Journal of Economics and Applied Statistics
Volume 17 number 3, December 2020
ISSN : 1112-234X EISSN : 2600-664
Pages : 154- 165

STEIN UNBIASED RISK ESTIMATE AS A MODEL SELECTION ESTIMATOR

[*]**Nihad NOURI**
National Higher School of Statistics and Applied Economy,
LAMOPS, Kolea 42003 Algeria
Email: nouri.enssea@gmail.com

Fatiha MEZOUED
National Higher School of Statistics and Applied Economy,
LAMOPS, Kolea 42003 Algeria
Email: famezoued@yahoo.fr

Received : 2020/06/16 **Accepted** 2020/09/21 **Published online** : 2020/12/31

ABSTRACT: To restore a low-rank structure from a noisy matrix, many recent authors has used and studied truncated singular value decomposition. So thus, according to these studies, the image can be better estimated by shrinking the singular values as well. In this paper, we are interested in the performance of the model proposed by Candès (2012) for other thresholding function (Minimax Concave Penalty (MCP)), and under the assumption that the distribution of data matrix Y belongs to an elliptically distribution family which extends the Gaussian case. Under this distributional context, we propose to apply stein unbiased risk estimate (SURE) improved by S. Canu and D. Fourdrinier (2017), in order to select the best thresholding function between MCP and Soft-thresholding, and the optimal shrinking parameter λ from the data Y. Numerical results reveal that the risk estimate SURE is good, the minima are reached for the same λ ($\lambda^* = \hat{\lambda} = 5218.4$), the difference between the estimated (SURE) and the usual (Mean Square Error (MSE)) risks is small, and that the risk of MCP is lower than the one of Soft.

Keywords : elliptical distribution; mean square error; minimax concave penalty; singular value decomposition; stein's unbiased risk estimate;

JEL Classification: (Time New Roman, 9, normal).

1. INTRODUCTION:

While methods of developing image enhancement systems have seen remarkable progress in recent years, their ability to manipulate a large volume of data has remained rather modest (ALIN A.et ANASTASIOS B. et PANAGIOTIS T. 2001, 772–783), (ALIN

[*] Author Corresponding

A.et ANASTASIOS B. et PANAGIOTIS T. 2001). The main purpose of image enhancement is to improve the quality and the information content of the original data by eliminating noise without significant loss of information, from a statistical point of view this problem is can presented as follows.

Let Y be an observed n×m matrix that we seek to decompose into singular values with m < n, under the multivariate additive model

$$Y = M + e \ , \quad e \sim e\left(0_{nm}, I_n \otimes \Sigma\right) \qquad (1.1)$$

Where e is a noise matrix and e $(0_{nm}, I_n \otimes \Sigma)$ denotes the elliptically distributions with covariance matrix proportional to $I_n \otimes \Sigma$, here, Σ is an m x m invertible scale matrix and I_n is the n-dimentional identity matrix, M is an unknown n×m matrix to be estimated, which contains the information. A large number of substantial distributions are covered by the elliptically symmetric distributions such as the Gaussian, Exponential, Cauchy, t-Student, Logistic and Weibull.

It is assumed that the information contained in matrix M is redundant; therefore, this matrix M is of low rank. This assumption has been considered by many authors, see, (FUCHS J J. 2005), (THOMAS R. 2015) and (YUNG X 2019) that is

$$rank \quad M = q < m, \qquad (1.2)$$

Many authors gave differents appraoches in order to improve the estimators $\widehat{M}$ of M in model (1.1) through the unbiased risk estimator of Stein, known as SURE (Steins Unbiased Risk Estimate) we refer, for example, to the monographs of (CANDES E J. et SING-LONG C A. et TRZASKO J D. 2013, 4643–4657), (LUISIER F. et BLU T. et UNSER M. 2007, 593–606) and (ZHANG X P. et DESAI M D. 1998, 265–267), while they had adopted in their studies the Gaussian approach. Assuming that the noise e follows a normal distribution is really restricted, since, the real data does not necessarily and even very rarely follows this distribution, hence we are interested in reformalizing the problem for a noise according to an elliptically distribution as in model (1.1) see, (CANU S. et FOURDRINIER D. 2017).

