Özlem Tuğfe Demir and Emil Björnson

The Bussgang Decomposition of Nonlinear Systems

Basic theory and MIMO extensions

Many of the systems in various signal processing applications are nonlinear due to, for example, hardware impairments, such as nonlinear amplifiers and finite-resolution quantization. The Bussgang decomposition is a popular tool used when analyzing the performance of systems that involve such nonlinear components. In a nutshell, the decomposition provides an exact probabilistic relationship between the output and the input of a nonlinearity: the output is equal to a scaled version of the input plus uncorrelated distortion. The decomposition can be used to compute either exact performance results or lower bounds, where the uncorrelated distortion is treated as independent noise. This lecture note explains the basic theory, provides key examples, extends the theory to complex-valued vector signals, and clarifies some potential misconceptions.

Relevance

The origin of the decomposition is a technical report by Julian J. Bussgang in 1952 [1]. Interestingly, the decomposition is not explicitly stated in his report but, rather, was a consequence of his results. In fact, it is mainly nontrivial extensions of his results that are utilized in current research; for example, applications to complex-valued multiple-input, multiple-output (MIMO) systems are popular in the communication community. There is no standard reference that presents and proves those extended results, and it can be hard to differentiate between which results are exact and which are mere approximations. This lecture note fills these gaps.

Prerequisites

This lecture note requires basic knowledge of random variables, linear algebra, signals and systems, and estimation theory.

Problem statement

Let us consider two jointly Gaussian continuous-time stationary random processes $f(t)$ and $g(t)$. One of the processes, say $f(t)$, undergoes a nonlinear memoryless distortion represented by the function $U(\cdot)$. The resulting non-Gaussian random process is $F(t) = U(f(t))$. The problem at hand is to obtain the cross correlation between samples of the distorted process $F(t)$ and $g(t)$ as well as $f(t)$ in a tractable form. The aim is to analyze the impact of nonlinearities commonly encountered in signal processing applications. Two further problems are to extend the results to MIMO systems and to generalize to the case of a non-Gaussian input process $f(t)$.

Solution: Bussgang decomposition

In the original paper [1], Bussgang computed the cross correlation of the two random variables obtained by sampling $F(t)$ and $g(t)$ at specific time instances. Let $x = f(t_1) \in \mathbb{R}$ and $y = g(t_2) \in \mathbb{R}$ denote the zero-mean Gaussian random variables obtained by sampling at time t_1 and t_2, respectively. Moreover, let $z = F(t_1) = U(x) \in \mathbb{R}$ be the sampled output of the nonlinear distortion function. We then have the following main result from [1, Sec. III].

Theorem 1: The Bussgang theorem

The cross correlation of $z = U(x)$ and y is

$$C_{zy} = \mathbb{E}\{U(x)y\} = \underbrace{\frac{\mathbb{E}\{U(x)x\}}{\mathbb{E}\{x^2\}}}_{\triangleq B}\mathbb{E}\{xy\}$$
$$= BC_{xy}, \tag{1}$$

where B is called the Bussgang gain, and $C_{xy} \triangleq \mathbb{E}\{xy\}$ is the cross correlation of x and y.

The Bussgang theorem shows that the cross correlation between two Gaussian signals is the same before and after one of them has passed through a nonlinear function, except for a scaling factor B. The value of B depends on the choice of $U(\cdot)$ but the theorem holds for any function. After the original paper [1], the Bussgang theorem was recognized as a special case of the Price theorem [2], which provides an alternative computation method for the Bussgang gain that we will return to later. In the remainder of this lecture note, we focus on the

Digital Object Identifier 10.1109/MSP.2020.3025538
Date of current version: 24 December 2020

sampled random variables $x = f(t_1)$, $y = g(t_2)$, and $z = F(t_1)$. However, it is important to remember that the discrete-time random variables are obtained from underlying continuous-time random processes, that the distortion is memoryless, and that x and z are samples taken at the same time.

A consequence of Theorem 1 is that the output signal can be decomposed as

$$z = U(x) = Bx + \eta, \tag{2}$$

where η is a zero-mean random variable that is uncorrelated to both x and y. This can be shown by multiplying both sides of (2) with x or y, taking the expectation, and using Theorem 1. Hence, the classical relationship in (2) has given rise to the name *Bussgang decomposition.* The Bussgang decomposition in its elementary form shows that the output contains the useful part Bx and the distortion part η. In other words, the output of a nonlinear function is equal to a scaled version of the input plus the uncorrelated distortion η. Note that η and x are not independent. Since $\eta = U(x) - Bx$ is a deterministic function of x, the distortion term is non-Gaussian distributed and statistically dependent on x. Even if the Bussgang decomposition is named after Bussgang, the decomposition is not explicitly stated in [1].

