Another Useful Theorem for Non-Linear Transformations of Gaussian Random Variables

Paolo Banelli, Member, IEEE

Abstract

This paper presents a useful theorem for non-linear transformations of the sum of independent, zero-mean, Gaussian random variables. It is proved that the linear regression coefficient of the non-linear transformation output with respect to the overall input is identical to the linear regression coefficient with respect to any Gaussian random variable that is part of the input. As a side-result, the theorem is useful to simplify the computation of the partial regression coefficient also for non-linear transformations of Gaussian-mixtures. Due to its generality, and the wide use of Gaussians, and Gaussian-mixtures, to statistically model several phenomena, the potential use of the theorem spans multiple disciplines and applications, including communication systems, as well as estimation and information theory. In this view, the paper highlights how the theorem can be exploited to facilitate the derivation of fundamentals performance limits such as the SNR, the MSE and the mutual information in additive non-Gaussian (possibly non-linear) channels.

Index Terms

Gaussian random variables, Gaussian-mixtures, non-linearity, linear regression, SNR, MSE, mutual information.

I. INTRODUCTION

Non-linear transformations of Gaussian random variables, and processes, is a classical subject of probability theory, with particular emphasis in communication systems. Several results are available in the literature to statistically characterize the non-linear transformation output, for both real [1], [2], [3], [4], [5], [8], [9], [11] and complex [14], [15], [16] Gaussian-distributed input processes.

The author is with the Department of Electronic and Information Engineering, University of Perugia, 06125 Perugia, Italy (e-mail: paolo.banelli@diei.unipg.it).
If the input to the non-linear transformation is the sum of two, or more, Gaussian random variables, then the overall input is still Gaussian and, consequently, the statistical characterization can still exploit the wide classical literature on the subject. For instance, a key point is to establish the equivalent input-output linear-gain (or linear regression coefficient) of the non-linearity. Anyway, if the interest is to infer only a part of the input by the overall output, and to establish a partial regression coefficient (or linear-gain) with respect to this part of the input, it is necessary to compute multiple-folded integrals involving the non-linear transformation. This task is in general tedious and, sometimes, also prohibitive. This paper proves that, if the non-linear transformation input is the sum of zero-mean, independent, Gaussian random variables, all the partial regression coefficients are identical, and equal to the overall input-output regression coefficient. Thus, the theorem highly simplifies the computation of the partial regression coefficient, which can be performed by a single-folded integral over the Gaussian probability density function (pdf) of the overall input.

To the best of the author knowledge, the theorem is new, or at least well hidden in the technical literature. Due to its potential usefulness in several disciplines, it deserves to be highlighted to the scientific community, which is the major scope of this paper. As a valuable side-product, the theorem lets to simplify the computation of the partial linear-gain, also when the non-linearity input is the sum of Gaussian-mixtures [25]. Gaussian-mixtures are widely used in multiple disciplines, such as to model electromagnetic interference [17], images background noise [19], financial assets returns [20], and, more generally, to statistically model clustered data sets. Actually, it is the similarity of the theoretical results for suboptimal estimators of Gaussian sources impaired by a Gaussian-mixture (impulsive) noise in [21], with those of non-linear transformations of Gaussian random variables in [13], [16], [15], that led to conjecture the existence of the theorem, which is proved and analyzed in this paper. Throughout the paper $E\{\cdot\}$ is used for statistical expectation, interchangeably with $E_{X_1,\ldots,X_N}\{\cdot\}$, which is used, when necessary, to highlight the (joint) probability density function pdf $f_{X_1,\ldots,X_N}(\cdot)$ involved in the expectation integral.

II. Statistical Model

Let’s consider the random variable

$$Y = X + N,$$  \(1\)

which is the sum of two independent zero-mean real random variables, $X$ and $N$, distributed according to a Gaussian pdf, with variances $\sigma_X^2$ and $\sigma_N^2$, as expressed by $f_X(x) = G(x; \sigma_X^2)$, $f_N(n) = G(n; \sigma_N^2)$,
Fig. 1. The statistical model

and

\[ G(\alpha; \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\alpha^2}{2\sigma^2}}. \]  

(2)

It is well known, that also \( Y \) is Gaussian distributed [23] with \( f_Y(y) = G(y; \sigma_Y^2) \), and \( \sigma_Y^2 = \sigma_X^2 + \sigma_N^2 \). Let also assume that \( Z \) is a non-linear transformation of \( Y \), as graphically shown in Fig. 1, and summarized by

\[ Z = g(Y) = g(X + N). \]  

(3)

### III. INPUT-OUTPUT LINEAR REGRESSION

For any \( Y \) and any non-linear transformation \( g(\cdot) \), the output random variable \( Z \) can be decomposed as the sum of a scaled version of the input \( Y \) with an uncorrelated distortion term \( W_y \), as expressed by

\[ Z = g(Y) = k_y Y + W_y, \]  

(4)

where

\[ k_y = \frac{E\{ZY\}}{E\{Y^2\}} \]  

(5)

is the input-output linear gain, or regression coefficient, that grants the orthogonality between \( Y \) and \( W_y \), i.e., \( E\{YW_y\} = 0 \).

The coefficient \( k_y \) in (5), is the same coefficient that appears in the Bussgang theorem [11], which extends (4), preserving the orthogonality of the distortion, only to special random processes, such as the
Gaussian ones. Indeed, for the class of stationary Bussgang processes \cite{10}, \cite{12}, it holds true that

\[ Z(t) = k_y Y(t) + W_y(t), \]  

where

\[ k_y = \frac{R_{ZY}(0)}{R_{YY}(0)} = \frac{E\{Z(t + \tau)Y(t)\}}{E\{Y^2(t)\}}, \forall t, \forall \tau, \]  

\[ R_{ZY}(\tau) = E\{Z(t)Y(t + \tau)\} \] is the classical cross-correlation function for stationary random processes, and \( R_{W_y Y}(\tau) = 0 \). For instance, the Bussgang theorem can be exploited to characterize the power spectral density of the output of a non linearity with Gaussian input processes. This fact induced an extensive technical literature, with closed form solutions for the computation of \( k_y \) for a wide class of non liner distortions \( g(\cdot) \), as detailed in \cite{11}, \cite{12}, \cite{13}, \cite{14}, \cite{15}, \cite{16}, \cite{8}, \cite{9}, for real Gaussian inputs, and in \cite{14}, \cite{15}, \cite{16} for complex Gaussian inputs.

