SMOOTH HARD-THRESHOLDING FOR SINGULAR VALUES WITH STEIN'S UNBIASED RISK ESTIMATE

GUANZHONG YANG

Abstract. Low-rank matrix denoising is a central primitive in patch-based image restoration and many other inverse problems. Classical SVD-based image denoising methods often choose a truncation rank by matching residual singular-value energy with an estimated noise energy, but this rule is not a finite-sample risk principle because a fitted low-rank approximation inevitably absorbs part of the noise. This paper develops a mathematically rigorous alternative based on Stein's unbiased risk estimate (SURE). Since singular value hard thresholding is discontinuous and does not satisfy the hypotheses of Stein's lemma, we introduce a logistic smooth hard-threshold spectral estimator. We prove that the smooth shrinker satisfies the regularity conditions required by a spectral-estimator version of Stein's lemma, and therefore admits an exactly unbiased fixed-threshold risk estimate under Gaussian noise. For a fixed observed matrix and a finite set of candidate thresholds separated from the observed singular values, the ordering of the fixed-threshold smooth SURE objective eventually agrees with a simple limiting score. The limiting score has the same algebraic form as the biased hard-threshold SURE formula, but here it is used only as a computational device for ranking finite candidates. Selecting the minimizing threshold is a data-adaptive tuning step; the selected SURE value should not be interpreted as an unbiased risk estimate of the finally selected estimator.

1. Introduction

Notation.

For matrices A, B of the same size, we write $\langle A, B\rangle = \mathrm{tr}(A^T B)$ and $\|A\|_F = \langle A, A\rangle^{1/2}$ for the Frobenius inner product and Frobenius norm. All expectations are taken with respect to the Gaussian noise W, while the signal matrix X is treated as deterministic.

2020 *Mathematics Subject Classification.* Primary 62H12; Secondary 62G08, 15A18, 68U10.

Key words and phrases. Stein's unbiased risk estimate, singular value thresholding, spectral estimators, low-rank matrix denoising, smooth hard thresholding.

Observation model.

Let

$$Y = X + W \in \mathbb{R}^{m\times n}, \tag{1.1}$$

where X is an unknown deterministic matrix, usually assumed to be low-rank or approximately low-rank, and $W_{ij} \overset{\text{i.i.d.}}{\sim} N(0,\tau^2)$. Write $k = \min(m,n)$ and use the thin singular value decomposition

$$\begin{gathered} Y = U\,\mathrm{diag}(\sigma_1,\dots,\sigma_k)V^T, \\ U \in \mathbb{R}^{m\times k}, \qquad V \in \mathbb{R}^{n\times k}, \qquad U^TU = V^TV = I_k. \end{gathered} \tag{1.2}$$

with $\sigma_1 \geq \cdots \geq \sigma_k \geq 0$. For continuous Gaussian noise, the singular values are distinct and positive with probability one whenever the usual full-rank condition is satisfied; throughout the main derivation we therefore work on the simple full-rank set

$$\sigma_1 > \sigma_2 > \cdots > \sigma_k > 0. \tag{1.3}$$

SVD energy matching.

SVD-based low-rank approximation has long been used in image denoising. Guo, Zhang, Zhang and Liu proposed an efficient image denoising method that groups similar patches by block matching, arranges each group as a matrix, and applies SVD truncation to exploit the resulting low-rank structure [4]. The mathematical motivation is the Eckart–Young–Mirsky theorem: keeping the largest singular values gives the best low-rank approximation in Frobenius norm. In their denoising rule, the truncation rank is selected by comparing discarded singular-value energy with the expected noise energy, approximately $mn\tau^2$.

Finite-sample bias of the residual.

This energy-matching principle is attractive but theoretically limited. If $\widehat{X} = \widehat{X}(Y)$ is fitted from the noisy observation, then

$$\begin{aligned} \mathbb{E}\left\|Y - \widehat{X}\right\|_F^2 &= \mathbb{E}\left\|X + W - \widehat{X}\right\|_F^2 \\ &= \mathbb{E}\left\|\widehat{X} - X\right\|_F^2 + \mathbb{E}\,\|W\|_F^2 - 2\mathbb{E}\left\langle \widehat{X} - X, W\right\rangle. \end{aligned} \tag{1.4}$$

A denoising estimator that adapts to Y generally fits some component of the noise, so the correlation term is not zero. Consequently, residual energy should not be treated as an exact finite-sample surrogate for the noise energy or for the mean squared error (MSE)

$$R(\widehat{X}, X) = \mathbb{E}\left\|\widehat{X}(Y) - X\right\|_F^2. \tag{1.5}$$

SURE and the hard-thresholding obstruction.

A principled way to estimate risk under Gaussian noise is Stein's unbiased risk estimate. Candès, Sing-Long and Trzasko established SURE formulae for singular value thresholding and, more generally, spectral estimators satisfying mild differentiability and integrability assumptions [3]. The hard-threshold family is a natural target in the present setting because classical SVD denoising is truncation-based: the leading singular components are kept at their observed magnitudes, while smaller components are discarded. Soft thresholding has a well-developed SURE theory, but it shrinks every retained singular value and therefore represents a different spectral shrinkage family. If the aim is to stay close to rank-truncated SVD denoising while replacing residual-energy matching by a risk-based finite-candidate tuning principle, the natural computational goal is to compare candidate thresholds, equivalently possible retained ranks, by SURE-type scores and select the lowest-scoring candidate. The corresponding hard-threshold spectral rule is

$$\widehat{X}_\lambda(Y) = U\,\mathrm{diag}(f_\lambda(\sigma_1),\dots,f_\lambda(\sigma_k))V^T, \qquad f_\lambda(x) = x\mathbf{1}\{x>\lambda\}. \tag{1.6}$$

This rule, however, contains a jump discontinuity at $x=\lambda$. Hansen emphasized that differentiability almost everywhere is not enough for Stein's lemma; in particular, singular value hard thresholding does not satisfy the relevant condition, and the corresponding hard-threshold SURE expression is not an unbiased estimate of risk [10].

Fixed-rank comparison.

There is a related but distinct fixed-rank viewpoint. For a deterministic rank h, the reduced-rank estimator that keeps exactly the largest h singular components has its own degrees-of-freedom and SURE theory; such formulae were studied by Mukherjee, Chen, Wang and Zhu in multivariate regression, and Hansen clarified that the fixed-rank reduced-rank estimator satisfies the relevant Stein conditions [7, 10]. Consequently, for the finite-candidate comparison just described, one could obtain a deterministic-rank ranking score through fixed-rank SURE. The present paper nevertheless follows the smooth-threshold route because it addresses a different question. Fixed-rank SURE begins by taking the retained rank as the parameter; the truncation problem above begins from a threshold rule and asks how a hard-threshold-looking score can arise from legitimate fixed-threshold SURE formulae despite the discontinuity of singular value hard thresholding. This observation does not remove the tuning problem, because the rank or

threshold must still be chosen. We therefore construct fixed-threshold smooth SURE formulae and then use realized finite candidate values only as plug-in scores for ranking possible retained ranks, with the data-dependence caveats made explicit below.

Contribution and organization.

The goal of this paper is therefore not to deny the biasedness of hard-threshold SURE. Instead, we introduce a smooth hard-threshold approximation, prove that it satisfies the hypotheses under which SURE is exactly unbiased for every fixed threshold, and then use a limiting finite-candidate argument to obtain a simple and computationally efficient threshold-selection rule. The resulting procedure is a SURE-based tuning strategy, not a claim that the minimized SURE value is an unbiased risk estimate after threshold selection. Section 2 defines the smooth spectral estimator and proves the regularity needed for Stein's identity. Section 3 derives the finite-candidate limiting rule. Section 4 clarifies the precise interpretation of the limiting score and the post-selection caveat. Section 5 reports numerical checks of the fixed-threshold identity and post-selection caveat, evaluates the oracle-style performance of the ranking rule, describes the complete SVD denoising pipelines, gives a paired BSD68 statistical comparison of the two SVD rank-selection rules, and then gives an illustrative fixed-realization comparison. Section 6 concludes.

