MMSE DENOISING OF SPARSE AND NON-GAUSSIAN AR(1) PROCESSES

Pouria Tohidi[†] *Emrah Bostan*[‡] *Pedram Pad*[‡] *Michael Unser*[‡]

[†] Electrical Engineering Department, Sharif University of Technology, Tehran, Iran
[‡]Biomedical Imaging Group, École Polytechinque Fédérale de Lausanne, Lausanne, Switzerland

ABSTRACT

We propose two minimum-mean-square-error (MMSE) estimation methods for denoising non-Gaussian first-order autoregressive (AR(1)) processes. The first one is based on the message passing framework and gives the exact theoretic MMSE estimator. The second is an iterative algorithm that combines standard wavelet-based thresholding with an optimized non-linearity and cycle-spinning. This method is more computationally efficient than the former and appears to provide the same optimal denoising results in practice. We illustrate the superior performance of both methods through numerical simulations by comparing them with other well-known denoising schemes.

***Index Terms*—** Auto-regressive, Non-Gaussian, Denoising, Minimum Mean Square Error, Message Passing, Operator-Like Wavelets, Consistent Cycle Spinning

1. INTRODUCTION

Gaussian as well as non-Gaussian auto-regressive (AR) processes are widely studied in the theory of stochastic processes and commonly used in different fields of signal processing [1, 2, 3]. Thus, the problem of finding the optimal denoising method has a special importance for these processes (optimality in the sense of mean square error (MSE)). A further motivation is to come up with a MMSE goldstandard against which to test existing denoising methods. However, this problem remains unsolved for the general non-Gaussian case until now.

In [4], the authors considered the problem of denoising Lévy processes [5], which may be interpreted as the (unstable) limit of a first order auto-regressive (AR(1)) process (i.e., with pole at the origin) [6]. Based on the property that Lévy processes have independent increments, the authors proposed a message-passing estimation method to denoise such processes. Since the corresponding graphical model is cycle-free, the obtained estimator is the exact theoretical MMSE estimator. However, the algorithm is very slow for signals of practical length.

On the other hand, continuous-time Lévy processes are self-similar and whitened by the derivative operator [6, 7]. This suggests that their coefficients in a standard wavelet basis are nearly decoupled. Exploiting this point, Kazerouni et al. derived an optimized wavelet-based denoiser for this class of processes and demonstrated empirically that it provides the MMSE solution [8]. Their algorithm combines standard wavelet-domain thresholding [9], [10] and an idea called consistent cycle-spinning. It also relies on Haar wavelets which are ideally matched to Lévy processes. This method is much faster than the one based on message passing and can be used for longer signal lengths.

In this paper, we consider the denoising problem of AR(1) processes in the general non-Gaussian setting. These processes are not whitened by the derivative operator anymore (except for the case of Lévy processes). This implies that the decomposition in a standard wavelet basis is no longer appropriate. Also, the lack of independence of their increments makes the message-passing scheme proposed in [4] inadequate.

The contribution of this work is to extend the optimal denoising methods available for Lévy processes to the richer class of non-Gaussian AR(1) processes. The non-obvious aspect is that the previous solutions take advantage of the very specific properties of Lévy processes; in particular, the independence of their increments. The leading idea of our approach is to replace finite differences by linear-prediction residuals and the Haar wavelets of [8] by an alternative operator-like basis that is matched to the processes [11]. The justification for this latter choice is that these operator-like wavelets have been shown recently to provide an *independent component analysis* for the sub-class of α-stable AR(1) processes [12]. A remarkable outcome is that the performance of our new wavelet-based denoiser matches that of our generalized message-passing-based algorithm which is guaranteed to be MMSE-optimal for the general AR(1) processes (since the factor graph has no loops).

The paper is organized as follows: In Section 2, we discuss the continuous-time and discrete-time AR(1) processes and give the exact mathematical model that we consider in this paper. In Section 3, we describe our extended message-passing-based MMSE estimator. In Section 4, we introduce our corresponding wavelet-based denoiser. Section 5 is dedicated to the experimental results and contains graphs in which the performances of our estimators are compared with some

other widely used denoising methods. At the end, we summarize the results and shortly discuss the possible future studies.