The Bussgang decomposition can be viewed as the linear minimum-mean squared error (MMSE) estimate of z given x, with η being the estimation error. Hence, the decomposition holds even if x is not Gaussian distributed, but the Bussgang decomposition also guarantees that the distortion signal η is uncorrelated to any other jointly Gaussian random variable y, which does not hold when considering the non-Gaussian distributed x and y.

Bussgang decomposition for complex random variables

The Bussgang theorem was extended to the complex case in [3]. We present this result and then provide a direct proof from [4] that uses the linear MMSE estimator. For notational convenience, in the remainder of this lecture note, we use $C_x \triangleq \mathbb{E}\{|x|^2\}$ to denote the power of a signal x and we use $C_{xy} \triangleq \mathbb{E}\{xy^*\}$ to denote the cross correlation between x and y.

Theorem 2: The complex Bussgang theorem

Consider the jointly circularly symmetric complex Gaussian random variables $x \in \mathbb{C}$ and $y \in \mathbb{C}$. Let $z = U(x) \in \mathbb{C}$ be the output of a deterministic function. The cross correlations $C_{zy} \triangleq \mathbb{E}\{zy^*\}$ and $C_{xy} \triangleq \mathbb{E}\{xy^*\}$ are then related as

$$\begin{aligned} C_{zy} &= \mathbb{E}\{U(x)y^*\} = \underbrace{\frac{\mathbb{E}\{U(x)x^*\}}{\mathbb{E}\{|x|^2\}}}_{\triangleq B = C_{zx}/C_x} \mathbb{E}\{xy^*\} \\ &= BC_{xy}. \end{aligned} \tag{3}$$

Proof

We begin by decomposing y into two parts:

$$y = \frac{\mathbb{E}\{yx^*\}}{\mathbb{E}\{|x|^2\}}x + \underbrace{\left(y - \frac{\mathbb{E}\{yx^*\}}{\mathbb{E}\{|x|^2\}}x\right)}_{\triangleq \epsilon}, \tag{4}$$

which is equivalent to computing an MMSE estimate of y given x, with ϵ representing the estimation error. Hence, it follows that the second part, ϵ, in (4) is uncorrelated with x:

$$\begin{aligned} \mathbb{E}\{\epsilon x^*\} &= \mathbb{E}\left\{\left(y - \frac{\mathbb{E}\{yx^*\}}{\mathbb{E}\{|x|^2\}}x\right)x^*\right\} \\ &= \mathbb{E}\{yx^*\} - \frac{\mathbb{E}\{yx^*\}}{\mathbb{E}\{|x|^2\}}\mathbb{E}\{|x|^2\} \\ &= 0. \end{aligned} \tag{5}$$

Since x and y are jointly Gaussian, the fact that x and ϵ are uncorrelated implies that they are also independent complex Gaussian variables. By using the decomposition in (4), it follows that

$$\begin{aligned} C_{zy} &= \mathbb{E}\{U(x)y^*\} \\ &= \frac{\mathbb{E}\{U(x)x^*\}}{\mathbb{E}\{|x|^2\}}\mathbb{E}\{xy^*\} + \underbrace{\mathbb{E}\{U(x)\epsilon^*\}}_{=0} \\ &= BC_{xy}, \end{aligned} \tag{6}$$

by using that the independence between x and ϵ implies $\mathbb{E}\{U(x)\epsilon^*\} = \mathbb{E}\{U(x)\}\mathbb{E}\{\epsilon^*\} = 0$. ■

The complex Bussgang theorem is the natural complex-valued extension of Theorem 1. The corresponding complex Bussgang decomposition is given by (2) with the only exception that the Bussgang gain is now computed as $B = C_{zx}/C_x = \mathbb{E}\{U(x)x^*\}/\mathbb{E}\{|x|^2\}$ instead.

A first use case of the Bussgang decomposition is to quantify the signal-to-distortion ratio (SDR) at the output of the distortion function. The SDR is the power ratio of the desired signal Bx to the additive distortion η:

$$\mathrm{SDR} = \frac{\mathbb{E}\{|Bx|^2\}}{\mathbb{E}\{|\eta|^2\}} = \frac{|B|^2 C_x}{C_z - |B|^2 C_x}, \tag{7}$$

where we have used that the additive distortion η is uncorrelated with the desired signal x.

A second use case is to analyze the performance of a communication system where $x \sim \mathcal{N}_{\mathbb{C}}(0, C_x)$ is the transmitted information signal. Suppose the receiver obtains the noisy distorted signal $U(x) + y = z + y$, where $U(\cdot)$ models the hardware distortion and y is thermal noise with power σ^2, which is uncorrelated to x. The hardware distortion might, for example, be caused of a sequence of nonideal blocks in the receiver hardware [5], as illustrated in Figure 1. The first block is the low-noise amplifier (LNA), which can distort both the amplitude and phase of the complex input signal. In the yellow part of the figure, the amplitude distortion is exemplified and clipping occurs for input signals with large amplitudes. The second block is the in-phase/quadrature (I/Q) demodulator that might have mismatches between its branches leading to I/Q imbalance. In the green curve, the effect of I/Q imbalance on a quadrature phase-shift keying constellation is shown, where the actual transmitted points are affected by leakage from the mirror subcarriers. Finally, in the analog-to-digital converter (ADC) block, the real and imaginary parts of the received signal are quantized to be represented by a finite number of bits. Quantization distortion is inevitable even if a large number of ADC bits are used [6]. We can use the Bussgang decomposition in (2) to rewrite the received signal as

$$U(x) + y = \underbrace{Bx}_{\text{Desired signal}} + \underbrace{\eta + y}_{\text{Uncorrelated signal}}. \tag{8}$$

