Article

On the Distribution of the Information Density of Gaussian Random Vectors: Explicit Formulas and Tight Approximations

Jonathan E. W. Huffmann [1] and **Martin Mittelbach** [2,*]

[1] Lehrstuhl für Theoretische Informationstechnik, Fakultät für Elektrotechnik und Informationstechnik, Technische Universität München, 80290 München, Germany; jonathan.huffmann@tum.de

[2] Lehrstuhl für Theoretische Nachrichtentechnik, Fakultät für Elektrotechnik und Informationstechnik, Technische Universität Dresden, 01062 Dresden, Germany

[*] Correspondence: martin.mittelbach@tu-dresden.de

Abstract: Based on the canonical correlation analysis, we derive series representations of the probability density function (PDF) and the cumulative distribution function (CDF) of the information density of arbitrary Gaussian random vectors as well as a general formula to calculate the central moments. Using the general results, we give closed-form expressions of the PDF and CDF and explicit formulas of the central moments for important special cases. Furthermore, we derive recurrence formulas and tight approximations of the general series representations, which allow efficient numerical calculations with an arbitrarily high accuracy as demonstrated with an implementation in PYTHON publicly available on GITLAB. Finally, we discuss the (in)validity of Gaussian approximations of the information density.

Keywords: information density; information spectrum; probability density function; cumulative distribution function; central moments; Gaussian random vector; canonical correlation analysis

1. Introduction and Main Theorems

Let ξ and η be arbitrary random variables on an abstract probability space $(\Omega, \mathcal{F}, \mathrm{P})$ such that the joint distribution $\mathrm{P}_{\xi\eta}$ is absolutely continuous w. r. t. the product $\mathrm{P}_{\xi} \otimes \mathrm{P}_{\eta}$ of the marginal distributions P_{ξ} and P_{η}. If $\frac{\mathrm{dP}_{\xi\eta}}{\mathrm{dP}_{\xi} \otimes \mathrm{P}_{\eta}}$ denotes the Radon–Nikodym derivative of $\mathrm{P}_{\xi\eta}$ w. r. t. $\mathrm{P}_{\xi} \otimes \mathrm{P}_{\eta}$, then

$$\mathrm{i}(\xi; \eta) = \log\left(\frac{\mathrm{dP}_{\xi\eta}}{\mathrm{dP}_{\xi} \otimes \mathrm{P}_{\eta}}(\xi, \eta)\right)$$

is called the information density of ξ and η. The expectation $\mathrm{E}(\mathrm{i}(\xi; \eta)) = I(\xi; \eta)$ of the information density, called mutual information, plays a key role in characterizing the asymptotic channel coding performance in terms of channel capacity. The non-asymptotic performance, however, is determined by the higher-order moments of the information density and its probability distribution. Achievability and converse bounds that allow a finite blocklength analysis of the optimum channel coding rate are closely related to the distribution function of the information density, also called information spectrum by Han and Verdú [1,2]. Moreover, based on the variance of the information density tight second-order finite blocklength approximations of the optimum code rate can be derived for various important channel models. First work on a non-asymptotic information theoretic analysis was already published in the early years of information theory by Shannon [3], Dobrushin [4], and Strassen [5], among others. Due to the seminal work of Polyanskiy et al. [6], considerable progress has been made in this area. The results of [6] on the one hand and the requirements of current and future wireless networks regarding latency and reliability on the other hand stimulated a significant new interest in this type of analysis (Durisi et al. [7]).

The information density $i(\xi;\eta)$ in the case when ξ and η are jointly Gaussian is of special interest due to the prominent role of the Gaussian distribution. Let $\xi = (\xi_1, \xi_2, \ldots, \xi_p)$ and $\eta = (\eta_1, \eta_2, \ldots, \eta_q)$ be real-valued random vectors with nonsingular covariance matrices R_ξ and R_η and cross-covariance matrix $R_{\xi\eta}$ with rank $r = \text{rank}(R_{\xi\eta})$. (For notational convenience, we write vectors as row vectors. However, in expressions where matrix or vector multiplications occur, we consider all vectors as column vectors.) Without loss of generality for the subsequent results, we assume the expectation of all random variables to be zero. If $(\xi_1, \xi_2, \ldots, \xi_p, \eta_1, \eta_2, \ldots, \eta_q)$ is a Gaussian random vector, then Pinsker [8], Ch. 9.6 has shown that the distribution of the information density $i(\xi;\eta)$ coincides with the distribution of the random variable

$$\nu = \frac{1}{2} \sum_{i=1}^{r} \varrho_i \big(\tilde{\xi}_i^2 - \tilde{\eta}_i^2 \big) + I(\xi;\eta). \tag{1}$$

In this representation $\tilde{\xi}_1, \tilde{\xi}_2, \ldots, \tilde{\xi}_r, \tilde{\eta}_1, \tilde{\eta}_2, \ldots, \tilde{\eta}_r$ are independent and identically distributed (i.i.d.) Gaussian random variables with zero mean and unit variance and the mutual information $I(\xi;\eta)$ in (1) has the form

$$I(\xi;\eta) = \frac{1}{2} \sum_{i=1}^{r} \log \left(\frac{1}{1 - \varrho_i^2} \right). \tag{2}$$

Moreover, $\varrho_1 \geq \varrho_2 \geq \ldots \geq \varrho_r > 0$ denote the positive canonical correlations of ξ and η in descending order, which are obtained by a linear method called canonical correlation analysis that yields the maximum correlations between two sets of random variables (see Section 3). The rank r of the cross-covariance matrix $R_{\xi\eta}$ satisfies $0 \leq r \leq \min\{p,q\}$, and for $r = 0$ we have $i(\xi;\eta) \equiv 0$ almost surely and $I(\xi;\eta) = 0$. This corresponds to $P_{\xi\eta} = P_\xi \otimes P_\eta$ and the independence of ξ and η such that the resulting information density is deterministic. Throughout the rest of the paper, we exclude this degenerated case when the information density is considered and assume subsequently the setting and notation introduced above with $r \geq 1$. As customary notation, we further write $\mathbb{R}$, $\mathbb{N}_0$, and $\mathbb{N}$ to denote the set of real numbers, non-negative integers, and positive integers.

Main contributions. Based on (1), we derive in Section 4 series representations of the probability density function (PDF) and the cumulative distribution function (CDF) as well as explicit general formulas for the central moments of the information density $i(\xi;\eta)$ given subsequently in Theorems 1 to 3. The series representations are useful as they allow tight approximations with errors as low as desired by finite sums as shown in Section 5.2. Moreover, we derive recurrence formulas in Section 5.1 that allow efficient numerical calculations of the series representations in Theorems 1 and 2.

Theorem 1 (PDF of information density). *The PDF $f_{i(\xi;\eta)}$ of the information density $i(\xi;\eta)$ is given by*

$$f_{i(\xi;\eta)}(x) = \frac{1}{\varrho_r \sqrt{\pi}} \sum_{k_1=0}^{\infty} \sum_{k_2=0}^{\infty} \cdots \sum_{k_{r-1}=0}^{\infty} \left[\prod_{i=1}^{r-1} \frac{\varrho_r}{\varrho_i} \frac{(2k_i)!}{(k_i!)^2 4^{k_i}} \left(1 - \frac{\varrho_r^2}{\varrho_i^2} \right)^{k_i} \right] \times$$

$$\frac{\text{K}_{\frac{r-1}{2}+k_1+k_2+\cdots+k_{r-1}} \left(\frac{|x-I(\xi;\eta)|}{\varrho_r} \right)}{\Gamma\big(\frac{r}{2} + k_1 + k_2 + \cdots + k_{r-1}\big)} \left(\frac{|x - I(\xi;\eta)|}{2\varrho_r} \right)^{\big(\frac{r-1}{2}+k_1+k_2+\cdots+k_{r-1}\big)}, \quad x \in \mathbb{R}\setminus\{I(\xi;\eta)\}, \tag{3}$$

where $\Gamma(\cdot)$ denotes the gamma function [9], Sec. 5.2.1 and $\text{K}_\alpha(\cdot)$ denotes the modified Bessel function of second kind and order α [9], Sec. 10.25(ii). If $r \geq 2$, then $f_{i(\xi;\eta)}(x)$ is also well defined for $x = I(\xi;\eta)$.

Theorem 2 (CDF of information density). *The CDF $F_{i(\xi;\eta)}$ of the information density $i(\xi;\eta)$ is given by*

$$F_{i(\xi;\eta)}(x) = \begin{cases} \dfrac{1}{2} - V(I(\xi;\eta) - x) & \text{if} \quad x \le I(\xi;\eta) \\[2mm] \dfrac{1}{2} + V(x - I(\xi;\eta)) & \text{if} \quad x > I(\xi;\eta) \end{cases},$$

with $V(z)$ defined by

$$V(z) = \sum_{k_1=0}^{\infty} \sum_{k_2=0}^{\infty} \cdots \sum_{k_{r-1}=0}^{\infty} \left[\prod_{i=1}^{r-1} \frac{\varrho_r}{\varrho_i} \frac{(2k_i)!}{(k_i!)^2 4^{k_i}} \left(1 - \frac{\varrho_r^2}{\varrho_i^2} \right)^{k_i} \right] \frac{z}{2\varrho_r} \times$$
$$\left[K_{\frac{r-1}{2}+k_1+k_2+\cdots+k_{r-1}}\left(\frac{z}{\varrho_r} \right) L_{\frac{r-3}{2}+k_1+k_2+\cdots+k_{r-1}}\left(\frac{z}{\varrho_r} \right) \right.$$
$$\left. + K_{\frac{r-3}{2}+k_1+k_2+\cdots+k_{r-1}}\left(\frac{z}{\varrho_r} \right) L_{\frac{r-1}{2}+k_1+k_2+\cdots+k_{r-1}}\left(\frac{z}{\varrho_r} \right) \right], \quad z \ge 0, \tag{4}$$

where $L_\alpha(\cdot)$ denotes the modified Struve L function of order α [9], Sec. 11.2.

