香港城市大學
City University of Hong Kong

Inverse Problems on Large Scales
November 28-December 2, 2022

RICAM
JOHANN-RADON-INSTITUTE
FOR COMPUTATIONAL AND APPLIED MATHEMATICS

Selecting Regularization Parameters for Nuclear Norm Type Minimization Problems

Raymond Chan
Department of Mathematics

KX LI, YW Wen, HW LI
Hunan Normal University, China
Capital Normal University, China

專業 創新 胸懷全球
Professional · Creative
For The World

Supported by HKRGC

Outline

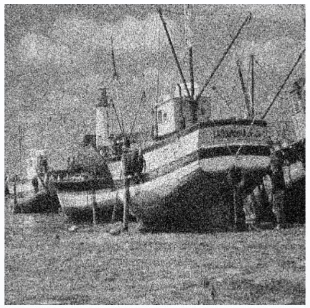
- **Introduction**
- **Choosing the parameters for the regularization problems**
- **Applications**
- **Conclusion**

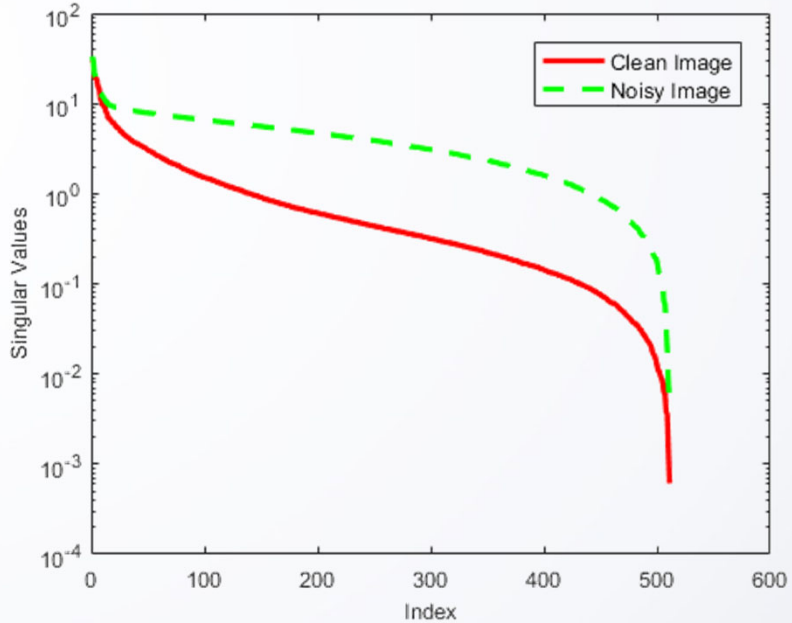
- **Introduction**
- Choosing the parameters for the regularization problems
- Applications
- Conclusion

Introduction

Low-rank matrix recovery problems arise in many applications:

- ☐ Image Denoising
- ☐ Image inpainting
- ☐ Matrix completion
- ☐ Background subtraction





The regularized model and constrained model

The noisy observation $Y \in \mathbb{R}^{m \times n}$ can be modeled as

$$Y = X + W,$$

where X is the clean low-rank matrix, and W is the Gaussian white noise with variance τ^2.

- **The constrained model:** $$\min_{\|X-Y\|_F^2 \leq \eta} f(X)$$

- **The regularization model:** $$\min_X f(X) + \frac{1}{2\lambda}\|X - Y\|_F^2$$

$$\eta \longleftrightarrow \lambda$$

The choice of regularization function for low-rank recovery

A rank minimization approach to recovery X is to set

$$f(X) = \text{rank}(X)$$

However, the rank function is nonconvex and NP-hard!

Relaxation:

- Nuclear norm $f(X) = \|X\|_* \equiv \sum_{i=1}^{m} \sigma_{X,i}$
- Truncated nuclear norm $f(X) = \|X\|_{r,*} \equiv \sum_{i=r+1}^{m} \sigma_{X,i}$
- Weighted nuclear norm $f(X) = \|X\|_{\mathbf{w},*} \equiv \sum_{i=1}^{m} w_i \sigma_{X,i}$

where $\sigma_{X,1} \geq \sigma_{X_2} \geq \cdots \geq \sigma_{X,m}$ are singular values of X.

The optimal regularization parameter

- [M. Gavish *et al.* 2014] The optimal **hard threshold** parameter λ in

$$\min_X \operatorname{rank}(X) + \frac{1}{2\lambda}\|X - Y\|_F^2$$

  is chosen as $\lambda = (4/\sqrt{3})\sqrt{n}\tau$ when the noise level τ is known or $\lambda = 0.2858\sigma_{Y,\text{median}}$ when τ is unknown.

- [E. Candes *et al.* 2013] The optimal **soft threshold** parameter λ in

$$\min_X \|X\|_* + \frac{1}{2\lambda}\|X - Y\|_F^2$$

  can be estimated by minimizing

$$\text{SURE}(\lambda) = -mn\tau^2 + \sum_{i=1}^{m} \min\left(\lambda^2, \sigma_{Y,i}^2\right) + 2\tau^2 \operatorname{div}_Y \left(\text{SVT}_\lambda(Y)\right).$$

Other methods: GCV, L-curve, discrepancy principle . . .

Discrepancy principle

- Mathematically, the two minimization models are equivalent:

$$\min_{\|X-Y\|_F^2 \le \eta} f(X) \quad \text{and} \quad \min_X f(X) + \frac{1}{2\lambda}\|X-Y\|_F^2,$$

- Given $\|Y\|_F^2 > \eta > 0$ (exclude $X = 0$), there exists a corresponding optimal $\lambda > 0$ satisfying

$$\|X(\lambda) - Y\|_F^2 = \eta \qquad (*)$$

such that the solution of the former is also a solution of the latter, and vice versa.

- Choosing λ is difficult, while determining η is easier. Can we obtain λ for any given η through solving the nonlinear equation $(*)$?