2. PRELIMINARIES AND MATHEMATICAL MODEL

In this section, we first present our extended class of non-Gaussian (or sparse) continuous-time AR(1) processes using the innovation model proposed in [13] and [14]. The model boils down to the linear filtering of a non-Gaussian white noise (innovation) by using a suitable (1st order) shaping filter. Then, we move on to the discrete-time AR(1) process and we give the exact model that we consider for our problem.

A continuous-time AR(1) process is whitened by operator $\mathrm{D}+\kappa\mathrm{I}$ where D and I stand for the derivative and the identity operators, respectively, and $\kappa \geq 0$ is the AR(1) parameter; i.e., s is an AR(1) process if

$$(\mathrm{D}+\kappa\mathrm{I})\{s\} = w \tag{1}$$

where w is a continuous-time white Lévy noise [13]. The crucial extension here is that the continuous-time innovation w is not necessarily Gaussian. Equivalently, for the case of $\kappa > 0$, we can write

$$s = h * w \tag{2}$$

in which $*$ is the convolution operator and

$$h(t) = \mathrm{e}^{-\kappa t}\mathbf{1}_+(t) \tag{3}$$

is the Green's function of $\mathrm{D}+\kappa\mathrm{I}$ where $\mathbf{1}_+(t)$ is the unit step function. However, in the case that $\kappa = 0$ which leads to the generation of Lévy processes, there are some subtle theoretical points that need to be taken care of [13].

As it is shown in [12], if we sample the process $s(t)$ ideally and uniformly with period T, the obtained discrete process is also a discrete AR(1) process. Precisely, if we define

$$s[k] = s(kT), \tag{4}$$

for $k \in \mathbb{Z}$, then we have the linear prediction formula

$$s[k] = \rho s[k-1] + w[k] \tag{5}$$

in which $\rho = \mathrm{e}^{-\kappa T}$ and

$$w[k] = \int_{(k-1)T}^{kT} \mathrm{e}^{-\kappa(\tau-kT)} w(\tau) d\tau. \tag{6}$$

According to (6), the linear prediction error $w[k]$ is a sequence of (non-Gaussian) i.i.d. random variables whose distribution depends on the underlying innovation w, the parameter of AR(1) process κ and the sampling period T. Thus, the impulse response of the whitening operator of the discrete-time signal is

$$h[n] = \begin{cases} 1 & n = 0 \\ -\rho & n = 1 \\ 0 & \text{otherwise} \end{cases}. \tag{7}$$

Define the vector of N samples of s as

$$\mathbf{s} = \begin{bmatrix} s[0] & s[1] & \cdots & s[N-1] \end{bmatrix}. \tag{8}$$

Now, the problem considered here is that we are given

$$\mathbf{y} = \mathbf{s} + \mathbf{n} \tag{9}$$

where $\mathbf{n}$ is an N-dimensional Gaussian noise vector with i.i.d. entries with zero mean and variance σ^2 and we want to find the MMSE estimator of $\mathbf{s}$.

3. MMSE ESTIMATOR BASED ON MESSAGE PASSING

Our message passing algorithm is derived from a factor graph that summarizes the statistical dependencies of the discrete process. Since for any value of κ there is a Markov dependency between entries of $\mathbf{s}$, we notice that the structure of the factor-graph is exactly the same as that of Lévy processes ($\kappa = 0$) described in [4]. This implies that the network through which the messages are passed remains the same. Therefore, the pseudo-code of the algorithm can be written as in Algorithm 1. The crucial extension here is that the messages now depend on κ which somewhat complicates the implementation due to the scaling by $\rho \neq 1$.

