Advanced Image Processing HW4

Anilesh Bansal (22b0928) Parth Pujari (210100106)

1 Question 1

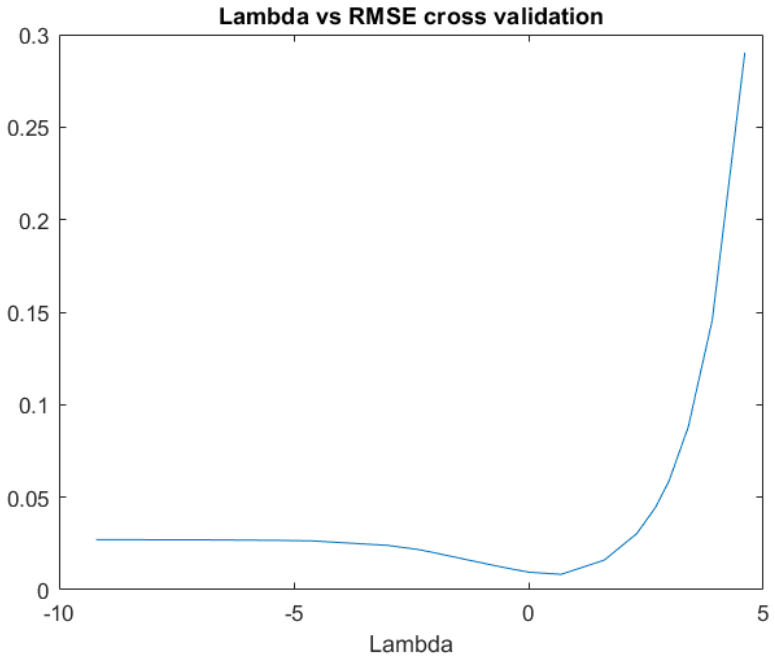

Figure 1: 5 sparse

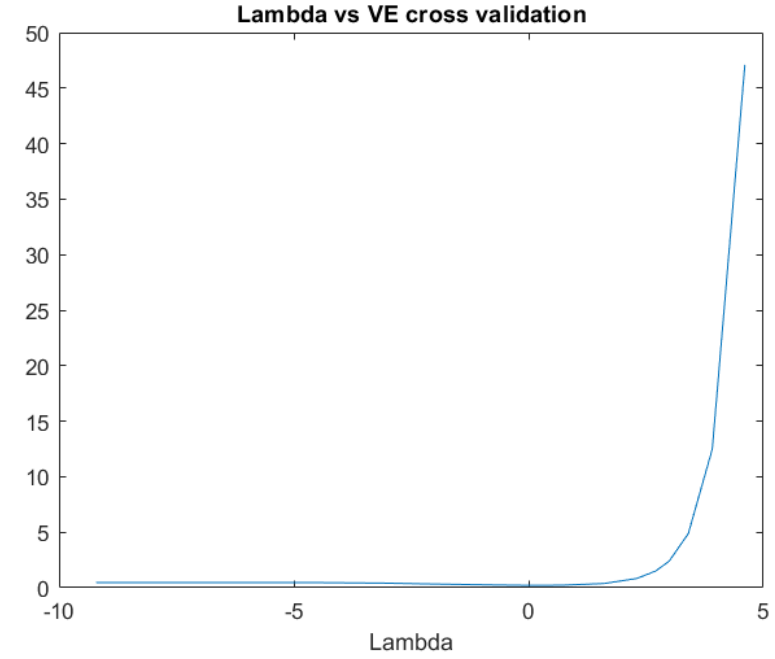

Figure 2: 5 sparse

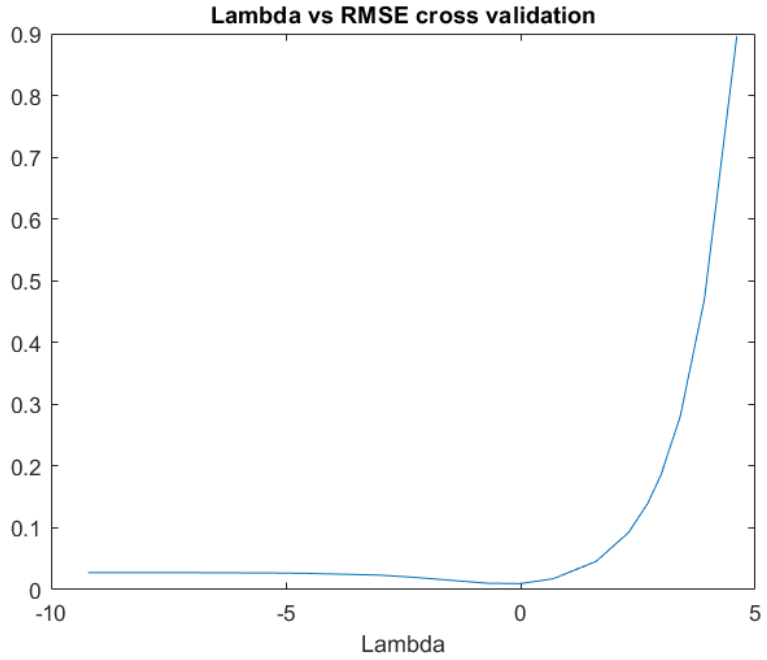

Figure 3: 10 sparse

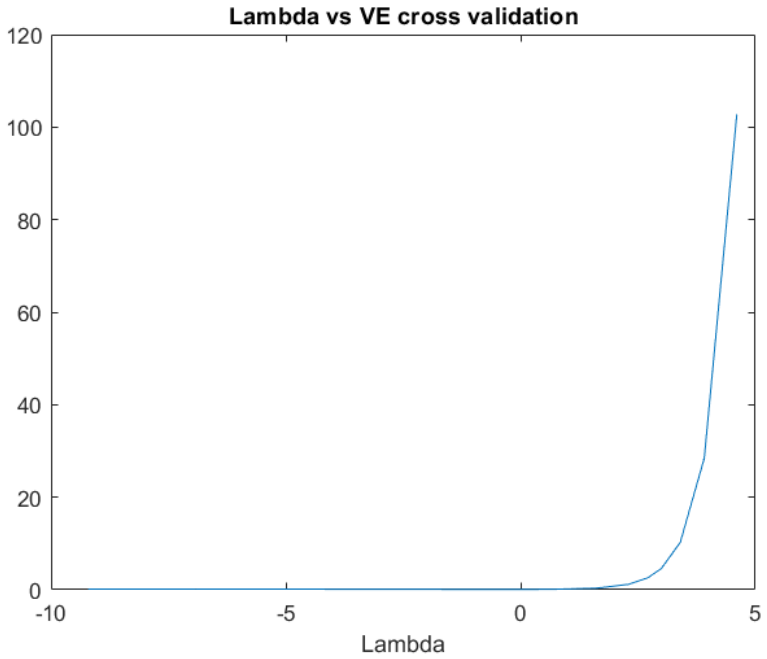

Figure 4: 10 sparse

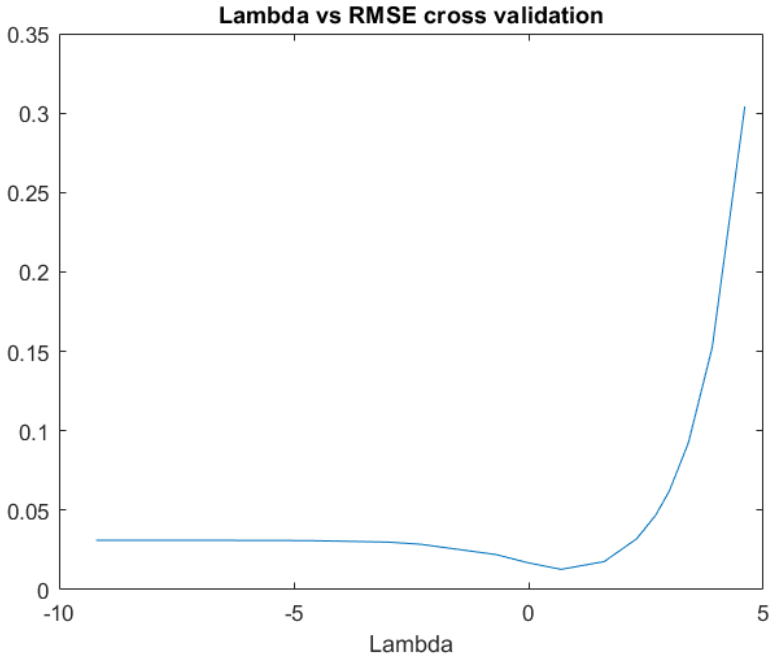

Figure 5: 15 sparse

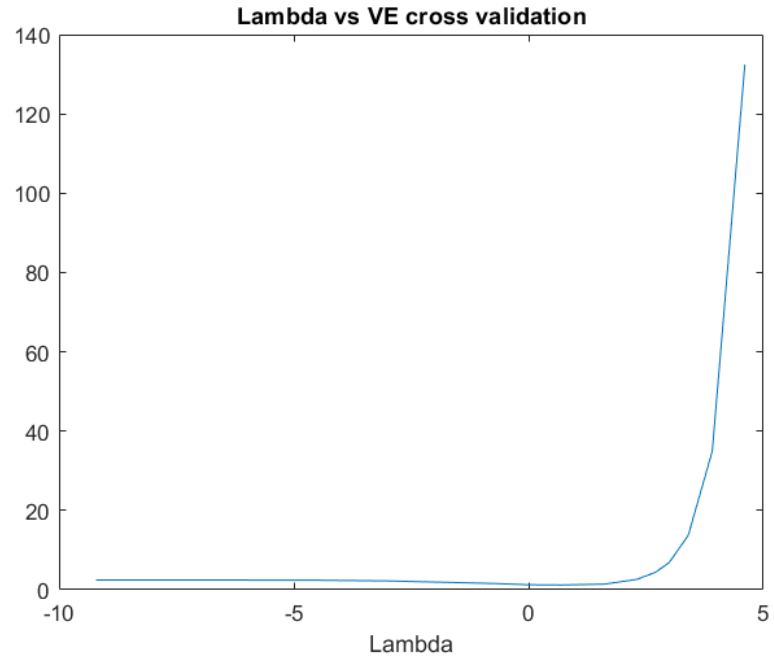

Figure 6: 15 sparse

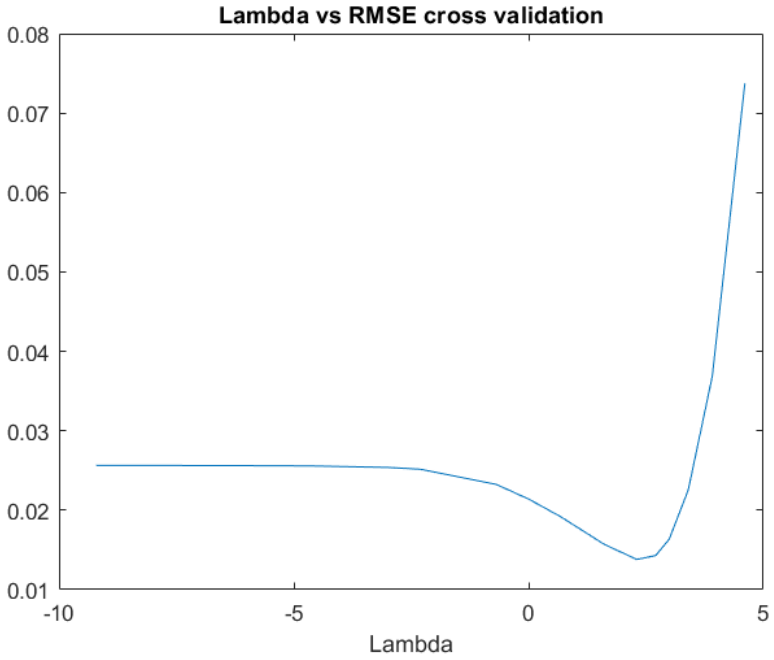

Figure 7: 20 sparse

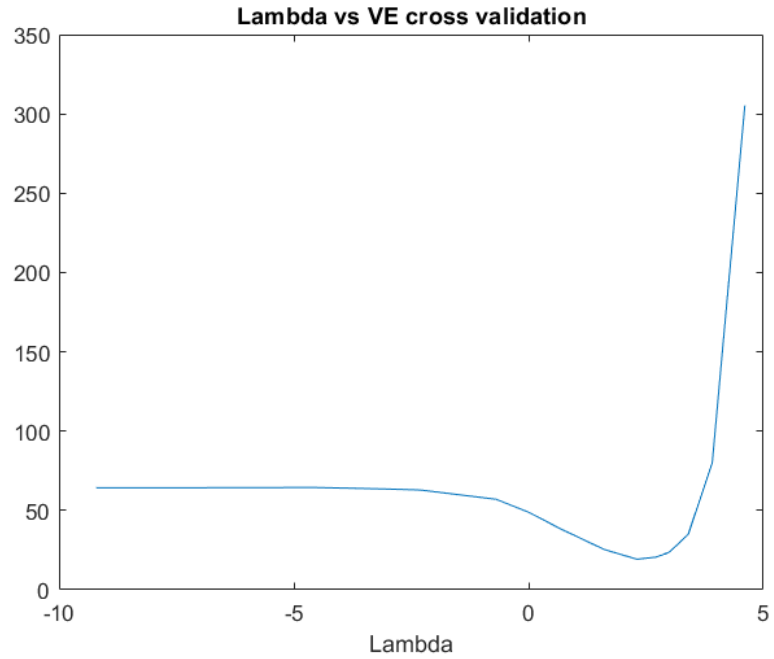

Figure 8: 20 sparse

a) The plots remain almost constant when $log(\lambda) < 0$ or $\lambda < 1$. At $\lambda = 1$, the value of RMSE error drops to a minima, and then starts to rise again with a very high slope. The two graphs obtained are very similar in their shape as well as the labellings. The minima of RMSE occurs at the same values of λ in both the graphs.