This signal contains a desired part Bx and an uncorrelated additive "noise" term $\eta + y$. Since the latter term is uncorrelated with x, we can utilize the *worst-case uncorrelated additive noise* theorem from [7] to compute an achievable data rate. That theorem states that the worst distribution of $\eta+y$ from a rate perspective is independent complex Gaussian, in which case, the rate is

$$\log_2\left(1+\frac{\mathbb{E}\{|Bx|^2\}}{\mathbb{E}\{|\eta|^2\}+\sigma^2}\right)$$
$$=\log_2\left(1+\frac{|B|^2C_x}{C_z-|B|^2C_x+\sigma^2}\right)$$
$$\text{bits per channel use,} \tag{9}$$

where we have used that the distortion noise η is uncorrelated with y by Theorem 2. In fact, this is the exact expression for the generalized mutual information of the channel under Gaussian inputs and nearest-neighbor decoding [8].

An alternative computation of the Bussgang gain and two examples

If the distortion function $U(x)$ is differentiable or has finite jump discontinuities where the first derivative can be represented using the Dirac function, there is an alternative way of computing the Bussgang gain B that might be easier. We exemplify this way in the real-valued case where $x \sim \mathcal{N}(0,C_x)$ has the probability density function $p(x)=\left(1/\sqrt{2\pi C_x}\right)e^{-x^2/(2C_x)}$. Since its derivative is $p'(x)=-(x/C_x)p(x)$, we can then rewrite the Bussgang gain as

$$B=\frac{\mathbb{E}\{U(x)x\}}{C_x}=\int_{-\infty}^{\infty}\frac{U(x)x}{C_x}p(x)\,dx$$
$$\overset{(a)}{=}-\int_{-\infty}^{\infty}U(x)p'(x)\,dx$$
$$\overset{(b)}{=}\int_{-\infty}^{\infty}U'(x)p(x)\,dx=\mathbb{E}\{U'(x)\}, \tag{10}$$

where we identify $p'(x)$ in (a) and integrate by parts to get (b). The last expression in (10) reveals that the Bussgang gain can be also computed as the expected value of the first derivative of the distortion function. This result is a special case of the Price theorem [2].

Example 1: One-bit quantization [2, Sec. III]

Consider a real-valued signal $x \sim \mathcal{N}(0,C_x)$ that enters the nonlinear distortion function $z=U(x)=sgn(x)$, which represents one-bit quantization. The Bussgang gain can then be found as $B=\mathbb{E}\{U'(x)\}=2\mathbb{E}\{\delta(x)\}=2p(0)=\sqrt{2/(\pi C_x)}$, where $\delta(x)$ is the Dirac function. The same Bussgang gain can be computed as $B=\mathbb{E}\{U(x)x\}/\mathbb{E}\{x^2\}=\mathbb{E}\{|x|\}/C_x$.

A similar alternative way of computing the Bussgang gain exists in the complex-valued case, where the derivative of the distortion function $U(x)$ is defined as [9]

$$\frac{\partial U(x)}{\partial x}=\frac{1}{2}\left(\frac{\partial U(x)}{\partial \Re\{x\}}-j\frac{\partial U(x)}{\partial \Im\{x\}}\right). \tag{11}$$

One can then show that the Bussgang gain can be computed as [9]

$$B=\mathbb{E}\left\{\frac{\partial U(x)}{\partial x}\right\}. \tag{12}$$

Example 2: Third-order nonlinearity

Consider a complex-valued signal $x \sim \mathcal{N}_{\mathbb{C}}(0,C_x)$ that enters the third-order nonlinear distortion function $z=U(x)=|x|^2x$, which might model a nonlinear amplifier [4], [10]. The Bussgang gain can be obtained as $B=\mathbb{E}\{|x|^4\}/\mathbb{E}\{|x|^2\}=2C_x$. The same number is found by evaluating $B=\mathbb{E}\{\partial U(x)/\partial x\}=\mathbb{E}\{2|x|^2\}=2C_x$ using (11).

The additive quantization noise model is nothing but the Bussgang decomposition

The Bussgang decomposition is unique in the sense that it is the only decomposition $z=U(x)=Bx+\eta$ of a distorted signal having the property that the additive distortion noise η is uncorrelated with the input signal x and any other jointly Gaussian signal y. No other value of B can be used to achieve that.

