Bayesian Denoising: From MAP to MMSE Using Consistent Cycle Spinning

Abbas Kazerouni, Ulugbek S. Kamilov, *Student Member, IEEE*, Emrah Bostan, *Student Member, IEEE*, and Michael Unser, *Fellow, IEEE*

***Abstract*—We introduce a new approach for the implementation of minimum mean-square error (MMSE) denoising for signals with decoupled derivatives. Our method casts the problem as a penalized least-squares regression in the redundant wavelet domain. It exploits the link between the discrete gradient and Haar-wavelet shrinkage with cycle spinning. The redundancy of the representation implies that some wavelet-domain estimates are inconsistent with the underlying signal model. However, by imposing additional constraints, our method finds wavelet-domain solutions that are mutually consistent. We confirm the MMSE performance of our method through statistical estimation of Lévy processes that have sparse derivatives.**

***Index Terms*—Augmented Lagrangian, MMSE estimation, total variation denoising, wavelet denoising.**

I. Introduction

CONSIDER the problem of estimating an unknown signal $\mathbf{x} \in \mathbb{R}^n$ from noisy measurements $\mathbf{y} = \mathbf{x} + \mathbf{n}$, where the vector $\mathbf{n}$ represents additive white Gaussian noise (AWGN). We would like to design an estimator $\mathbf{f} : \mathbb{R}^n \mapsto \mathbb{R}^n$ that maps the data $\mathbf{y}$ to the estimate $\widehat{\mathbf{x}} = \mathbf{f}(\mathbf{y})$. A standard methodology for estimation is to solve the regularized least-squares regression

$$\widehat{\mathbf{x}} = \mathbf{f}(\mathbf{y}) = \underset{\mathbf{x}\in\mathbb{R}^n}{\arg\min} \left\{ \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|_2^2 + \Phi(\mathbf{x}) \right\}, \qquad (1)$$

where the function Φ is a *regularizer* that promotes solutions with desirable properties. The choice of a suitable regularizer Φ is a crucial part for designing a good denoising method and the challenge is often to find the simplest Φ that results in highest-quality estimates.

One of the most popular regularizers is total variation (TV) [1], which penalizes the ℓ_1-norm of the discrete gradient of the signal. TV was introduced in the context of denoising [1], but has since been successfully extended to solve more-general inverse problems [2], [3]. From a Bayesian perspective, TV denoising can be interpreted as a maximum-a-posteriori (MAP) estimator assuming a Laplace prior on the signal gradient. Although TV regularization works well for general signals, there is potentially a large performance gap between the TV regularization and the minimum mean-squared error (MMSE) estimator with the correct prior [4].

The difficulty is often in finding tractable methods for performing MMSE estimation. Louchet and Moisan [5] proposed to use Monte Carlo Markov Chain (MCMC) Metropolis sampler to compute the posterior expectation based on the Bayesian interpretation of TV. They have shown that their algorithm, which in fact evaluates the MMSE counterpart of TV denoising, improves upon the standard TV approach. Recently, Gribonval [6] showed that it is possible to recast the MMSE estimation problem for AWGN into a MAP-like form that is compatible with (1).

In this letter, we take the Bayesian perspective and propose a new approach for efficiently computing MMSE-TV denoising. It consists in an adaptation of our recent *consistent cycle spinning* (CCS) algorithm [7], which was successfully used to compute MAP estimators for sparse stochastic signals. The algorithm is based on cycle spinning [8] and exploits the link between the discrete gradient and Haar-wavelet shrinkage. In essence, we extend the original CCS algorithm by replacing MAP shrinkage functions with their MMSE counterparts.

Cycle spinning implies a redundant representation. Thus, not every set of wavelet-domain coefficients can be perfectly inverted back to the signal domain. A problem arises whenever the estimated coefficients violate the invertibility condition. It can be resolved through *consistency*, which refers to a technique whereby the solutions obtained by CCS are made perfectly invertible. This is achieved by restricting the wavelet-domain solution to the space spanned by the basis function of the transform, which results in a norm equivalence between signal-domain and wavelet-domain estimations. A key advantage of CCS is its computational simplicity, which reduces the difficult estimation problem to the iterative application of a scalar shrinkage function in the shift-invariant Haar-wavelet domain. To keep notations simple, we concentrate on denoising 1D signals; however, the method can be extended to 2D images, as was done in [7].

