Noise Level Estimation for Overcomplete Dictionary Learning Based on Tight Asymptotic Bounds

Rui Chen[1(✉)] and Changshui Yang[2]

[1] Tianjin University, Tianjin, China
ruichen@tju.edu.cn
[2] Peking University, Beijing, China
csyang@pku.edu.cn

Abstract. In this paper, we address the problem of estimating Gaussian noise level from the trained dictionaries in update stage. We first provide rigorous statistical analysis on the eigenvalue distributions of a sample covariance matrix. Then we propose an interval-bounded estimator for noise variance in high dimensional setting. To this end, an effective estimation method for noise level is devised based on the boundness and asymptotic behavior of noise eigenvalue spectrum. The estimation performance of our method has been guaranteed both theoretically and empirically. The analysis and experiment results have demonstrated that the proposed algorithm can reliably infer true noise levels, and outperforms the relevant existing methods.

Keywords: Dictionary learning · Sample covariance matrix
Random matrix theory · Noise level estimation

1 Introduction

The dictionary learning is a matrix factorization problem that amounts to finding the linear combination of a given signal $\mathbf{Y} \in \mathbb{R}^{N \times M}$ with only a few atoms selected from columns of the dictionary $\mathbf{D} \in \mathbb{R}^{N \times K}$. In an overcomplete setting, the dictionary matrix $\mathbf{D}$ has more columns than rows $K > N$, and the corresponding coefficient matrix $\mathbf{X} \in \mathbb{R}^{K \times M}$ is assumed to be sparse. For most practical tasks in the presence of noise, we consider a contamination form of the measurement signal $\mathbf{Y} = \mathbf{DX} + \mathbf{w}$, where the elements of noise $\mathbf{w}$ are independent realizations from the Gaussian distribution $\mathcal{N}(0, \sigma_n^2)$. The basic dictionary learning problem is formulated as:

$$\min_{\mathbf{D}, \mathbf{X}} \|\mathbf{Y} - \mathbf{DX}\|_F^2 \quad s.t. \ \|\mathbf{x}_i\|_0 \leq L \ \forall i \tag{1}$$

Therein, L is the maximal number of non-zero elements in the coefficient vector $\mathbf{x}_i$. Starting with an initial dictionary, this minimization task can be solved by the popular alternating approaches such as the method of optimal directions (MOD) [1] and K-SVD [2]. The dictionary training on noisy samples can incorporate the denoising together

into one iterative process [3]. For a single image, the K-SVD algorithm is adopted to train a sparsifying dictionary and the developed method in [3] denoises the corrupted image by alternating between the update stages of the sparse representations and the dictionary. In general, the residual errors of learning process are determined by noise levels. Noise incursion in a trained dictionary can affect the stability and accuracy of sparse representation [4]. So the performance of dictionary learning highly depends on the estimation accuracy of unknown noise level σ_n^2 when the noise characteristics of trained dictionaries are unavailable.

The main challenge of estimating the noise level lies in effectively distinguishing the signal from noise by exploiting sufficient prior information. The most existing methods have been developed to estimate the noise level from image signals based on specific image characteristics [5–8]. Generally, these works assume that a sufficient amount of homogeneous areas or self-similarity patches are contained in natural images. Thus empirical observations, singular value decomposition (SVD) or statistical properties can be applied on carefully selected patches. However, it is not suitable for estimating the noise level in dictionary update stage because only few atoms for sparse representation cannot guarantee the usual assumptions. To enable wider applications and less assumptions, more recent methods estimate the noise level based on principal component analysis (PCA) [9, 10]. These methods underestimate the noise level since they only take the smallest eigenvalue of block covariance matrix. Although later work [11] has made efforts to tackle these problems by spanning low dimensional subspace, the optimal estimation for true noise variance is still not achieved due to the inaccuracy of subspace segmentation. As for estimating the noise variance techniques, the scaled median absolute deviation of wavelet coefficients has been widely adopted [12]. Leveraging the results from random matrix theory (RMT), the median of sample eigenvalues is also used as an estimator of noise variance [13]. However, these estimators are no longer consistent and unbiased when the dictionary matrix has high dimensional structure.