One seemingly different decomposition is the additive quantization noise model (AQNM) originally proposed in [6] to model quantization errors. This model is sometimes described as an alternative decomposition; however, the AQNM is nothing but the Bussgang decomposition tailored to quantization. In [6, Lemma 1], a scalar quantizer function $Q(\cdot)$ is considered, which has the property $\mathbb{E}\{x|Q(x)\}=Q(x)$, which means that each quantization interval is represented by its mean value. When the input is $x \sim \mathcal{N}_{\mathbb{C}}(0,C_x)$, the AQNM says that the output can be expressed as a summation of a scaled version of x plus an uncorrelated distortion term η:

$$z=Q(x)=(1-\beta)x+\eta, \tag{13}$$

where $\beta=\mathbb{E}\{|x-z|^2\}/C_x$ and $\mathbb{E}\{|\eta|^2\}=\beta(1-\beta)C_x$.

We will show that (13) equals the Bussgang decomposition $x=Bx+\eta$, where the Bussgang gain $B=C_{zx}/C_x$ equals $1-\beta$. Using the assumption $\mathbb{E}\{x|Q(x)\}=Q(x)$ from [6], we have

$$C_{zx}=\mathbb{E}\{Q(x)x^*\}=\mathbb{E}\{\mathbb{E}\{Q(x)x^*|Q(x)\}\}$$
$$=\mathbb{E}\{Q(x)Q^*(x)\}=C_z. \tag{14}$$

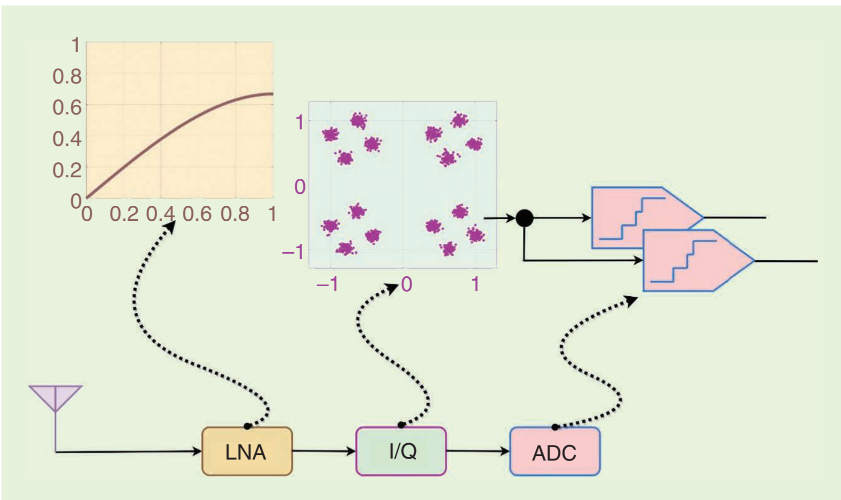

FIGURE 1. Some common sources of hardware impairments in a wireless receiver.

By utilizing this result, the scaling $1-\beta$ in (13) can be rewritten as

$$1-\beta = 1-\frac{\mathbb{E}\{|x-z|^2\}}{C_x} = 1-\frac{C_x+C_z-C_{zx}-C_{zx}^*}{C_x} = \frac{C_{zx}}{C_x} = B. \tag{15}$$

Hence, the AQNM is a special case of the Bussgang decomposition for distortion functions that satisfy the condition $\mathbb{E}\{x \mid Q(x)\} = Q(x)$. The bottom line is that the Bussgang decomposition is unique but that the value of B depends on the distortion function. We also note that the Gaussianity of the input signal x is only required for η to be uncorrelated to any other jointly Gaussian random variable y. The same decomposition can be used for the non-Gaussian x, which follows from utilizing the linear MMSE estimator, but η in that case will only be uncorrelated to x.

Extension to MIMO systems

In recent years, it has become popular to analyze MIMO systems that are subject to hardware impairments, in particular, in MIMO communications [11], [12]. In this part, we extend the Bussgang results to be applicable to such cases.

Consider two jointly circularly symmetric Gaussian random vectors $\mathbf{x} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{C}_x)$ and $\mathbf{y} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{C}_y)$, which both have length M. The correlation matrices are denoted as $\mathbf{C}_x = \mathbb{E}\{\mathbf{x}\mathbf{x}^{\mathrm{H}}\}$ and $\mathbf{C}_y = \mathbb{E}\{\mathbf{y}\mathbf{y}^{\mathrm{H}}\}$ and are assumed to have full rank. The cross-correlation matrix is denoted as $\mathbf{C}_{xy} = \mathbb{E}\{\mathbf{x}\mathbf{y}^{\mathrm{H}}\}$. Using this notation, we can generalize the Bussgang theorem as follows.

