Stochastic Models and Techniques for Sparse Signals

Arash Amini[1] Michael Unser[1]

[1]Biomedical Imaging Group (BIG)
Institute of Microengineering
Swiss Federal Institute of Technology Lausanne (EPFL), Switzerland
{arash.amini,michael.unser}@epfl.ch

1 Introduction

Real-world data such as multimedia, biomedical, and telecommunication signals are formed of specific structures. However, these structures only determine some general properties of the data while the unknown or unpredictable parts are assumed to be random. This fact suggests that we can use stochastic models to explain real-world signals. Processes such as Gaussian white noise or Gaussian ARMA processes are well-known examples which are extensively used in modeling some components of the natural signals.

Although Gaussian models are strong tools that offer simplicity in analysis, they fall short of describing sparse or compressible structures. To overcome this issue, it is common to consider Gaussian mixture models. The drawback of this technique is that it is restricted to the discrete-time signals and cannot be generalized to continuous-time data.

The recent framework introduced in [1] shows that it is possible to generalize continuous-time Gaussian stochastic processes to non-Gaussian models with almost the same structure. The interesting point is that a large subset of the non-Gaussian part refers to compressible/sparse models based on the definition in [2]. The new family exhibits desirable properties such as being closed with respect to linear combinations or linear filtering, similar to the Gaussian models.

In this paper, we adhere to the model in [1] and present a stochastic framework for sparse and non-sparse processes. Based on the statistical information of the model, we revisit the classical denoising problem and show that the optimal MMSE estimator is achievable in some cases. Thus, when possible, we use it as the gold standard to evaluate the performance of common variational methods for recovering sparse signals. The results show that, by fitting the sparsifying penalty function to the signal statistics, we can almost achieve the MMSE performance.

2 Stochastic Model

The stochastic model discussed in this paper is depicted in Figure 1. The signal of interest, $s(x)$, is formed by applying the linear shaping operator L^{-1} to the white innovations $w(x)$. Also, the linear shift-

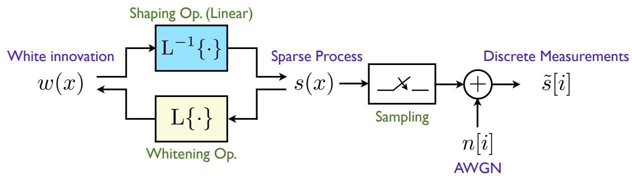

Fig. 1 The general stochastic model.

invariant operator L whitens back the process $s(x)$. In this paper, we restrict the choice of L to the differential operators of the form $\sum_{i=0}^{n} \lambda_i D^i$, where D and $D^0 = I$ represent the first-order derivative and identity operators, respectively. It is known that such nth-order differential operators are not invertible for $n \geq 1$. To define L^{-1} uniquely, we need to introduce n boundary conditions. We assume that the boundary conditions depend only on $w(x)$ for $x \leq 0$ and are set such that the operator L^{-1} is linear (not necessarily shift-invariant). A simple example would be zero boundary conditions at $x = 0$.

The model in Figure 1 is classical for Gaussian innovations. However, one of the main contributions in [1] is to show that it is possible to replace Gaussian innovations with other non-Gaussian innovations found through the Lévy-Khintchine representation theorem [3]. Note that, for many differential operators L, the inverse operator L^{-1} (after including proper boundary conditions) is unstable. This creates difficulties in showing that $L^{-1}w$ is well-defined for non-Gaussian innovations. In order to benefit from the Lévy-Khintchine representation theorem, we need to use the notion of generalized stochastic processes developed by Gelfand. In his theory, instead of the conventional point-wise definition, a random process is defined through inner products (e.g., $\langle w, \varphi\rangle$) with a space of test functions. The characteristic form defined by

$$\hat{\mathscr{P}}_w(\varphi) = \mathscr{E}\left\{e^{-j\langle w,\varphi\rangle}\right\} \quad (1)$$

plays a key role in Gelfand's theory. Now, a simplified version of the Lévy-Khintchine representation theorem reads as follows:

The function $\hat{\mathscr{P}}_w(\varphi) = exp\Big(\int_{\mathbb{R}} f\big(\varphi(x)\big)dx\Big)$ is a valid characteristic form of a stationary white process if

$$f(\omega) = jb_1\omega - \frac{b_2^2}{2}\omega^2 + \int_{\mathbb{R}\setminus\{0\}} \big(cos(a\omega) - 1\big)v(a)da, \quad (2)$$

where b_1 and b_2 are arbitrary constants and $v(\cdot)$ is a symmetric function satisfying

$$\int_{\mathbb{R}\setminus\{0\}} min(1, a^2)v(a)da < \infty. \quad (3)$$

The functions f and v are called Lévy function and Lévy density, respectively. Three special white processes characterized by (2) are:

1. Gaussian white process: By setting $v \equiv 0$, all the inner products will follow Gaussian laws.
2. Impulsive Poisson: By setting $b_1 = b_2 = 0$ and $v(a) = \lambda p_a(a)$, where p_a is a symmetric probability density function, we obtain a white process equivalent to having a random stream of Diracs such that the location of impulses obey a Poisson distribution and there are λ impulses in a unit interval on average. Furthermore, the amplitude of the impulses follow the probability law p_a. It is easy to check that this white process results in signals $s(x)$ that have a finite rate of innovation [4].
3. Symmetric α-stable: By setting $b_1 = b_2 = 0$ and $v(a) = \frac{c_\alpha}{|a|^{\alpha+1}}$, where c_α is a positive constant and $0 < \alpha < 2$, we obtain α-stable innovations. Due to their heavy-tail distributions, they are known to have compressible realizations [2].

These three examples show that the type of white innovations greatly influences the characteristics of the desired continuous-time process $s(x)$. The last thing to mention about the model is that we assume to have only a finite number of noisy/noiseless samples of $s(x)$, denoted by $\{\tilde{s}[k]\}_{k=0}^m = \{s(x = k) + n[k]\}_{k=0}^m$. The additive noise on the samples, if present, is assumed to be Gaussian.

3 Denoising

The denoising problem defined as estimating the noiseless values by observing the noisy samples, is a classical problem. To employ the statistical informations provided by the model, we need to obtain the a priori distribution of the noiseless samples. The characteristic form of the white innovations is the key to obtaining the joint distribution. However, working with the joint pdf at high dimensions is impractical, unless there is an efficient way of factorizing it.

In spline theory, it is known that the nth-order differential operator L has a discrete counterpart in form of an FIR filter which we represent by $\{d[k]\}_{k=0}^n$. Furthermore, the impulse response of the filter formed by

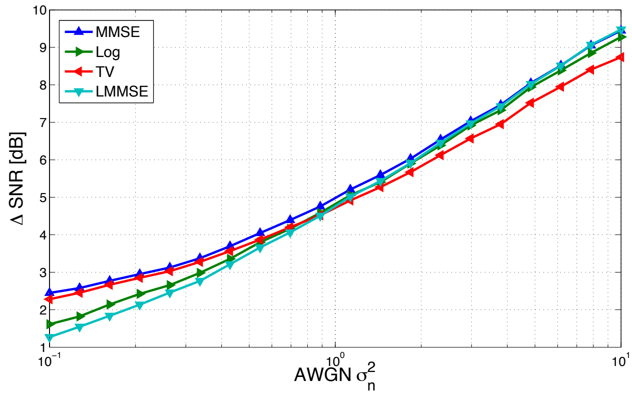

Fig. 2 SNR improvement *vs.* variance of the additive noise for Gaussian impulsive Poisson innovations. The denoising methods are: MMSE estimator, Log regularization, TV regularization, and LMMSE estimator.

composition of the discretized operator with L^{-1} results in a compact-support function, $\beta_L(x)$, which is known as the L-spline [5, 6]. Let us define

$$u[i] = (d * s)[i] = \sum_{k=0}^{n} d[k]s[i-k]. \quad (4)$$

The $u[i]$ are referred to as generalized increments and they are useful in factorizing the a priori distribution.

Theorem 1. *The joint a priori distribution factorizes as*

$$p_s\big(s[m], \dots, s[0]\big) = \prod_{\theta=2n-1}^{m} |d[0]|\; p_u\big(u[\theta]\big|\{u[\theta-i]\}_{i=1}^{n-1}\big) \times p_s(s[2n-2], \dots, s[0]), \quad (5)$$

where $p_s(s[2n-2], \dots, s[0])$ depends on the boundary conditions and the conditional probabilities are given by

$$p_u\big(u[\theta]\big|\{u[\theta-i]\}_{i=1}^{n-1}\big) = \quad (6)$$

$$\frac{\mathcal{F}_{\omega_i}^{-1}\left\{e^{I(\omega_0,\dots,\omega_{n-1})}\right\}\left(\{u[\theta-i]\}_{i=0}^{n-1}\right)}{\mathcal{F}_{\omega_i}^{-1}\left\{e^{I(0,\omega_1,\dots,\omega_{n-1})}\right\}\left(\{u[\theta-i]\}_{i=1}^{n-1}\right)}. \quad (7)$$