Determining an upper bound for η

Theorem 1. [J. L. Mead *et al.*, 2013] *Assume that the error matrix* $Y - X = W \sim \mathcal{N}(0, \tau^2 I)$. *Then* $\|W\|_F^2$ *is* χ^2*-distributed with variance* τ^2 *and* mn *degrees of freedom. Moreover, we have*

$$\mathbb{E}\|W\|_F^2 = mn\tau^2.$$

Theorem 2. [N. Halko *et al.*, 2010] *Given a Gaussian matrix* $W \sim \mathcal{N}(0, \tau^2 I)$, *we have*

$$\mathbb{P}\{\|W\|_F \geq \mathbb{E}\|W\|_F + \tau t\} \leq \mathrm{e}^{-t^2/2}.$$

⇒ $\mathbb{P}\{\|W\|_F \leq \mathbb{E}\|W\|_F + t\tau\} \geq 1 - \mathrm{e}^{-t^2/2}$. When $t \geq 4$, the noise is bounded with probability greater than 0.9996.

CityU

The upper bound η

- Set the upper bound to

$$\eta = cmn\tau^2,$$

$\eta \longrightarrow \lambda$

where $c \simeq 1$ can be adjusted appropriately to suit the applications.

- If the noise level is not given, τ can be estimated using the median rule [D. L. Donoho *et al.*, 1994]:

$$\tau = \frac{\text{median}(\tilde{Y}_{H,H})}{0.6745}$$

where $\tilde{Y}_{H,H}$ is the high-high coefficient of Y at the finest wavelet transform level.

CityU

- Introduction
- **Choosing the parameters for the regularization problems**
- Applications
- Conclusion

Nuclear Norm Minimization (NNM)

NNM: [J. F. Cai *et al.* 2010] $\min_{X} \|X\|_* + \frac{1}{2\lambda}\|Y - X\|_F^2.$

Lemma 1. *Let* $Y = \sum_{i=1}^{m} \sigma_{Y,i} u_i v_i^T$ *be the SVD of the matrix* Y*. Then the singular value thresholding* (SVT) *operator*

$$\mathrm{SVT}_{\lambda}(Y) \equiv \sum_{i=1}^{m} (\sigma_{Y,i} - \lambda)_+ u_i v_i^T$$

is the minimizer of NNM.

How to find λ when given η?

Nuclear norm minimization

Theorem 3. *Let*

$$\widehat{X}(\lambda) = \arg\min_{X} \|X\|_* + \frac{1}{2\lambda}\|Y - X\|_F^2,$$

i.e., $\widehat{X}(\lambda) = \mathrm{SVT}_\lambda(Y)$, *then the nonlinear function*

$$\psi(\lambda) \equiv \|Y - \widehat{X}(\lambda)\|_F^2$$

is a positive, continuous, and strictly increasing function of λ *in* $[0, \sigma_{Y,1}]$.

Proof:

$$\psi(\lambda) = j\lambda^2 + \sum_{i=j+1}^{m} \sigma_{Y,i}^2, \quad \lambda \in (\sigma_{Y,j+1}, \sigma_{Y,j}], \quad j = 1, \ldots, m.$$

Nuclear norm minimization

Theorem 4. *Let* $\widehat{X}(\lambda) = \mathrm{SVT}_\lambda(Y)$, *then the nonlinear equation* $\psi(\lambda) \equiv \|Y - \widehat{X}(\lambda)\|_F^2 = \eta$ *has a unique solution* $\lambda > 0$ *for any* η *satisfying* $0 < \eta < \|Y\|_F^2$.

Proof:

- Define a new sequence $b_j = j\sigma_{Y,j}^2 + \sum_{i=j+1}^m \sigma_{Y,i}^2, j = 1, ..., m$.
- $\{b_j\}$ is non-increasing: $b_1 \geq b_2 \geq \ldots \geq b_m \geq b_{m+1} = 0$.
- $\exists$ an k s.t. $b_{k+1} < \eta \leq b_k$ and recall $\psi(\lambda) = k\lambda^2 + \sum_{i=k+1}^m \sigma_{Y,i}^2$:

$$\lambda = \sqrt{(\eta - \sum_{i=k+1}^{m} \sigma_{Y,i}^2)/k}$$

Truncated nuclear norm minimization (TNNM)

TNNM: [Y. Hu *et al.*, 2012; T. H. Oh *et al.*, 2013]

$$\min_X \|X\|_r + \frac{1}{2\lambda}\|Y - X\|_F^2,$$

where $\|X\|_r = \sum_{i=r+1}^{m} \sigma_{X,i}$ is the truncated nuclear norm (TNN).

Lemma 2. *For any* $\lambda > 0$*, a global solution of TNNM with a target rank* $r < m$ *is given by the partial singular value thresholding* (PSVT) *operator*

$$\mathrm{PSVT}_{r,\lambda}(Y) \equiv \sum_{i=1}^{r} \sigma_{Y,i} u_i v_i^T + \sum_{i=r+1}^{m} (\sigma_{Y,i} - \lambda)_+ u_i v_i^T.$$

Truncated nuclear norm minimization

- Denote $Y_r = \sum_{i=1}^{r} \sigma_{Y,i} u_i v_i^T$ and $Z = Y - Y_r$, then

$$\min_X \|X\|_r + \frac{1}{2\lambda}\|Y - X\|_F^2 \quad \Rightarrow \quad \min_X \|X\|_* + \frac{1}{2\lambda}\|Z - X\|_F^2.$$

- The solution of TNNM is given by

$$X = Y_r + \mathrm{SVT}_\lambda(Z).$$

Therefore, although $\|X\|_r$ is non-convex, we can solve it easily with the help of convex nuclear norm optimization.

Generalized weighted nuclear norm minimization (GWNNM)

GWNNM: [Huang *et al.*, 2019]:

$$\min_X \|X\|_{\mathbf{w},*} + \frac{1}{2\lambda}\|Y - X\|_F^2$$

where $\|X\|_{\mathbf{w},*} = \sum_{i=1}^{m} w_i \sigma_{X,i}$ is the generalized weighted nuclear norm, with the weights

$$w_i = (\sigma_{Y,i} + \epsilon)^{p-1}, i = 1, ..., m,$$

$0 \leq p < 1$ and ϵ is a sufficiently small positive number.