2. Smooth Hard-Threshold Spectral Estimator

2.1. Estimator and candidate thresholds

Estimator.

For fixed threshold $\lambda \geq 0$ and smoothness parameter $\omega > 0$, define

$$f_{\omega,\lambda}(x) = \frac{x}{1 + \exp[-\omega(x - \lambda)]}, \qquad x \in \mathbb{R}. \tag{2.1}$$

The associated spectral estimator is

$$\widehat{X}_{\omega,\lambda}(Y) = U \operatorname{diag}(f_{\omega,\lambda}(\sigma_1), \ldots, f_{\omega,\lambda}(\sigma_k))V^T. \tag{2.2}$$

For every $x \neq \lambda$,

$$f_{\omega,\lambda}(x) \longrightarrow f_\lambda(x) = x\mathbf{1}\{x > \lambda\} \qquad \text{as } \omega \to \infty. \tag{2.3}$$

Thus $f_{\omega,\lambda}$ is a smooth approximation to hard thresholding away from the discontinuity.

Candidate thresholds.

Given a simple full-rank observation Y, define the nonzero-rank threshold candidates by

$$\mathcal{B}(Y) = \left\{ \lambda_h = \frac{\sigma_h + \sigma_{h+1}}{2} : h = 1, \dots, k-1 \right\} \cup \left\{ \lambda_k = \frac{\sigma_k}{2} \right\}. \quad (2.4)$$

For the zero-rank candidate, fix any realized numerical threshold $\lambda_0 > \sigma_1$, equivalently the candidate that discards all singular components, and write

$$\mathcal{B}_0(Y) = \{\lambda_0\} \cup \mathcal{B}(Y).$$

For λ_h with $h \geq 1$, exactly the first h singular values are retained by the limiting hard-threshold rule.

Remark 2.1 (Data-dependent candidate values)**.** The thresholds in $\mathcal{B}_0(Y)$ are computed from the observed singular values and are therefore random before Y is observed. The fixed-threshold SURE identity below applies only to deterministic thresholds. Thus, if $\lambda_h(Y)$ is regarded as part of the estimator $Y \mapsto \widehat{X}_{\omega,\lambda_h(Y)}(Y)$, the fixed-threshold divergence formula does not by itself give an unbiased risk estimate for that data-dependent estimator. In the ranking step below, we use the realized numbers $\lambda_h(Y)$ only as plug-in candidate values at the fixed observed matrix and compare the corresponding fixed-threshold SURE formula values. No unbiasedness claim is made for the random scores $Y \mapsto \mathrm{SURE}_{\omega,\lambda_h(Y)}(Y)$.

2.2. Regularity and Unbiased SURE for Fixed Threshold

Regularity of estimator.

Lemma 2.2 (Global Lipschitz continuity of the scalar shrinker)**.** *For every $\omega > 0$ and $\lambda \geq 0$, the function $f_{\omega,\lambda}$ is globally Lipschitz on $\mathbb{R}$.*

Proof. Derivative bound. The derivative is

$$f'_{\omega,\lambda}(x) = \frac{1}{1 + e^{-\omega(x-\lambda)}} + \frac{\omega x e^{-\omega(x-\lambda)}}{(1 + e^{-\omega(x-\lambda)})^2}. \quad (2.5)$$

Compact-tail decomposition. This derivative is continuous on $\mathbb{R}$. Moreover,

$$\lim_{x\to\infty} f'_{\omega,\lambda}(x) = 1, \qquad \lim_{x\to-\infty} f'_{\omega,\lambda}(x) = 0. \quad (2.6)$$

Hence there exists $M > 0$ such that $|f'_{\omega,\lambda}(x)| \leq 2$ for $|x| > M$. On the compact interval $[-M, M]$, continuity gives $\sup_{|x|\leq M} |f'_{\omega,\lambda}(x)| < \infty$. Therefore $\sup_{x\in\mathbb{R}} |f'_{\omega,\lambda}(x)| < \infty$, and the mean value theorem implies global Lipschitz continuity. □

Lemma 2.3 (Regularity of the spectral estimator)**.** *For every fixed $\omega > 0$ and $\lambda \geq 0$, the map $Y \mapsto \widehat{X}_{\omega,\lambda}(Y)$ is globally Lipschitz as a matrix-valued spectral function and satisfies $\widehat{X}_{\omega,\lambda}(0) = 0$.*

Proof. Origin condition. The identity $\widehat{X}_{\omega,\lambda}(0) = 0$ is immediate from $f_{\omega,\lambda}(0) = 0/(1 + e^{\omega\lambda}) = 0$. *Spectral Lipschitz transfer.* By Lemma 2.2, the restriction of $f_{\omega,\lambda}$ to $[0, \infty)$ is Lipschitz and satisfies $f_{\omega,\lambda}(0) = 0$. We use the operator-Lipschitz estimate for the singular value functional calculus of Andersson, Carlsson and Perfekt [2]: if $f : [0, \infty) \to \mathbb{R}$ is Lipschitz and $f(0) = 0$, then the singular-value map

$$G_f(Y) = U \operatorname{diag}(f(\sigma_1(Y)), \ldots, f(\sigma_k(Y)))V^T \tag{2.7}$$

is Lipschitz with respect to the Frobenius norm. Hence $Y \mapsto \widehat{X}_{\omega,\lambda}(Y)$ is globally Lipschitz. □

Unbiased SURE identity.

Proposition 2.4 (Unbiased SURE for smooth hard thresholding)**.** *Assume $Y = X + W$ with independent $N(0, \tau^2)$ entries. For fixed $\omega > 0$ and $\lambda \geq 0$, the SURE functional*

$$\mathrm{SURE}_{\omega,\lambda}(Y) = -mn\tau^2 + \left\|\widehat{X}_{\omega,\lambda}(Y) - Y\right\|_F^2 + 2\tau^2 \operatorname{div}(\widehat{X}_{\omega,\lambda})(Y) \tag{2.8}$$

satisfies

$$\mathbb{E}\,\mathrm{SURE}_{\omega,\lambda}(Y) = \mathbb{E}\left\|\widehat{X}_{\omega,\lambda}(Y) - X\right\|_F^2. \tag{2.9}$$

Proof. Application of Stein's identity. Candès, Sing-Long and Trzasko show that Lipschitz spectral estimators with the required integrability are weakly differentiable and satisfy Stein's identity [3]. Hansen also stresses that global Lipschitz continuity is a sufficient condition for Stein's lemma in the matrix setting, after identifying $\mathbb{R}^{m\times n}$ with $\mathbb{R}^{mn}$ [10]. Lemma 2.3 gives this Lipschitz condition and the origin condition. To make the integrability explicit, let $L_{\omega,\lambda} < \infty$ be a Lipschitz constant for $Y \mapsto \widehat{X}_{\omega,\lambda}(Y)$. Since $\widehat{X}_{\omega,\lambda}(0) = 0$,

$$\left\|\widehat{X}_{\omega,\lambda}(Y)\right\|_F \leq L_{\omega,\lambda} \|Y\|_F, \tag{2.10}$$

and therefore $\mathbb{E}\left\|\widehat{X}_{\omega,\lambda}(Y) - Y\right\|_F^2 < \infty$ under Gaussian noise. By Rademacher's theorem the Lipschitz map is differentiable almost everywhere and weakly differentiable; its weak Jacobian is essentially bounded by $L_{\omega,\lambda}$, so

$$|\operatorname{div}(\widehat{X}_{\omega,\lambda})(Y)| \leq mnL_{\omega,\lambda} \quad \text{for almost every } Y. \tag{2.11}$$

Thus $\mathbb{E}|\operatorname{div}(\widehat{X}_{\omega,\lambda})(Y)| < \infty$. The hypotheses of the Gaussian Stein identity are therefore satisfied, and (2.8) is an unbiased estimate of the fixed-threshold risk. □

Closed-form divergence.