In the above expression, $\mathcal{F}^{-1}$ stands for the inverse Fourier operator and the multivariate function $I(\cdot)$ is defined as

$$I(\omega_0, \dots, \omega_{n-1}) = \int_{\mathbb{R}} f\left(\sum_{i=0}^{n-1} \omega_i \beta_L(x-i)\right) dx. \quad (8)$$

The factorization in Theorem 1 enables us to apply statistical denoising methods such as minimum mean-square error (MMSE). In Figures 2 and 3 we show the denoising results for samples of Lévy processes ($L = D$) with impulsive Poisson and α-stable innovations, respectively. The MMSE method is implemented using a message-passing algorithm, while the rest of the methods are variational techniques.

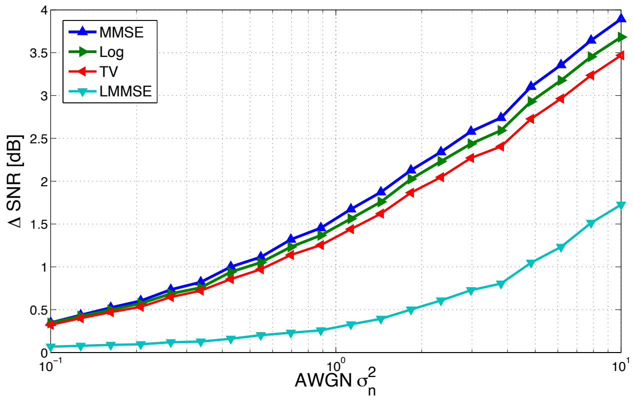

Fig. 3 SNR improvement *vs.* variance of the additive noise for Cauchy (α-stable with $\alpha = 1$) innovations. The denoising methods are: MMSE estimator, Log regularization, TV regularization (which is equivalent to MAP here), and LMMSE estimator.

- Log:

$$\{\hat{s}[k]\} = arg\min_{s[k]} \sum_{k=0}^{m} \left|\tilde{s}[k] - s[k]\right|^2 + \tau \sum_{k=1}^{m} \log\left(1 + \left|s[k] - s[k-1]\right|^2\right). \quad (9)$$

- TV:

$$\{\hat{s}[k]\} = arg\min_{s[k]} \sum_{k=0}^{m} \left|\tilde{s}[k] - s[k]\right|^2 + \tau \sum_{k=1}^{m} \left|s[k] - s[k-1]\right|. \quad (10)$$

- LMMSE:

$$\{\hat{s}[k]\} = arg\min_{s[k]} \sum_{k=0}^{m} \left|\tilde{s}[k] - s[k]\right|^2 + \tau \sum_{k=1}^{m} \left|s[k] - s[k-1]\right|^2. \quad (11)$$

The results in Figures 2 and 3 show that properly-tuned variational methods can almost achieve the MMSE performance. However, it should be noted that both the innovation statistics and additive noise power should be taken into account.

Acknowledgments

The author is supported by the FUN-SP ERC grant.

References

[1] M. Unser, P. Tafti, and Q. Sun, "A unified formulation of Gaussian vs. sparse stochastic processes: Part I—Continuous-domain theory", *arXiv:1108.6150v1*.

[2] A. Amini, M. Unser and F. Marvasti, "Compressibility of deterministic and random infinite sequences," *IEEE Trans. on Sig. Proc.*, vol. 59, no. 11, pp. 5193-5201, Nov. 2011.

[3] K. I. Sato, *Lévy Processes and Infinitely Divisible Distributions*. Chapman & Hall, 1994.

[4] M. Vetterli, P. Marziliano and T. Blu, "Sampling signals with finite rate of innovation," *IEEE Trans. Sig. Proc.*, vol. 50, no. 6, pp. 1417-1428, Jun. 2002.

[5] I. Khalidov and M. Unser, "From differential equations to the construction of new wavelet-like bases," *IEEE Trans. Sig. Proc.*, vol. 54, no. 4, pp. 1256-1267, Apr. 2006.

[6] M. Unser, P. Tafti, A. Amini, and H. Kirshner, "A unified formulation of Gaussian vs. sparse stochastic processes: Part II—Discrete-Domain theory," *arXiv:1108.6152v1*.