Generalized weighted nuclear norm minimization

Lemma 3. [S. Gu *et al.*, CVPR, 2014] *For any* $\lambda > 0$, *if* $\{w_i\}_{i=1}^m$ *satisfy* $0 \leq w_1 \leq w_2 \leq ... \leq w_m$, *then a global solution of GWNNM is given:*

$$\mathrm{WSVT}_\lambda(Y) \equiv \sum_{i=1}^{m} (\sigma_{Y,i} - \lambda w_i)_+ u_i v_i^T.$$

Theorem 5. *Let* $\widehat{X}_{\mathbf{w}}(\lambda) = \arg\min_X \|X\|_{\mathbf{w},*} + \frac{1}{2\lambda}\|Y - X\|_F^2$, *then*

$$\phi(\lambda) = \|Y - \widehat{X}_{\mathbf{w}}(\lambda)\|_F^2$$

is a positive, continuous, and strictly increasing function of λ *in* $[0, \sigma_{Y,1}]$.

Generalized weighted nuclear norm minimization

Theorem 6. *Let* $\widehat{X}_{\mathbf{w}}(\lambda) = \mathrm{WSVT}_{\lambda}(Y)$*, then* $\|Y - \widehat{X}_{\mathbf{w}}(\lambda)\|_F^2 = \eta$ *has a unique solution* $\lambda > 0$ *for any* η *satisfying* $0 < \eta < \|Y\|_F^2$.

Proof:

- $\phi(\lambda) = \lambda^2 \sum_{i=1}^{k} w_i^2 + \sum_{i=k+1}^{m} \sigma_{Y,i}^2$.
- Define the sequence $b_j = (\frac{\sigma_{Y,j}}{w_j})^2 \sum_{i=1}^{j} w_i^2 + \sum_{i=j+1}^{m} \sigma_{Y,j}^2, j = 1, ..., m$.
- $\{b_j\}$ is non-increasing and there exists a k such that $b_{k+1} < \eta \le b_k$.

$$\lambda = \sqrt{(\eta - \sum_{i=k+1}^{m} \sigma_{Y,i}^2) / \sum_{i=1}^{k} w_i^2}$$

CityU

Randomized method

- SVD cost for $Y \in \mathbb{R}^{m\times n}$ is $O(mn\min(m,n))$ and is expensive.

- Let $Q_\ell \in \mathbb{R}^{n\times\ell}$, $\ell \le m$, be an orthogonal matrix. Then for all $A \in \mathbb{R}^{m\times\ell}$ and $X = AQ_\ell^\top$, we have

$$\|Y - X\|_F^2 = \|YQ_\ell - A\|_F^2 + a.$$

  Here $a = \|Y\|_F^2 - \|YQ_\ell\|_F^2$.

- Hence we can consider the reduced problem:

$$\min_A f(A) + \frac{1}{2\lambda}\|YQ_\ell - A\|_F^2.$$

  The size of the matrix YQ_ℓ ($\in \mathbb{R}^{m\times\ell}$) is smaller than that of Y ($\in \mathbb{R}^{m\times n}$).

- Introduction
- Choosing the parameters for the regularization problems
- **Applications**
- Conclusion

Simulations on synthetic data

- The s-rank matrix $X_\dagger = MN^\top \in \mathbb{R}^{m\times n}$ where $M \in \mathbb{R}^{m\times s}$ and $N \in \mathbb{R}^{n\times s}$ are uniform-random matrices.
- $s = \lfloor \rho \cdot \min(m, n) \rfloor$ with ρ from 1% to 40%.
- A noise matrix $W \sim \mathcal{N}(0, \tau^2 I)$ is added to $X_\dagger$ to get the observed matrix Y, i.e., $Y = X_\dagger + W$.
- We recover the matrix $X_\dagger$ from its noisy observation matrix Y.
- The bound $\eta = mn\tau^2$.
- For the randomized algorithms, $\ell = s + 5$.

CityU

Methods to compare

- [M. Gavish *et al.* 2014] The optimal **hard threshold** parameter λ in

$$\min_X \operatorname{rank}(X) + \frac{1}{2\lambda}\|X - Y\|_F^2$$

where $\lambda = (4/\sqrt{3})\sqrt{n}\tau$.

- [E. Candes *et al.* 2013] The optimal **soft threshold** parameter λ in

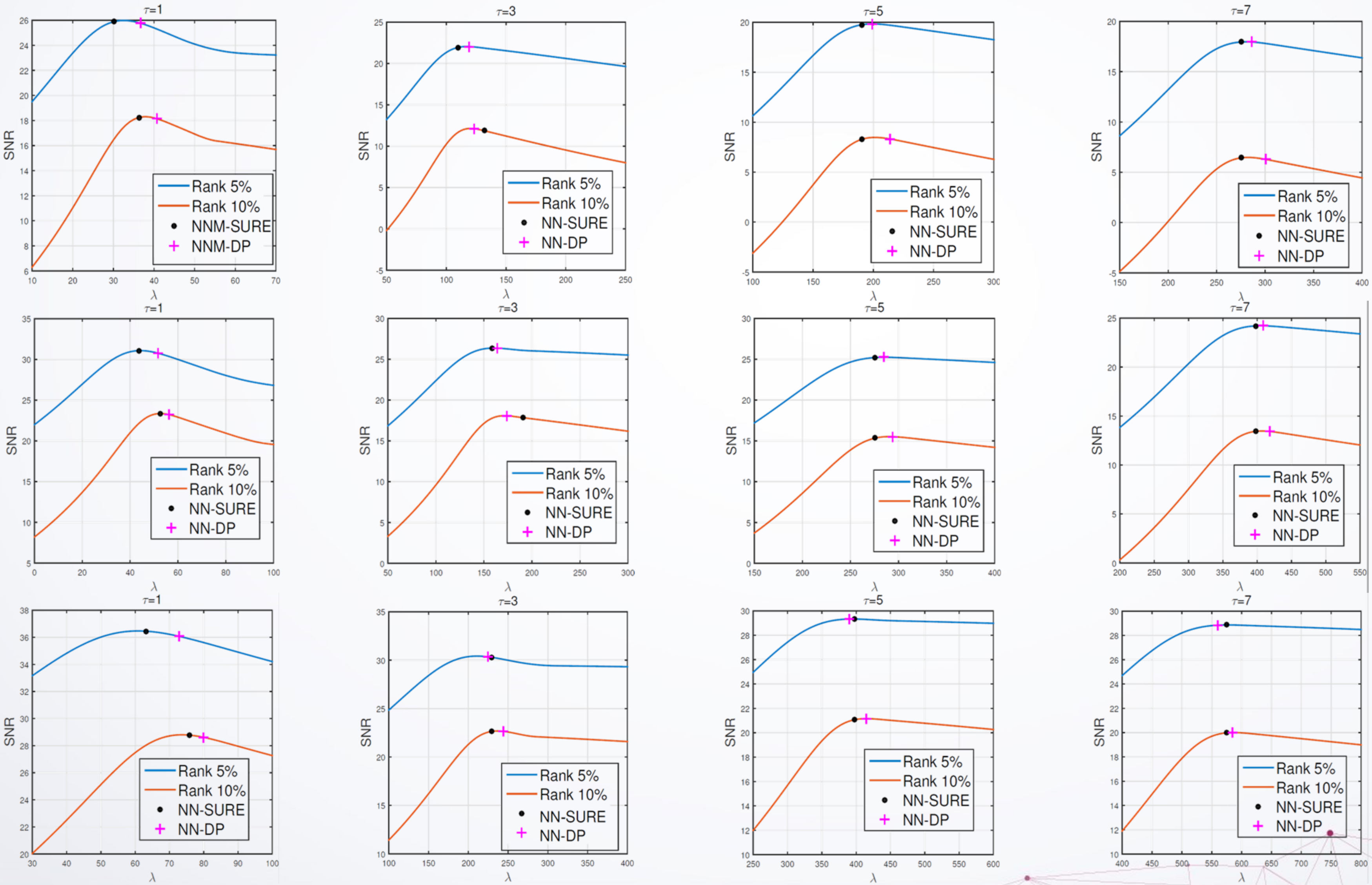
$$\min_X \|X\|_* + \frac{1}{2\lambda}\|X - Y\|_F^2$$

