The Analysis of Missing Samples in Signals Sparse in the Hermite Transform Domain

Miloš Brajović, *Student Member, IEEE*, Irena Orović, *Member, IEEE*, Miloš Daković, *Member, IEEE,* Srđan Stanković, *Senior Member, IEEE*

Abstract **— The influence of missing samples in signals which exhibit sparsity in the domain of Hermite transform is analyzed. The study provides theoretical concepts for the efficient reconstruction of the signals with missing samples. Single component signals are analyzed, and the main results guarantee further generalization of the presented concepts to the case of multicomponent signals. The theoretical contributions are confirmed through several numerical examples.**

Keywords **— Compressive sensing, Digital signal processing, Hermite functions, Sparse signal processing.**

I. Introduction

COMPRESSIVE sensing (CS) deals with an incomplete set of randomly selected signal samples. Reduced observation in CS is usually a consequence of sampling strategy. Moreover, the elimination of corrupted noisy signal samples using different techniques such as L-estimation can be treated as an equivalent problem. The mathematical foundation of CS lies in fact that it is possible to reconstruct a sparse signal solving an undetermined linear system of equations [1]-[8]. It is known that such systems may have infinitely many solutions, but the idea behind compressive sensing is to try to find the sparsest one [4]-[8]. One way for finding the solution is to use known optimization algorithms, based on l_0 norm minimization or l_1-norm minimization (which is used in practical applications) [3]. In several approaches, concentration measures are used as measures of signal sparsity [10]. Well known algorithms include linear programming based solvers, for example, Primal-dual interior-point methods and greedy algorithms such as Matching Pursuit (MP) or Orthogonal Matching Pursuit (OMP) which are based on iterative finding of an approximate solution [1]-[3]. Other approaches and algorithms based on signal processing techniques and noise analysis have been also proposed [4].

The main motivation for the following analysis is to derive the expression which relates the number of missing samples of the analyzed signal with the statistics of the noise induced by the incomplete set of data. The view of the CS problem from the noise perspective is already introduced in the literature [4]. The analysis is performed for the single component sparse signal. The considered sparse transform domain is the domain of the Hermite transform [9]-[15]. The results might be used for the construction of an automated algorithm for missing samples recovery [5]. The importance of the Hermite transform is confirmed through a number of different applications, including image processing, tomography, analysis of protein structure, biomedicine [11], [12], [14]. Especially interesting is the suitability of the transform for representing QRS complexes of ECG signals [9], [10].

The paper is organized as follows: theoretical background is given in the Section II. Statistical properties of the noise induced by missing samples are analyzed in the Section III. Section IV presents the amplitude analysis based on the derived statistics. Section V presents experimental results.

II. The Hermite Transform

Hermite functions are defined as [3],[1],[9]-[15]:

$$\psi_0(x)=\frac{1}{\sqrt[4]{\pi}}e^{-x^2/2},\psi_1(x)=\frac{\sqrt{2}x}{\sqrt[4]{\pi}}e^{-x^2/2},$$

$$\psi_p(x)=x\sqrt{\frac{2}{p}}\psi_{p-1}(x)-\sqrt{\frac{p-1}{p}}\psi_{p-2}(x),\forall p\geq 2. \quad (1)$$

If we introduce $c_p(x)$ as the p-th Hermite coefficient:

$$c_p=\int_{-\infty}^{\infty}f(x)\psi_p(x)dx\,. \quad (2)$$

the decomposition by using N Hermite basis functions can be defined as follows:

$$f(n)=\sum_{p=0}^{N-1}c_p\psi_p(n)$$

It will be assumed that the considered signal is of length M. Integrals of the form (2) can be numerically calculated by using Gauss-Hermite quadrature [3], [10]:

$$c_p\approx\frac{1}{M}\sum_{m=1}^{M}\mu_{M-1}^{p}(x_m)f(x_m) \quad (3)$$

with points x_m obtained as zeros of M-th order Hermite polynomial, where the M is signal length. The constants $\mu_{M-1}^{p}(x_m)$ are calculated using Hermite functions:

$$\mu_{M-1}^{p}(x_m)=\frac{\psi_p(x_m)}{(\psi_{M-1}(x_m))^2}\,.$$

It will be assumed that the Hemite expansion is performed using exactly $N=M$ Hermite functions, for

This work is supported by the Montenegrin Ministry of Science, project grant: CS-ICT "New ICT Compressive sensing based trends applied to: multimedia, biomedicine and communications".