Theorem 3: The Bussgang theorem for MIMO distortions

Consider the jointly circularly symmetric Gaussian random vectors $\mathbf{x}$ and $\mathbf{y}$. Let $\mathbf{U}: \mathbb{C}^M \to \mathbb{C}^M$ denote a distortion function and $\mathbf{z} = \mathbf{U}(\mathbf{x})$ is the distorted signal when using $\mathbf{x}$ as input. The cross-correlation matrix $\mathbf{C}_{zy} = \mathbb{E}\{\mathbf{z}\mathbf{y}^{\mathrm{H}}\}$ of $\mathbf{z}$ and $\mathbf{y}$ is a linear transformation of the cross-correlation matrix $\mathbf{C}_{xy}$ of $\mathbf{x}$ and $\mathbf{y}$:

$$\mathbf{C}_{zy} = \mathbf{C}_{zx}\mathbf{C}_x^{-1}\mathbf{C}_{xy}. \tag{16}$$

Proof

The proof is a matrix extension of the proof of Theorem 2. Let us express $\mathbf{y}$ as a summation of the MMSE estimate of it given $\mathbf{x}$ and the estimation error $\boldsymbol{\epsilon} \in \mathbb{C}^M$:

$$\mathbf{y} = \mathbf{C}_{yx}\mathbf{C}_x^{-1}\mathbf{x} + \boldsymbol{\epsilon}, \tag{17}$$

where $\boldsymbol{\epsilon}$ is defined as $\boldsymbol{\epsilon} = \mathbf{y} - \mathbf{C}_{yx}\mathbf{C}_x^{-1}\mathbf{x}$. If we multiply both sides of (17) by $\mathbf{x}^{\mathrm{H}}$ from the right and take the expectation, we obtain

$$\mathbf{C}_{yx} = \mathbf{C}_{yx}\mathbf{C}_x^{-1}\mathbf{C}_x + \mathbb{E}\{\boldsymbol{\epsilon}\mathbf{x}^{\mathrm{H}}\} = \mathbf{C}_{yx} + \mathbb{E}\{\boldsymbol{\epsilon}\mathbf{x}^{\mathrm{H}}\}, \tag{18}$$

from which it follows that $\mathbb{E}\{\boldsymbol{\epsilon}\mathbf{x}^{\mathrm{H}}\} = \mathbf{0}$. Hence, $\boldsymbol{\epsilon}$ and $\mathbf{x}$ are uncorrelated, which implies that they are also independent since these are jointly Gaussian variables. Finally, we obtain (16) as

$$\mathbf{C}_{zy} = \mathbb{E}\{\mathbf{z}\mathbf{y}^{\mathrm{H}}\} = \mathbb{E}\{\mathbf{z}\mathbf{x}^{\mathrm{H}}\}\mathbf{C}_x^{-1}\mathbf{C}_{yx}^{\mathrm{H}} + \mathbb{E}\{\mathbf{z}\boldsymbol{\epsilon}^{\mathrm{H}}\} = \mathbf{C}_{zx}\mathbf{C}_x^{-1}\mathbf{C}_{xy} \tag{19}$$

by utilizing that $\mathbf{C}_{yx}^{\mathrm{H}} = \mathbf{C}_{xy}$ and that $\mathbb{E}\{\mathbf{z}\boldsymbol{\epsilon}^{\mathrm{H}}\} = \mathbf{0}$ since $\mathbf{z}$ and $\boldsymbol{\epsilon}$ are independent. ■

From this theorem we notice that the Bussgang gain is represented by the matrix

$$\mathbf{B} = \mathbf{C}_{zx}\mathbf{C}_x^{-1}, \tag{20}$$

and we call it a MIMO extension since the distortion function takes multiple inputs and provide multiple outputs. It is possible to extend the result to the case where $\mathbf{C}_x$ is rank deficient, in which case the inverse in (20) is replaced by a pseudo-inverse; see [4, Sec. II.A] for details.

A consequence of Theorem 3 is the Bussgang decomposition for MIMO functions:

$$\mathbf{z} = \mathbf{U}(\mathbf{x}) = \mathbf{B}\mathbf{x} + \boldsymbol{\eta}, \tag{21}$$

where the additive distortion term $\boldsymbol{\eta}$ is uncorrelated both with $\mathbf{x}$ and any other Gaussian random vector $\mathbf{y}$ that is correlated with $\mathbf{x}$. This result is illustrated in Figure 2(a).

