On positivity of time-frequency distributions.

Janssen, A.J.E.M.; Claasen, T.A.C.M.

Published in: IEEE Transactions on Acoustics, Speech, and Signal Processing

DOI: 10.1109/TASSP.1985.1164622

Published: 01/01/1985

Citation for published version (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal?

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
This implicit solution can be solved for \( d(t) \) and \( d'(t) \), respectively, for a specific time function for \( d'(t) \) and \( d(t) \), respectively if the inverse function for \( d'(t) \) and \( d(t) \) respectively exists. Equation (8) gives examples for specific time functions as the linear and quadratic case. These are most commonly used for approximations of complex physical time functions for time varying delays.  

Linear case:  
Given  
\[
\frac{d(t)}{d'(t)} = a + bt\tag{8a}
\]
with  
\[
t \geq a/(1 - b)
\]
for guarantee causality, then follows  
\[
d'(t) = (a + bt)/(1 - b). \tag{8b}
\]

Quadratic case:  
Given  
\[
d(t) = a + bt + ct^2\tag{8c}
\]
with  
\[
t \geq \frac{(1 - b)^2/4c^2}{a + t/2c}\tag{8d}
\]
for causality and  
\[
t \leq (1 - b)^2/4c - a\tag{8d}
\]
for guaranteeing uniqueness.  
It then follows that  
\[
d'(t) = \frac{1 - b}{2c} - \sqrt{\left(1 - b\right)^2/4c^2 - a + t/2c} - t. \tag{8d}
\]
To guarantee causality and uniqueness time varying delays defined with respect to the delayed signal are only valid for a bounded time interval. Otherwise it would be possible that "wavefronts pass other wavefronts." This means that if two wavefronts enter such time varying delay elements then they have changed their order at the output of the time varying delay element.  

IV. THE EXPLICIT SOLUTION  
The implicit solution for \( \epsilon(t) \), (3c) shall be transformed to an explicit solution using the results found in Section III. For convenience substitution, (9a) shall be used  
\[
d_s(t) = \Delta_s(t) + \tau_s(t - \Delta_s(t)). \tag{9a}
\]
Where \( \Delta_s(t) \) is the time varying delay between \( s_i(t) \) and \( s_i(t) \), (3c) can be written as  
\[
\epsilon(t) + \tau_s(t - \epsilon(t)) = \Delta_s(t). \tag{9b}
\]
With (7a) it follows that  
\[
\tau_s(t) = \tau_s(t - \tau_s(t)). \tag{9c}
\]
Inserting (9c) into (9b) yields  
\[
d_s(t) = \epsilon(t) + \tau_s(t - \epsilon(t)) \tag{9d}
\]
Comparing (9b) and (9d), it follows that  
\[
\epsilon(t) = d_s(t) - \tau_s(t - d_s(t)). \tag{9c}
\]
Equation (9c) is the explicit solution for \( \epsilon(t) \) as a function of \( d_s(t) \), see (9a) and \( \tau_s(t) \). Here \( \tau_s(t) \) is defined with respect to the source signal, \( s_i(t) \). Adams et al.'s solution (4) is an approximation of (9e) assuming that \( d_s(t) \) and \( \tau_s(t) \) can be approximated by a linear function with small delay rates (10).  
Given  
\[
d_s(t) = mt\tag{10a}
\]
\[
\tau_s(t) = nt\tag{10b}
\]
then follows from (9c) and (8b)  
\[
\epsilon(t) = mt - nt\tag{10c}
\]
or  
\[
\epsilon(t) = d_s(t) - \tau_s(t)\tag{10d}
\]
where (10d) represents the solution presented in [1].  

V. SUMMARY  
Designing estimators for time varying delays often needs the specification of appropriate signal generation models. Here the signal generation model in [1] has been investigated and it has been shown that it is not correct in the general case. However, Adams et al. [1] discuss a specific problem and for small delay rate their solution is a good approximation. A detailed discussion of two possible definitions of a time varying delay is given with the intention of improving the understanding of signal generation models with several time varying delays.  