Algorithm 1 Message Passing MMSE Denoising for AR(1) Processes

$\mu^{-}_{s[0]}(x) \leftarrow p_w(x)$
$\mu^{+}_{s[N-1]}(x) \leftarrow 1$
for $i = 1$ **to** $N-1$ **do**
  Compute $\mu^{-}_{s[i]}(x)$ from $\mu^{-}_{s[i-1]}(x)$
  Compute $\mu^{+}_{s[N-1-i]}(x)$ from $\mu^{+}_{s[N-i]}(x)$
end
for $k = 0$ **to** $N-1$ **do**
  $p_{s[k]|\mathbf{y}}(x|\mathbf{y}) \propto \mu^{-}_{s[k]}(x)\, p_{y[k]|\mathbf{s}}(y_k|\mathbf{s})\, \mu^{+}_{s[k]}(x)$
  $\hat{s}_{\mathrm{MMSE}}[k] = \int_{\mathbb{R}} x p_{s[k]|\mathbf{y}}(x|\mathbf{y}) dx$
end
$\hat{\mathbf{s}}_{\mathrm{MMSE}} \leftarrow \begin{bmatrix} \hat{s}_{\mathrm{MMSE}}[0] & \cdots & \hat{s}_{\mathrm{MMSE}}[N-1] \end{bmatrix}$
return $\hat{\mathbf{s}}_{\mathrm{MMSE}}$

The expression for computing forward messages is

$$\mu^{-}_{s[i]}(x) \propto \int_{\mathbb{R}} \mu^{-}_{s[i-1]}(z) p_{\mathbf{y}|\mathbf{s}}(y_i|z) p_w(x-\rho z) dz \tag{10}$$

and for the backward messages is

$$\mu^{+}_{s[N-1-i]}(x) \propto \int_{\mathbb{R}} \mu^{+}_{s[N-i]}(z) p_{\mathbf{y}|\mathbf{s}}(y_{N-i+1}|z) p_w(z-\rho x) dz \tag{11}$$

in which y_i is the ith entry of $\mathbf{y}$, p denotes the probability density function of the random variable in the subscript and the

symbol $\propto$ indicates equality after normalization that forces the μ-functions to integrate to 1.

In the case of the Lévy processed considered in [4] where $\kappa = 0$ and $\rho = 1$, the authors could evaluate (10) and (11) simply by saving all messages and probability density functions over a fine enough grid and using a Riemann sum approximation of the integrals (notice that when $\rho = 1$ the integrals are simple convolutions). In the more general AR(1) scenario for which $\kappa > 0$ and $\rho < 1$, if we start with a certain grid on x for storing the values of the μ functions, the grid will expand in the case of μ^- and shrink in the case of μ^+ by a factor of ρ after each iteration of the algorithm. Thus, after a few iterations, the numerical approximation of the underlying integrals becomes meaningless.

We overcome this problem by switching to the Fourier domain to evaluate the convolution-like integrals as multiplications. Our algorithm is justified as follows. First, we define

$$q(x) = p_w(\rho x). \tag{12}$$

For the μ^- functions, we have

$$\mu^-_{s[i]}(x) \propto \int_{\mathbb{R}} \underbrace{\mu^-_{s[i-1]}(z) p_{\mathbf{y}|\mathbf{s}}(y_{i-1}|z)}_{f_{i-1}(z)} p_w(x - \rho z) dz \tag{13}$$

$$= \int_{\mathbb{R}} f_{i-1}(z) q\big(\frac{x}{\rho} - z\big) dz \tag{14}$$

which can be summarized in a convolution form

$$\mu^-_{s[i]}(x) \propto \{f_{i-1} * q\}(\frac{x}{\rho}) \tag{15}$$

In the Fourier domain, this translates into

$$\mu^-_{s[i]}(x) \propto \mathcal{F}^{-1}\{\mathcal{F}\{f_{n-1}\} \times \mathcal{F}\{q\}\}(\frac{x}{\rho}) \tag{16}$$

where $\mathcal{F}$ and $\mathcal{F}^{-1}$ are the Fourier transform and its inverse, respectively. F

Similarly, for the μ^+ functions, we obtain

$$\mu^+_{s[i]}(x) \propto \int_{\mathbb{R}} \underbrace{\mu^+_{s[i+1]}(z) p_{y_{i+1}|\mathbf{s}}(y_{i+1}|z)}_{g_{i+1}(z)} p_w(z - \rho x) dz \tag{17}$$

and finally end up with

$$\mu^+_{s[i]}(x) \propto \mathcal{F}^{-1}\left\{\mathcal{F}\{g_{i+1}\} \times \mathcal{F}\{p_w\}\right\}(\rho x) \tag{18}$$

Hence, by implementing the steps 4 and 5 in the algorithm according to calculations (16) and (18), we are able to implement the MMSE estimator for general non-Gaussian AR(1) processes.