The method to obtain the result in Theorem 1 is adopted from Mathai [10], where a series representation of the PDF of the sum of independent gamma distributed random variables is derived. Previous work of Grad and Solomon [11] and Kotz et al. [12] goes in a similar direction as Mathai [10]; however, it is not directly applicable since only the restriction to positive series coefficients is considered there. Using Theorem 1, the series representation of the CDF of the information density in Theorem 2 is obtained. The details of the derivations of Theorems 1 and 2 are provided in Section 4.

Theorem 3 (Central moments of information density). *The m-th central moment $E\big([i(\xi;\eta) - I(\xi;\eta)]^m\big)$ of the information density $i(\xi;\eta)$ is given by*

$$E\big([i(\xi;\eta) - I(\xi;\eta)]^m\big)$$
$$= \begin{cases} \displaystyle\sum_{(m_1,m_2,\cdots,m_r) \in \mathcal{K}_{m,r}^{[2]}} m! \prod_{i=1}^{r} \frac{(2m_i)!}{4^{m_i}(m_i!)^2} \varrho_i^{2m_i} & \text{if} \quad m = 2\tilde{m} \\[4mm] 0 & \text{if} \quad m = 2\tilde{m} - 1 \end{cases}, \tag{5}$$

for all $\tilde{m} \in \mathbb{N}$, where $\mathcal{K}_{m,r}^{[2]} = \big\{ (m_1, m_2, \ldots, m_r) \in \mathbb{N}_0^r : 2m_1 + 2m_2 + \cdots + 2m_r = m \big\}$.

Pinsker [8], Eq. (9.6.17) provided a formula for $\sum_{i=1}^{r} E\big(\big[\frac{\varrho_i}{2}(\tilde{\xi}_i^2 - \tilde{\eta}_i^2) \big]^m \big)$, which he called "derived m-th central moment" of the information density, where $\tilde{\xi}_i$ and $\tilde{\eta}_i$ are given as in (1). These special moments coincide for $m = 2$ with the usual central moments considered in Theorem 3.

The rest of the paper is organized as follows: In Section 2, we discuss important special cases which allow simplified and explicit formulas. In Section 3, we provide some background on the canonical correlation analysis and its application to the calculation of the information density and mutual information for Gaussian random vectors. The proofs of the main Theorems 1 to 3 are given in Section 4. Recurrence formulas, finite sum approximations, and uniform bounds of the approximation error are derived in Section 5, which allow efficient and accurate numerical calculations of the PDF and CDF of the information density. Some examples and illustrations are provided in Section 6, where also the (in)validity of Gaussian approximations is discussed. Finally, Section 7 summarizes the paper. Note that a first version of this paper was published on ARXIV as preprint [13].

2. Special Cases

2.1. Equal Canonical Correlations

A simple but important special case for which the series representations in Theorems 1 and 2 simplify to a single summand and the sum of products in Theorem 3 simplifies to a single product is considered in the following corollary.

Corollary 1 (PDF, CDF, and central moments of information density for equal canonical correlations). *If all canonical correlations are equal, i.e.,*

$$\varrho_1 = \varrho_2 = \ldots = \varrho_r,$$

then we have the following simplifications.
(i) The PDF $f_{\mathrm{i}(\xi;\eta)}$ of the information density $\mathrm{i}(\xi;\eta)$ simplifies to

$$f_{\mathrm{i}(\xi;\eta)}(x) = \frac{1}{\varrho_r\sqrt{\pi}\Gamma\left(\frac{r}{2}\right)}\mathrm{K}_{\frac{r-1}{2}}\left(\frac{|x - I(\xi;\eta)|}{\varrho_r}\right)\left(\frac{|x - I(\xi;\eta)|}{2\varrho_r}\right)^{\frac{r-1}{2}}, \quad x \in \mathbb{R}\backslash\{I(\xi;\eta)\}, \quad (6)$$

where $I(\xi;\eta)$ is given by

$$I(\xi;\eta) = -\frac{r}{2}\log\left(1 - \varrho_r^2\right).$$

If $r \geq 2$, then $f_{\mathrm{i}(\xi;\eta)}(x)$ is also well defined for $x = I(\xi;\eta)$.
(ii) The CDF $F_{\mathrm{i}(\xi;\eta)}$ of the information density $\mathrm{i}(\xi;\eta)$ is given by

$$F_{\mathrm{i}(\xi;\eta)}(x) = \begin{cases} \dfrac{1}{2} - \mathrm{V}(I(\xi;\eta) - x) & \text{if} \quad x \leq I(\xi;\eta) \\ \dfrac{1}{2} + \mathrm{V}(x - I(\xi;\eta)) & \text{if} \quad x > I(\xi;\eta) \end{cases}, \quad (7)$$

with $\mathrm{V}(z)$ defined by

$$\mathrm{V}(z) = \frac{z}{2\varrho_r}\left[\mathrm{K}_{\frac{r-1}{2}}\left(\frac{z}{\varrho_r}\right)\mathrm{L}_{\frac{r-3}{2}}\left(\frac{z}{\varrho_r}\right) + \mathrm{K}_{\frac{r-3}{2}}\left(\frac{z}{\varrho_r}\right)\mathrm{L}_{\frac{r-1}{2}}\left(\frac{z}{\varrho_r}\right)\right], \quad z \geq 0. \quad (8)$$

(iii) The m-th central moment $\mathrm{E}\left([\mathrm{i}(\xi;\eta) - I(\xi;\eta)]^m\right)$ of the information density $\mathrm{i}(\xi;\eta)$ has the form

$$\mathrm{E}\left([\mathrm{i}(\xi;\eta) - I(\xi;\eta)]^m\right) = \begin{cases} \dfrac{m!}{(m/2)!}\left(\prod_{j=1}^{m/2}\left(\frac{r}{2} + j - 1\right)\right)\varrho_r^m & \text{if} \ \ m = 2\tilde{m} \\ 0 & \text{if} \ \ m = 2\tilde{m} - 1 \end{cases}, \quad (9)$$

for all $\tilde{m} \in \mathbb{N}$.

Clearly, if all canonical correlations are equal, then the only nonzero term in the series (3) and (4) occur for $k_1 = k_2 = \ldots = k_{r-1} = 0$. For this single summand, the product in squared brackets in (3) and (4) is equal to 1 by applying $0^0 = 1$, which yields the results of part (i) and (ii) in Corollary 1. Details of the derivation of part (iii) of the corollary are provided in Section 4.

Note, if all canonical correlations are equal, then we can rewrite (1) as follows:

$$\nu = \frac{\varrho_r}{2}\left(\sum_{i=1}^{r}\tilde{\xi}_i^2 - \sum_{i=1}^{r}\tilde{\eta}_i^2\right) + I(\xi;\eta).$$

This implies that ν coincides with the distribution of the random variable

$$\nu_* = \frac{\varrho_r}{2}(\zeta_1 - \zeta_2) + I(\xi; \eta),$$

where ζ_1 and ζ_2 are i.i.d. χ^2-distributed random variables with r degrees of freedom. With this representation, we can obtain the expression of the PDF given in (6) also from [14], Sec. 4.A.4.

Special cases of Corollary 1. The case when all canonical correlations are equal is important because it occurs in various situations. The subsequent cases follow from the properties of canonical correlations given in Section 3.

(i) Assume that the random variables $\xi_1, \xi_2, \ldots, \xi_p, \eta_1, \eta_2, \ldots, \eta_q$ are pairwise uncorrelated with the exception of the pairs $(\xi_i, \eta_i), i = 1, 2, \ldots, k \leq \min\{p, q\}$ for which we have $\mathrm{cor}(\xi_i, \eta_i) = \rho \neq 0$, where $\mathrm{cor}(\cdot, \cdot)$ denotes the Pearson correlation coefficient. Then, $r = k$ and $\varrho_i = |\rho|$ for all $i = 1, 2, \ldots, r$. Note, if $p = q = k$, then for the previous conditions to hold, it is sufficient that the two-dimensional random vectors (ξ_i, η_i) are i.i.d. However, the identical distribution of the (ξ_i, η_i)'s is not necessary. In Laneman [15], the distribution of the information density for an additive white Gaussian noise channel with i.i.d. Gaussian input is determined. This is a special case of the case with i.i.d. random vectors (ξ_i, η_i) just mentioned. In Wu and Jindal [16] and in Buckingham and Valenti [17], an approximation of the information density by a Gaussian random variable is considered for the setting in [15]. A special case very similar to that in [15] is also considered in Polyanskiy et al. [6], Sec. III.J. To the best of the authors' knowledge, explicit formulas for the general case as considered in this paper are not available yet in the literature.