REFERENCES  

On Positivity of Time-Frequency Distributions  
A. J. E. M. JANSSEN and T. A. C. M. CLAASEN  

Abstract—This correspondence addresses the problem of how to regard the fundamental impossibility with time-frequency energy distribution of Cohen's class always to be nonnegative and, at the same time, to have correct marginal distributions. It is shown that the Wigner distribution is the only member of a large class of bilinear time-frequency distributions that becomes nonnegative after smoothing in the time-frequency plane by means of Gaussian weight functions with BT product equal to unity.  

Manuscript received November 28, 1983; revised November 16, 1984. The authors are with Philips Research Laboratories, Eindhoven, The Netherlands.
I. INTRODUCTION

The need for a tool to adequately describe signals in time and frequency simultaneously has been felt for a long time. In the past few decades several functions depending on time and frequency $\omega$ have been proposed to meet the needs of signal analysis on this point. Among these proposals are the distributions of Wigner, the spectrogram distribution and, as the most well known, the spectrogram (see [1]-[4]). All these distributions have been shown to be members of one large class of time-frequency distributions, viz. Cohen’s class (see [5]-[6]). The signal analyst uses these functions to obtain an idea of the distribution of the energy over time and frequency of nonstationary signals. This is done to get an impression of the “physically” occurring phenomena such as frequency-dependent signal delays, instantaneous frequencies, time-varying resonances. Therefore, one would like such a function, $C_f(t, \omega)$, to satisfy a number of conditions.

1) Just as with the instantaneous power $|f|^2$ and the power spectral density $|F|^2$ of the signal, $C_f$ should be bilinear in $f$ so that the global property

$$E_{f_1,f_2} = |\alpha|^2 E_{f_1} + |\beta|^2 E_{f_2} + 2Re \alpha \beta^* C_{f_1,f_2}, \tag{1}$$

where $E_f$ stands for either $|f|^2$ or $|F|^2$, and $E_{f_1,f_2}$ is a cross term reflected locally by

$$C_{f_1,f_2} = |\alpha|^2 C_{f_1} + |\beta|^2 C_{f_2} + 2Re \alpha \beta^* C_{f_1,f_2}, \tag{2}$$

for all $\alpha$, $\beta$, $f_1$, and $f_2$.

2) $C_f$ should have correct marginal distributions. That is, integration over all frequencies at a certain time $t$ should yield the instantaneous power

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} C_f(t, \omega) \, d\omega = |f(t)|^2 \tag{3}$$

and similarly, integration over all time should yield the power spectral density

$$\int_{-\infty}^{\infty} C_f(t, \omega) \, dt = |F(t)|^2. \tag{4}$$

3) $C_f(t, \omega)$ should be nonnegative for all $t$ and $\omega$.

Now both the Wigner distribution and Rihaczek’s distribution have correct marginal distributions, but, unfortunately, they may take negative values (and Rihaczek’s distribution even complex values). The spectrogram on the other hand does not have negative values, but fails to have correct marginals. This is no surprise as it has been shown [7] that these requirements (correct marginals and positivity) are incompatible within Cohen’s class, which is, in a sense, a manifestation of the time-frequency uncertainty principle, expressed by Heisenberg’s uncertainty relation [8].

Probability density functions that are always nonnegative and that do have correct marginals have been given by Cohen and Zaporovanny [9]. As opposed to the members of Cohen’s class, these distributions involve the signal in a nonbilinear way. Although this need not hamper an interpretation as a probability density, it makes an interpretation as an energy density function doubtful. Furthermore, the density functions in [9] contain a normalized probability function on the unit square as a sort of parameter. It is, however, hard to see how to obtain, by clever choices of this parameter, satisfactory representations for signals in sufficiently large classes (e.g., FM signals or impulse response functions of linear systems, cf. [6], [10], [11]). Therefore, a logical restriction in the context of signal analysis is to only consider distributions within Cohen’s class.

It is the aim of this correspondence to show that in this class the Wigner distribution comes closest to the utopia of both having correct marginals and being nonnegative, and that it stands alone in this respect among the members of Cohen’s class. This will be demonstrated by proving that the Wigner distribution is the only member of Cohen’s class with correct marginals that always becomes nonnegative after smoothing with a two-dimensional Gaussian weight function with BT product equal to unity. As an example, certain distributions of Rihaczek’s type are considered for the chirp, the signal whose instantaneous frequency changes linearly with time.

