ON THE OPTIMALITY OF OPERATOR-LIKE WAVELETS FOR SPARSE AR(1) PROCESSES

Pedram Pad and Michael Unser

Biomedical Imaging Group, EPFL, Switzerland

ABSTRACT

Sinusoidal transforms such as the DCT are known to be optimal—that is, asymptotically equivalent to the Karhunen-Loève transform (KLT)—for the representation of Gaussian stationary processes, including the classical AR(1) processes. While the KLT remains applicable for non-Gaussian signals, it loses optimality and, is outperformed by the independent-component analysis (ICA), which aims at producing the most-decoupled representation. In this paper, we consider an extension of the classical AR(1) model that is driven by symmetric-alpha-stable (SαS) noise which is either Gaussian ($\alpha = 2$) or sparse ($0 < \alpha < 2$). For the sparse (non-Gaussian) regime, we prove that an expansion in a proper wavelet basis (including the Haar transform) is much closer to the optimal orthogonal ICA solution than the classical Fourier-type representations. Our criterion for optimality, which favors independence, is the Kullback-Leibler divergence between the joint pdf of the original signal and the product of the marginals in the transformed domain. We also observe that, for very sparse AR(1) processes ($\alpha \leq 1$), the operator-like wavelet transform is indistinguishable from the ICA solution that is determined through numerical optimization.

Index Terms— Operator-like wavelets, Independent component analysis, Auto-regressive processes

1. INTRODUCTION

Wavelets are widely used for signal and image processing. Typical examples of application are JPEG2000 for image compression [1] and shrinkage methods for denoising [2]. However, despite this popularity in practice, there are few theoretical results about their optimality for the representation of stochastic processes.

It is known that that wavelets are optimal (up to some constant) for the N-term approximation of deterministic functions in Besov spaces [3]. There is also some empirical evidence of the statistical optimality of wavelet-like functions. In [4], Cardoso and Donoho performed two independent component analysis (ICA) experiments with realizations of Meyer's ramp process, which is non-stationary with the same second-order statistics as Brownian motion [5], and the sawtooth process—a stationary variant of the former. For both cases, they observed that the basis vectors of ICA had a multiresolution structure similar to wavelets. Another well-known study is Olshausen and Field's ICA experiment on a huge database of natural images. These authors pointed out that the resulting components have properties that are reminiscent of 2D wavelets and/or Gabor functions and made interesting connections with visual perception [6].

Recently, Unser et al. have specified a broad class of sparse stochastic processes based on a generalized continuous-domain innovation model [7, 8]. These processes are specified as solutions of linear stochastic differential equations (SDE) driven by *general* continuous-domain white noise that is not necessarily Gaussian.

In this paper, we rely on a generalized innovation model to establish the optimality of a class of wavelets in some stochastic sense. We focus on the sparse AR(1) processes that are defined by first-order SDE driven by α-stable noise. AR(1) systems and α-stable distributions are at the core of signal modeling and probability theory. Since stable processes have heavy-tailed statistics for $\alpha < 2$, they are prototypical representatives for sparse signals [9], while one recovers the classical Gaussian processes for $\alpha = 2$.

Specifically, we are interested in determining the best orthogonal expansion of these processes with minimal dependency between transform-domain coefficients. Our measure of quality is based on the Kullback-Leibler divergence. Classically, we know that, for $\alpha = 2$ (Gaussian input), Fourier-type transforms (FT) such as the DCT are asymptotically equivalent to the Karhunen-Loéve transform [10, 11] and thus result in a fully decoupled (independent) representation. In this paper, we shall see that, for $\alpha < 2$, this classical result does not hold anymore and that the operator-like wavelet transforms proposed in [12] outperform the FT. Also, by finding the optimal transform for different values of α, we shall observe that, for α less than some threshold, operator-like wavelet are optimal.