For a simple full-rank matrix, the divergence formula for the real spectral estimator

$$G_f(Y) = U \operatorname{diag}(f(\sigma_1), \dots, f(\sigma_k))V^T$$

is

$$\operatorname{div}(G_f)(Y) = |m-n| \sum_{i=1}^{k} \frac{f(\sigma_i)}{\sigma_i} + \sum_{i=1}^{k} f'(\sigma_i) + 2 \sum_{\substack{i,j=1 \\ i \neq j}}^{k} \frac{\sigma_i f(\sigma_i)}{\sigma_i^2 - \sigma_j^2}, \quad (2.12)$$

which is the real-valued divergence formula for spectral estimators derived in [3]. Substituting $f = f_{\omega,\lambda}$ gives

$$\begin{aligned} \operatorname{SURE}_{\omega,\lambda}(Y) = & -mn\tau^2 + \sum_{i=1}^{k} \big(\sigma_i - f_{\omega,\lambda}(\sigma_i)\big)^2 \\ & + 2\tau^2 \Bigg\{ |m-n| \sum_{i=1}^{k} \frac{f_{\omega,\lambda}(\sigma_i)}{\sigma_i} + \sum_{i=1}^{k} f'_{\omega,\lambda}(\sigma_i) \\ & \qquad + 2 \sum_{\substack{i,j=1 \\ i \neq j}}^{k} \frac{\sigma_i f_{\omega,\lambda}(\sigma_i)}{\sigma_i^2 - \sigma_j^2} \Bigg\}. \end{aligned} \quad (2.13)$$

3. Threshold Selection over a Finite Candidate Set

3.1. Limiting scores

We next show how the fixed-threshold smooth SURE formula can be ranked efficiently after the candidate values in $\mathcal{B}_0(Y)$ have been computed for a fixed observed matrix Y. The role of SURE here must be interpreted carefully. For any deterministic threshold $\lambda \geq 0$, Proposition 2.4 gives an unbiased risk estimate. The candidate thresholds in $\mathcal{B}_0(Y)$, however, are data-dependent before observation. Therefore, Section 3 does not assert that $\operatorname{SURE}_{\omega,\lambda_h(Y)}(Y)$ is an unbiased risk estimate for the random-threshold estimator. Instead, for the realized matrix Y, we treat each $\lambda_h(Y)$ as a numerical input to the fixed-threshold formula and compare the resulting plug-in scores. The analysis below is a deterministic pointwise ranking statement for the realized matrix:

as $\omega \to \infty$, the ordering of these plug-in scores agrees with the ordering of the limiting scores $S_h(Y)$, whenever the limiting inequalities are strict.

Definition 3.1 (Limiting candidate score)**.** Define the zero-rank score by

$$S_0(Y) = \|Y\|_F^2 - mn\tau^2. \tag{3.1}$$

For $h = 1, \dots, k$, define

$$S_h(Y) = -mn\tau^2 + \sum_{i=h+1}^{k} \sigma_i^2 + 2\tau^2 \left\{ |m-n|h + h + 2\sum_{i=1}^{h} \sum_{\substack{j=1 \\ j\neq i}}^{k} \frac{\sigma_i^2}{\sigma_i^2 - \sigma_j^2} \right\}, \tag{3.2}$$

with the convention that an empty sum is zero. The score S_0 is the pointwise limit of the fixed-threshold smooth SURE formula for any realized numerical threshold larger than σ_1; it represents the candidate that discards all singular components.

For $h = 0$, use the zero-rank candidate λ_0 introduced above. For $h \geq 1$, this expression is obtained from (2.13) by replacing $f_{\omega,\lambda_h}(\sigma_i)$ with $\sigma_i \mathbf{1}\{i \leq h\}$ and $f'_{\omega,\lambda_h}(\sigma_i)$ with $\mathbf{1}\{i \leq h\}$, which is valid pointwise as $\omega \to \infty$ because $\lambda_h \notin \{\sigma_1, \dots, \sigma_k\}$.

Lemma 3.2 (Pointwise convergence on candidates)**.** *For every $h = 0, 1, \dots, k$,*

$$\mathrm{SURE}_{\omega,\lambda_h}(Y) \longrightarrow S_h(Y) \qquad \text{as } \omega \to \infty. \tag{3.3}$$

Proof. Avoiding the discontinuity. For $h = 0$, choose any realized numerical threshold $\lambda_0 > \sigma_1$, so that all singular components are discarded in the limiting rule. For $h \geq 1$, λ_h lies strictly between adjacent singular values, or below σ_k when $h = k$. Hence every difference $\sigma_i - \lambda_h$ is nonzero. Therefore

$$f_{\omega,\lambda_h}(\sigma_i) \to \sigma_i \mathbf{1}\{i \leq h\}, \qquad f'_{\omega,\lambda_h}(\sigma_i) \to \mathbf{1}\{i \leq h\}. \tag{3.4}$$

Finite-sum limit. The formula (2.13) contains only finitely many algebraic operations and finite sums over the fixed singular values. Passing the limit through these finite sums gives (3.1) for $h = 0$ and (3.2) for $h \geq 1$. □

Theorem 3.3 (Eventual preservation of strict ordering)**.** *Let $h, \ell \in \{0, 1, \dots, k\}$. If $S_h(Y) < S_\ell(Y)$, then there exists $\Omega_{h\ell} < \infty$ such that for all $\omega > \Omega_{h\ell}$,*

$$\mathrm{SURE}_{\omega,\lambda_h}(Y) < \mathrm{SURE}_{\omega,\lambda_\ell}(Y). \tag{3.5}$$

Consequently, because the extended candidate set including $h = 0$ is finite, if h^* is the unique minimizer of $S_h(Y)$ over $h = 0, 1, \ldots, k$, then λ_{h^*} is also the minimizer of $\mathrm{SURE}_{\omega,\lambda_h}(Y)$ over the corresponding plug-in thresholds for all sufficiently large ω.

Proof. Pairwise comparison. By Lemma 3.2,

$$\mathrm{SURE}_{\omega,\lambda_h}(Y) - \mathrm{SURE}_{\omega,\lambda_\ell}(Y) \longrightarrow S_h(Y) - S_\ell(Y) < 0. \tag{3.6}$$

Finite minimization. The definition of convergence implies that the left-hand side is negative for all sufficiently large ω. The minimizer statement follows by applying this argument to the finitely many pairs (h^*, ℓ) with $\ell \neq h^*$ and taking the maximum of the corresponding thresholds $\Omega_{h^*\ell}$. □

Remark 3.4 (Ties)**.** If several $S_h(Y)$ attain the same minimum, the limiting ordering alone does not select a unique candidate. For computational parsimony one may choose the largest threshold among the minimizers, equivalently the smallest retained rank. This tie-breaking rule affects only exactly tied limiting scores. In numerical implementation, the denominators $\sigma_i^2 - \sigma_j^2$ in (3.2) should also be protected when singular values are nearly tied. A concrete rule is to set a tolerance

$$\varepsilon_{\mathrm{sv}} = c_{\mathrm{sv}}\, \epsilon_{\mathrm{mach}} \max\{\sigma_1^2, 1\}, \tag{3.7}$$

with c_{sv} a modest safety constant, and to regard a pair as numerically tied whenever $|\sigma_i^2 - \sigma_j^2| \leq \varepsilon_{\mathrm{sv}}$. When such a near tie is detected, the affected limiting-score comparisons should be treated as numerically unreliable rather than as reliable evidence for a particular global minimizer. In implementation one should either avoid splitting nearly tied singular values or fall back to a conservative rule, such as choosing the smallest retained rank among candidates whose scores remain distinguishable without near-tie denominators. Exact ties have probability zero under the continuous Gaussian model but can occur up to floating-point precision in patch groups.