II. THE TIME-FREQUENCY DISTRIBUTIONS OF COHEN

In this section some known facts about Cohen’s class of distributions are presented with special attention for the particularly good properties of the Wigner distribution. We start with the definition of the Wigner distribution. When $f(t)$ is a continuous-time signal, its Wigner distribution $W_f(t, \omega)$ is defined as

$$W_f(t, \omega) = \int_{-\infty}^{\infty} e^{-j\omega t} f(t + \tau/2) f^*(t - \tau/2) \, d\tau. \tag{5}$$

Although introduced a long time ago [2], [3], the Wigner distribution has been proved to be a useful tool in signal analysis only rather recently (see [6], [12], and [11] for an application in loudspeaker evaluation).

Any member $C_f$ of Cohen’s class can be expressed in terms of the Wigner distribution of $f$ with the aid of what may be called a kernel function [6], [13]

$$C_f(t, \omega; \varphi) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \varphi(t - \tau, \omega - \xi) W_f(\tau, \xi) \, d\tau \, d\xi. \tag{6}$$

To retain the bilinearity property, $\varphi$ may not depend on $f$ as is the case in [9]. For the members of Cohen’s class one has the property [6] that shifts in time and frequency of the signal are reflected in the distribution by similar shifts, i.e.,

$$C_{f,R_b}(t, \omega; \varphi) = C_f(t + a, \omega + b; \varphi + a, b) \tag{7}$$

where $T_{R_b}f(t)$ is given by $(T_{R_b}f)(t) = \exp(-j\omega t) f(t + a)$.

Taking $\varphi(t, \omega) = \varphi_w(t, \omega) = 2\pi \delta(\omega)$ in (6), one obtains the Wigner distribution. When one takes $\varphi(t, \omega) = \varphi_w(t, \omega) = 2\exp(j\omega t)$, one obtains Rihaczek’s distribution $R_f(t, \omega) = F^2(t) F(\omega)$ with $F(\omega)$ its Fourier transform of $f$. And when one takes $\varphi(t, \omega) = \varphi_g(t, \omega) = W_f(t, \omega)$ (where $W_f$ is the Wigner distribution of some function $g$), one gets the spectrogram

$$S_f(t, \omega) = \left| \int_{-\infty}^{\infty} g(t - \tau) f(\tau) \, d\tau \right|^2 \tag{8}$$

in which $g$ acts as the window function.

The condition of having correct marginals means that (3) and (4) hold for any $f(t)$. It can be shown [6] that this is the case if and only if

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} \varphi(t, \omega) \, d\omega = \delta(t), \quad \int_{-\infty}^{\infty} \varphi(t, \omega) \, dt = 2\pi \delta(\omega). \tag{9}$$

It can easily be checked that $\varphi_w$ and $\varphi_g$ satisfy these conditions, so both the Wigner distribution and Rihaczek’s distribution have correct marginals. Since no Wigner distribution, and hence no $\varphi_g$ can ever satisfy both relations in (9) simultaneously, no spectrogram can yield correct marginals.

The fact that the distribution functions in Cohen’s class with correct marginals cannot be anywhere nonnegative for all $f$ would not be overly impractical if the set of signals for which negative values occur would be small or “esoteric” and/or if the negative values would only occur in very restricted regions of the $(t, \omega)$-plane. In practice, it turns out that the distribution functions are everywhere nonnegative only in exceptional cases, and that the regions where negative values are attained are so large that considerable smoothing is required to get a nonnegative distribution. That signals with distribution functions that attain negative values are abound can also be inferred from the proof in [7, Sect. 1].

It is remarkable that suitable averages of the Wigner distribution over the time-frequency plane are nonnegative for all $f(t)$. In particular, it can be shown [14] that
\[
\int_{-\infty}^{\infty} K(t - \tau, \omega - \xi) W_f(\tau, \xi) \, d\xi \geq 0 \quad (10)
\]
for all \( f(t) \), here
\[
K(t, \omega) = \frac{1}{\Omega T} \exp \left( -\frac{t^2}{2T^2} - \frac{\omega^2}{\Omega^2} \right) \quad (11)
\]
with \( \Omega T \geq 1 \) (see [15] for more results of this type).