b) The validation error $VE = \frac{\sum_{i \in \mathcal{V}} (y_i - \Phi^i \hat{x_g})^2}{|\mathcal{V}|}$ where $\hat{x_g}$ is the minimizer to the loss $J = ||y - \Phi x||^2 + \lambda ||\theta||_1$

If the sets $\mathcal{R}$ and $\mathcal{V}$ are exactly the same, then the first term in the loss $||y - \Phi x||$ becomes η and is constant, and the validation error then is simply $\frac{1}{|\mathcal{V}|}||\eta||^2$ which would be same for all lambda, and would not give us any relevant information for which λ to choose for the least error $||\hat{x_g} - x||$.

This is similar in essence to Testing and Training Data in ML, we want the testing data to be different so that the model is not overfitting and falling in place with the actual observed results.

c) Theorem 1 in the linked paper mentiones this proxying ability. Particularly it states that provided m_{cv} is sufficiently large, the following holds with high probability

$$h(\lambda, +)\epsilon_{cv} - \sigma_n^2 \leq \epsilon_x \leq h(\lambda, -)\epsilon_{cv} - \sigma_n^2$$

where m_{cv} are the number of cross validation measurements, ϵ_{cv} is the cross validation error and ϵ_x is the actual residual error.

The difference of the upper and the lower bound

$$\frac{m}{m_{cv}} \frac{2\lambda\sqrt{2}}{\sqrt{m_{cv}} - \frac{2\lambda^2}{\sqrt{m_{cv}}}} \epsilon_{cv}$$

is roughly proportional to $1/m_{cv}^{2/3}$ and becomes tighter as m_{cv} increases.

d) Although the book mentions that $\lambda_N = 2\sigma\sqrt{\tau \frac{\log p}{N}}$ for some $\tau > 2$ works as a valid choice for the regulator parameter with very high probability, it doesn't give an exact value of what λ works best, just gives an estimate of the order of the magnitude and a lower bound on λ. Further the bound is theoretical and may not always hold well in all practical scanerios. Since the above cross validation method for choosing λ depends on some part of the data given itself, it ensures a better recovery than just the theoretical guarantee stated.

e) The advantages of using the Morozov's Discrepancy principle are that it doesn't require the splitting of data into reconstruction and validation sets, and thus is easier, effective and may result in better recovery in some cases, if not all.

However, it relies on knowing the noise model beforehand which may not always be the case. Further, it's implementation is highly dependent on the number of measurements m.