Element-wise distortion for MIMO systems

The Bussgang decomposition for MIMO functions has been widely used to model the hardware impairments in multiple-antenna communication systems [11]. In this case, M is the number of receive antennas and the distortion function represents

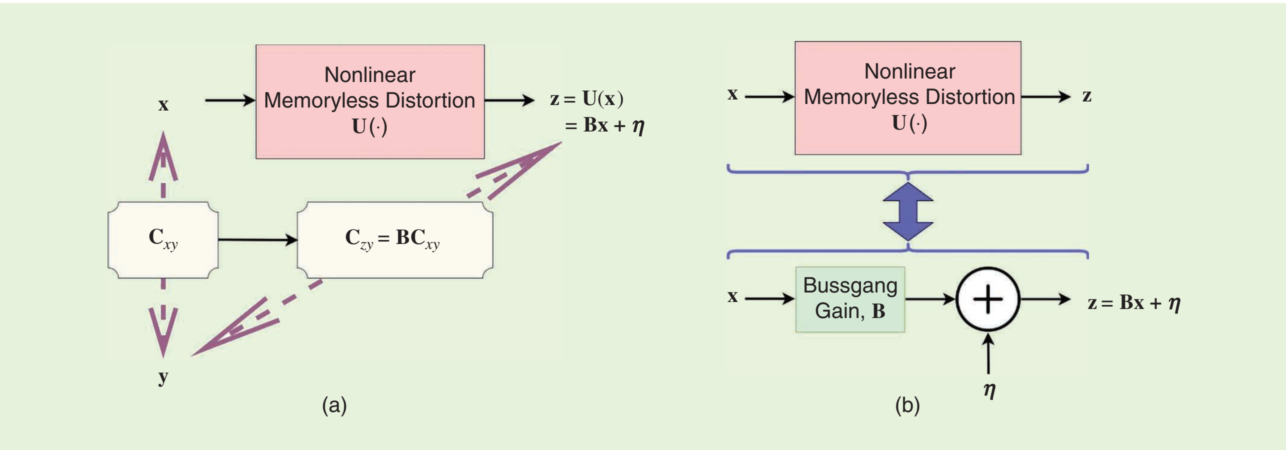

FIGURE 2. The Bussgang decomposition for a nonlinear memoryless distortion function $\mathbf{U}(\cdot)$. (a) the Bussgang decomposition for jointly Gaussian random vectors $\mathbf{x}$ and $\mathbf{y}$ and (b) the generalized Bussgang decomposition for a non-Gaussian random vector $\mathbf{x}$.

impairments in the antenna branches. A common assumption is that there is no crosstalk between the branches, so that each one can be separately modeled in the way shown in Figure 1. The distortion function then has the form

$$\mathbf{z}=\mathbf{U}(\mathbf{x})=\begin{bmatrix} U_1(x_1) \\ \vdots \\ U_M(x_M) \end{bmatrix}, \quad (22)$$

where x_m denotes the mth element of $\mathbf{x}$. Hence, each output is a distorted version of only the input having the same index. We can then simplify the Bussgang matrix. We note that all elements of $\mathbf{x}$ are jointly Gaussian and the distortion noise for each element of $\mathbf{z}$ is thus uncorrelated to $\mathbf{x}$ by Theorem 2. Then, it follows that $\mathbf{C}_{zx}=\mathbf{D}\mathbf{C}_x$, where $\mathbf{D}=\mathrm{diag}(d_1,\ldots,d_M)$ is a diagonal matrix and $d_m=\mathbb{E}\{U_m(x_m)x_m^*\}/\mathbb{E}\{|x_m|^2\}$ is the Bussgang gain corresponding to the mth component of the distortion function, i.e., $z_m=U_m(x_m)$. Hence, the Bussgang gain matrix of the overall MIMO distortion becomes $\mathbf{B}=\mathbf{C}_{zx}\mathbf{C}_x^{-1}=\mathbf{D}$ from (20), and we obtain the simplified Bussgang decomposition

$$\mathbf{z}=\mathbf{D}\mathbf{x}+\boldsymbol{\eta}=\begin{bmatrix} d_1x_1 \\ \vdots \\ d_Mx_M \end{bmatrix}+\boldsymbol{\eta}. \quad (23)$$

Hence, when an element-wise distortion function affects the Gaussian signal $\mathbf{x}$, the output $\mathbf{z}$ is an element-wise scaled version of $\mathbf{x}$ plus a distortion vector $\boldsymbol{\eta}$ that is uncorrelated with $\mathbf{x}$. We reiterate that the distortion vector $\boldsymbol{\eta}$ is uncorrelated to any other vector $\mathbf{y}$ that is jointly Gaussian with $\mathbf{x}$, which is the part of this result that critically requires $\mathbf{x}$ and $\mathbf{y}$ to be Gaussian distributed.

Are the elements of the distortion noise η uncorrelated?

