GENERATING RANDOM SIGNALS AND SPARSE AND COMPRESSIBLE VECTORS

J. Vybíral

RICAM, Austrian Academy of Sciences
Altenbergerstraße 69, A-4040 Linz, Austria
Email: jan.vybiral@oeaw.ac.at

ABSTRACT

From various reasons, it is sometimes useful both in theory and in praxis to consider random models of signals. The most popular method is the Bernoulli-Gaussian model, where each coordinate of a vector $x \in \mathbb{R}^n$ is given as a product of independent Bernoulli and Gaussian variables. We propose another model, where each coordinate is taken randomly with respect to a density $c_{p,\beta} t^\beta e^{-t^p}$, where $\beta > -1$ and $0 < p < \infty$ are real parameters. We show, that (on average) the coordinates of such a vector decay very fast. Theoretical results are also illustrated by numerical experiments.

Keywords— best m-term approximation, average widths, random sparse vectors, Bernoulli-Gaussian model

1. INTRODUCTION

It the theory of signal processing, noise is usually modelled by random vectors. For example, the *white noise* is given by $x = \varepsilon\omega \in \mathbb{R}^n$, where $\varepsilon > 0$ is a positive real number and $\omega = (\omega_1, \dots, \omega_n)$ is a vector of independent Gaussian variables. To distinguish more between the role of ε (size, or energy of x) and ω (direction of x), we may also consider the vector $\tilde{x}$ given by

$$\tilde{x}_i = \varepsilon \cdot \frac{\omega_i}{\left(\sum_{j=1}^n \omega_j^2\right)^{1/2}} = \frac{\varepsilon\omega_i}{\|\omega\|_2}, \quad i = 1, \dots, n. \tag{1}$$

Then $\varepsilon = \|\tilde{x}\|_2$ is the size of $\tilde{x}$ and $\omega/\|\omega\|_2$ is a random vector in $\mathbb{S}^{n-1}$, the unit sphere of $\mathbb{R}^n$. If also the size of the vector should be a random quantity, one needs only to replace $\varepsilon > 0$ through an appropriate random variable. Both these simple constructions of random noise turned out to be extremely successful in the theory of image processing but also in numerous real life applications.

It is the main purpose of this work to address the random generation of a structured signal. This seems to be a more delicate and more complicated task. It is nowadays a common knowledge, that structured signals usually posses a sparse (or nearly sparse) representation in a suitable bases or frame appropriately adapted to the specific class of signals. We shall therefore consider models generating random vectors in $\mathbb{R}^n$. Our wish is, that the properties of those vectors should resemble as much as possible the properties of the frame decomposition coefficients of a typical signal.

Probably the most common random model to generate sparse vectors, cf. [2, 5], is the so-called *Bernoulli-Gaussian model*. Let again $\varepsilon > 0$ be a real number and let $\omega = (\omega_1, \dots, \omega_n)$ be a vector of independent Gaussian variables. Furthermore, let $0 < p \ll 1$ be a real number and let $\varrho = (\varrho_1, \dots, \varrho_n)$ be a vector of independent Bernoulli variables

$$\varrho_i = \begin{cases} 1, & \text{with probability } p, \\ 0, & \text{with probability } 1-p. \end{cases}$$

The *Bernoulli-Gaussian vector* $x \in \mathbb{R}^n$ is then given by

$$x_i = \varepsilon \varrho_i \cdot \omega_i, \quad i = 1, \dots, n. \tag{2}$$

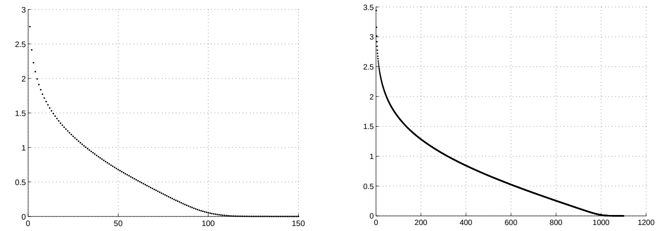
Let us point out, that this construction could be again rephrased (i.e. renormalized) in the sense of (1). Obviously, the number of non-zero elements of x is almost surely given by $\sum_{j=1}^n \varrho_j$. The estimated value of this expression is $k := pn$ and using Hoeffding's inequality we observe, that