4. WAVELET-BASED ESTIMATION USING CONSISTENT CYCLE SPINNING

We now describe our new iterative wavelet-based algorithm for the estimation of $\mathbf{s}$. The key idea here is to *decouple* the AR(1) process (*i.e.*, rendering its independent components). According to [12], the wavelet-like basis that best approximates an independent component analysis (ICA) for AR(1) processes is the operator-like wavelet that matches the operator $\mathrm{D} + \kappa\mathrm{I}$. We thus use the operator-like wavelet basis and construct a wavelet frame that is tight and shift-invariant. We then solve the following optimization problem:

$$\hat{\mathbf{v}} = \arg\min_{\mathbf{v}} \left(\frac{1}{2}\|\mathbf{u} - \mathbf{v}\|_2^2 + \Phi(\mathbf{v}) \,:\, \mathbf{H}^\top \mathbf{H}\mathbf{v} = \mathbf{v} \right). \tag{19}$$

Let us look at (19) in more details. The operator $\mathbf{H}$ corresponds to the operator-like wavelet transform and $\mathbf{u} = \mathbf{H}\mathbf{y} \in \mathbb{R}^M$ with $M > N$. The constraint in (19) that $\mathbf{H}^\top\mathbf{H}\mathbf{v} = \mathbf{v}$ makes the solution *consistent* in the sense that the wavelet transform of the final solution $\hat{\mathbf{s}} = \mathbf{H}^\top\hat{\mathbf{v}}$ is necessarily equal to $\hat{\mathbf{v}}$ that is the solution of (19). Note that—since $\mathbf{H}$ is redundant—there exist some $\mathbf{v}$ such that $\mathbf{H}^\top\mathbf{H}\mathbf{v} \neq \mathbf{v}$, which are excluded by our formulation. As for the regularization of the solution, we choose

$$\Phi(\mathbf{v}) = \sum_{i=1}^{N} \phi_{\mathrm{MMSE}}(v_d^i), \tag{20}$$

where v_d^i denotes the detail coefficient at location i, and ϕ_{MMSE} is a scalar-valued potential function whose proximal operator corresponds to the MMSE shrinkage (see Stein's formula [15]):

$$\hat{w}_i = \nu_{\mathrm{MMSE}}(u_i) = \mathbb{E}(v|u_i)$$
$$= u_i + \sigma_i^2 \frac{\mathrm{d}}{\mathrm{d}u_i} \log p_{u_i}(u_i), \tag{21}$$

where σ_i^2 is the noise variance and p_{u_i} is the pdf of the wavelet coefficient of the noisy signal at location i. The latter is given by the convolution

$$p_{u_i}(t) = p_{v_i}(t) * \frac{1}{\sqrt{2\pi\sigma_i^2}} \mathrm{e}^{-\frac{t^2}{2\sigma_i^2}} \tag{22}$$

with p_{v_i} being the pdf of the detail coefficient of the original signal. Note that p_{v_i} is determined by the pdf of the samples of the innovation process. There is also the theoretical guarantee that the corresponding MMSE potential function ϕ_{MMSE} exists [16]. From a computational point of view, (21) can be pre-computed and stored as a look-up table (LUT). In particular, when the process is Gaussian, the solution is a simple linear (Wiener) estimator; otherwise it is a generalized threshold. For the resolution of (19), one can rely on the augmented Lagrangian approach, which results in an iterative framework. Otherwise, the algorithmic steps are the same as those of the method in [8].