The fact that (10) is always nonnegative when \( \Omega T \geq 1 \) might raise the question whether averaging the Wigner distribution over a region in the time-frequency plane with area \( 1 \) always produces nonnegative values. The answer is no and as a matter of fact, it can be shown [15] that there is no \( K(t, \omega) \) that vanishes outside a bounded set in the time-frequency plane such that (10) is always nonnegative. As a consequence, there is no signal \( f(t) \) whose Wigner distribution vanishes outside a bounded set (for otherwise (6) and (8) would contradict the statement just made). One can show that this is just another way of saying that there are no signals that are both time-limited and band limited.

In connection with (10) and (11) the following facts about the Wigner distribution are worth mentioning [16]. When \( W_f(t, \omega) \geq 0 \) everywhere if and only if \( f(t) = \exp (-\alpha t^2 + 2\beta t + \gamma) \) or \( \exp \left( -t^2 + \omega \right) \) with a real, \( \alpha, \beta, \gamma \) complex and \( \Re \alpha \geq 0 \). More recently, this result has been strengthened as follows [17]. When (10) is nonnegative everywhere, where \( K(t, \omega) \) as in (11) with \( \Omega T < 1 \), then \( f(t) \) must also have the special form given above. That is, to obtain a nonnegative distribution from the Wigner distribution by smoothing with a Gaussian weight function \( K \) as in (11), one has to take the parameters \( T \) and \( \Omega \) such that \( \Omega T \geq 1 \) (unless one deals with a very exceptional case). Thus, even the slightest deviation of \( f(t) \) from being of the above special form causes the Wigner distribution to be negative at places to an extent that considerable smoothing is required to get an everywhere nonnegative distribution.

The striking phenomenon noted above for the Wigner distribution has not been proved to hold for the general time-frequency distributions of Cohen's class. What we shall show, however, in the next section is the following.

1) There is no distribution \( C_f \) in Cohen's class with correct marginals that becomes nonnegative for every \( f(t) \) after smoothing as in (10) and (11) if \( \Omega T < 1 \).

2) Of all members of Cohen's class with correct marginals, only the Wigner distribution will yield a nonnegative result for all \( f(t) \) after smoothing as in (10) and (11) with \( \Omega T = 1 \).

III. DISTRIBUTIONS OF COHEN'S CLASS WITH CORRECT MARGINALS

In this section a schematic proof of the result announced at the end of the previous section is given; in [18] a completely detailed proof and various extensions of the result can be found.

Assume that
\[
\frac{1}{\Omega T} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp \left( -\frac{t^2}{2T^2} - \frac{\omega^2}{\Omega^2} \right) C_f(t, \omega; \varphi) \, dt \, d\omega \geq 0 \quad (12)
\]
for all \( f(t) \). In view of (7) this is equivalent with assuming that (10) \[ \text{with } C_f(t, \xi; \varphi) \text{ instead of } W_f(t, \xi) \] is nonnegative for all \( f(t) \). Assume, in addition, that \( C_f(t, \omega; \varphi) \) has correct marginals for all \( f(t) \). By (6) we can write (12) as
\[
\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} G(\tau, \xi) W_f(\tau, \xi) \, d\xi \geq 0 \quad (13)
\]
for all \( f(t) \), where
\[
G(\tau, \xi) = \frac{1}{2\pi \Omega T} \int_{-\infty}^{\infty} \exp \left( -\frac{(t - \tau)^2}{2T^2} - \frac{\omega^2}{\Omega^2} \right) \varphi(t - \tau, \omega - \xi) \cdot \exp \left( -\frac{t^2}{2T^2} - \frac{\omega^2}{\Omega^2} \right) \, dt \, d\omega. \quad (14)
\]
It can be shown, on the assumption that \( G(\varphi, \xi) \) is (square) integrable, that nonnegativity of (13) for all \( f(t) \) implies the existence of numbers \( c_n \geq 0 \) and orthonormal functions \( f_n(t) \) such that
\[
G(\tau, \xi) = \sum c_n f_n(t) f_n(\xi). \quad (15)
\]
Indeed, this can easily be deduced from the theorem [19] that a square integrable function \( H(t, s) \) satisfying
\[
\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} H(t, s) f(t) f^*(s) \, dt \, ds \geq 0 \quad (16)
\]
for all \( f(t) \) admits a representation \( H(t, s) = \sum c_n f_n(t) f_n(s) \) with \( c_n \) and \( f_n \) as above.