For the objective evaluation of the algorithm, we perform estimation of *Lévy processes* [9], [10] that have independent and stationary increments. Recently, we have developed a message-passing algorithm for computing the MMSE estimator for Lévy processes [4]. We show that CCS performs as well as the message-passing algorithm [4], while running substantially faster.

II. Background on Signal Denoising

A. Wavelet-Based Denoising

The estimation in wavelet-based denoising algorithms is performed in three steps.

Manuscript received October 25, 2012; revised January 14, 2013; accepted January 17, 2013. Date of publication January 23, 2013; date of current version February 04, 2013. This work was supported by the European Commission under Grant ERC-2010-AdG 267439-FUN-SP. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Anna Scaglione.

The authors are with the Biomedical Imaging Group, École polytechnique fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland (e-mail: abbas.kazerouni@epfl.ch; ulugbek.kamilov@epfl.ch; emrah.bostan@epfl.ch; michael.unser@epfl.ch).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2013.2242061

1070-9908/$31.00 © 2013 IEEE

1) Perform the wavelet transform of the data $\mathbf{u} = \mathbf{W}\mathbf{y} \in \mathbb{R}^m$, where, in general, $m \geq n$. The equality $m = n$ happens when the transform is orthogonal.
2) Apply a denoising algorithm represented by $\mathbf{f}_{\mathbf{W}} : \mathbb{R}^m \rightarrow \mathbb{R}^m$ that maps the data into the estimate $\widehat{\mathbf{w}} = \mathbf{f}_{\mathbf{W}}(\mathbf{u})$.
3) Return to the image domain by applying the inverse transform to obtain $\widehat{\mathbf{x}} = \mathbf{W}^T\widehat{\mathbf{w}} \in \mathbb{R}^n$.

In the sequel we restrict our attention to *tight* wavelet frames, which implies that the transform preserves inner-products and is self-reversible

$$\forall \mathbf{x} \in \mathbb{R}^n, \qquad \mathbf{W}^T\mathbf{W}\mathbf{x} = \mathbf{x}. \tag{2}$$

This also implies that the transform is norm-preserving, with $\|\mathbf{x}\|_2^2 = \|\mathbf{W}\mathbf{x}\|_2^2$. Traditionally, the estimation problem in the wavelet domain is expressed as

$$\widehat{\mathbf{w}} = \mathbf{f}_{\mathbf{W}}(\mathbf{u}) = \underset{\mathbf{w}\in\mathbb{R}^m}{\arg\min} \left\{ \frac{1}{2}\|\mathbf{w} - \mathbf{u}\|_2^2 + \Phi_{\mathbf{W}}(\mathbf{w}) \right\}. \tag{3}$$

The least-squares term in (3) is justified by the fact that the transform $\mathbf{W}$ is norm preserving. Nevertheless, we show in the sequel that (3) is suboptimal when $\mathbf{W}$ is redundant.

The powerful aspect of wavelet decomposition is that it has a decorrelating effect on many natural signals. Thus, it is very common to select the regularizer $\Phi_{\mathbf{W}}$ to be separable. For example, the popular wavelet soft-thresholding algorithm [11] is obtained by using the non-smooth convex function $\Phi_{\mathbf{W}}(\cdot) = \|\cdot\|_1$ that favors sparse wavelet-domain solutions and admits a closed form solution $\widehat{\mathbf{w}} = \eta(\mathbf{u}; \lambda)$. The soft-thresholding function η is applied component-wise on the detail wavelet coefficients.

