SIP (2020), vol. 9, e3, page 1 of 16 © The Authors, 2020.

doi:10.1017/ATSIP.2019.27

ORIGINAL PAPER

Signal denoising using the minimum-probability-of-error criterion

JISHNU SADASIVAN,[1] SUBHADIP MUKHERJEE[2] AND CHANDRA SEKHAR SEELAMANTULA[1]

We consider signal denoising via transform-domain shrinkage based on a novel risk criterion called the minimum probability of error (MPE), which measures the probability that the estimated parameter lies outside an ϵ-neighborhood of the true value. The underlying parameter is assumed to be deterministic. The MPE, similar to the mean-squared error (MSE), depends on the ground-truth parameter, and therefore, has to be estimated from the noisy observations. The optimum shrinkage parameter is obtained by minimizing an estimate of the MPE. When the probability of error is integrated over ϵ, it leads to the expected ℓ_1 distortion. The proposed MPE and ℓ_1 distortion formulations are applicable to various noise distributions by invoking a Gaussian mixture model approximation. Within the realm of MPE, we also develop a specific extension to subband shrinkage. The denoising performance of MPE turns out to be better than that obtained using the minimum MSE-based approaches formulated within Stein's unbiased risk estimation (SURE) framework, especially in the low signal-to-noise ratio (SNR) regime. Performance comparisons with three benchmarking algorithms carried out on electrocardiogram signals and standard test signals taken from the Wavelab toolbox show that the MPE framework results in SNR gains particularly for low input SNR.

Keywords: Minimum probability of error, Shrinkage estimator, Risk estimation, Expected ℓ_1 distortion, Gaussian mixture model (GMM)

Received 22 August 2019; Revised 18 December 2019

1. INTRODUCTION

Signal denoising algorithms are often developed with the objective of minimizing a chosen distortion function that quantifies the distance between an estimate and the ground-truth. The most widely used distortion measure in the literature is the mean-squared error (MSE). The ground-truth parameter may be deterministic or stochastic with a known prior distribution. The latter formalism leads to Bayesian estimators. Within the deterministic signal estimation paradigm, which is also the formalism considered in this paper, one typically desires that the estimator has minimum variance and is unbiased (MVU) [1,2]. An MVU estimator may not always exist, and if it does, it can be obtained using the theory of sufficient statistics. Eldar and Kay [2] showed that, when it comes to minimizing the MSE, biased estimates can outperform the MVU estimate. For example, one could shrink the MVU estimate and optimally select the shrinkage parameter to minimize the MSE.

In this paper, we consider the problem of estimating a deterministic signal corrupted by additive white noise. The noise distribution is assumed to be known, but not restricted to be Gaussian. We propose a new distortion measure based on the probability of error and develop transform-domain shrinkage estimators. The proposed risk criterion leads to shrinkage estimators that are adapted to the noise type, meaning that the shrinkage operators corresponding to different noise distributions having the same variance turn out to be different. This behavior is in contrast with shrinkage estimators obtained using the MSE, wherein one obtains the same shrinkage profile corresponding to different noise distributions having the same variance. Before proceeding with the developments, we review important literature related to the problem at hand.

A) Prior art

The MSE is by far the most widely used measure for obtaining the optimum shrinkage parameter. Since the MSE is a function of the ground-truth, directly minimizing it results in a practically unrealizable estimate, in the sense that the estimate depends on the unknown parameter/or its statistics. However, in some cases, it is possible to find the optimum shrinkage parameter, for example, using a min-max approach [2], where the parameter is constrained to a known set. An optimum shrinkage estimator, when the variance of the unbiased estimate is a scaled version of the square of the parameter, with a known scaling, is proposed in [2].

[1]Department of Electrical Engineering, Indian Institute of Science, Bangalore 560012, India
[2]Presently with Department of Mathematics, KTH Royal Institute of Technology, Stockholm, Sweden

Corresponding author:
Subhadip Mukherjee.
Email: subhadipju@gmail.com

Optimum shrinkage estimators have also been computed based on *risk estimation*, where an unbiased estimate of the MSE, which depends only on the noisy observations, is obtained and subsequently minimized to compute the optimal shrinkage. Under the assumption of Gaussian noise, an unbiased estimate of the MSE, namely Stein's unbiased risk estimator (SURE), was developed based on Stein's lemma [3], and has been successfully employed in numerous denoising applications. The shrinkage estimator of the mean of a multivariate Gaussian distribution with diagonal covariance matrix, obtained by minimizing SURE, dominates the classical least-squares estimate when the observation dimension exceeds three [4].

A risk minimization approach for denoising using a linear expansion of elementary thresholding functions has been developed in [5–9], wherein the combining weights are chosen optimally to minimize the SURE objective. SURE-optimized wavelet-domain thresholding techniques have been proposed in [10–12]. Atto et al. [13,14] have investigated the problem of signal denoising based on optimally selecting the parameters of a wavelet-domain smooth sigmoidal shrinkage function by minimizing the SURE criterion. Ramani et al. [15] developed a Monte Carlo technique to select the parameters of a generic denoising operator based on SURE. An image denoising algorithm based on non-local means (NLM) is proposed in [16], where the NLM parameters are optimized using SURE. Notable denoising algorithms that aim to optimize the SURE objective include wavelet-domain multivariate shrinkage [17], local affine transform for image denoising [18], SURE-optimized blockwise shrinkage for image denoising [19], SURE-optimized Savitzky-Golay filter [20], etc. The SURE approach has also found applications in image deconvolution [21] and compressive sensing [22].

The original formulation of SURE, which assumes independent Gaussian noise has been extended to certain distributions in continuous and discrete exponential families in [23,24], respectively, with the assumption of independence left unchanged. Eldar generalized SURE (GSURE) for distributions belonging to the non-i.i.d. multivariate exponential family [4]. Giryes et al. [25] used a projected version of GSURE for selecting parameters in the context of solving inverse problems. Luisier et al. [26] proposed an unbiased estimate of MSE under Poisson noise and minimized it to obtain the optimum wavelet-domain image denoising function. A detailed discussion of Gaussian parameter estimation using shrinkage estimators, together with a performance comparison of SURE with the maximum-likelihood (ML) and soft-thresholding-based estimators can be found in [27] (Chapter 2). It is shown in [27] that the soft-thresholding-based estimator dominates the James-Stein shrinkage estimator in terms of MSE if the parameter vector to be estimated is highly sparse, such as a spike. On the other hand, the shrinkage estimator dominates if the parameter vector has nearly equal entries.

B) This paper

We propose a signal estimation method based on the probability of error (PE) criterion, which we first introduced in [28] and demonstrated application to electrocardiogram (ECG) signal denoising. Recently, in [29] we proposed a speech denoising framework based on [28]. In this paper, we expand further on the idea and carry out a detailed investigation on various aspects related to the PE criterion. In particular, we consider different noise distributions such as Laplacian, Student's-*t*, Gaussian, a Gaussian mixture model (GMM), expected ℓ_1 distortion, analysis and comparison of the implicit shrinkage function, and performance benchmarking with standard denoising techniques. The PE quantifies the probability of the estimate falling outside an ϵ-neighborhood of the true parameter value. Since the PE risk depends on the ground truth, we consider a surrogate, which may be biased, and minimize it to obtain the shrinkage parameter (Section 2). The optimization is carried out in the discrete cosine transform (DCT) domain, either in a pointwise fashion or on a subband basis. The resulting estimator is referred to as the shrinkage estimator based on the minimum probability of error (MPE). We derive the PE risk for Gaussian, Laplacian, and Student's-*t* noise distributions (Sections 2-2.1 and 2.5). In practical applications, where the noise distribution may be multimodal and not known explicitly, we propose to use a Gaussian mixture model (GMM) approximation [30,31] (Section 2-2.4). We show the performance of the MPE-based denoising technique on the *Piece-Regular* signals taken from the *Wavelab* toolbox in Gaussian, Student's-*t*, and Laplacian noise contaminations (Section 3). Proceeding further, we also consider the probability of error integrated over $0 < \epsilon < \infty$ (Section 4), which results in the *expected* ℓ_1 *distortion* between the parameter and its estimate. The ℓ_1 distance has been used in the Bayesian setting and in regression problems as a robust distortion metric [1,32]. The estimators for the expected ℓ_1 distortion are also derived by invoking the GMM approximation (Section 4-4.1). The noise-adaptive shrinkage behavior resulting from the proposed PE and ℓ_1 distortion-based risk functions are demonstrated by plotting the shrinkage profile as a function of the parameter value (Section 4-4.3). We also assess the denoising performance of the shrinkage estimator obtained by minimizing the ℓ_1 distortion for different input SNRs and for different number of noisy realizations (Section 5).

To further boost the denoising performance of the ℓ_1 distortion-based estimator, we develop an iterative algorithm to successively refine the cost function and the resulting estimate, starting with the noisy signal as the initialization (Section 5). The iterations lead to an improvement of 2–3 dB in the output signal-to-noise ratio (SNR) (Section 5-5.1).

For performance evaluation, we conduct experiments on the *Piece-Regular* and the *HeaviSine* signals from the *Wavelab* toolbox [33], and ECG signals from the *PhysioBank* database [34]. We consider three techniques for comparison: (i) wavelet-domain soft-thresholding [35]; (ii)

https://doi.org/10.1017/ATSIP.2019.27 Published online by Cambridge University Press

SURE-based orthonormal wavelet thresholding using a linear expansion of thresholds (SURE-LET) [5]; and (iii) SURE-based smooth sigmoid shrinkage (SS) in wavelet domain [13], all assuming Gaussian noise contamination (Section 6) in order to facilitate a fair comparison.

II. THE PE RISK

Consider the observation model $\boldsymbol{x} = \boldsymbol{s} + \boldsymbol{w}$, where $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{s} \in \mathbb{R}^n$ denote the noisy and clean signals, respectively. The noise vector $\boldsymbol{w}$ is assumed to have i.i.d. entries with zero mean and variance σ^2. The goal is to estimate $\boldsymbol{s}$ from $\boldsymbol{x}$ by minimizing a suitable cost function. The signal model is considered in an appropriate transform domain, where the signal admits a parsimonious representation, but noise does not. We consider two types of shrinkage estimators: (i) pointwise, where a shrinkage factor $a_i \in [0, 1]$ is applied to x_i in order to obtain the estimate $\widehat{s}_i = a_i x_i$, which means that $\widehat{s}_i$ does not depend on $\widehat{s}_j$ for $i \neq j$; and (ii) subband-based estimator, wherein a single shrinkage factor a_J is applied to a group of coefficients x_i, $i \in J$ in subband $J \subset 1, 2, \ldots, n$. Shrinkage estimators may also be interpreted as premultiplication of $\boldsymbol{x}$ by a diagonal matrix.

A) PE risk for pointwise shrinkage

Since the estimate of s_i does not depend on x_j, for $j \neq i$, in the pointwise shrinkage scenario, we drop the index i for brevity of notation. The PE risk is defined as

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

where $\epsilon > 0$ is a predefined tolerance parameter. The risk $\mathcal{R}$ quantifies the estimation error using the probability measure and implicitly takes into account the noise distribution beyond the first- and second-order statistics. On the contrary, the MSE relies only on the first- and second-order statistics of noise for shrinkage estimators. Substituting $\widehat{s} = ax = a(s + w)$, the risk $\mathcal{R}$ evaluates to

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

where $F(\cdot)$ is the cumulative distribution function (c.d.f.) of the noise. Since $\mathcal{R}$ depends on s, which is the parameter to be estimated, it is impractical to optimize it directly over a. To circumvent the problem, we minimize an estimate of $\mathcal{R}$, which is obtained by replacing s with x (which is also the ML estimate of s). Such an estimate $\widehat{\mathcal{R}} = \mathcal{R}(x; a)$ takes the form

$$\widehat{\mathcal{R}} = 1 - F\left(\frac{\epsilon - (a-1)x}{a}\right) + F\left(-\frac{\epsilon + (a-1)x}{a}\right), \tag{2}$$

and correspondingly, the optimal shrinkage parameter is obtained as $a_{\text{opt}} = \arg\min_{0 \le a \le 1} \widehat{\mathcal{R}}$. A grid search is performed to optimize $\widehat{\mathcal{R}}$ over $a \in [0, 1]$, and the clean signal is obtained as $\widehat{s} = a_{\text{opt}} x$. We next derive explicit formulae for the risk function for Gaussian, Laplacian, and Student's-t noise distributions.

