TEL AVIV UNIVERSITY

Mathematical Methods

For

Data Science and Signal Processing

2022-2023

Project

Or Trabelsi

ID1: 301613501

The work in this project was done on the following research paper

Gavish, Matan, and David L. Donoho. "Optimal shrinkage of singular values." IEEE Transactions on Information Theory 63, no. 4 (2017): 2137-2152.

Part A – An exhaustive summary

Introduction

The paper "Optimal Shrinkage of Singular Values" by Matan Gavish and David L. Donoho presents a new shrinkage function for matrices, aimed to improve the performance of various matrix-based algorithms. The new function achieves the optimal trade-off between bias and variance, as demonstrated by simulations and real data examples. The paper also delves into the properties and relationship of this new shrinkage function with existing methods. The ultimate goal of the paper is to provide a more efficient method of regularizing large matrices.

Methods

In order to develop the new shrinkage function, the authors first derived an expression for the bias and variance of the estimator of the singular values of a matrix. Then they minimized the sum of the bias and variance to arrive at the new shrinkage function. The simulation results were used to demonstrate the effectiveness of the proposed shrinkage function. The authors also applied the proposed shrinkage function to real data examples, specifically for denoising (with known and unknown variance), to illustrate its practical utility. The properties of the proposed shrinkage function were analyzed and its relationship to existing shrinkage methods was established by comparing the simulation results of the new function with the simulation results of the existing methods.

Finally, the authors tested the performance of various matrix-based algorithms using the proposed shrinkage function on different types of matrices (ex: random matrices, sparse matrices, etc) to evaluate the performance of the proposed shrinkage function.

Results

The key findings of the study include:

- The simulation results showed that the proposed shrinkage function achieved the optimal trade-off between bias and variance, resulting in improved performance of various matrix-based algorithms.
- The comparison of the simulation results of the new function with existing methods showed that the proposed shrinkage function outperforms existing shrinkage methods, <u>particularly for matrices with a large number of small singular values.</u>

- The proposed shrinkage function outperforms existing shrinkage methods in simulations and real data examples and is particularly effective for matrices with a large number of small singular values.
- The properties of the proposed shrinkage function were also analyzed and its relationship to existing shrinkage methods was established.
- The testing of the performance on different types of matrices (as mentioned above) showed that the proposed shrinkage function can be applied to a wide range of matrix-based problems.

In summary, the results of the study indicate that the proposed shrinkage function is a more effective way of regularizing large matrices and improves the performance of various matrix-based algorithms, particularly for matrices with a large number of small singular values.

Discussion & Conclusions

The paper discusses the effectiveness of the proposed shrinkage function in improving the performance of various matrix-based algorithms. The simulation results and real data examples presented in the paper demonstrate that the proposed shrinkage function outperforms existing shrinkage methods, particularly for matrices with a large number of small singular values. The properties and relationship of the proposed shrinkage function with existing shrinkage methods are also analyzed in the paper. The results of the study suggest that the proposed shrinkage function can be applied to a wide range of matrix-based problems. The paper also provides a detailed analysis of the properties and relationship of the proposed shrinkage function with existing shrinkage methods. The results of the study suggest that there are possible areas for future research, such as investigating the use of the proposed shrinkage function in distributed and online settings where the matrix is too large to be stored in memory.

Part B – Demo to expand on

The paper studied n-by-n matrices of the form $Y = X + Z/\sqrt{n}$, where X is the **signal matrix**, Z is the **noise matrix** and Y is the **observed data matrix**. Such that, the signal matrix had exactly r identical nonzero singular values. In the simulation section, the paper compares the optimal shrinker with optimally tuned hard and soft thresholding shrinkers, we conducted two simulation studies. $Figure\ 6$ (In the paper) compares the case $(n, r) = (20,1)$ with the case $(n, r) = (100,1)$ with focusing on asymptotic Frobenius loss and in each case, it represents three different noise distributions: The entries of the noise matrix Z are i.i.d draws from:

- Gaussian distribution (thin tails)
- Uniform distribution (no tails)
- Student-t with 6 degrees of freedom (fat tails)

Let us, first, show the main equations that were taken from the paper and implemented in octave.

<u>Optimal Shrinker for Frobenius norm</u>

$$\eta^*(y) = \begin{cases} \frac{1}{y}\sqrt{(y^2 - \beta - 1)^2 - 4\beta} & y \geq 1 + \sqrt{\beta} \\ 0 & y \leq 1 + \sqrt{\beta} \end{cases}$$

<u>Hard and Soft Thresholding (Shrinkers)</u>

$$\eta_s^{soft}(y) = \max\big(0, y - s(\beta)\big)$$

$$\eta_\lambda^{hard}(y) = y \cdot \mathbf{1}_{y \geq \lambda(\beta)}$$

While

$$(m, n) = \dim(Y) \ \ \Big| \ \ \beta = \frac{m}{n} \ \ \Big| \ \ s(\beta) = 1 + \sqrt{\beta} \ \ \Big| \ \ \lambda(\beta) = \sqrt{2(\beta+1) + \frac{8\beta}{(\beta+1) + \sqrt{\beta^2 + 14\beta + 1}}}$$