(ii) Assume that the conditions of part (i) are satisfied. Furthermore, assume that $\hat{A}$ is a real nonsingular matrix of dimension $p \times p$ and $\hat{B}$ is a real nonsingular matrix of dimension $q \times q$. Then, the random vectors

$$\hat{\xi} = \hat{A}\,\xi \qquad \text{and} \qquad \hat{\eta} = \hat{B}\,\eta$$

have the same canonical correlations as the random vectors ξ and η, i.e., $\varrho_i = |\rho|$ for all $i = 1, 2, \ldots, k \leq \min\{p, q\}$.

(iii) If $r = 1$, i.e., if the cross-covariance matrix $R_{\xi, \eta}$ has rank 1, then Corollary 1 obviously applies. Clearly, the most simple special case with $r = 1$ occurs for $p = q = 1$, where $\varrho_1 = |\mathrm{cor}(\xi_1, \eta_1)|$.

As a simple multivariate example, let the covariance matrix of the random vector $(\xi_1, \xi_2, \ldots, \xi_p, \eta_1, \eta_2, \ldots, \eta_q)$ be given by the Kac-Murdock–Szegö matrix

$$\begin{pmatrix} R_\xi & R_{\xi\eta} \\ R_{\xi\eta} & R_\eta \end{pmatrix} = \left(\rho^{|i-j|} \right)_{i,j=1}^{p+q}$$

which is related to the covariance function of a first-order autoregressive process, where $0 < |\rho| < 1$. Then, $r = \mathrm{rank}(R_{\xi\eta}) = 1$ and $\varrho_1 = |\rho|$.

(iv) As yet another example, assume $p = q$ and $R_{\xi\eta} = \rho R_\xi^{1/2} R_\eta^{1/2}$ for some $0 < |\rho| < 1$. Then, $\varrho_i = |\rho|$ for $i = 1, 2, \ldots, r = q$. Here, $A^{1/2}$ denotes the square root of the real-valued positive semidefinite matrix A, i.e., the unique positive semidefinite matrix B such that $BB = A$.

2.2. More on Special Cases with Simplified Formulas

Let us further evaluate the formulas given in Corollary 1 and Theorem 3 for some relevant parameter values.

(i) *Single canonical correlation coefficient.* In the most simple case, there is only a single non-zero canonical correlation coefficient, i.e., $r = 1$. (Recall, in the beginning of the paper,

we have excluded the degenerated case when all canonical correlations are zero.) Then, the formulas of the PDF and the m-th central moment in Corollary 1 simplify to the form

$$f_{i(\xi;\eta)}(x) = \frac{1}{\varrho_1 \pi} K_0\left(\frac{|x - I(\xi;\eta|)}{\varrho_1}\right), \quad x \in \mathbb{R}\backslash\{I(\xi;\eta)\},$$

and

$$E\left([i(\xi;\eta) - I(\xi;\eta)]^m\right) = \begin{cases} \left(\dfrac{m!}{(m/2)!}\right)^2 \left(\dfrac{\varrho_1}{2}\right)^m & \text{if } m = 2\tilde{m} \\ 0 & \text{if } m = 2\tilde{m} - 1 \end{cases}, \tag{10}$$

for all $\tilde{m} \in \mathbb{N}$. A formula equivalent to (10) is also provided by Pinsker [8], Lemma 9.6.1 who considered the special case $p = q = 1$, which implies $r = 1$.

(ii) *Second and fourth central moment.* To demonstrate how the general formula given in Theorem 3 is used, we first consider $m = 2$. In this case, the summation indices $m_1, m_2, \ldots, m_r$ have to satisfy $m_i = 1$ for a single $i \in \{1, 2, \ldots, r\}$, whereas the remaining m_i's have to be zero. Thus, (5) evaluates for $m = 2$ to

$$E\left([i(\xi;\eta) - I(\xi;\eta)]^2\right) = \text{var}\left(i(\xi;\eta)\right) = \sum_{i=1}^{r} \varrho_i^2. \tag{11}$$

As a slightly more complex example, let $m = 4$. In this case, either we have $m_i = 2$ for a single $i \in \{1, 2, \ldots, r\}$, whereas the remaining m_i's are zero or we have $m_{i_1} = m_{i_2} = 1$ for two $i_1 \neq i_2 \in \{1, 2, \ldots, r\}$, whereas the remaining m_i's have to be zero. Thus, (5) evaluates for $m = 4$ to

$$E\left([i(\xi;\eta) - I(\xi;\eta)]^4\right) = 9 \sum_{i=1}^{r} \varrho_i^4 + 6 \sum_{i=2}^{r} \sum_{j=1}^{i-1} \varrho_i^2 \varrho_j^2.$$

(iii) *Even number of equal canonical correlations.* As in Corollary 1, assume that all canonical correlations are equal and additionally assume that the number r of canonical correlations is even, i.e., $r = 2\tilde{r}$ for some $\tilde{r} \in \mathbb{N}$. Then, we can use [9], Secs. 10.47.9, 10.49.1, and 10.49.12 to obtain the following relation for the modified Bessel function $K_\alpha(\cdot)$ of a second kind and order α

$$K_{\frac{r-1}{2}}(y) = \sqrt{\frac{\pi}{2}} \exp(-y) \sum_{i=0}^{r/2-1} \frac{(r/2 - 1 + i)!}{(r/2 - 1 - i)!\, i!\, 2^i} y^{-(i+\frac{1}{2})}, \quad y \in (0, \infty). \tag{12}$$

Plugging (12) into (6) and rearranging terms yields the following expression for the PDF of the information density:

$$f_{i(\xi;\eta)}(x) = \frac{1}{\varrho_r 2^{r-1}(r/2 - 1)!} \exp\left(-\frac{|x - I(\xi;\eta)|}{\varrho_r}\right) \times$$

$$\sum_{i=0}^{r/2-1} \frac{(2(r/2 - 1) - i)!\, 2^i}{(r/2 - 1 - i)!\, i!} \left(\frac{|x - I(\xi;\eta)|}{\varrho_r}\right)^i, \quad x \in \mathbb{R}.$$

By integration, we obtain for the function $V(\cdot)$ in (8) the expression

$$V(z) = \frac{1}{2} - \frac{1}{2^{r-1}(r/2 - 1)!} \exp\left(-\frac{z}{\varrho_r}\right) \times$$

$$\sum_{i=0}^{r/2-1} \frac{(2(r/2 - 1) - i)!\, 2^i}{(r/2 - 1 - i)!} \sum_{j=0}^{i} \frac{1}{(i-j)!} \left(\frac{z}{\varrho_r}\right)^{i-j}, \quad z \geq 0.$$

Note that these special formulas can also be obtained directly from the results given in [14], Sec. 4.A.3.

To illustrate the principal behavior of the PDF and CDF of the information density for equal canonical correlations, it is instructive to consider the specific value $r = 2$ in the above formulas, which yields

$$f_{i(\xi;\eta)}(x) = \frac{1}{2\varrho_r} \exp\left(-\frac{|x - I(\xi;\eta)|}{\varrho_r}\right), \quad x \in \mathbb{R},$$

$$V(z) = \frac{1}{2}\left(1 - \exp\left(-\frac{z}{\varrho_r}\right)\right), \qquad z \geq 0$$

and $r = 4$, for which we obtain

$$f_{i(\xi;\eta)}(x) = \frac{1}{4\varrho_r} \exp\left(-\frac{|x - I(\xi;\eta)|}{\varrho_r}\right)\left(1 + \frac{|x - I(\xi;\eta)|}{\varrho_r}\right), \quad x \in \mathbb{R},$$

$$V(z) = \frac{1}{2}\left(1 - \exp\left(-\frac{z}{\varrho_r}\right)\left(1 + \frac{z}{2\varrho_r}\right)\right), \qquad z \geq 0.$$

3. Mutual Information and Information Density in Terms of Canonical Correlations

First introduced by Hotelling [18], the canonical correlation analysis is a widely used linear method in multivariate statistics to determine the maximum correlations between two sets of random variables. It allows a particularly simple and useful representation of the mutual information and the information density of Gaussian random vectors in terms of the so-called canonical correlations. This representation was first obtained by Gelfand and Yaglom [19] and further extended by Pinsker [8], Ch. 9. For the convenience of the reader, we summarize in this section the essence of the canonical correlation analysis and demonstrate how it is applied to derive the representations in (1) and (2).

The formulation of the canonical correlation analysis given below is particularly suitable for implementations. The corresponding results are given without proof. Details and thorough discussions can be found, e.g., in Härdle and Simar [20], Koch [21], or Timm [22].