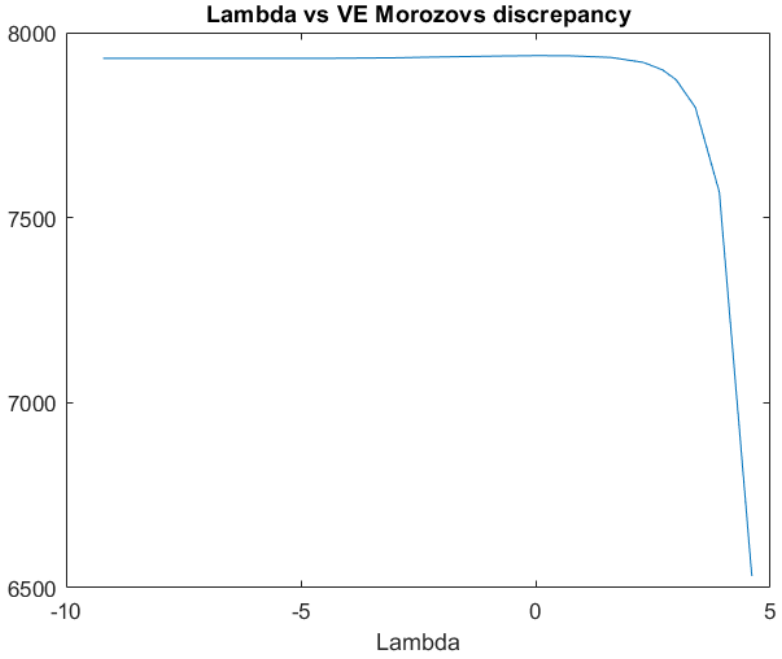

Figure 9: 5 sparse

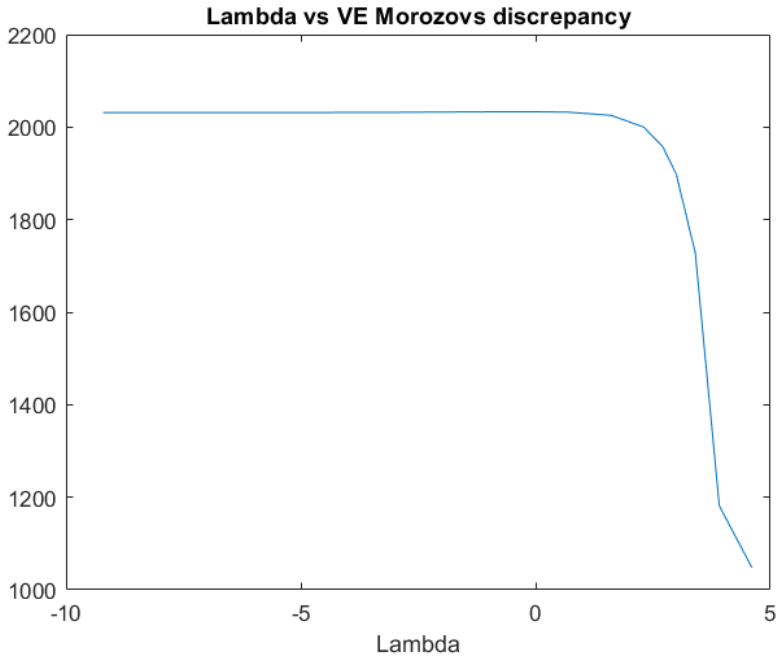

Figure 10: 10 sparse

2 Question 2

We observe that (Figure) for lower values of rank, the reconstruction is better as well as for higher values of f (retained pixels). These trends are as expected. Theorem 2 tells us that larger number of entries should be known for larger ranked matrices for recovery.

We use **Cross validation** to choose lambda by checking the reconstruction error of some pixels of the image and minimizing rmse of those using the same support matrix. We then use the median lambda for the results.

