AN OPTIMUM SHRINKAGE ESTIMATOR BASED ON MINIMUM-PROBABILITY-OF-ERROR CRITERION AND APPLICATION TO SIGNAL DENOISING

Jishnu Sadasivan, Subhadip Mukherjee, and Chandra Sekhar Seelamantula

Department of Electrical Engineering, Indian Institute of Science, Bangalore 560012, India
Emails: jishnus@ece.iisc.ernet.in, subhadip@ee.iisc.ernet.in, chandra.sekhar@ieee.org

ABSTRACT

We address the problem of designing an optimal pointwise shrinkage estimator in the transform domain, based on the minimum probability of error (MPE) criterion. We assume an additive model for the noise corrupting the clean signal. The proposed formulation is general in the sense that it can handle various noise distributions. We consider various noise distributions (Gaussian, Student's-t, and Laplacian) and compare the denoising performance of the estimator obtained with the mean-squared error (MSE)-based estimators. The MSE optimization is carried out using an unbiased estimator of the MSE, namely *Stein's Unbiased Risk Estimate* (SURE). Experimental results show that the MPE estimator outperforms the SURE estimator in terms of SNR of the denoised output, for low ($0 - 10$ dB) and medium values ($10 - 20$ dB) of the input SNR.

***Index Terms*—** Risk estimator, Stein's unbiased risk estimation, minimum probability of error, shrinkage function.

1. INTRODUCTION

An unbiased estimate of mean squared error (MSE), namely *Stein's Unbiased Risk Estimator* (SURE), was proposed in a seminal paper by Stein [1] for estimating the mean of an independent and identically distributed (i.i.d.) multivariate Gaussian distribution. He showed that the resulting shrinkage-type estimator of mean, obtained by minimizing the SURE, dominates classical least squares estimate when the number of data points exceeds 3. Since MSE is a function of the unknown parameters to be estimated, direct minimization of it results in an unrealizable estimator. The fundamental philosophy behind the risk estimation methodology is to replace MSE by its unbiased estimate, which depends only on the observations. Recently, many applications such as image and speech denoising have successfully deployed this approach to find an estimate of the clean signal buried in noise [2,13]. It is possible to obtain a biased estimate of the parameter that has a lower MSE than an unbiased estimate of it by scaling that unbiased estimate with a scalar between zero and one. The resulting biased estimator is called shrinkage estimator [14] and the corresponding multiplier can be obtained by minimizing SURE. The original formulation of SURE based on the assumption of independent Gaussian noise was later extended to certain distributions in continuous and discrete exponential families in [15] and [16], respectively. Both [15] and [16] rely on the assumption of independence of observations. SURE for non-i.i.d. multivariate distributions in the exponential family was recently developed by Eldar [9].

Even though the parametric model of the distribution of observation is known, incorporation of the prior knowledge in classical SURE framework is limited only up to the estimation of second-order statistics. Irrespective of the distribution of the observations, the shrinkage estimator obtained through minimization of SURE will depend only on second-order statistics (c.f. Appendix). Assuming that the parametric form of the noise distribution is known, we consider a new cost function for denoising based on minimizing the probability of error between the estimate and the true parameter exceeding a threshold. Prior knowledge of the distribution enables analytical computation of such a cost. We then develop a risk estimator for the minimum-probability-of-error (MPE) criterion and obtain the optimal shrinkage parameter. We consider applications to electrocardiogram (ECG) signal denoising in various noise conditions – Gaussian, Student's-t, and Laplacian. Notably, the Gaussian and Student's-t distributional behavior of noise is preserved by a linear orthonormal transformation [17]. In Section 2, we develop the theory for Gaussian statistics first and then extend it to other distributions. In Section 3, we present results related to denoising of a synthesized signal and ECG signal.