$$\mathbb{P}\left(\left|\sum_{j=1}^n \varrho_j - k\right| > sk\right) \le 2\exp\left(-\frac{2s^2k^2}{n}\right)$$

for arbitrary $s > 0$. It follows, that if k is small (i.e. $k^2 \ll n$) then the concentration of the number of non-zero elements of x around k is not very strong. Unfortunately, if k gets larger, the effects of the theory of *concentration of measure* [6] come into play and the random vectors generated by (2) resemble more and more the vectors of n-dimensional white noise restricted to the coordinates, where $\varrho_i = 1$. In this sense, (2) represents rather a *randomly filtered white noise* then a structured signal. Especially, the first (let us say) $k/2$ largest coordinates of x are approximately of the same order, cf. Figure 1.

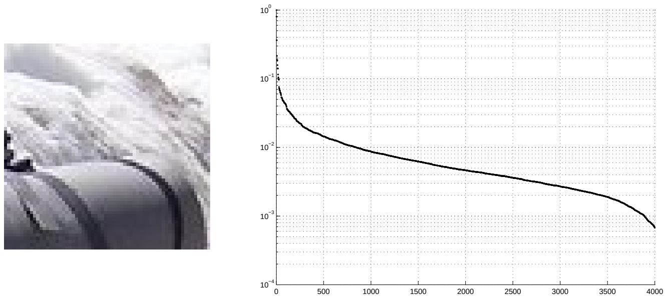
As shown in a one simple example in Figure 2, the typical signal exhibits a different behaviour. Namely, it is rather the *decay* of $x^* = (x_1^*, \dots, x_n^*)$ (the non increasing rearrangement of x) then its *sparsity*, what characterizes a structured signal. Therefore, we are looking for a random model, where a typical vector would reproduce this effect. To be able to formulate this idea more rigorously, we use the notion of best m-term approximation in a certain average setting.

The author acknowledges the financial support provided by the START-award "Sparse Approximation and Optimization in High Dimensions" of the Fonds zur Förderung der wissenschaftlichen Forschung (FWF, Austrian Science Foundation).



(a) $k = 100$ (b) $k = 1000$

Fig. 1: We generated $N = 1000$ random vectors in $\mathbb{R}^n, n = 10^6$ with respect to Bernoulli-Gaussian model with $k = 100$ and $k = 1000$. The graphs show the average of the rearrangements of their absolute values. Observe, that the ratio $x_1^*/x_{k/2}^*$ is approximately 4 or 5, respectively.



(a) Real-life picture (b) Rearranged coefficients

Fig. 2: A real-life 64x64 picture and a rearrangement of the absolute value of coefficients of its 2D discrete Fourier transform. The rearranged coefficients were divided by the largest one (to make the curve start in 1) to allow a comparison later on.

Before we come to that, let us sketch another interesting point of view on this subject. Sparse vectors play a crucial role in the novel and vastly growing area of *compressed sensing*. It was observed in [3], that the theory of compressed sensing applies also to vectors, which are *compressible*, i.e. $\|x\|_p$ is small for (preferably small) $0 < p < 1$. Indeed, the simple formula

$$\sup_{x:\|x\|_p\le 1} \inf\{\|x-y\|_\infty : \#\text{supp}\, y \le k\} \le k^{-1/p}$$

tells us, that such a vector x may be very well approximated by sparse vectors. This suggest in an intuitive way, that a typical vector of the unit ball of ℓ_p^n should be (almost) sparse. It was one of the objectives of [7], to show, that this is wrong for arbitrary $p > 0$, even for $p \ll 1$. Namely, it turned out, that all the measures usually considered in the (non-convex) geometrical functional analysis in connection with ℓ_p^n spaces are "bad" – a typical vector with respect to any of them does not involve much structure and corresponds rather to a noise then a signal. Therefore, we are looking for a new type of measures (cf. Definition 3), which would behave better from this point of view.