Let $\boldsymbol{Y = U\Delta V^T}$ be the singular values decomposition of the matrix $\boldsymbol{Y}$ in model (1.1), where $\boldsymbol{U}$ is an n×n orthogonal matrix whose columns are the eigenvectors of $\boldsymbol{Y\,Y^T}$, $\boldsymbol{V}$ is an m×m orthogonal matrix whose columns are the eigenvectors of $\boldsymbol{Y^TY}$, and $\boldsymbol{\Delta}$ is an n×m diagonal matrix with $\boldsymbol{\Delta = (\Delta_1 \quad 0)}$ where $\boldsymbol{\Delta_1 = Diag(\sigma_i)}$, where $\boldsymbol{\sigma_i}$ denotes the ith singular value of $\boldsymbol{Y^TY}$.

We consider a family of estimators given as follows

$$\hat{M} = \hat{M}_{\lambda}\left(Y\right) = SVT_{\lambda}\left(Y\right) = \sum_{i=1}^{m} \varphi_{\lambda}\left(\sigma_i\right) u_i v_i^T \qquad (1.3)$$

which is obtained from a family of shrinkage functions $\varphi_\lambda(\sigma_i)$ see., Fan and Li (FAN J. et LI R. 2001, 1348–1360), such as the soft-thresholding and the Minimax concave penalty, the ones we are interested in this paper. The model selection problem here is to find the optimal shrinkage parameter λ and the right estimator $\widehat{M}_{\lambda}(Y)$ which minimizes the risk MSE (mean square error) formulated as follows

$$R\left(\hat{M}, M\right) = E_{M,\Sigma}\left[L\left\{\hat{M}\left(Y\right), M\right\}\right], \qquad (1.4)$$

Where $\boldsymbol{E_{M,\Sigma}}$ denotes the expectation with respect to the distribution of $\boldsymbol{Y}$ in model (1.1), and $\boldsymbol{L}$ is the invariant quadratic loss given by

$$\begin{aligned} L\left\{\hat{M}\left(Y\right), M\right\} &= \tau^{-2} tr\left[\left\{M - \hat{M}\left(Y\right)\right\}\Sigma^{-1}\left\{M - \hat{M}\left(Y\right)\right\}^T\right] \\ &= \tau^{-2}\left\|M - \hat{M}\left(Y\right)\right\|^2_{\Sigma^{-1}} \end{aligned} \qquad (1.5)$$

here ***tr*** denotes the trace of the matrix, $\|.\|_F$ the Forbenius norm and $\boldsymbol{\tau^2}$ is the proportionality constant of the covariance matrix defined through $\boldsymbol{cov() = \tau^2 I_n \otimes \Sigma}$ (see reference (CANU S. et FOURDRINIER D. 2017, 3-5).

The main obstacle of this minimization problem is that we do not know M, so we can not find the estimate $\widehat{M}$ which minimizes the risk (1.4). The idea is then to estimate this risk. In this case, we consider the steins unbiased risk estimate (SURE) proposed by Charles Stein (STEIN C M. 1981, 1135–1151), whose expectation is equal to the risk, that is,

$$R\left(\hat{M}, M\right) = E_{M,\Sigma}\left\{\delta_0(Y)\right\} \qquad (1.6)$$

The paper is organized as follows, in section 2, we recall basic notions of risk estimate SURE (Stein Unbiased Risk Estimator) and complexity control, sections 3 and 4 contain our main contribution which consist in some experiments on simulated data of the SURE established by Canu and Fourdrinier (2017) when $\boldsymbol{\Sigma = \sigma^2 I_m}$ is known and proportional to the identity matrix with ε follows a multivariate t-Student distribution. Finally, in section 5, we give some conclusions and perspectives.

2. Complexity control and model selection

From the observed matrix ***Y*** given in (1.1) our main objective is to propose an estimator $\widehat{M}$ of M which minimize the MSE risk in (1.4), which depends on the trade-off between the ***Bias*** and the ***variance***.

$$MSE = R\left(\hat{M}, M\right) = \mathrm{var}(\hat{M}) + Bias^2$$

and it can be written as function of the empirical error $||Y - \widehat{M}||_{\Sigma}^{2}$

$$\left\|M - \hat{M}\right\|_{\Sigma}^{2} = \left\|Y - \hat{M}\right\|_{\Sigma}^{2} + pen(Y, \hat{M}) \qquad (2.1)$$

we recall that, the expression in (2.1) represents stein unbaised risk estimate (SURE), where *pen(Y,* $\widehat{M}$*)* is a penalty function, this penalty is a function of the model's complexity which goes up when the complexity goes up.