We start with preliminaries about the exact signal model and the concept of operator-like wavelets. In Section 3, we describe our performance criterion and provide an iterative procedure for finding the optimal basis (ICA). Results for different AR(1) processes and different transform domains are

This work was supported by the European Commission under Grant ERC-2010-AdG 267439- FUN-SP.

discussed in Section 4. The last two sections are dedicated to the recapitulation of the main results, and the relation to prior work.

2. PRELIMINARIES

2.1. Continuous and Discrete SαS AR(1) Processes

A continuous-domain symmetric-α-stable AR(1) process s can be generated by applying a first-order differential system to a white noise excitation as

$$s(t) = (\rho_\kappa * w)(t) = \int_{-\infty}^{t} \mathrm{e}^{\kappa(t-\tau)} w(\tau)\mathrm{d}\tau, \quad (1)$$

where w is an SαS innovation process; that is, a continuous-domain α-stable white noise that is formally equivalent to the weak derivative of an SαS Lévy process [7, 13]. The impulse response of the system is the causal exponential $\rho_\kappa(t) = \mathrm{e}^{\kappa t}\mathbf{1}_+(t)$, where $\mathbf{1}_+(t)$ is the unit step. The AR(1) process is well-defined for $\kappa < 0$. The limit case $\kappa = 0$ can also be handled by changing the lower limit of integration in (1) from $-\infty$ to 0, which results in a Lévy process that is non-stationary.

It can be shown (see [7]) that (1) is formally equivalent to the innovation model[1]

$$\mathrm{L}s = w, \quad (2)$$

where $\mathrm{L} = \frac{\mathrm{d}}{\mathrm{d}t} + \kappa \mathrm{I}$ is the whitening operator of the continuous-domain AR(1) process.

Now, if we sample s at the integers, we get a sequence $\{s_k = s(k)\}_{k\in\mathbb{Z}}$ that satisfies the first-order difference equation

$$s_k - \mathrm{e}^{\kappa} s_{k-1} = w_k \quad (3)$$

with

$$w_k = \langle w, \beta_\kappa(\cdot - k)\rangle,$$

where $\beta_\kappa(t) = \mathbf{1}_{[0,1)}\mathrm{e}^{\kappa t}$ is the exponential B-spline with parameter κ [8]. Since the kernels $\{\beta_\kappa(\cdot - k)\}_{k\in\mathbb{Z}}$ have disjoint support, according to the properties of the white noise, $\{w_n\}_{k\in\mathbb{Z}}$ is an i.i.d. sequence of SαS random variables with the common characteristic function

$$\mathbb{E}\left\{\mathrm{e}^{\mathrm{j}\omega\langle \mathrm{w},\beta_\kappa\rangle}\right\} = \mathrm{e}^{-|\|\beta_\kappa\|_\alpha \omega|^\alpha} \quad (4)$$

and width parameter $\|\beta_\kappa\|_\alpha$, which is the L_α (pseudo)norm of the B-spline. The conclusion is that the continuous-domain model (1) maps into the discrete AR(1) process $\{s_k\}_{k=1}^{\infty}$ that is uniquely specified by (3).

[1]The proper mathematical interpretation of this equation is in the weak sense of generalized functions, with $\langle\varphi, \mathrm{L}s\rangle = \langle\varphi, w\rangle$ for all φ in Schwartz class of smooth and rapidly decreasing test functions.

We now consider n consecutive samples of the process and define the random vectors

$$\mathbf{s} = \begin{bmatrix} s_1 \\ \vdots \\ s_n \end{bmatrix} \quad \text{and} \quad \mathbf{w} = \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix}.$$

This allows us to rewrite (3) as

$$\mathbf{s} = \mathbf{L}\mathbf{w} \quad (5)$$

in which $\mathbf{L} = [l_{ij}]_{n\times n}$ and

$$l_{ij} = \mathrm{e}^{\kappa(j-i)} \cdot \mathbf{1}_{\{j\geq i\}}.$$

This yields the discrete-domain counterpart of the innovation model (2).