2. AVERAGE BEST M-TERM APPROXIMATION

We recall the notion of average best m-term approximation as developed in [7]. We assume, that the signs of the components of a random vector are chosen uniformly in $\{-1, +1\}$. If one wishes to obtain complex vectors, we take signs uniformly distributed in $\{z \in \mathbb{C} : |z| = 1\}$. Therefore, we restrict ourselves to vectors with positive coordinates in the following definition.

Furthermore, we prefer to split the properties of the random vector into its size (which correspond to the real parameter $\varepsilon > 0$ used already above) and its direction. We shall therefore assume, that the random vector is chosen randomly on the ℓ_p unit sphere in $\mathbb{R}^n$.

Definition 1 *Let $0 < p \le \infty$ and let $n \ge 2$ and $1 \le m \le n$ be natural numbers.*

(i) We set

$$\Delta_p^n := \begin{cases} \Big\{t \in [0,\infty)^n : \sum_{j=1}^n t_j^p = 1\Big\} \text{ if } p < \infty, \\ \Big\{t \in [0,\infty)^n : \max_{j=1,\dots,n} t_j = 1\Big\} \text{ if } p = \infty. \end{cases}$$

(ii) Let μ be a Borel probability measure on Δ_p^n. We put

$$\sigma_m^p(\mu) := \int_{\Delta_p^n} x_m^* d\mu(x).$$

The quantity $\sigma_{m-1}^p(\mu)$ were called the average best m-term width of $id : \ell_p^n \to \ell_\infty^n$ with respect to μ in [7]

The prominent role in the geometry of ℓ_p^n spaces is played by the (normalized) cone measure and the (normalized) Hausdorff measure. Let us recall their definitions and their basic properties.

Definition 2 *Let $0 < p \le \infty$ and $n \ge 2$.*

(i) Then the normalized cone measure on Δ_p^n is defined by

$$\mu_p(\mathcal{A}) = \frac{\lambda([0,1]\cdot\mathcal{A})}{\lambda([0,1]\cdot\Delta_p^n)}, \quad \mathcal{A} \subset \Delta_p^n$$

Here, $[0,1]\cdot\mathcal{A} = \{t\cdot x : 0 \le t \le 1, x \in \mathcal{A}\}$.

(ii) The normalized $n-1$ dimensional Hausdorff measure on Δ_p^n is defined by

$$\varrho_p(\mathcal{A}) = \frac{\mathcal{H}(\mathcal{A})}{\mathcal{H}(\Delta_p^n)}, \qquad \mathcal{A} \subset \Delta_p^n,$$

where $\mathcal{H}$ is the usual $n-1$ dimensional Hausdorff measure in $\mathbb{R}^n$.

Let us mention, that for $p \in \{1, 2, \infty\}$ these measures coincide. The cone measure enjoys two fundamental properties. The first one is its connection to the Lebesgue measure λ, which is described by the so-called *polar decomposition identity*, cf. [1],

$$\frac{\int_{\mathbb{R}^n_+} f(x)d\lambda(x)}{\lambda([0,1]\cdot\Delta_p^n)} = n\int_0^\infty r^{n-1}\int_{\Delta_p^n} f(rx)d\mu_p(x)dr, \qquad (3)$$

which holds for every $f\in L_1(\mathbb{R}^n_+)$.