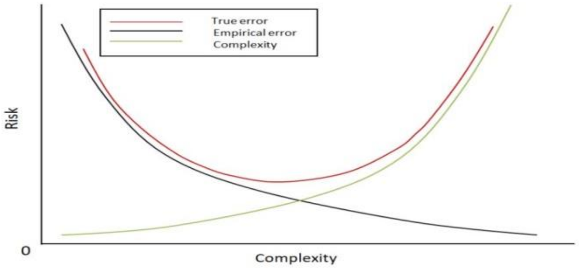


Figure N°1: Illustration of the relation between true error and empirical error

As we can see in the above figure (Figure N°1**)**, the empirical error always decreases, but true error (MSE) decreases up to some points and then it starts to increase; and the optimal model would be a model which has a minimum true error not empirical error.

Recently, Canu, S and Fourdrinier, D developed various SURE- type estimators for a quadratic loss which correspond to different situations depending on whether the covariance matrix Σ is known or unknown, and that the noise distribution e in (1.1) is Gaussian or not (for more details, please refer to (CANU S. et FOURDRINIER D. 2017, 60–72)), we are interested in using the one when $\boldsymbol{\Sigma = \sigma^2 I_m}$ is known and proportional to the identity matrix with e is spherically symmetric

$$\delta_0^{d}(Y) = \frac{1}{t^2 s^2}\left\|Y - \hat{M}\right\|_F^2 + 2div_Y(\hat{M}) - nm \qquad (2.2)$$

Where

$$div_Y(\hat{M}) = \sum_{i=1}^{m}\left\{1_{s_i>1} + |m-n|\left(1-\frac{1}{s_i}\right)\right\} + 2\sum_{i=1}^{m}\sum_{\substack{j=1\\ j\neq i}}^{m}\frac{s_i(s_i-1)_+}{s_i^2 - s_j^2},$$

As for M's estimator, many challenges have been made to improve the native estimate $\widehat{M}$ = Y under the quadratic loss using trancated singular value decomposition obtained from shrinkage functions family $\varphi_\lambda(\sigma_i)$, see., (BIGOT J. et DELDALLE C. et FRAL D. 2017, 4991–5040), (CANDES E J. et SING-LONG C A. et TRZASKO J D. 2013, 4643–4657) and (SARWAR B. et KARYPIS G. et KONSTAN J. et RIEDL J. 2002, 28), here, we are interested in two shrinkage functions Soft-thresholding and MCP given respectively as

$$j_{1}(s_i) = (s_i - 1)_+ \qquad (2.3)$$

and

$$\varphi_\lambda(\sigma_i) = \begin{cases} 0 & si\sigma_i \le \lambda \\ \frac{\mu}{\mu-1}(\sigma_i - \lambda) & si\lambda < \sigma_i \le \mu\lambda \\ \sigma_i & si\mu\lambda < \sigma_i \end{cases} \qquad (2.4)$$

where μ is greater than 1 ($\mu > 1$)

so thus, the optimal thresholding values $\boldsymbol{\lambda}^*$ and $\hat{\lambda}$ are the ones which satisfy

$$\lambda^* = \arg\min_{\lambda}\left\|M - \hat{M}_\lambda(Y)\right\|_\Sigma^2$$

and

$$\hat{\lambda} = \arg\min_{\lambda}\left\|Y - \hat{M}_\lambda(Y)\right\|_\Sigma^2 + 2div_Y(\hat{M}_\lambda(Y))$$

Now for the coefficient proportionality we give the following lemma.