Authors are with the University of Montenegro, Faculty of Electrical Engineering, Dzordza Vasingtona bb, 81000 Podgorica, Montenegro. Corresponding author is Miloš Brajović, email: milosb@ac.me (phone: +382 69 486 639).

signal of length M. The coefficients of the Hermite expansion (3) can be written in the following form:

$$c_p \approx \frac{1}{M}\sum_{m=1}^{M}\mu_{M-1}^{p}(x_m)s(x_m) = \frac{1}{M}\sum_{m=1}^{M}\frac{\psi_p(x_m)}{(\psi_{M-1}(x_m))^2}s(x_m).$$

Note that in order to satisfy the completeness of the representation, discrete signals have to be sampled at points equal to zeros of the M-th order Hermite polynomial [3], [15].

III. MISSING SAMPLES INFLUENCE

Consider the sparse signal of length M, analytically defined as:

$$s(x_m) = A\psi_{p_0}(x_m),$$

which can be written in the domain of the Hermite expansion coefficients in following form:

$$c_p \approx \sum_{m=1}^{M}\frac{A}{M}\frac{\psi_p(x_m)\psi_{p_i}(x_m)}{(\psi_{M-1}(x_m))^2},$$

where A is the signal amplitude and p_0 is the given order of the Hermite function corresponding to the signal. Normalized signal components are multiplied with the basis function $\mu_{M-1}^{p}(x_m)$ to produce:

$$z(x_m) = z(m) = \frac{A}{M}\frac{\psi_p(x_m)\psi_{p_0}(x_m)}{(\psi_{M-1}(x_m))^2}$$

Note that the discrete Hermite expansion satisfies the orthonormality property [13]:

$$\frac{1}{M}\sum_{m=1}^{M}\frac{\psi_p(x_m)}{(\psi_{M-1}(x_m))^2}\psi_k(x_m) = \delta(p-k). \quad (4)$$

For the sake of simplicity, index m will be further used to denote Hermite polynomial zero x_m.

Since the orhonormality property (4) holds, it is obvious that the relation among the members of a set $\mathbf{\Phi}$ containing values of $z(m)$ holds:

$$z(1) + z(2) + ... + z(M) = 0,$$

for all $p \neq p_0$.

In order to analyze the CS signal case, a subset of $M_A \leq M$ randomly positioned available samples from the set $\mathbf{\Phi}$ is considered:

$$\mathbf{\Xi} = \{z(m_1), z(m_2), ..., z(m_{M_A})\} \subset \mathbf{\Phi},$$

with $\{m_1, m_2, \ldots, m_{M_A}\}$ being the set of random positions from the set $\{1, 2, \ldots, M\}$. Since the Hermite expansion can be considered as a linear discrete transform, omitting some of the signal samples produces the same result as if these samples assume zero values [4]. A transform with a reduced number of signal samples can be considered as a transform of complete set of samples, with some of these samples affected with a noise [4]. Without loss of generality, we will assume that $A = 1$. In that case, the disturbance (noise) can be expressed as:

$$\eta(m) = \begin{cases} 0, \text{ for } m \in \{m_1, m_2, \ldots, m_{M_A}\} \\ -z(m), \text{ for } m \notin \{m_1, m_2, \ldots, m_{M_A}\}. \end{cases}$$

The Hermite expansion over the available set of samples from $\mathbf{\Xi}$ can be written in the following form:

$$Z_p = \sum_{i=1}^{M_A} z(m_i) = \sum_{m=1}^{M}[z(m) + \eta(m)].$$

It is a random variable, formed as a sum of M_A randomly positioned samples.