2. PROBLEM FORMULATION AND PROPOSED METHOD

Consider the vector signal model $\mathbf{x} = \mathbf{s} + \mathbf{w}$, where $\mathbf{s} \in \mathbb{R}^n$ denotes the clean signal vector and $\mathbf{x}$ is the observed signal corrupted by additive and white Gaussian noise $\mathbf{w}$, with known covariance matrix $\sigma^2\mathbf{I}$. We assume that the estimator of $\mathbf{s}$ from $\mathbf{x}$ is a pointwise shrinkage function, that is, the estimate of s_i, the i^{th} entry of $\mathbf{s}$, is of the form $s_i = a_i x_i$, where a_is, with $0 \leq a_i \leq 1$, are shrinkage parameters to be obtained by minimizing a suitable cost function, popularly referred to as *risk* in the statistics literature. Since the estimate of s_i depends only on x_i and not on x_js for $j \neq i$, we drop the index i in the remainder of the analysis, in the interest of notational brevity. We propose a risk function of the form

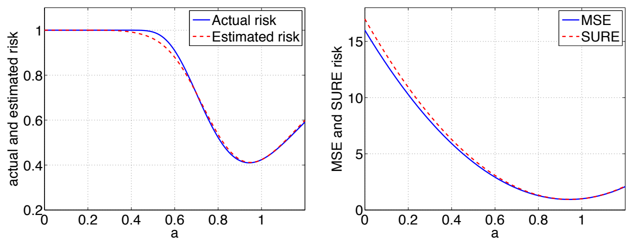
$$\mathcal{R} \quad = \quad \mathcal{P}\left(|\hat{s} - s| > \epsilon\right), \tag{1}$$

for a suitably chosen $\epsilon > 0$, which directly captures the probability that the estimated value lies outside an ϵ-radius of the actual parameter value s. Since $\hat{s} = ax$, and x follows a Gaussian distribution with mean s and variance σ^2, we have that $z \triangleq \hat{s} - s = ax - s$ is distributed as $\mathcal{N}\left((a-1)s, a^2\sigma^2\right)$. As a consequence, the expression of $\mathcal{R}$ simplifies to

$$\begin{aligned} \mathcal{R} &= \mathcal{P}\left(|z| > \epsilon\right) \\ &= Q\left(\frac{\epsilon - (a-1)s}{a\sigma}\right) + Q\left(\frac{\epsilon + (a-1)s}{a\sigma}\right), \end{aligned} \tag{2}$$

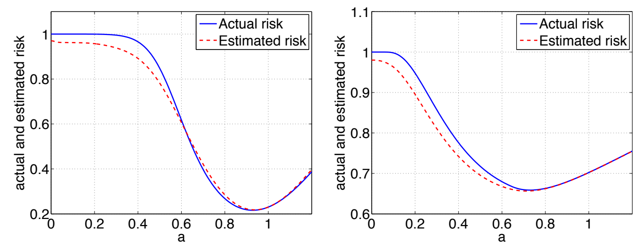
where $Q(\cdot)$ denotes the tail probability of the standard Gaussian distribution, given by $Q(u) = \frac{1}{\sqrt{2\pi}} \int_u^\infty \exp\left(-\frac{t^2}{2}\right) \mathrm{d}t$. The expres-

978-1-4799-2893-4/14/$31.00 ©2014 IEEE



(a) (b)

Fig. 1. (Colour online) Variation of the proposed risk-estimate (averaged over 100 noise realizations) and the SURE with respect to the shrinkage parameter a. The value of ϵ chosen in (a) is $\epsilon = 3\sigma$.



(a) Student's-t noise, $\sigma^2 = 2$. (b) Laplacian noise, $\sigma^2 = 1$.