is given by minimizing

$$\text{SURE}(\lambda) = -mn\tau^2 + \sum_{i=1}^{m} \min\left(\lambda^2, \sigma_{Y,i}^2\right) + 2\tau^2 \operatorname{div}_Y\left(\text{SVT}_\lambda(Y)\right)$$

over 101 equi-spaced λ's in $\log[10^{-1}, 10^7]$.

Regularization parameter selection (NNM-DP VS. NNM-SURE)



$m = n = 500$

$m = n = 1000$

$m = n = 2000$

SNR(dB) for different methods with the standard algorithm

		$f(X)$	Rank(X)	$\|X\|_*$		$\|X\|_r (r=1)$		$\|X\|_{\mathbf{w},*}$	
$m=n$	τ	ρ	HardT	NN-SURE	NN-DP	TNN-SURE	TNN-DP	GWNN-SURE	GWNN-DP
500	3	1%	14.47	11.78	*11.96*	***14.47***	***14.47***	*13.49*	13.05
		5%	**22.99**	21.84	*21.94*	***22.99***	***22.99***	*22.82*	*22.82*
		10%	26.25	25.69	*25.72*	26.25	***26.29***	*26.22*	26.21
		20%	29.43	*29.30*	*29.30*	***29.59***	***29.59***	29.51	*29.54*
		40%	32.48	32.72	*32.75*	32.84	***32.89***	*32.84*	*32.84*
	5	1%	**11.35**	*8.53*	8.51	***11.35***	***11.35***	*9.96*	9.31
		5%	**21.80**	19.71	*19.75*	***21.80***	21.79	*21.30*	21.26
		10%	**25.57**	24.16	*24.20*	***25.57***	25.53	*25.34*	25.32
		20%	**29.06**	28.26	*28.27*	***29.06***	29.03	*28.97*	28.94
		40%	32.29	*31.92*	31.91	***32.31***	***32.31***	*32.27*	32.26
Average SNR			24.57	23.39	*23.43*	***24.62***	***24.62***	*24.27*	24.16
Average time (sec)			0.054	1.124	0.059	1.178	0.060	1.144	0.058
1000	3	1%	19.22	17.69	*17.94*	19.22	***19.25***	18.90	*18.97*
		5%	26.46	*26.34*	*26.34*	26.64	***26.65***	26.61	*26.62*
		10%	29.51	29.71	*29.72*	29.87	***29.88***	29.84	*29.85*
		20%	32.54	*33.14*	33.10	***33.20***	33.18	*33.16*	33.15
		40%	35.39	*36.67*	36.59	***36.71***	36.63	*36.65*	36.59
	5	1%	**17.76**	15.34	*15.48*	***17.76***	***17.76***	*16.98*	16.90
		5%	**26.08**	25.22	*25.25*	***26.08***	26.05	*25.97*	25.95
		10%	**29.32**	*28.89*	28.88	***29.32***	29.31	*29.30*	29.26
		20%	32.44	*32.29*	*32.29*	***32.52***	32.51	*32.47*	32.46
		40%	35.51	35.57	*35.63*	35.68	***35.73***	35.67	*35.69*
Average SNR			28.42	28.09	*28.12*	***28.70***	***28.70***	*28.56*	28.54
Average time (sec)			0.240	5.646	0.261	5.992	0.267	5.895	0.249

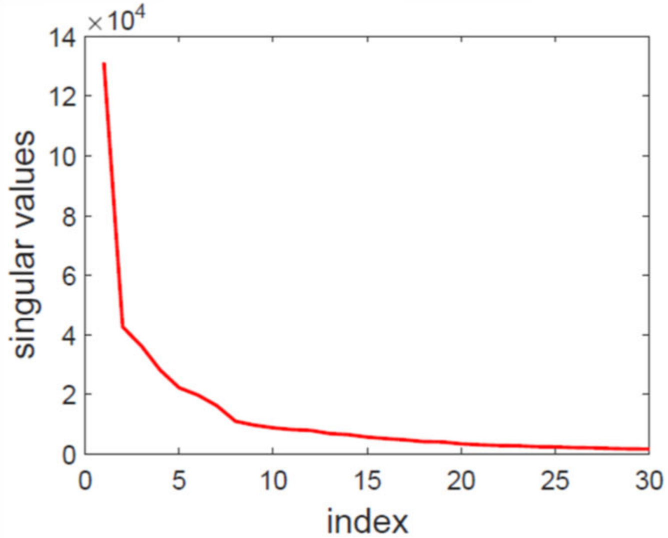
SNR(dB) for different methods with the randomized algorithm