In the next sections, we are going to study linear transforms applied to the signal s (or $\mathbf{s}$). Here we recall a fundamental property of stable distributions that we shall use in our derivations.

Property 1 (Linear combination SαS random variables) *If $r_1, \ldots, r_k$ are i.i.d. random variables with symmetric α-stable distributions around 0 with width parameter 1, and if $a_1, \ldots, a_k$ are k real numbers, then $a_1 r_1 + \cdots + a_k r_k$ has the same distribution as $(\sum_{i=1}^{k} |a_i|^\alpha)^{1/\alpha} r_1$.*

2.2. Operator-Like Wavelets

It is well-known that conventional wavelet bases act like smoothed versions of a derivative operator. If we want to apply a wavelet-like transform to uncouple the AR(1) signal (1), we need to select basis functions that essentially behave like the whitening operator L in (2). The good news is that such wavelet-like basis functions exist and that they can be tailored to any given differential operator L [12]. Specifically, the operator-like wavelet at scale i and location k is given by

$$\psi_{i,k} = \mathrm{L}^* \phi_i(\cdot - 2^i k),$$

where ϕ_i is a scale-dependent smoothing kernel. Based on the fact that $s = \mathrm{L}^{-1}w$ and the orthogonality of $\{\psi_{i,k}\}$, we can compute the wavelet coefficients of the signal as

$$\begin{aligned} v_{i,k} &= \langle s, \psi_{i,k}\rangle = \langle \mathrm{L}^{-1}w, \psi_{i,k}\rangle \\ &= \langle w, \mathrm{L}^{-1*}\mathrm{L}^*\phi_i(\cdot - 2^i k)\rangle = \langle w, \phi_i(\cdot - 2^i k)\rangle, \end{aligned} \quad (6)$$

from which we can deduce two properties (see [7]). First, the wavelet coefficients at scale i follow an SαS distribution with width parameter $\|\phi_i\|_\alpha$. Second, since w is independent at every point, the level of decoupling is directly dependent upon the degree of overlap of the smoothing kernels $\phi_i(\cdot - 2^i k)$. In the case of a first-order operator, the operator-like wavelets of Khalidov et al [12] are very similar to Haar wavelets, except that they are piecewise exponential instead of piecewise constant (for $\kappa = 0$). They are orthogonal with non-overlapping support within a given scale. This allows us to conclude that the operator-like wavelet coefficients are independent and identically (SαS) distributed within each scale. This property suggest that this type of transform is an excellent candidate for decoupling AR(1) processes.

3. DEPENDENCY OF THE COEFFICIENTS OF REPRESENTATION OF AR(1) PROCESSES IN TRANSFORM DOMAINS

The transformed representation of the signal is denoted by $\tilde{\mathbf{s}} = [\tilde{s}_1 \ \cdots \ \tilde{s}_n]^\top = \mathbf{Hs}$, where $\mathbf{H} = [h_{ij}]_{n\times n}$ is the underlying orthogonal transformation matrix (DCT, wavelet transform, or ICA). We take the Kullback-Leibler Distance (KLD) between the exact probability density function $p_{\tilde{\mathbf{s}}}(\tilde{\mathbf{s}})$ of $\tilde{\mathbf{s}}$ and the product $p_{\tilde{s}_1}(\tilde{s}_1)\cdots p_{\tilde{s}_n}(\tilde{s}_n)$ of its marginals as the measure of quality of the transformation $\mathbf{H}$. It is given by

$$\mathrm{R}(\mathbf{H}) = \frac{1}{n}\mathbb{D}(p_{\tilde{\mathbf{s}}}(\tilde{\mathbf{s}}) \,\|p_{\tilde{s}_1}(\tilde{s}_1)\cdots p_{\tilde{s}_n}(\tilde{s}_n)), \quad (7)$$

where $\mathbb{D}(\cdot\|\cdot)$ is the KLD function. The function $\mathrm{R}(\mathbf{H})$ depends continuously on $\mathbf{H}$. Since $\mathrm{R}(\mathbf{H}) \geq 0$ and since the equality holds if and only if the $\tilde{s}_i$ are completely independent, a lower value of R means that the $\tilde{s}_i$ are less dependent. This criterion is commonly used in ICA to find the most-independent representation [14].