To solve the aforementioned problems, we propose to accurately estimate the noise variance in a trained dictionary by using exact eigenvalues of a sample covariance matrix. The proposed method can also be applied to estimate the noise level for the noisy image. As a novel contribution, we construct the tight asymptotic bounds of extreme eigenvalues to separate the subspaces between the signal and the noise based on random matrix theory (RTM). Moreover, in order to eliminate the possible bias caused by the high-dimensional settings, a corrected estimator is derived to provide the consistent inference on the noise variance for a trained dictionary. Based on these asymptotic results, we develop an optimal variance estimator which can well deal with the settings with different sample sizes and dimensions. The practical usefulness of our method is numerically illustrated.

2   Tight Bounds for Noise Eigenvalue Distributions

In this section, we analyze the asymptotical distribution of the ratio of extreme eigenvalues of a sample covariance matrix based on the limiting RTM law. Then a tight bound is derived.

2.1   Eigenvalue Subspaces of Sample Covariance Matrix

We consider the sparse approximation of each observed sample $\mathbf{y}_i \in \mathbb{R}^N$ with s prototype atoms selected from learned dictionary $\mathbf{D}$. With respect to the sparse model (1), we aim at estimating the noise level σ_n^2 for an elementary trained dictionary $\mathbf{D}_s$ containing a subset of the atoms $\{\mathbf{d}_i\}_{i=1}^s$. Note that $\mathbf{D}_s = \mathbf{D}_S^0 + \mathbf{w}_S$, where $\mathbf{D}_S^0$ denotes original dictionary and $\mathbf{w}_S$ is the additive Gaussian noise. At each iterative step, the noise level σ_n^2 goes gradually to zero when updating towards true dictionary $\mathbf{D}_S^0$ [14]. The known noise variance is helpful to avoid noise incursion and determine the sample size, the sparsity degree and even the performance of the true underlying dictionary [15]. To derive the relationship between the eigenvalues and noise level, we first construct the sample covariance matrix of dictionary $\mathbf{D}_s$ as follows:

$$\Sigma_S = \frac{1}{s-1} \sum_{i=1}^s (\mathbf{d}_i - \overline{\mathbf{d}})(\mathbf{d}_i - \overline{\mathbf{d}})^{\mathrm{T}}, \quad \overline{\mathbf{d}} = \frac{1}{s} \sum_{i=1}^s \mathbf{d}_i \tag{2}$$

According to (2), the square matrix $\sum_s$ has N dimensions with the sparse condition $N \gg s$. Based on the symmetric property, this matrix is decomposed into the product of three matrices: an orthogonal matrix $\mathbf{U}$, a diagonal matrix and a transpose matrix $\mathbf{U}^T$, which can be selected by satisfying $\mathbf{U}^T\mathbf{U} = \mathbf{I}$. Here, this transform process is written as:

$$\mathbf{U}^{\mathrm{T}}\Sigma_S\mathbf{U} = \mathbf{diag}(\lambda_1, \ldots, \lambda_m, \lambda_{m+1}, \ldots, \lambda_N) \tag{3}$$

Given $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_N$, we exploit the eigenvalue subspaces to enable the separation of atoms from noise. To be more specific, we divide the eigenvalues into two sets $\mathbf{S} = \mathbf{S}_1 \cup \mathbf{S}_2$ by finding the appropriate bound in a spiked population model [16]. Most structures of an atom lie in low-dimension subspace and thus the leading eigenvalues in set $\mathbf{S}_1 = \{\lambda_i\}_{i=1}^m$ are mainly contributed by atom itself. The redundant-dimension subspace $\mathbf{S}_2 = \{\lambda_i\}_{i=m+1}^N$ is dominated by the noise. Because the atoms contribute very little to this later portion, we take all the eigenvalues of $\mathbf{S}_2$ into consideration to estimate the noise variance while eliminating the influence of trained atoms. Moreover, the random variables $\{\lambda_i\}_{i=m+1}^N$ can be considered as the eigenvalues of pure noise covariance matrix $\Sigma_{\mathbf{w}}$, whose dimensions are N.