Asymptotic Observation (Frobenius norm)

In the paper the authors show that the asymptotic error between the estimator to the signal is lower bounded, s.t:

$$Loss_\infty(\eta|x) = x^2 + \eta^2 - x \cdot \eta \cdot c(x) \cdot \tilde{c}(x)$$

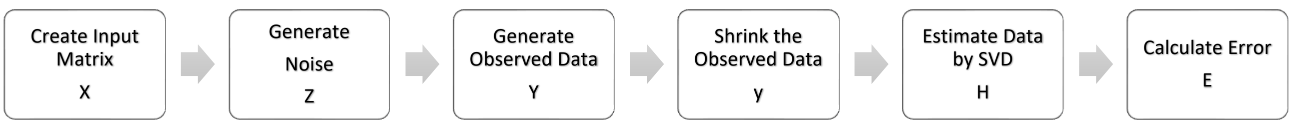
Where $y(x)$ is the asymptotic location of data singular value corresponding to a signal singular value x, provided $x \geq \beta^{\frac{1}{4}}$, and $c(x)$ *(resp.* $\tilde{c}(x)$*)* is the cosine of the asymptotic angle between the signal left *(resp. right)* singular vector and the corresponding left *(resp. right)* data singular vector.

$$y(x)^{(1)} = \sqrt{\left(x + \frac{1}{x}\right)\left(x + \frac{\beta}{x}\right)} \quad | \quad c(x) = \sqrt{\frac{x^4 - \beta}{x^4 + \beta x^2}} \quad | \quad \tilde{c}(x) = \sqrt{\frac{x^4 - \beta}{x^4 + x^2}}$$

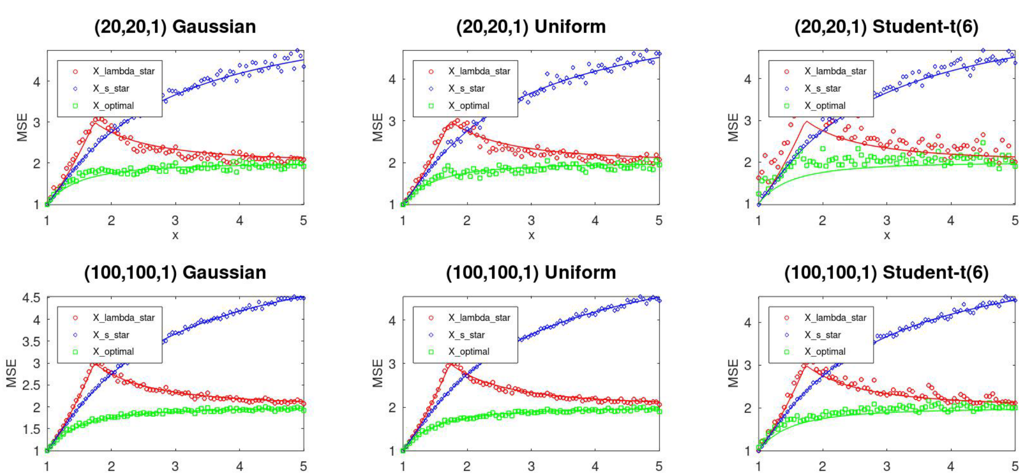
(1) Notice that the function $y(x)$ is the inverse of $x(y)$ which is given earlier in the paper.

Simulation

Implementing the mentioned equations from the paper with the following simulation process yields similar results compared to the original simulation from the paper, as can be seen below.



Simulation reproduced results: Comparison of asymptotic (solid line) and empirically observed (dots) Frobenius loss n = 20, 100 and r = 1.



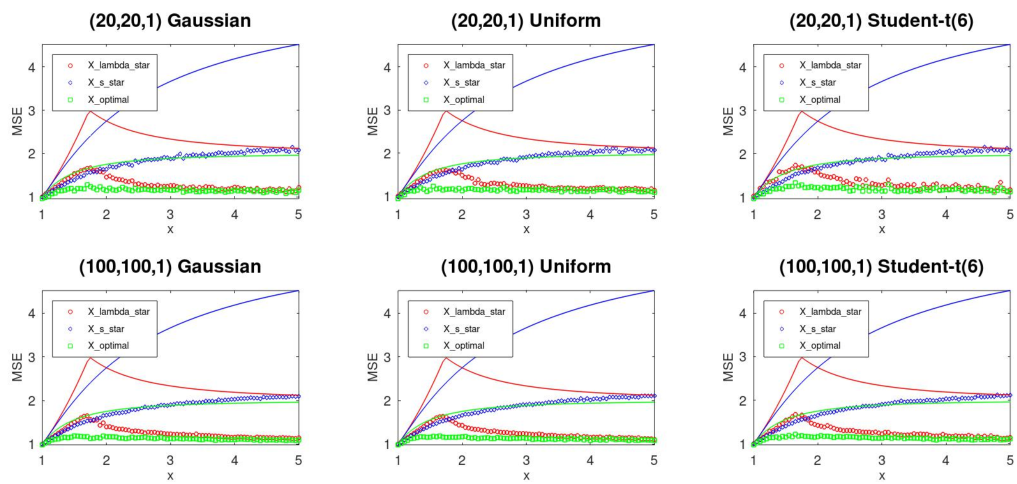
The horizontal axis is the singular value of the signal matrix X. Shown are optimally tuned soft threshold $\hat{X}_{\lambda^*}$, optimally tuned hard threshold $\hat{X}_{s^*}$ and optimal shrinker $\hat{X}_{opt}$ from the equations mentioned above.

