BLIND OPTIMIZATION OF ALGORITHM PARAMETERS FOR SIGNAL DENOISING BY MONTE-CARLO SURE

Sathish Ramani, Thierry Blu and Michael Unser

Biomedical Imaging Group,
Ecole Polytechnique Fédérale de Lausanne (EPFL),
CH-1015 Lausanne, Switzerland.

ABSTRACT

We consider the problem of optimizing the parameters of an arbitrary denoising algorithm by minimizing *Stein's Unbiased Risk Estimate* (SURE) which provides a means of assessing the true mean-squared-error (MSE) purely from the measured data assuming that it is corrupted by Gaussian noise. To accomplish this, we propose a novel Monte-Carlo technique based on a black-box approach which enables the user to compute SURE for an arbitrary denoising algorithm with some specific parameter setting. Our method only requires the response of the denoising algorithm to additional input noise and does not ask for any information about the functional form of the corresponding denoising operator. This, therefore, permits SURE-based optimization of a wide variety of denoising algorithms (global-iterative, pointwise, etc). We present experimental results to justify our claims.

Index Terms— Stein's unbiased risk estimate, Monte-Carlo estimation, Total variation denoising, and wavelet soft-thresholding.

1. INTRODUCTION

Denoising algorithms aim at eliminating noise from measured data while trying to preserve the important signal features (such as texture and edges) as much as possible. Any denoising algorithm can be interpreted as a mapping $\mathbf{f}_{\boldsymbol{\lambda}}$ (that depends on some parameters $\boldsymbol{\lambda}$) which takes as input the measured data $\mathbf{y}$ to yield the signal estimate $\tilde{\mathbf{x}} = \mathbf{f}_{\boldsymbol{\lambda}}(\mathbf{y})$. When applying a particular algorithm, the user is faced with the difficult task of adjusting $\boldsymbol{\lambda}$ for best performance. To achieve this, researchers usually resort to empirical alternatives or pose the problem in a Bayesian framework. Bayesian and empirical methods are popular in the context of variational denoising where one of the key problems is the selection of the "best" regularization parameter [1–4].

In the context of denoising, the mean-squared-error (MSE) of the signal estimate is obviously the preferred measure of quality to optimize $\boldsymbol{\lambda}$. Unfortunately, the MSE depends on the noise-free signal which is generally unavailable or unknown a priori. A practical approach, therefore, is to replace the true MSE by an estimate in the scheme of things. For Gaussian noise, *Stein's Unbiased Risk Estimate* (SURE) provides a means for unbiased estimation of the true MSE solely using the given data and some description of the first order dependence of the denoising operator with respect to the data; specifically, the divergence of $\mathbf{f}_{\boldsymbol{\lambda}}$ with respect to $\mathbf{y}$ [5]. The computation of this divergence is analytically feasible only in a few special cases such as when the signal estimate is obtained by a linear transformation of the noisy data (linear filtering [2]), a pointwise non-linear mapping or a combination of both (e.g., wavelet thresholding [6–8]). Especially challenging are the cases where the functional form of the denoising operator is not known explicitly and the denoised output is the result of an iterative optimization procedure. Since most variational and Bayesian approaches fall into this latter category, there is a large variety of algorithms for which the analytical evaluation of the required divergence term is not tractable mathematically nor even feasible numerically.

In this paper, we address this limitation by proposing a practicable black-box approach for estimating SURE for a general denoising scenario. Our method is based on Monte-Carlo simulation: the denoising algorithm is probed with additive noise and the response signal is used to estimate the desired divergence—the method completely relies on the output of the operator and does not need any information about its functional form.

In what follows, we briefly review the SURE formulation and then present our Monte-Carlo method together with a practicable algorithm. We then move on to validate the proposed scheme by presenting experimental results for two non-trivial problems: the optimization of the regularization parameter for total variational denoising (TVD) and the adjustment of the threshold for redundant wavelet denoising (RWD) by universal soft-thresholding. In particular, we demonstrate that SURE computed using the new Monte-Carlo strategy closely mimics the behaviour of the true MSE and correctly yields the optimal regularization parameter for the cases considered.