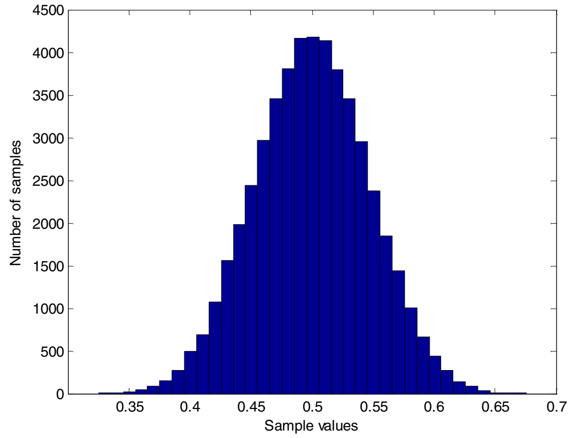

Figure 1: Histogram of the value $Z_{p=135}$, calculated based on 50000 realizations of signal with random missing samples positions, for M_A = 100 and M = 200.

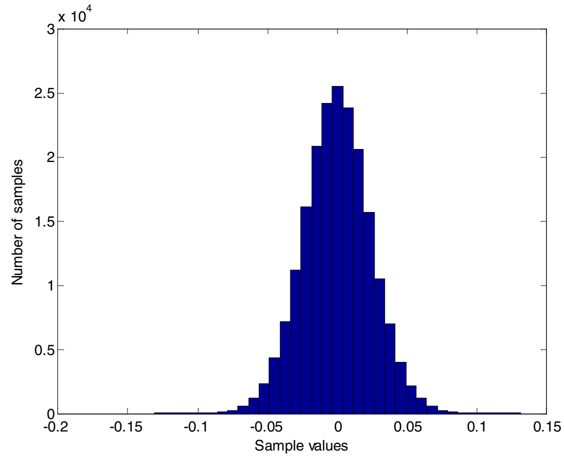

Figure 2: Histogram of the random variable $Z_{p \neq p_0}$ calculated based on 500 realizations of signal with random missing samples positions

A. Mean value and variance at the signal position

First, we will analyze the statistics of the considered random variable, i.e. Hermite expansion coefficients of the CS signal, for the case when $p = p_0$. Note that $Z_{p=p_0}$ is a random variable, with normal distribution, according to the Central limit theorem. This statement will be confirmed through a numerical example. From (4) it obviously holds:

$$E\{z(1) + z(2) + ... + z(M)\} = 1.$$

In the case of $z(m)$ with M_A available randomly positioned samples from the set $\mathbf{\Phi}$, the expected (mean) value of the random variable $Z_{p=p_0}$ follows:

$$\mu_{zs} = E\{z(m_1) + z(m_2) + ... + z(m_{M_A})\} = M_A / M \quad (5)$$

Having in mind that the mean value is not zero, variance can be calculated with the following approximation:

$$\begin{aligned} \text{var}\{Z_{p=p_0}\} &= E\{|Z_{p=p_0} - \mu_{zs}|^2\} = E\{|Z_{p=p_0}|^2\} - |\mu_{zs}|^2 \\ &\approx \log\left(\frac{M}{3}\right)\frac{M_A M - M_A^2}{M^2(M-1)}. \end{aligned} \quad (6)$$

The variance (6) is obtained heuristically and it is evaluated through numerical experiments. When $A \neq 1$ the variance (6) is multiplied with A^2.

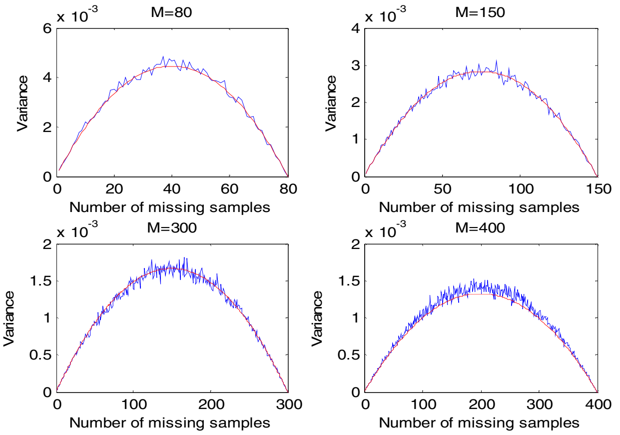

Figure 3: Variance dependence on the number of available samples for the variable $Z_{p=p_0}$: solid line – experimental results, dashed line – theoretical result