(i) Gaussian distribution: In this case, the noisy observation x also follows a Gaussian distribution, and therefore, $\widehat{s} - s$ is distributed as $\mathcal{N}((a-1)s, a^2\sigma^2)$. The PE risk estimate is given as

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

where $Q(u) = (1/\sqrt{2\pi}) \int_u^\infty e^{-t^2/2}\, \mathrm{d}t$.

(ii) Student's-t distribution: Consider the case where the noise follow a Student's-t distribution with parameter $\lambda > 2$ and the probability density function (p.d.f.) of noise is given by

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

The variance of w is $\sigma^2 = \lambda/(\lambda - 2)$. The expression for $\widehat{\mathcal{R}}$ is the one given in (2) with

$$F(w) = \frac{1}{2} + w\Gamma\left(\frac{\lambda+1}{2}\right) \frac{G_1\left(\frac{1}{2}, \frac{\lambda+1}{2}; \frac{3}{2}; -\frac{w^2}{\lambda}\right)}{\sqrt{\lambda\pi}\,\Gamma\left(\frac{\lambda}{2}\right)},$$

where G_1 is the hypergeometric function defined as

$$G_1(a, b; c; z) = \sum_{k=0}^{\infty} \frac{(a)_k (b)_k}{(c)_k} \frac{z^k}{k!},$$

and $(q)_k$ denotes the Pochhammer symbol:

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

(iii) Laplacian distribution: Considering the noise to be i.i.d. Laplacian with zero-mean and parameter b (variance $\sigma^2 = 2b^2$), with the p.d.f. $f(w) = (1/2b)\exp(-|w|/b)$, the PE risk can be obtained by using the following expression for $F(w)$ in (2):

$$F(w) = \frac{1}{2} + \frac{1}{2}\mathrm{sgn}(w)\left(1 - \exp\left(-\frac{|w|}{b}\right)\right). \tag{4}$$

B) Closeness of $\widehat{\mathcal{R}}$ to $\mathcal{R}$

To measure the closeness of $\widehat{\mathcal{R}}$ to $\mathcal{R}$, consider the example of estimating a scalar $s = 4$ from a noisy observation x. The PE risk estimate $\widehat{\mathcal{R}}$ is obtained by setting $s = x$. In Figs 1(a),1(b), and 1(c), we show the variation of the actual risk $\mathcal{R}$ and its estimate $\widehat{\mathcal{R}}$ with a, averaged over 100 independent trials, for Gaussian, Student's-t, and Laplacian noise distributions, respectively. The noise has zero mean, and the variance is taken as $\sigma^2 = 1$ for Gaussian and Laplacian models, whereas for Student's-t model, the variance is $\sigma^2 = 2$. The value of ϵ is set equal to σ while computing the PE risk. We observe that $\widehat{\mathcal{R}}$ is a good approximation to $\mathcal{R}$, particularly in the vicinity of the minima. The deviation of

https://doi.org/10.1017/ATSIP.2019.27 Published online by Cambridge University Press

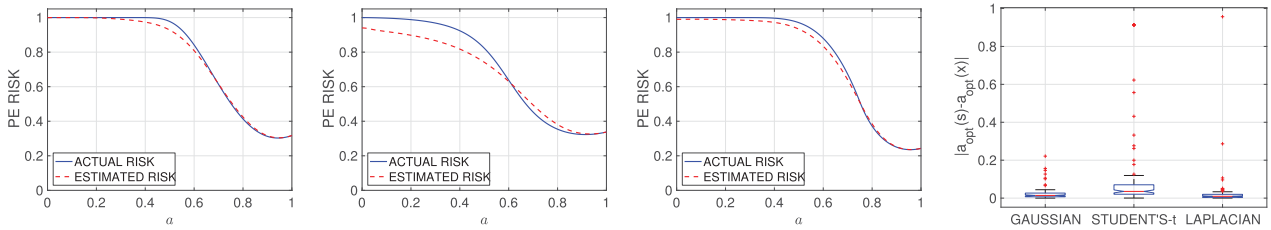

Fig. 1. (Color online) The PE risk averaged over 100 realizations for: (a) Gaussian, (b) Student's-*t*, and (c) Laplacian noise, versus the shrinkage parameter a; and (d) the percentiles of error in minima.

the shrinkage parameter $a_{\text{opt}}(x)$, obtained by minimizing $\widehat{\mathcal{R}}$, with respect to its true value $a_{\text{opt}}(s)$ resulting from the minimization of $\mathcal{R}$, is shown in Fig. 1(d) for the three noise models under consideration. The central red lines in Fig. 1(d) indicate the medians, whereas the black lines on the top and bottom denote the 25 and the 75 percentile points, respectively. We observe that $a_{\text{opt}}(x)$ is concentrated around $a_{\text{opt}}(s)$, especially for Gaussian and Laplacian noise, barring a few outliers.

C) Perturbation probability of the location of minimum

The location of the minimum of the PE risk determines the shrinkage parameter. Therefore, one must ensure that it does not deviate too much from its actual value, with high probability, when s is replaced by x in the original risk $\mathcal{R}$. Let $a_{\text{opt}}(s) = \arg\min_{0\le a\le 1} \mathcal{R}(s;a)$ denote the argument that minimizes the true risk $\mathcal{R}$. Consider the probability of deviation given by

$$P_e^{\text{MPE}} = \mathbb{P}\left(\left|a_{\text{opt}}(s) - a_{\text{opt}}(x)\right| \ge \delta\right), \tag{5}$$

for some $\delta > 0$. Using a first-order Taylor-series approximation of $a_{\text{opt}}(x)$ about s, and substituting $x = s + w$, we obtain $a_{\text{opt}}(x) \approx a_{\text{opt}}(s) + w\, a'_{\text{opt}}(s)$, where $'$ denotes the derivative. We would like to point out that the Taylor-series approximation is not rigorously justified here but it is reasonable to guide our analysis. The deviation probability P_e^{MPE} in (5) simplifies to $P_e^{\text{MPE}} = \mathbb{P}(|w| \ge \delta/\left|a'_{\text{opt}}(s)\right|)$. For additive Gaussian noise w with zero mean and variance σ^2, placing the Chernoff bound [36,37] on P_e^{MPE} leads to

$$P_e^{\text{MPE}} \le 2\exp\left(-\frac{\delta^2}{2\sigma^2\left|a'_{\text{opt}}(s)\right|^2}\right).$$

To ensure that P_e^{MPE} is less than α, for a given $\alpha \in (0,1)$, it suffices to have

$$\left|a'_{\text{opt}}(s)\right|^2 \le \frac{\delta^2}{2\sigma^2 \log\left(\frac{2}{\alpha}\right)}, \tag{6}$$

which translates to a lower-bound on the input SNR. Since there is no closed-form expression available for $a'_{\text{opt}}(s)$ in the context of the PE risk, we empirically obtain the range of input SNR values s^2/σ^2, for which (6) is satisfied.

Analogously, to satisfy an upper bound on the deviation probability P_e^{SURE} of the minimum in the case of SURE, for a given deviation $\delta > 0$, one must ensure that

$$\frac{s^6}{8\sigma^6}\left(\delta - \frac{\sigma^4}{(s^2+\sigma^2)\,s^2}\right)^2 \ge \log\left(\frac{2}{\alpha}\right). \tag{7}$$

The proof of (7) is given in Appendix A.

The minimum input SNR required to ensure $P_e \le \alpha$ for both SURE- and MPE-based shrinkage estimators is shown in Fig. 2, for different values of α and δ. The PE risk estimate is obtained by replacing s with x and setting $\epsilon = \sigma$. We observe that reducing the amount of deviation δ for a given probability α, or vice versa, leads to a higher input SNR requirement for both SURE and MPE. We also observe from Fig. 2 that, for given δ and α, SURE requires a higher input SNR than MPE to keep the δ-deviation probability under α. Also, for a given input SNR, the δ-deviation probability of the estimated shrinkage parameter $a_{\text{opt}}(x)$ from the optimum $a_{\text{opt}}(s)$ is smaller for MPE than SURE, thereby indicating that the MPE-based shrinkage is comparatively more reliable and robust than the SURE-based one at low input SNRs.

D) Unknown noise distributions

In practical applications, the distribution of noise may not be known in a parametric form and may also be multimodal. At best, one would have access to realizations of the noise, from which the distribution has to be estimated. In such cases, approximation of the noise p.d.f. using a GMM is a viable alternative [30,31], wherein one can estimate the parameters of the GMM using the expectation-maximization algorithm [38]. Gaussian mixture modeling is attractive as it comes with the guarantee that, asymptotically, as the number of Gaussians increases, the GMM approximation to a p.d.f. that has a finite number of discontinuities converges uniformly except at the points of the discontinuity [31]. The GMM approximation can be used even for non-Gaussian, unimodal distributions. For the GMM-based noise p.d.f.

$$f(w) = \sum_{m=1}^{M} \frac{\alpha_m}{\sigma_m\sqrt{2\pi}} \exp\left(-\frac{(w-\theta_m)^2}{2\sigma_m^2}\right), \tag{8}$$

https://doi.org/10.1017/ATSIP.2019.27 Published online by Cambridge University Press

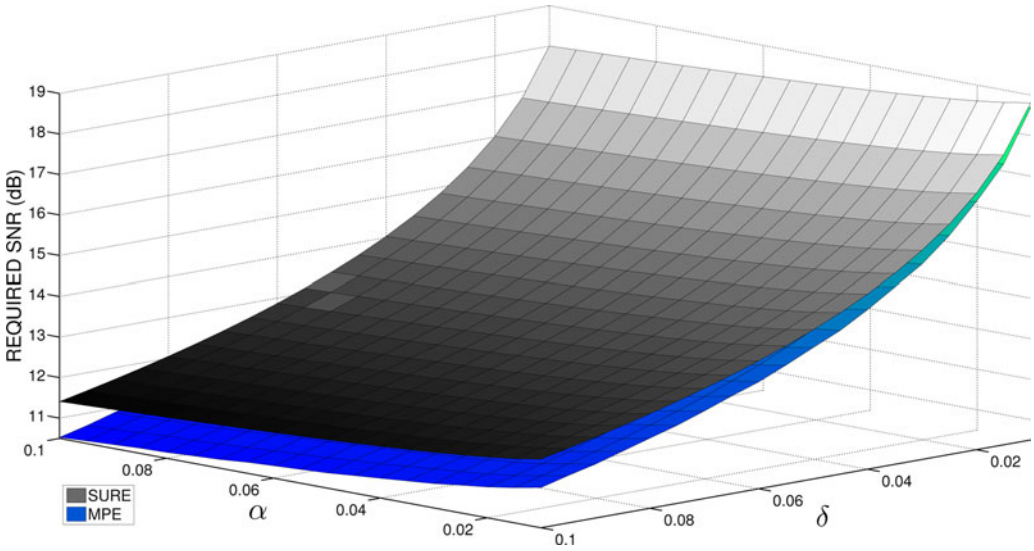

Fig. 2. (Color online) Input SNR requirement for SURE (black) and MPE (blue) to ensure that the probability of δ-perturbation of the minima is less than or equal to α.

the PE risk turns out to be

$$\widehat{\mathcal{R}} = \sum_{m=1}^{M} \alpha_m \left[Q\left(\frac{\epsilon - (a-1)s - \theta_m}{a\sigma_m} \right) + Q\left(\frac{\epsilon + (a-1)s + \theta_m}{a\sigma_m} \right) \right], \tag{9}$$

using (3). For illustration, consider the estimation of a scalar $s = 4$ in the transform domain from its noisy observation x. The additive noise is Laplacian distributed with zero mean and variance $\sigma^2 = 1$. The noise distribution is modeled using a GMM with $M = 4$ components and the corresponding PE risk estimate is obtained using (9) by setting $s = x$. In Fig. 3(a), we show a Laplacian p.d.f. and its GMM approximation. Fig. 3(b) shows the GMM approximation to a multimodal distribution. Figure 4(a) shows the PE risk based on the original Laplacian distribution as well as the GMM approximation, as a function of the shrinkage parameter a. The close match between the two indicates that the GMM is a viable alternative when the noise distribution is unknown or follows a complicated model. In Fig. 4(b), we plot the GMM-based PE risk and its estimate averaged over 100 independent trials. We observe that the locations of the minima of the actual risk and its estimate are in good agreement, thereby justifying the minimization of $\widehat{\mathcal{R}}$. The PE risk and its estimate are shown in Fig. 4(c) for the multimodal p.d.f. of Fig. 3(b).