Based on the nonsingular covariance matrices R_ξ and R_η of the random vectors $\xi = (\xi_1, \xi_2, \ldots, \xi_p)$ and $\eta = (\eta_1, \eta_2, \ldots, \eta_q)$, and the cross-covariance matrix $R_{\xi\eta}$ with rank $r = \text{rank}(R_{\xi\eta})$ satisfying $0 \leq r \leq \min\{p, q\}$, define the matrix

$$M = R_\xi^{-\frac{1}{2}} R_{\xi\eta} R_\eta^{-\frac{1}{2}},$$

where the inverse matrices $R_\xi^{-1/2} = \left(R_\xi^{1/2}\right)^{-1}$ and $R_\eta^{-1/2} = \left(R_\eta^{1/2}\right)^{-1}$ can be obtained from diagonalizing R_ξ and R_η. Then, the matrix M has a singular value decomposition

$$M = UDV^{\text{T}},$$

where V^{T} denotes the transpose of V. The only non-zero entries $d_{1,1}, d_{2,2}, \ldots, d_{r,r} > 0$ of the matrix $D = (d_{i,j})_{i,j=1}^{p,q}$ are called canonical correlations of ξ and η, denoted by $\varrho_i = d_{i,i}, i = 1, 2, \ldots, r$. The singular value decomposition can be chosen such that $\varrho_1 \geq \varrho_2 \geq \ldots \geq \varrho_r$ holds, which is assumed throughout the paper.

Define the random vectors

$$\hat{\xi} = (\hat{\xi}_1, \hat{\xi}_2, \ldots, \hat{\xi}_p) = A\,\xi \qquad \text{and} \qquad \hat{\eta} = (\hat{\eta}_1, \hat{\eta}_2, \ldots, \hat{\eta}_q) = B\,\eta,$$

where the nonsingular matrices A and B are given by

$$A = U^{\text{T}} R_\xi^{-\frac{1}{2}} \qquad \text{and} \qquad B = V^{\text{T}} R_\eta^{-\frac{1}{2}}.$$

Then, the random variables $\hat{\xi}_1, \hat{\xi}_2, \ldots, \hat{\xi}_p, \hat{\eta}_1, \hat{\eta}_2, \ldots, \hat{\eta}_q$ have unit variance and they are pairwise uncorrelated with the exception of the pairs $(\hat{\xi}_i, \hat{\eta}_i), i = 1, 2, \ldots, r$ for which we have $\mathrm{cor}(\hat{\xi}_i, \hat{\eta}_i) = \varrho_i$.

Using these results, we obtain for the mutual information and the information density

$$I(\xi; \eta) = I(A\xi; B\eta) = I(\hat{\xi}; \hat{\eta}) = \sum_{i=1}^{r} I(\hat{\xi}_i; \hat{\eta}_i) \tag{13}$$

$$\mathrm{i}(\xi; \eta) = \mathrm{i}(A\xi; B\eta) = \mathrm{i}(\hat{\xi}; \hat{\eta}) = \sum_{i=1}^{r} \mathrm{i}(\hat{\xi}_i; \hat{\eta}_i) \qquad \text{(P-almost surely).} \tag{14}$$

The first equality in (13) and (14) holds because A and B are nonsingular matrices, which follows, e.g., from Pinsker [8], Th. 3.7.1. Since we consider the case where ξ and η are jointly Gaussian, $\hat{\xi}$ and $\hat{\eta}$ are jointly Gaussian as well. Therefore, the correlation properties of $\hat{\xi}$ and $\hat{\eta}$ imply that all random variables $\hat{\xi}_i, \hat{\eta}_j$ are independent except for the pairs $(\hat{\xi}_i, \hat{\eta}_i)$, $i = 1, 2, \ldots, r$. This implies the last equality in (13) and (14), where $\mathrm{i}(\hat{\xi}_1; \hat{\eta}_1), \mathrm{i}(\hat{\xi}_2; \hat{\eta}_2), \ldots, \mathrm{i}(\hat{\xi}_r; \hat{\eta}_r)$ are independent. The sum representations follow from the chain rules of mutual information and information density and the equivalence between independence and vanishing mutual information and information density.

Since $\hat{\xi}_i$ and $\hat{\eta}_i$ are jointly Gaussian with correlation $\mathrm{cor}(\hat{\xi}_i, \hat{\eta}_i) = \varrho_i$, we obtain from (13) and the formula of mutual information for the bivariate Gaussian case the identity (2). Additionally, with $\hat{\xi}_i$ and $\hat{\eta}_i$ having zero mean and unit variance, the information density $\mathrm{i}(\hat{\xi}_i; \hat{\eta}_i)$ is further given by

$$\mathrm{i}(\hat{\xi}_i; \hat{\eta}_i) = -\frac{1}{2} \log(1 - \varrho_i^2) - \frac{\varrho_i^2}{2(1 - \varrho_i^2)} \left(\hat{\xi}_i^2 - \frac{2\, \hat{\xi}_i \hat{\eta}_i}{\varrho_i} + \hat{\eta}_i^2 \right), \quad i = 1, 2, \ldots, r. \tag{15}$$

Now assume $\tilde{\xi}_1, \tilde{\xi}_2, \ldots, \tilde{\xi}_r, \tilde{\eta}_1, \tilde{\eta}_2, \ldots, \tilde{\eta}_r$ are i.i.d. Gaussian random variables with zero mean and unit variance. Then, the distribution of the random vector

$$\frac{1}{\sqrt{2}} \left(\sqrt{1 + \varrho_i}\, \tilde{\xi}_i + \sqrt{1 - \varrho_i}\, \tilde{\eta}_i, \; \sqrt{1 + \varrho_i}\, \tilde{\xi}_i - \sqrt{1 - \varrho_i}\, \tilde{\eta}_i \right)$$

coincides with the distribution of the random vector $(\hat{\xi}_i, \hat{\eta}_i)$ for all $i = 1, 2, \ldots, r$. Plugging this into (15), we obtain together with (14) that the distribution of the information density $\mathrm{i}(\xi; \eta)$ coincides with the distribution of (1).

4. Proof of Main Results

4.1. Auxiliary Results

To prove Theorem 1, the following lemma regarding the characteristic function of the information density is utilized. The results of the lemma are also used in Ibragimov and Rozanov [23] but without proof. Therefore, the proof is given below for completeness.

Lemma 1 (Characteristic function of (shifted) information density)**.** *The characteristic function of the shifted information density* $\mathrm{i}(\xi; \eta) - I(\xi; \eta)$ *is equal to the characteristic function of the random variable*

$$\tilde{v} = \frac{1}{2} \sum_{i=1}^{r} \varrho_i \big(\tilde{\xi}_i^2 - \tilde{\eta}_i^2 \big), \tag{16}$$

where $\tilde{\xi}_1, \tilde{\xi}_2, \ldots, \tilde{\xi}_r, \tilde{\eta}_1, \tilde{\eta}_2, \ldots, \tilde{\eta}_r$ *are i.i.d. Gaussian random variables with zero mean and unit variance, and* $\varrho_1, \varrho_2, \ldots, \varrho_r$ *are the canonical correlations of* ξ *and* η. *The characteristic function of* $\tilde{v}$ *is given by*

$$\varphi_{\tilde{v}}(t) = \prod_{i=1}^{r} \frac{1}{\sqrt{1 + \varrho_i^2 t^2}}, \quad t \in \mathbb{R}. \tag{17}$$

Proof. Due to (1), the distribution of the shifted information density $\mathrm{i}(\xi;\eta) - I(\xi;\eta)$ coincides with the distribution of the random variable $\tilde{v}$ in (16) such that the characteristic functions of $\mathrm{i}(\xi;\eta) - I(\xi;\eta)$ and $\tilde{v}$ are equal.

It is a well known fact that $\tilde{\xi}_i^2$ and $\tilde{\eta}_i^2$ in (16) are chi-squared distributed random variables with one degree of freedom from which we obtain that the weighted random variables $\varrho_i \tilde{\xi}_i^2/2$ and $\varrho_i \tilde{\eta}_i^2/2$ are gamma distributed with a scale parameter of $1/\varrho_i$ and shape parameter of $1/2$. The characteristic function of these random variables therefore admits the form

$$\varphi_{\frac{\varrho_i}{2}\tilde{\xi}_i^2}(t) = (1 - \varrho_i j t)^{-\frac{1}{2}}.$$

Further, from the identity $\varphi_{-\varrho_i \tilde{\xi}_i^2/2}(t) = \varphi_{\varrho_i \tilde{\xi}_i^2/2}(-t)$ for the characteristic function and from the independence of $\tilde{\xi}_i$ and $\tilde{\eta}_i$, we obtain the characteristic function of $\tilde{v}_i = \varrho_i(\tilde{\xi}_i^2 - \tilde{\eta}_i^2)/2$ to be given by

$$\varphi_{\tilde{v}_i}(t) = (1 - \varrho_i j t)^{-\frac{1}{2}}(1 + \varrho_i j t)^{-\frac{1}{2}} = \left(1 + \varrho_i^2 t^2\right)^{-\frac{1}{2}}.$$

Finally, because $\tilde{v}$ in (16) is given by the sum of the independent random variables $\tilde{v}_i$, the characteristic function of $\tilde{v}$ results from multiplying the individual characteristic functions of the random variables $\tilde{v}_i$. By doing so, we obtain (17). $\qquad\square$

As further auxiliary result, the subsequent proposition providing properties of the modified Bessel function K_α of second kind and order α will be used to prove the main results.