The computation of \( k_y \) requests to compute a single-folded integral, as expressed by

\[ k_y = \frac{1}{P_Y} E_Y\{g(Y)Y\} = \frac{1}{P_Y} \int_{-\infty}^{+\infty} yg(y)G(y; \sigma_y^2)dy, \]  

where, \( P_Y = E\{Y^2\} \), is used in the following for notation compactness. Note that, (8) is in general much easier to compute than its equivalent double-folded integral

\[ k_y = \frac{1}{P_Y} E_{XN}\{g(X + N)(X + N)\} \]
\[ = \frac{1}{P_Y} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x + n)g(x + n)f_X(x)f_N(n)dxdn. \]  

Additionally, it is also possible to express the non-linearity output as a linear regression with respect to a single input \( X \), or \( N \), as expressed by

\[ Z = g(X + N) = k_x X + W_x, \]  

\[ Z = g(X + N) = k_n N + W_n, \]  

where

\[ k_x = \frac{E\{ZX\}}{E\{X^2\}} = \frac{E_{XN}\{g(X + N)X\}}{P_X}, \]  

\[ k_n = \frac{E\{ZN\}}{E\{N^2\}} = \frac{E_{XN}\{g(X + N)N\}}{P_N}. \]
and \( E\{XW_x\} = E\{NW_n\} = 0 \). The relationship between the three regression coefficients \( k_y, k_x, \) and \( k_n, \) for generic random variables, e.g., also non Gaussian, is summarized by

\[
P_Y k_y = E_{XN}\{g(X + N)(X + N)\} = E_{XN}\{g(X + N)X\} + E_{XN}\{g(X + N)N\} = P_X k_x + P_N k_n,
\]

which highlights that the linear gain of the overall input is a weighted sum of the linear gains of each input component, as expressed by

\[
k_y = \frac{P_X}{P_X + P_N + 2E\{XN\}} k_x + \frac{P_N}{P_X + P_N + 2E\{XN\}} k_n. \tag{14}
\]

Note that, for special cases when \( k_x = k_n, \) and \( X, N \) are orthogonal (i.e., \( E\{XN\} = 0 \)), then (14) induces also \( k_y = k_x = k_n. \)

### IV. EQUAL-GAIN THEOREMS

A case when \( k_y = k_x = k_n, \) for any non-linear transformation \( g(\cdot) \), is summarized by the following Theorem 1 and Lemmas:

**Theorem 1**: If \( X \) and \( N \) are two independent zero-mean Gaussian random variables, \( Y = X + N, \) and \( g(\cdot) \) any non-linear single-valued regular function, then

\[
\frac{E\{ZY\}}{\sigma_Y^2} = \frac{E\{ZX\}}{\sigma_X^2} = \frac{E\{ZN\}}{\sigma_N^2} = k. \tag{15}
\]

**Proof**: See Appendix.

**Lemma 1**: If \( Y = \alpha_x X + \alpha_n N, \) with \( \alpha_x, \alpha_n \in \mathbb{R}, \) then

\[
\frac{E\{ZY\}}{\sigma_Y^2} = \frac{1}{\alpha_x} \frac{E\{ZX\}}{\sigma_X^2} = \frac{1}{\alpha_n} \frac{E\{ZN\}}{\sigma_N^2}.
\]

**Proof**: By Theorem 1 with \( \tilde{X} = \alpha_x X \) and \( \tilde{N} = \alpha_n N. \)

**Lemma 2**: If \( Y = \sum_{j=1}^{J} \alpha_j X_j, \) \( \alpha_j \in \mathbb{R}, \) and \( X_j \) are independent zero-mean Gaussian random variables, then

\[
\frac{E\{ZY\}}{\sigma_Y^2} = \frac{1}{\alpha_i} \frac{E\{ZX_i\}}{\sigma_{X_i}^2}, \forall i.
\]

**Proof**: By Theorem 1 and Lemma 1 with \( X = \alpha_i X_i \) and \( N = \sum_{(j \neq i)} \alpha_j X_j. \)
In general, by equations (4), (10), and (11), it is possible to observe that,

\[
E\{W_yX\} = E\{(Z - k_y(X + N))X\} \\
= E\{ZX\} - k_yE\{X^2\} - k_yE\{NX\} \\
= k_xE\{X^2\} + E\{W_xX\} - k_yE\{X^2\} \\
= (k_x - k_y)P_X,
\]

and analogously

\[
E\{W_yN\} = (k_n - k_y)P_N.
\]  

Due to the fact that in the derivations of (16) and (17), it is only necessary to assume \( X, N \) to be orthogonal (i.e., \( E\{NX\} = 0 \)), and not necessarily Gaussian, it is demonstrated the following more general theorem

**Theorem 2:** If \( X \) and \( N \) are two orthogonal random variables, \( Y = X + N \), \( g(\cdot) \) is any single-valued regular function, by the definitions (4), (10), and (11)

\[
E\{W_yX\} = E\{W_yN\} = 0 \quad \text{iff} \quad k_y = k_x = k_n.
\]

The property \( E\{W_yX\} = E\{W_yN\} = 0 \) in Theorem 2, highlights the key element that distinguishes independent zero-mean Gaussian random inputs, with respect to the general situation, when \( X \) and \( N \) are characterized by arbitrary pdf's. Indeed, for zero-mean Gaussian inputs, by means of Theorem 1 and the sufficient condition in Theorem 2, the distortion term \( W_y \) is orthogonal to both the input components \( X \) and \( N \), while in general it is orthogonal only to their sum \( Y = X + N \). This means that, in the general case, it is only possible to state that

\[
E\{W_yX\} = -E\{W_yN\} \neq 0,
\]

which is equivalent to link the tree linear gains by (14), rather than by (15), which is just a special case. Another special case is summarized in the following

**Theorem 3:** If \( X \) and \( N \) are two independent zero-mean random variables with identical probability density functions \( f_X(\cdot) = f_N(\cdot) \), \( Y = X + N \), \( g(\cdot) \) is any single-valued regular function, then in (4), (10), and (11)

\[
k_y = k_x = k_n.
\]
Proof: By observing the definitions of \(k_x\) and \(k_n\) in (12) and (13), it is straightforward to conclude that \(k_x = k_n\), when \(f_X(\cdot)\) is identical to \(f_N(\cdot)\) (note that also \(\sigma_X^2 = \sigma_N^2\)) and, consequently, due to \(E\{XN\} = E\{X\}E\{N\} = 0\), (20) follows from (14).