If one integrates the identity (15) over all \( \xi \), one gets by (3)
\[
\sum c_n |f_n(\tau)|^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} G(\tau, \xi) \, d\xi \quad (17)
\]
for all \( \tau \). Similarly,
\[
\sum c_n |f_n(\xi)|^2 = \frac{\sqrt{\pi}}{\Omega} \exp \left( -\frac{\xi^2}{2\Omega^2} \right) \quad (18)
\]
for all \( \xi \). Alternatively, \( F_n(\xi) \) is the Fourier transform of \( f_n(t) \). The right-hand sides of (17) and (18) can be evaluated by inserting (8) into (14) and interchanging integrals. One ends up with the identities
\[
\sum c_n |F_n(\tau)|^2 = \frac{1}{2\pi \sqrt{\pi}} \exp \left( -\frac{\tau^2}{2\pi} \right) \quad (19)
\]
for all \( \tau \), and
\[
\sum c_n |F_n(\xi)|^2 = \frac{\sqrt{\pi}}{\Omega} \exp \left( -\frac{\xi^2}{2\Omega^2} \right) \quad (20)
\]
for all \( \xi \). Since all \( c_n \geq 0 \), one thus obtains
\[
c_n |f_n(\tau)|^2 \leq \frac{1}{2\pi \sqrt{\pi}} \exp \left( -\frac{\tau^2}{2\pi} \right), \quad (21)
\]
for all \( n, \tau, \xi \). Now it can be shown [20, Theorem 128] that for \( \Omega T < 1 \) the two inequalities in (21) are incompatible, unless \( c_n = 0 \), in view of a version of Heisenberg's uncertainty relation. For \( \Omega T = 1 \) these two inequalities are just compatible, but force each \( c_n f_n(\tau) \) to a be a multiple of the Gaussian \( \exp \left( -\tau^2/2T^2 \right) \). In the first case we conclude that \( c_n = 0 \) for all \( n \), so that \( G = 0 \), and thus \( \varphi = 0 \). In the second case we see that \( G(\tau, \xi) \) must be a multiple of \( 2\pi \exp \left( -\tau^2/2T^2 + \xi^2/2\Omega^2 \right) \), the Wigner distribution of \( \exp \left( -\tau^2/2T^2 + \xi^2/2\Omega^2 \right) \). In view of (14) the proof is easily completed now.

IV. GENERALIZED RIHACZEK DISTRIBUTIONS

In this section it is shown that the situation as concerns positivity is often far worse than is indicated by the main result of this correspondence. Take, for example,
\[
\varphi(t, \omega) = \varphi_0(t, \omega) = 2\pi \delta(t) \delta(\omega) \text{ or } \alpha^{-1} \exp \left( j\alpha^{-1} t \omega \right) \quad (22)
\]
according as \( \alpha = 0 \) or \( \alpha \neq 0 \). The distributions corresponding to \( \varphi_0 \) with \( \alpha = 0 \) (generalized Rihaczek distribution) and the ones corresponding to \( \alpha = 0 \) (Wigner distribution) were compared in [10] with respect to spread. It was shown [10] that for all \( f(t) \) and all \( (t_0, \omega_0) \) the minimum of
\[
\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[ \frac{(t - t_0)^2}{T^2} + (\omega - \omega_0)^2 \Omega^2 \right] C_f(t, \omega; \varphi_0) \, dt \, d\omega \quad (23)
\]
as a function of \( \alpha \) occurs at \( \alpha = 0 \). Taking for \( (t_0, \omega_0) \) the center of gravity,
one sees that the Wigner distribution is, in a sense, best concentrated among the \( \mathcal{C}_f(t, \omega; \varphi_0) \)'s. In this section it is demonstrated that, at least for the chirp, the Wigner distribution also behaves best with respect to positivity. Consider the chirp signal \( f(t) = \exp \left( \frac{j \pi T}{2} t^2 \right) \). One gets by a straightforward calculation