Proposition 1 (Properties related to the function K_α). *For all $\alpha \in \mathbb{R}$, the function*

$$y \mapsto y^\alpha \mathrm{K}_\alpha(y), \quad y \in (0, \infty),$$

where $\mathrm{K}_\alpha(\cdot)$ denotes the modified Bessel function of second kind and order α [9], Sec. 10.25(ii), is strictly positive and strictly monotonically decreasing. Furthermore, if $\alpha > 0$, then we have

$$\lim_{y \to +0} y^\alpha \mathrm{K}_\alpha(y) = \sup_{y \in (0,\infty)} y^\alpha \mathrm{K}_\alpha(y) = \Gamma(\alpha) 2^{\alpha-1}. \tag{18}$$

Proof. If $\alpha \in \mathbb{R}$ is fixed, then $\mathrm{K}_\alpha(y)$ is strictly positive and strictly monotonically decreasing w. r. t. $y \in (0, \infty)$ due to [9], Secs. 10.27.3 and 10.37. Furthermore, we obtain

$$\frac{\mathrm{d} y^\alpha \mathrm{K}_\alpha(y)}{\mathrm{d} y} = -y^\alpha \mathrm{K}_{\alpha-1}(y), \quad y \in (0, \infty)$$

by applying the rules to calculate derivatives of Bessel functions given in [9], Sec. 10.29(ii). It follows that $y^\alpha \mathrm{K}_\alpha(y)$ is strictly positive and strictly monotonically decreasing w. r. t. $y \in (0, \infty)$ for all fixed $\alpha \in \mathbb{R}$.

Consider now the Basset integral formula as given in [9], Sec. 10.32.11

$$\mathrm{K}_\alpha(yz) = \frac{\Gamma\left(\alpha + \frac{1}{2}\right)(2z)^\alpha}{y^\alpha \sqrt{\pi}} \int_{u=0}^{\infty} \frac{\cos(uy)}{(u^2 + z^2)^{\alpha+\frac{1}{2}}} \, \mathrm{d}u \tag{19}$$

for $|\arg(z)| < \pi/2$, $y > 0$, $\alpha > -\frac{1}{2}$ and the integral

$$\int_{u=0}^{\infty} \frac{1}{(u^2 + 1)^{\alpha+\frac{1}{2}}} \, \mathrm{d}u = \frac{\sqrt{\pi}\, \Gamma(\alpha)}{2\,\Gamma\left(\alpha + \frac{1}{2}\right)} \tag{20}$$

for $\alpha > 0$, where the equality holds due to [24], Secs. 3.251.2 and 8.384.1. Using (19) and (20), we obtain

$$
\lim_{y \to +0} y^\alpha K_\alpha(y) = \lim_{y \to +0} \frac{\Gamma\left(\alpha + \frac{1}{2}\right) 2^\alpha}{\sqrt{\pi}} \int_{u=0}^\infty \frac{\cos(uy)}{\left(u^2 + 1\right)^{\alpha + \frac{1}{2}}} \, \mathrm{d}u
$$

$$
= \frac{\Gamma\left(\alpha + \frac{1}{2}\right) 2^\alpha}{\sqrt{\pi}} \int_{u=0}^\infty \frac{1}{\left(u^2 + 1\right)^{\alpha + \frac{1}{2}}} \, \mathrm{d}u = \Gamma(\alpha) 2^{\alpha - 1},
$$

for all $\alpha > 0$, where we also applied the dominated convergence theorem, which is possible due to $\cos(uy) / \left(u^2 + 1\right)^{\alpha + 1/2} \le 1 / \left(u^2 + 1\right)^{\alpha + 1/2}$. Using the previously derived monotonicity, we obtain (18). $\square$

4.2. Proof of Theorem 1

To prove Theorem 1, we calculate the PDF $f_{\tilde{v}}$ of the random variable $\tilde{v}$ introduced in Lemma 1 by inverting the characteristic function $\varphi_{\tilde{v}}$ given in (17) via the integral

$$
f_{\tilde{v}}(v) = \frac{1}{2\pi} \int_{-\infty}^\infty \varphi_{\tilde{v}}(t) \exp\left(-\jmath t v\right) \mathrm{d}t, \quad v \in \mathbb{R}. \tag{21}
$$

Shifting the PDF of $\tilde{v}$ by $I(\xi; \eta)$, we obtain the PDF $f_{\mathrm{i}(\xi; \eta)} = f_{\tilde{v}}(x - I(\xi; \eta))$, $x \in \mathbb{R}$, of the information density $\mathrm{i}(\xi; \eta)$.

The method used subsequently is based on the work of Mathai [10]. To invert the characteristic function $\varphi_{\tilde{v}}$, we expand the factors in (17) as

$$
\left(1 + \varrho_i^2 t^2\right)^{-\frac{1}{2}} = \left(1 + \varrho_r^2 t^2\right)^{-\frac{1}{2}} \frac{\varrho_r}{\varrho_i} \left(1 + \left(\frac{\varrho_r^2}{\varrho_i^2} - 1\right)\left(1 + \varrho_r^2 t^2\right)^{-1}\right)^{-\frac{1}{2}} \tag{22}
$$

$$
= \left(1 + \varrho_r^2 t^2\right)^{-\frac{1}{2}} \sum_{k=0}^\infty (-1)^k \binom{-1/2}{k} \frac{\varrho_r}{\varrho_i} \left(1 - \frac{\varrho_r^2}{\varrho_i^2}\right)^k \left(1 + \varrho_r^2 t^2\right)^{-k}. \tag{23}
$$

In (23), we have used the binomial series

$$
(1 + y)^a = \sum_{k=0}^\infty \binom{a}{k} y^k \tag{24}
$$

where $a \in \mathbb{R}$. The series is absolutely convergent for $|y| < 1$ and

$$
\binom{a}{k} = \prod_{\ell=1}^k \frac{a - \ell + 1}{\ell}, \qquad k \in \mathbb{N}, \tag{25}
$$

denotes the generalized binomial coefficient with $\binom{a}{0} = 1$. Since

$$
\left| \left(1 - \frac{\varrho_r^2}{\varrho_i^2}\right)\left(1 + \varrho_r^2 t^2\right)^{-1} \right| < 1 \tag{26}
$$

holds for all $t \in \mathbb{R}$, the series in (23) is absolutely convergent for all $t \in \mathbb{R}$. Using the expansion in (23) and the absolute convergence together with the identity

$$
\binom{-1/2}{k} = \frac{(-1)^k (2k)!}{(k!)^2 4^k} \tag{27}
$$

we can rewrite the characteristic function $\varphi_{\tilde{v}}$ as

$$\varphi_{\tilde{v}}(t) = \sum_{k_1=0}^{\infty} \sum_{k_2=0}^{\infty} \cdots \sum_{k_{r-1}=0}^{\infty} \left[\prod_{i=1}^{r-1} \frac{\varrho_r}{\varrho_i} \frac{(2k_i)!}{(k_i!)^2 4^{k_i}} \left(1 - \frac{\varrho_r^2}{\varrho_i^2}\right)^{k_i} \right] \times$$
$$\left(1 + \varrho_r^2 t^2\right)^{-\left(\frac{r}{2} + k_1 + k_2 + \cdots + k_{r-1}\right)}, \quad t \in \mathbb{R}. \tag{28}$$

To obtain the PDF $f_{\tilde{v}}$, we evaluate the inversion integral (21) based on the series representation in (28). Since every series in (28) is absolutely convergent, we can exchange summation and integration. Let $\beta = \frac{r}{2} + k_1 + k_2 + \cdots + k_{r-1}$. Then, by symmetry, we have for the integral of a summand

$$\int_{t=-\infty}^{\infty} \frac{\exp(-\jmath t v)}{(1 + \varrho_r^2 t^2)^{\beta}} \, dt = 2 \int_{t=0}^{\infty} \frac{\cos(t v)}{(1 + \varrho_r^2 t^2)^{\beta}} \, dt = \frac{2}{\varrho_r} \int_{u=0}^{\infty} \frac{\cos(u v / \varrho_r)}{(1 + u^2)^{\beta}} \, du, \tag{29}$$

where the second equality is a result of the substitution $t = u/\varrho_r$. By setting $z = 1$, $\alpha = \beta - \frac{1}{2} \geq 0$ and $y = v/\varrho_r$ in the Basset integral formula given in (19) in the proof of Proposition 1 and using the symmetry with respect to v, we can evaluate (29) to the following form:

$$\int_{t=-\infty}^{\infty} \frac{\exp(-\jmath t v)}{(1 + \varrho_r^2 t^2)^{\beta}} \, dt = \frac{\sqrt{\pi}}{\Gamma(\beta) 2^{\beta - \frac{3}{2}} \varrho_r^{\beta + \frac{1}{2}}} \mathrm{K}_{\beta - \frac{1}{2}} \left(\frac{|v|}{\varrho_r}\right) |v|^{\beta - \frac{1}{2}}, \quad v \in \mathbb{R} \backslash \{0\}. \tag{30}$$

Combining (21), (28), and (30) yields

$$f_{\tilde{v}}(v) = \frac{1}{2\sqrt{\pi}} \sum_{k_1=0}^{\infty} \sum_{k_2=0}^{\infty} \cdots \sum_{k_{r-1}=0}^{\infty} \left[\prod_{i=1}^{r-1} \frac{\varrho_r}{\varrho_i} \frac{(2k_i)!}{(k_i!)^2 4^{k_i}} \left(1 - \frac{\varrho_r^2}{\varrho_i^2}\right)^{k_i} \right] \times$$
$$\frac{\mathrm{K}_{\frac{r-1}{2} + k_1 + k_2 + \cdots + k_{r-1}} \left(\frac{|v|}{\varrho_r}\right) |v|^{\left(\frac{r-1}{2} + k_1 + k_2 + \cdots + k_{r-1}\right)}}{\Gamma\left(\frac{r}{2} + k_1 + k_2 + \cdots + k_{r-1}\right) 2^{\left(\frac{r-3}{2} + k_1 + k_2 + \cdots + k_{r-1}\right)} \varrho_r^{\left(\frac{r+1}{2} + k_1 + k_2 + \cdots + k_{r-1}\right)}}, \quad v \in \mathbb{R} \backslash \{0\}. \tag{31}$$

Slightly rearranging terms and shifting $f_{\tilde{v}}(\cdot)$ by $I(\xi; \eta)$ yields (3).