5. NUMERICAL RESULTS

Based on our developments, we consider the problem of denoising AR(1) processes. We choose $\kappa = 0.4$ and generate

signals of length $N = 128$. We consider the cases of Gaussian, Laplace and Cauchy distributed innovation processes. The AR(1) signal is degraded by additive white Gaussian noise (AWGN) with various variance values.

We then compare the denoising performance of the following methods:
1) MP: The method refers to the message-passing algorithm derived in Section 3.
2) CCS-MMSE: The method refers to the estimate $\hat{\mathbf{s}} = \mathbf{H}^\top \hat{\mathbf{v}}$ where $\hat{\mathbf{v}}$ is the solution of (19).
3) TV: The method refers to the total variation denoising algorithm.
4) CCS-MAP: The method is similar to CCS-MMSE. The MMSE shrinkage operator is replaced with the MAP shrinkage function[1].
5) ST: The method refers to soft thresholding denoising in the orthogonal wavelet domain where we used the operator-like wavelets for the signal expansion.
6) W-MMSE: The method is similar to ST. The soft thresholding operator is replaced with the MMSE shrinkage function.
7) W-MAP: The method is similar to ST. The soft thresholding operator is replaced with the MAP shrinkage function.

The algorithms are compared in terms of SNR improvement, which simply refers to the SNR difference between the denoised and the noisy signals. For iterative denoising algorithms —*i.e.*, CCS-MMSE, CCS-MAP, and TV—, we perform 250 iterations, where TV denoising is implemented by following the FISTA framework [17]. For the deterministic methods, the parameters (namely, the threshold value for ST and the regularization parameter for TV) is optimized for the best-possible SNR performance using an oracle. The experiment is repeated 300 times and the results are given in Figure 1.

By looking at the results, we see that the SNR improvement curves of the MP and CCS-MMSE are indistinguishable from each other and outperform the rest of the algorithms. Also, it is seen that this behaviour is consistent and independent of the underlying innovation process. This shows that the experimentally CCS-MMSE achieves the exact MMSE estimation performance of MP. We also see that the performance of TV is better than CCS-MAP. This reflects the importance of choosing appropriate shrinkage functions. Finally, we note that the non-iterative algorithms—W-MMSE, W-MAP and ST—are not competitive compared to the iterative denoising algorithms considered in the experiments.

6. CONCLUSION

We intoduced two algorithms for computing the MMSE estimator of an AR(1) processes contaminated with additive white Gaussian noise. To the best of our knowledge, these

[1]Note that for the Gaussian innovation case, MMSE and MAP shrinkage functions are the same.