		$f(X)$	Rank(X)	$\|X\|_*$		$\|X\|_r(r=1)$		$\|X\|_{\mathbf{w},*}$	
$m=n$	τ	ρ	HardT	NN-SURE	NN-DP	TNN-SURE	TNN-DP	GWNN-SURE	GWNN-DP
500	3	1%	**14.47**	11.78	*12.28*	***14.47***	***14.47***	*13.62*	13.15
		5%	**22.99**	21.75	*22.02*	***22.99***	22.97	22.75	*22.81*
		10%	**26.25**	25.52	*25.72*	***26.25***	***26.25***	26.15	*26.18*
		20%	29.43	29.05	*29.29*	29.43	***29.57***	29.39	*29.52*
		40%	32.48	32.19	*32.74*	32.48	***32.88***	32.47	*32.84*
	5	1%	**11.35**	8.19	*8.79*	***11.35***	***11.35***	*9.97*	9.45
		5%	**21.80**	19.39	*19.90*	***21.80***	21.79	*21.30*	21.29
		10%	**25.57**	23.90	*24.25*	***25.57***	25.52	*25.34*	25.32
		20%	**29.06**	28.04	*28.27*	***29.06***	29.03	28.93	*28.94*
		40%	32.29	31.72	*31.91*	32.29	***32.31***	32.25	*32.26*
Average SNR			24.57	23.15	*23.52*	24.57	***24.61***	*24.22*	24.18
Average time (sec)			0.005	0.116	0.006	0.123	0.007	0.117	0.006
1000	3	1%	**19.22**	17.71	*18.11*	***19.22***	19.18	18.90	*18.95*
		5%	26.46	26.00	*26.25*	26.46	***26.51***	26.43	*26.47*
		10%	29.51	29.32	*29.64*	29.51	***29.78***	29.50	*29.75*
		20%	32.54	32.53	*33.07*	32.54	***33.14***	32.54	*33.12*
		40%	35.56	35.76	*36.59*	35.82	***36.62***	35.89	*36.59*
	5	1%	**17.76**	15.56	*15.80*	***17.76***	17.75	*17.19*	16.99
		5%	**26.08**	25.04	*25.30*	***26.08***	26.01	*25.97*	25.92
		10%	**29.32**	28.72	*28.88*	***29.32***	29.28	*29.27*	29.23
		20%	32.44	32.12	*32.28*	32.44	***32.49***	32.41	*32.45*
		40%	35.51	35.35	*35.63*	35.51	***35.73***	35.51	*35.68*
Average SNR			28.44	27.81	*28.16*	28.47	***28.65***	28.36	*28.52*
Average time (sec)			0.020	0.366	0.022	0.394	0.023	0.393	0.021

CityU

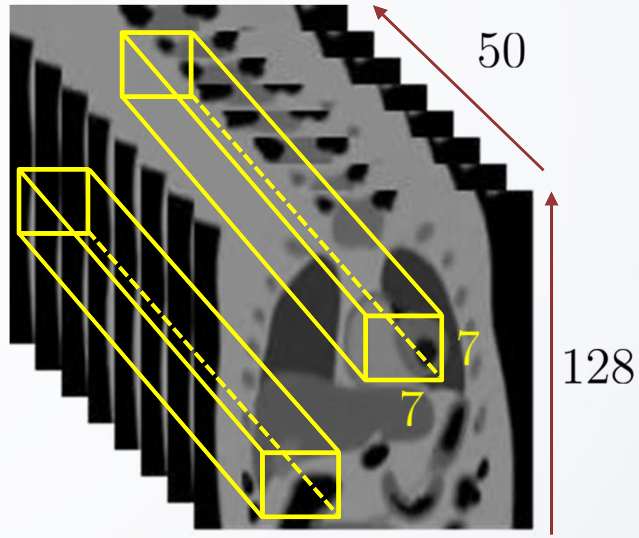
Simulations on PINCAT numerical phantom

The PINCAT phantom simulates a first-pass myocardial perfusion real-time MRI series. There are 50 images of size 128×128 taken at 50 time-steps.