It remains to show that $f_{\mathrm{i}(\xi;\eta)}(x)$ is also well defined for $x = I(\xi; \eta)$ if $r \geq 2$. Indeed, if $r \geq 2$, then we can use Proposition 1 to obtain

$$\lim_{x \to I(\xi;\eta)} f_{\mathrm{i}(\xi;\eta)}(x) = \frac{1}{2\varrho_r \sqrt{\pi}} \sum_{k_1=0}^{\infty} \sum_{k_2=0}^{\infty} \cdots \sum_{k_{r-1}=0}^{\infty} \left[\prod_{i=1}^{r-1} \frac{\varrho_r}{\varrho_i} \frac{(2k_i)!}{(k_i!)^2 4^{k_i}} \left(1 - \frac{\varrho_r^2}{\varrho_i^2}\right)^{k_i} \right] \times$$
$$\frac{\Gamma\left(\frac{r-1}{2} + k_1 + k_2 + \cdots + k_{r-1}\right)}{\Gamma\left(\frac{r-1}{2} + k_1 + k_2 + \cdots + k_{r-1} + \frac{1}{2}\right)}$$

where we used the exchangeability of the limit and the summation due to the absolute convergence of the series. Since $\Gamma(\alpha)/\Gamma(\alpha + \frac{1}{2})$ is decreasing w. r. t. $\alpha \geq \frac{1}{2}$, we have

$$\frac{\Gamma\left(\frac{r-1}{2} + k_1 + k_2 + \cdots + k_{r-1}\right)}{\Gamma\left(\frac{r-1}{2} + k_1 + k_2 + \cdots + k_{r-1} + \frac{1}{2}\right)} \leq \frac{\Gamma\left(\frac{r-1}{2}\right)}{\Gamma\left(\frac{r-1}{2} + \frac{1}{2}\right)} \leq \sqrt{\pi}.$$

Then, with (69) in the proof of Theorem 4, it follows that $\lim_{x \to I(\xi;\eta)} f_{\mathrm{i}(\xi;\eta)}(x)$ exists and is finite. $\square$

4.3. Proof of Theorem 2

To prove Theorem 2, we calculate the CDF $F_{\tilde{v}}$ of the random variable $\tilde{v}$ introduced in Lemma 1 by integrating the PDF $f_{\tilde{v}}$ given in (31). Shifting the CDF of $\tilde{v}$ by $I(\xi; \eta)$, we obtain the CDF $F_{i(\xi;\eta)}(x) = F_{\tilde{v}}(x - I(\xi; \eta))$, $x \in \mathbb{R}$, of the information density $i(\xi; \eta)$. Using the symmetry of $f_{\tilde{v}}$, we can write

$$
F_{\tilde{v}}(z) = \mathrm{P}(\tilde{v} \leq z) = \begin{cases} \dfrac{1}{2} - \displaystyle\int_{v=0}^{-z} f_{\tilde{v}}(v) \, \mathrm{d}v & \text{for} \quad z \leq 0 \\ \dfrac{1}{2} + \displaystyle\int_{v=0}^{z} f_{\tilde{v}}(v) \, \mathrm{d}v & \text{for} \quad z > 0 \end{cases}.
$$

It is therefore sufficient to evaluate the integral

$$
\mathrm{V}(z) := \int_{v=0}^{z} f_{\tilde{v}}(v) \, \mathrm{d}v \tag{32}
$$

for $z \geq 0$. To calculate the integral (32), we plug (31) into (32) and exchange integration and summation, which is justified by the monotone convergence theorem. To evaluate the integral of a summand, consider the following identity

$$
\int_{x=0}^{z} x^{\alpha} \mathrm{K}_{\alpha}(x) \, \mathrm{d}x = 2^{\alpha-1} \sqrt{\pi} \Gamma\left(\alpha + \frac{1}{2}\right) z \left[\mathrm{K}_{\alpha}(z) \mathrm{L}_{\alpha-1}(z) + \mathrm{K}_{\alpha-1}(z) \mathrm{L}_{\alpha}(z) \right] \tag{33}
$$

for $\alpha > -1/2$ given in [25], Sec. 1.12.1.3, where $\mathrm{L}_{\alpha}(\cdot)$ denotes the modified Struve L function of order α [9], Sec. 11.2. Using (33) with $\alpha = \frac{r-1}{2} + k_1 + k_2 + \cdots + k_{r-1} \geq 0$, we obtain (4). $\square$

4.4. Proof of Theorem 3

Using the random variable

$$
\tilde{v} = \sum_{i=1}^{r} \tilde{v}_i \qquad \text{with} \qquad \tilde{v}_i = \frac{\varrho_i}{2}(\tilde{\xi}_i - \tilde{\eta}_i)
$$

introduced in Lemma 1 and the well-known multinomial theorem [9], Sec. 26.4.9

$$
(y_1 + y_2 + \dots y_r)^m = \sum_{(\ell_1, \ell_2, \dots, \ell_r) \in \mathcal{K}_{m,r}} m! \prod_{i=1}^{r} \frac{y_i^{\ell_i}}{\ell_i!},
$$

where $\mathcal{K}_{m,r} = \left\{ (\ell_1, \ell_2, \dots, \ell_r) \in \mathbb{N}_0^r : \ell_1 + \ell_2 + \cdots + \ell_r = m \right\}$, we can write the m-th central moment of the information density $i(\xi; \eta)$ as

$$
\mathrm{E}\left([i(\xi; \eta) - I(\xi; \eta)]^m \right) = \mathrm{E}\left(\left[\sum_{i=1}^{r} \tilde{v}_i \right]^m \right)
$$

$$
= \sum_{(\ell_1, \ell_2, \dots, \ell_r) \in \mathcal{K}_{m,r}} m! \prod_{i=1}^{r} \frac{\mathrm{E}(\tilde{v}_i^{\ell_i})}{\ell_i!}. \tag{34}
$$

To obtain the second equality in (34), we have exchanged expectation and summation and additionally used the identity $\mathrm{E}\left(\prod_{i=1}^{r} \tilde{v}_i^{\ell_i} \right) = \prod_{i=1}^{r} \mathrm{E}\left(\tilde{v}_i^{\ell_i} \right)$, which holds due to the independence of the random variables $\tilde{v}_1, \tilde{v}_2, \dots, \tilde{v}_r$.

Based on the relation between the ℓ-th central moment of a random variable and the ℓ-th derivative of its characteristic function at 0, we further have

$$
\mathrm{E}(\tilde{v}_i^{\ell_i}) = (-j)^{\ell_i} \frac{\mathrm{d}^{\ell_i}}{\mathrm{d}t^{\ell_i}} \varphi_{\tilde{v}_i}(t) \Big|_{t=0}, \tag{35}
$$

where $\varphi_{\tilde{v}_i}(t) = \left(1 + \varrho_i^2 t^2\right)^{-1/2}, t \in \mathbb{R}$, is the characteristic function of the random variable $\tilde{v}_i$ derived in the proof of Lemma 1. As in the proof of Theorem 1, consider now the binomial series expansion using (24)

$$\varphi_{\tilde{v}_i}(t) = \left(1 + \varrho_i^2 t^2\right)^{-\frac{1}{2}} = \sum_{m_i=0}^{\infty} \binom{-1/2}{m_i} (\varrho_i t)^{2m_i}.$$

The series is absolutely convergent for all $t < \varrho_i^{-1}$. Furthermore, consider the Taylor series expansion of the characteristic function $\varphi_{\tilde{v}_i}$ at the point 0

$$\varphi_{\tilde{v}_i}(t) = \sum_{\ell_i=0}^{\infty} \left(\left. \frac{\mathrm{d}^{\ell_i}}{\mathrm{d}t^{\ell_i}} \varphi_{\tilde{v}_i}(t) \right|_{t=0} \right) \frac{t^{\ell_i}}{\ell_i!}.$$

Both series expansions must be identical in an open interval around 0 such that we obtain by comparing the series coefficients

$$\left. \frac{\mathrm{d}^{\ell_i}}{\mathrm{d}t^{\ell_i}} \varphi_{\tilde{v}_i}(t) \right|_{t=0} = \begin{cases} \ell_i! \binom{-1/2}{\ell_i/2} \varrho_i^{\ell_i} & \text{if} \quad \ell_i = 2m_i \\ 0 & \text{if} \quad \ell_i = 2m_i - 1 \end{cases}$$

for all $m_i \in \mathbb{N}$. With this result, (35) evaluates to

$$\mathrm{E}\big(\tilde{v}_i^{\ell_i}\big) = \begin{cases} \dfrac{(\ell_i!)^2}{((\ell_i/2)!)^2 \, 4^{\frac{\ell_i}{2}}} \varrho_i^{\ell_i} & \text{if} \quad \ell_i = 2m_i \\ 0 & \text{if} \quad \ell_i = 2m_i - 1 \end{cases} \tag{36}$$

for all $m_i \in \mathbb{N}$, where we have additionally used the identity (27).