V. A SIMPLE INTERPRETATION

An intuitive interpretation of the cases summarized by Theorem 1, Theorem 2, and Theorem 3 is that the non-linear function \(g(\cdot)\) statistically handles each input component in the same way, in the sense that it does not privilege or penalize any of the two, with respect to the uncorrelated distortion. In order to clarify this intuitive statement, let’s assume that \(X\) and \(N\) are zero-mean and uncorrelated, i.e., \(E\{XN\} = 0\), \(g(\cdot)\) is an odd function, i.e., \(g(y) = g(-y)\), and that the goal is to linearly infer either \(X\), or \(N\), or their sum \(Y = X + N\), from the observation \(Z\). Obviously, in this simplified set-up, also \(Z\) is zero-mean, and consequently the best (in the MMSE sense) linear estimators of, \(X\), \(N\), and \(Y\) are expressed by

\[
\hat{X}(Z) = \frac{\sigma_X}{\sigma_Z} \rho_{XZ} Z = k_x \frac{\sigma_X^2}{\sigma_Z^2} Z, \quad (21)
\]

\[
\hat{N}(Z) = \frac{\sigma_N}{\sigma_Z} \rho_{NZ} Z = k_n \frac{\sigma_N^2}{\sigma_Z^2} Z, \quad (22)
\]

\[
\hat{Y}(Z) = \frac{\sigma_Y}{\sigma_Z} \rho_{YZ} Z = k_y \frac{\sigma_X^2 + \sigma_N^2}{\sigma_Z^2} Z = \hat{X}(Z) + \hat{N}(Z), \quad (23)
\]

where \(\rho_{XZ} = E\{XZ\}/\sigma_X \sigma_Z\), \(\rho_{NZ}\), and \(\rho_{YZ}\) are the cross-correlation coefficients for zero-mean random variables. Note that, as well known [24], the equality \(\hat{Y}(Z) = \hat{X}(Z) + \hat{N}(Z)\) in (23) holds true also when \(k_y \neq k_x \neq k_n\). Equations (21)–(23) highlight that, if the two zero-mean inputs \(X\) and \(N\) equally contribute to the input in the average power sense, i.e., when \(\sigma_X^2 = \sigma_N^2\), and their non-Gaussian, and non-identical \(pdfs\) \(f_X(x)\), and \(f_N(n)\), induce \(k_x > k_n\) (or \(k_x < k_n\)), then \(X\) (or \(N\)) appears less undistorted in the output \(Z\) and, consequently, it gives an higher contribution to the estimation of the sum, by \(\hat{X}\) (or \(\hat{N}\)).

VI. COUNTER EXAMPLE AND CONJECTURES

This section describes a possible way to test if the property in (20) may hold true, or not, with respect to a wider class of \(pdfs\). To this end, let’s assume that \(X\) is still distributed as a Gaussian \(pdf\), while \(N\) is a zero-mean Gaussian-mixture, as expressed by

\[
f_N(n) = \sum_{l=0}^{L} \beta_l f_l(n) = \sum_{l=0}^{L} \frac{\beta_l}{\sqrt{2\pi\sigma_{N,l}^2}} e^{-\frac{n^2}{2\sigma_{N,l}^2}}, \quad (24)
\]
where $\sigma^2_N = \sum_{l=0}^{L} \beta_l \sigma^2_{N,l}$ is the variance, and $\sum_{l=0}^{L} \beta_l = 1$, i.e., $\beta_l \geq 0$ are the probability-masses associated to a discrete random variable, in order to grant that $f_N(n)$ is a proper pdf with unit area.

This scenario models a wide class of symmetric, zero-mean random variables, and represents a way to control how much $N$ departs, from a Gaussian distribution, depending on the choice of $L$ and $\beta_l$. For instance, this quite general framework includes an impulsive noise $N$, characterized by the Middleton’s Class-A canonical model [17], where $L = \infty$, $\beta_l = e^{-A/\Gamma}l!$ are Poisson-distributed weights, $\sigma^2_{N,l} = l/A + \Gamma + \sigma^2_N$, with $A$ and $\Gamma$ the canonical parameters that control the impulsiveness of the noise [18]. Conversely, observe that when $L = 0$, and $\beta_0 = 1$, the hypotheses of Theorem 1 hold true, and consequently (20) is verified.

If $X$ and $N$ are independent, $Y = X + N$ is also distributed as a Gaussian-mixture, as expressed by

$$f_Y(y) = f_N(y) * f_X(y) = \sum_{l=0}^{L} \beta_l[f_l(y)*f_X(y)] = \sum_{l=0}^{L} \beta_l G(Y; \sigma^2_{Y,l}),$$

(25)
due to the fact that the convolution of two zero-mean Gaussian functions, still produces a zero-mean Gaussian function, with variance equal to $\sigma^2_{Y,l} = \sigma^2_X + \sigma^2_{N,l}$. Thus, the linear regression coefficient $k_y$ can be expressed by

$$k_y = \frac{E_Y\{g(Y)Y\}}{\sigma^2_Y} = \frac{1}{\sigma^2_Y} \sum_{l=0}^{L} \beta_l E_Y\{g(Y)Y\},$$

(26)

where $Y_l = X + N_l$ stands for the $l$-th “virtual” Gaussian random variable that is possible to associate to the $l$-th Gaussian pdf in (25). Equation (26) suggests that, in this case, $k_y$ can be interpreted as a weighted sum of other $L + 1$ regression coefficients

$$k^{(l)}_y = \frac{E_Y\{g(Y)Y_l\}}{\sigma^2_{Y,l}},$$

(27)
as expressed by

$$k_y = \sum_{l=0}^{L} \frac{\sigma^2_{Y,l}}{\sigma^2_Y} \beta_l k^{(l)}_y.$$  

(28)

Each gain $k^{(l)}_y$ in (28), is associated to the virtual output $Z_l = g(Y_l)$, generated by the non-linearity $g(\cdot)$, when it is applied to the Gaussian-distributed virtual input $Y_l$. Analogously

$$k_X = \frac{1}{\sigma^2_X} E_XN\{g(X + N)X\} = \sum_{l=0}^{L} \beta_l k^{(l)}_X,$$