3.2. Recursive ranking

The limiting scores can be compared recursively. From (3.1) and (3.2), for $h = 0, \ldots, k-1$,

$$\begin{aligned} S_{h+1}(Y) - S_h(Y) = & -\sigma_{h+1}^2 \\ & + 2\tau^2 \left\{ |m-n| + 1 + 2 \sum_{\substack{j=1 \\ j \neq h+1}}^{k} \frac{\sigma_{h+1}^2}{\sigma_{h+1}^2 - \sigma_j^2} \right\}. \end{aligned} \tag{3.8}$$

Thus all candidate scores may be ranked by computing the initial value $S_0(Y)$ and then accumulating the adjacent differences. The selected rank is

$$\widehat{h} \in \arg\min_{0\le h\le k} S_h(Y), \qquad \widehat{\lambda} = \lambda_{\widehat{h}}. \tag{3.9}$$

After the ranking step, one can then report the corresponding hard-truncated approximation $\widehat{X}_{\widehat{\lambda}}$, with the convention that $\widehat{h} = 0$ returns the zero matrix, when a computationally sharper low-rank output is desired. The fixed-threshold SURE identity remains rigorous for deterministic λ; after the data-adaptive choice $\widehat{\lambda} = \widehat{\lambda}(Y)$ has been made, the selected SURE value should be regarded as a tuning criterion rather than as an unbiased estimate of the selected estimator's risk.

4. Interpretation of the Limiting Score

Role of the limiting score.

The score (3.2) has the same algebraic form as the expression obtained by formally substituting hard thresholding into the spectral divergence formula. Hansen showed that such a hard-threshold expression is not, in general, an unbiased risk estimate for singular value hard thresholding [10]. The present use of (3.2) is therefore deliberately narrower: it is a finite-candidate ranking score obtained as the large-ω limit of valid fixed-threshold smooth SURE formulae, not a new unbiased risk identity for hard thresholding.

Relation to fixed-rank SURE.

The same limiting score is also closely related to the fixed-rank reduced-rank estimator. For deterministic h, retaining the first h singular components gives the fixed-rank estimator

$$Y \mapsto U\,\mathrm{diag}(\sigma_1, \dots, \sigma_h, 0, \dots, 0)V^T, \tag{4.1}$$

which is distinct from fixed-threshold hard thresholding as a statistical parametrization. Hansen's positive result for fixed-rank reduced-rank estimators therefore does not contradict his negative result for hard thresholding [10]. Algebraically, (3.2) agrees with the fixed-rank reduced-rank score: the retained-retained part of the double sum pairs to $h(h-1)$, giving

$$\begin{aligned} S_h(Y) = & -mn\tau^2 + \sum_{i=h+1}^{k} \sigma_i^2 \\ & + 2\tau^2 \left\{ h(|m-n|+h) + 2\sum_{i=1}^{h}\sum_{j=h+1}^{k} \frac{\sigma_i^2}{\sigma_i^2 - \sigma_j^2} \right\}. \end{aligned} \tag{4.2}$$

This identifies an algebraic connection with fixed-rank SURE, while the threshold-based derivation above gives the interpretation used in this paper.

Why use the smooth-threshold route?

The smooth-threshold construction is not intended to be a stronger alternative to the fixed-rank degrees-of-freedom theory. Rather, it answers a more specific question. Classical SVD denoising is naturally phrased as a cutoff rule: keep large observed singular values and remove small ones. Soft thresholding has a clean SURE theory, but it changes this truncation principle by shrinking every retained singular value. Hard thresholding preserves the truncation principle, but its discontinuity prevents the formal hard-threshold SURE expression from being unbiased. The smooth route gives a controlled way to stay within a threshold-parametrized family: for every deterministic λ and finite ω the estimator satisfies the Stein regularity assumptions, and the limiting finite-candidate argument then recovers the simple ranking score. Thus the mathematical contribution is modest but precise: it justifies a hard-threshold-looking ranking rule through smooth fixed-threshold SURE, rather than claiming a new unbiased SURE formula for discontinuous or data-selected hard thresholding.

SURE tuning and post-selection risk.

The construction separates three statements:

(i) for fixed deterministic λ, (2.13) is unbiased;
(ii) for realized finite candidates, the plug-in smooth scores converge to $S_h(Y)$;
(iii) strict inequalities among finitely many limiting scores are eventually preserved.

Only the first statement is an unbiased-risk identity. It applies to a deterministic threshold:

$$\mathbb{E}\,\mathrm{SURE}_{\omega,\lambda}(Y) = \mathbb{E}\left\|\widehat{X}_{\omega,\lambda}(Y) - X\right\|_F^2. \tag{4.3}$$

Once the midpoint candidates $\lambda_h(Y)$ are computed from the observed singular values, or once a threshold is selected by minimizing the same scores, this identity no longer applies automatically. The divergence of the resulting map would include the dependence of the chosen threshold on Y, and the selected score has the usual optimism of model and tuning-parameter selection [8, 9]. Thus SURE minimization is used here as a tuning rule, not as an unbiased post-selection risk estimator.

5. Empirical Evaluation

This section reports three matrix-level checks and two image-level comparisons. The matrix experiments verify fixed-threshold SURE, illustrate post-selection optimism, and compare the induced rank rule with an oracle and residual-energy matching. The image experiments insert the same rank-selection rule into the original SVD denoising pipeline. The empirical contribution is the local rank-rule replacement, not a new patch-search or aggregation architecture: the Set12 table gives fixed-realization context, while the BSD68 experiment gives a paired statistical comparison of energy matching and the SURE-modified SVD rule.

For the matrix-level experiments, eight Set12 images are normalized to $[0, 1]$, overlapping 32×32 patches are extracted with stride four, and three query patches are sampled from each image. For each query patch, normalized cross-correlation is computed against all other patches from the same image, and the top 50 most similar patches are retained for the diagnostic patch-group display; this matching step is separate from the full image-denoising patch search in Algorithm 1. In the numerical risk experiments, the clean matrix $X \in \mathbb{R}^{32\times 32}$ is the query patch itself, with no rank truncation before adding independent Gaussian noise

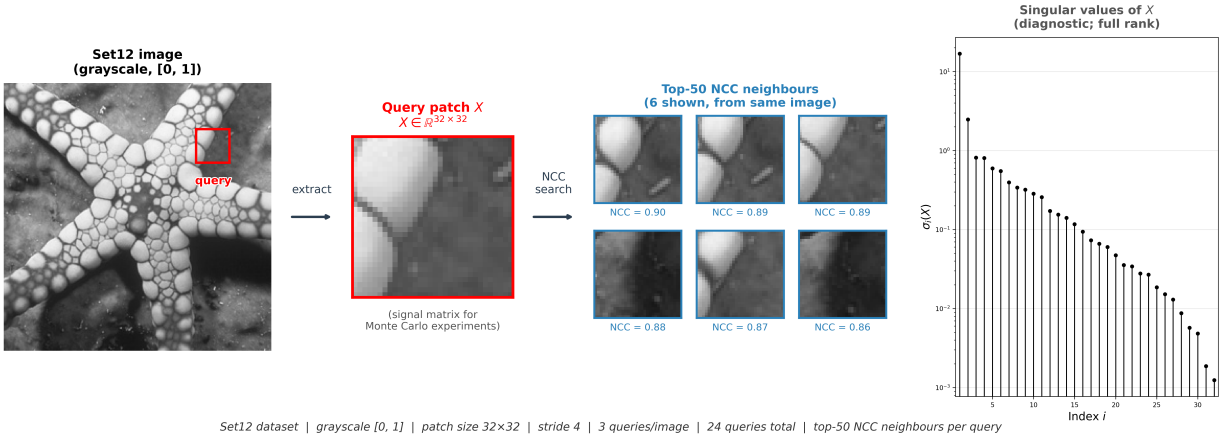
$$Y_r = X + W_r, \qquad (W_r)_{ij} \overset{\text{i.i.d.}}{\sim} N(0, \tau^2), \qquad r = 1, \dots, N. \tag{5.1}$$

Losses are measured by the Frobenius error

$$L(\widehat{X}; X, Y_r) = \left\| \widehat{X}(Y_r) - X \right\|_F^2. \tag{5.2}$$

Oracle ranks use the clean matrix X only as a benchmark and are not available to a practical denoising method. Implementation details such as random seeds, grids, tie breaking and near-tie tolerances are recorded in the replication code at `https://github.com/ECFDPB/SURE-SVD-Denoising`.

The patch-group illustration documents the data construction, but it is not one of the three validation plots. It is included here as an unnumbered setup display so that the numbering of the three main empirical figures remains unchanged.