From (34) and (36) we now obtain $\mathrm{E}\Big([\mathrm{i}(\xi;\eta) - I(\xi;\eta)]^m \Big) = 0$ for all $m = 2\tilde{m} - 1$ with $\tilde{m} \in \mathbb{N}$ because, if m is odd, then for all $(\ell_1, \ell_2, \ldots, \ell_r) \in \mathcal{K}_{m,r}$ at least one of the ℓ_i's has to be odd. If $m = 2\tilde{m}$ with $\tilde{m} \in \mathbb{N}$, we obtain from (34) and (36)

$$\begin{aligned} \mathrm{E}\Big([\mathrm{i}(\xi;\eta) - I(\xi;\eta)]^m \Big) &= \sum_{(\ell_1,\ell_2,\ldots,\ell_r) \in \mathcal{K}_{m,r}} m! \prod_{i=1}^{r} \frac{1}{\ell_i!} \frac{(\ell_i!)^2}{((\ell_i/2)!)^2 \, 4^{\frac{\ell_i}{2}}} \varrho_i^{\ell_i} \\ &= \sum_{(m_1,m_2,\ldots,m_r) \in \mathcal{K}_{m,r}^{[2]}} m! \prod_{i=1}^{r} \frac{(2m_i)!}{(m_i!)^2 \, 4^{m_i}} \varrho_i^{2m_i}. \qquad \square \end{aligned}$$

4.5. Proof of Part (iii) of Corollary 1

Using the random variable $\tilde{v}$ as in the proof of Theorem 3, we can write the m-th central moment of the information density $\mathrm{i}(\xi;\eta)$ as

$$\mathrm{E}\Big([\mathrm{i}(\xi;\eta) - I(\xi;\eta)]^m \Big) = \mathrm{E}(\tilde{v}^m) = (-\jmath)^m \left. \frac{\mathrm{d}^m}{\mathrm{d}t^m} \varphi_{\tilde{v}}(t) \right|_{t=0},$$

where the characteristic function $\varphi_{\tilde{v}}$ of $\tilde{v}$ is given by $\varphi_{\tilde{v}}(t) = \left(1 + \varrho_r^2 t^2\right)^{-r/2}, t \in \mathbb{R}$, due to Lemma 1 and the equality of all canonical correlations. Using the binomial series and the Taylor series expansion as in the proof of Theorem 3, we obtain

$$\left. \frac{\mathrm{d}^m}{\mathrm{d}t^m} \varphi_{\tilde{v}}(t) \right|_{t=0} = \begin{cases} m! \binom{-r/2}{m/2} \varrho_r^m & \text{if} \quad m = 2\tilde{m} \\ 0 & \text{if} \quad m = 2\tilde{m} - 1 \end{cases}$$

for all $\tilde{m} \in \mathbb{N}$. Collecting terms and additionally using the definition of the generalized binomial coefficient given in (25) in the proof of Theorem 1 yields (9). □

5. Recurrence Formulas and Finite Sum Approximations

If there are at least two distinct canonical correlations, then the PDF $f_{i(\xi;\eta)}$ and CDF $F_{i(\xi;\eta)}$ of the information density $i(\xi;\eta)$ are given by the infinite series in Theorems 1 and 2. If we consider only a finite number of summands in these representations, then we obtain approximations amenable in particular for numerical calculations. However, a direct finite sum approximation of the series in (3) and (4) is rather inefficient since modified Bessel and Struve L functions have to be evaluated for every summand. Therefore, we derive in this section recursive representations, which allow efficient numerical calculations. Furthermore, we derive uniform bounds of the approximation error. Based on the recurrence relations and the error bounds, an implementation in the programming language PYTHON has been developed, which provides an efficient tool to numerically calculate the PDF and CDF of the information density with a predefined accuracy as high as desired. The developed source code as well as illustrating examples are made publicly available in an open access repository on GITLAB [26].

Subsequently, we adopt all the previous notation and assume $r \geq 2$ and at least two distinct canonical correlations (since otherwise we have the case of Corollary 1, where the series reduce to a single summand).

5.1. Recurrence Formulas

The recursive approach developed below is based on the work of Moschopoulos [27], which extended the work of Mathai [10]. First, we rewrite the series representations of the PDF and CDF of the information density given in Theorem 1 and Theorem 2 in a form, which is suitable for recursive calculations. To begin with, we define two functions appearing in the series representations (3) and (4), which involve the modified Bessel function K_α of second kind and order α and the modified Struve L function L_α of order α. Let us define for all $k \in \mathbb{N}_0$ the functions U_k and D_k by

$$U_k(z) = \frac{K_{\frac{r-1}{2}+k}(z)}{\Gamma\left(\frac{r}{2}+k\right)}\left(\frac{z}{2}\right)^{\frac{r-1}{2}+k}, \quad z \geq 0 \tag{37}$$

and

$$D_k(z) = \frac{z}{2\varrho_r}\left[K_{\frac{r-1}{2}+k}\left(\frac{z}{\varrho_r}\right)L_{\frac{r-3}{2}+k}\left(\frac{z}{\varrho_r}\right) + K_{\frac{r-3}{2}+k}\left(\frac{z}{\varrho_r}\right)L_{\frac{r-1}{2}+k}\left(\frac{z}{\varrho_r}\right)\right], \quad z \geq 0. \tag{38}$$

Furthermore, we define for all $k \in \mathbb{N}_0$ the coefficient δ_k by

$$\delta_k = \sum_{(k_1, k_2, \ldots, k_{r-1}) \in \mathcal{K}_{k,r-1}} \left[\prod_{i=1}^{r-1} \frac{(2k_i)!}{(k_i!)^2 4^{k_i}}\left(1 - \frac{\varrho_r^2}{\varrho_i^2}\right)^{k_i}\right], \tag{39}$$

where $\mathcal{K}_{k,r-1} = \left\{(k_1, k_2, \ldots, k_{r-1}) \in \mathbb{N}_0^{r-1} : k_1 + k_2 + \cdots + k_{r-1} = k\right\}$. With these definitions, we obtain the following alternative series representations of (3) and (4) by observing that the multiple summations over the indices $k_1, k_2, \ldots, k_{r-1}$ can be shortened to one summation over the index $k = k_1 + k_2 + \ldots + k_{r-1}$.

Proposition 2 (Alternative representation of PDF and CDF of the information density). *The PDF $f_{i(\xi;\eta)}$ of the information density $i(\xi;\eta)$ given in Theorem 1 has the alternative series representation*

$$f_{i(\xi;\eta)}(x) = \frac{1}{\varrho_r\sqrt{\pi}}\left(\prod_{i=1}^{r-1}\frac{\varrho_r}{\varrho_i}\right)\sum_{k=0}^{\infty}\delta_k U_k\left(\frac{|x - I(\xi;\eta)|}{\varrho_r}\right), \quad x \in \mathbb{R}. \tag{40}$$

The function $V(\cdot)$ *specifying the CDF* $F_{i(\xi;\eta)}$ *of the information density* $i(\xi;\eta)$ *as given in Theorem* 2 *has the alternative series representation*

$$V(z) = \left(\prod_{i=1}^{r-1} \frac{\varrho_r}{\varrho_i}\right) \sum_{k=0}^{\infty} \delta_k D_k(z), \qquad z \geq 0. \tag{41}$$

Based on the representations in Proposition 2 and with recursive formulas for $U_k(\cdot)$, $D_k(\cdot)$ and δ_k, we are in the position to calculate the PDF and CDF of the information density by a single summation over completely recursively defined terms. In the following, we will derive recurrence relations for $U_k(\cdot)$, $D_k(\cdot)$ and δ_k, which allow the desired efficient calculations.

Lemma 2 (Recurrence formula of the function U_k)**.** *If for all* $k \in \mathbb{N}_0$ *the function* U_k *is defined by* (37), *then* $U_k(z)$ *satisfies for all* $k \geq 2$ *and* $z \geq 0$ *the recurrence formula*

$$U_k(z) = \frac{z^2}{(r+2k-2)(r+2k-4)} U_{k-2}(z) + \frac{r+2k-3}{r+2k-2} U_{k-1}(z). \tag{42}$$

Proof. First, assume $z = 0$. Based on Proposition 1, we obtain for all $k \in \mathbb{N}_0$

$$\lim_{z \to +0} U_k(z) = \frac{\Gamma\left(\frac{r-1}{2} + k\right)}{2\,\Gamma\left(\frac{r}{2} + k\right)}, \tag{43}$$

such that $U_k(0)$ is well defined and finite. Using the recurrence relation $\Gamma(y+1) = y\Gamma(y)$ for the Gamma function [24], Sec. 8.331.1 we have

$$\frac{\Gamma\left(\frac{r-1}{2} + k\right)}{2\,\Gamma\left(\frac{r}{2} + k\right)} = \frac{\left(\frac{r-1}{2} + k - 1\right)}{\left(\frac{r}{2} + k - 1\right)} \cdot \frac{\Gamma\left(\frac{r-1}{2} + k - 1\right)}{2\,\Gamma\left(\frac{r}{2} + k - 1\right)}.$$

This shows together with (43) that the recurrence formula (42) holds for $U_k(0)$ and $k \geq 2$.