Fig. 2. (Colour online) Variation of the proposed risk-estimate (averaged over 100 realizations) for Student's-t and the Laplacian noise statistics with respect to the shrinkage parameter a. The value of ϵ chosen in (a) and (b) is $\epsilon = \sigma$.

sion of risk depends on s, which is the parameter to be estimated, and consequently the value of a obtained by minimizing R also depends on s, thereby resulting in an unrealizable estimator. To alleviate this problem, we replace s by its maximum-likelihood (ML) estimate, namely, $\hat{s}_{ML} = x$ in the expression of risk in (2). Hence, we obtain an estimate of the actual risk R given by

$$\hat{\mathcal{R}} = Q\left(\frac{\epsilon - (a-1)x}{a\sigma}\right) + Q\left(\frac{\epsilon + (a-1)x}{a\sigma}\right), \quad (3)$$

which is minimized over a to obtain the optimum shrinkage parameter a_{opt}. Therefore, we have that

$$a_{\text{opt}} = \arg\min_{0\le a\le 1} Q\left(\frac{\epsilon - (a-1)x}{a\sigma}\right) + Q\left(\frac{\epsilon + (a-1)x}{a\sigma}\right). \quad (4)$$

Subsequently, the estimate of the parameter s is obtained by multiplying x with the optimum shrinkage parameter, that is, $\hat{s} = a_{\text{opt}}x$. We deploy the steepest descent method to solve the optimization problem in (4). Starting from an initial guess $a^{(0)}$, the value of a is updated as $a^{(t+1)} = a^{(t)} - \mu^{(t)}\frac{\mathrm{d}\hat{R}}{\mathrm{d}a}$, where $\mu^{(t)} > 0$ is the step size chosen in iteration t. The optimum value of the parameter obtained by minimizing the SURE is $a_{SURE} = \max\left\{0, 1 - \frac{\sigma^2}{x^2}\right\}$. To illustrate how the actual risk $\mathcal{R}$ and and its estimate $\hat{\mathcal{R}}$ behave as a function of the shrinkage parameter a, the following experiment is carried out. We consider the problem of estimating a scalar $s = 4$ in additive Gaussian noise of zero mean and variance $\sigma^2 = 1$. The estimated risk $\hat{R}$ is calculated for $\epsilon = 3\sigma$. In Figure 1, we show $\mathcal{R}$ and $\hat{\mathcal{R}}$ as functions of a, averaged over 100 independent trials. Although $\hat{\mathcal{R}}$ differs slightly from the actual risk $\mathcal{R}$, the values of a where they attain minima are approximately equal.

2.1. Extension to non-Gaussian distributions

The proposed approach is generalizable to the case where the additive noise samples w_ns have zero mean and follow a non-Gaussian density. In that case, the expression for the risk $\mathcal{R}$ becomes

$$\begin{aligned}\mathcal{R} &= \mathcal{P}\left(|a(s+w) - s| > \epsilon\right)\\ &= 1 - F_W\left(\frac{\epsilon - (a-1)s}{a}\right) + F_W\left(-\frac{\epsilon + (a-1)s}{a}\right),\end{aligned}$$

where $F_W(w) = \int_{-\infty}^{w} f_W(t)\mathrm{d}t$ is the cumulative distribution function of the additive noise. The risk $\hat{R}$ can be estimated by replacing s with its ML estimate x in the expression for $\mathcal{R}$. We consider the following two cases where the noise distribution is non-Gaussian.

2.1.1. *Case 1*

Let us consider the case where w_ns follow Student's-t distribution with parameter $\lambda > 2$, that is,

$$f_W(w) = \frac{\Gamma\left(\frac{\lambda+1}{2}\right)}{\sqrt{\lambda\pi}\,\Gamma\left(\frac{\lambda}{2}\right)}\left(1 + \frac{w^2}{\lambda}\right)^{-\frac{\lambda+1}{2}}.$$