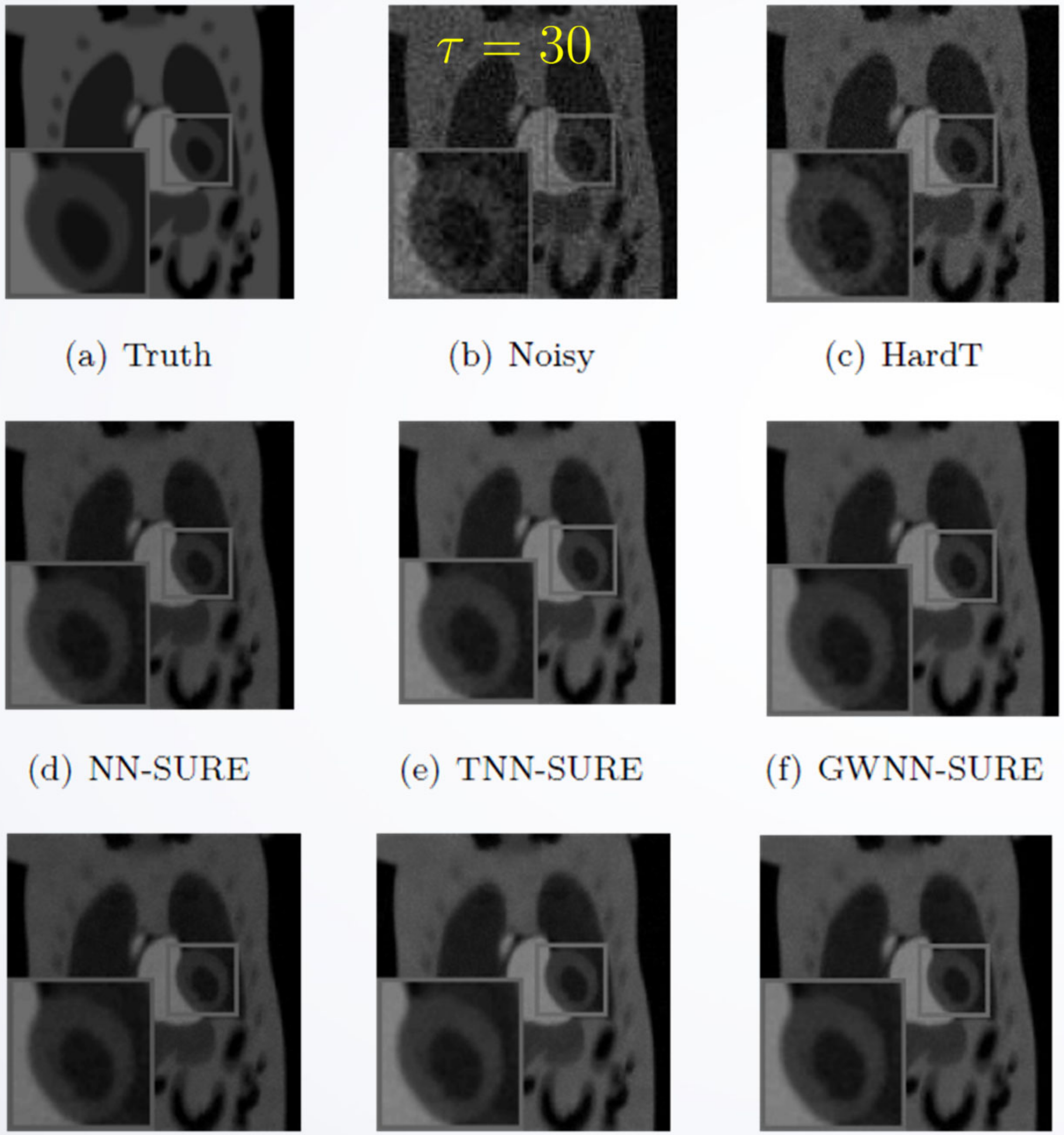


(a) Global Casorati matrix

$128^2 \times 50$ matrix



Denoised with a locally low-rank method with sliding windows [J. Trzasko, A. Manduca, and E. Borisch, 2011].