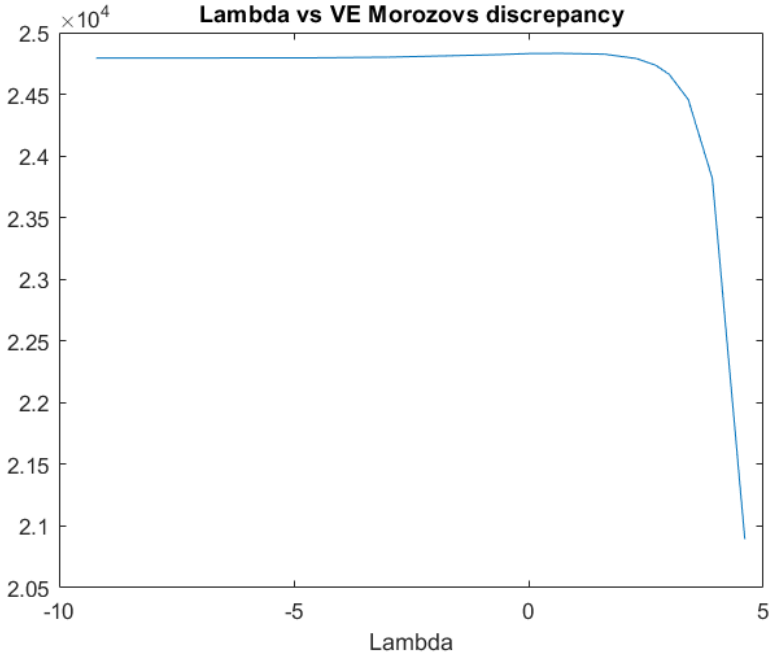

Figure 11: 15 sparse

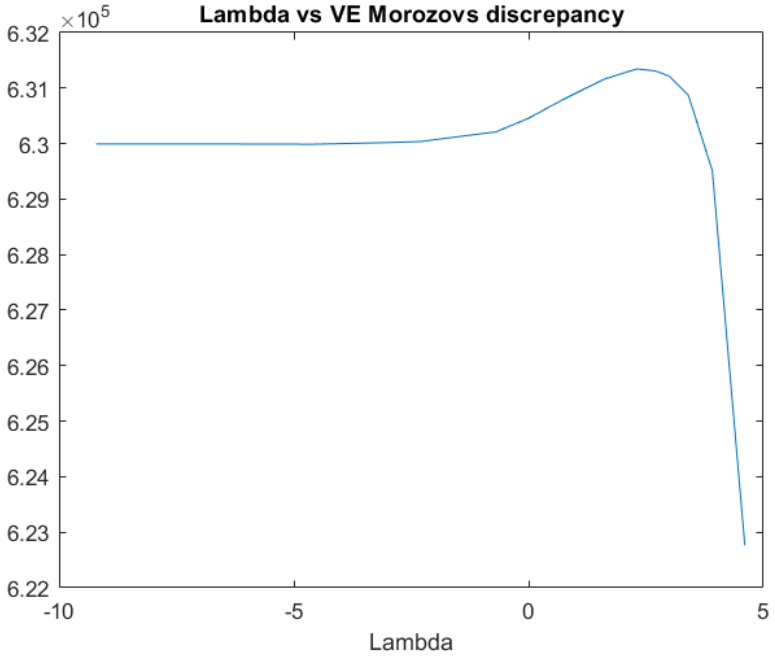

Figure 12: 20 sparse

(The values were all close to 1).