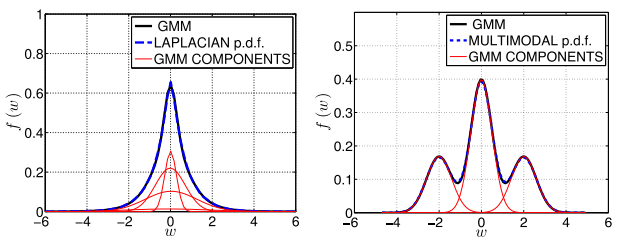

Fig. 3. (Color online) Original noise distribution and a GMM approximation: (a) Laplacian p.d.f. and its approximation using a four-component GMM; and (b) A multimodal p.d.f. and its three-component GMM approximation.

E) PE risk for subband shrinkage

Let a_J be the shrinkage factor applied to the set of coefficients x_i, $i \in J$ in subband J. The estimate $\widehat{\boldsymbol{s}}_J$ of the clean signal is obtained by $\widehat{\boldsymbol{s}}_J = a_J \boldsymbol{x}_J$, where $\boldsymbol{x}_J \in \mathbb{R}^{|J|}$ and $a_J \in [0, 1]$. For notational brevity, we drop the subscript J, as we did for pointwise shrinkage, and express the estimator as $\widehat{\boldsymbol{s}} = a\boldsymbol{x}$, where boldface letters indicate vectors. Analogous to pointwise shrinkage, the PE risk for subband shrinkage is defined as $\mathcal{R} = \mathbb{P}(\|\widehat{\boldsymbol{s}} - \boldsymbol{s}\|_2 > \epsilon)$, which, for $\widehat{\boldsymbol{s}} = a\boldsymbol{x}$, becomes $\mathcal{R} = \mathbb{P}(\|a\boldsymbol{w} + (a-1)\boldsymbol{s}\|_2 > \epsilon)$. For $\boldsymbol{w} \sim \mathcal{N}(0, \sigma^2 I)$,

$$\mathcal{R} = 1 - F(\theta | k, \lambda), \tag{10}$$

where $k = |J|$, $\lambda = \sum_{j=1}^{k} ((1-a)^2 s_j^2 / a^2\sigma^2)$, $\theta = (\epsilon / a\sigma)^2$, and $F(\theta \,|\, k, \lambda)$ is the c.d.f. of the non-central χ^2 distribution, given by

$$F(\theta \mid k, \lambda) = \sum_{m=0}^{\infty} \frac{\lambda^m e^{-\lambda/2}}{2^m m!} \frac{\gamma\left(\frac{\theta}{2}, \frac{k+2m}{2}\right)}{\Gamma\left(\frac{k+2m}{2}\right)},$$

where $\Gamma(a) = \int_0^\infty \exp(-t) t^{a-1}\, \mathrm{d}t$ and $\gamma(x, a) = \int_0^x \exp(-t) t^{a-1}\, \mathrm{d}t$.

Similar to pointwise shrinkage, we propose to obtain an estimate $\widehat{\mathcal{R}}$ of $\mathcal{R}$ for subband shrinkage estimators by replacing s_j with x_j. The optimum subband shrinkage factor is obtained by minimizing $\widehat{\mathcal{R}}$.

Fig. 5 shows the subband PE risk and its estimate versus a, where the underlying clean signal $\mathbf{s} \in \mathbb{R}^{|J|}$ is corrupted by Gaussian noise and the subband size is chosen to be $|J| = k = 8$. The clean signal $\mathbf{s}$ is generated by drawing samples from $\mathcal{N}(2 \times \mathbf{1}_k, I_k)$, where $\mathbf{1}_k$ and I_k denote a k-length vector of all ones and a $k \times k$ identity matrix, respectively. The observation $\boldsymbol{x}$ is obtained by adding zero-mean i.i.d. Gaussian noise to $\mathbf{s}$, with an input

https://doi.org/10.1017/ATSIP.2019.27 Published online by Cambridge University Press

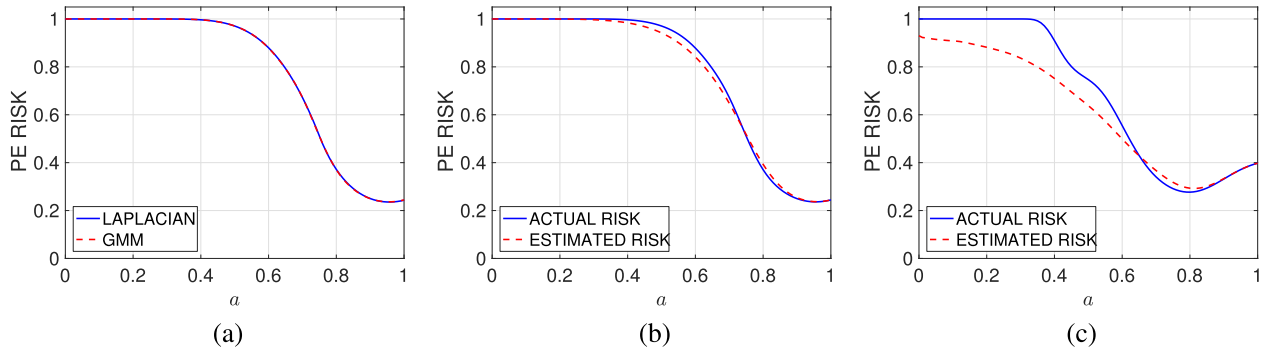

Fig. 4. (Color online) The PE risk estimate versus the shrinkage parameter a: (a) PE risk for Laplacian noise, considering the Laplacian p.d.f. and its GMM approximation; (b) GMM-based PE risk estimate for Laplacian noise; and (c) PE risk estimate for multimodal noise; for $\epsilon = \sigma$. The risk estimates are averaged over 100 Monte Carlo realizations.

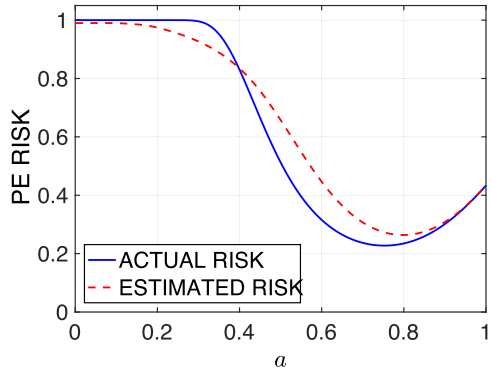

Fig. 5. (Color online) The PE risk and its estimate averaged over 100 Monte Carlo trials for the subband shrinkage estimator versus a; where $\epsilon = \sqrt{k}\sigma$, with k=8. The additive noise is Gaussian with $\text{SNR}_{\text{in}} = 5$ dB. In each trial, $\boldsymbol{s}$ is generated by drawing samples from $\mathcal{N}(2 \times \mathbf{1}_k, I_k)$.

SNR of 5 dB, where the input SNR is defined as $\text{SNR}_{\text{in}} = 10\log_{10}((1/k\sigma^2)\sum_{n=1}^{k} s_n^2)$ dB. The PE risk estimate is obtained by replacing $\boldsymbol{s}$ with $\boldsymbol{x}$ in (10), which does not drastically alter the minimum (cf. Fig. 5).

III. EXPERIMENTAL RESULTS FOR MPE-BASED DENOISING

The performance of the MPE-based pointwise and subband shrinkage estimator is validated on a synthesized harmonic signal (of length $N = 2048$) in Gaussian noise and the *Piece-Regular* signal (of length $N = 4096$) in Gaussian, Student's-*t*, multimodal, and Laplacian noise. The *Piece-Regular* signal has both smooth and rapidly-varying regions, making it a suitable candidate for the assessment of denoising performance.

A) Performance of pointwise-shrinkage estimator

1) Harmonic signal denoising

Consider the signal

$$s_n = \cos\left(\frac{5\pi n}{2048}\right) + 2\sin\left(\frac{10\pi n}{2048}\right), \quad 0 \le n \le 2047, \tag{11}$$

Table 1. Comparison of MPE, SURE-based shrinkage estimator and Wiener filter (WF) for different input SNRs. The output SNR values are averaged over 100 noise realizations.

Input SNR (dB)	Output SNR (dB)				
	MPE			SURE	WF
	$\epsilon = 3.5\sigma$	$\epsilon = 2.5\sigma$	$\epsilon = 1.5\sigma$		
−5.0	11.67	5.99	−0.18	−0.27	1.44
−2.5	14.42	8.62	2.34	2.23	3.96
0	17.02	10.96	4.80	4.71	6.35
2.5	19.08	13.36	7.31	7.21	8.79
5.0	21.25	15.52	9.72	9.64	11.09
7.5	22.93	18.26	12.32	12.23	13.60
10.0	25.34	20.57	14.77	14.69	15.92
12.5	26.91	22.79	17.26	17.17	18.20
15.0	28.77	25.05	19.66	19.59	20.33
17.5	30.74	27.44	22.20	22.12	22.57
20.0	32.65	29.61	24.61	24.54	24.60

in additive white Gaussian noise, with zero mean and variance σ^2. Since the denoising is carried out in the DCT [39] domain, the Gaussian noise statistics remain unaltered. For the purpose of illustration, we assume that σ^2 is known. In practice, σ^2 may not be known *a priori* and could be replaced by the robust median estimate [40] or the trimmed estimate [41]. The clean signal is estimated using inverse DCT after applying the optimum shrinkage. The denoising performance of the MPE and SURE-based approaches is compared in Table 1. In case of the Wiener filter, the power spectrum of the clean signal is estimated using the standard spectral subtraction technique [42,43]. We observe that MPE-based shrinkage with $\epsilon = 3.5\sigma$ is superior to SURE and Wiener filter (WF) by 8–12 dB. The comparison also shows that the performance of the MPE depends critically on the choice of ϵ.

2) Piece-Regular signal denoising

We consider several noisy realizations of the *Piece-Regular* signal, taken from the *Wavelab* toolbox [33], under Gaussian, Student's-*t*, multimodal (Fig. 3(b)), and Laplacian noise contaminations. The noise variance is assumed to be known. Notably, the Gaussian and Student's-*t* distributions of noise are preserved by an orthonormal transform [44], unlike the Laplacian distribution. Therefore, the

https://doi.org/10.1017/ATSIP.2019.27 Published online by Cambridge University Press

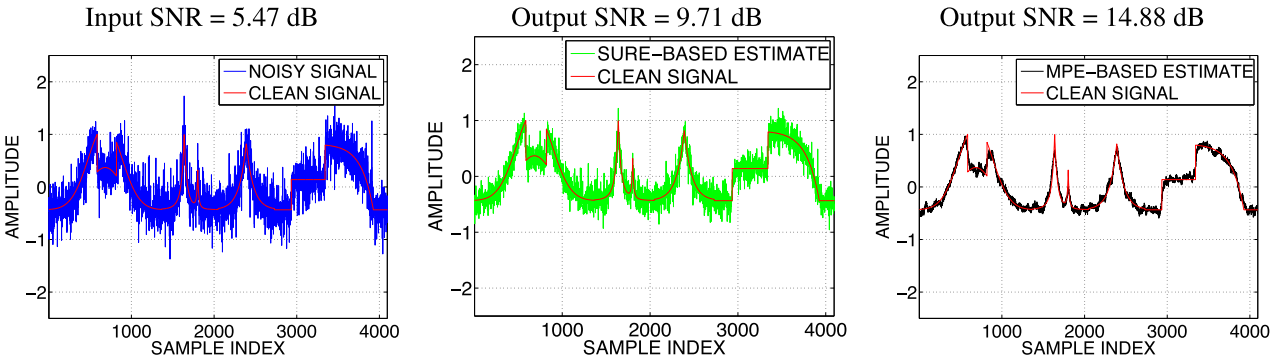

Fig. 6. (Color online) Denoising performance of the MPE- and SURE-based pointwise shrinkage estimators for the *Piece-Regular* signal corrupted by Laplacian noise. The PE risk is calculated by using a GMM approximation and setting $\epsilon = 3\sigma$.