From a Bayesian MAP perspective, the soft-thresholding algorithm inherently assumes the detail wavelet coefficients to be i.i.d. Laplace random variables. Alternatively, one could assign other i.i.d. priors p_w to the wavelet coefficients and identify a corresponding MAP proximity operator η_{MAP} [12]. Better yet, one can choose the pointwise MMSE estimator for specific priors [13], [14]. A convenient way of determining the optimal shrinkage function in the AWGN scenario is to rely on Stein's formula [6], [15]

$$\begin{aligned}\widehat{w}_i = \eta_{\mathrm{MMSE}}(u_i) &= \mathbb{E}\left[w|u_i\right] \\ &= u_i + \sigma_i^2 \frac{d}{du} \log p_u(u_i),\end{aligned} \tag{4}$$

where σ_i^2 is the variance of the noise of the ith wavelet-domain coefficient and $p_u = p_w * g_{\sigma_i}$ is the probability distribution of the noisy measurement with the noise distribution

$$g_\sigma(w) = \frac{1}{\sigma\sqrt{2\pi}} e^{-w^2/2\sigma^2}. \tag{5}$$

We illustrate in Fig. 1 the shrinkage functions η_{MAP} and η_{MMSE} for Laplace-distributed coefficients in AWGN of unit variance.

A major drawback of the wavelet-based approach is that MMSE estimation in the wavelet domain is not in general equivalent to MMSE estimation in the signal domain. The exception happens *only* when the wavelet transform is orthogonal. However, empirical results reported over the years show that wavelet-based algorithms using overcomplete transforms are more effective than those using orthogonal transforms [8].

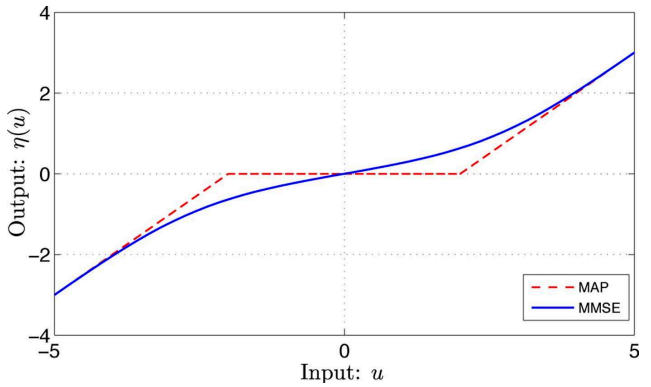

Fig. 1. Comparison of η_{MAP} (dashed) and η_{MMSE} (solid) shrinkage functions for a Laplace prior with parameter $b = 1/2$ and AWGN of variance $\sigma^2 = 1$.

B. Total-Variation Regularization

TV denoising operates directly in the signal domain by solving the optimization problem (1) with the regularizer $\Phi(\mathbf{x}) = \lambda\|D(\mathbf{x})\|_1$, where $\lambda > 0$ is the regularization parameter and $D(\mathbf{x}) \in \mathbb{R}^n$ is a vector that contains the discrete gradient of $\mathbf{x}$ at all positions, defined component-wise as

$$[D(\mathbf{x})]_i = x_{i+1} - x_i, \qquad \forall i \in [1 \ldots n-1], \tag{6}$$

where we assume that $x_0 = 0$.

III. Consistent Cycle Spinning

In this section, we present the CCS algorithm, which is in some sense an extension to both conventional wavelet denoising and TV regularization with regularized least-squares data terms. In particular, we are interested in adapting CCS to use the MMSE shrinkage in (4).

A. Formulation of the Problem

As mentioned in Section II-A, the fact that the frame is tight implies that (2) always holds. However, since $m > n$, there are some $\mathbf{w} \in \mathbb{R}^m$ such that the converse equality $\mathbf{W}\mathbf{W}^T\mathbf{w} = \mathbf{w}$, does *not* hold. This in turn implies that the solution obtained by solving (3) in the tight wavelet-frame domain is not *consistent*, in the sense that the wavelet transform of the final solution $\widehat{\mathbf{x}} = \mathbf{W}^T\widehat{\mathbf{w}}$ is not necessarily equal to $\widehat{\mathbf{w}}$ obtained from (3). However, since the desired solution always lives in the signal domain, it makes sense to constraint the wavelet-domain estimate to the consistent subspace, which also makes it perfectly invertible. Such estimation can be performed by solving the following constrained optimization problem