Simplifying (7), we write

$$\begin{aligned}\mathrm{R}(\mathbf{H}) &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{H}(\tilde{s}_i) - \frac{1}{n}\mathbb{H}(\tilde{\mathbf{s}}) \quad (8)\\ &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{H}(\tilde{s}_i) - \mathbb{H}(w_1) - \frac{1}{n}\log\det\mathbf{HL},\end{aligned}$$

where $\mathbb{H}(\cdot)$ is the entropy function. The first observation is that $\log\det\mathbf{HL} = 0$. In addition, since the w_i are α-stable, the distribution of $\tilde{s}_i$ is the same as the distribution of $\bar{h}_i w_1$, where $\bar{h}_i$ is the α-(pseudo)norm of the i^{th} row of $\mathbf{HL}$ (see Property 1). Thus,

$$\mathrm{R}(\mathbf{H}) = \frac{1}{n}\sum_{i=1}^{n}\log\bar{h}_i, \quad (9)$$

where

$$\bar{h}_i = \left(\sum_{r=1}^{n}\left|\sum_{k=1}^{n}h_{ik}l_{kr}\right|^{\alpha}\right)^{\frac{1}{\alpha}}. \quad (10)$$

These simple formulas can be calculated for any given $\mathbf{H}$.

While ICA is usually determined empirically based on the observations of a process, we take advantage of the underlying stochastic model to derive an optimal solution based on the minimization of (9), which involves the computation of ℓ_α norms of the transformation matrix. Specifically, we implemented the following iterative algorithm, which finds the optimal transform $\mathbf{H}$ for different values of κ, α, and n:

- Initialize $\mathbf{H}$ and $\eta > 0$.
- Repeat
  - $*$ $\tilde{h}_{ij} \leftarrow h_{ij} - \eta\frac{\partial \mathrm{R}}{\partial h_{ij}}$ for all $i, j = 1, \ldots, n$.
  - $*$ Set $\mathbf{H}$ to the projection of $\tilde{\mathbf{H}} = \left[\tilde{h}_{ij}\right]_{n\times n}$ onto the space of unitary matrices.

  until *convergence*.

The algorithm requires the computation of the partial derivatives of $\mathrm{R}(\mathbf{H})$ whose closed formula

$$\frac{\partial \mathrm{R}}{\partial h_{ij}} = \frac{1}{n\bar{h}_i^{\alpha}}\sum_{r=1}^{n}l_{jr}\mathrm{sgn}\left(\sum_{k=1}^{n}h_{ik}l_{kr}\right)\left|\sum_{k=1}^{n}h_{ik}l_{kr}\right|^{\alpha-1}$$

is derived from (9) and (10).

4. RESULTS

First, we investigate the case of a Lévy process (i.e., $\kappa = 0$) for which the operator-like wavelet transform coincides with the classical Haar transform. We give in Figure 1 the value of our performance criterion as a function of α for various transforms with $n = 64$: the identity (which is used as baseline), the discrete cosine transform (DCT), the Haar wavelet transform (HWT), and the optimal solution (ICA) provided by our algorithm. For $\alpha = 2$ (Gaussian scenario), the process s is a Brownian motion whose KLT is a sinusoidal transform that is known analytically. In that case, we observe that R vanishes for the DCT and the optimal transform, which is consistent with the fact that they both converge to the KLT. The latter achieves perfect decorrelation, but this is equivalent to independence only in the Gaussian case. By contrast, as α decreases, the DCT becomes less favorable while the performance of the HWT gets closer to the optimal one. In fact, it becomes indistinguishable from that of ICA for $\alpha \leq 1$, which is a remarkable finding since ICA is tuned to the data while HWT is not.