The variance of noise is given by $\sigma^2 = \frac{\lambda}{\lambda-2}$. The expression for R can be obtained by using

$$F_W(w) = \frac{1}{2} + w\Gamma\left(\frac{\lambda+1}{2}\right)\frac{F_1\left(\frac{1}{2}, \frac{\lambda+1}{2}; \frac{3}{2}; -\frac{w^2}{\lambda}\right)}{\sqrt{\lambda\pi}\,\Gamma\left(\frac{\lambda}{2}\right)},$$

where $F_1(\cdot)$ denotes the hypergeometric function given by

$$F_1(a, b; c; z) = \sum_{n=0}^{\infty}\frac{(a)_n(b)_n}{(c)_n}\frac{z^n}{n!},$$

where, in turn, $(q)_n$ is defined as

$$(q)_n \triangleq \begin{cases}1 \text{ for } n = 0,\\ q(q+1)(q+2)\cdots(q+n-1) \text{ for } n > 0.\end{cases}$$

2.1.2. *Case 2*

In the case where w_ns follow an independent zero-mean Laplacian distribution with parameter b, that is, $f_W(w) = \frac{1}{2b}\exp\left(-\frac{|w|}{b}\right)$, the expression for R is computed by setting

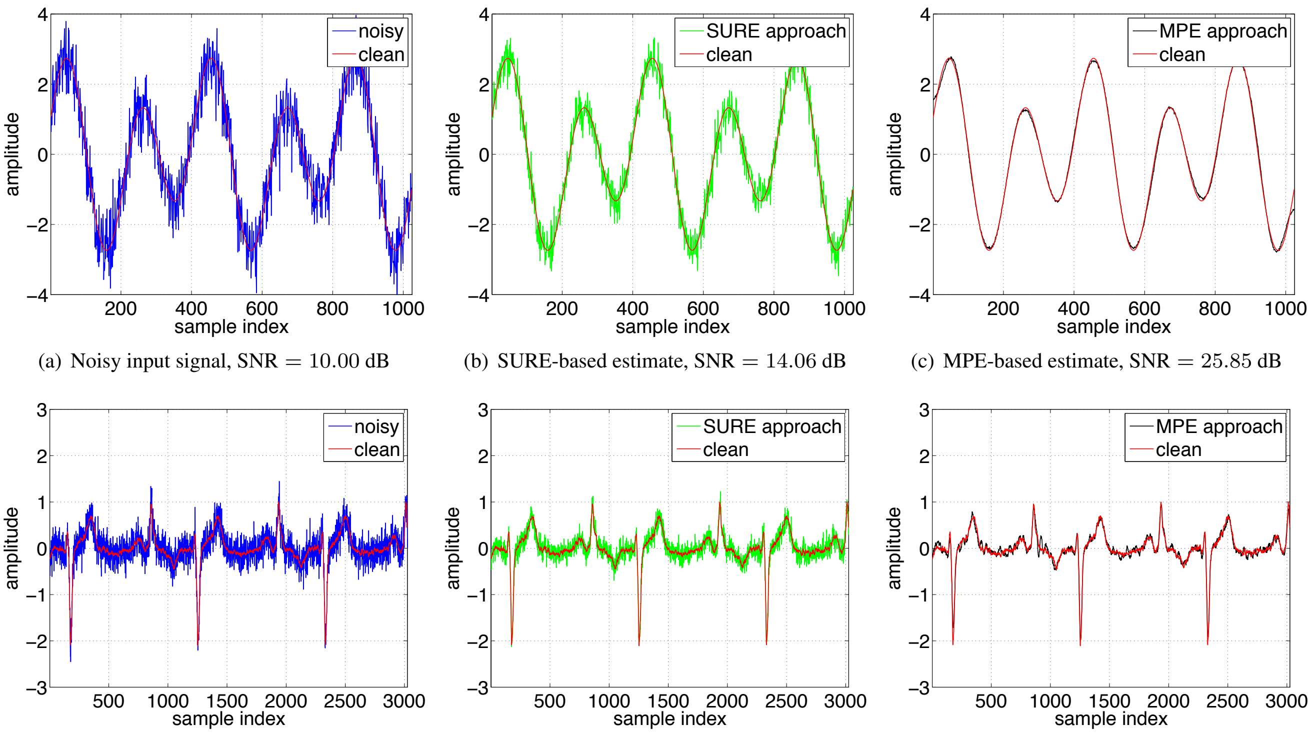
$$F_W(w) = \frac{1}{2} + \frac{1}{2}\text{sign}(w)\left(1 - \exp\left(-\frac{|w|}{b}\right)\right).$$

The variance of noise is related to b as $\sigma^2 = 2b^2$.