The authors thank the Swiss National Science Foundation (SNSF) for supporting this work under the grant 200020-109415.

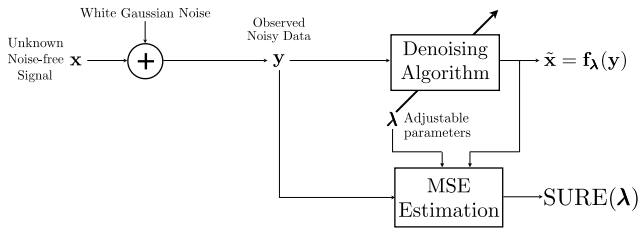

Fig. 1. Schematic of the denoising problem: $\tilde{\mathbf{x}}$ is obtained by the application of the denoising algorithm on the data $\mathbf{y}$. The MSE estimation box then computes an estimate of the MSE of $\tilde{\mathbf{x}}$ (i.e., SURE) as a function of $\boldsymbol{\lambda}$ knowing only $\mathbf{y}$ and $\mathbf{f}_{\boldsymbol{\lambda}}(\mathbf{y})$.

2. PROBLEM FORMULATION AND SURE

We adopt the standard vector formulation of a denoising problem: we measure noisy data $\mathbf{y} \in \mathbb{R}^N$ given by,

$$\mathbf{y} = \mathbf{x} + \mathbf{b}, \tag{1}$$

where $\mathbf{x} \in \mathbb{R}^N$ represents the vector containing the samples of the unknown deterministic noise-free signal and $\mathbf{b} \in \mathbb{R}^N$ denotes the vector containing the zero-mean white Gaussian noise of variance σ^2, respectively. We are given a denoising algorithm which is represented by the operator $\mathbf{f}_{\boldsymbol{\lambda}} : \mathbb{R}^N \rightarrow \mathbb{R}^N$ that maps the input data $\mathbf{y}$ on to the signal estimate $\tilde{\mathbf{x}}$:

$$\tilde{\mathbf{x}} = \mathbf{f}_{\boldsymbol{\lambda}}(\mathbf{y}), \tag{2}$$

where $\boldsymbol{\lambda}$ represents the set of parameters characterizing $\mathbf{f}_{\boldsymbol{\lambda}}$; these should be adjusted to yield the best estimate of the signal [see Figure 1]. Our primal aim in this work is to optimize $\boldsymbol{\lambda}$ knowing only $\mathbf{y}$ and $\tilde{\mathbf{x}} = \mathbf{f}_{\boldsymbol{\lambda}}(\mathbf{y})$ as illustrated by the "MSE estimation" box in Figure 1. To achieve this, we propose the use of SURE as a reliable estimate of the true MSE.

In the sequel, we will assume that $\mathbf{f}_{\boldsymbol{\lambda}}$ is a bounded and continuous operator (i.e., the input-output mapping is continuous and a small perturbation of the input necessarily yields a small perturbation in the output). In particular, we do require that the divergence of $\mathbf{f}_{\boldsymbol{\lambda}}$ with respect to the data $\mathbf{y}$ given by

$$\mathrm{div}_{\mathbf{y}}\{\mathbf{f}_{\boldsymbol{\lambda}}(\mathbf{y})\} = \sum_{k=1}^{N} \frac{\partial \mathrm{f}_{\boldsymbol{\lambda}k}(\mathbf{y})}{\partial y_k}, \tag{3}$$

where $\mathrm{f}_{\boldsymbol{\lambda}k}(\mathbf{y})$ and y_k are the k^{th} component of the vectors $\mathbf{f}_{\boldsymbol{\lambda}}(\mathbf{y})$ and $\mathbf{y}$ respectively, is well-defined in the weak sense. Then the SURE corresponding to $\tilde{\mathbf{x}} = \mathbf{f}_{\boldsymbol{\lambda}}(\mathbf{y})$ is a random variable given by