\[
C_f(t, \omega; \varphi_0) = \pi T \delta \left( \frac{t}{T} - \omega \right) \quad \text{or} \\
T \int_{-\infty}^{\infty} \left( \frac{t}{T} - \omega \right)^2 \exp \left( - \frac{1}{4 \alpha} \frac{t^2}{\omega^2} \right) \, dt \, d\omega
\]

according as \( \alpha = 0 \) or \( \alpha \neq 0 \). Note that in all cases \( C_f(t, \omega; \varphi_0) \) is constant along lines \( \omega = \omega_0 \) instead of \( \omega = \omega_0 - \omega_0 \) as \( \omega = \omega_0 \) in \( \mathcal{C}_f(t, \omega; \varphi_0) \) (this can be achieved by choosing \( \varphi_0(t, \omega) = -\omega^2 \) for \( \omega \neq 0 \)). Now let \( \Omega > 0 \) and calculate the convolution of \( C_f(t, \omega; \varphi_0) \) and \( (1/\Omega) \exp \left( - \frac{(t/T)^2}{\Omega^2} \right) \). One gets

\[
\frac{1}{\Omega T} \int_{-\infty}^{\infty} \left( \frac{t}{T} - \omega \right)^2 \exp \left( - \frac{(t - \xi)^2}{\xi^2} \right) \, dt \, d\omega
\]

for some complex constant \( C \). It can be checked that the right-hand side decays rapidly as \( |(t/T) - \xi| \to \infty \). It is also apparent that it takes negative values, no matter how large the product \( \Omega T \) is. Hence, for no value of \( \Omega T \), will smoothing as in (10) and (11) yield a nonnegative result for all \( f(t) \).

**References**


**Efficient Methods to Estimate Correlation Functions of Gaussian Processes and Their Performance Analysis**

TAIHO KOH AND EDWARD J. POWERS

**Abstract—** New efficient methods to estimate crosscorrelation functions of Gaussian signals are studied. In these methods, the “covariance property” of the Gaussian distribution is utilized such that the correlation estimates can be computed with only additions. To evaluate the performances of the methods, exact expressions for the bias and variance of these estimators are formulated and utilized in comparing these methods with the conventional correlation estimator. As a result, we point out that these new methods can give estimates which are comparable to the conventional approach.

**I. INTRODUCTION**

One of the well-known properties of Gaussian distributions is the equivalence between the statistical independence and uncorrelatedness. This latter property, which is a direct consequence of this property, is useful in evaluating the effect of a nonlinear transformation of Gaussian random vectors on their covariance matrix. In the subsequent discussion, \( E \) and \( \langle \cdot \rangle \) denote expectation and transposition, respectively.

**Lemma:** Let \( X \) and \( Y \) be jointly Gaussian random vectors in \( \mathbb{R}^m \) and \( \mathbb{R}^n \), respectively, with zero means and finite second moments in which the \( n \) by \( n \) matrix \( E[YY'] \) is positive definite with its inverse \( E^{-1}[YY'] \). Then, for any Borel function \( G \) from \( \mathbb{R}^m \) to \( \mathbb{R}^n \) with \( E[G'(Y)Y(Y')^T] < \infty \),

\[
E[XG(Y)] = E[XY'] E^{-1}[YY'] E[YG(Y)]
\]

**Proof:** Let \( U = X - AY \) where \( A = E[XY'] \) and \( E^{-1}[YY'] \) is an \( m \) by \( n \) constant matrix. Then \( E[UU'] = 0 \). So \( U \) and \( Y \) are independent of each other. Therefore, from the nesting property of conditional expectation, it follows that \( E[U'G(Y)] = E[U] \) \( G(Y) = 0 \). Since \( X = U + AY \), we have \( E[XG(Y)] = AE[YG(Y)] \).

The above lemma is also closely related to the well-known Büssegang’s theorem [1] which shows the invariance (except for a constant)