Patch construction schematic: a query patch is extracted from a normalized grayscale image, compared with patches from the same image by normalized cross-correlation, and associated with its nearest patch neighbours. Pixel intensities are normalized in this diagnostic display to compare relative eigenvalue magnitudes; this normalization is separate from the $[0, 255]$ image-denoising pipeline. The matrix-level experiments below use the query patch itself as the clean matrix X.

5.1. Monte Carlo check of fixed-threshold unbiasedness

The first experiment tests Proposition 2.4 with fixed deterministic (ω, λ). Over independent replicates we compare

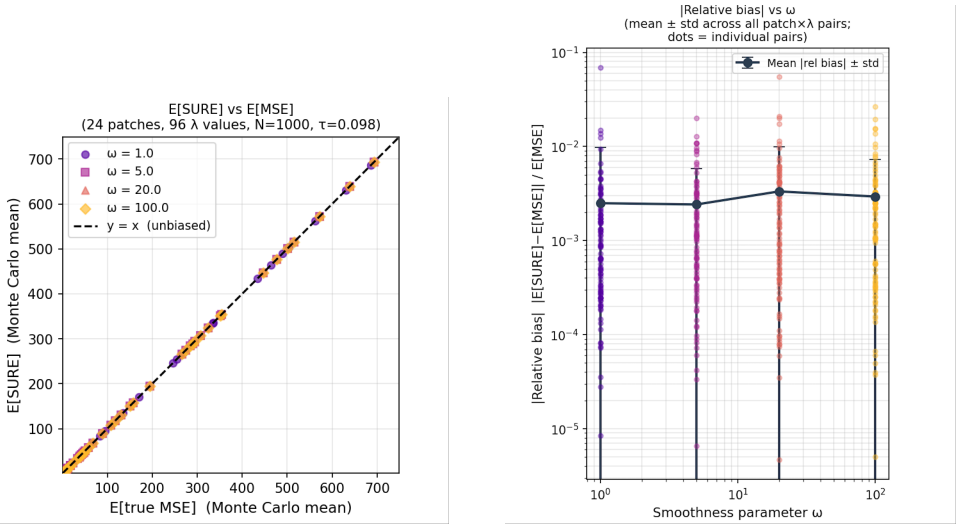
$$\widehat{R}_{\omega,\lambda} = \frac{1}{N}\sum_{r=1}^{N} \left\| \widehat{X}_{\omega,\lambda}(Y_r) - X \right\|_F^2, \qquad \widehat{S}_{\omega,\lambda} = \frac{1}{N}\sum_{r=1}^{N} \mathrm{SURE}_{\omega,\lambda}(Y_r) \tag{5.3}$$

which should agree up to Monte Carlo error.

Figure 1 shows the expected agreement. Panel (a) compares Monte Carlo mean SURE with Monte Carlo mean loss, and panel (b) reports the absolute relative discrepancy

$$\frac{|\widehat{S}_{\omega,\lambda} - \widehat{R}_{\omega,\lambda}|}{\widehat{R}_{\omega,\lambda}} \tag{5.4}$$

as a function of ω. The discrepancies remain small over the tested range of smoothing parameters, including large ω where the logistic shrinker is close to hard thresholding away from the threshold. This is consistent with the theoretical statement. The observed nonzero discrepancies are finite Monte Carlo errors, not evidence of systematic bias.



(a) Monte Carlo mean SURE versus Monte Carlo mean true loss for fixed (λ, ω).

(b) Absolute relative discrepancy as a function of the smoothness parameter ω.

FIGURE 1. Monte Carlo verification of the fixed-threshold identity in Proposition 2.4. The thresholds are deterministic in this experiment, so the setup matches the hypotheses of the SURE theorem. The agreement between the Monte Carlo mean SURE and the Monte Carlo mean true loss confirms the implementation of (2.13) up to simulation error.

5.2. Post-selection optimism and low-risk rank selection

The second experiment illustrates the caveat in Section 4. For each noisy matrix, the limiting scores are minimized over a finite rank set $\mathcal{H}$ to select rank:

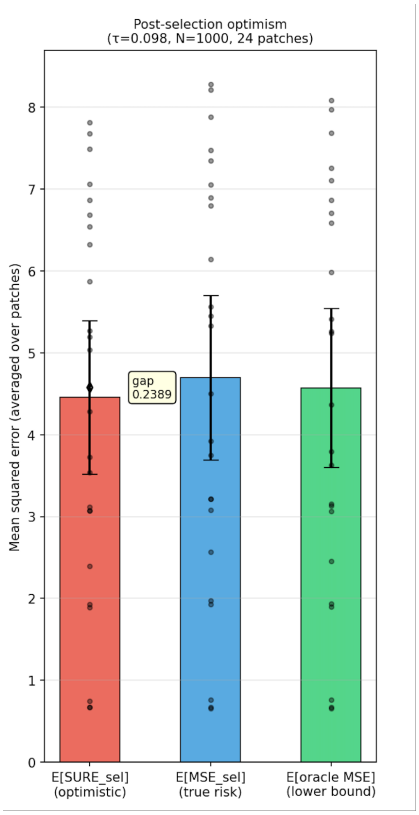
$$\widehat{h}_{\mathrm{SURE}}(Y_r) \in \arg\min_{h \in \mathcal{H}} S_h(Y_r). \tag{5.5}$$

The selected score $S_{\widehat{h}_{\mathrm{SURE}}}(Y_r)$ is compared with the true loss of the selected estimator and with the oracle rank

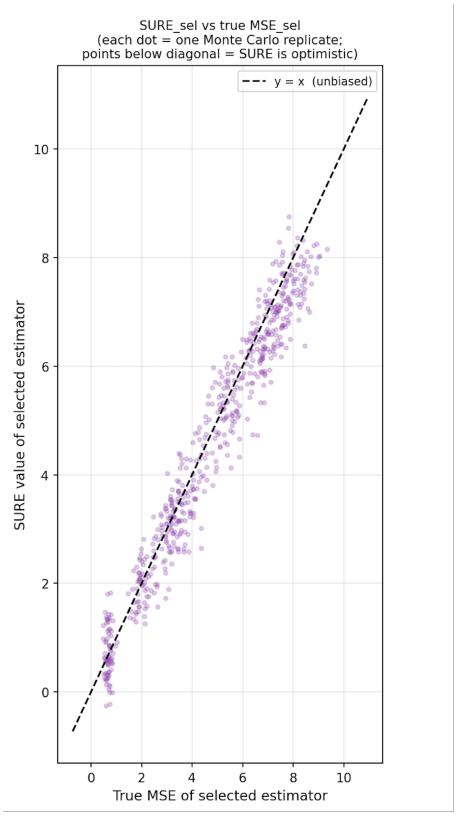
$$h_{\mathrm{or}}(Y_r) \in \arg\min_{h \in \mathcal{H}} L(\widehat{X}_h; X, Y_r), \tag{5.6}$$

where $\widehat{X}_h$ is the hard rank-h truncation. This oracle is used only to measure how much excess loss is incurred by the data-driven rank selection.

Figure 2 shows both effects: the selected SURE score is optimistic as a risk estimate, but the SURE-selected rank has true loss close to the oracle. Thus the score is used as a tuning criterion, not as an unbiased post-selection risk estimate.



(a) Selected SURE score, selected true loss, and oracle loss.



(b) Replicate-level selected SURE score versus selected true loss.

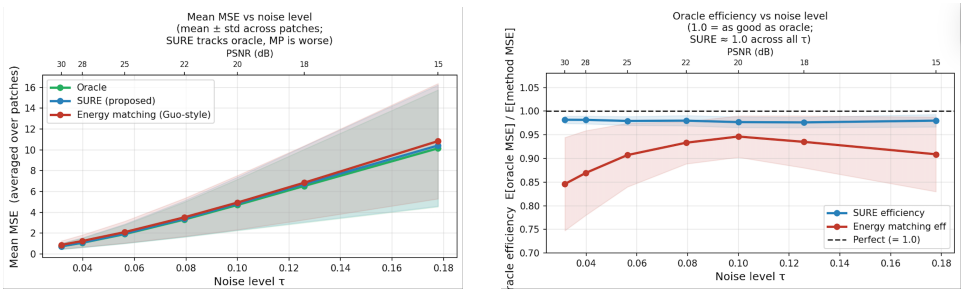
FIGURE 2. Post-selection behaviour of the plug-in ranking step. The selected SURE value is optimistic after minimization over candidate ranks, as shown by the gap between selected SURE and selected true loss and by the scatter plot lying below the diagonal. At the same time, the true loss of the SURE-selected rank is close to the oracle loss, showing that the ranking rule can still choose a low-risk rank.