2.2   Asymptotic Bounds for Noise Eigenvalues

Suppose the sample matrix $\Sigma_{\mathbf{w}}$ has the form $(s-1)\Sigma_{\mathbf{w}} = \mathbf{HH}^{\mathrm{T}}$, where the sample entries of $\mathbf{H}$ are independently generated from the distribution $\mathcal{N}(0, \sigma_n^2)$. Then the real matrix $\mathbf{M} = \mathbf{HH}^{\mathrm{T}}$ follows a standard Wishart distribution [17]. The ordered eigenvalues of $\mathbf{M}$ are denoted by $\overline{\lambda}_{\max}(\mathbf{M}) \geq \cdots \geq \overline{\lambda}_{\min}(\mathbf{M})$. In the high dimensional situation: $N/s \to \gamma \in [0, \infty)$ as $s, N \to \infty$, the Tracy-Widom law gives the limiting distribution of the largest eigenvalue of the large random matrix $\mathbf{M}$ [18]. Then we have the following asymptotic expression:

$$\Pr\left\{\frac{\bar{\lambda}_{\max}/\sigma_n^2 - \mu}{\xi} \leq z\right\} \rightarrow F_{\text{TW1}}(z) \tag{4}$$

where $F_{\text{TW1}}(z)$ indicates the cumulative distribution function with respect to the Tracy-Widom random variable. In order to improve both the approximation accuracy and convergence rate, even only with few atom samples, we need choose the suitable centering and scaling parameters μ, ξ [19]. By the comparison between different values, such parameters are defined as

$$\begin{cases} \mu = 1/s \cdot \left(\sqrt{s-1/2} + \sqrt{N-1/2}\right)^2 \\ \xi = 1/s \cdot \left(\sqrt{s-1/2} + \sqrt{N-1/2}\right)\left(\frac{1}{\sqrt{s-1/2}} + \frac{1}{\sqrt{N-1/2}}\right)^{1/3} \end{cases} \tag{5}$$

The empirical distribution of the eigenvalues of the large sample matrix converges almost surely to the Marcenko-Pastur distribution on a finite support [20]. Based on the generalized result in [21], when $N \rightarrow \infty$ and $\gamma \in [0, \infty)$, with probability one, we derive limiting value of the smallest eigenvalue as

$$\bar{\lambda}_{\min}/\sigma_n^2 \rightarrow (1 - \sqrt{\gamma})^2 \tag{6}$$

According to the asymptotic distributions described in the theorems (4) and (6), we further quantify the distribution of the ratio of the maximum eigenvalue to minimum eigenvalue in order to detect the noise eigenvalues. Let T_1 be a detection threshold. Then we find T_1 by the following expression:

$$\begin{aligned} \Pr\left\{\frac{\bar{\lambda}_{\max}}{\bar{\lambda}_{\min}} \leq T_1\right\} &= \Pr\left\{\frac{\bar{\lambda}_{\max}}{\sigma_n^2} \leq T_1 \cdot \frac{\bar{\lambda}_{\min}}{\sigma_n^2}\right\} \approx \Pr\left\{\frac{\bar{\lambda}_{\max}}{\sigma_n^2} \leq T_1 \cdot \left(1 - \sqrt{N/s}\right)^2\right\} \\ &= \Pr\left\{\frac{\bar{\lambda}_{\max}/\sigma_n^2 - \mu}{\xi} \leq \frac{T_1 \cdot \left(1 - \sqrt{N/s}\right)^2 - \mu}{\xi}\right\} \approx F_{\text{TW1}}\left\{\frac{T_1 \cdot \left(1 - \sqrt{N/s}\right)^2 - \mu}{\xi}\right\} \end{aligned} \tag{7}$$

Note that there is no closed-form expression for the function F_{TW1}. Fortunately, the values of F_{TW1} and the inverse F_{TW1}^{-1} can be numerically computed at certain percentile points [16]. For a required detection probability α_1, this leads to

$$\frac{T_1 \cdot \left(1 - \sqrt{N/s}\right)^2 - \mu}{\xi} = F_{\text{TW1}}^{-1}(\alpha_1) \tag{8}$$