The actual cost and the risk estimate for the Student-t and Laplacian noise cases is given in Figure 2. From the plots, we observe that the locations of the optima match closely. Hence, the proposed risk estimators are reliable and can be used to substitute for the actual risk.

3. EXPERIMENTAL RESULTS

We deploy the proposed MPE shrinkage estimator for denoising synthesized as well as real ECG signals.



(a) Noisy input signal, SNR $= 10.00$ dB (b) SURE-based estimate, SNR $= 14.06$ dB (c) MPE-based estimate, SNR $= 25.85$ dB

(d) Noisy input signal, SNR $= 5.00$ dB (e) SURE-based estimate, SNR $= 9.05$ dB (f) MPE-based estimate, SNR $= 12.88$ dB

Fig. 3. (Colour online) Comparison of denoising performance of the MPE estimator and the SURE estimator for additive Gaussian noise. The first and second rows correspond to the experiments done with the synthesized signal and the ECG signal, respectively.

3.1. Denoising of synthesized signal

We synthesize a signal s_n, $0 \leq n \leq N-1$, of the form $s_n = \cos\left(\frac{5\pi n}{N}\right) + 2\sin\left(\frac{10\pi n}{N}\right)$, where $N = 1024$, and consider the task of estimating s_ns from their noisy measurement $x_n = s_n + w_n$. We assume that the noise w_n corrupting the signal s_n are i.i.d. samples from a Gaussian distribution with zero mean and variance σ^2. A pointwise shrinkage function of the form $\hat{S}_k = aX_k$, where k denotes the discrete cosine transform (DCT) coefficient index, is applied to the DCT coefficients of the noisy signal x_n, and subsequently inverse DCT is computed to obtain the estimate $\hat{s}_n$ of the clean signal s_n. The optimum parameter a_{opt} is obtained by minimizing the proposed estimate $\hat{\mathcal{R}}$ of the actual risk over a. Since DCT is an orthonormal transform, the DCT coefficients of the noise are also i.i.d. samples following a Gaussian distribution with identical mean and variance as w_n. We carry out the task of denoising for different values of input SNR, defined as $\text{SNR}_{\text{in}} = \dfrac{1}{N\sigma^2}\displaystyle\sum_{n=0}^{N-1} s_n^2$. The performance of the proposed MPE based estimator is compared with the standard SURE estimate and the denoised output signals corresponding to $\text{SNR}_{\text{in}} = 10$ dB are shown in Figures 3(a), 3(b), and 3(c). We observe that the MPE estimator results in an improvement of approximately 10 dB in output SNR over the standard SURE-based estimator.