Since the Bussgang gain matrix is diagonal when having element-wise distortions, one may tend to think that the elements of the distortion $\boldsymbol{\eta}$ will also be uncorrelated, so that we effectively get one separate Bussgang decomposition per received signal. However, this is generally not the case, as we show next. Let $\mathbf{C}_\eta=\mathbb{E}\{\boldsymbol{\eta}\boldsymbol{\eta}^{\mathrm{H}}\}\in\mathbb{C}^{M\times M}$ denote the correlation matrix of the distortion vector $\boldsymbol{\eta}$. Using the fact that $\boldsymbol{\eta}$ is uncorrelated with $\mathbf{x}$, it can be computed as

$$\mathbf{C}_\eta=\mathbf{C}_z-\mathbf{B}\mathbf{C}_x\mathbf{B}^{\mathrm{H}}. \quad (24)$$

Whenever the input signal $\mathbf{x}$ contains correlated elements, such that $\mathbf{C}_x$ is nondiagonal, the correlation matrix will likely also be nondiagonal. This is intuitively quite clear: If two (almost) identical signals are sent through identical hardware components, then the distortion should also be (almost) identical. This type of correlation typically appears in wireless communications since each receive antenna observes a different linear combination of the same transmitted information signals. Some conditions for when the correlation can be neglected, so that $\mathbf{C}_\eta$ is approximately diagonal, are derived in [4]. However, it is rather common that the correlation is neglected without motivation (cf. [12]), which might lead to substantial approximation errors.

As an example, we consider a setup where a four-antenna receiver quantizes the real and imaginary parts of each entry in the received signal $\mathbf{x}$ using identical b-bit ADCs. The input signal is generated as $\mathbf{x}=\mathbf{H}\mathbf{s}$, where $\mathbf{H}\in\mathbb{C}^{4\times4}$ is the MIMO channel matrix from a four-antenna transmitter. We consider Rayleigh fading where $\mathbf{H}$ has independent $\mathcal{N}_{\mathbb{C}}(0,1)$-distributed entries. For each channel realization, $\mathbf{H}$ is assumed perfectly known and the transmitted signal is $\mathbf{s}\sim\mathcal{N}_{\mathbb{C}}(\mathbf{0},\mathbf{I}_4)$, so $\mathbf{x}$ is conditionally complex Gaussian distributed. The Bussgang decomposition then says that the ADC output can be written as $\mathbf{z}=\mathbf{D}\mathbf{x}+\boldsymbol{\eta}$. To demonstrate that the elements of $\boldsymbol{\eta}$ are correlated, Figure 3 shows the cumulative distribution function (CDF) of the normalized off-diagonal elements of $\mathbf{C}_\eta$ (i.e., the correlation coefficients) for different number of ADC bits. When the ADC resolution is low, most of the correlation coefficients are nonzero and some are rather large. However, when the ADC resolution is high, the off-diagonal elements are almost zero and can potentially be approximated as zero when quantifying communication rates.

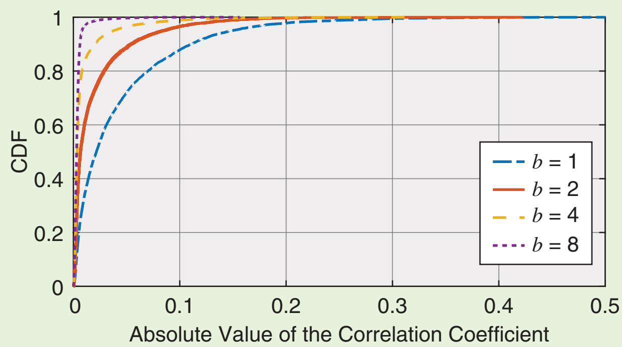

FIGURE 3. The cumulative distribution function (CDF) of the absolute value of the correlation coefficient between elements in $\boldsymbol{\eta}$.

The generalized Bussgang decomposition for non-Gaussian input signals

In the Bussgang theorem, we are utilizing that $\mathbf{x}$ and $\mathbf{y}$ are Gaussian signals. The main result cannot be generalized to non-Gaussian signals. However, we can always decompose the distorted signal according to (21) using the Bussgang gain matrix $\mathbf{B}=\mathbf{C}_{zx}\mathbf{C}_x^{-1}$, but it generally will not be a diagonal matrix, even if an element-wise distortion of the type in (22) is used. As mentioned previously, the intuition is that $\mathbf{B}\mathbf{x}$ is the linear MMSE estimate of $\mathbf{z}$ given a non-Gaussian distributed observation $\mathbf{x}$. In this analogy, $\boldsymbol{\eta}$ is the estimation error which is uncorrelated with $\mathbf{x}$ since

$$\begin{aligned}\mathbb{E}\{\boldsymbol{\eta}\mathbf{x}^{\mathrm{H}}\}&=\mathbb{E}\{(\mathbf{z}-\mathbf{C}_{zx}\mathbf{C}_x^{-1}\mathbf{x})\mathbf{x}^{\mathrm{H}}\}\\&=\mathbf{C}_{zx}-\mathbf{C}_{zx}\mathbf{C}_x^{-1}\mathbf{C}_x=\mathbf{0}. \quad (25)\end{aligned}$$

The generalized Bussgang decomposition for the non-Gaussian input $\mathbf{x}$ is illustrated in Figure 2(b). It is suitable both for quantifying the SDR and for analyzing the performance of nonlinear communication systems. For example, [10] did this using practically modulated data signals. That article also showed that although treating the uncorrelated distortion η as independent Gaussian noise is convenient, one can increase the performance by exploiting its information content.