Plugging the definitions of μ and ξ into the Eq. (8), we finally obtain the threshold

$$T_1 = \frac{s\left(\sqrt{s-1/2} + \sqrt{N-1/2}\right)^2}{\left(\sqrt{s} - \sqrt{N}\right)^2} \cdot \left(\frac{\left(\sqrt{s-1/2} + \sqrt{N-1/2}\right)^{-2/3}}{(s-1/2)^{1/6}(N-1/2)^{1/6}} \cdot F_{\text{TW1}}^{-1}(\alpha_1) + 1\right) \tag{9}$$

When the detection threshold T_1 is known in the given probability, it means that an asymptotic upper bound can also be obtained for determining the noise eigenvalues of the matrix $\Sigma_{\mathbf{w}}$ because the equality $\lambda_{m+1}/\lambda_N = \overline{\lambda}_{\max}/\overline{\lambda}_{\min}$ holds. In general, the noise eigenvalues in the set $\mathbf{S}_2$ surround the true noise variance as it follows the Gaussian distribution. The estimated largest eigenvalue λ_{m+1} should be no less than σ_n^2. The known smallest eigenvalue λ_N is no more than σ_n^2 by the theoretical analysis [11]. The location and value of λ_{m+1} in $\mathbf{S}$ are obtained by checking the bound $\lambda_{m+1} \leq T_1 \cdot \lambda_N$ with high probability α_1. In addition, λ_1 cannot be selected as noise eigenvalue λ_{m+1}.

3   Noise Variance Estimation Algorithm

3.1   Bounded Estimator for Noise Variance

Without requiring the knowledge of signal, the threshold T_1 can provide good detection performance for finite s, N even when the ratio N/s is not too large. Based on this result, more accurate estimation can be obtained by averaging all elements in $\mathbf{S}_2$. Hence, the maximum likelihood estimator of σ_n^2 is

$$\hat{\sigma}_n^2 = \frac{1}{N-m} \sum_{j=m+1}^{N} \lambda_j \tag{10}$$

In the low dimensional setting where N is relatively small compared with s, the estimator $\hat{\sigma}_n^2$ is consistent and unbiased as $s \to \infty$. It follows asymptotically normal distribution as

$$\sqrt{s}\left(\hat{\sigma}_n^2 - \sigma_n^2\right) \to \mathcal{N}(0, \ t^2), \ \ t^2 = \frac{2\sigma_n^4}{N-m} \tag{11}$$

When N is large with respect to the sample size s, the sample covariance matrix shows significant deviations from the underlying population covariance matrix. In this context, the estimator $\hat{\sigma}_n^2$ might have a negative bias, which leads to overestimation of true noise variance [22, 23]. We investigate the distribution of another eigenvalue ratio. Namely, the ratio of the maximum eigenvalue to the trace of the eigenvalues is

$$U = \frac{\lambda_{m+1}}{1/(N-m) \cdot \mathrm{tr}(\Sigma_{\mathbf{w}})} = \frac{\lambda_{m+1}}{1 \big/ (N-m) \cdot \sum_{j=m+1}^{N} \lambda_j} \tag{12}$$

According to the result in (4), the ratio U also follows a Tracy-Widom distribution as both N, $s \to \infty$. The denominator in the definition of U is distributed as an independent $\sigma_n^2 \chi_N^2/N$ random variable, and thus has $\mathrm{E}(\hat{\sigma}_n^2) = \sigma_n^2$ and $\mathrm{Var}(\hat{\sigma}_n^2) = 2\sigma_n^4/(N \cdot s)$. It is easy to show that replacing σ_n^2 by $\hat{\sigma}_n^2$ results in the same limiting distribution in (4). Then we have

$$\Pr\left\{\frac{\lambda_{m+1}/\hat{\sigma}_n^2 - \mu}{\xi} \leq z\right\} \rightarrow F_{\mathrm{TW1}}(z) \tag{13}$$