Next, we switch to a stationary AR(1) process with $\mathrm{e}^{\kappa} = 0.9$ and $n = 64$. For $\alpha = 2$, we get the classical Gaussian AR(1) process which is well studied in the literature and for which the DCT is known to be asymptotically optimal [10, 11]. The performance curves for the DCT, the HWT, the operator-like wavelet matched to the process, and the optimal ICA solution are plotted in Figure 2. Here too, the trend is essentially the same as before, with the operator-like wavelet transform mimicking the optimal ICA solution as $\alpha \leq 1$ (see Figure 3). Also, note that the operator-like wavelet transform generally outperforms the HWT. This highlights the benefits of tuning the wavelet to the differential characteristics of the process.

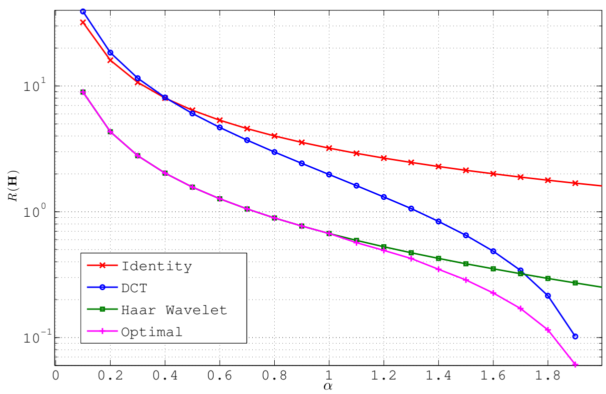

Fig. 1. $\mathrm{R}(\mathbf{H})$ of Lévy processes versus α when $n = 64$ for different $\mathbf{H}$.

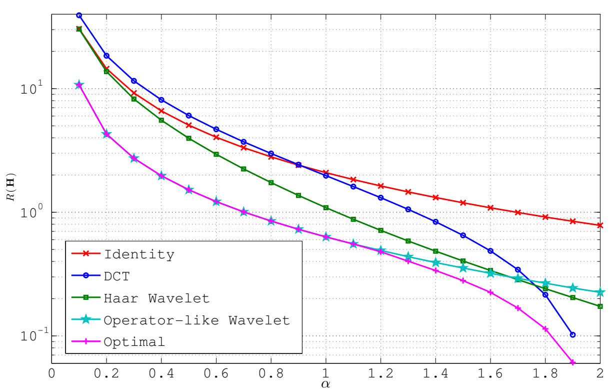

Fig. 2. $\mathrm{R}(\mathbf{H})$ versus α when $\mathrm{e}^{\kappa} = 0.9$ and $n = 64$ for different $\mathbf{H}$.

To substantiate those findings, we could also prove the theorem below, which states that, for any $\alpha < 2$, the operator-like wavelet transform outperforms the DCT (or, equivalently, the KLT associated with the Gaussian member of the family) as the block-size n tends to infinity.

Theorem 1 *If* $\alpha < 2$ *and* $\mathrm{Re}(\kappa) \leq 0$*, we have*

$$\lim_{n\to\infty} R(\mathit{OpWT}) < \lim_{n\to\infty} R(\mathit{DCT}) = \infty, \qquad (11)$$

where OpWT stands for the operator-like wavelet transform.

The proof is omitted in the interest of conciseness.

5. SUMMARY AND FUTURE STUDIES

In this paper, we have shown that operator-like wavelets are the optimal basis for representing very sparse ($\alpha \lesssim 1$) α-stable AR(1) processes. Also, we saw that, for any $\alpha < 2$, these wavelets almost systematically outperform sinusoidal transforms. This result is unexpected because the DCT is known to be asymptotically optimal for $\alpha = 2$. The criterion we used is the KLD between the distribution of the coefficients in the transform domain and the product of its marginals.