Lessons learned

The Bussgang decomposition establishes that the output of a nonlinear function is a scaled version of the random input signal plus an uncorrelated distortion term. It is an exact and unique representation. The distortion is not independent and not Gaussian, but it can be treated as that to obtain a lower bound on the communication performance. The decomposition can be extended to MIMO systems, but then, the entries of the distortion vector are generally mutually correlated.

Acknowledgments

The authors were partially supported by the Excellence Center at Linköping–Lund in Information Technology (ELLIIT) and the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation.

Authors

Özlem Tuğfe Demir (ozlem.tugfe.demir@liu.se) received her Ph.D. degree in electrical and electronics engineering from Middle East Technical University, Ankara, Turkey, in 2018. She is currently a postdoctoral researcher at Linköping University, Linköping, Sweden. Her research interests focus on signal processing and optimization in wireless communications; massive multiple-input, multiple-output systems; deep learning; and green communications. She is a Member of IEEE.

Emil Björnson (emil.bjornson@liu.se) received his Ph.D. degree in telecommunications from the KTH Royal Institute of Technology, Sweden, in 2011. He is currently an associate professor at Linköping University, Linköping, Sweden. He has authored *Optimal Resource Allocation in Coordinated Multi-Cell Systems* (2013) and *Massive MIMO Networks: Spectral, Energy, and Hardware Efficiency* (2017). He received the 2018 IEEE Marconi Prize Paper Award in Wireless Communications, the 2019 European Association for Signal Processing Early CAREER Award, the 2019 IEEE Communications Society Fred W. Ellersick Prize, and the 2019 IEEE Signal Processing Magazine Best Column Award. In September 2020, he became a part-time visiting full professor at the KTH Royal Institute of Technology. He is a Senior Member of IEEE.

References

[1] J. J. Bussgang, "Crosscorrelation functions of amplitude-distorted Gaussian signals," Research Lab. Electronics, Massachusetts Inst. Technology, Cambridge, MA, Tech. Rep. 216, 1952. [Online]. Available: http://hdl.handle.net/1721.1/4847

[2] H. E. Rowe, "Memoryless nonlinearities with Gaussian inputs: Elementary results," *Bell Syst. Tech. J.*, vol. 61, no. 7, pp. 1519–1525, 1982. doi: 10.1002/j.1538-7305.1982.tb04356.x.

[3] J. Minkoff, "The role of AM-to-PM conversion in memoryless nonlinear systems," *IEEE Trans. Commun.*, vol. 33, no. 2, pp. 139–144, 1985. doi: 10.1109/TCOM.1985.1096262.

[4] E. Björnson, L. Sanguinetti, and J. Hoydis, "Hardware distortion correlation has negligible impact on UL massive MIMO spectral efficiency," *IEEE Trans. Commun.*, vol. 67, no. 2, pp. 1085–1098, Feb. 2019. doi: 10.1109/TCOMM.2018.2877331.

[5] T. Schenk, *RF Imperfections in High-Rate Wireless Systems: Impact and Digital Compensation*. Berlin: Springer-Verlag, 2008.

[6] A. K. Fletcher, S. Rangan, V. K. Goyal, and K. Ramchandran, "Robust predictive quantization: Analysis and design via convex optimization," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 4, pp. 618–632, 2007. doi: 10.1109/JSTSP.2007.910622.

[7] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, 2003. doi: 10.1109/TIT.2003.809594.

[8] W. Zhang, Y. Wang, C. Shen, and N. Liang, "A regression approach to certain information transmission problems," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 11, pp. 2517–2531, 2019. doi: 10.1109/JSAC.2019.2933964.

[9] W. McGee, "Circularly complex Gaussian noise—A price theorem and a Mehler expansion," *IEEE Trans. Inf. Theory*, vol. 15, no. 2, pp. 317–319, 1969. doi: 10.1109/TIT.1969.1054293.

[10] Ö. T. Demir and E. Björnson, "Channel estimation in massive MIMO under hardware non-linearities: Bayesian methods versus deep learning," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 109–124, 2020. doi: 10.1109/OJCOMS.2019.2959913.

[11] E. Björnson, J. Hoydis, M. Kountouris, and M. Debbah, "Massive MIMO systems with non-ideal hardware: Energy efficiency, estimation, and capacity limits," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 7112–7139, 2014. doi: 10.1109/TIT.2014.2354403.

[12] L. Xu, X. Lu, S. Jin, F. Gao, and Y. Zhu, "On the uplink achievable rate of massive MIMO system with low-resolution ADC and RF impairments," *IEEE Commun. Lett.*, vol. 23, no. 3, pp. 502–505, 2019. doi: 10.1109/LCOMM.2019.2895823.

SP