Unfortunately, the asymptotic approximation present in (13) is inaccurate for small and even moderate values of N [24]. This approximation is not a proper distribution function. The simulation observations imply that the major factor contributing to the poor approximation is the asymptotic error caused by the constant ξ [24]. Therefore, a more accurate estimate for the standard deviation of $\lambda_{m+1}/\hat{\sigma}_n^2$ will provide a significant improvement. For finite samples, we have

$$\mathrm{E}\left(\frac{\lambda_{m+1}}{\sigma_n^2}\right) = \mu, \quad \mathrm{E}\left(\frac{\lambda_{m+1}^4}{\sigma_n^4}\right) = \mu^2 + \xi^2 \tag{14}$$

Using these asymptotic results, we get the corrected deviation

$$\xi' = \sqrt{\frac{N \cdot s}{2 + N \cdot s}\left(\xi^2 - \frac{2}{N \cdot s}\mu^2\right)} \tag{15}$$

Note that this formula in (15) has corrected the overestimation in the high dimensional setting. thus the better approximation for the probabilities of the ratio is

$$\Pr\left\{\frac{\lambda_{m+1}/\hat{\sigma}_n^2 - \mu}{\xi'} \geq z\right\} \approx 1 - F_{\mathrm{TW1}}(z) \tag{16}$$

The determination of the distribution for the ratio U is devoted to the correction of the variance estimator. In order to complete the detection of the large deviations of the initial estimator $\hat{\sigma}_n^2$, we provide a procedure to set the threshold T_2. Based on the result in (16), an approximate expression for the overestimation probability is given by

$$\Pr\left\{\frac{\hat{\sigma}_n^2}{\lambda_{m+1}} \leq T_2\right\} = \Pr\left\{\frac{\lambda_{m+1}/\hat{\sigma}_n^2 - \mu}{\xi'} \geq \frac{1/T_2 - \mu}{\xi'}\right\} \approx 1 - F_{\mathrm{TW1}}\left(\frac{1/T_2 - \mu}{\xi'}\right) \tag{17}$$

Hence, for a desired probability level α_2, the above equation can be numerically inverted to find the decision threshold. After some simplified manipulations, we obtain

$$T_2 = \frac{1}{\xi' \cdot F_{\mathrm{TW1}}^{-1}(1 - \alpha_2) + \mu} \tag{18}$$

Asymptotically, the spike eigenvalue λ_{m+1} converges to the right edge of the support $\sigma_n^2(1 + \sqrt{N/s})$ as N, s go to infinity. According to the expression in (18), this function turns out to have a simple approximation $T_2 = 1/\mu$ in the high probability case. Then the upper bound $T_2 \cdot \lambda_{m+1}$ for the known $\hat{\sigma}_n^2$ yields a bias estimation. Finally, the following expectation holds true:

$$E\left(\frac{\mu \cdot T_2 \cdot \lambda_{m+1}}{1 + \sqrt{N/s}}\right) \approx \sigma_n^2 \ll \hat{\sigma}_n^2 \tag{19}$$

By analyzing the statistical result in (19), the correction for $T_2 \cdot \lambda_{m+1}$ can be approximated as the better estimator than $\hat{\sigma}_n^2$ because this bias-corrected estimator is closer to the true variance under the high dimensional conditions. If $\hat{\sigma}_n^2$ can satisfy the requirement of no excess of the bound $T_2 \cdot \lambda_{m+1}$, the sample eigenvalues are consistent estimates of their population counterparts. Hence, the optimal estimator is given by

$$\hat{\sigma}_*^2 = \min\left\{\hat{\sigma}_n^2, \ \frac{\mu \cdot T_2 \cdot \lambda_{m+1}}{1 + \sqrt{N/s}}\right\} \tag{20}$$

3.2   Implementation

Based on the construction of two thresholds, we propose a noise estimation algorithm for dictionary learning as follows:

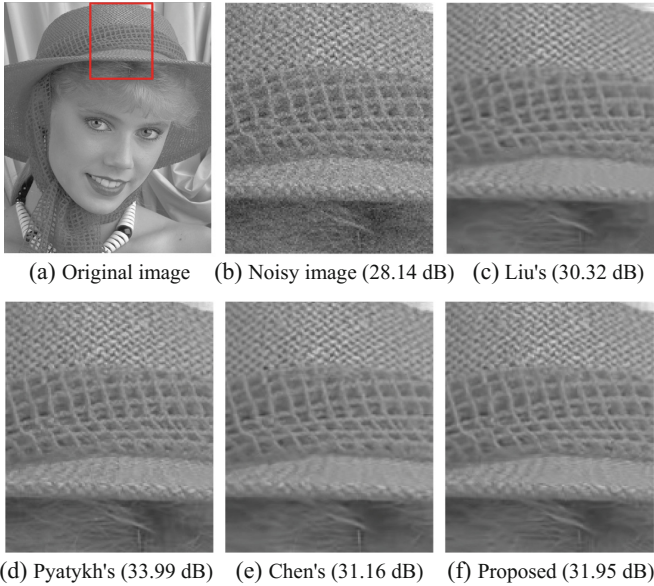
Algorithm 1 Noise Estimation for Dictionary Learning

1: **Input:** Noisy dictionary $\mathbf{D}_s$, the dimension N, the sample number s, the probability levels α_1 and α_2.

2: **Compute** the eigenvalues $\{\lambda_i\}_{i=1}^N$ of the sample covariance matrix Σ_s, and order $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_N$.

3: **Compute** two thresholds T_1 and T_2.

4: **for** $i = 1 : N-1$ **do**

   if $\lambda_{i+1} \leq T_1 \cdot \lambda_N$ **then**

      Obtain the location $m+1 = i+1$, $\lambda_{m+1} = \lambda_{i+1}$ and **break**

   end if

  end for

5: **Estimate** an initial noise variance $\hat{\sigma}_n^2$ using (10).

6: **Compare** the values of two estimators of (20) and select the minimum as an optimal estimator $\hat{\sigma}_*^2$.

7: **Output:** noise level estimation $\sigma_n^2 = \hat{\sigma}_*^2$.

---

4   Numerical Experiments

The proposed estimation method is evaluated on two benchmark datasets: Kodak [7] and TID2008 [9]. The subjective experiment is to compare our method with three state-of-the-art estimation methods by Liu et al. in [8], Pyatykh *et al.* in [9] and Chen *et al.* in [11], which are relevant in SVD domain. The testing images are added to the
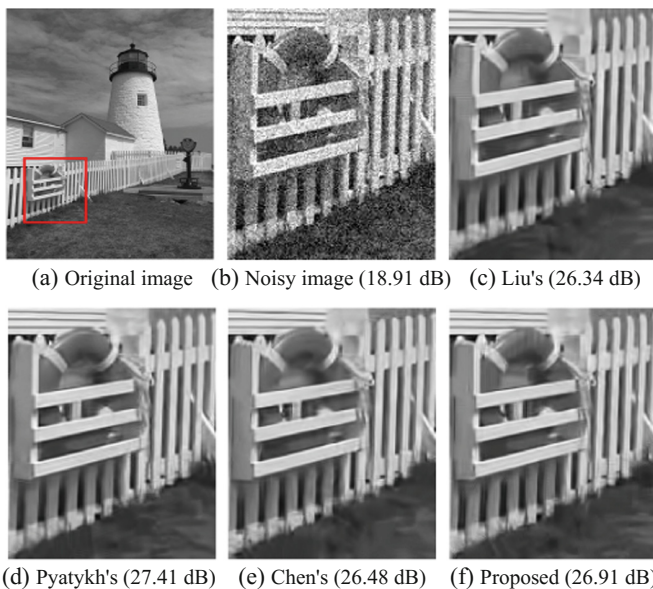
independent white Gaussian noise with deviation level 10 and 30, respectively. We set the probabilities α_1, $\alpha_2 = 0.97$ and choose $N = 256$, $s = 3$. In general, a higher noise estimation accuracy leads to a higher denoising quality. We use the K-SVD method to denoise the images [3]. Figures 1 and 2 show the results using our method outperform other competitors. Moreover, our peak signal-to-noise ratios (PSNRs) are nearest to true values, 32.03 dB and 27.01 dB, respectively.