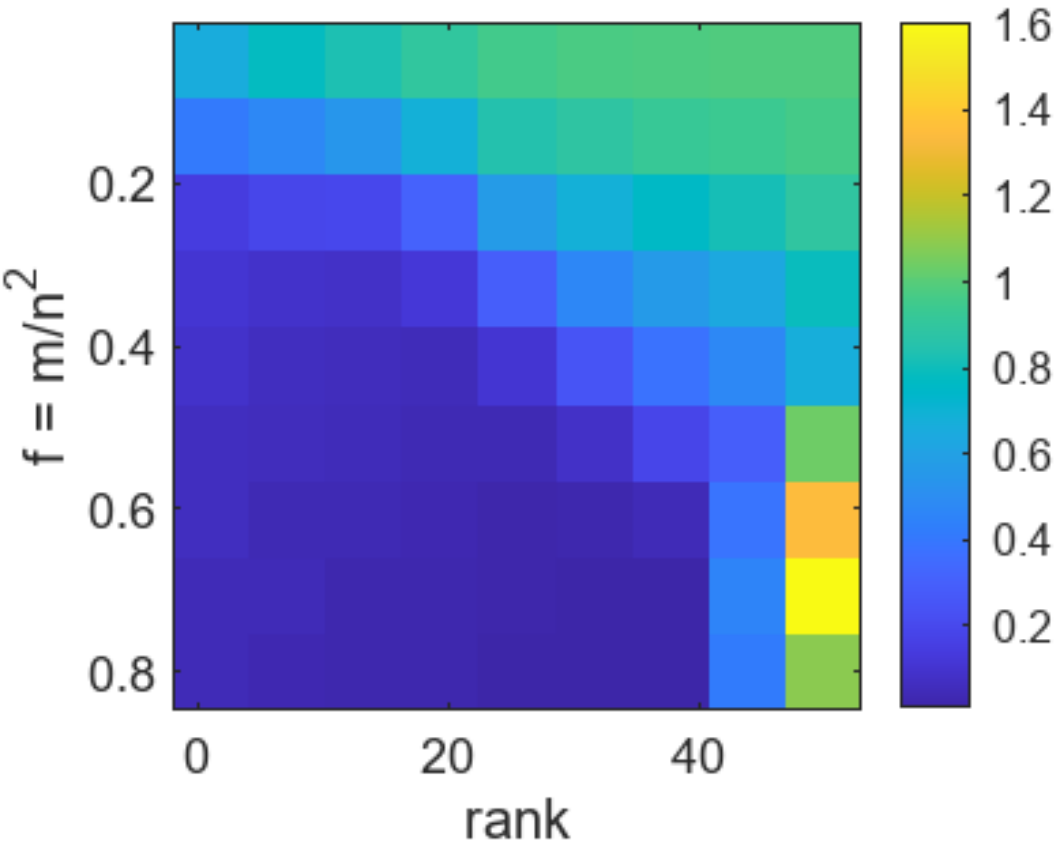

Figure 13: f and rank vs error

3 Question 3

RMSE for brick: 0.54
RMSE for sand: 0,54

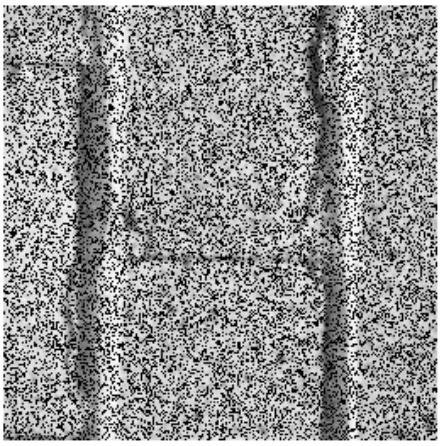

Figure 14: Brick

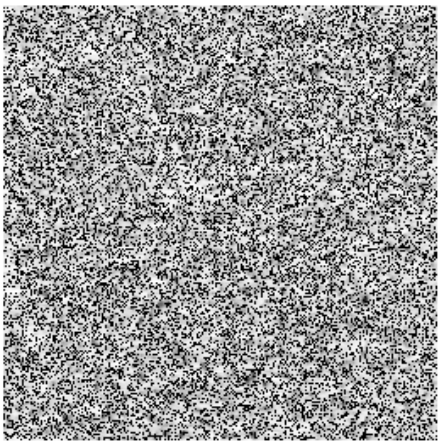

Figure 15: Sand

4 Question 4

Paper: Low-rank matrix recovery problem minimizing a new ratio of two norms approximating the rank function then using an ADMM-type solver with applications.
Publishing: 31 August 2022, Journal of Computational and Applied Mathematics
Authors: Kaixin Gao, Zheng-Hai Huang , Lulu Guo

In the paper, the authors propose a new nonconvex approximation of the rank function, that is the ratio of the nuclear norm and the Frobenius norm (N/F)

The nuclear norm of a matrix A denoted $||A||_*$, is defined as the sum of its singular values

$$||A||_* = \sum_i \sigma_i(A)$$

The norm can be computed from the singular value decomposition of A.