3.2. Denoising of ECG signal

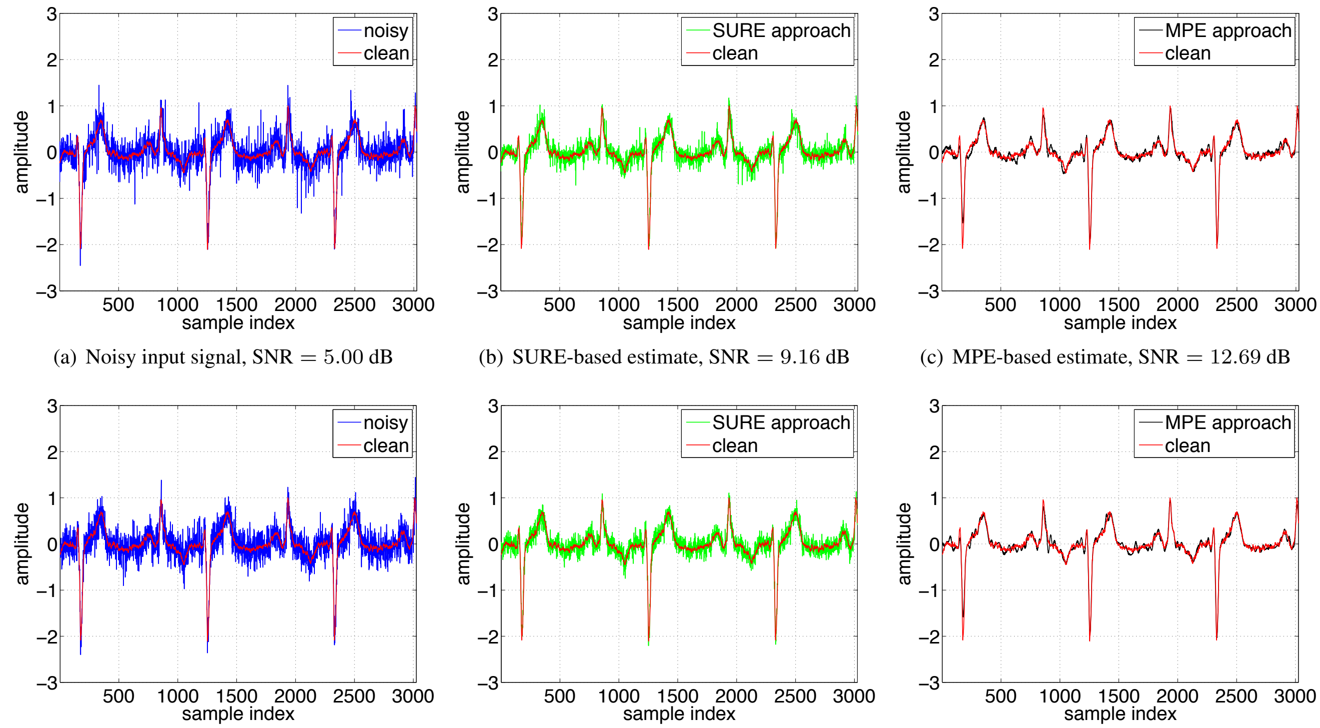
We next consider the task of denoising real ECG signals corrupted by additive zero-mean noise following the Gaussian, Student's-t, and the Laplacian distribution with known variances. In practice, the variance can be estimated quite reliably using ML estimators or median-based robust estimators [18, 19]. The ECG signals used in our experiment are taken from the PhysioBank database [20]. For the purpose of denoising, pointwise shrinkage estimate is obtained in the DCT domain, and the optimum shrinkage a_{opt} is obtained by minimizing the respective MPE risk estimates depending on the noise statistics. The denoised signals obtained by minimizing MPE and SURE are shown in Figures 3(d)-(f) and 4(a)-(f), respectively for Gaussian and non-Gaussian noise statistics. The output SNR values for the MPE and SURE-based estimates (averaged over 50 independent realizations of Gaussian noise) are reported in Table 1, corresponding to various input SNR values.

3.3. Choice of the parameter ϵ

Appropriate choice of the parameter ϵ plays an important role in determining the performance of the MPE estimator. It is difficult to find an expression for the optimum value of ϵ in closed form that will lead to maximum SNR in the denoised output signal. We plot the ensemble-averaged (over 20 trials) output SNR as a function of ϵ in Figure 5, for the case where the noise samples follow an i.i.d. Gaussian distribution. We observe that, for almost all values of input SNR, output SNR values exhibit a peak approximately at $\beta = \frac{\epsilon}{\sigma} = 3.5$ and $\beta = 3$, for the synthesized signal and ECG signals, respectively.