(29)

where

$$k^{(l)}_X = \frac{E_XN_l\{g(X + N_l)X\}}{\sigma^2_X}.$$  

(30)
Due to the fact that $X$, $N_l$, and $Y_l = X + N_l$, satisfy the hypotheses of Theorem 1, it is possible to conclude that

$$k_x^{(l)} = k_y^{(l)},$$  (31)

which plugged in (28) leads to

$$k_y = \sum_{l=0}^{L} \frac{\sigma^2_Y}{\sigma^2_Y} \beta_l k_x^{(l)}.  \quad (32)$$

By direct inspection of (32) and (29), it is possible to conclude that $k_y \neq k_x$, as soon as $L > 0$, for any value of the weights $\beta_l$, and any non-linear function $g(\cdot)$. Thus, it is reasonable the following

**Conjecture 1:** If $X \sim \mathcal{N}(0, \sigma^2_X)$, and $N$ is an independent random variable, $Y = X + N$, $g(\cdot)$ is any single-valued regular function, in the definitions (4), (10), and (11)

$$k_y = k_x = k_n \quad \text{iff} \quad N \sim \mathcal{N}(0, \sigma^2_N),$$

i.e., if and only if also $N$ is zero-mean Gaussian-distributed.

A stronger conjecture is the following

**Conjecture 2:** If $X$ and $N$ are two random variables, $Y = X + N$, $g(\cdot)$ is any single-valued regular function, in the definitions (4), (10), and (11)

$$k_y = k_x = k_n \quad \text{iff} \quad \begin{cases} N \sim \mathcal{N}(0, \sigma^2_N) \\ X \sim \mathcal{N}(0, \sigma^2_X), \ E\{XN\} = 0 \end{cases}$$

i.e., if and only if the two random variables are both zero-mean, Gaussian, and independent.

The conjectures, which as such are left without a proof, imply that the hypotheses of Theorem 1 are not only sufficient, but also necessary.

**VII. COMPUTER-AIDED SIMULATIONS**

This section reports some computer-aided simulation, to confirm the Theorems, and also to give some further strength to the Conjectures. To this end, it is considered a simple soft-limiting non-linearity, as expressed by

$$g(y) = \begin{cases} y, & |y| < y_{th} \\ y_{th}\text{sign}(y), & |y| \geq y_{th} \end{cases}.$$  (33)

The clipping threshold has been fixed as $y_{th} = 1$, and the average input power is always set to $P_Y = 10$, in order to evidence the non-linear behavior, by frequently clipping the input $Y = X + N$. Samples of
the random variables $X$ and $N$ have been generated according to either a zero-mean Gaussian (i.e., with $f(\alpha) = G(\alpha; \sigma^2)$) or to a zero-mean Laplace pdf (e.g., with analogous notation $f(\alpha) = L(\alpha; 2/\lambda^2) = 0.5\lambda e^{-\lambda|\alpha|}$). The regression coefficients $k_y, k_x$, and $k_n$ have been estimated by substituting each expected value in (5), (12) and (13), with the corresponding sample-mean over $10^6$ samples.

Fig. 2 Fig. 5 plot the linear-regression coefficients versus the mean square ratio $\rho_p = P_X/(P_X + P_N)$, which represents the power percentage of $Y = X + N$ that is absorbed by $X$, when $X$ and $N$ are independent.

Fig. 2 where the input of the soft-limiter is the sum of two independent zero-mean Gaussians, confirms Theorem 1, with all the three regression coefficients that are identical, independently of how the input power $P_Y = P_X + P_N$ is split between $X$ and $N$.

Conversely, in Fig. 3 the input is the sum of two (zero-mean) independent Laplace random variables, and $k_y \neq k_x \neq k_n$. However, when $\rho_p = 0.5$, i.e., when the input power $P_Y$ is equally split between $X$ and $N$, the three coefficients are equal, as predicted by Theorem 3.

---

Fig. 2. Linear regression coefficients versus the input power ratio, when the inputs are independent and Gaussians, i.e., $X \sim G(0, P_X)$, $N \sim G(0, P_N)$, $P_Y = 10$. 

February 2, 2012 DRAFT
Fig. 3. Linear regression coefficients versus the input power ratio, when the inputs are independent and Laplace, i.e., $X \sim L(0, P_x), N \sim L(0, P_N), P_Y = 10$.

In Fig. 4, where $X$ is zero-mean Gaussian while $N$ is an independent zero-mean Laplacian, it is clearly shown that $k_y \neq k_x \neq k_n$ for any $\rho_p$, confirming Conjecture 1. Actually, the three coefficient tend to be equal when $\rho_p \to 1$, because in this case the Gaussian $X$ is dominant, $Y = X + N$ is almost Gaussian and the situation tends to that one in Fig. 2.

In Fig. 5, differently from Fig. 2, the two Gaussian inputs $X$ and $N$ are not independent, and they are correlated with a correlation coefficient $\rho_{XN} = 0.3$. It is observed that, in this case, all the regression coefficients are different, except when $\rho_p = 0.5$, i.e., when $P_X = P_N$ and each variable absorbs a fraction equal to $(1 - 2\rho_{XN})/2$ of the total power $P_Y$. Note however that, when $P_X = P_N$, $k_y < k_x$ due to (14), which becomes $k_y = k_x/(1 + \rho_{XN})$. Additionally, it is possible to observe that $k_y$ in Fig. 5 should be equal to the value in Fig. 2 because the non-linearity in both cases has a Gaussian input $Y$, with the same power $P_Y = \sigma_Y^2 = 10$. Another interpretation of this result is the following: due to the correlation $\rho_{XN}$, it is possible to express each separate component, for instance $N$, as a function of the other one, i.e., $N = \rho_{XN}X + \varepsilon$, with $\varepsilon \sim G(0, \sigma_\varepsilon^2)$, $\varepsilon$ independent of $X$, and $\sigma_\varepsilon^2$ such that $P_Y = (1 + \rho_{XN})^2\sigma_X^2 + \sigma_\varepsilon^2$. 
Thus, for $Y = U + \varepsilon$, $U = (1 + \rho_{XN})X$ the hypotheses of Theorem 1 are satisfied and consequently $k_y = k_u = k_\varepsilon$, where by straightforward substitutions $k_u = \frac{E\{ZU\}}{P_U} = \frac{k_x}{(1 + \rho_{XN})}$.

VIII. INFORMATION AND ESTIMATION THEORETICAL IMPLICATIONS

This section is dedicated to identify a (non-exhausting) framework that is pertinent to communications, information theory, and estimation theory, where Theorem 1 can find a useful application. Actually, the model in Fig. 1 is quite common in several communication systems, where $X$ may represent the useful information, $N$ the noise or interference, and $g(\cdot)$ either a distorting non-linear device (such as an amplifier, a limiter, an analog-to-digital converter, etc.), or an estimator/detector that is supposed to contrast the detrimental effect of $N$ on $X$.