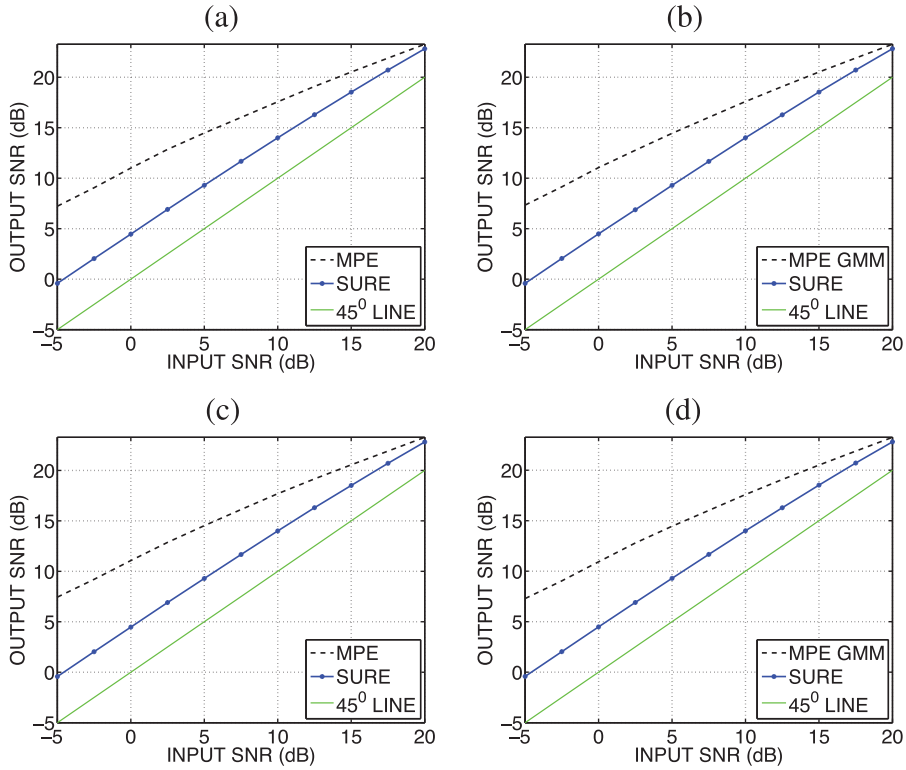

Fig. 7. (Color online) Output SNR versus input SNR corresponding to the MPE- and SURE-based pointwise shrinkages, under various noise distributions. The output SNR values are calculated by averaging over 100 independent noise realizations. (a) Gaussian noise, (b) Laplacian noise, (c) Student's-t noise, and (d) Multimodal noise.

PE estimates for Laplacian noise and multimodal noise are computed based on a four-component and a three-component GMM approximation, respectively, in the DCT domain. The denoised signal corresponding to Laplacian noise is shown in Fig. 6 to serve as an illustration. The MPE estimates are better than SURE estimates. The SNR plots in Fig. 7 indicate that the MPE outperforms SURE for the noise types under consideration and that the gains are particularly high in the input SNR range of -5 to 20 dB and tend to reduce beyond 20 dB.

3) Effect of ϵ on the denoising performance of MPE

Obtaining a closed-form expression for the ϵ that maximizes the output SNR is not straightforward. We determine the optimum ϵ empirically by measuring the SNR gain as a function of ϵ (cf. Fig. 8) for i.i.d. Gaussian noise. We observe that the output SNR exhibits a peak approximately at $\beta = \epsilon/\sigma = 3.5$ for the harmonic signal in (11) and at $\beta = 3$ for the *Piece-Regular* signal. As a rule of thumb, we recommend $\epsilon = 3\sigma$ for pointwise shrinkage estimators. Some insights into the role of ϵ and its choice will be presented in Section 4-4.3.

B) Performance of subband MPE shrinkage

To validate the performance of the MPE-based subband shrinkage estimator (cf. Section 2-2.5), we consider denoising of the *Piece-Regular* signal in additive Gaussian noise. The clean signal and its noisy measurement are shown in Fig. 9(a). Denoising is carried out by grouping k adjacent DCT coefficients to form a subband. The denoised signals

https://doi.org/10.1017/ATSIP.2019.27 Published online by Cambridge University Press

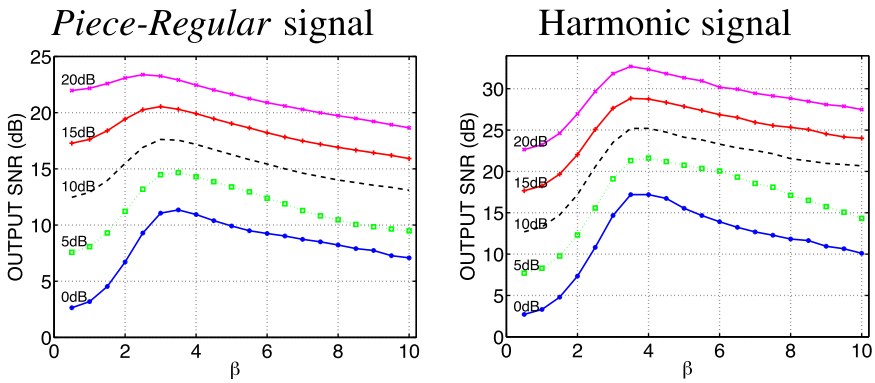

Fig. 8. (Color online) Average output SNR of the pointwise MPE shrinkage as a function of $\beta = \epsilon/\sigma$, for different values of the input SNR. The output SNR curves peak when $\beta \approx 3$.

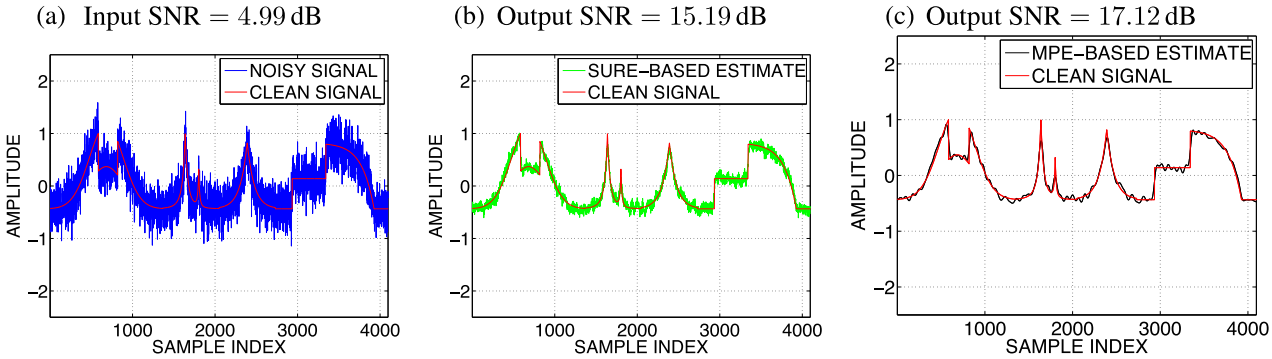

Fig. 9. (Color online) Comparison of denoising performance of the subband shrinkage estimators using MPE and SURE, for the *Piece-Regular* signal corrupted by additive Gaussian noise. The subband size is taken as $k = 16$ and the value of ϵ is $1.75\sqrt{k}\sigma$.

obtained using SURE and MPE are shown in Fig. 9(b) and 9(c), respectively. The subband size k is chosen to be 16 and the parameter ϵ is set equal to $1.75\sqrt{k}\sigma$, a value that was determined experimentally and found to be nearly optimal. We observe that the MPE gives 1 dB improvement in SNR than the SURE approach.

The variation of the output SNR is also studied as a function of k (cf. Fig. 10). We experimented with $\epsilon = 3\sigma$, $\epsilon = 1.75\sqrt{k}\sigma$, and $\epsilon = 1.25\sqrt{k}\sigma$ corresponding to subband sizes $k = 1$, $k \in [2, 16]$, and $k > 16$, respectively. For both SURE and MPE, as k increases, the output SNR also increases and eventually saturates for $k \geq 40$. For input SNR below 15 dB, MPE gives a comparatively higher SNR than SURE, and the margin diminishes with increase in input SNR or the subband size k. The degradation in performance of SURE for low SNRs is due to the large error in estimating the MSE at such SNRs. The SURE-based estimate of MSE becomes increasingly reliable as k increases, thereby leading to superior performance.

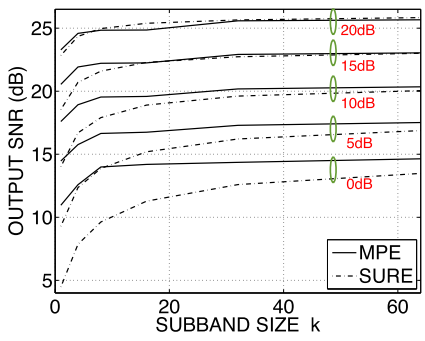

Fig. 10. (Color online) Output SNR versus subband size k, averaged over 100 noise realizations, for different input SNRs. The output SNR of MPE is consistently superior to that obtained using SURE, especially when $k \leq 40$ and the input SNR is below 15 dB.

IV. THE EXPECTED ℓ_1 DISTORTION

To eliminate the dependence of MPE on ϵ, we consider the accumulated probability of error, namely $\int_0^\infty \mathbb{P}(|\widehat{s} - s| > \epsilon)\mathrm{d}\epsilon$ as the risk to be minimized. For a nonnegative random variable Y, we know that $\mathcal{E}\{Y\} = \int_0^\infty \mathbb{P}(Y > \epsilon)\mathrm{d}\epsilon$. Therefore, the accumulated probability of error is the expected ℓ_1 distortion:

$$\mathcal{R}_{\ell_1}(a, s) = \mathcal{E}\{|\widehat{s} - s|\} = \int_0^\infty \mathbb{P}\left(|\widehat{s} - s| > \epsilon\right) \mathrm{d}\epsilon. \qquad (12)$$

For Gaussian noise distribution,

$$\mathcal{R}_{\ell_1}(a, s) = \int_0^\infty Q\left(\frac{\epsilon - (a-1)s}{a\sigma}\right) \mathrm{d}\epsilon + \int_0^\infty Q\left(\frac{\epsilon + (a-1)s}{a\sigma}\right) \mathrm{d}\epsilon. \qquad (13)$$

https://doi.org/10.1017/ATSIP.2019.27 Published online by Cambridge University Press

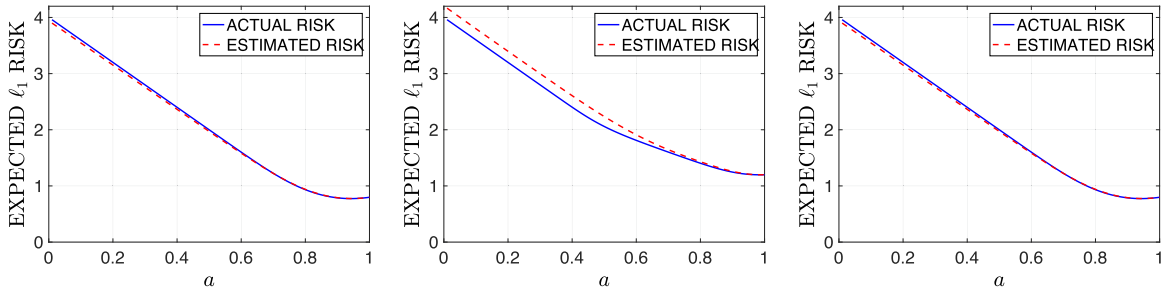

Fig. 11. (Color online) The expected ℓ_1-risk and its estimate versus a, averaged over 100 noise realizations for: (a) Gaussian noise with $\sigma^2 = 1$; (b) three-component GMM; and (c) a four-component GMM approximation to the Laplacian distribution.