4. CONCLUSION

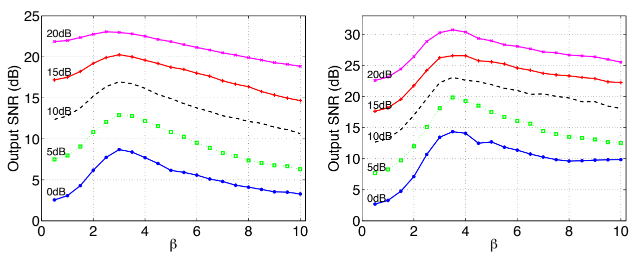
We proposed an optimum pointwise shrinkage estimator by minimizing a risk function based on the MPE criterion. Our formulation is applicable to scenarios where the noise samples are independent



(a) Noisy input signal, SNR $= 5.00$ dB (b) SURE-based estimate, SNR $= 9.16$ dB (c) MPE-based estimate, SNR $= 12.69$ dB

(d) Noisy input signal, SNR $= 5.00$ dB (e) SURE-based estimate, SNR $= 9.10$ dB (f) MPE-based estimate, SNR $= 12.85$ dB

Fig. 4. (Colour online) Comparison of denoising performance of the MPE estimator and the SURE estimator for ECG signals corrupted by noise following non-Gaussian statistics. The first and second rows correspond to the cases where the noise follows Student's-t and Laplacian distributions, respectively. We observed that $\epsilon = 3.5\sigma$ is optimum.



(a) ECG signal (b) Synthesized signal

Fig. 5. (Colour online) Variation of output SNR (averaged over 50 independent noise realizations) as a function of $\beta = \frac{\epsilon}{\sigma}$, for different values of input SNR indicated on the respective plots.

Input SNR	Output SNR (MPE)	Output SNR (SURE)
-5.00	4.76	-0.60
-2.50	6.56	1.80
0	8.65	4.23
2.50	10.83	6.66
5.00	12.92	9.09
7.50	14.88	11.46
10.00	16.93	13.83
12.50	18.62	16.10
15.00	20.21	18.35
17.50	21.69	20.49
20.00	22.98	22.57

Table 1. Comparison of the MPE and SURE estimates for different values of input SNR (dB). The output SNR values (dB) are averaged over 50 independent noise realizations; $\epsilon = 3.5\sigma$.

and follow an additive model, but it can handle noise following a non-Gaussian distribution. We assumed that noise has zero-mean and its variance is known. As the experimental results show, shrinkage estimator based on the proposed risk estimate outperforms the estimator based on MSE minimization using SURE. As an application, we considered the task of denoising ECG signals corrupted by additive noise following various statistical models, to illustrate the efficacy of the MPE estimator over its SURE-based counterpart. The improvement in performance in terms of SNR of the denoised output is attributed to the fact that the MPE framework incorporates knowledge of the distribution of the observations, which goes beyond the second-order statistics considered in MSE-based optimization.

Appendix : SURE for pointwise shrinkage estimator

Suppose $x = s+w$ is the noisy observation, where w has zero-mean and variance σ^2, and $\hat{s} = ax$ is an estimate of s. MSE of $\hat{s}$ is defined as $\mathcal{R} = \mathcal{E}\left\{|\hat{s} - s|^2\right\} = a^2\sigma^2 + (a-1)^2\left(\mathcal{E}\left\{x^2\right\} - \sigma^2\right)$. An unbiased estimate of $\mathcal{R}$ is given by $\hat{\mathcal{R}} = a^2\sigma^2 + (a-1)^2\left(x^2 - \sigma^2\right)$. Minimizing $\hat{\mathcal{R}}$ with respect to a yields $a_{\text{opt}} = 1 - \frac{\sigma^2}{x^2}$.

5. REFERENCES

[1] C. M. Stein, "Estimation of the mean of a multivariate normal distribution," *Ann. Stat.*, vol. 9, no. 6, pp. 1135–1151, Nov. 1981.