The Frobenius norm, sometimes also called the Euclidean norm, is matrix norm of an $m \times n$ matrix A defined as the square root of the sum of the absolute squares of its elements,

$$||A||_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2}$$

Let $X \in \mathbb{R}^{n_1 \times n_2}$ be the unknown matrix, $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \longrightarrow \mathbb{R}^m (m \leq n_1 n_2)$ is a linear map, similar to the one defined in the paper given in Q5, and $y \in \mathbb{R}^m$ is the observed vector.

Then the paper looks at the problem of $\frac{N}{F}$ minimization, in both the constrained and unconstrained forms.

Constrained :

$$\underbrace{\mathbf{min}}_{x} \frac{||X||_*}{||X||_F} \text{ s.t. } ||\mathcal{A}(X) - y||_2 \leq \tau$$

where $\tau \geq 0$ represents the magnitude of disturbance in observations.

Unconstrained:

$$\underbrace{\mathbf{min}}_{x} \lambda \frac{||X||_*}{||X||_F} + \frac{1}{2}||\mathcal{A}(X) - y||_2$$

where $\lambda > 0$ is the regularization perimeter.

The paper also proposes a alternating direction method of multipliers (ADMM) type algorithm for solving the above cost.

Advantages: As per the results in the paper, this method provides a better recovery compared to SVT in all scanerios over a wide variety of ranks and noise levels.

Disadvantages: Since this is a nonconvex optimization problem, the ADMM used to solve it may not converge always. To handle this, the inner loop of the algorithm terminates after some max iterations. Thus at times, we may not get the recovery.

5 Question 5

a) In the Fourier Phase retrieval problem, we receive $y = |Ax_0|^2$ for the signal x_0 for some known matrix $A \in \mathbb{C}^{m \times n}$ and solve for x subject to $y = |Ax|^2, x \in S$ where $S \subseteq \mathbb{C}^n$ corresponds to the set with the required properties in x. For the Fourier phase retrival, we take S as the set of k-sparse vectors.

Since $|Fz|^2 = |Fx|^2$, $z = e^{\imath\phi}x$ for $\phi \in [0, 2\pi)$ also comes as the solution to the general phase retrieval problem. Instead, we restrict our x to just real values and ignore the global phase difference i.e. $x \in \mathbb{R}^n$. This is the real-valued phase retrieval problem.

b) Taking matrix $X = xx^*$ and vector $y \in \mathbb{R}^m$ such that

$$y[i] = a_i^* X a_i = a_i^* x x^* a_i = |x^* a_i|^2 = |a_i^* x|^2$$

$$\implies y = |Ax|^2$$

where $|.|^2$ is taken component-wise.

Since this is the same as the real-valued phase retriaval problem, we model the phase retrival problem as

$$\text{find X subject to } \mathcal{A}(X) = y, X \succeq 0, rank(X) = 1$$

where $\mathcal{A}$ is the function as defined in the paper.

c) The restriction that rank(X) comes directly from the fact that $X = xx^*$, and thus all the rows are some constant multiple of the first row, which implies that the rank is 1.

Further $X \succeq 0$ since $v^*Xv = v^*xx^*v = |v^*x|^2 \geq 0$ for all $v \in \mathbb{R}^n$

d) If X is a positive semidefinite (PSD) matrix, it means that all of its eigenvalues are non-negative. In this case, the trace of X, denoted as $\text{Tr}(X)$, is the sum of its eigenvalues.

Additionally, for any matrix, the trace norm (nuclear norm) $||X||_*$ is the sum of its singular values.

Now, if X is a PSD matrix, then its singular values coincide with the square roots of its eigenvalues. This is a result of the spectral theorem, which states that for any PSD matrix, there exists an orthogonal matrix U such that $X = UDU^T$, where D is a diagonal matrix with non-negative eigenvalues of X on its diagonal.

Since the singular values of X are the square roots of its eigenvalues, for a PSD matrix, the trace norm $||X||_*$ is equal to $\text{Tr}(X)$. Since we are only dealing with matrices X such that $X \succeq 0$, minimizing the trace is the same as minimizing the nucleur norm.

e) The lower bound on the number of measurements for the matrix $A \in \mathbb{C}^{m \times n}$ drawn from an appropriate random distribution, say Gaussian or Fourier is $m = \Omega(n \text{ polylog n})$ using the PhaseLift method.