Not All ℓ_p-Norms Are Compatible with Sparse Stochastic Processes

Arash Amini and Michael Unser
Biomedical Imaging Group, EPFL, Lausanne, Switzerland
Email: {arash.amini and michael.unser}@epfl.ch

***Abstract*—We adopt the framework of sparse stochastic processes of [1] and investigate the sparse/compressible priors obtained by linearly measuring the processes. We show such priors are necessarily infinitely divisible. This property is satisfied by many priors used in statistical learning such as Gaussian, Laplace, and a wide range of fat-tailed distributions such as Student's-t and α-stable laws. However, it excludes some popular priors used in compressed sensing including all distributions that decay like $\exp(-\mathcal{O}(|x|^p))$ for $1 < p < 2$. This fact can be considered as evidence against the usage of ℓ_p-norms for $1 < p < 2$ in regularization techniques involving sparse priors.**

I. Introduction

In many applications, the signals of interest admit sparse/compressible representation in some transform domains. For analyzing such signals, it is befitting to establish sparse/compressible signal models. The typical example is to assume independent and identically distributed (i.i.d.) coefficients in a transform domain with Bernoulli-Gaussian law.

The framework of sparse stochastic processes introduced in [1] is an alternative modeling approach that assumes continuous-space stochastic processes. In this framework, the discrete-space signals of interest are modeled as the (generalized) samples of the continuous-space random processes. One of the advantages of this approach is that the model lends itself to the derivation of the statistics in all transform domains.

The sparse stochastic processes represent a wide spectrum of signals, from conventional Gaussian models to sparse/compressible models with fat-tailed distributions. The key ingredient is the innovation process (commonly known as white excitation noise) upon which the sparse processes are built.

In this presentation, we focus on sparse/compressible priors that arise from linearly projecting sparse stochastic processes. In particular, we characterize the rate of decay of such priors and show that none decays proportionally to $\exp(-\mathcal{O}(|x|^p))$ for $1 < p < 2$. This forbids the usage of ℓ_p-norms for $1 < p < 2$ in the MAP estimators associated with these priors.

II. Results

Theorem 1: Let $s = \mathrm{L}^{-1}w$ be a well-defined (sparse) stochastic process such that w is a white Lévy noise. Then, $\langle s, \varphi_k \rangle$ is infinitely divisible (id) for all basis functions φ_k such that $\mathrm{L}^{-1^*}\varphi_k$ satisfies some mild regularity conditions.

An id random variable X can be written as $X_1 + \cdots + X_n$ in distribution for any integer n, where $X_1, \dots, X_n$ are i.i.d. and their distribution depends on both n and X. Although

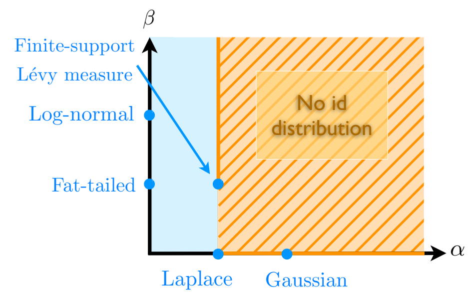

Fig. 1. Identification of id distributions with respect to their tail probabilities in the form of $\exp\big(-\mathcal{O}(|x|^\alpha(\log|x|)^\beta)\big)$. Examples include $(\alpha = 2, \beta = 0)$ for Gaussians; $(\alpha = 1, \beta = 0)$ for Laplace; $(\alpha = 0, \beta = 1)$ for all the fat-tailed laws; $(\alpha = 0, \beta = 2)$ for log-normal distributions; and $(\alpha = 1, \beta = 1)$ for all id laws with non-zero but finitely supported Lévy measures. The only id distributions in the hashed area are Gaussians.

the family of id distributions includes many sparse and fat-tailed priors (which are compressible according to [2], [3], [4]), it excludes some popular sparse priors such as Bernoulli-Gaussian [5]. Here, we reveal an interesting property of this family: a gap between the rate of decay of the Gaussian priors and that of the rest of the family.

Theorem 2: The only id distributions that decay faster than $\mathrm{e}^{-\mathcal{O}(|x|\log|x|)}$ are the Gaussians.

The implication of Theorem 2 is best understood by considering the denoising problem $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}$. The variational form of the MAP denoiser is known to be

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \ \frac{1}{2\sigma_n^2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 \underbrace{-\log p_X(\mathbf{x})}_{J(\mathbf{x})}.$$

The special case $J(\mathbf{x}) = \|\mathbf{x}\|_p^p$ requires that the elements of $\mathbf{x}$ be i.i.d. with a distribution of the form $\mathrm{e}^{-\mathcal{O}(|x|^p)}$, which is not feasible with Theorem 2 for $1 < p < 2$. The gap is depicted in Figure 1.

Acknowledgment

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

References

[1] M. Unser, P. Tafti, and Q. Sun, *arXiv:1108.6150v1 [cs.IT]*, 2011.
[2] V. Cevher, *NIPS2009*, pp. 169–176.
[3] A. Amini, M. Unser, and F. Marvasti, *IEEE TSP*, vol. 59, no. 11, pp. 5193–5201, 2011.
[4] R. Gribonval, V. Cevher, and M. E. Davies, *IEEE TIT*, vol. 58, no. 8, pp. 5016–5034, 2012.
[5] A. Amini, U. S. Kamilov, and M. Unser, *ITW2012*.