Typical parameters, to assess performance of such a non-linear communication system, are the signal-to-noise power ratio (SNR), the maximal mutual information (capacity), and the mean square estimation error (MSE), whose link has attracted several research efforts in the last decade (see [28], [29] and references therein).
Fig. 5. Linear regression coefficients versus the input power ratio, when the inputs are correlated Gaussians, i.e., $X \sim G(0, P_X)$, $N \sim G(0, P_N)$, $\rho_{XN} = 0.3$, $P_Y = 10$.

A. SNR considerations

In order to define a meaningful SNR, it is useful to separate the non-linear device output as the sum of the useful information with an uncorrelated distortion, as in (10). For simplicity, we assume in the following that all the random variables are zero-mean, i.e., $P_X = \sigma_X^2$. Thus, the SNR at the non-linearity output, is expressed by

$$SNR_x = k_x^2 \frac{E\{X^2\}}{E\{W_x^2\}} = \frac{k_x^2 \sigma_X^2}{E\{Z^2\} - k_x^2 \sigma_X^2} = \left( \frac{E_Y \{g^2(Y)\}}{k_x^2 \sigma_X^2} - 1 \right)^{-1},$$

(34)

where the second equality is granted by the orthogonality between $X$ and $W_x$.

In the general case, in order to obtain a closed form expression for (34), it would be necessary to solve the double folded integral in (12), for the computation of $k_x$. However, if $X$ and $N$ are zero-mean, independent, and Gaussian, by Theorem 1 the computation can be simplified by exploiting that $k_x = k_y$. 

February 2, 2012 DRAFT
and, consequently, the computation of the SNR would request to solve only single-folded integrals, e.g., (8) and \( E_Y\{g^2(Y)\} \). Note that, in this case also \( Y = X + N \) would be Gaussian and, consequently, the computations of these single-folded integrals can benefit of the results available in the literature for a wide class of non-linearities \( g(\cdot) \) [1], [2], [4], [8], [9], [11], [13], [15], [16].

Actually, it could be argued that the SNR may be also defined by exploiting (4) rather than (8). Indeed, by rewriting (4) as
\[
Z = g(X + N) = k_y X + k_y N + W_y
\] (35)
another SNR could be expressed as
\[
\text{SNR}_y = \frac{k_y^2 E\{X^2\}}{k_y^2 E\{N^2\} + E\{W_y^2\}} = \frac{k_y^2 \sigma_X^2}{E_Y\{g^2(Y)\} - k_y^2 \sigma_X^2}
\]
\[
= \left( \frac{E_Y\{g^2(Y)\}}{k_y^2 \sigma_X^2} - 1 \right)^{-1} \quad \text{(36)}
\]

Note that, Theorem 1 states that the two SNRs in (36) and (34) are identical if \( X \) and \( N \) are zero-mean, independent, and Gaussian. Conjecture 2, claims that this is also the only situation where it is correct to use (36) in place of (34).

When \( N \) is non-Gaussian, it is possible to approximate its continuous pdf by a Gaussian-mixture, with infinite accuracy [26]. Thus, the example in Section [VI] represents a wide class of zero-mean and symmetrical noise pdfs, summarized by the Gaussian-mixture in (24), where \( k_x \neq k_y \), and (34) should be used instead of (36). However, despite (36) cannot be used to compute the SNR, as detailed in Section [VI] Theorem 1 turns to be useful to compute \( k_x \), due to the fact that
\[
k_x = \sum_{l=0}^{L} \beta_l k_x^{(l)} \quad k_y^{(l)} = k_y \frac{E_{Y_l}\{g(Y_l)Y_l\}}{\sigma_{Y,l}^2}. \quad \text{(37)}
\]

B. MSE considerations

The definition of the error at the non-linearity output may depend on the non-linearity purpose. If the non-linearity \( g(\cdot) \) represents an estimator of \( X \) given the observation \( Y = X + N \), as expressed by
\[
\hat{X} = g(X + N) = k_x X + W_x
\] (38)
the estimation error is defined as
\[
e = \hat{X} - X = (k_x - 1) X + W_x \quad \text{(39)}
\]
Exploiting the uncorrelation between $X$ and $N$,

$$E\{W_x^2\} = E_Y \{g^2(Y)\} - k_x^2 E\{X^2\}, \quad (40)$$

the MSE at the non-linearity output is defined by

$$\text{MSE} = E\{e^2\} = (k_x - 1)^2 E\{X^2\} + E\{W_x^2\}$$

$$= (1 - 2k_x) E\{X^2\} + E\{g^2(Y)\}. \quad (41)$$

However, looking at (38) from another point of view, it is also possible to consider $g(\cdot)$ as a distorting device that scales by $k_x$ the useful information $X$, that is (41) represents the MSE of a (conditionally) biased estimator. In this view, it is possible to define an unbiased estimator $\hat{X}_u = \hat{X} / k_x$ and the associated unbiased estimation error as

$$e_u = \hat{X} / k_x - X = W_x / k_x, \quad (42)$$

whose mean square-value is expressed by

$$\text{MSE}_u = E\{e_u^2\} = E\{W_x^2\} / k_x^2$$

$$= E_Y \{g^2(Y)\} / k_x^2 - E\{X^2\}. \quad (43)$$

Thus, it is straightforward to prove that, for a given information power $E\{X^2\}$, the non-linearities that maximize the two MSE are different, as expressed by

$$g_{mmse}(\cdot) = \arg \min_{g(\cdot)} \text{MSE} = \arg \min_{g(\cdot)} \log(\text{MSE})$$

$$= \arg \min_{g(\cdot)} \left[ E\{g^2(Y)\} / k_x \right], \quad (44)$$

and

$$g_{u-mmse}(\cdot) = \arg \min_{g(\cdot)} \text{MSE}_u = \arg \min_{g(\cdot)} \left[ E\{g^2(Y)\} / k_x^2 \right]. \quad (45)$$

The first criterion corresponds to the classical Bayesian minimum MSE (MMSE) estimator, that is $g_{mmse}(Y) = E_{X|Y} \{X\}$. By means of (34) and (45), the second criterion, which is the unbiased-MMSE (UMMSE) estimator, is equivalent to the maximum-SNR (MSNR) criterion. Note that $k_x$ depends on $g(\cdot)$ by (12) and consequently, in general

$$g_{u-mmse}(\cdot) \neq \frac{g_{mmse}(\cdot)}{k_x^{(mmse)}}. \quad (46)$$

Indeed, the right-hand term in (46) is a (conditionally) unbiased estimator, but not the optimal one, because it has been obtained by first optimizing the MSE, and by successively compensating the biasing gain, while $g_{u-mmse}(Y)$ should be obtained the other way around, as expressed by (42) and (45). The two
criteria tend to be quite similar when the functional derivative \( \frac{\delta k}{\delta g(\cdot)} \approx 0 \) in the neighborhood of the optimal solution \( g_{\text{mmse}}(\cdot) \).