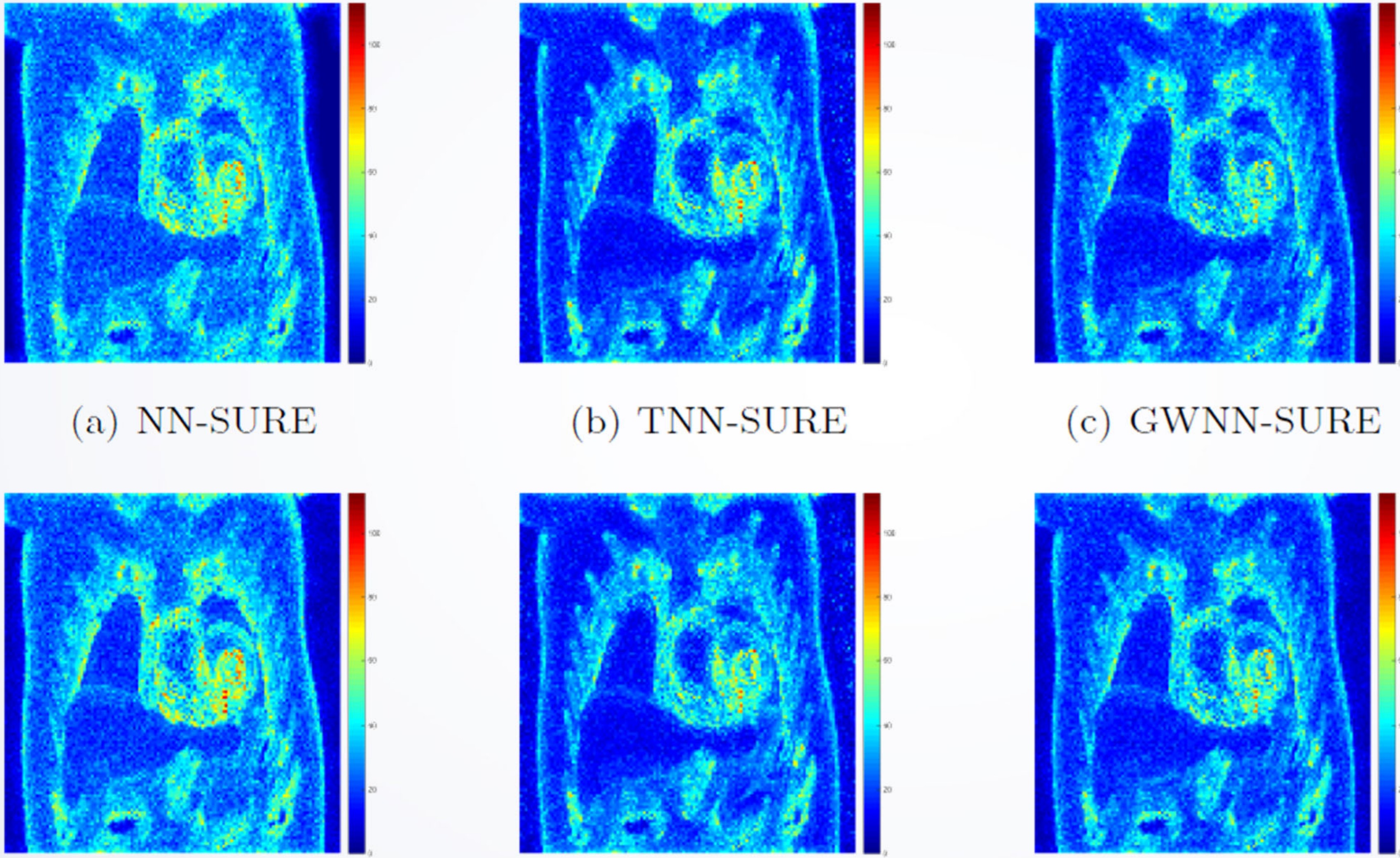
(a) Truth (b) Noisy (c) HardT

(d) NN-SURE (e) TNN-SURE (f) GWNN-SURE

(g) NN-DP (h) TNN-DP (i) GWNN-DP

Need to solve 16,384 low-rank problems of size $7^2 \times 50$

Images at $T = 10$ with an enlarged portion cropped



(a) NN-SURE (b) TNN-SURE (c) GWNN-SURE

(d) NN-DP (e) TNN-DP (f) GWNN-DP

The worst-case error through time of different methods

SNRs of different methods for the PINCAT Numerical Phantom

τ	Noisy	HardT	NN-SURE	NN-DP	TNN-SURE	TNN-DP	GWNN-SURE	GWNN-DP
5	27.50	33.19	*33.76*	33.71	34.34	*34.42*	***35.03***	35.02
10	21.48	27.60	*28.79*	28.73	29.64	*29.69*	***30.13***	***30.13***
15	17.96	24.33	*25.90*	25.85	26.89	*27.01*	***27.33***	27.31
20	15.46	22.01	*23.88*	23.83	25.09	*25.14*	***25.34***	25.31
25	13.52	20.21	22.23	*22.26*	*23.73*	23.70	***23.81***	23.76
30	11.94	18.74	*21.02*	20.98	***22.54***	22.53	***22.54***	22.48
35	10.60	17.49	19.84	*19.90*	***21.59***	21.55	*21.46*	21.40
40	9.44	16.41	*18.97*	18.96	***20.77***	20.70	*20.52*	20.46
45	8.42	15.45	18.11	*18.13*	***19.99***	19.94	*19.68*	19.62
50	7.50	14.59	17.29	*17.39*	***19.34***	19.27	*18.92*	18.88
Average	14.38	21.00	*22.98*	22.97	24.39	*24.40*	***24.48***	24.44
Average Time	–	19.01	733.34	20.12	761.91	20.06	754.24	20.81

Need to solve 16,384 low-rank problems of size $7^2 \times 50$

CityU

- Introduction
- Choosing the parameters for the regularization problems
- Applications
- **Conclusion**

Conclusion

- We considered the regularized models for various nuclear norm type regularizers.
- We derived simple formulas for the λ for these models in terms of the noise level τ.
- Our approach is competitive to the SURE methods in terms of noise removal but is much faster in terms of the CPU time.

References and Related Work

- Y.W. Wen and R.H. Chan, Parameter Selection for Total Variation Based Image Restoration Using Discrepancy Principle, by *IEEE Trans. Image Process.*, 21 (2012), 1770–1781.
- T. Teuber, G. Steidl, and R.H. Chan, Minimization and Parameter Estimation for Seminorm Regularization Models with I-Divergence Constraints, *Inverse Problems*, 29 (2013) 035007.
- Y. Wen and R. Chan, Using Generalized Cross Validation to Select Regularization Parameter for Total Variation Regularization Problems, *Inverse Problems and Imaging*, 7 (2018) 1103–1120.
- K.X. Li, H.W. Li, R.H. Chan, and Y.W. Wen, Selecting Regularization Parameters for Nuclear Norm Type Minimization Problems, accepted by *SIAM Journal on Scientific Computing.*

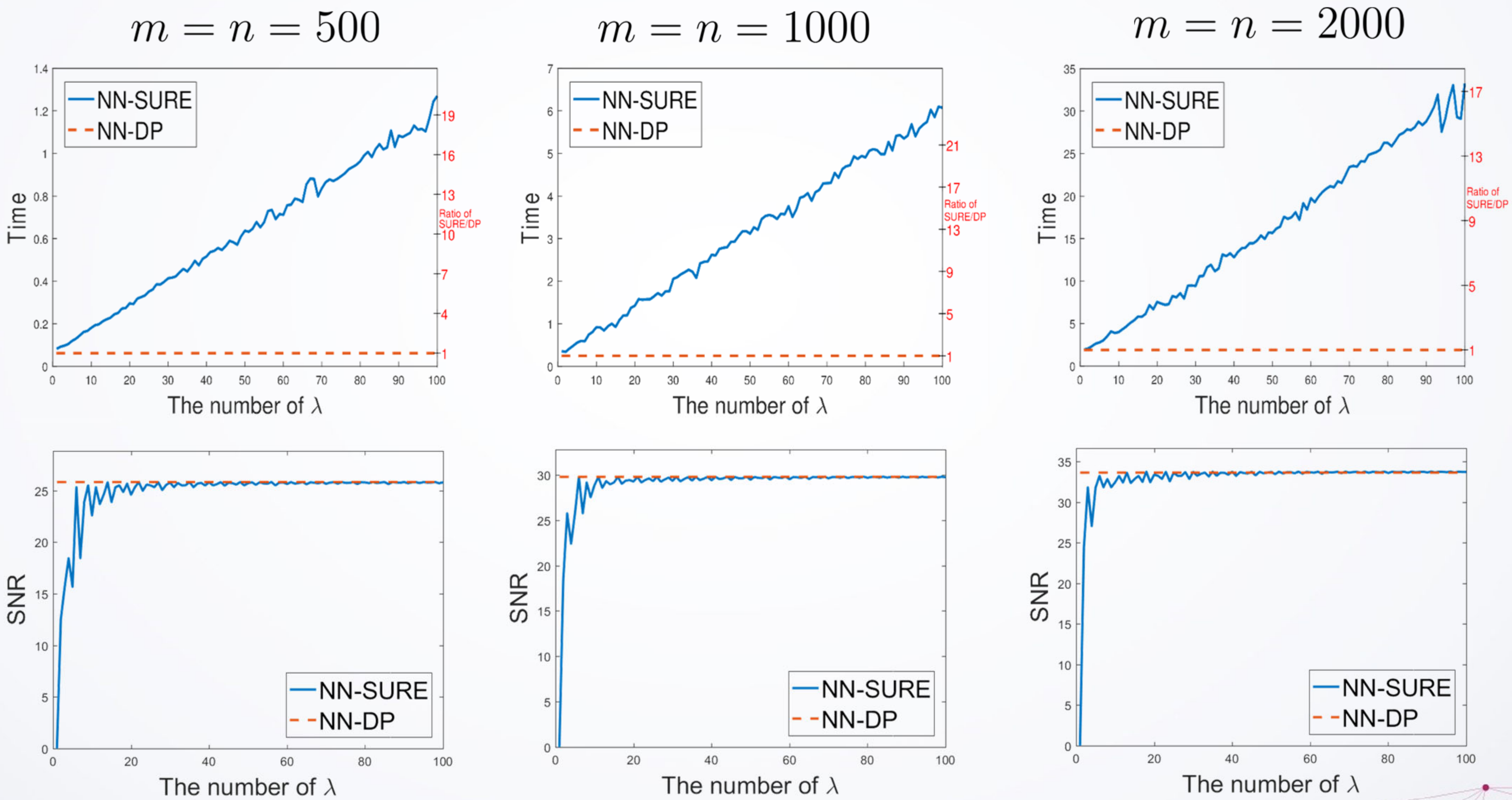
CityU

Thank you!

