{Var}(Z) = \sum_{i=1}^n \text{Var}(S_i)$$

then we have

$$\mathbb{P}(|Z| \geq t) \leq 2e^{-\frac{t^2/2}{\text{Var}(Z) + Lt/3}}$$

Matrix Concentration Inequalities I

From Scalar Random Variables to Random Matrices

- In general, it is hard to extend the bounds derived for scalar random variables to random matrices. In our case if we can represent our random matrices by a sum of bounded i.i.d random matrices, we can apply the matrix Bernstein bound.

Matrix Concentration Inequalities II

From Scalar Random Variables to Random Matrices

Matrix Bernstein bound If $\mathbf{S}_1, \dots, \mathbf{S}_n$ are centered i.i.d random matrices such that $\forall i, \mathbf{S}_i \in \mathbb{R}^{d_1 \times d_2}$ and $\|\mathbf{S}_i\| \leq L$ and

$$\mathbf{Z} = \sum_{i=1}^n \mathbf{S}_i$$

If we denote $V(\mathbf{Z})$ as the co-variance matrix where:

$$V(\mathbf{Z}) = \max\{\|\mathbb{E}[\mathbf{Z}^* \mathbf{Z}]\|, \|\mathbb{E}[\mathbf{Z} \mathbf{Z}^*]\|\}$$

Then the concentration bound is:

$$\mathbb{P}(\|\mathbf{Z}\| \geq t) \leq (d_1 + d_2) e^{-\frac{t^2/2}{V(\mathbf{Z}) + Lt/3}}$$

Also we can get a bound for the expectation of \mathbf{Z} :

$$\mathbb{E}[\|\mathbf{Z}\|] \leq \sqrt{2V(\mathbf{Z}) \log(d_1 + d_2)} + \frac{1}{3}L \log(d_1 + d_2)$$

Using Random Matrix Bounds I

Our Attempt to Bound the Solution

We can simplify the optimal and the proposed solution to

$$\begin{aligned}\mathbf{X}^* &= [\mathbf{I} - \sigma^2 m (\mathbf{A}^\top \mathbf{A} + \sigma^2 m \mathbf{I})^{-1}]_r \\ \tilde{\mathbf{X}} &= [\mathbf{I} - \sigma^2 m (\mathbf{C}^\top \mathbf{C})^{-1}]_r\end{aligned}$$

So for $\tilde{\mathbf{X}}$ to be concentrating around \mathbf{X}^* it has to that $\mathbf{C}^\top \mathbf{C}$ concentrates around $\mathbf{A}^\top \mathbf{A} + \sigma^2 m \mathbf{I}$

Using Random Matrix Bounds II

Our Attempt to Bound the Solution

We have

$$\begin{aligned} & \left\| \mathbf{C}^\top \mathbf{C} - (\mathbf{A}^\top \mathbf{A} + \sigma^2 m \mathbf{I}) \right\| \\ &= \left\| \mathbf{A}^\top \mathbf{A} + \mathbf{G}^\top \mathbf{A} + \mathbf{G} \mathbf{A}^\top + \mathbf{G}^\top \mathbf{G} - \mathbf{A}^\top \mathbf{A} - \sigma^2 m \mathbf{I} \right\| \\ &\leq 2 \|\mathbf{A}\| \|\mathbf{G}\| + \left\| \mathbf{G}^\top \mathbf{G} - \sigma^2 m \mathbf{I} \right\| \\ &\leq 2 \|\mathbf{A}\| \|\mathbf{G}\| + \|\mathbf{G}\|^2 + \sigma^2 m \\ &\leq 2 \|\mathbf{A}\| (\sqrt{n} + \sqrt{t} + \sqrt{m}) + (\sqrt{n} + \sqrt{t} + \sqrt{m})^2 + \sigma^2 m \\ &= (\sqrt{n} + \sqrt{t} + \sqrt{m})(2 \|\mathbf{A}\| + \sqrt{n} + \sqrt{t} + \sqrt{m}) + \sigma^2 m \end{aligned}$$

with probability at least $1 - e^{-\frac{t^2}{2}}$

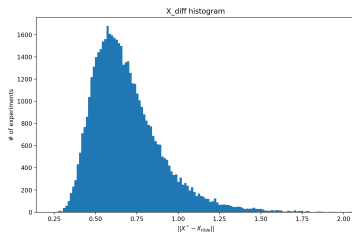
Table of Contents

- 1 Motivation & Background
- 2 Problem & Modeling
- 3 From Random Variables to Random Matrices
- 4 Simulation & Results**
- 5 Conclusion

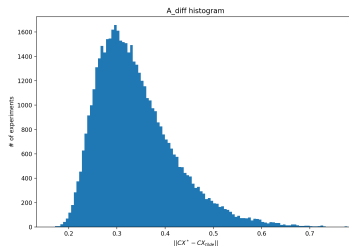
Empirical Results

Experiments

Experiment settings: $\mathbf{A} \sim \mathcal{N}(5, 1)$, $\mathbf{G} \sim \mathcal{N}(0, 0.01)$. We sample \mathbf{A} for 50000 times and plot the histogram for both $\|\mathbf{X}^* - \tilde{\mathbf{X}}\|_2$ and $\|\mathbf{C}\mathbf{X}^* - \mathbf{C}\tilde{\mathbf{X}}\|_2$



(a) $\|\mathbf{X}^* - \tilde{\mathbf{X}}\|_2$



(b) $\|\mathbf{C}\mathbf{X}^* - \mathbf{C}\tilde{\mathbf{X}}\|_2$

Table of Contents

- 1 Motivation & Background
- 2 Problem & Modeling
- 3 From Random Variables to Random Matrices
- 4 Simulation & Results
- 5 Conclusion**

Conclusion

- Both the experiment and the theory shows that if $\sigma_g < \frac{\sigma_r(\mathbf{A})}{\sqrt{m}\|\mathbf{A}\|_2}$ we have

$$\mathbb{P}\left(\left\|\mathbf{C}\mathbf{X}^* - \mathbf{C}\tilde{\mathbf{X}}\right\|_2 \geq \varepsilon\right) \leq \delta$$

ε, δ and σ_g are unsure and need to be determined but there should be some according to theory