$$\widehat{\mathbf{w}} = \underset{\mathbf{w}\in\mathbb{R}^m}{\arg\min}\, \mathcal{J}(\mathbf{u}, \mathbf{w}) \quad \text{s.t} \quad \mathbf{W}\mathbf{W}^T\mathbf{w} = \mathbf{w}, \tag{7a}$$

with

$$\mathcal{J}(\mathbf{u}, \mathbf{w}) = \frac{1}{2}\|\mathbf{w} - \mathbf{u}\|_2^2 + \Phi_{\mathbf{W}}(\mathbf{w}). \tag{7b}$$

The estimation problem (7) combines better approximation capabilities of redundant representations with wavelet-domain solutions that behave as if the transform were truly orthogonal.

Reconstruction with undecimated wavelet transforms is commonly referred to as cycle spinning [8]. If we consider an undecimated single-level Haar expansion, which involves sums and differences [16], then it is clear that each element of the

discrete gradient can be uniquely mapped to the detail coefficients of $\mathbf{w} = \mathbf{W}\mathbf{x}$ by replacing each finite difference by some corresponding Haar coefficient. Let $\mathbf{w}^a$ and $\mathbf{w}^d$ denote the approximation and detail coefficients in $\mathbf{w}$, respectively. Then, for the case of $\Phi_{\mathbf{W}}(\mathbf{w}) = \lambda\sqrt{2}\|\mathbf{w}^d\|_1$, the optimization in (7) exactly performs TV denoising [7]. The optimization also admits a simple interpretation as estimation of a sparse and consistent wavelet-domain solution.

B. Iterative Consistent MMSE Shrinkage

We now consider the MMSE version of TV denoising where we set the regularizer to

$$\Phi_{\mathbf{W}}(\mathbf{w}) = \sum_{i=1}^{n} \phi_{\mathrm{MMSE}}(w_i^d) \tag{8}$$

and where the potential function ϕ_{MMSE} corresponds to MMSE shrinkage (4). Observe that, when $\phi_{\mathrm{MMSE}}(\cdot) = |\cdot|$, the problem reduces to TV denoising. In general, ϕ_{MMSE} does not necessarily admit an analytic formula. However, one can formally characterize it as [6]

$$\phi_{\mathrm{MMSE}}(w) = -\frac{1}{2}(\eta_{\mathrm{MMSE}}^{-1}(w) - w)^2 - \log p_u(\eta_{\mathrm{MMSE}}^{-1}(w)), \tag{9}$$

for all w in the image of η_{MMSE}, where $\eta_{\mathrm{MMSE}}^{-1}$ denotes the inverse of the MMSE shrinkage.

A practical optimization scheme for the constrained minimization problem (7) with potential function ϕ_{MMSE} can be obtained by using augmented-Lagrangian approach [17], which casts the constrained problem as a sequence of unconstrained problems. The idea is to replace the objective function $\mathcal{J}$ with the new penalty function

$$\mathcal{L}(\mathbf{w}, \mathbf{x}) = \mathcal{J}(\mathbf{u}, \mathbf{w}) + \frac{\tau}{2}\|\mathbf{w} - \mathbf{W}\mathbf{x}\|_2^2 - \boldsymbol{\mu}^T(\mathbf{w} - \mathbf{W}\mathbf{x}), \tag{10}$$

where $\tau > 0$ is the penalty parameter and $\boldsymbol{\mu} \in \mathbb{R}^m$ is the vector of Lagrange multipliers. The condition $\mathbf{w} = \mathbf{W}\mathbf{x}$ asserted by the penalty function constraints $\mathbf{w}$ to the column space of $\mathbf{W}$, which is equivalent to the consistency condition $\mathbf{w} = \mathbf{W}\mathbf{W}^T\mathbf{w}$. Although the quadratic penalty term in $\mathcal{J}$ does not influence the final solution, it typically improves the convergence behavior of the iterative optimization [17]. To solve the minimization of the objective (10), we alternate between solving the problem for $\mathbf{w}$ with $\mathbf{x}$ fixed and *vice versa*.