5.3. Oracle-style comparison with energy matching

The third experiment compares oracle, SURE and residual-energy rank selection over several noise levels τ. The oracle rule chooses the best rank using X and is included only as a benchmark.

For a method M, define its empirical oracle efficiency at noise level τ by

$$\mathrm{Eff}_\tau(M) = \frac{\widehat{\mathbb{E}}_\tau\, L(\widehat{X}_{h_{\mathrm{or}}}; X, Y)}{\widehat{\mathbb{E}}_\tau\, L(\widehat{X}_{h_M}; X, Y)}. \tag{5.7}$$



(a) Mean MSE as a function of the noise standard deviation τ.

(b) Oracle efficiency as a function of τ.

FIGURE 3. Oracle-style comparison of rank-selection rules. The proposed SURE rule remains close to the oracle benchmark across the tested noise levels, while the residual-energy matching baseline has larger excess risk and lower oracle efficiency. The oracle is not implementable in practice; it is used only to quantify the excess loss of the data-driven rules.

Values close to one indicate near-oracle performance, while smaller values indicate larger excess risk.

Figure 3 shows that SURE closely tracks the oracle curve and has higher oracle efficiency than residual-energy matching, consistent with the degrees-of-freedom correction in the SURE score.

5.4. Complete SVD denoising pipelines

The preceding experiments isolate rank selection for one noisy matrix. The image-denoising experiment applies the same local SVD truncation primitive to patch-group matrices. In the synthetic setting one observes

$$Y = X + W, \qquad W_{ab} \overset{\text{i.i.d.}}{\sim} N(0, \tau^2), \tag{5.8}$$

computes the SVD of Y, selects a rank h, and returns the hard rank-h truncation. In the image setting this operation is applied to group matrices $P_j \in \mathbb{R}^{m \times n_j}$ formed from similar patches. This is an algorithmic transfer, not a direct application of the fixed deterministic-matrix model, because grouping, overlap, aggregation and back projection are data-dependent pipeline operations.

The original method of Guo et al. [4] uses residual-energy matching: for $P_j \in \mathbb{R}^{m \times n_j}$ with $k_j = \min(m, n_j)$ and singular values $\sigma_{j,1} \geq \cdots \geq \sigma_{j,k_j}$, it chooses the smallest rank satisfying

$$\sum_{i=h+1}^{k_j} \sigma_{j,i}^2 \leq m n_j \tau^2. \tag{5.9}$$

The SURE-modified method keeps the pipeline fixed but replaces this local rank rule by

$$\begin{aligned} S_0(P_j) &= \|P_j\|_F^2 - mn_j\tau^2, \\ S_h(P_j) &= -mn_j\tau^2 + \sum_{i=h+1}^{k_j} \sigma_{j,i}^2 \\ &\quad + 2\tau^2 \Bigg\{ |m-n_j|h + h \\ &\qquad + 2\sum_{i=1}^{h}\sum_{\substack{\ell=1 \\ \ell\neq i}}^{k_j} \frac{\sigma_{j,i}^2}{\sigma_{j,i}^2 - \sigma_{j,\ell}^2} \Bigg\}, \quad h = 1, \ldots, k_j, \end{aligned} \tag{5.10}$$

and selects

$$\widehat{h}_j \in \arg\min_{0\le h\le k_j} S_h(P_j). \tag{5.11}$$

The denoised group is the rank-$\widehat{h}_j$ hard truncation, so the only algorithmic substitution from the original pipeline to the proposed pipeline is the replacement of residual-energy matching by SURE limiting-score minimization.

Figure 4 and Algorithm 1 summarize the modified pipeline and its implementation-level steps.

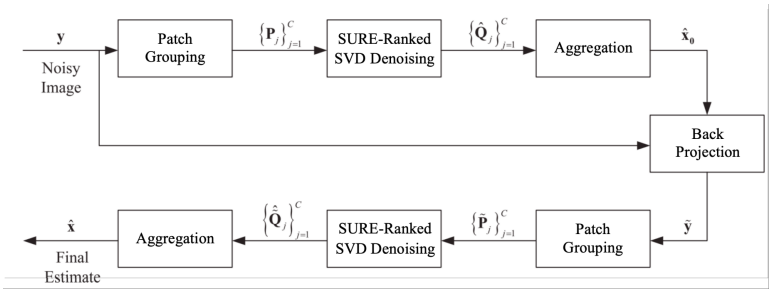


FIGURE 4. SURE-modified two-pass SVD denoising pipeline. The method uses the original patch grouping, SVD truncation, aggregation, back projection, and refinement structure; only the local rank-selection rule inside each grouped-patch SVD block is changed.

Algorithm 1. Proposed SURE-modified SVD denoising

Require: Noisy image $\mathbf{y} \in \mathbb{R}^{H\times W}$ on the 8-bit intensity scale, noise standard deviation τ on the same scale

Ensure: Denoised image $\widehat{\mathbf{x}}$

Stage 1: initial denoising.

Extract overlapping $\sqrt{m}\times\sqrt{m}$ patches from $\mathbf{y}$; the reference-patch stride selects reference locations, and the search-window half-size restricts candidates. For each reference patch, select the L closest patches by $S(\mathbf{y}_j,\mathbf{y}_c)=\|\mathbf{y}_j-\mathbf{y}_c\|_2^2$ and form $P_j=[\mathbf{y}_j,\mathbf{y}_{c,1},\dots,\mathbf{y}_{c,L}]\in\mathbb{R}^{m\times n_j}$, where $n_j=L+1$.

for each group matrix P_j, $j=1,\dots,C$ **do**

Compute $P_j=U_j\,\mathrm{diag}(\sigma_{j,1},\dots,\sigma_{j,k_j})V_j^T$, where $k_j=\min(m,n_j)$.

Choose the SURE rank $\widehat{h}_j\in\arg\min_{h\in\{0,1,\dots,k_j\}}S_h(P_j)$, where

$$S_0(P_j)=\|P_j\|_F^2-mn_j\tau^2,$$

$$S_h(P_j)=-mn_j\tau^2+\sum_{i=h+1}^{k_j}\sigma_{j,i}^2 + 2\tau^2\left\{|m-n_j|h+h+2\sum_{i=1}^{h}\sum_{\substack{\ell=1\\ \ell\neq i}}^{k_j}\frac{\sigma_{j,i}^2}{\sigma_{j,i}^2-\sigma_{j,\ell}^2}\right\},\quad h\geq 1.$$

Set $\widehat{Q}_j=U_j\,\mathrm{diag}(\sigma_{j,1},\dots,\sigma_{j,\widehat{h}_j},0,\dots,0)V_j^T$.

Set $w_j=1-\widehat{h}_j/(L+1)$ if $\widehat{h}_j<L+1$, and $w_j=1/(L+1)$ otherwise.

end for

Aggregate all $\widehat{Q}_j$ by weighted averaging to obtain $\widehat{\mathbf{x}}_0$.

Back projection: form $\widetilde{\mathbf{y}}=\widehat{\mathbf{x}}_0+\delta(\mathbf{y}-\widehat{\mathbf{x}}_0)$ with $\delta=0.5$, and update $\widetilde{\tau}=\gamma\sqrt{\max\{\tau^2-\|\widetilde{\mathbf{y}}-\widehat{\mathbf{x}}_0\|_F^2/(HW),0\}}$, where γ is a scaling factor.

Stage 2: refinement. Repeat Stage 1 on $\widetilde{\mathbf{y}}$ with noise level $\widetilde{\tau}$ to obtain $\widehat{\mathbf{x}}$.