Denoting $u = \epsilon - (a-1)s/a\sigma$ and $\mu = -(a-1)s/a\sigma$, the first integral in (13) is evaluated as

$$\begin{aligned}
&\int_0^\infty Q\left(\frac{\epsilon-(a-1)s}{a\sigma}\right) d\epsilon \\
&= a\sigma \int_\mu^\infty Q(u)\, du \\
&= a\sigma \left(\int_0^\infty Q(u)\, du - \int_0^\mu Q(u)\, du\right) \\
&= a\sigma \left(\frac{1}{\sqrt{2\pi}} - \mu Q(\mu) - \frac{1}{\sqrt{2\pi}}\left(1 - e^{-\mu^2/2}\right)\right) \\
&= a\sigma \left(\frac{e^{-\mu^2/2}}{\sqrt{2\pi}} - \mu Q(\mu)\right). \qquad (14)
\end{aligned}$$

The second term in (13) can be evaluated by replacing μ with $-\mu$ in (17). Combining both integrals, we obtain the expression for the *expected* ℓ_1 *distortion*:

$$\begin{aligned}
\mathcal{R}_{\ell_1}(a,s) &= a\sigma\left[\sqrt{2/\pi}\, e^{-\mu^2/2} - \mu Q(\mu) + \mu Q(-\mu)\right] \\
&= a\sigma\left[\sqrt{\frac{2}{\pi}} \exp\left(-\frac{(a-1)^2 s^2}{2a^2\sigma^2}\right) + 2\frac{(a-1)s}{a\sigma}\right. \\
&\quad \left. \times Q\left(-\frac{(a-1)s}{a\sigma}\right) - \frac{(a-1)s}{a\sigma}\right].
\end{aligned}$$

An estimate of the expected ℓ_1 distortion is calculated by replacing s with x. In Fig. 11(a), we show the variation of the original ℓ_1 distortion and its estimate obtained by setting $s = x$, as functions of a, averaged over 100 independent realizations of $\mathcal{N}(0,1)$ noise. The actual parameter value is $s = 4$. The figure shows that the minimum of the expected ℓ_1 risk is close to that of its estimate.

In principle, one could iteratively minimize the ℓ_1 distortion by starting with $\widehat{s} = x$ and successively refining it. Such an approach is given in Algorithm 1. An illustration of the denoising performance of the iterative algorithm is deferred to Section 5.

Algorithm 1 Iterative minimization of the expected ℓ_1 distortion.

1. **Initialization:** Set $j \leftarrow 1, \widehat{s}^{(j)} \leftarrow x$, and N_{iter} = Maximum iterationcount.
2. **Iterate until** j **exceeds** N_{iter}**:**
   - Find $a_{\text{opt}}^{(j)} = \arg\min_{0 \le a \le 1} \mathcal{R}_{\ell_1}(a, \widehat{s}^{(j)})$ by a grid-search.
   - $j \leftarrow j + 1$.
   - Compute $\widehat{s}^{(j)} = a_{\text{opt}}^{(j-1)} x$.
3. **Output:** Denoised estimate $\widehat{s}^{(j)}$.

A) Expected ℓ_1 risk using GMM approximation

For the GMM p.d.f. in (8), the expected ℓ_1 distortion evaluates to (cf. Appendix B for the derivation)

$$\mathcal{R}_{\ell_1} = \sum_{m=1}^{M} a\alpha_m\sigma_m \left(\sqrt{\frac{2}{\pi}} e^{-\mu_m^2/2} - 2\mu_m Q(\mu_m) + \mu_m\right), \qquad (15)$$

where $\mu_m = -((a-1)s + \theta_m)/a\sigma_m$. The expected ℓ_1 risk and its estimate for a multimodal (cf. Fig. 3(b)) and Laplacian noise p.d.f.s are shown in Figs 11(b) and 11(c), respectively. We observe that, in both cases, the locations of the minima of the true risk and its estimate are in good agreement.

B) Optimum shrinkage versus posterior SNR

We next study the behavior of a_{opt} for different input SNRs to compare the denoising capabilities of the MPE and the expected ℓ_1-distortion-based shrinkage estimators. The optimal pointwise shrinkage parameter a_{opt} for Gaussian noise statistics, obtained by minimizing SURE, PE risk estimate, and the estimated ℓ_1 risk, for different values of the *a posteriori* SNR x^2/σ^2 is plotted in Fig. 12(a). To illustrate the effect of ϵ, the variation of a_{opt} versus *a posteriori* SNR for MPE corresponding to Gaussian noise is shown in Fig. 12(b), for different ϵ. We observe that the shrinkage profiles are characteristic of a reasonable denoising algorithm, as Fig. 12(a) and 12(b) exhibit that the shrinkage parameters increase as the *a posteriori* SNR increases. Whereas in

https://doi.org/10.1017/ATSIP.2019.27 Published online by Cambridge University Press

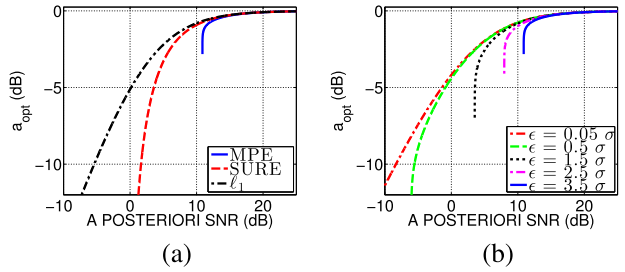

Fig. 12. (Color online) Shrinkage parameter profiles as a function of the *a posteriori* SNR, corresponding to different risk functions: (a) MPE, SURE, expected ℓ_1 distortion; and (b) MPE for different values of ϵ. The shrinkage factor a_{opt} is plotted on a log scale mainly to highlight the fine differences among various attenuation profiles.

the case of MPE, the choice of ϵ is crucial, the expected ℓ_1 distortion does not require tuning such a parameter.

C) Shrinkage profiles of estimators based on MPE/ℓ_1 vis-à-vis SURE

To gain insight into the denoising behavior of the MPE and the ℓ_1-minimization-based shrinkage estimators, it is instructive to study the resulting thresholding functions shown in Fig. 13. Figures 13(a) and 13(b) correspond to Gaussian noise. The comparison of the shrinkage functions based on MPE, ℓ_1 risk, and SURE (cf. Fig. 13(a)) reveals that MPE results in a hard-thresholding-type shrinkage, whereas the ℓ_1 risk and SURE produce shrinkage profiles similar to a soft-thresholding function (the threshold operator is called "soft" as it is a continuous function of the input data [27]). The shrinkage functions resulting from MPE with different ϵ are shown in Fig. 13(b). We observe that increasing ϵ is tantamount to applying a larger attenuation on the noisy signal. In Fig. 13(c), we demonstrate the variation of the MPE-based shrinkage with respect to two different noise distributions, namely, Gaussian and Laplacian, having the same first- and second-order moments. We observe that the MPE-based estimator corresponding to Laplacian noise attenuates larger amplitudes more than its Gaussian counterpart. Such an attenuation profile helps suppress Laplacian noise more effectively, considering the heavier tail of the Laplacian distribution. A distribution-specific shrinkage profile also distinguishes MPE from the conventional MSE-based estimators. In contrast, the MSE-based shrinkage estimator, given by [28]: $\hat{s}_{\text{MSE}} = \max\{0, 1 - \sigma^2/x^2\}x$, relies only on moments up to second order.

An inspection of the shrinkage functions corresponding to different ϵ values in Fig. 13(b) reveals a particularly interesting property: Transform coefficients with magnitudes smaller than ϵ are set to zero by the MPE-based shrinkage function. Thus, the PE risk essentially results in a distribution-specific threshold operator parameterized by ϵ. In this context, we recall an observation made by Johnstone et al. in [45]: A threshold value of 3σ to 4σ or higher works well for signals that are sparse in the transform domain, whereas a threshold value of 2σ or lower is more appropriate for a dense signal. The optimal ϵ values obtained empirically for the harmonic and the *Piece-Regular* signals (cf. Section 3-3.1.3.1.3, Fig. 8) also support this statement. Since the harmonic signal is sparser than the *Piece-Regular* signal in the DCT domain, the optimal ϵ tends to be higher for the harmonic signal as compared with that for the *Piece-Regular* signal. Thus, the shrinkage functions shown in Fig. 13(b) provide us with valuable insights on how to fix ϵ. If the underlying signal is sparse in the transform considered, one should fix ϵ in the range $[3\sigma, 4\sigma]$, whereas for dense signals, one should set ϵ to about 2σ.

V. PERFORMANCE OF THE EXPECTED ℓ_1 DISTORTION-BASED POINTWISE SHRINKAGE ESTIMATOR

In a practical denoising application, we have only one noisy realization from which the clean signal has to be estimated. However, it is instructive to consider the case of multiple realizations as it throws some light on the performance comparisons vis-à-vis other estimators such as the ML estimator. Consider the observation model $\boldsymbol{x}^{(m)} = \boldsymbol{s} + \boldsymbol{w}^{(m)}$ in $\mathbb{R}^n$, $1 \le m \le M$, where one has access to M noisy realizations of the signal $\mathbf{s}$, and the noise vectors $\boldsymbol{w}^{(m)}$ are drawn independently from the $\mathcal{N}(\mathbf{o}, \sigma^2 I_n)$ distribution. The ML estimator of the i^{th} signal coefficient s_i is given by $\hat{s}_{\text{ML},i} = (1/M)\sum_{m=1}^{M} x_i^{(m)}$, where $x_i^{(m)}$ is the i^{th} component of $\mathbf{x}^{(m)}$. Dropping the subscript i, as each coefficient is treated independently of the others, the shrinkage estimator takes the

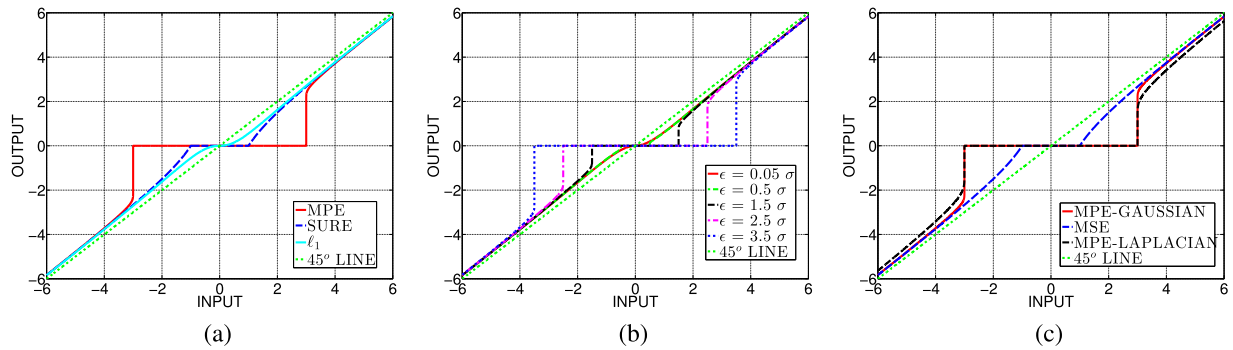

Fig. 13. (Color online) Thresholding functions corresponding to different risk functions: (a) MPE, SURE, expected ℓ_1 distortion; (b) MPE for various values of ϵ; and (c) MPE, MSE. In (a) and (b), the noise considered is Gaussian. In (c), both Laplacian and Gaussian noise types have been considered with $\sigma = 1$ and $\epsilon = 3\sigma$.

https://doi.org/10.1017/ATSIP.2019.27 Published online by Cambridge University Press

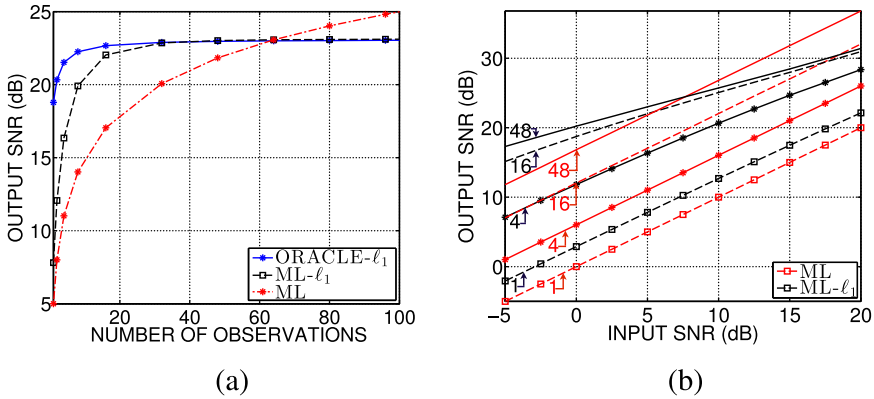

Fig. 14. (Color online) Comparison of denoising performance as a function of number of observations for a *Piece-Regular* signal in Gaussian noise: (a) Output SNR corresponding to input SNR 5 dB; and (b) Output SNR as a function of the number of observations M and input SNR. The numerical values on the curves indicate the corresponding values of M. In both (a) and (b), the results are averaged over 100 independent noise realizations.