Lemma 1: let e be a random noise as in model (1.1). The proportionatity coefficient τ equals to

$$\tau = E\|M\|^2 \big/ SNR\sqrt{nm}. \qquad (2.5)$$

Proof. From model (1.1), we have

$$\text{ɘ} \sim \text{e} \left(0_{nm}, I_n \otimes \Sigma\right)$$

see reference (CANU S. et FOURDRINIER D. 2017, 4) we have also

$$\tau^2 = E[R^2]/nm.$$

with $R^2 = \|\varepsilon\|_F^2$

so thus

$$\|\varepsilon\|^2 = nm\tau^2$$

since from the fact that

$$SNR = E\|M\|^2 / \sqrt{E(R^2)}$$

where SNR denotes the signal to noise ratio, we have

$$SNR = E\|M\|^2 / \tau\sqrt{nm}.$$

the result in (2.5) follows.

3. Experiments on simulation data

This section consists on two points, first; we compare the risk estimator SURE with the MSE for the two thresholding functions, MCP and Soft-thresholding. Then, we compare SURE risk for the two different functions, under the assumption that the random noise in model (1.1) follows t-Student distribution with degree of freedom equal to 5, its generative function is given by

$$f(t)\mu\ (1-t^2)^{-\frac{1}{2}(s-1)}, t \geq 0$$

where s denotes the freedom degree.

we define the mean-squared error (MSE) of SVTλ(Y) as

$$MSE(Y) = E\left\|M - SVT_l\,(Y)\right\|_F^2, \tag{3.1}$$

an unbiased risk estimator (SURE) of (3.1) is

$$\delta_0^{\text{d}}(Y) = \frac{1}{\text{t}^2\text{s}^2}\left\|Y - SVT_l\,(Y)\right\|_F^2 + 2div_Y(SVT_l\,(Y)) - nm, \tag{3.2}$$

Thresholding algorithm

Our simulation study consists in following these steps

step1: Create noise instance,
step2: Built the noisy data using model (1.1),
step3: Compute SVD of the noisy matrix ***Y***,
step4: Generate linearly spaced vector of lambda λ,
step5: Calculate the value of the proportionality coefficient τ^2 using equation (2.5),
step6: Compute the singular value thresholding (SVT) using the two different shrinkage functions in (2.3) and (2.4),
step7: Compute MSE and SURE for SVT using formulas (3.1) and (3.2).

3.1. Experimental protocol

In the following, we assess the behavior of SURE given by equations (3.2) to the one of the MSE in equation (3.1) as a model selector. recall that, The experimental protocol developed here is inspired by the one proposed by Candès et al (CANDES E J. et SING-LONG C A. et TRZASKO J D. 2013), considering t-Student noise distributions for ə in model (1.1), where here the covariance matrix Σ is known with $\boldsymbol{\Sigma = \sigma^2 I_m}$. Let Y be an n x m observed matrix, which is constructed according to the model (1.1) with m = 200 and n = 500.

We test the different models corresponding to the different proposed shrinkage function, for a set of matrices M randomly generated according to a t-Student distribution of freedom degree equal to 5, with a set of different ranks, namely {m/4 ; m/2 ; m}, as well as for different SNR (signal noise ratio) values {0.2; 0.8; 1.2}

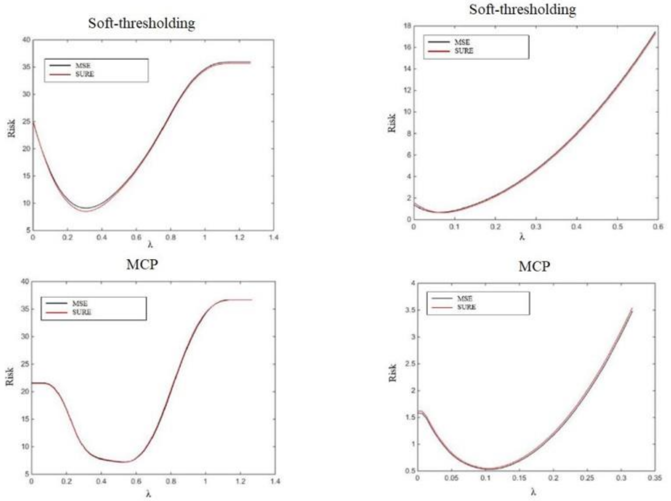
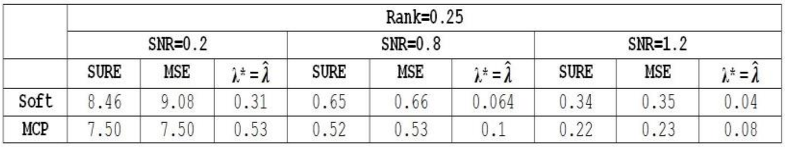
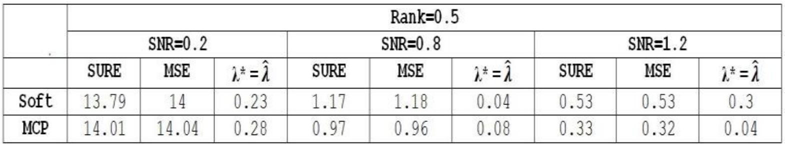