$$\eta(\mathbf{f}_{\boldsymbol{\lambda}}(\mathbf{y})) = \frac{1}{N}\|\mathbf{y} - \mathbf{f}_{\boldsymbol{\lambda}}(\mathbf{y})\|^2 - \sigma^2 + \frac{2\sigma^2}{N}\,\mathrm{div}_{\mathbf{y}}\{\mathbf{f}_{\boldsymbol{\lambda}}(\mathbf{y})\}, \tag{4}$$

where $\|\cdot\|^2$ represents the Euclidean norm. The following theorem, due to Stein, states that η is an unbiased estimate of the true MSE given by

$$\mathrm{MSE}(\mathbf{f}_{\boldsymbol{\lambda}}(\mathbf{y})) = \frac{1}{N}\|\mathbf{x} - \mathbf{f}_{\boldsymbol{\lambda}}(\mathbf{y})\|^2. \tag{5}$$

Theorem 1 *(cf. [5]) The random variable $\eta(\mathbf{f}_{\boldsymbol{\lambda}}(\mathbf{y}))$ is an unbiased estimator of MSE$(\mathbf{f}_{\boldsymbol{\lambda}}(\mathbf{y}))$, that is,*

$$E_{\mathbf{b}}\{MSE(\mathbf{f}_{\boldsymbol{\lambda}}(\mathbf{y}))\} = E_{\mathbf{b}}\{\eta(\mathbf{f}_{\boldsymbol{\lambda}}(\mathbf{y}))\}, \tag{6}$$

where $E_{\mathbf{b}}\{\cdot\}$ represents the expectation with respect to $\mathbf{b}$.

3. MONTE-CARLO SURE

As noted in (4), the divergence term, $\mathrm{div}_{\mathbf{y}}\{\mathbf{f}_{\boldsymbol{\lambda}}(\mathbf{y})\}$, plays a pivotal role in the computation of SURE. The divergence can be calculated analytically and has a closed form expression only in some special cases such as when $\mathbf{f}_{\boldsymbol{\lambda}}$ is linear or when $\mathbf{f}_{\boldsymbol{\lambda}}$ is a pointwise operator in an orthogonal transform domain [6–8]. For a general $\mathbf{f}_{\boldsymbol{\lambda}}$, the evaluation of the divergence may not be tractable analytically and worse, it may even be numerically infeasible, especially if $\mathbf{f}_{\boldsymbol{\lambda}}$ is implemented in an iterative fashion (as is the case with most variational or PDE-based denoising methods). We circumvent this difficulty by proposing a novel technique that is based on the following theorem which allows us to estimate the required divergence (and thus SURE) for an arbitrary $\mathbf{f}_{\boldsymbol{\lambda}}$.

Theorem 2 *Let $\mathbf{f}_{\boldsymbol{\lambda}}(\mathbf{z})$ be the output of $\mathbf{f}_{\boldsymbol{\lambda}}$ corresponding to $\mathbf{z} = \mathbf{y} + \mathbf{b}'$, where $\mathbf{b}'$ is a zero-mean i.i.d random vector (that is independent of $\mathbf{y}$) with covariance $\varepsilon^2\mathbf{I}$. Then*

$$\mathrm{div}_{\mathbf{y}}\{\mathbf{f}_{\boldsymbol{\lambda}}(\mathbf{y})\} = \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon^2} E_{\mathbf{b}'}\{\mathbf{b}'^{\mathrm{T}}(\mathbf{f}_{\boldsymbol{\lambda}}(\mathbf{z}) - \mathbf{f}_{\boldsymbol{\lambda}}(\mathbf{y}))\}, \tag{7}$$

provided that $\mathbf{f}_{\boldsymbol{\lambda}}$ admits a well-defined second order Taylor expansion.