Given the noisy wavelet-domain measurements $\mathbf{u} \in \mathbb{R}^m$, penalty parameter $\tau > 0$, regularizer $\Phi_{\mathbf{W}}$ in (8), and initial signal-domain solution $\widehat{\mathbf{x}}^0$, CCS estimation proceeds as:

1) *Initialize*: Set $t = 0$ and $\boldsymbol{\mu}^0 = \mathbf{0}$.
2) *Update* $\widehat{\mathbf{w}}$: Minimize $\mathcal{L}$ with respect to $\mathbf{w}$ with $\mathbf{x}$ fixed

$$\begin{aligned}\widehat{\mathbf{w}}^{t+1} &= \underset{\mathbf{w}\in\mathbb{R}^m}{\arg\,min}\, \mathcal{L}(\mathbf{w}, \widehat{\mathbf{x}}^t) \\ &= \underset{\mathbf{w}\in\mathbb{R}^m}{\arg\,min} \left\{ \frac{1}{2}\|\mathbf{w} - \tilde{\mathbf{u}}\|_2^2 + \tilde{\lambda}\Phi_{\mathbf{W}}(\mathbf{w}) \right\} \\ &= \eta_{\mathrm{MMSE}}(\tilde{\mathbf{u}}),\end{aligned} \tag{11}$$

where $\tilde{\mathbf{u}} = (\mathbf{u} + \tau\mathbf{W}\widehat{\mathbf{x}}^t + \boldsymbol{\mu}^t)/(1+\tau)$ and $\tilde{\lambda} = 1/(1+\tau)$. The function η_{MMSE} is applied componentwise on the detail coefficients of $\widehat{\mathbf{w}}$. It is a 1D shrinkage operator and can be implemented as a lookup table.

3) *Update* $\widehat{\mathbf{x}}$: Minimize $\mathcal{L}$ with respect to $\mathbf{x}$ with $\mathbf{w}$ fixed

$$\begin{aligned}\widehat{\mathbf{x}}^{t+1} &= \underset{\mathbf{x}\in\mathbb{R}^n}{\arg\,min}\, \mathcal{L}(\widehat{\mathbf{w}}^{t+1}, \mathbf{x}) \\ &= \mathbf{W}^T\left(\widehat{\mathbf{w}}^{t+1} - \frac{\boldsymbol{\mu}^t}{\tau}\right).\end{aligned} \tag{12}$$

4) *Update* $\boldsymbol{\mu}$: Lagrange multipliers are updated according to the simple rule

$$\boldsymbol{\mu}^{t+1} = \boldsymbol{\mu}^t - \tau(\widehat{\mathbf{w}}^{t+1} - \mathbf{W}\widehat{\mathbf{x}}^{t+1}). \tag{13}$$

5) Set $t = t + 1$ and proceed to Step 2.

For each iteration $t = 1, 2, \ldots$, the proposed update rules produce estimates $\widehat{\mathbf{x}}^t$ of the true signal $\mathbf{x}$. The computational complexity of the algorithm is $O(n)$ per iteration, since it reduces to the evaluation of wavelet transforms and pointwise non-linearities.

We conclude this section with some remarks regarding the convergence of the algorithm. In general, its analysis is complicated by the fact that the potential function ϕ_{MMSE} is not necessarily convex. However, the MMSE cost function has a favorable property that its unique global minimum is also its unique stationary point [6]. This suggests that the optimization will not get trapped in a local minimum, which is corroborated by the experimental results in Section IV.

IV. Experimental Results

In this section, we experimentally validate our iterative algorithm by comparing its performance against the MMSE estimator that was recently proposed for Lévy processes [4]. The defining property of a Lévy process, which can be viewed as a continuous-time analog of a random walk, is the independence and stationarity of its increments [9], [10]. This property implies that the applications of finite-difference operator (6) on samples of a Lévy process decouples it into a sequence of independent random variables. The algorithm in [4] is based on the *belief propagation* method, which efficiently computes the marginals of the joint posterior distribution by performing local computations on a tree [18].