B. Mean value and variance at non-signal positions

For $p \neq p_0$ the random variable $Z_{p \neq p_0}$ corresponds to a new additive noise [4]. Due to the orthonormality properties the random variable $Z_{p \neq p_0}$ has a zero mean value, i.e.:

$$\mu_{zn} = E\{Z(p \neq p_0)\} = 0. \quad (7)$$

It can be easily shown that the variance at non-signal position can be calculated starting from:

$$\operatorname{var}\{Z_{p \neq p_0}\} = E\{(z(m_1) + z(m_2) + ... + z(m_{M_A}))^2\}$$

and has the following form:

$$\operatorname{var}\{Z_{p \neq p_0}\} = \frac{M_A M - M_A^2}{M^2(M-1)}. \quad (8)$$

IV. Amplitude Analysis in the Domain of the Hermite Expansion

According to the central limit theorem, both random variables $Z_{p=p_0}$ and $Z_{p \neq p_0}$ have Gaussian distribution, i.e. behave as Gaussian variables with different mean values and variances, as shown in previous section. A ratio of the value of the coefficients of the Hermite expansion at the signal position $p = p_0$ and the values at the other positions which are a consequence of the described noise can be found. This ratio can be used as a measure of the wrong signal detection.

In order to simplify our analysis, we will make a rough assumption that the random variable $|Z_{p=p_0}|$ is deterministic, and equal to its mean value with subtracted its one standard deviation. The signal will be miss-detected if $|Z_{p=p_0}|$ is smaller than $\mu_s - \sigma_s$ (note that $|Z_{p=p_0}|$ is higher than $\mu_s - \sigma_s$ with high probability) and/or noisy random variable $|Z_{p \neq p_0}|$ is larger than $2\sigma_N$ (which occurs with probability of about 0.9545). The ratio:

$$R_{95} = \frac{Z(p \neq p_0)}{Z(p = p_0)} < \frac{2\sigma_N}{\mu - \sigma_s} \approx \frac{2}{\sqrt{\frac{M_A(M-1)}{M - M_A}} - \sqrt{\log_{10}\left(\frac{M}{3}\right)}} \quad (9)$$

can give a range of M_A necessary for a CS algorithm to perform the reconstruction. It is obvious that for a small number of available samples, i.e. a large number of missing samples this ratio is large. It means that the values $Z_{p \neq p_0}$ caused by missing samples are large.

Equivalently, the R_{99} ratio can be derived using $3\sigma_N$ instead of $2\sigma_N$ in (9) as:

$$R_{95} = \frac{Z(p \neq p_0)}{Z(p = p_0)} < \frac{3\sigma_N}{\mu - \sigma_s} \approx \frac{3}{\sqrt{\frac{M_A(M-1)}{M - M_A}} - \sqrt{\log_{10}\left(\frac{M}{3}\right)}}, \quad (10)$$

which guarantees, with probability of about 0.99, that the signal component will be detected. Note that these ratios are only illustrative approximations, and our further research will lead to more precise formulations of these probabilistic expressions.

V. Examples

Example 1: In order to illustrate statistics of $Z_{p=p_0}$, consider a one component discrete signal $s(m) = \psi_{135}(m)$, for $m = 1,...,200$, i.e. $M = 200$. The number of randomly positioned available samples is $M_A = 100$. The histogram of the value $Z_{p=135}$ calculated based on 50000 independent realizations of signal with random missing samples positions is shown on Fig. 1. Note that the expected value is $\mu_{zs} = M_A / M = 0.5$, according to (5), which obviously corresponds to the largest number of samples in the histogram. The histogram shows theoretically expected normal distribution of the random variable $Z_{p=135}$.

Example 2: Consider the one component discrete signal $s(m) = \psi_{140}(m)$, for $m = 1,...,M$, where $M = 400$, with missing samples. The number of randomly positioned available samples is $M_A = 150$. Histogram of the random variable $Z_{p \neq 150}$ calculated based on 500 independent realizations of signal with random missing samples (i.e. 400×500 samples) is shown on the Fig. 2. The zero mean value corresponds to the highest value of the histogram (largest number of samples), i.e. it is expected value. It can be noted that the random variable has normal distribution, as it is expected according to the Central limit theorem.