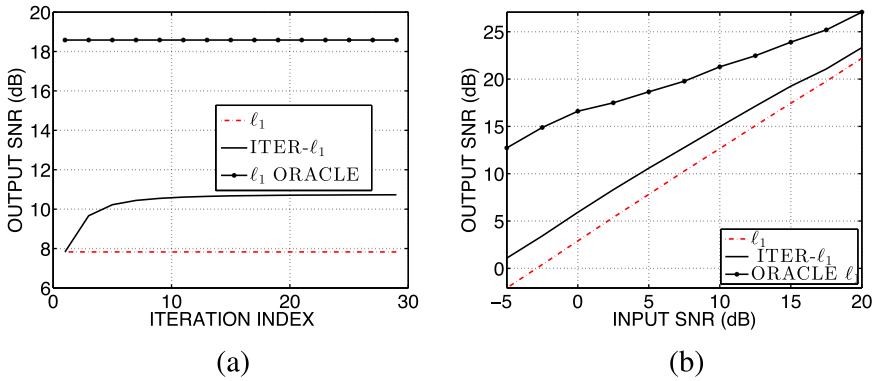

Fig. 15. (Color online) Performance of ℓ_1-risk-minimization-based pointwise shrinkage estimator (*Piece-Regular* signal in Gaussian noise): (a) Output SNR versus iterations (input SNR 5 dB); and (b) Output SNR versus input SNR, averaged over 100 independent noise realizations. The number of iterations in Algorithm 1 is fixed at $N_{\text{iter}} = 20$.

form $\widehat{s} = a_{\text{opt}}\hat{s}_{\text{ML}}$. To study the behavior of the estimate with respect to M, we consider two variants: (i) where a_{opt} is obtained by minimizing $\mathcal{R}_{\ell_1}(s, a)$, referred to as the oracle-ℓ_1; and (ii) where a_{opt} is chosen to minimize $\mathcal{R}_{\ell_1}(\hat{s}_{\text{ML}}, a)$, referred to as ML-$\ell_1$. The output SNR as a function of M for the *Piece-Regular* signal, corresponding to an input SNR of 5 dB, is shown in Fig. 14(a). For all three estimators, namely, oracle-ℓ_1, ML-ℓ_1, and the ML estimate, the output SNR increases with M. However, for the oracle-ℓ_1 and the ML-ℓ_1 estimators, the output SNR stagnates as M increases beyond 40. For $M \leq 60$, the oracle-ℓ_1 and the ML-ℓ_1 shrinkage estimators exhibit better performance compared with the ML estimator. As one would expect, the performance of the ML-ℓ_1 estimator matches with that obtained using the oracle-ℓ_1 as M becomes large, because the ML estimate converges in probability to the true parameter. For $M = 1$, which is often the case in practice, the ML-ℓ_1 estimate significantly dominates the ML estimator as seen in Fig. 14(a). The SNR gain over the ML estimator could be further improved by using the iterative minimization algorithm introduced in Section 4 (cf. Algorithm 1). The performance of the ML-ℓ_1 and the ML estimators, for different values of M and input SNR is shown in Fig. 14(b). The figures show that for small values of SNR and M, the ML-ℓ_1 estimate outperforms the ML estimator. This is of significant importance in a practical setting where we have only one noisy realization ($M = 1$).

A) Iterative minimization of the expected ℓ_1-risk

When $M = 1$, the ML-ℓ_1 estimator is obtained by minimizing $\mathcal{R}_{\ell_1}(x, a)$, where x is the noisy version of s. We refer to this estimate as the non-iterative ℓ_1-based shrinkage estimator. Following Algorithm 1, one could iteratively refine the estimate, starting from x. We compare the non-iterative ℓ_1-based estimator with its iterative counterpart (ITER-ℓ_1), and present the results in Figs 15,16, and 17, corresponding to Gaussian, multimodal (Fig. 3(b)), and a GMM approximation to the Laplacian noise, respectively. The output SNR obtained using the *oracle*-ℓ_1 estimator, calculated by minimizing $\mathcal{R}_{\ell_1}(s, a)$, is also shown for benchmarking the performance.

We make the following observations from Figs 15,16, and 17: (i) the output SNR increases with iterations, albeit marginally after about 10 iterations; (ii) the iterative method consistently dominates the non-iterative one, with an overall SNR improvement of about 2–3 dB, for input SNR in the range −5 dB to 20 dB; and (iii) the SNR gain of the iterative

https://doi.org/10.1017/ATSIP.2019.27 Published online by Cambridge University Press

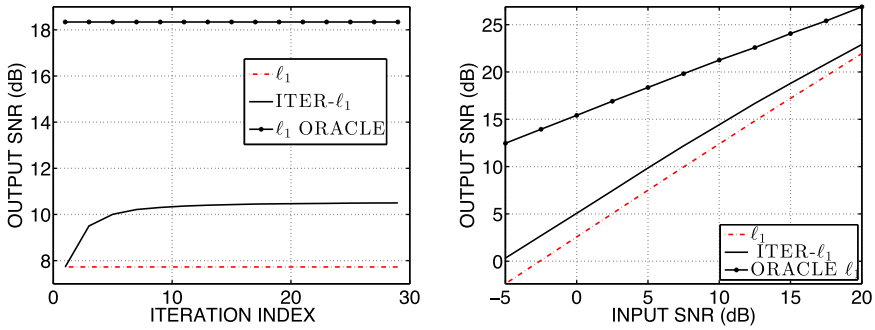

Fig. 16. (Color online) Performance of pointwise shrinkage estimator based on ℓ_1 risk minimization: (a) Output SNR versus iterations, corresponding to the *Piece-Regular* signal contaminated by noise (SNR 5 dB) whose p.d.f. is given in Fig. 3(b); and (b) Output SNR versus input SNR (averaged over 100 independent noise realizations). The number of iterations in Algorithm 1 is set to $N_{\text{iter}} = 20$.

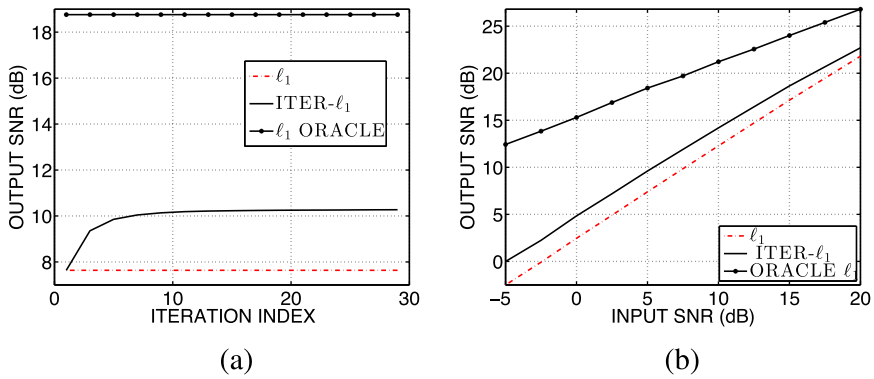

Fig. 17. (Color online) Performance of pointwise shrinkage estimator obtained by ℓ_1 risk minimization: (a) Output SNR versus iterations for the *Piece-Regular* signal in Laplacian noise with an input SNR of 5 dB; and (b) Output SNR versus input SNR for $N_{\text{iter}} = 20$. In (a) and (b), the Laplacian distribution is modeled using a four-component GMM to calculate the ℓ_1-risk estimate. The results displayed in (b) are obtained after averaging over 100 realizations.

technique also reduces for higher input SNR, similar to other denoising algorithms.

VI. PERFORMANCE ASSESSMENT OF MPE AND ℓ_1-RISK MINIMIZATION ALGORITHMS VERSUS BENCHMARK DENOISING ALGORITHMS

We compare the MPE and the ℓ_1-based shrinkage estimators with three denoising algorithms: (i) wavelet soft-thresholding[1] (ST) [35]; (ii) the SURE-LET denoising algorithm[2] [5]; and (iii) smooth sigmoid shrinkage (SS) [13] in the wavelet domain[3]. In [35], a wavelet-based soft-thresholding scheme is used for denoising, with the threshold selected as $\tau = \sigma\sqrt{2\ln(N)}$ for an N-length signal. The SURE-LET technique employs a *linear expansion of thresholds* (LET), which is a linear combination of elementary denoising functions and optimizes for the coefficients by minimizing the SURE criterion. In [13], a smooth sigmoid shrinkage is applied on the wavelet coefficients to achieve denoising, and the parameters of the sigmoid, which control the degree of attenuation, are obtained by minimizing the SURE objective. We consider ECG signals taken from the *PhysioBank* database, and the *HeaviSine* and *Piece-Regular* signals taken from *Wavelab* toolbox for performance evaluation.

The noise is assumed to follow a Gaussian distribution and the output SNR values are averaged over 100 independent realizations. The noise variance is estimated using a median-based estimator [40], which was also employed by Luisier et al.[1] and Donoho[2]. In SURE-LET, SS, and ST techniques, denoising is performed using Symmlet-4, with three levels of decomposition, as these settings were found to be the best for the ECG signal (following [20]). In the case of MPE and ℓ_1-based shrinkage estimators, denoising is performed in the DCT domain. We use the shorthand notations MPE and MPE-SUB to denote the pointwise and subband shrinkage estimators, respectively. The corresponding SURE-based subband shrinkage estimator is denoted as SURE-SUB. We set $k = 16$ and $\epsilon = 1.75\sqrt{k}\sigma$ for computing the subband shrinkage parameters. These parameters have not been specifically optimized; however, experimentally they were found to work well. The output SNR as a function of the input SNR, obtained using various algorithms, is shown in Fig. 18.

[1]A Matlab implementation is included in the *Wavelab* toolbox available at: http://statweb.stanford.edu/~wavelab/.

[2]A MATLAB implementation of the SURE-LET algorithm is available at: http://bigwww.epfl.ch/demo/suredenoising.