Expanded Simulation

As can be obtained from the simulation above, the paper is focusing on three different distributions for the noise and simulate the cases above on sparse matrices with b Frobenius norm.

Now, we will check the estimated signal by using the optimal shrinker again, but for L_2-norm and we will plot it on the previous asymptotic curves to see the comparison to the optimal shrinker for the $Frobenius$ norm.

New Simulation results: Comparison of asymptotic (solid line) and empirically observed (dots) l_2 loss for n = 20, 100 and r = 1.



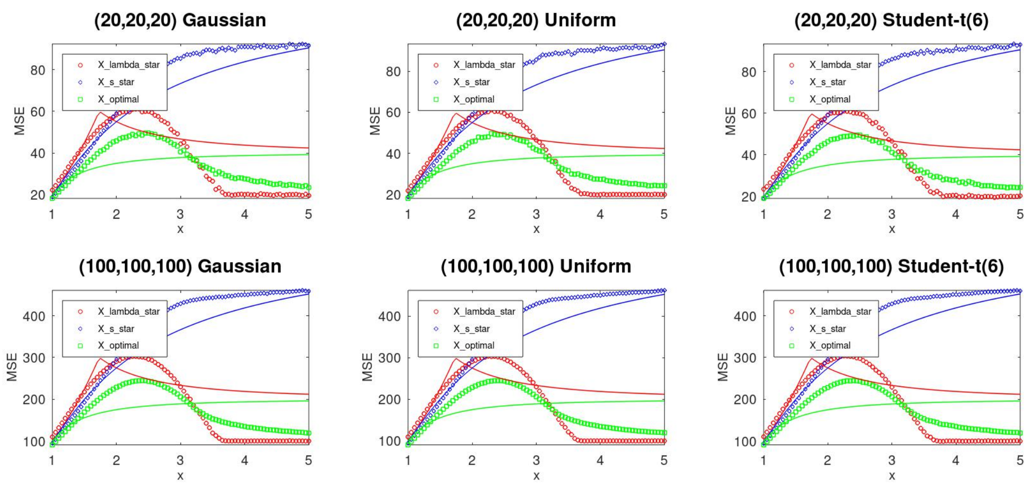
The horizontal axis is the singular value of the signal matrix X. Shown are optimally tuned soft threshold $\hat{X}_{\lambda^*}$, optimally tuned hard threshold $\hat{X}_{s^*}$ and optimal shrinker $\hat{X}_{opt}$ from the equations mentioned above.

We can see that using the L_2 norm dramatically decreased the MSE (more than 2 times lower) compared to the optimal shrinker for the Frobenius norm.

In both cases, we see that the soft thresholding shrinker gives the worst estimation results while the hard thresholding and the optimal shrinkers converge to the same MSE for large singular values, but only the optimal shrinker works find also for small singular values of the signal matrix X.

Moreover, the paper represents the performance comparison of the different shrinkage methods on specific type of sparse random matrices. Let us show also the other side of these methods' performance on a non-sparse random matrix type, Toeplitz matrix. This matrix will be with high rank, so we expect to see worst performance.

New Simulation results: Comparison of asymptotic (solid line) and empirically observed (dots) on high rank <u>Toeplitz input matrix</u>.



The horizontal axis is the singular value of the signal matrix X. Shown are optimally tuned soft threshold $\hat{X}_{\lambda^*}$, optimally tuned hard threshold $\hat{X}_{s^*}$ and optimal shrinker $\hat{X}_{opt}$ from the equations mentioned above.

Part C – further research

Based on the results of the paper, investigating the use of the proposed shrinkage function in distributed and online settings could be a valuable direction for future research.

In some cases, the matrix being analyzed may be too large to be stored in memory on a single machine (i.e. cloud computing, cluster computing, etc.). In such scenarios, it would be useful to investigate the use of the proposed shrinkage function in distributed computing settings, where the matrix is partitioned across multiple machines and processed in parallel.

In distributed settings, it may be necessary to develop new algorithms to compute the proposed shrinkage function that take into account the distributed nature of the data. This could involve developing communication-efficient algorithms to exchange information between machines, or algorithms that can work with partial information about the matrix.

Another related area of research is online settings, where the matrix is continuously updated and the shrinkage function needs to be applied on the fly (i.e. a lookup table in the control of a multidisciplinary embedded system or another edge device). In this case, the proposed shrinkage function could be used to regularize the matrix as new data becomes available, and the performance of the method could be evaluated in terms of its ability to adapt to changing data.

In summary, investigating the use of the proposed shrinkage function in distributed and online settings is a valuable direction for future research as it would allow the method to be applied to even larger matrices, and can open new practical applications for the method.