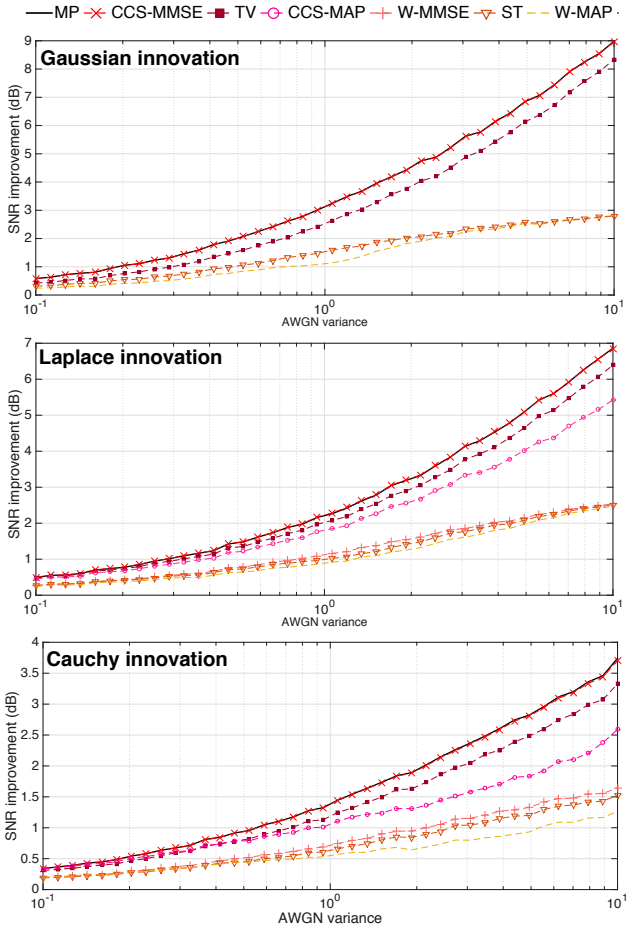

Fig. 1. SNR improvement versus noise variance for different estimation schemes for denoising AR(1) processes with different distributions.

are the first practical solutions for the optimal denoising of non-Gaussian stationary processes. The first method is based on the message passing framework and is guaranteed to yield the theoretical MMSE estimate. The second is a wavelet-based method that is computationally more efficient, and its results are found to be indistiguishable from the MMSE solution. However, we have not yet been able to establish its optimality theoretically. The results for AR(1) processes with different input distributions are also plotted and compared to some of the well-known denoising schemes, confirming the superiority of our approaches.

7. REFERENCES

[1] S. I. Resnick, "Heavy tail modeling and teletraffic data: special invited paper," *The Annals of Statistics*, vol. 25, no. 5, pp. 1805–1869, 1997.

[2] C. M Gallagher, "A method for fitting stable autoregressive models using the autocovariation function," *Statistics & Probability Letters*, vol. 53, no. 4, pp. 381–390, 2001.

[3] S. Ling, "Self-weighted least absolute deviation estimation for infinite variance autoregressive models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 67, no. 3, pp. 381–393, 2005.

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

[5] K. Sato, *Lévy Processes and Infinitely Divisible Distributions*, Cambridge University Press, 1999.

[6] M. Unser and P. D. Tafti, *An Introduction to Sparse Stochastic Processes*, Cambridge University Press, 2014.

[7] E. Bostan, U. S. Kamilov, M. Nilchian, and M. Unser, "Sparse stochastic processes and discretization of linear inverse problems," *IEEE Transactions on Image Processing*, vol. 22, no. 7, pp. 2699–2710, July 2013.

[8] A. Kazerouni, U. S. Kamilov, E. Bostan, and M. Unser, "Bayesian denoising: From MAP to MMSE using consistent cycle spinning," *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 249–252, March 2013.

[9] D. L. Donoho, "Nonlinear wavelet methods for recovery of signals, densities, and spectra from indirect and noisy data," in *Proceedings of Symposia in Applied Mathematics*. American Mathematical Society, 1993, vol. 47, pp. 173–205.

[10] C. Taswell, "The what, how, and why of wavelet shrinkage denoising," in *Computing in Science Engineering*, May/June 2000, vol. 2, no. 3, pp. 12–19.

[11] I. Khalidov and M. Unser, "From differential equations to the construction of new wavelet-like bases," in *IEEE Transactions on Signal Processing*, April 2006, vol. 54, no. 4, pp. 1256–1267.

[12] P. Pad and M. Unser, "Optimality of operator-like wavelets for representing sparse AR(1) processes," *IEEE Transactions on Signal Processing*, vol. 63, no. 18, pp. 4827–4837, September 2015.

[13] M. Unser, P. Tafti, and Q. Sun, "A unified formulation of Gaussian vs. sparse stochastic processes - part i: Continuous-domain theory," *IEEE Transactions on Information Theory*, vol. 60, no. 3, pp. 1945–1962, March 2014.

[14] M. Unser, P. D. Tafti, A. Amini, and H. Kirshner, "A unified formulation of Gaussian vs. sparse stochastic processes - part ii: Discrete-domain theory," *IEEE Transactions on Information Theory*, vol. 60, no. 5, pp. 3036 – 3051, May 2014.

[15] C. M. Stein, "Estimation of the mean of a multivariate normal distribution," in *Annals of Statistics*, November 1981, vol. 9, no. 6, pp. 1135–1151.

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

[17] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.