[2] F. Luisier, T. Blu, and M. Unser, "A new SURE approach to image denoising: Interscale orthonormal wavelet thresholding," *IEEE Trans. Image Process.*, vol 16, no. 3, pp. 593–606, Mar. 2007.

[3] D. L. Donoho and I. M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *J. Amer. Stat. Assoc.*, vol. 90, no. 432, pp.1200–1224, Dec. 1995.

[4] A. Benazza-Benyahia and J. C. Pesquet, "Building robust wavelet estimators for multicomponent images using Steins principle," *IEEE Trans. Image Process*, vol. 14, no. 11, pp. 1814–1830, Nov. 2005

[5] J. C. Pesquet, A. Benazza-Benyahia, and C. Chaux, "A SURE approach for digital signal/image deconvolution problems," *IEEE Trans. Signal Process.*, vol. 57, no. 12, pp. 4616–4632, Dec. 2009

[6] R. Giryes, M. Elad, and Y. C. Eldar, " The projected GSURE for automatic parameter tuning in iterative shrinkage methods," *Appl. Comput. Harmon. Anal.*, vol. 30, no. 3, pp. 407–422, May 2011.

[7] A. Benazza-Benyahia, and C. Chaux, "Building robust wavelet estimators for multicomponent images using Stein's principle," *IEEE Trans. Image Process*, vol. 14, no. 11, pp. 18141830, Nov. 2005

[8] T. Blu and and F. Luisier, "The SURE-LET approach to image denoising," *IEEE Trans. Image Process*, vol. 16, no. 11, pp. 2778–2786, Nov. 2007.

[9] Y. C. Eldar, "Generalized SURE for exponential families: Applications to regularization," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 471–481, Feb. 2009.

[10] N. R. Muraka and C. S. Seelamantula, "A risk-estimation-based comparison of mean square error and Itakura-Saito distortion measures for speech enhancement," in *Proc. Interspeech*, Florence, Italy, pp. 349-352, Aug. 2011.

[11] S. R. Krishnan and C. S. Seelamantula, "A generalized Stein's estimation approach for speech enhancement based on perceptual criteria," in *Proc. Workshop on Statistical and Perceptual Audition (SAPA)–Speech Communication with Adaptive Learning (SCALE).*, Portland, Sep. 7-8, 2012.

[12] S. R. Krishnan and C. S. Seelamantula, "On the selection of optimum Savitzky-Golay filters," *IEEE Trans. Signal Process.*, vol. 61, no. 2, pp. 380–391, Jan. 2013.

[13] N. Zheng, X. Li, T. Blu, and T. Lee, "SURE-MSE speech enhancement for robust speech recognition," in *Proc.* 7^{th} *Int. Symp. Chinese Spoken Language Process.*, pp. 271–274, Shenzhen, China, Nov. 2010.

[14] S. Kay and Y. C. Eldar, "Rethinking biased estimation," *IEEE Signal Process. Magazine*, vol. 25, no. 3, pp. 133–136, May 2008.

[15] H. M. Hudson, "A natural identity for exponential families with applications in multiparameter estimation," *Ann. Stat.*, vol. 6, no. 3, pp. 473–484, 1978.

[16] J. T. Hwang, "Improving upon standard estimators in discrete exponential families with applications to Poisson and negative binomial cases," *Ann. Stat.*, vol. 10, no. 3, pp. 857–867, 1982.

[17] B. M. G. Kibria and A. H. Joarder, "A short review of multivariate t-distribution," *Journal of Statistical Research*, vol. 40, no. 1, pp. 59–72, 2006.

[18] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice Hall, 2010.

[19] S. Mallat, *A Wavelet Tour of Signal Processing*, 3rd edition, Academic Press, 2009.

[20] PhysioBank database [Online].
Available: http://www.physionet.org/ physiobank/database/aami-ec13/