(a) Original image   (b) Noisy image (28.14 dB)   (c) Liu's (30.32 dB)

(d) Pyatykh's (33.99 dB)   (e) Chen's (31.16 dB)   (f) Proposed (31.95 dB)

Fig. 1. Denoising results on the *Woman* image using K-SVD.

To quantitatively evaluate the accuracy of noise estimation, the average of standard deviations, mean square error (MSE), mean absolute difference (MAD) are computed by randomly selecting 1500 image patches from 20 testing images. The results shown in Table 1 indicate that the proposed method is more accurate and stable than other methods. Next, we compare our optimal estimator $\hat{\sigma}_*^2$ with $\hat{\sigma}_n^2$ and other two existing estimators in the literatures. The simulated realization of a sample covariance matrix is followed a Gaussian distribution with different variances. As presented in Table 2, the performance of $\hat{\sigma}_*^2$ is invariably better than other estimators. To test robustness of our estimation method, we further obtain the empirical probabilities of the estimated eigenvalues at typical confidence levels. Figure 3 illustrates that two asymptotic bounds can achieve very high success probabilities.

(a) Original image   (b) Noisy image (18.91 dB)   (c) Liu's (26.34 dB)

(d) Pyatykh's (27.41 dB)   (e) Chen's (26.48 dB)   (f) Proposed (26.91 dB)

Fig. 2. Denoising results on the *House* image using K-SVD.

Table 1. Estimation results of different methods (Best results are highlighted).

σ_n	Liu's [8]	Pyatykh's [9]	Chen's [11]	Proposed
1	2.18	1.34	0.59	**1.16**
5	7.30	3.83	5.41	**5.27**
10	13.86	7.19	11.83	**10.19**
15	16.72	13.91	15.92	**15.17**
20	20.99	18.75	20.62	**19.90**
25	26.64	23.29	24.34	**25.06**
30	32.38	27.27	31.98	**30.12**
MAD	3.30	1.59	0.98	**0.15**
MSE	4.84	3.22	1.39	**0.03**

Table 2. Estimation results of four estimators (Best results are highlighted).

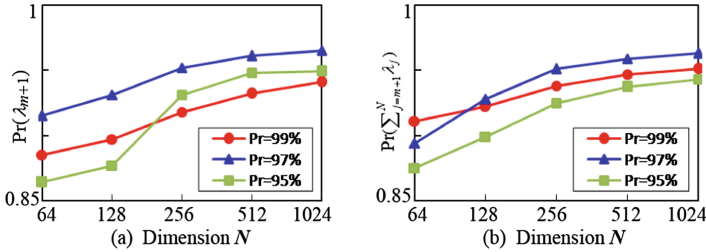| σ_n	$\hat{\sigma}_{\mathrm{median}}$ [23]	$\hat{\sigma}_{\mathrm{US}}$ [13]	$\hat{\sigma}_n$	$\hat{\sigma}_*$
1	1.27	1.99	1.14	**1.06**
5	4.59	5.27	6.24	**5.18**
10	8.76	11.28	**9.97**	9.94
15	15.22	14.29	16.17	**14.93**
20	20.85	19.14	20.96	**20.10**
25	25.87	25.98	26.31	**25.28**
30	30.59	30.37	31.16	**30.11**
MAD	0.64	0.78	0.86	**0.12**
MSE	0.52	0.72	0.99	**0.02**

Fig. 3. Empirical probabilities of exact noise eigenvalue estimation.

5 Conclusions

In this paper, we have shown how to infer the noise level from a trained dictionary. The eigen-spaces of the signal and noise are transformed and separated well by determining the eigen-spectrum interval. In addition, the developed estimator can effectively eliminate the estimation bias of a noise variance in high dimensional context. Our noise estimation technique has low computational complexity. The experimental results have demonstrated that our method outperforms the relevant existing methods over a wide range of noise level conditions.