The proof of this theorem will be presented elsewhere. This is a powerful result since (7) does not require any knowledge of the functional form of $\mathbf{f}_{\boldsymbol{\lambda}}$, thus making it applicable for a wide variety of algorithms. The important point is that $\mathbf{f}_{\boldsymbol{\lambda}}$ is treated as a black-box, meaning that we only need the output of the operator irrespective of how it is implemented. Equation (7) forms the basis of our Monte-Carlo approach for computing SURE for a general $\mathbf{f}_{\boldsymbol{\lambda}}$. Since, in practice, the limit in equation (7) cannot be implemented due to finite machine precision, we propose to use the following approximation:

$$\mathrm{div}_{\mathbf{y}}\{\mathbf{f}_{\boldsymbol{\lambda}}(\mathbf{y})\} \approx \frac{1}{\varepsilon^2}\mathbf{b}'^{\mathrm{T}}(\mathbf{f}_{\boldsymbol{\lambda}}(\mathbf{y} + \mathbf{b}') - \mathbf{f}_{\boldsymbol{\lambda}}(\mathbf{y})). \tag{8}$$

The idea is to add a small amount of noise (of variance ε^2) to $\mathbf{y}$ and evaluate $\mathbf{f}_{\boldsymbol{\lambda}}(\mathbf{y} + \mathbf{b}')$. The difference $\mathbf{f}_{\boldsymbol{\lambda}}(\mathbf{y} + \mathbf{b}') - \mathbf{f}_{\boldsymbol{\lambda}}(\mathbf{y})$ is then used according to (8) to obtain an estimate of the divergence. The schematics of implementing the rhs of (8) is illustrated in Figure 2.

We will demonstrate numerically that the approximation in (8) is quite reasonable and yields excellent numerical results. The validity of the approximation in (8) depends on how small ε can be made. In practice, we must select ε small enough to mimic the limit, but still large enough so as to avoid

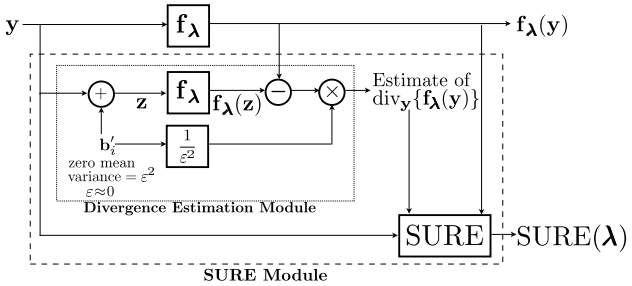

Fig. 2. The dotted box depicts the module which estimates $\mathrm{div}_{\mathbf{y}}\{\mathbf{f}_{\boldsymbol{\lambda}}(\mathbf{y})\}$ according to (8) using the realization $\mathbf{b}'_i$. The dashed box represents the SURE module (depicted as the MSE estimation box in Figure 1) which computes the SURE according to equation (4) for a given $\boldsymbol{\lambda}$.

numerical round-off errors in $\mathbf{f}_{\boldsymbol{\lambda}}(\mathbf{y}+\mathbf{b}')$. It turns out that the estimation procedure is remarkably robust with respect to ε (cf. Section 4).

We now develop an algorithm based on (8) for estimating $\mathrm{div}_{\mathbf{y}}\{\mathbf{f}_{\boldsymbol{\lambda}}(\mathbf{y})\}$ for an arbitrary $\mathbf{f}_{\boldsymbol{\lambda}}$. The algorithm assumes that a suitable ε has been selected and that a set of K independent random vectors $\{\mathbf{b}'_i\}_{i=1}^{K}$ where each $\mathbf{b}'_i$ (independent of $\mathbf{y}$) is zero-mean i.i.d with variance ε^2 has been generated.