Parameters: patch size 9×9 if $\tau<20$, 10×10 if $20\leq\tau<40$, and 11×11 if $\tau\geq 40$; these thresholds use the 8-bit scale; similar patches $L=85$; search-window half-size 35; reference-patch stride 3.

Remark: The only modification relative to Guo et al. [4] is the rank-selection rule: energy matching is replaced by SURE-based limiting-score minimization. Patch grouping, SVD factorisation, truncation, aggregation and back projection remain identical. The branch $w_j=1/(L+1)$ for a full-rank group is the positive fallback aggregation weight inherited from the original LRA-SVD rule, rather than the continuous extension $1-\widehat{h}_j/(L+1)=0$.

TABLE 1. Paired BSD68 comparison between energy matching and the SURE-modified SVD pipeline using one recorded noisy realization per image–noise pair. Here Δ denotes SURE minus energy matching, and the p-values are one-sided Wilcoxon signed-rank tests for a positive signed-rank location shift at each noise level. Runtime entries are mean wall-clock seconds per image in this implementation.

σ	N	ΔPSNR	PSNR win	p_{PSNR}	ΔSSIM	SSIM win	p_{SSIM}	Time EM/SURE
10	68	-0.517	4%	1.000	-0.0068	10%	1.000	77.8/78.5
30	68	-0.050	44%	0.978	-0.0016	44%	0.992	83.8/84.8
50	68	+0.118	65%	0.041	+0.0129	68%	0.042	89.3/90.6

5.5. Paired BSD68 comparison and Wilcoxon tests

To test whether the fixed-realization pattern is stable across more images, we also compare only the two controlled SVD pipelines on BSD68. For each image and noise level in the recorded result file, the same noisy realization is reused by both rank-selection rules. Thus the paired unit is the image at a fixed noise level, and the comparison isolates the effect of replacing residual-energy matching by SURE ranking. For PSNR and SSIM we apply a one-sided Wilcoxon signed-rank test to the paired differences

$$\Delta = \text{SURE-modified SVD} - \text{energy-matching SVD}$$

at each noise level, testing for a positive signed-rank location shift of the paired-difference distribution.

Table 1 shows that the SURE modification is not uniformly beneficial. At the low-noise setting $\sigma = 10$, the paired differences are negative, so the one-sided tests give no evidence for a SURE advantage. At $\sigma = 30$ the two rules are close, with slightly negative average differences. At $\sigma = 50$, however, SURE gives positive average gains in both PSNR and SSIM, with win rates of 65% and 68% and nominal one-sided Wilcoxon p-values below 0.05. The runtime overhead is small in this implementation: averaged over the reported BSD68 runs, the SURE-modified pipeline takes 84.65 seconds per image versus 83.60 seconds for energy matching, about a 1.3% increase. These results support a noise-regime interpretation: the SURE correction is most useful in the high-noise setting, rather than a uniformly dominant replacement for energy matching.

5.6. Illustrative comparison with SVD-related methods

We finally give an illustrative fixed-realization comparison involving K-SVD [5], LPG-PCA [6], the original energy-matching SVD pipeline, and the SURE-modified SVD pipeline on Set12. The displayed methods are all connected to the singular-value-decomposition viewpoint: K-SVD uses SVD in its dictionary-update step, LPG-PCA uses local PCA, and the two SVD pipelines use local rank-truncated SVD. The tested 8-bit noise levels are

$$\sigma \in \{10, 30, 50\}. \tag{5.12}$$

These three levels are used as representative low, moderate, and high noise regimes. For image index i, the noisy image is generated after setting

$$\mathrm{rng}(100i + \sigma), \qquad y = x + \sigma Z, \qquad Z_{ab} \overset{\text{i.i.d.}}{\sim} N(0, 1). \tag{5.13}$$

For each image–noise pair, all methods use the same single noisy realization. K-SVD uses Ron Rubinstein's MATLAB implementation (`ksvdbox13`, `ompbox10`) with parameters

$$\begin{gathered}\texttt{blocksize} = 8, \quad \texttt{dictsize} = 256, \quad \texttt{sigma} = \sigma, \quad \texttt{maxval} = 255, \\ \texttt{trainnum} = 40000, \quad \texttt{iternum} = 10, \quad \texttt{memusage} = \texttt{'high'}.\end{gathered} \tag{5.14}$$

LPG-PCA uses the official MATLAB code of Zhang et al. with profile `'fast'` and $K = 0$. The energy-matching and SURE columns are the controlled comparison of interest: they use identical Python pipeline code and differ only in the rank-selection function. K-SVD and LPG-PCA are included only to provide context for the fixed-realization image-denoising scale within the family of SVD- or PCA-related methods. The shared parameters of the two SVD pipelines are patch size 9×9, 10×10, or 11×11 according to noise level; $L = 85$ similar patches; search-window half-size 35; reference-patch stride 3; back-projection parameter $\delta = 0.5$; and, in the experiments, noise-update scaling factor $\gamma = 0.65$. The full implementation and scripts are available at `https://github.com/ECFDPB/SURE-SVD-Denoising`.

PSNR is computed on the full image on the 8-bit scale. SSIM uses an 11×11 Gaussian window with standard deviation 1.5, constants

$$C_1 = (0.01 \cdot 255)^2, \qquad C_2 = (0.03 \cdot 255)^2, \tag{5.15}$$

and the spatial average of the SSIM map, with equivalent MATLAB and Python convolution implementations.

TABLE 2. Illustrative fixed-realization PSNR and SSIM comparison on Set12 under additive white Gaussian noise. Noise levels are reported on the 8-bit intensity scale. The controlled comparison of interest is between the energy-matching and SURE-modified SVD columns; K-SVD and LPG-PCA are included as SVD- or PCA-related reference points. Boldface marks the best PSNR and the best SSIM separately among the displayed methods in each image–noise row, with ties included.

Image	σ	K-SVD		LPG-PCA		Energy matching		SURE proposed	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Cameraman	10	33.64	0.9292	33.66	0.9258	**34.08**	**0.9357**	33.65	0.9345
	30	28.01	0.8242	27.94	0.7958	**28.26**	**0.8275**	**28.26**	0.8237
	50	25.53	0.7543	25.47	0.7009	25.32	0.7017	**25.76**	**0.7684**
House	10	36.01	0.9116	36.28	0.9162	36.57	0.9178	**36.61**	**0.9190**
	30	31.26	0.8403	31.17	0.8106	31.81	0.8534	**32.26**	**0.8584**
	50	28.05	0.7780	28.27	0.7151	28.65	0.8058	**29.59**	**0.8203**
Pepper	10	34.19	0.9278	34.06	0.9219	34.56	0.9314	**34.64**	**0.9320**
	30	28.67	0.8443	28.40	0.8139	28.67	0.8481	**28.94**	**0.8529**
	50	26.15	0.7851	25.72	0.7191	25.79	0.7881	**26.42**	**0.7992**
Fishstar	10	33.00	0.9306	33.17	0.9284	**33.57**	**0.9340**	33.11	0.9295
	30	27.23	0.8153	27.31	0.8034	27.53	0.8224	**27.68**	**0.8274**
	50	24.28	0.7195	24.41	0.6954	24.32	0.7225	**24.53**	**0.7303**
Monarch	10	33.75	0.9527	33.99	0.9526	**34.59**	**0.9612**	34.31	0.9598
	30	27.80	0.8759	27.76	0.8582	28.08	0.8828	**28.35**	**0.8929**
	50	25.17	0.8051	24.96	0.7610	25.14	0.8090	**25.58**	**0.8225**
Airplane	10	33.00	0.9249	33.00	0.9202	**33.26**	**0.9283**	32.65	0.9255
	30	27.19	0.8336	27.10	0.7997	**27.38**	**0.8391**	27.32	0.8378
	50	24.56	0.7606	24.55	0.6900	24.62	0.7674	**24.73**	**0.7682**
Parrot	10	33.18	0.9281	33.42	0.9261	**33.65**	**0.9336**	33.10	0.9292
	30	27.54	0.8259	27.68	0.8020	27.97	0.8316	**27.99**	**0.8324**
	50	25.40	0.7704	25.29	0.7095	25.77	0.7783	**25.82**	**0.7810**
Barbara	10	34.39	0.9361	35.02	0.9404	**35.41**	**0.9446**	35.29	0.9444
	30	28.56	0.8266	29.24	0.8411	29.62	0.8619	**29.98**	**0.8729**
	50	25.38	0.7170	26.41	0.7373	26.36	0.7654	**26.94**	**0.7827**
Ship	10	33.64	0.8868	33.72	0.8880	**33.93**	**0.8896**	33.74	0.8857
	30	28.34	0.7472	28.27	0.7393	28.32	0.7474	**28.46**	**0.7524**
	50	25.88	0.6656	25.87	0.6340	25.71	0.6643	**26.05**	**0.6744**
Man	10	33.54	0.9018	33.71	0.9025	**33.96**	**0.9077**	33.71	0.9058
	30	28.17	**0.7525**	28.18	0.7413	28.16	0.7506	**28.19**	0.7506
	50	26.02	0.6697	26.05	0.6356	25.87	0.6695	**26.21**	**0.6809**
Couple	10	33.52	0.9004	33.62	0.9011	**33.93**	**0.9063**	33.71	0.9013
	30	27.94	0.7509	28.08	0.7542	28.18	0.7643	**28.39**	**0.7726**
	50	25.31	0.6425	25.50	0.6343	25.25	0.6504	**25.58**	**0.6665**
Average		29.39	0.8235	29.49	0.8042	29.71	0.8302	**29.81**	**0.8358**