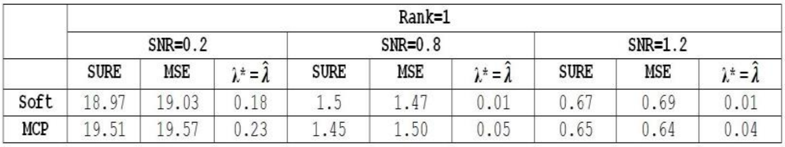
Figure N°2: Comparison of the Monte Carlo estimates of the risk (black) with the SURE estimator (red) versus λ, rank(M)=50, SNR=0,2 (left) and rank(M)=50, SNR=0,8 (right)



	Rank=0.25								
	SNR=0.2			SNR=0.8			SNR=1.2		
	SURE	MSE	$\lambda^* = \hat{\lambda}$	SURE	MSE	$\lambda^* = \hat{\lambda}$	SURE	MSE	$\lambda^* = \hat{\lambda}$
Soft	8.46	9.08	0.31	0.65	0.66	0.064	0.34	0.35	0.04
MCP	7.50	7.50	0.53	0.52	0.53	0.1	0.22	0.23	0.08



	Rank=0.5								
	SNR=0.2			SNR=0.8			SNR=1.2		
	SURE	MSE	$\lambda^* = \hat{\lambda}$	SURE	MSE	$\lambda^* = \hat{\lambda}$	SURE	MSE	$\lambda^* = \hat{\lambda}$
Soft	13.79	14	0.23	1.17	1.18	0.04	0.53	0.53	0.3
MCP	14.01	14.04	0.28	0.97	0.96	0.08	0.33	0.32	0.04



	Rank=1								
	SNR=0.2			SNR=0.8			SNR=1.2		
	SURE	MSE	$\lambda^* = \hat{\lambda}$	SURE	MSE	$\lambda^* = \hat{\lambda}$	SURE	MSE	$\lambda^* = \hat{\lambda}$
Soft	18.97	19.03	0.18	1.5	1.47	0.01	0.67	0.69	0.01
MCP	19.51	19.57	0.23	1.45	1.50	0.05	0.65	0.64	0.04

Table N°1: Estimation of the optimal risk in relation to SNR and the rank of *M* for the two proposed thresholding functions

The experiment above [Figure N°2] illustrates the difference between the risk MSE and the estimated risk SURE. Otherwise, it exhibits the attitude of SURE in (3.2) to estimate the invariant loss in (1.5) when the noise is spherically symmetric and non-Gaussian. We note that the risk estimate is good, the minima are reached for the same λ and the difference between the estimated and MSE risks is low, for the different values of SNR and rank.

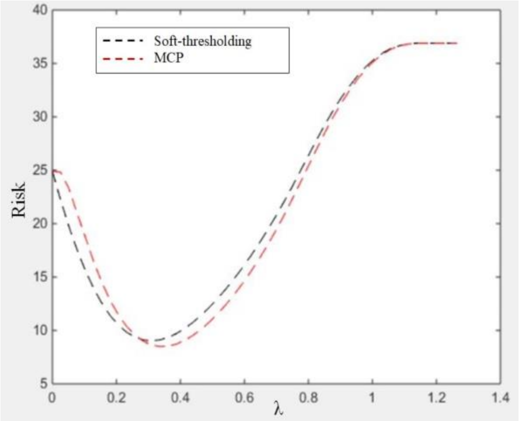


Figure N°3: Comparison of the estimated risk versus λ for the different proposed thresholding functions (MCP and Soft) (with Rank (M) = 50, SNR = 0.2)