The second property of μ_p is its description in terms of random variables. Let $p=2$ and let $\omega_1,\dots,\omega_n$ be independent normally distributed Gaussian random variables. Then

$$\varrho_p(\mathcal{A}) = \mu_p(\mathcal{A}) = \mathbb{P}\left(\frac{(|\omega_1|,\dots,|\omega_n|)}{\left(\sum_{j=1}^n \omega_j^2\right)^{1/2}}\in\mathcal{A}\right), \qquad \mathcal{A}\subset\Delta_p^n.$$

As noted in [8], this relation may be generalized to all values of p with $0<p<\infty$. Let $\omega_1,\dots,\omega_n$ be independent random variables on $\mathbb{R}_+$ each with density $c_pe^{-t^p},\ t\ge 0$ with respect to the Lebesgue measure, where $c_p=\frac{p}{\Gamma(1/p)}$ is the normalizing constant. Then, cf. [8, Lemma 1],

$$\mu_p(\mathcal{A}) = \mathbb{P}\left(\frac{(\omega_1,\dots,\omega_n)}{\left(\sum_{j=1}^n \omega_j^p\right)^{1/p}}\in\mathcal{A}\right), \qquad \mathcal{A}\subset\Delta_p^n. \quad (4)$$

Using (3) and (4), the following theorem was proven in [7].

Theorem 1 *Let $0<p\le\infty$ and let $n\ge 2$ and $1\le m\le n$ be natural numbers. Then*

$$\sigma_m^{p,\infty}(\mu_p) \lesssim \left[\frac{\log\left(\frac{en}{m}\right)}{n}\right]^{1/p} \qquad (5)$$

and

$$\sigma_1^{p,\infty}(\varrho_p) \lesssim \left[\frac{\log(en)}{n}\right]^{1/p}. \qquad (6)$$

Let us comment briefly on (5) and (6). The average value of the components of a fixed $x\in\Delta_p^n$ is obviously $n^{-1/p}$. Theorem 1 states, that the average value of the maximum component of x (taken with respect to μ_p or ϱ_p) is only slightly larger (namely $(\log(en))^{1/p}$ times larger). This effect is numerically illustrated in Figure 3 (a).

Altogether, we observe, that (unfortunately) the concept of average best m-term approximation is of a very limited use in connection with classical measures μ_p and ϱ_p. We shall see in the rest of this paper, that it becomes useful, when applied to other measures, which give more weight to vectors with a strong decay of their components.

3. TENSOR PRODUCT MEASURES

In this section, we propose a new class of measures defined on Δ_p^n through its density with respect to the cone measure μ_p. Essentially, we follow the idea, that this new measures should "promote sparsity", i.e. the density should be large, if more and more of the components of x tend to zero.

Definition 3 *Let $0<p<\infty$, $\beta>-1$ and $n\ge 2$. Then we define the probability measure $\theta_{p,\beta}$ on Δ_p^n by*

$$\frac{d\theta_{p,\beta}}{d\mu_p}(x) = C_{p,\beta}^{-1}\cdot\prod_{i=1}^n x_i^\beta, \quad x\in\Delta_p^n, \qquad (7)$$

where

$$C_{p,\beta} = \int_{\Delta_p^n}\prod_{i=1}^n x_i^\beta d\mu_p(x) \qquad (8)$$

is the normalizing constant.

Intuitively, the more the parameter $\beta>-1$ gets closer to -1, the stronger is the singularity of x_i^β near zero and the bigger role is played by the nearly sparse vectors, cf. Figure 4. The condition $\beta>-1$ ensures, that (8) is finite.

The following result (cf. [7]) is demonstrated numerically in Figure 3 (b). It shows, that (with the choice $\beta=p/n-1>-1$) the components of a typical vector x taken with respect to $\theta_{p,p/n-1}$ decay even exponentially.