[3]Pastor et al. kindly provided the MATLAB implementation of their denoising technique [13], which facilitated the comparisons reported in this paper.

https://doi.org/10.1017/ATSIP.2019.27 Published online by Cambridge University Press

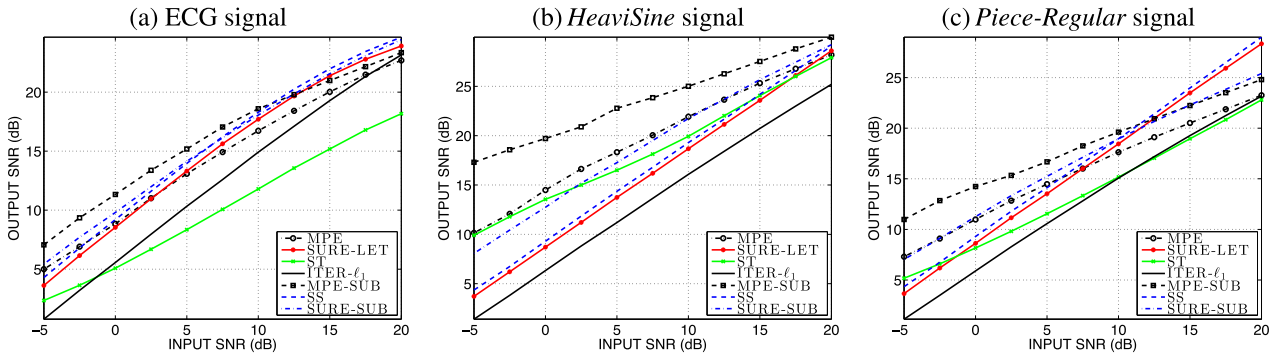

Fig. 18. (Color online) Output SNR of various denoising algorithms, averaged over 100 noise realizations, corresponding to different input SNRs.

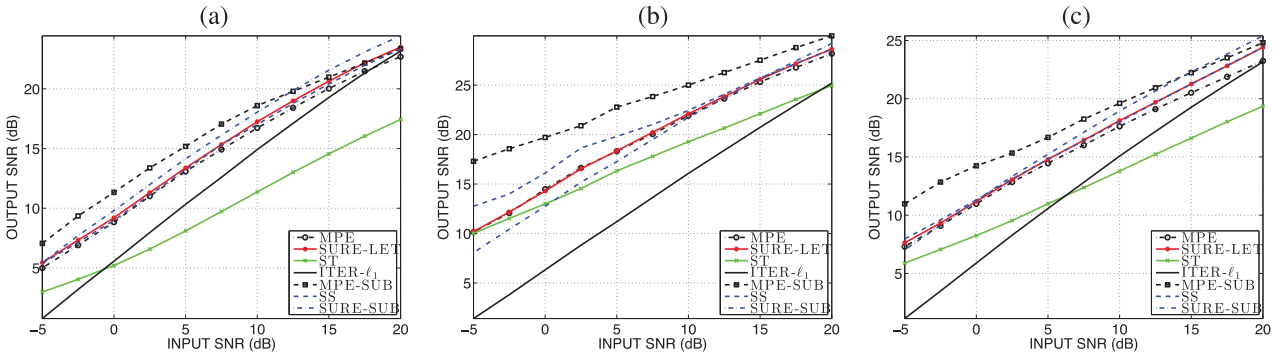

Fig. 19. (Color online) DCT-domain performance comparison of various denoising algorithms for different input SNRs, results are averaged over 100 noise realizations. (a) ECG signal, (b) *HeaviSine* signal, and (c) *Piece-Regular* signal

From the ECG signal denoising performance shown in Fig. 18(a), we observe that the MPE estimate consistently dominates soft-thresholding-based denoising for input SNRs ranging from −5 dB to 20 dB. The iterative ℓ_1-distortion-based shrinkage estimator (20 iterations) yields lower output SNR compared with the MPE-based estimate for input SNR values in the range −5 to 17.5 dB, but surpasses it for relatively higher values of input SNR (17.5–20 dB). The SURE-LET and SS algorithms dominate both MPE and ℓ_1-based shrinkage estimators, because they use more sophisticated denoising functions in the transform domain, thereby offering greater flexibility. For input SNR in the range 0–20 dB, the expected ℓ_1-distortion-based shrinkage estimator consistently outperforms the soft-thresholding-based technique. We observe from Fig. 18 that the MPE-subband estimator is better than the competing algorithms at low input SNR.

In the preceding discussion, we considered DCT-domain shrinkage for the proposed methods and wavelet-based denoising for the benchmark methods, because these are the best-case performance scenarios. In order to maintain uniformity in comparison, we have also considered the scenarios where all the techniques operate either in the DCT or wavelet domain. Figure 19 shows a performance comparison in the case of DCT. For input SNRs in the range −5 to 15 dB, MPE-SUB shows the best denoising performance. Next, we address wavelet-domain denoising using MPE. We consider the same denoising framework as that of Atto and

Table 2. Performance comparison of wavelet domain denoising of ECG signal. The output SNR values are averaged over 100 noise realizations.

Input	Output SNR (dB)						
SNR (dB)	SURE-LET	SS	SURE-SUB	ST	MPE	MPE-SUB	ITER-ℓ_1
−5.00	3.61	4.29	4.50	2.34	3.32	**4.55**	−0.09
−2.50	6.15	6.80	6.99	3.63	5.83	**7.06**	2.41
0	8.54	9.19	9.40	5.08	8.29	**9.47**	4.88
2.50	10.96	11.60	11.79	6.68	10.67	**11.85**	7.35
5.00	13.31	13.96	14.09	8.34	13.06	**14.17**	9.83
7.50	15.61	16.22	16.34	10.06	15.33	**16.38**	12.25
10.00	17.70	**18.29**	18.23	11.79	17.42	18.28	14.68
12.50	19.66	**20.25**	20.04	13.56	19.39	19.98	17.00
15.00	21.38	**21.97**	21.04	15.18	21.13	21.34	19.28
17.50	22.78	**23.43**	22.88	16.79	22.57	22.44	21.38
20.00	23.90	**24.65**	24.05	18.17	23.69	23.46	23.26

https://doi.org/10.1017/ATSIP.2019.27 Published online by Cambridge University Press

Table 3. Performance comparison for wavelet domain denoising of *HeaviSine* signal. Results presented are averaged over 100 noise realizations.

Input	Output SNR (dB)						
(l)2-8 SNR (dB)	SURE-LET	SS	SURE-SUB	ST	MPE	MPE-SUB	ITER-ℓ_1
−5.00	3.71	4.35	4.57	**9.91**	3.41	4.65	−0.04
−2.50	6.19	6.73	7.07	**11.78**	5.87	7.11	2.40
0	8.70	9.30	9.58	**13.54**	8.39	9.63	4.91
2.50	11.20	11.84	12.10	**14.99**	10.97	12.16	7.45
5.00	13.74	14.35	14.54	**16.49**	13.42	14.67	9.89
7.50	16.18	16.79	17.03	**18.13**	15.86	17.10	12.41
10.00	18.67	19.29	19.54	**19.93**	18.34	19.59	14.93
12.50	21.14	21.67	21.33	22.01	20.77	**22.05**	17.40
15.00	23.58	24.19	24.32	24.05	23.25	**24.41**	19.92
17.50	26.10	26.61	26.69	26.00	25.72	**26.79**	22.39
20.00	28.60	**29.08**	28.89	27.90	28.18	29.02	24.88

Table 4. Performance comparison for wavelet domain denoising of the *Piece-Regular* signal. The output SNR values are averaged over 100 noise realizations.

Input	Output SNR (dB)						
SNR (dB)	SURE-LET	SS	SURE-SUB	ST	MPE	MPE-SUB	ITER-ℓ_1
−5.00	3.65	4.33	4.53	**5.17**	3.37	4.59	−0.08
−2.50	6.18	6.74	7.02	6.59	5.83	**7.08**	2.37
0	8.62	9.30	9.50	8.16	8.34	**9.56**	4.90
2.50	11.14	11.78	11.97	9.81	10.77	**12.03**	7.39
5.00	13.52	14.11	14.30	11.54	13.19	**14.36**	9.87
7.50	16.00	16.57	16.68	13.33	15.57	**16.75**	12.35
10.00	18.44	**18.95**	18.82	15.16	17.96	18.93	14.86
12.50	20.97	**21.41**	20.91	17.04	20.42	20.98	17.34
15.00	23.49	**23.97**	22.78	18.95	22.99	22.78	19.85
17.50	25.92	**26.48**	24.40	20.83	25.41	24.22	22.29
20.00	28.34	**28.99**	25.93	22.79	27.86	25.55	24.82

Pastor [13], but replace their sigmoid-shrinkage function with MPE-based shrinkage in every subband. The parameter ϵ has been set to $\epsilon = 3\sigma$ for MPE and $\epsilon = 1.25\sigma\sqrt{k}$ for MPE-SUB, where k is the subband signal length. Tables 2 and 4 show that for input SNR below 10 dB, MPE-SUB gives better results than the competing algorithms. For the *HeaviSine* signal (cf. Table 3), soft-thresholding technique gave a better performance than the other techniques. However, MPE-SUB leads to a better performance than SURE-LET, SURE-SUB, and SS. The wavelet-domain denoising results are shown in a tabular form and DCT-domain denoising in figures, since the relative margin of improvement is smaller in the case of wavelet. When all the techniques operate in the DCT domain, the margin of improvement is maximum in the case of MPE.

VII. CONCLUSIONS

We have proposed a new framework for signal denoising based on a novel criterion, namely the probability of error. Our framework is applicable to scenarios where the noise is i.i.d. and additive. Denoising is performed by means of optimum transform-domain shrinkage in the sense of minimum probability of error. We have considered both pointwise and subband shrinkage estimators within the MPE paradigm. The performance of the proposed MPE estimators depends on the choice of the error-tolerance parameter ϵ, which was determined empirically considering SNR gain as the quantity of interest. The parameter ϵ also acts as the threshold for the shrinkage function. We also proposed an alternative by integrating the probability of error for $\epsilon \in (0, +\infty)$, which led to the expected ℓ_1 distortion, which was minimized iteratively. We demonstrated that iterations improve the denoising performance and the resulting shrinkage estimator outperforms the classical ML estimator when the number of observations is small or the input SNR is low.

For performance comparisons with benchmarking denoising algorithms, we considered real ECG signals and *Wavelab* signals in Gaussian noise. Experimentally, we have found that an increase in subband size leads to a higher output SNR, and saturates beyond a point. When the subband size is small or the input SNR is low, MPE outperforms SURE.

We demonstrated that the optimum shrinkage parameter obtained by minimizing estimates of the PE/ℓ_1 distortions increases monotonically with increase in the *a posteriori* SNR. This behavior of the shrinkage parameter is essential for denoising. A theoretical characterization of this behavior is needed and may lead to interesting inferences, which could lead to a rigorous convergence proof for the proposed iterative expected ℓ_1 distortion minimization technique. Another important observation is that, for lower input

https://doi.org/10.1017/ATSIP.2019.27 Published online by Cambridge University Press

SNRs, the proposed denoising framework yields a higher output SNR than MSE-based techniques. The improvement may be 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 conventional MSE-based optimization. To the best of our knowledge, this is the first attempt at demonstrating competitive denoising performance with probability of error chosen as the distortion measure in a non-Bayesian setting.

Following our preliminary contribution [28], recently, Kudryavtsev and Shestakov considered the probability that the maximum error between the estimates and true wavelet transform coefficients exceeds a critical value and analyzed the asymptotic behavior of the resulting optimal minimax threshold value considering specifically hard- and soft-thresholding schemes [46]. However, the advantages of such a strategy in the context of a denoising task are yet to be ascertained.

ACKNOWLEDGMENTS

This work was supported by Indian Space Research Organization – Indian Institute of Science (ISRO-IISc) Space Technology Cell, and an IMPRINT project (No. 6000) funded by the Ministry of Human Resource Development and the Ministry of Health and Family Welfare, Government of India.

REFERENCES

1 Poor H.V.: An Introduction to Signal Detection and Estimation, Springer, New York, 1994.

2 Kay S.; Eldar Y.C.: Rethinking biased estimation. *IEEE Signal Process. Mag.*, **25** (3) (2008), 133–136.

3 Stein C.M.: Estimation of the mean of a multivariate normal distribution. *Ann. Stat.*, **9** (6) (1981), 1135–1151.

4 Eldar Y.C.: Generalized SURE for exponential families: applications to regularization. *IEEE Trans. Signal Process.*, **57** (2) (2009), 471–481.

5 Luisier F.; Blu T.; Unser M.: A new SURE approach to image denoising: interscale orthonormal wavelet thresholding. *IEEE Trans. Image Process.*, **16** (3) (2007), 593–606.