Figure N°3 presents results for a choice of fixed parameters (namely SNR = 0.2 and rank (M) = 50). While, Table N°1 presents results obtained from tests for all of the proposed parameter values, for each value of SNR, and for each value of rank of M. It presents optimal risks of each method

3.2. Magnetic resonance imaging.

Here, we consider a sharp image [Figure N°4 (a)] (a cardiac perfusion MRI) used by Candès et al (CANDES E J. et SING-LONG C A. et TRZASKO J D. 2013) in the Gaussian case, for which we add a Multivariate Student noise, [Figure N°4 (b)].

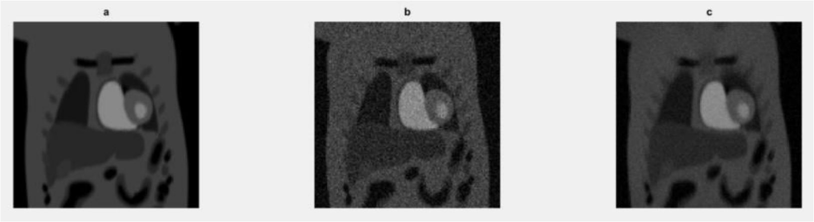


Figure N°4: (a) truth image, (b) noisy image, (c) SVT denoising image

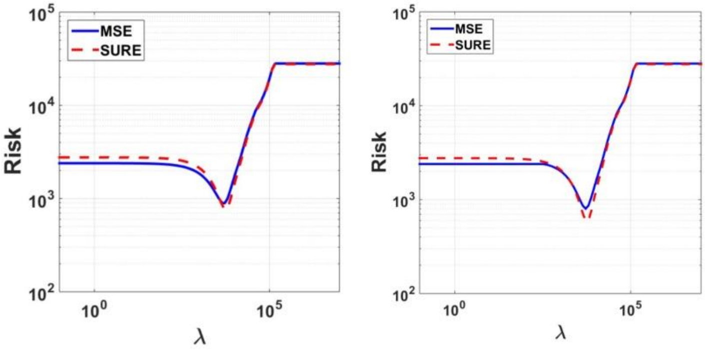
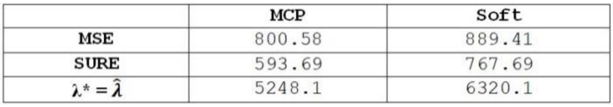


Figure N°5: Plots of MSE and SURE for SVT as a function of thresholdding value, λ, using Soft-thresholding (left) and MCP (right); the optimal thresholding value λ correspondent to MCP shrinkage function is used to generate the images in [Figure 4 (c)]



	MCP	Soft
MSE	800.58	889.41
SURE	593.69	767.69
$\lambda^* = \hat{\lambda}$	5248.1	6320.1

Table N°2: Estimation of the optimal risk for the two proposed thresholding functions

4. Results and Discussion

First, we find that SOFT and MCP tend to obtain a lower relative risk when the rank of the matrix M is small, and when the SNR increases, see table N°1. Indeed for an SNR equal to 1.2, and a rank equal to 50, the two functions obtain the smallest risk. However, from the comparison of the risk, we note that the risk of MCP is better from the one given by Soft, since; the risk associated to MCP is in all cases lower than the one associated to Soft.

In conclusion, we recommend to use the MCP type function for estimating the unknown matrix ***M***.

5. Conclusion and perspectives

Truncated singular value decomposition remains powerful and useful to recover a reduced rank matrix from noisy data which is a quite interesting topic that has excite the scientific community for a few years. When it comes to shrinkage, a recurring problem that has been little addressed until now is the choice of the shrinkage function as well as the choice of the optimal threshold, for this reason, we are interested in the problem of model selection by minimizing the risk estimator SURE for singular value thresholding. The use of SURE in (2.2) developed by Fourdrinier and Canu is interesting since it is adapted for a large class of density, namely the elliptical class, as it contains many distributions that are more leptokurtic than the normal distribution, it allows to model more structures in the real data.