Example 3: Dependence of the variance (6) on the number of available samples M_A, is numerically tested, for four different values of signal lengths M: 80, 150, 300 and 400. Consider a one component discrete signal $s(m) = \psi_{50}(m)$, for $m = 1,...,80$. The number of randomly positioned available samples is varied between 1 and M with step 1. For each value of M_A the experiment is repeated 1000 times, and the variance value is calculated for the every current M_A. Note that the number of samples of the original signal, M, and the order of the Hermite function is fixed. The results are shown on the Fig. 3 along with theoretical curve calculated based on (6). The same experiment is performed for $M = 150$, $M = 300$ and $M = 400$, with $p_0 = 100$. Results are shown on Fig.3.

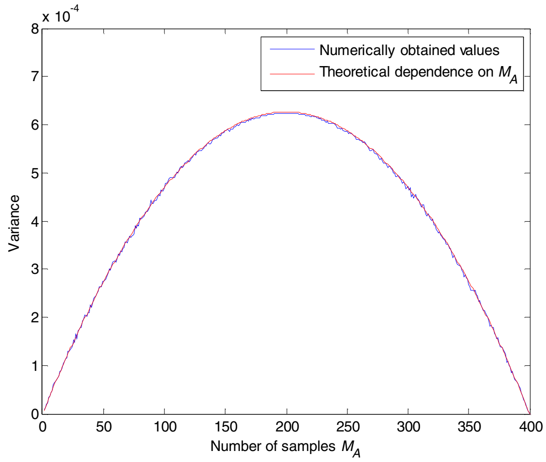

Figure 4: Variance of the random variable $Z_{p \neq p_0}$ as a function of the number of missing samples (blue line), and the theoretical value (red line).

Strictly speaking, the variance is also a function of the order of the Hermite function p_0. The strict mathematical foundation of this fact is a topic of our further research.

Example 4: The variance at non-signal position is experimentally tested. One component signal $s(m) = \psi_{200}(m)$, $m = 1,...,M$ is considered. Signal length is $M = 400$. Variable parameter is the number of available samples M_A, and it is varied between 1 and 400. For each value of M_A, the experiment is repeated 1000 times, and the variance is calculated. Obtained numerical results, along with the theoretical dependence on M_A of the variance are shown on the Fig. 4.

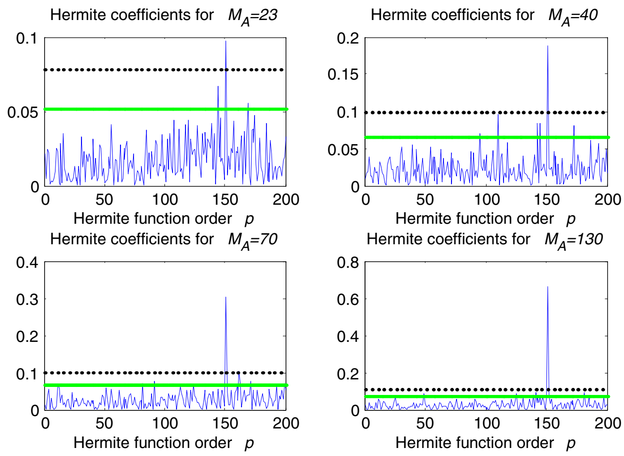

Figure 5: Hermite transform of mono-component signal with missing samples along with the theoretical threshold levels $R_{95} \mid Z_{p=p_0} \mid$ (solid line) and $R_{99} \mid Z_{p=p_0} \mid$ (dotted line)

Example 5: Mono-component signal $s(m) = \psi_{150}(m)$ with length M = 200 is considered. Number of available samples was varied, and four cases are considered: $M_A = 22$, $M_A = 40$, $M_A = 70$ and $M_A = 130$. Hermite expansion coefficients are shown on Fig. 5. The reference levels $R_{95} \mid Z_{p=p_0} \mid$ and $R_{99} \mid Z_{p=p_0} \mid$ are calculated using the mean value of the random variable $\mid Z_{p=p_0} \mid$ and they are shown on Fig. 5, solid and dotted line, respectively. Threshold level $R_{99} \mid Z_{p=p_0} \mid$ enables the detection with the probability of about 0.99, while the threshold $R_{95} \mid Z_{p=p_0} \mid$ guaranties the detection probability of about 0.95. Note that the thresholds adapt according to the given number of available samples. The increase of the number of available samples decreases the noise level at non-signal positions. It is important to emphasize that the thresholds are calculated based on the rough assumption about deterministic nature of the random variable $Z_{p=p_0}$.