Theorem 2 *Let $0<p<\infty$ and let $n\ge 2$ and $1\le m\le n$ be integers. Then for every fixed $m\in\mathbb{N}$,*

$$\frac{1}{\left(\frac{1}{p}+1\right)^m} \lesssim \liminf_{n\to\infty}\sigma_m^p(\theta_{p,p/n-1}) \qquad (9)$$

$$\le \limsup_{n\to\infty}\sigma_m^p(\theta_{p,p/n-1}) \lesssim \frac{1}{\left(\frac{1}{p}+1\right)^m} + \frac{e^{-m}}{m!},$$

where the constants do not depend on m, but may depend on p.

Obviously, $\theta_{p,0}=\mu_p$. For the cases $0>\beta>p/n-1$, there are no theoretical results up to now.

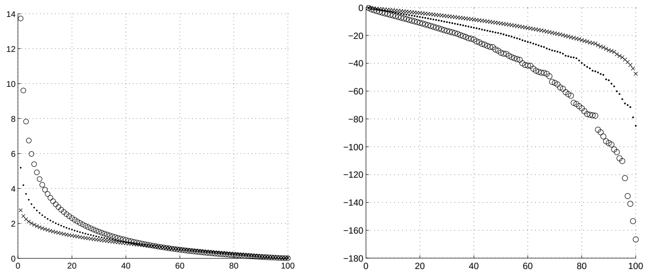
4. NUMERICAL EXPERIMENTS

4.1. Cone measure

Let us describe, how the numerical experiments were performed and implemented. We start with the cone measure. The key role is played by (4). It shows, that a random point with respect to μ_p may be generated in the following way. First, we generate $\omega_1,\dots,\omega_n$ with respect to the density $c_pe^{-t^p}, t>0$ and then calculate

$$\frac{(\omega_1,\dots,\omega_n)}{\left(\sum_{j=1}^n \omega_j^p\right)^{1/p}}\in\Delta_p^n.$$

The running time of this algorithm is linear on n. Furthermore, the values of ω_i are easy to obtain. For example the package *GNU Scientific Library* [4] implements a modification of the Marsaglia-Tsang random number generator with respect to the gamma distribution. In this way, we generated 10^8 random points $x\in\Delta_p^n$ for $n=100$ and $p\in\{0.5,1,2\}$ to approximate numerically the value of $n^{1/p}\cdot\int_{\Delta_p^n}x_m^*d\mu_p(x)$. The result may be found in the Figure 3 (a).



(a) $n^{1/p} \cdot \int_{\Delta_p^n} x_m^* d\mu_p(x)$ (b) $\log_{10}(\int_{\Delta_p^n} x_m^* d\theta_{p,p/n-1})$

Fig. 3: Approximations of $n^{1/p} \cdot \int_{\Delta_p^n} x_m^* d\mu_p(x)$ (left) and $\log_{10}(\int_{\Delta_p^n} x_m^* d\theta_{p,p/n-1})$ (right) for $n = 100$, $p = 0.5(\circ)$, $p = 1(\bullet)$ and $p = 2(\times)$ based on sampling of 10^8 random points.

4.2. Tensor measures

It was observed already in [1], that the measures $\theta_{p,\beta}$ allow a formula similar to (4). We plug the function $f(x) = \chi_{[0,\infty)\cdot A} \prod_{i=1}^n x_i^\beta e^{-\|x\|_p^p}$ into (3), where A is any μ_p-measurable subset of Δ_p^n, and obtain

$$\int_{[0,\infty)\cdot A} \prod_{i=1}^n x_i^\beta e^{-\|x\|_p^p} d\lambda(x) = \lambda([0,1]\cdot \Delta_p^n)\cdot \cdot n \cdot \int_0^\infty r^{n-1+n\beta} e^{-r^p} dr \cdot \int_A \prod_{i=1}^n x_i^\beta d\mu_p(x).$$

Furthermore, a similar formula for $A = \Delta_p^n$, leads to

$$\int_A 1 d\,\theta_{p,\beta} = \frac{\int_A \prod_{i=1}^n x_i^\beta d\mu_p(x)}{\int_{\Delta_p^n} \prod_{i=1}^n x_i^\beta d\mu_p(x)} = \frac{\int_{[0,\infty)\cdot A} \prod_{i=1}^n x_i^\beta e^{-\|x\|_p^p} dx}{\int_{\mathbb{R}_+^n} \prod_{i=1}^n x_i^\beta e^{-\|x\|_p^p} dx}.$$