Note that, an image is often cut into ***K*** blocks, in order to carry out its processing. Consequently, we plan to adapt this model selection on analyzes made on block-wise images rather than globally, in order to obtain a better estimate of the noisy matrix. Otherwise, it is really interesting to propose a new rule to select two parameters, the one related to the thresholding λ and the block-wize parameter ***K*** from the data, for example, for $\mathcal{E} \sim \mathcal{N}(0_{nm},I_n \otimes \Sigma)$ with $\Sigma = \sigma^2 I_m$ and k=4, k=10 we have

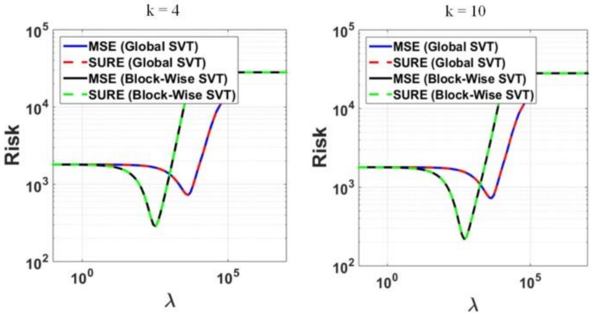


Figure N°6: Plots of MSE and SURE for SVT as a function of threshold value, for k=4 (left) and k=10 (right)

As we can see from the illustration given in Figure N°6, the choice of the value of ***K*** affects the value of the optimal lambda λ.

Acknowledgment

We would like to express our gratitude to professors S.CANU and D.FOURDRINIER for fruitful conversations, and to the whole team of the LITIS laboratory at the University of Rouen, France.

BIBLIOGRAPHY:

1. ALIN A., ANASTASIOS B. et PANAGIOTIS T., «*Novel Bayesian multiscale method for speckle removal in medical ultrasound images*.», IEEE, transaction medical imaging journal, n°20, 2001, pp.772-783.

2. BIGOT J., DELDALLE C. et FRAL D., «*Generalized SURE for optimal shrinkage of singular values in low-rank matrix denoising.*», the journal of machine learning research, n°18, 2017, pp.4991-5040.

3. CANDES E J., SING-LONG C A. et TRZASKO J D., «*Unbiased risk estimate for singular value thresholding and spectral estimators.*», IEEE, transactions on signal processing, n° 61, 2013, pp.4643-4657.

4. CANU S. et FOURDRINIER D., «*Unbaised risk estimates for matrix estimation in the elliptical case.*», Elsevier, Journal of Multivariate Analysis, n°158, 2017, pp.60-72.

5. FAN J. et LI R., «*Variable selection via nonconcave penalized likelihood and its oracle properties.*», Journal of the American statistical Association, n° 96, 2001, pp.1348-1360.

6. FUCHS J J., «*Vers une nouvelle décomposition de matrice*.», GRETSI, Groupe dEtudes du Traitement du Signal et des Images, 2005.

7. LUISIER F., BLU T. et UNSER M., «*A new SURE approach to image denoising: Interscale orthonormal wavelet thresholding*.», IEEE, Transactions on image processing, n° 16, 2007, pp.593-606.

8. MEHTA R. et RANA K., «*Evolution of singular value decomposition in recommendation systems: a review.*», International journal of business intelligence and data mining, n°14, 2019, pp.528-547.

9. SARWAR B., KARYPIS G., KONSTAN J. et RIEDL J., «*Incremental singular value decomposition algorithms for highly scalable recommender systems*.», Citeseer, Fifth international conference on computer and information science, n° 27, 2002, pp.27-28.

10. STEIN C M., «*Estimation of the mean of a multivariate normal distribution.*», JSTOR. The Annals of Statistics, 1981, pp.1135-1151.

11. THOMAS R., «*Decomposition en valeurs singulieres et approximation de rang faible*.» 2015.

12. YUNG X., «*Singular value decomposition based recommendation using imputed data*.» Knowledge-Based System, *n*° 163, 2019, pp.485-494.

13. ZHANG X P. DESAI M D., «*Adaptive denoising based on SURE risk*.», IEEE, signal processing letters, n°5, 1998, pp. 265-267.