A challenging topic that deserves further investigation is finding a closed form for the optimal transform for any given α, κ, and n. Our results also call for extensions for broader classes of stochastic processes within the boundaries of generalized innovation model, such as other white noises or/and higher-order differential operators. In addition, there is a strong incentive for studying the problem in the original continuous domain.

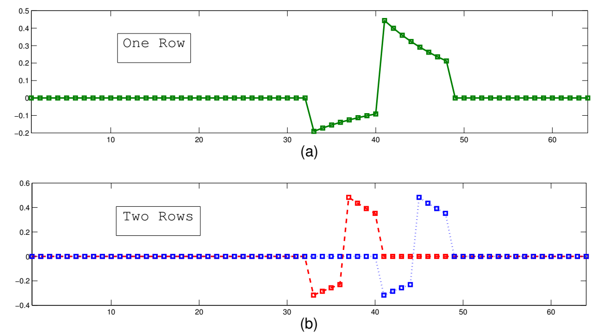

Fig. 3. Three rows of the optimal $\mathbf{H}$ for $\alpha = 1$ and $n = 64$. Parts (a) and (b) show the dyadic structure of the wavelets.

6. RELATION TO PRIOR WORKS

The present study builds upon the theory of sparse stochastic processes proposed by Unser et al., where the argument is made that continuous-domain innovation models do induce a sparse behavior when switching to a non-Gaussian excitation within the limit of mathematical admissibility [7]. Here, we focus on the simplest non-Gaussian version of this model (first-order differential system) with an SαS excitation. Due to the underlying innovation model and the properties of SαS laws, we are able to obtain an explicit characterization of the optimal transform for this particular class of stochastic processes via Equations (9) and (10), which is a novel model-based point of view for ICA. As far as we know, the conclusions that we draw are the first theoretical results on the optimality of wavelet-like basis functions for a given class of stochastic processes—this was confirmed to us by David Donoho, who is a pioneer of the field. His prior work provides empirical evidence that points to a similar direction with the help of the ICA machinery developed by Cardoso [4].

7. REFERENCES

[1] D. S. Taubman and M. W. Marcellin, *JPEG2000: Image Compression Fundamentals, Standards and Prac-*

tice. Kluwer Academic Publishers, 2001.

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

[3] R. A. Devore, "Nonlinear approximation," *Acta Numerica*, vol. 7, pp. 51–150, 1998.

[4] J. F. Cardoso and D. L. Donoho, "Some experiments on independent component analysis of non-gaussian processes," in *Proceedings of the IEEE Signal Processing Workshop on Higher-Order Statistics*, 1999, pp. 74 –77.

[5] Y. Meyer, "Wavelets and applications," in *Lecture at CIRM Luminy Meeting*, Luminy, France, March 1992.

[6] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, pp. 607–609, 1996.

[7] M. Unser, P. D. Tafti, and Q. Sun, "A unified formulation of gaussian vs. sparse stochastic processes-part i: Continuous-domain theory," *arXiv:1108.6150v1*.

[8] M. Unser, P. D. Tafti, A. Amini, and H. Kirshner, "A unified formulation of gaussian vs. sparse stochastic processes-part ii: Discrete-domain theory," *arXiv:1108.6152v1*.

[9] A. Amini, M. Unser, and F. Marvasti, "Compressibility of deterministic and random infinite sequences," *IEEE Transactions on Signal Processing*, vol. 59, no. 11, pp. 5193–5201, November 2011.

[10] J. Pearl, "On coding and filtering stationary signals by discrete fourier transforms," *IEEE Transactions on Information Theory*, vol. 19, no. 2, pp. 229–232, March 1973.

[11] M. Unser, "On the approximation of the discrete karhunen-loeve transform for stationary processes," *Signal Processing*, vol. 7, no. 3, pp. 231–249, 1984.

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

[13] G. Samorodnitsky and M. S. Taqqu, *Stable Non-Gaussian Random Processes*. Chapman and Hall, 1994.

[14] J. V. Stone, *Independent Component Analysis*. The MIT Press, September 2004.