To generate a random point on Δ_p^n with respect to $\theta_{p,\beta}$, we may therefore generate $\omega_1', \dots, \omega_n'$ with respect to the density $c_{p,\beta} t^\beta e^{-t^p}, t > 0$, where $c_{p,\beta}^{-1} = \int_0^\infty t^\beta e^{-t^p} dt$ is a normalizing constant and then we consider the vector

$$\frac{(\omega_1', \dots, \omega_n')}{\left(\sum_{j=1}^n (\omega_j')^p\right)^{1/p}} \in \Delta_p^n.$$

Using [4], we generated again 10^8 random points $x \in \Delta_p^n$ with respect to $\theta_{p,p/n-1}$ for $n = 100$ and $p \in \{0.5, 1, 2\}$ and used them to numerically approximate the expression $\log_{10}(\int_{\Delta_p^n} x_m^* d\theta_{p,p/n-1})$, cf. Figure 3 (b).

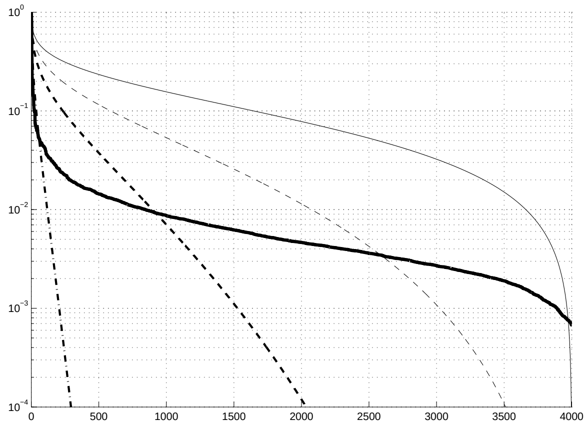


Fig. 4: Approximations of $\int_{\Delta_p^n} x_m^* d\theta_{1,\beta}$ for $n = 4000$, $p = 1$ and $\beta = 0$(——), $\beta = .3-1$ (– –), $\beta = .1-1$ (**– –**), $\beta = .01-1$ (**– · –**) based on $N = 10^4$ samples of x. The curve from Figure 2 is plotted by **——**.

5. REFERENCES

[1] F. Barthe, M. Csörnyei, and A. Naor. A note on simultaneous polar and cartesian decomposition. *Geometric Aspects of Functional Analysis, Lecture Notes in Mathematics,Springer, Berlin*, 2003.

[2] J. Bobin, J.-L. Starck, J. M. Fadili, Y. Moudden, and D. L. Donoho. Morphological component analysis: An adaptive thresholding strategy. *IEEE Trans. Image Process.*, 16 (11):2675 – 2681, 2007.

[3] A. Cohen, W. Dahmen, and R. DeVore. Compressed sensing and best k-term approximation. *J. Amer. Math. Soc.*, 22 (1):211 – 231, 2009.

[4] GNU. GNU Scientific Library, Software Package. `http://www.gnu.org/software/gsl/`.

[5] R. Gribonval, H. Rauhut, K. Schnass, and P. Vandergheynst. Atoms of all channels, unite! average case analysis of multi-channel sparse recovery using greedy algorithms. *J. Four. Anal. Appl.*, 14:655–687, 2008.

[6] M. Ledoux. *The concentration of measure phenomenon*. AMS, 2001.

[7] J. Vybíral. Average best m-term approximation. *preprint*, `http://arxiv.org/abs/1009.1751`, 2010.

[8] G. Schechtman and J. Zinn. On the volume of the intersection of two l_p^n balls. *Proc. AMS*, 110 (1):217–224, 1990.