We consider the denoising of three types of Lévy signals of length $n = 2048$, with three increment distributions

- *Gaussian*: This is the best-known example of a Lévy process. We set the distribution of increments to $\mathcal{N}(0, 1)$.
- *Laplace*: We set the location parameter to $\mu = 0$ and the scale parameter to $b = 1$.
- *Cauchy*: Similarly, we set the location parameter to $\mu = 0$ and the scale parameter to $\gamma = 1$.

In Fig. 2, we compare the reconstruction performance of our algorithm (CCS-MMSE) against TV denoising, LMMSE, and MMSE obtained through message-passing algorithm (MP-MMSE). We perform 10 random realizations of the problem for various noise levels σ^2 and plot the average MSE reduction $\Delta\mathrm{MSE} = 10\log_{10}\left(\|\mathbf{x} - \widehat{\mathbf{x}}\|_2^2/\|\mathbf{x} - \mathbf{y}\|_2^2\right)$ for each algorithm. The LMMSE, which corresponds to the MSE optimal linear estimator, was implemented by using the smoothing spline algorithm described in [19]. TV denoising

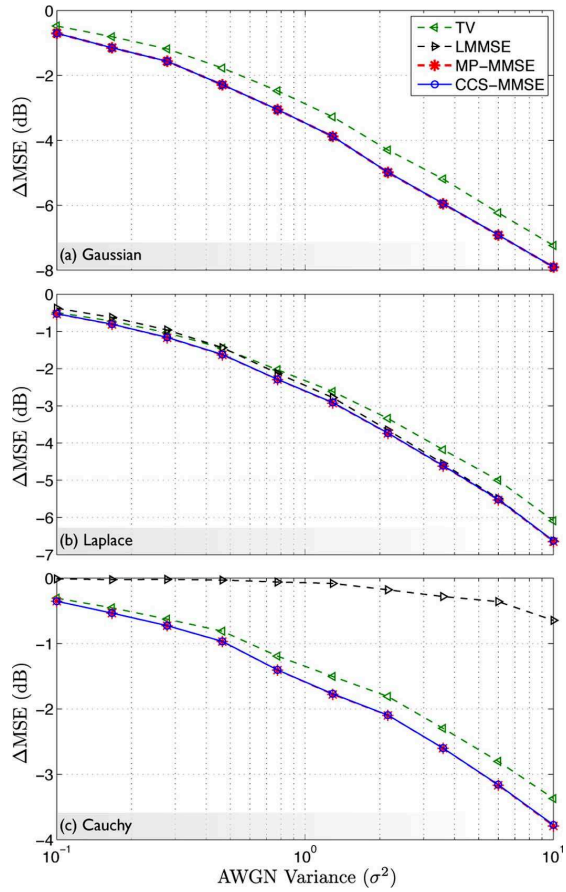

Fig. 2. Comparison of various estimation methods on Lévy processes. The average MSE reduction is plotted against AWGN variance σ^2 for Lévy processes with (a) Gaussian, (b) Laplace, and (c) Cauchy increments. In all simulations both MMSE estimators result in the best MSE reduction.

was performed using the FISTA algorithm described in [2]. We let all of the iterative methods run for 100 iterations[1].

As can be seen in Fig. 2, both CCS-MMSE and MP-MMSE result in the best MSE reduction for all signals and noise levels. It can also be concluded from Fig. 2 that the performance of LMMSE and TV methods heavily depends on the type of signal. For example, it is well known that, while the LMMSE estimator is optimal for Brownian motion (see Fig. 2(a)), it is not well-suited to the recovery of sparse signals (see Fig. 2(c)).

Although their performances coincide, CCS-MMSE is much cheaper than MP-MMSE computationally. For example, the average reconstruction speed of MP-MMSE across all noise levels was 8.14 and 51.36 seconds for Laplace and Cauchy increments, respectively, whereas the average reconstruction speed of CCS-MMSE was only 0.13 seconds for the same signals.