Timing and SNR w.r.t. number of λ's tried

$\tau = 3$

$\rho = 0.1$



40 λ's may be enough, yet our gain in speed is around 7-10 times.

How to find Q_ℓ in the randomized algorithm

How to comput matrix Q_ℓ? [N. Halko *et al.*, 2010]

- ☐ Draw a Gaussian random matrix $\Omega \in \mathbb{R}^{m\times\ell}$;
- ☐ Form the matrix $B = (Y^\top Y)^q Y^\top \Omega \in \mathbb{R}^{n\times\ell}$ ($q = 2$ is enough).
- ☐ Construct a matrix $Q_\ell \in \mathbb{R}^{n\times\ell}$ whose columns form an orthonormal basis for the range of B, e.g., the QR factorization $B = Q_\ell R$.

Rank of individual blocks in the MRI images

c)

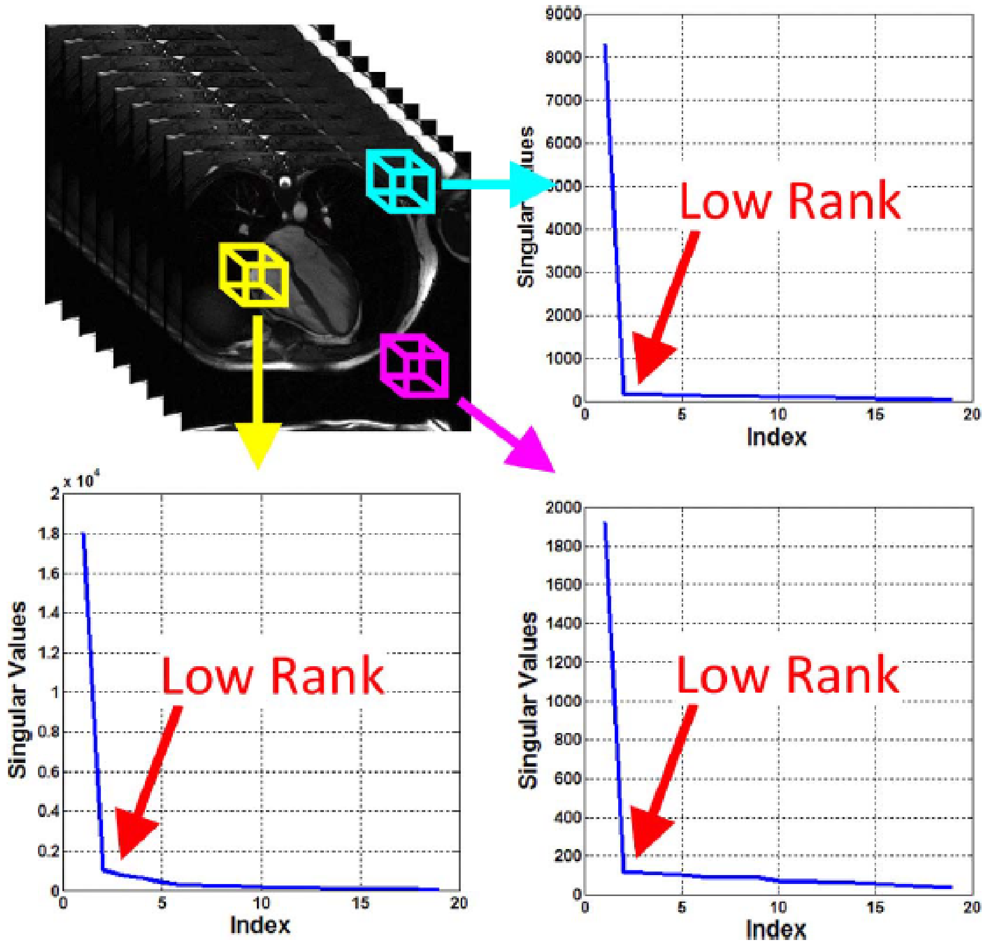


Fig. 2. Singular values of the Casorati matrices formed from three different $8 \times 8 \times 19$ block sets extracted from a cine cardiac sequence, including: a static region (cyan); a dynamic region (yellow); and background noise (magenta).

The block-wise operation in the MRI experiment is as follows:

1. We first pad the $128 \times 128 \times 50$ tensor $\mathbf{g}$, where only the bottom and right sides are filled. ($\mathbf{u}$ = padarray($\mathbf{g}$, [6 6 0], 'circular', 'post')). The dimension of $\mathbf{u}$ is $(128+6) \times (128+6) \times 50$.

2. We start to process each block, that is, transform the block with dimension $7 \times 7 \times 50$ into $7^2 \times 50$ and perform denoising processing. The step size is 1, and then perform the same processing on the next block.

3. Add the blocks together to get $\mathbf{v}$ of dimension $(128+6) \times (128+6) \times 50$, and we do the following for the pixels on the left and above to obtain $128 \times 128 \times 50$ $\mathbf{h}$:

```
h             =  v(1:128,1:128,:)
h(1:6,:,:)    =  h(1:6,:,:) + v(129:end,1:128,:)
h(:,1:6,:)    =  h(:,1:6,:) + v(1:128,129:end,:)
h(1:6,1:6,:)  =  h(1:6,1:6,:) + v(129:end,129:end,:)
```

4. In this way, each pixel value of $\mathbf{h}$ undergoes 7^2 operations, and we can take the average to get the finally denoised results ($\mathbf{h} = \mathbf{h}/7^2$).