6 Luisier F.; Blu T.; Unser M.: SURE-LET for orthonormal wavelet-domain video denoising. *IEEE Trans. Circ. Syst. Video Tech.*, **20** (6) (2010), 913–919.

7 Luisier F.; Blu T.: The SURE-LET multichannel image denoising: interscale orthonormal wavelet thresholding. *IEEE Trans. Image Process.*, **17** (4) (2008), 482–492.

8 Blu T.; Luisier F.: The SURE-LET approach to image denoising. *IEEE Trans. Image Process.*, **16** (11) (2007), 2778–2786.

9 Zheng N.; Li X.; Blu T.; Lee T.: SURE-MSE speech enhancement for robust speech recognition, in *Proc. 7th Int. Symp. Chinese Spoken Language Process.*, November 2010, 271–274.

10 Donoho D.L.; Johnstone I.M.: Adapting to unknown smoothness via wavelet shrinkage. *J. Amer. Stat. Assoc.*, **90** (432) (1995), 1200–1224.

11 Benazza-Benyahia A.; Pesquet J.C.: Building robust wavelet estimators for multicomponent images using Stein's principle. *IEEE Trans. Image Process.*, **14** (11) (2005), 1814–1830.

12 Zhang X.; Desai M.D.: Adapting denoising based on SURE risk. *IEEE Signal Process. Lett.*, **5** (10) (1998), 265–267.

13 Atto A.M.; Pastor D.; Mercier G.: Smooth adaptation by sigmoid shrinkage. *EURASIP J. Image and Video Process.*, 2009, 1-24, Article ID 532312, doi:10.1155/2009/532312.

14 Atto A.M.; Pastor D.; Mercier G.: Optimal SURE parameter for sigmoidal wavelet shrinkage, in *Proc. Eur. Signal Process. Conf.*, August 2009, 40–43.

15 Ramani S.; Blu T.; Unser M.: Monte Carlo SURE: a black-box optimization of regularization parameters for general denoising algorithms. *IEEE Trans. Image Process.*, **17** (9) (2008), 1540–1544.

16 De Ville D.V.; Kocher M.: Nonlocal means with dimensionality reduction and SURE-based parameter selection. *IEEE Trans. Image Process.*, **20** (9) (2011), 2683–2690.

17 Hsung T.; Lun D.; Ho K.C.: Optimizing the multiwavelet shrinkage denoising. *IEEE Trans. Signal Process.*, **53** (1) (2005), 240–251.

18 Qiu T.; Wang A.; Yu N.; Song A.: LLSURE: local linear SURE-based edge-preserving image filtering. *IEEE Trans. Image Process.*, **22** (1) (2013), 80–90.

19 Wu Y.; Tracey B.H.; Natarajan P.; P.Noonan J.: Fast blockwise SURE shrinkage for image denoising. *Signal Process.*, **103**, (2014), 45–59.

20 Krishnan S.R.; Seelamantula C.S.: On the selection of optimum Savitzky-Golay filters. *IEEE Trans. Signal Process.*, **61** (2) (2013), 380–391.

21 Pesquet J.C.; Benazza-Benyahia A.; Chaux C.: A SURE approach for digital signal/image deconvolution problems. *IEEE Trans. Signal Process.*, **57** (12) (2009), 4616–4632.

22 Zheng L.; Liu Q.; Wanga X.; Maleki A.: ℓ_p-based complex approximate message passing with application to sparse stepped frequency radar. *Signal Process.*, **134**, (2017), 249–260.

23 Hudson H.M.: A natural identity for exponential families with applications in multi-parameter estimation. *Ann. Stat.*, **6** (3) (1978), 473–484.

24 Hwang J.T.: Improving upon standard estimators in discrete exponential families with applications to Poisson and negative binomial cases. *Ann. Stat.*, **10** (3) (1982), 857–867.

25 Giryes R.; Elad M.; Eldar Y.C.: The projected GSURE for automatic parameter tuning in iterative shrinkage methods. *Appl. Comput. Harmon. Anal.*, **30** (3) (2011), 407–422.

26 Luisier F.; Vonesch C.; Blu T.; Unser M.: Fast interscale wavelet denoising of Poisson-corrupted images. *Signal Process.*, **90**, (2010), 415–427.

27 Johnstone I.M.: Gaussian estimation: sequence and wavelet models, June 2013. Available at: http://statweb.stanford.edu/imj/GE06-11-13.pdf. Last Accessed on: Dec. 17, 2019.

28 Sadasivan J.; Mukherjee S.; Seelamantula C.S.: An optimum shrinkage estimator based on minimum-probability-of-error criterion and application to signal denoising, in *Proc. IEEE Intl. Conf. on Acoust. Speech and Signal Process.*, 2014, 4249–4253.

29 Sadasivan J.; Mukherjee S.; Seelamantula C.S.: Speech enhancement using the minimum-probability-of-error criterion, in *Proc. Interspeech*, September 2018, 1141–1145.

30 Plataniotis K.N.; Hatzinakos D.: Gaussian Mixtures and Their Applications to Signal Processing, CRC Press, Canada, 2000.

31 Sorenson H.W.; Alspach D.L.: Recursive Bayesian estimation using Gaussian sums. *Automatica*, **7**, (1971), 465–479.

32 Sovic A.; Sersic D.: Robustly adaptive wavelet filter bank using L1 norm, in *Proc. 18th Intl. Conf. on Systems, Signals and Image Process.*, 2011.

https://doi.org/10.1017/ATSIP.2019.27 Published online by Cambridge University Press

33 WAVELAB toolbox [Online]. Available at: https://statweb.stanford.edu/wavelab/. Last accessed on: Dec. 17, 2019.

34 PhysioBank database [Online]. Available at: https://www.physionet.org/physiobank/database/aami-ec13/. Last accessed on: Dec. 17, 2019.

35 Donoho D.L.: Denoising by soft thresholding. *IEEE Trans. Info. Theory.*, **41** (3) (1995), 613–627.

36 Mitzenmacher M.; Upfal E.: Probability and Computing: Randomized Algorithms and Probabilistic Analysis, Cambridge University Press, UK, 2005.

37 Chang S.; Cosmon P.C.; Milstein L.B.: Chernoff-type bounds for the gaussian error function. *IEEE Trans. Commun.*, **59** (11) (2011), 471–481.

38 Redner R.; Walker H.: Mixture densities, maximum likelihood and the EM algorithm. *SIAM Rev.*, **26** (2) (1984), 195–239.

39 Rao K.R.; Yip P.: Discrete Cosine Transform: Algorithms, Advantages, Applications, Academic, Boston, MA, 1990.

40 Mallat S.: A Wavelet Tour of Signal Processing, 3rd ed., Academic Press, Burlington, 2009.

41 Pastor D.; Socheleau F.: Robust estimation of noise standard deviation in presence of signals with unknown distributions and occurrences. *IEEE Trans. Signal Process.*, **60** (4) (2012), 1545–1555.

42 Boll S.: Suppression of acoustic noise in speech using spectral subtraction. *IEEE Trans. Acoust. Speech, Signal Process.*, **27** (2) (1979), 113–120.

43 Loizou P.: Speech Enhancement – Theory and Practice, CRC Press, Boca Raton, 2007.

44 Kibria B.M.G.; Joarder A.H.: A short review of multivariate *t*-distribution. *J. Stat. Res.*, **40** (1) (2006), 59–72.

45 Johnstone I.M.; Silverman B.W.: Empirical Bayes selection of wavelet thresholds. *Ann. Stat.*, **33** (4) (2005), 1700–1752.

46 Kudryavtsev A.A.; Shestakov O.V.: The asymptotic behavior of the optimal threshold minimizing the probability-of-error criterion. *J. Math. Sci.*, **234** (6) (2018), 810–816.

APPENDIX

A) PERTURBATION OF SURE-BASED POINTWISE SHRINKAGE

To analyze the perturbation in the location of the minimum of the SURE cost, in comparison with the true MSE, one needs to evaluate

$$P_e^{\text{SURE}} = \mathbb{P}\left\{\left|a_{\text{opt}}(s) - a_{\text{opt}}(x)\right| \geq \delta\right\},$$

where $a_{\text{opt}}(s) = s^2/(s^2+\sigma^2)$ and $a_{\text{opt}}(x) = 1 - \sigma^2/x^2$. Let

$$h(x) = a_{\text{opt}}(s) - a_{\text{opt}}(x) = \left(\frac{s^2}{s^2+\sigma^2} - 1 + \frac{\sigma^2}{x^2}\right).$$

The Taylor-series expansion of $h(x)$ about s yields

$$h(x) = \frac{\sigma^4}{s^2(s^2+\sigma^2)} - 2\frac{w\sigma^2}{s^3} + \sum_{n=2}^{\infty} \frac{h^{(n)}(s)}{n!} w^n,$$

where $h^{(n)}$ is the nth derivative h. Using the first-order Taylor series approximation $h(x) \approx h(s) + w\, h^{(1)}(s)$, we obtain

$$h(x) \approx \frac{\sigma^4}{s^2(s^2+\sigma^2)} - 2\frac{w\sigma^2}{s^3},$$

which, in turn, leads to an approximation of the perturbation probability P_e^{SURE}:

$$\begin{aligned} P_e^{\text{SURE}} &= \mathbb{P}\{|h(x)| \geq \delta\} \\ &\approx \mathbb{P}\left\{\left|\frac{\sigma^4}{s^2(s^2+\sigma^2)} - 2\frac{w\sigma^2}{s^3}\right| \geq \delta\right\}. \end{aligned}$$

Invoking $w \sim \mathcal{N}(0, \sigma^2)$, and using the Chernoff bound [36], we obtain

$$P_e^{\text{SURE}} \leq 2\exp\left(-\frac{s^6}{8\sigma^6}\left(\delta - \frac{\sigma^4}{(s^2+\sigma^2)s^2}\right)^2\right).$$

Consequently, to satisfy an upper bound on the deviation probability of the form $P_e^{\text{SURE}} \leq \alpha$, for a given $\delta > 0$, one must ensure that

$$\frac{s^6}{8\sigma^6}\left(\delta - \frac{\sigma^4}{(s^2+\sigma^2)s^2}\right)^2 \geq \log\left(\frac{2}{\alpha}\right). \tag{A.1}$$

The condition in (A.1) translates to an equivalent condition on the minimum required SNR s^2/σ^2 to achieve a certain P_e^{SURE}.

B) EXPECTED ℓ_1 RISK FOR GMM

For additive noise with the p.d.f. given in (8), we have

$$\begin{aligned} \mathcal{E}\{|\widehat{s} - s|\} = \sum_{m=1}^{M} \alpha_m \Bigg(&\int_0^{\infty} Q\left(\frac{\epsilon - (a-1)s - \theta_m}{a\sigma_m}\right) \mathrm{d}\epsilon \\ &+ \int_0^{\infty} Q\left(\frac{\epsilon + (a-1)s + \theta_m}{a\sigma_m}\right) \mathrm{d}\epsilon \Bigg), \end{aligned} \tag{A.2}$$

using (9) and (12). Letting $\mu_m = -((a-1)s + \theta_m)/a\sigma_m$ and $u_m = (\epsilon - (a-1)s - \theta_m)/a\sigma_m$, we get

$$\int_0^{\infty} Q(u_m)\, \mathrm{d}\epsilon = a\sigma_m \left(\frac{e^{-\mu_m^2/2}}{\sqrt{2\pi}} - \mu_m Q(\mu_m)\right). \tag{A.3}$$

Substituting (A.3) in (A.2) yields

$$\begin{aligned} \mathcal{E}\{|\widehat{s} - s|\} = &\sum_{m=1}^{M} a\alpha_m\sigma_m \\ &\times \left(\sqrt{\frac{2}{\pi}} e^{-\mu_m^2/2} - 2\mu_m Q(\mu_m) + \mu_m\right), \end{aligned}$$

which is the expression for the expected ℓ_1 distortion for noise following a GMM distribution.

https://doi.org/10.1017/ATSIP.2019.27 Published online by Cambridge University Press