References

1. Engan, K., Aase, S., Husoy, J.: Method of optimal directions for frame design. In: Proceedings of International Conference on Acoustics, Speech, and Signal Pattern Process (ICASSP), pp. 2443–2446 (1999)
2. Aharon, M., Elad, M., Bruckstein, A.: K-SVD: an algorithm designing overcomplete dictionaries for sparse representation. IEEE Trans. Signal Process. **54**(11), 4311–4322 (2006)
3. Elad, M., Aharon, M.: Image denoising via sparse and redundant representations over learned dictionaries. IEEE Trans. Image Process. **15**(12), 3736–3745 (2006)
4. Sahoo, S., Makur, A.: Enhancing image denoising by controlling noise incursion in learned dictionaries. IEEE Signal Process. Lett. **22**(8), 1123–1126 (2015)
5. Li, D., Zhou, J., Tang, Y.: Noise level estimation for natural images based on scale-invariant kurtosis and piecewise stationarity. IEEE Trans. Image Process. **26**(2), 1017–1030 (2017)
6. Hashemi, M., Beheshti, S.: Adaptive noise variance estimation in BayesShrink. IEEE Signal Process. Lett. **17**(1), 12–15 (2010)
7. Tang, C., Yang, X., Zhai, G.: Noise estimation of natural images via statistical analysis and noise injection. IEEE Trans. Circuit Syst. Video Technol. **25**(8), 1283–1294 (2015)
8. Liu, W., Lin, W.: Additive white gaussian noise level estimation in SVD domain for images. IEEE Trans. Image Process. **22**(3), 872–883 (2013)
9. Pyatykh, S., Hesser, J., Zhang, L.: Image noise level estimation by principal component analysis. IEEE Trans. Image Process. **22**(2), 687–699 (2013)
10. Liu, X., Tanaka, M., Okutomi, M.: Single-image noise level estimation for blind denoising. IEEE Trans. Image Process. **22**(12), 5226–5237 (2013)

11. Chen, G., Zhu, F., Heng, P.: An efficient statistical method for image noise level estimation. In: Proceedings of the International Conference on Computer Vision (ICCV), pp. 477–485 (2015)
12. Donoho, L., Johnstone, I.: Ideal spatial adaptation by wavelet shrinkage. Biometrika **81**, 425–455 (1994)
13. Ulfarsson, M., Solo, V.: Dimension estimation in noisy PCA with SURE and random matrix theory. IEEE Trans. Signal Process. **56**(12), 5804–5816 (2008)
14. Gribonval, R., Jenatton, R., Bach, F.: Sparse and spurious: dictionary learning with noise and outliers. IEEE Trans. Inf. Theory **61**(11), 6298–6319 (2015)
15. Jung, A., Eldar, Y., Gortz, N.: On the minimax risk of dictionary learning. IEEE Trans. Inf. Theory **62**(3), 1501–1515 (2016)
16. Johnstone, I.M.: On the distribution of the largest eigenvalue in principal components analysis. Ann. Stat. **29**(2), 295–327 (2001)
17. Chiani, M.: On the probability that all eigenvalues of Gaussian, Wishart, and double Wishart random matrices lie within an interval. IEEE Trans. Inf. Theory **63**(7), 4521–4531 (2017)
18. Karoui, N.E.: A rate of convergence result for the largest eigenvalue of complex white Wishart matrices. The Annals of Probability **34**(6), 2077–2117 (2006)
19. Ma, Z.M.: Accuracy of the Tracy-Widom limits for the extreme eigenvalues in white Wishart matrices. Bernoulli **18**(1), 322–359 (2012)
20. Marcenko, V.A., Pastur, L.A.: Distribution of eigenvalues for some sets of random matrices. Math. USSR-Sb. **1**(4), 457–483 (1967)
21. Bai, Z., Silverstein, J.: Spectral Analysis of Large Dimensional Random Matrices, 2nd edn. Springer, New York (2010). https://doi.org/10.1007/978-1-4419-0661-8
22. Kritchman, S., Nadler, B.: Determining the number of components in a factor model from limited noisy data. Chem. Int. Lab. Syst. **94**(1), 19–32 (2008)
23. Passemier, D., Li, Z., Yao, J.: On estimation of the noise variance in high dimensional probabilistic principal component analysis. J. R. Stat. Soc. B **79**(1), 51–67 (2017)
24. Nadler, B.: On the distribution of the ratio of the largest eigenvalue to the trace of a Wishart matrix. J. Multivar. Anal. **102**, 363–371 (2011)