Algorithm 1 *Estimation of* $\mathrm{div}_{\mathbf{y}}\{\mathbf{f}_{\boldsymbol{\lambda}}(\mathbf{y})\}$ *and computation of SURE for a given* $\boldsymbol{\lambda}=\boldsymbol{\lambda}_0$ *and fixed* ε*:*

Step 1: For $\boldsymbol{\lambda}=\boldsymbol{\lambda}_0$, evaluate $\mathbf{f}_{\boldsymbol{\lambda}}(\mathbf{y})$; $i=1$; $\mathrm{div}=0$
Step 2: Build $\mathbf{z}=\mathbf{y}+\mathbf{b}'_i$; Evaluate $\mathbf{f}_{\boldsymbol{\lambda}}(\mathbf{z})$ for $\boldsymbol{\lambda}=\boldsymbol{\lambda}_0$
Step 3: $\mathrm{div}=\mathrm{div}+\frac{1}{\varepsilon^2}\,\mathbf{b}_i'^{\mathrm{T}}\,(\mathbf{f}_{\boldsymbol{\lambda}}(\mathbf{z})-\mathbf{f}_{\boldsymbol{\lambda}}(\mathbf{y}))$; $i=i+1$
Step 4: If ($i\leq K$) go to Step 2; otherwise evaluate sample mean: $\mathrm{div}=\mathrm{div}/K$ and compute SURE($\boldsymbol{\lambda}_0$) using (4).

It is clear that our method needs only $O(N)$ storage while computationally it is K times as costly as the denoising algorithm itself. It should also be noted that to estimate $\mathrm{div}_{\mathbf{y}}\{\mathbf{f}_{\boldsymbol{\lambda}}(\mathbf{y})\}$ for a given set of parameters $\boldsymbol{\lambda}=\boldsymbol{\lambda}_0$, $\mathbf{f}_{\boldsymbol{\lambda}}(\mathbf{y})$ needs to be evaluated only once, while $\mathbf{f}_{\boldsymbol{\lambda}}(\mathbf{z})$ may be repeated with many realizations $\mathbf{b}'_i|_{i=1}^{K}$ for computing the sample mean in step 4. However, in practice, when N is large (especially images), it is usually sufficient to use a single realization (i.e., $K=1$) of $\mathbf{b}'$.

Let us now consider the special case of a linear denoising operator whose generic form is

$$\mathbf{f}_{\boldsymbol{\lambda}}(\mathbf{y})=\mathbf{F}_{\boldsymbol{\lambda}}\mathbf{y}, \tag{9}$$

where $\mathbf{F}_{\boldsymbol{\lambda}}$ is the matrix corresponding to the linear transformation. Then, the desired divergence is simply given by

$$\mathrm{div}_{\mathbf{y}}\{\mathbf{f}_{\boldsymbol{\lambda}}(\mathbf{y})\}=\mathrm{Trace}\{\mathbf{F}_{\boldsymbol{\lambda}}\}, \tag{10}$$

which may be explicitly evaluated if $\mathbf{F}_{\boldsymbol{\lambda}}$ is known. There are many scenarios, however, where the matrix is not known explicitly and where (9) is evaluated through a recursive process, in which case the trace computation can turn out to be difficult (especially for large data sets). For such cases, we can prove that Algorithm 1 yields an unbiased estimate of $\mathrm{Trace}\{\mathbf{F}_{\boldsymbol{\lambda}}\}$ irrespective of the value of ε. This provides another strong justification for dropping the limit in (7).