VI. Conclusion

The influence of missing samples of mono-component signals sparse in the Hermite transform domain is analyzed form the perspective of noise in the transform domain. The analysis provides statistical properties of Hermite coefficients at signal and non-signal positions. The R_{95} and R_{99} ratios have been approximated. Numerical study is provided through several suitable examples. The generalization of this analysis to the multi-component signals as well as more precise derivations of statistical parameters is the main topics of further research.

Refferences

[1] E. Candès, J. Romberg, T. Tao: "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information, " *IEEE Trans. on Information Theory*, 2006, 52, (2), pp. 489–509.

[2] D. Donoho: "Compressed sensing," *IEEE Trans. on Information Theory*, 2006, 52, (4), pp. 1289 – 1306

[3] S. Stanković, I. Orović, and E. Sejdić, *Multimedia signals and systems*, Springer - Verlag, 2012

[4] LJ. Stanković, S. Stanković, and M. Amin, "Missing Samples Analysis in Signals for Applications to L-estimation and Compressive Sensing, " *Signal Processing*, vol. 94, Jan 2014, pp. 401-408

[5] S. Stanković, I. Orović, and LJ. Stanković, "An Automated Signal Reconstruction Method based on Analysis of Compressive Sensed Signals in Noisy Environment, " *Signal Processing*, vol. 104, Nov 2014, pp. 43 - 50, 2014

[6] LJ. Stanković, I. Orović, S. Stanković, M. Amin, "Compressive Sensing Based Separation of Non-Stationary and Stationary Signals Overlapping in Time-Frequency," *IEEE Trans. on Signal Processing*, Vol. 61, no. 18, pp. 4562-4572, Sept. , 2013

[7] S. Stanković, I. Orović, LJ. Stanković, A. Draganić, "Single-Iteration Algorithm for Compressive Sensing Reconstruction," *Telfor Journal* , Vol. 6, No. 1, pp. 36-41, 2014

[8] LJ. Stanković, M. Daković, and T. Thayaparan, *Time-Frequency Signal Analysis with Applications,* Artech House, Boston, 2013

[9] A. Sandryhaila, S. Saba, M. Puschel, J. Kovacevic, "Efficient compression of QRS complexes using Hermite expansion, " *IEEE Tran. on Signal Proccessing,* 2012, vol.60, no.2, pp.947-955,

[10] M. Brajović, I. Orović, M. Daković, and S. Stanković, "Gradient-based signal reconstruction algorithm in the Hermite transform domain, " *Electronics Letters, in print*, 2015

[11] J.-B. Martens, "The Hermite transform—Theory, " *IEEE Trans. Acoust., Speech, Sig. Proc.*., vol. 38, no. 9, pp. 1595–1605, 1990.

[12] G. Leibon, D. N. Rockmore, W. Park, R. Taintor, G. S. Chirikjian, "A fast Hermite transform, " *Theoretical Computer Science*, Volume 409, Issue 2, 17 December 2008, Pages 211-228

[13] P. Lazaridis, G. Debarge, and P. Gallion, "Discrete orthogonal Gauss–Hermite transform for optical pulse propagation analysis," J. *Opt. Soc. Am. B 20*, 1508-1513 (2003)

[14] A I Rasiah, R. Togneri, Y. Attikiouzel, "Modelling 1-D signals using Hermite basis functions," *Vision, Image and Signal Processing, IEE Proceedings*, vol.144, no.6, pp.345,354, Dec 1997

[15] S. Thangavelu, *Lectures on Hermite and Laguerre Expansions*, Princeton, N.J.: Princeton University Press, 1993.