Table 2 reports this fixed-realization comparison. K-SVD and LPG-PCA are included as SVD- or PCA-related reference points, but the controlled comparison for this paper is between the two SVD columns, where every pipeline component is shared except the rank-selection rule. In this setting, replacing residual-energy rank selection by SURE changes the average from 29.71/0.8302 to 29.81/0.8358, with improvements over energy matching in 24 of 36 PSNR entries and 23 of 36 SSIM entries. Because these are single noisy realizations and the average gain is small, these numbers should be read as modest descriptive evidence for the rank-selection replacement, not as a claim of statistical significance or superiority over existing denoising methods. The gains are strongest at the larger tested noise levels and are not uniform at $\sigma = 10$. Figure 5 gives one qualitative high-noise example on House at $\sigma = 50$.

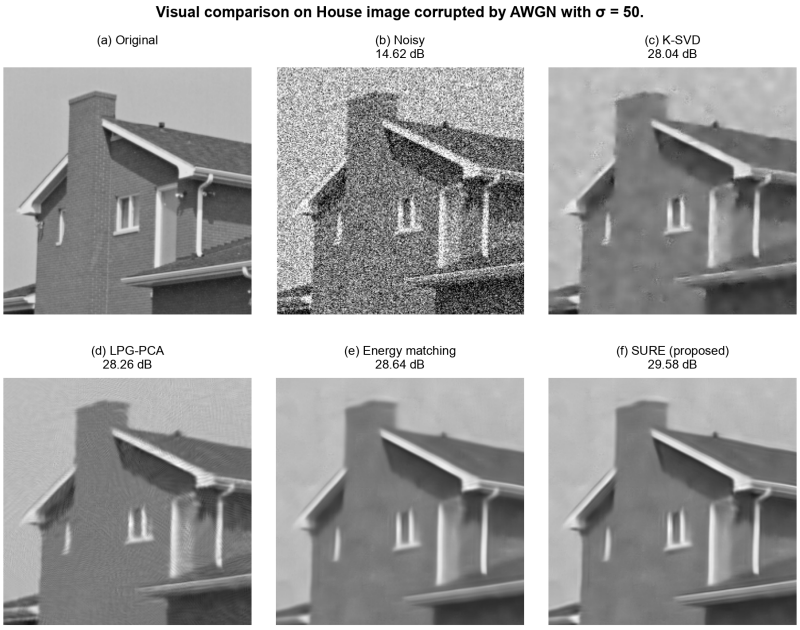


FIGURE 5. Illustrative visual comparison on the House image corrupted by additive white Gaussian noise with $\sigma = 50$. Panels show the original image, the shared noisy input, and the denoised outputs from K-SVD, LPG-PCA, energy matching, and the SURE-modified SVD pipeline. PSNR values are reported below the corresponding method labels for this single realization.

6. Conclusion and Discussion

Contribution.

This paper proposes a SURE-based threshold-ranking framework for SVD low-rank denoising. The central contribution is a reinterpretation of a biased hard-threshold SURE-like expression. We accept the negative result that singular value hard thresholding itself does not satisfy Stein's lemma [10]; nevertheless, by introducing a smooth hard-threshold approximation, we obtain a fixed-threshold estimator whose SURE is exactly unbiased under Gaussian noise. For a fixed observed matrix and a finite candidate set separated from the singular values, strict inequalities in the limiting scores are eventually preserved by the smooth SURE objective as $\omega \to \infty$. This provides a computational ranking rule for finite-candidate SURE-based threshold tuning under the separation and post-selection caveats stated above. As in other data-driven SURE tuning methods, the selected threshold should be distinguished from the fixed-threshold estimators used to construct the unbiased risk estimates.

Limitations.

The limitations are as follows. The ordering argument is finite-candidate and pointwise: thresholds must be separated from the observed singular values, and the result does not justify continuous optimization over all $\lambda \in \mathbb{R}$. The limiting score ranks candidates but is not an unbiased risk estimate for discontinuous hard thresholding or for a data-selected threshold. In the image pipeline, block matching and overlapping aggregation go beyond the fixed-matrix Gaussian model used in the proof. The Set12 table is a fixed-realization comparison, and the BSD68 Wilcoxon analysis uses one recorded noisy realization for each image–noise pair; it gives paired image-level evidence but does not quantify variability over alternative independent noise seeds, so a multiple-seed image benchmark is left for future work. The runtime measurements are implementation-dependent and are used only to indicate relative overhead.

Future work.

Future work includes multiple-seed image benchmarks, broader matrix and image tests, post-selection risk correction, extensions to other nonsmooth estimators, and a controlled continuous-threshold theory compatible with the fixed-threshold SURE interpretation.

References

[1] C. M. Stein. Estimation of the mean of a multivariate normal distribution. *The Annals of Statistics*, 9(6):1135–1151, 1981.

[2] F. Andersson, M. Carlsson, and K.-M. Perfekt. Operator-Lipschitz estimates for the singular value functional calculus. *Proceedings of the American Mathematical Society*, 144(5):1867–1875, 2016.

[3] E. J. Candès, C. A. Sing-Long, and J. D. Trzasko. Unbiased risk estimates for singular value thresholding and spectral estimators. *IEEE Transactions on Signal Processing*, 61(19):4643–4657, 2013.

[4] Q. Guo, C. Zhang, Y. Zhang, and H. Liu. An efficient SVD-based method for image denoising. *IEEE Transactions on Circuits and Systems for Video Technology*, 26(5):868–880, 2016.

[5] M. Elad and M. Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image Processing*, 15(12):3736–3745, 2006.

[6] L. Zhang, W. Dong, D. Zhang, and G. Shi. Two-stage image denoising by principal component analysis with local pixel grouping. *Pattern Recognition*, 43(4):1531–1549, 2010.

[7] A. Mukherjee, K. Chen, N. Wang, and J. Zhu. On the degrees of freedom of reduced-rank estimators in multivariate regression. *Biometrika*, 102(2):457–477, 2015.

[8] B. Efron. How biased is the apparent error rate of a prediction rule? *Journal of the American Statistical Association*, 81(394):461–470, 1986.

[9] B. Efron. The estimation of prediction error: covariance penalties and cross-validation. *Journal of the American Statistical Association*, 99(467):619–632, 2004.

[10] N. R. Hansen. A comment on Stein's unbiased risk estimate for reduced rank estimators. arXiv:1708.09657, 2017.

Department of Mathematics, Imperial College London, UK
Email address: gy625@ic.ac.uk