Actually, the MMSE and the MSNR criteria are equivalent from an information theoretic point of view only when \( g(\cdot) \) is linear [28], in which case \( g_{\text{u-mmse}}(\cdot) \) is equivalent to right-hand side of (46). For instance, this happens when \( X \) and \( N \) are both zero-mean, independent, and Gaussian as in Theorem 1, and consequently [24]

\[
\hat{X}_{\text{mmse}} = g_{\text{mmse}}(Y) = \frac{\sigma_X^2}{\sigma_X^2 + \sigma_N^2} Y = \frac{\sigma_X^2}{\sigma_X^2 + \sigma_N^2} (X + N) \tag{47}
\]

is just a scaled version of the UMMSE,

\[
\hat{X}_{\text{u-mmse}} = g_{\text{u-mmse}}(Y) = Y = X + N. \tag{48}
\]

By noting that the SNR is not influenced by a scaling coefficient, because it affects both the useful information and the noise, it is confirmed that for linear \( g(\cdot) \) the MMSE optimal solution is also MSNR optimal [28].

Conversely, when \( N \) is not Gaussian distributed, its pdf may be, or be approximated, by a Gaussian-mixture as in [24]. In this case, analogously to the consideration for the SNR computation, Theorem 1 turns to be useful to compute \( k_x \), and thus the MSE in (41), and (43), by the single-folded integrals involved in (37), rather than by the double-folded integrals in (30). The reader interested in this point, may find a deeper explanation in [22], where these considerations have been fully exploited to characterize the performance of MMSE and MSNR estimators for a Gaussian source impaired by impulsive Middleton’s Class-A noise.

C. Capacity considerations

The computation of the capacity of the non-linear information channel \( X \to Z = g(X + N) \) is in general prohibitive, due to the complicated expression for the pdf of the disturbance component in (10) or (35). Actually, when the noise is non-Gaussian it is even difficult to compute in closed form the mutual information \( I(X \to Y) \), in the absence of the non-linearity \( g(\cdot) \), and only bounds are in generale available [7]. For instance, when the noise \( N \) is the Gaussian-mixture summarized by [24], it is even difficult to compute its differential entropy \( h(N) \), which can only be bounded as suggested in [27]. In this case, the bounds in [27] can be used for a simple upper bound of the mutual information \( I(X, Z) \), which is based on the information processing inequality [30], as expressed by

\[
I(X, Z) = I(X \to Y \to Z) \leq I(X, Y). \tag{49}
\]
When $X$ is Gaussian, a simple lower-bound for $I(X, Z)$ is provided by the AWGN capacity of (10) and (35), when the disturbance is modeled as (the maximum-entropy [6]) zero-mean Gaussian distributed noise with variance equal to $E\{Z^2\} - k_x^2\sigma_X^2$ and $E\{Z^2\} - k_y^2\sigma_X^2$, respectively. Thus, exploiting (10) and (34), it is possible to conclude that

$$I(X, Z) \geq C_x^{(AWGN)} = \frac{1}{2} \log(1 + \text{SNR}_x), \quad (50)$$

while, by exploiting (35) and (36), it would be possible to conclude that

$$I(X, Z) \geq C_y^{(AWGN)} = \frac{1}{2} \log(1 + \text{SNR}_y). \quad (51)$$

By Theorem 1, the two lower-bounds are equivalent if $X$ and $N$ are zero-mean independent Gaussians. Otherwise, the correct SNR is (34) and the correct lower bound is (50). Note that, in both cases, either when $N$ is Laplace distributed and independent of $X$ (see Fig. 4), or when it is Gaussian distributed and positively correlated with $X$ (see Fig. 5), $k_x > k_y$ and consequently by (34) and (36), $C_x^{(AWGN)} > C_y^{(AWGN)}$.

It is also possible to derive a bound for the mutual information of the additive channel $Y = X + N$ by exploiting the interplay between MMSE and mutual information, as suggested in [29], where it is proved that, for non-Gaussian additive channels,

$$I(X, Y) \geq h(X) - \frac{1}{2} \log(2\pi e \text{MMSE}). \quad (52)$$

Actually, (52) can be also readily derived by the corollary of Theorem 8.6.6 in [30]. For a Gaussian source $X$, (52) simply becomes

$$I(X, Y) \geq \frac{1}{2} \log \left( \frac{\sigma_X^2}{\text{MMSE}} \right), \quad (53)$$

where by exploiting (41) it is possible to substitute

$$\text{MMSE} = \left( 1 - 2k_x^{(mmse)} \right) \sigma_X^2 + E_Y\{g_{\text{mmse}}^2(Y)\}. \quad (54)$$

Also in this case, if the noise $N$ can be modeled, or approximated, by the Gaussian-mixture in (24), Theorem 1 turns to be useful because it allows to compute the gain $k_x^{(mmse)}$ by (37), where only single-folded integrals are requested. However, it is worth to point out that in this case (37) and $E_Y\{g^2(Y)\}$ have to be computed with $g(Y) = g_{\text{mmse}}(Y)$, which is characterized by a rather involved expression (22), which prevents closed form solutions. Thus, the computation of the lower bound in (54) requests either (single-folded) numerical integration techniques, or to express $g_{\text{mmse}}(Y)$ as a series expansion by opportune functions (polynomials, hermite, etc.) that admit closed form expressions for their averages over Gaussian pdfs (see [4], [11], [15] and references therein). This is however out of the scope of this paper, and a possible subject for further investigations.
IX. CONCLUSIONS

The main contribution of this paper has been to prove and analyze a general and theoretically interesting theorem for non-linear transformations of the sum of zero-mean independent Gaussian random variables. Due to the widespread use of Gaussian random variables, the theorem can be useful in several fields, which include estimation theory, information theory, and non-linear system characterization. As a side-result, the theorem can be used to simplify the computations involved in the analysis of non-linear transformations of Gaussian-mixtures. Finally, the paper has highlighted its usefulness for the computation of the SNR, the MSE and bounds on the mutual information, associated with communication systems dealing with non-linear devices and estimators.