4. EXPERIMENTS

We illustrate the applicability of our Monte-Carlo method for two popular denoising algorithms: total variation denoising (TVD) [9] and redundant wavelet denoising (RWD) by universal soft-thresholding with the Haar transform. In both cases, the SURE computation is known to be non-trivial and most probably not feasible numerically. For our experiments, we used $K=1$ in Algorithm 1 with Gaussian random vectors. Our observation was that any $\varepsilon\in[10^{-12}\,,1]$ yielded agreeable results for both TVD and RWD. So we chose to be conservative with respect to round-off errors and selected $\varepsilon=0.1$ in all experiments. The performance of the methods was quantified by the signal-to-noise ratio (SNR) of the output $\mathbf{f}_{\boldsymbol{\lambda}}(\mathbf{y})$ computed as $\mathrm{SNR}=10\log_{10}\left(\frac{\|\mathbf{x}\|^2}{\|\mathbf{x}-\mathbf{f}_{\boldsymbol{\lambda}}(\mathbf{y})\|^2}\right)$. In all cases, the value of σ was set to achieve the desired input SNR computed by replacing $\|\mathbf{x}-\mathbf{f}_{\boldsymbol{\lambda}}(\mathbf{y})\|^2$ with $N\sigma^2$. The noise variance σ^2 was assumed to be known (it can be reliably estimated from $\mathbf{y}$ using the median estimator in [6]) for computing SURE in (4).

Figures 3 and 4 plot the true MSE and SURE as a function of λ for both TVD and RWD (for the Boats image, input SNR = 4 dB). In the case of TVD, the parameter λ represents the regularization parameter, while it denotes the soft-threshold value for RWD. It is clearly seen that SURE computed using our Monte-Carlo method approximates the true MSE curve remarkably well over the entire range of λ in both cases. Also significant is the fact that SURE yields correct values for the optimal λ in all cases. We observed the same trend for all test images and input SNRs which confirms the consistency of our method.

Some further denoising results are summarized in Table 1. The first value in each cell gives the SNR obtained by choosing λ based on the true MSE (oracle SNR), while the second corresponds to the result obtained by Monte-Carlo SURE optimization. Again, the two SNR values are in near perfect agreement for all the test images and noise levels. This confirms the validity of our choice of ε and K; it also demonstrates the reliability and robustness of our Monte-Carlo SURE optimization procedure.

5. CONCLUSIONS

We have developed a novel technique for computing SURE for an arbitrary denoising algorithm. A possible interpretation of our Monte-Carlo scheme is that of a random first order difference estimator of the divergence of an operator: it boils down to a randomly weighted summation of the difference between the restored signal and a perturbed version of it. In effect, this yields a black-box approach that uses only the

Table 1. Comparison of SNR obtained based on the true MSE and SURE

Images	Input SNR (dB)	4	8	12	16	20
Boats	TVD	(11.02, 11.01)	(13.12, 13.12)	(15.62, 15.62)	(18.38, 18.38)	(21.43, 21.43)
(512×512)	RWD	(11.90, 11.90)	(14.06, 14.06)	(16.49, 16.49)	(19.09, 19.09)	(21.92, 21.92)
Barbara	TVD	(9.44, 9.44)	(11.66, 11.66)	(14.48, 14.48)	(17.71, 17.71)	(21.16, 21.16)
(512×512)	RWD	(10.55, 10.55)	(12.87, 12.87)	(15.58, 15.58)	(18.61, 18.61)	(21.89, 21.89)
Peppers	TVD	(11.18, 11.18)	(13.70, 13.70)	(16.36, 16.36)	(19.18, 19.18)	(22.18, 22.18)
(256×256)	RWD	(12.03, 12.03)	(14.59, 14.59)	(17.26, 17.26)	(20.04, 20.04)	(22.88, 22.88)
Shepp-Logan	TVD	(15.21, 15.21)	(18.84, 18.82)	(22.71, 22.71)	(26.30, 26.30)	(30.14, 30.12)
(256×256)	RWD	(13.92, 13.92)	(17.51, 17.51)	(21.33, 21.33)	(24.96, 24.96)	(28.82, 28.82)