V. Conclusion

We have presented a promising approach for performing the MMSE denoising of signals that are whitened with a finite-difference operator. Although, the results presented here focus on 1D signals, the extensions to 2D or 3D data can be achieved by using higher-dimensional Haar-transforms. One can imagine concrete applications of the method for developing MSE optimal variational algorithms for denoising signals and images. Our numerical simulations provide strong empirical evidence that our algorithm reaches the MMSE solution. However, the theoretical confirmation of this observation requires an in-depth investigation, which remains a topic for future research. Another issue worth addressing in subsequent investigations is whether the method can be applied to general inverse problems of the type $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}$, which could potentially lead to substantial improvements in state-of-the-art methods.

[1]See supplementary material with plots that confirm the convergence at http://bigwww.epfl.ch/kamilov.

References

[1] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D*, vol. 60, no. 1–4, pp. 259–268, Nov. 1992.

[2] A. Beck and M. Teboulle, "Fast gradient-based algorithm for constrained total variation image denoising and deblurring problems," *IEEE Trans. Image Process.*, vol. 18, no. 11, pp. 2419–2434, Nov. 2009.

[3] V. Caselles, A. Chambolle, and M. Novaga, *Handbook of Mathematical Methods in Imaging*. Berlin, Germany: Springer, 2011, ch. Total Variation in Imaging, pp. 1017–1054.

[4] U. S. Kamilov, P. Pad, A. Amini, and M. Unser, "MMSE estimation of sparse Lévy processes," *IEEE Trans. Signal Process.*, vol. 61, no. 1, pp. 137–147, Jan. 2013.

[5] C. Louchet and L. Moisan, "Total variation denoising using posterior expectation," in *Eur. Signal Process. Conf*, Lausanne, Switzerland, Aug. 25–29, 2008.

[6] R. Gribonval, "Should penalized least squares regression be interpreted as maximum a posteriori estimation?," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2405–2410, May 2011.

[7] U. S. Kamilov, E. Bostan, and M. Unser, "Wavelet shrinkage with consistent cycle spinning generalizes total variation denoising," *IEEE Signal Process. Lett.*, vol. 19, no. 4, pp. 187–190, Apr. 2012.

[8] R. R. Coifman and D. L. Donoho, *Springer Lecture Notes in Statistics*. Berlin, Germany: Springer-Verlag, 1995, ch. Translation-invariant de-noising, pp. 125–150.

[9] K. Sato, *Lévy Processes and Infinitely Divisible Distributions*. Cambridge, U.K.: Cambridge Univ. Press, 1999.

[10] D. Applebaum, *Lévy Processes and Stochastic Calculus*. Cambridge, U.K.: Cambridge Univ. Press, 2009.

[11] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, Sep. 1994.

[12] P. L. Combettes and J.-C. Pesquet, "Proximal thresholding algorithm for minimization over orthonormal bases," *SIAM J. Optim.*, vol. 18, no. 4, pp. 1351–1376, 2007.

[13] A. Achim, A. Bezerianos, and P. Tsakalides, "Novel Bayesian multiscale method for speckle removal in medical ultrasound images," *IEEE Trans. Med. Imag.*, vol. 20, no. 8, pp. 772–783, Aug. 2001.

[14] J. M. Fadili and L. Boubchir, "Analytical form for a Bayesian wavelet estimator of images using the Bessel K form densities," *IEEE Trans. Image Process.*, vol. 14, no. 2, pp. 231–240, Feb. 2005.

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

[16] S. Mallat, *A Wavelet Tool of Signal Processing: The Sparse Way*, 3rd ed. San Diego, CA, USA: Academic, 2009.

[17] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd ed. Berlin, Germany: Springer, 2006.

[18] H.-A. Loeliger, J. Dauwels, J. Hu, S. Korl, L. Ping, and F. R. Kschischang, "The factor graph approach to model-based signal processing," *Proc. IEEE*, vol. 95, no. 6, pp. 1295–1322, Jun. 2007.

[19] M. Unser and T. Blu, "Generalized smoothing splines and optimal discretization of the wiener filter," *IEEE Trans. Signal Process.*, vol. 53, no. 6, pp. 2146–2159, Jun. 2005.