APPENDIX

Proof of Theorem 1: it is assumed that \( g(\cdot) \) is a regular function, in the sense it admits a power series representation, i.e., a Mc-Laurin expansion \( g(y) = \sum_{p=0}^{\infty} c_p \alpha^p \), with \( c_p = \left. \frac{g^{(p)}(\alpha)}{p!} \right|_{\alpha=0} \).

Thus, by the definition of the input-output linear-regression coefficient in (12)

\[
k_x = \frac{1}{\sigma^2_X} \int_{-\infty}^{+\infty} x \int_{-\infty}^{+\infty} g(x+n)f_N(n)dnf_X(x)dx
= \frac{1}{\sigma_X^2} \int_{-\infty}^{+\infty} x \int_{-\infty}^{+\infty} g(\alpha)f_N(\alpha-x)d\alpha f_X(x)dx
= \sum_{p=0}^{\infty} \frac{c_p}{\sigma_X^2} \int_{-\infty}^{+\infty} x m_p(x)f_X(x)dx,
\]

where \( m_p^{(\alpha)}(x) = E_{\alpha|x}\{\alpha^p\} \), is the non-central moment of order \( p \) of a Gaussian random variable with mean-value equal to \( m_{\alpha|x} = m_1^{(\alpha)}(x) = x \). The non-central moments can be computed by exploiting their well known relationship with the central moments \( \mu_k^{(\alpha)} = E\{(\alpha - m_{\alpha|x})^k\} \), expressed by

\[
m_p^{(\alpha)}(x) = \sum_{k=0}^{p} \binom{p}{k} \mu_k^{(\alpha)} [m_1^{(\alpha)}(x)]^{p-k},
\]

and the fact that for a Gaussian random variable, with variance \( \sigma^2 \), all the odd central moments are null, as expressed by

\[
\mu_k = \begin{cases} (k-1)!! \sigma^k & \text{, } k = 2l, \ l \in \mathbb{N} \\ 0 & \text{, } k = 2l + 1, \ l \in \mathbb{N}, \end{cases}
\]

February 2, 2012 DRAFT
where \((n)!! = n \cdot (n-2) \cdot (n-4) \ldots \cdot 1\), stands for the so called double-factorial of an integer number.

Thus, equation (55) becomes

\[
k_x = \sum_{p=0}^{\infty} \frac{c_p}{\sigma_X} \sum_{l=0}^{[p/2]} \left( \frac{p}{2l} \right) (2l - 1)!! \int_{-\infty}^{\infty} x^{p-2l+1} G(0; \sigma_x^2) dx
\]

\[
= \sum_{p=0}^{\infty} \frac{c_p}{\sigma_X} \sum_{l=0}^{[p/2]} \left( \frac{p}{2l} \right) (2l - 1)!! \sigma_n^{2l} \mu_{p-2l+1}^{(x)}
\]

(56)

Due to the fact \(\mu_{p-2l+1}^{(x)} = 0\) when \(p - 2l + 1\) is odd, that is when \(p = 2\gamma\) is even, (56) highlights the well known property that for zero-mean Gaussian inputs, the linear-gain (or regression coefficient) is imposed only by the odd part of the non-linearity, because only the terms in the series with \(p = 2\gamma + 1\) are different from zero. Thus, it is possible to conclude that

\[
k_x = \sum_{\gamma=0}^{\infty} c_{2\gamma+1} \frac{\sigma_X^{2\gamma}}{\sigma_n^{2\gamma}} \sum_{l=0}^{\gamma} \left( \frac{2\gamma + 1}{2l} \right) (2l - 1)!! (2\gamma + 1 - 2l)!! \left( \frac{\sigma_n}{\sigma_X} \right)^{2l}
\]

(57)

Analogously, due to the symmetry of the problem,

\[
k_n = \sum_{\gamma=0}^{\infty} c_{2\gamma+1} \frac{\sigma_X^{2\gamma}}{\sigma_n^{2\gamma}} \sum_{l=0}^{\gamma} \left( \frac{2\gamma + 1}{2l} \right) (2l - 1)!! (2\gamma + 1 - 2l)!! \left( \frac{\sigma_X}{\sigma_n} \right)^{2l}
\]

(58)

In order to prove Theorem 1, it is sufficient to prove \(k_x = k_n\), which substituted in (14), leads to \(k_y = k_x = k_n\).

To this end, observe that (58) can also be written as

\[
k_n = \sum_{\gamma=0}^{\infty} c_{2\gamma+1} \frac{\sigma_X^{2\gamma}}{\sigma_n^{2\gamma}} \sum_{l=0}^{\gamma} \left( \frac{2\gamma + 1}{2l} \right) (2l - 1)!! (2\gamma + 1 - 2l)!! \left( \frac{\sigma_n}{\sigma_X} \right)^{2\gamma - 2l}
\]

(59)

and by defining \(q = \gamma - l\)

\[
k_n = \sum_{\gamma=0}^{\infty} c_{2\gamma+1} \frac{\sigma_X^{2\gamma}}{\sigma_n^{2\gamma}} \sum_{q=0}^{\gamma} \left( \frac{2\gamma + 1}{2\gamma - 2l} \right) (2\gamma - 2q - 1)!! (2q + 1)!! \left( \frac{\sigma_n}{\sigma_X} \right)^{2q}
\]

(60)

Exploiting the property \(\binom{n}{k} = \binom{n}{n-k}\), and renaming for convenience the summation index \(q\) by \(l\), (60) becomes

\[
k_n = \sum_{\gamma=0}^{\infty} c_{2\gamma+1} \frac{\sigma_X^{2\gamma}}{\sigma_n^{2\gamma}} \sum_{l=0}^{\gamma} \left( \frac{2\gamma + 1}{2l + 1} \right) (2\gamma - 2l - 1)!! (2l + 1)!! \left( \frac{\sigma_X}{\sigma_n} \right)^{\gamma}
\]

(61)

By exploiting \(n! = n!(n-1)!!\), it is straightforward to verify that

\[
\left( \frac{2\gamma + 1}{2l + 1} \right) (2l + 1)!! (2\gamma - 2l - 1)!! = \left( \frac{2\gamma + 1}{2l} \right) (2l - 1)!! (2\gamma - 2l + 1)!!
\]

(62)
Consequently, (62) coincides with (57) and, due to the fact that \( E\{XN\} = E\{X\}E\{N\} = 0 \) by hypothesis, \( k_y = k_x = k_n \), by (14).