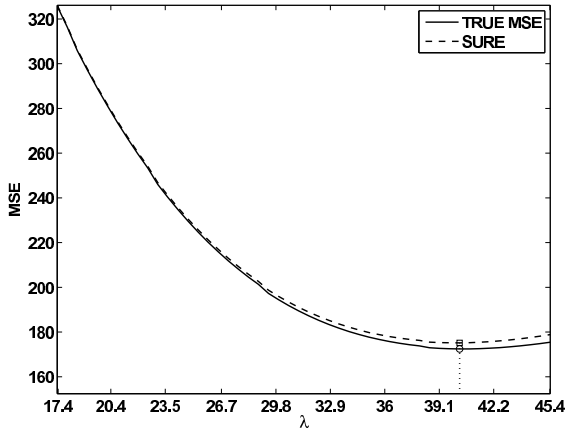

Fig. 3. True MSE and SURE as a function of λ for TVD.

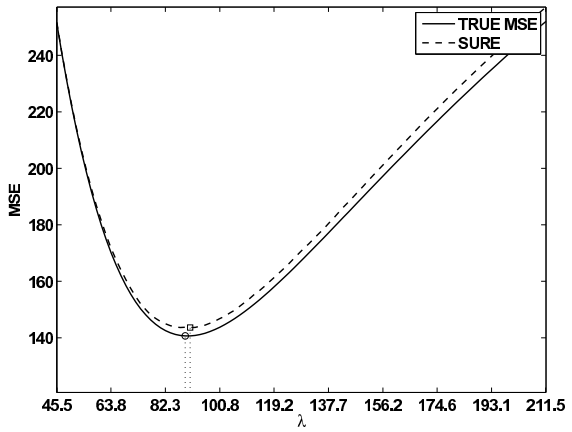

Fig. 4. True MSE and SURE as a function of λ for RWD.

output of the denoising algorithm and does not require any knowledge of its internal working. We did illustrate and validate the method by optimization of the parameters of some popular denoising algorithms. We found that SURE computed using our method perfectly predicts the true MSE in all the cases tested. Moreover, the SNR obtained by SURE-based optimization is in almost perfect agreement with the oracle solution (minimum MSE). This suggests that Monte-Carlo SURE can be reliably employed for data-driven adjustment of parameters in a large variety of denoising problems provided that the data is corrupted by Gaussian noise.

6. REFERENCES

[1] R. Molina, A. K. Katsaggelos, and J. Mateos, "Bayesian and regularization methods for hyperparameter estimation in image restoration," *IEEE Trans. Image Process.*, vol. 8, no. 2, pp. 231–246, 1999.

[2] N. P. Galatsanos and A. K. Katsaggelos, "Methods for choosing the regularization parameter and estimating the noise variance in image restoration and their relation," *IEEE Trans. Image Process.*, vol. 1, no. 3, pp. 322–336, 1992.

[3] W. C. Karl, "Regularization in image restoration and reconstruction," in *Handbook of Image & Video Processing*, A. Bovik, Ed., pp. 183–202. ELSEVIER, 2nd edition, 2005.

[4] G. Gilboa, N. Sochen, and Y. Y. Zeevi, "Estimation of optimal PDE-based denoising in the SNR sense," *IEEE Trans. Image Process.*, vol. 15, no. 8, pp. 2269–2280, 2006.

[5] C. Stein, "Estimation of the mean of a multivariate normal distribution," *Ann. Statist.*, vol. 9, pp. 1135–1151, 1981.

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

[7] X. -P. Zhang and M. D. Desai, "Adaptive denoising based on SURE risk," *IEEE Signal Process. Lett.*, vol. 5, no. 10, pp. 265–267, 1998.

[8] 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, 2007.

[9] M. A. T. Figueiredo, J. B. Dias, J. P. Oliveira, and R. D. Nowak, "On total variation denoising: A new Majorization-Minimization algorithm and an experimental comparison with wavalet denoising," *Proceedings of IEEE International Conference on Image Processing (ICIP 2006), Atlanta, GA, USA*, pp. 2633–2636, October 2006.