**REFERENCES**

[1] J. J. Bussgang, “Cross-correlation function of amplitude-distorted Gaussian signals,” Research Lab. of Electronics, M.I.T., Cambridge, MA, Tech. Rep. 216, March 1952.

[2] R. F. Baum, “The correlation function of smoothly limited Gaussian noise,” *IRE Trans. Inform. Theory*, vol. 3, pp. 193-197, Sep. 1957.

[3] R. Price, “A useful theorem for nonlinear devices having Gaussian inputs,” *IRE Trans. Inform. Theory*, vol. 4, pp. 69-72, Jun. 1958.

[4] W. B. Davenport and W. L. Root, *An Introduction to the Theory of Random Signals and Noise*, New York, McGraw-Hill, 1958.

[5] N. M. Blachman, “The uncorrelated output components of a non linearity,” *IEEE Trans. Inform. Theory*, vol. 14, n. 2, pp. 250-255, Mar. 1968.

[6] S. N. Diggavi and T. M. Cover, “The worst additive noise under a covariance constraint,” *IEEE Trans. Inform. Theory*, vol. 47, n. 7, pp. 3072-3081, Nov. 2001.

[7] S. Ihara, “On the capacity of channels with additive non-gaussian noise,” *Inform. Contr.*, vol. 37, n. 1, pp. 34-39, Apr. 1978.

[8] J. H. Van Vleck and D. Middleton, “The spectrum of clipped noise,” *Proc. of IEEE*, vol. 54, pp. 2-19, Jan. 1966.

[9] R. F. Baum, “The correlation function of Gaussian noise passed through nonlinear devices,” *IEEE Trans. Inform. Theory*, vol. 15, n. 4, pp. 448-456, Jul. 1969.

[10] A. H. Nuttall, “Theory and application of the separable class of random processes,” *Research Lab. of Electronics, Tech. Rept. 343*, M.I.T., Cambridge, MA., May 1958.

[11] B. Levine, *Fondements theoriques de la radiotechnique statistique*, New York, Editions de Moscou, 1973.

[12] F. Rocca, B. Godfrey, and F. Muir, “Bussgang Processes,” Stanford Exploration Project, SEP-Report n. 16, pp. 275-280, Apr. 1979 (available at http://sepwww.stanford.edu).

[13] H. E. Rowe, “Memoryless nonlinearities with Gaussian inputs: Elementary results,” *Bell Syst. Tech. Journal*, vol. 61, n. 7, pp. 1520–1523, Sep. 1982.

[14] J. Minkoff, “The role of AM-to-PM conversion in memoryless nonlinear systems,” *IEEE Trans. Commun.*, vol. 33, pp. 139–143, Feb. 1985.

[15] P. Banelli and S. Cacopardi, “Theoretical analysis and performance of OFDM signals in nonlinear AWGN channels,” *IEEE Trans. Commun.*, vol. 48, n. 3, pp. 430–441, Mar. 2000.

[16] D. Dardari, V. Tralli, A Vaccari, “A theoretical characterization of nonlinear distortion effects in OFDM systems,” *IEEE Trans. Commun.*, vol. 48, n. 10, pp. 1755-1764, Oct. 2000.

[17] D. Middleton, “Statistical-physical models of urban radio-noise environments Part I: Foundations,” *IEEE Trans. Electromagn. Compat.*, vol. 14, pp. 38-56, May 1972.

[18] L. A. Berry, “Understanding Middleton’s canonical formula for Class A noise,” *IEEE Trans. Electromagn. Compat.*, vol. 23, n. 4, pp. 337-344, Nov. 1981.

[19] D.S. Lee, “Effective Gaussian mixture learning for video background subtraction,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 7, n. 5, pp. 827-832, May 2005.
[20] I. Buckley, D. Saunders, and L. Seco, “Portfolio optimization when asset returns have the Gaussian mixture distribution,” European Jour. of Operational Research, vol. 185, n. 3, pp. 1434-1461, 2008.

[21] S. V. Zhidkov, “Performance analysis and optimization of OFDM receiver with blanking nonlinearity in impulsive noise environment,” IEEE Trans. Veh. Technol., vol. 55, n. 1, pp. 234-242, Jan. 2006.

[22] P. Banelli, “Bayesian Estimation of Gaussian sources in Middleton’s Class-A Impulsive Noise,” submitted to IEEE Trans. Signal Process., available on arXiv:1111.6828v1 (cs.IT) http://arxiv.org/abs/1111.6828v1 Nov. 2011.

[23] A. Papoulis, Probability, Random Variables and Stochastic Processes, McGraw-Hill, New York, 3rd edition, 1991.

[24] S. M. Kay, Fundamentals of Statistical Signal Processing. Vol. 1, Estimation Theory, Prentice-Hall, Englewood Cliffs, NJ, 1993.

[25] S. V. Vaseghi, Advanced digital signal processing and noise reduction, 4th edition, John Wiley & Son’s, Chichester, UK, 2009

[26] V. Maz’ya and G. Schmidt, “On approximate approximations using gaussian kernels,” IMA J. Numer. Anal., vol. 16, pp. 13-29, 1996.

[27] M. F Huber, T. Bailey, H. Durrant-Whyte, U. D. Hanebeck, ”On entropy approximation for Gaussian mixture random vectors,” Proc. of IEEE Int. Conf. on Multis. Fusion and Integr. for Intel1. Syst. 2008, pp. 181-188, Aug. 2008.

[28] D. Guo, S. Shamai, S. Verdu, ”Mutual information and minimum mean-square error in Gaussian channels,” IEEE Trans. Inform. Theory, vol. 51, n. 4, pp. 1261-1282, Apr. 2005

[29] S. Prasad, “Certain Relations between Mutual Information and Fidelity of Statistical Estimation,” available on arXiv: arXiv:1010.1508v1, (stat.AP) http://arxiv.org/abs/1010.1508v1., Oct. 2010.

[30] T. M. Cover, J. A. Thomas, Elements of Information Theory, 2nd ed., John Wiley & Sons, Inc., 2006.