Now, assume $z > 0$ and consider the recurrence formula

$$zK_\alpha(z) = zK_{\alpha-2}(z) + 2(\alpha-1)K_{\alpha-1}(z) \tag{44}$$

for the modified Bessel function of the second kind and order α [24], Sec. 8.486.10. Plugging (44) into (37) for $\alpha = \frac{r-1}{2} + k$ yields for $k \geq 2$

$$U_k(z) = \frac{K_{\frac{r-1}{2}+k-2}(z)}{\Gamma\left(\frac{r}{2}+k\right)} \left(\frac{z}{2}\right)^{\frac{r-1}{2}+k-2} \left(\frac{z}{2}\right)^2$$

$$+ \frac{\left(\frac{r-1}{2}+k-1\right) K_{\frac{r-1}{2}+k-1}(z)}{\Gamma\left(\frac{r}{2}+k\right)} \left(\frac{z}{2}\right)^{\frac{r-1}{2}+k-1}. \tag{45}$$

Using again the relation $\Gamma(y+1) = y\Gamma(y)$, we obtain

$$\Gamma\left(\tfrac{r}{2}+k\right) = \left(\tfrac{r}{2}+k-1\right)\Gamma\left(\tfrac{r}{2}+k-1\right) = \left(\tfrac{r}{2}+k-1\right)\left(\tfrac{r}{2}+k-2\right)\Gamma\left(\tfrac{r}{2}+k-2\right),$$

which yields together with (45) and (37) the recurrence formula (42) for $U_k(z)$ if $z > 0$ and $k \geq 2$. $\square$

Lemma 3 (Recurrence formula of the function D_k). *If, for all $k \in \mathbb{N}_0$, the function D_k is defined by (38), then $D_k(z)$ satisfies for all $k \geq 1$ and $z \geq 0$ the recurrence formula*

$$D_k(z) = D_{k-1}(z) - \frac{1}{2\sqrt{\pi}\left(\frac{r}{2} + k - 1\right)}\left(\frac{z}{\varrho_r}\right)U_{k-1}\left(\frac{z}{\varrho_r}\right), \tag{46}$$

with $U_k(\cdot)$ as defined in (37).

Proof. First, assume $z = 0$. We have $D_k(0) = 0$ for all $k \in \mathbb{N}_0$ and from the proof of Lemma 2 we have $U_k(0) = \Gamma\left(\frac{r-1}{2} + k\right)/2\,\Gamma\left(\frac{r}{2} + k\right)$ for all $k \in \mathbb{N}_0$. Thus, the left-hand side and the right-hand side of (46) are both zero, which shows that (46) holds for $z = 0$ and $k \geq 1$.

Now, assume $z > 0$ and consider the recurrence formula

$$z L_\alpha(z) = z L_{\alpha-2}(z) - 2(\alpha - 1)L_{\alpha-1}(z) - \frac{2^{1-\alpha}z^\alpha}{\sqrt{\pi}\Gamma\left(\alpha + \frac{1}{2}\right)}$$

for the modified Struve L function of order α [9], Sec. 11.4.25. Together with the recurrence formula (44) for the modified Bessel function of the second kind and order α, we obtain

$$z L_\alpha(z)K_{\alpha-1}(z) = z L_{\alpha-2}(z)K_{\alpha-1}(z) - 2(\alpha - 1)L_{\alpha-1}(z)K_{\alpha-1}(z)$$
$$- \frac{2^{1-\alpha}z^\alpha}{\sqrt{\pi}\Gamma\left(\alpha + \frac{1}{2}\right)}K_{\alpha-1}(z), \tag{47}$$

$$z K_\alpha(z)L_{\alpha-1}(z) = z K_{\alpha-2}(z)L_{\alpha-1}(z) + 2(\alpha - 1)K_{\alpha-1}(z)L_{\alpha-1}(z). \tag{48}$$

Plugging (47) and (48) into (38) for $\alpha = \frac{r-1}{2} + k$ yields for $k \geq 1$

$$D_k(z) = \frac{z}{2\varrho_r}\left[K_{\frac{r-1}{2}+k-1}\left(\frac{z}{\varrho_r}\right)L_{\frac{r-3}{2}+k-1}\left(\frac{z}{\varrho_r}\right) + K_{\frac{r-3}{2}+k-1}\left(\frac{z}{\varrho_r}\right)L_{\frac{r-1}{2}+k-1}\left(\frac{z}{\varrho_r}\right)\right]$$
$$- \frac{1}{\sqrt{\pi}\,\Gamma\left(\frac{r}{2} + k\right)}\left(\frac{z}{2\varrho_r}\right)^{\frac{r-1}{2}+k}K_{\frac{r-1}{2}+k-1}\left(\frac{z}{\varrho_r}\right).$$

Together with (38), the identity $\Gamma\left(\frac{r}{2} + k\right) = \left(\frac{r}{2} + k - 1\right)\Gamma\left(\frac{r}{2} + k - 1\right)$, and the definition of the function $U_k(\cdot)$ in (37), we obtain the recurrence formula (46) for $D_k(z)$ if $z > 0$ and $k \geq 1$. $\square$

Lemma 4 (Recursive formula of the coefficient δ_k). *The coefficient δ_k defined by (39) satisfies for all $k \in \mathbb{N}_0$ the recurrence formula*

$$\delta_{k+1} = \frac{1}{k+1}\sum_{j=1}^{k+1} j\,\gamma_j\,\delta_{k+1-j}, \tag{49}$$

where $\delta_0 = 1$ and

$$\gamma_j = \sum_{i=1}^{r-1}\frac{1}{2j}\left(1 - \frac{\varrho_r^2}{\varrho_i^2}\right)^j. \tag{50}$$

For the derivation of Lemma 4, we use an adapted version of the method of Moschopoulos [27] and the following auxiliary result.

Lemma 5. *For $k \in \mathbb{N}_0$, let g be a real univariate $(k+1)$-times differentiable function. Then, we have the following recurrence relation for the $(k+1)$-th derivative of the composite function $h = \exp(g)$*

$$h^{(k+1)} = \sum_{j=1}^{k+1} \binom{k}{j-1} g^{(j)} h^{(k-j+1)}, \tag{51}$$

where $f^{(i)}$ denotes the i-th derivative of the function f with $f^{(0)} = f$.

Proof. We prove the assertion of Lemma 5 by induction over k. First, consider the base case for $k = 0$. In this case, formula (51) gives

$$h^{(1)} = g^{(1)} h,$$

which is easily seen to be true.

Assuming formula (51) holds for $h^{(k)}$, we continue with the case $k+1$. Application of the product rule leads to

$$h^{(k+1)} = \left(h^{(k)} \right)^{(1)} = \left(\sum_{j=1}^{k} \binom{k-1}{j-1} g^{(j)} h^{(k-j)} \right)^{(1)}$$

$$= \sum_{j=1}^{k} \binom{k-1}{j-1} g^{(j+1)} h^{(k-j)} + \sum_{j=1}^{k} \binom{k-1}{j-1} g^{(j)} h^{(k-j+1)}.$$

Substitution of $j' = j+1$ in the first term gives

$$h^{(k+1)} = \sum_{j'=2}^{k+1} \binom{k-1}{j'-2} g^{(j')} h^{(k-j'+1)} + \sum_{j=1}^{k} \binom{k-1}{j-1} g^{(j)} h^{(k-j+1)}.$$

With this representation and the identity,

$$\binom{k-1}{j-2} + \binom{k-1}{j-1} = \binom{k}{j-1}$$

We finally have

$$h^{(k+1)} = g^{(1)} h^{(k)} + \sum_{j=2}^{k} \left[\binom{k-1}{j-1} + \binom{k-1}{j-2} \right] g^{(j)} h^{(k-j+1)} + g^{(k+1)} h$$

$$= \binom{k}{0} g^{(1)} h^{(k)} + \sum_{j=2}^{k} \binom{k}{j-1} g^{(j)} h^{(k-j+1)} + \binom{k}{k} g^{(k+1)} h$$

$$= \sum_{j=1}^{k+1} \binom{k}{j-1} g^{(j)} h^{(k-j+1)}.$$

This completes the proof of Lemma 5. □

Proof of Lemma 4. To prove the recurrence formula (49), we consider the characteristic function

$$\varphi_{\bar{v}}(t) = \prod_{i=1}^{r} \left(1 + \varrho_i^2 t^2 \right)^{-\frac{1}{2}}, \qquad t \in \mathbb{R} \tag{52}$$

of the random variable $\tilde{\nu}$ introduced in Lemma 1. On the one hand, the series representation of $\varphi_{\tilde{\nu}}$ given in (28) in the proof of Theorem 1 can be rewritten as follows using the coefficient δ_k defined in (39):

$$\varphi_{\tilde{\nu}}(t) = \left(1 + \varrho_r^2 t^2\right)^{-\frac{r}{2}} \left(\prod_{i=1}^{r-1} \frac{\varrho_r}{\varrho_i}\right) \sum_{\ell=0}^{\infty} \delta_\ell \left(1 + \varrho_r^2 t^2\right)^{-\ell}, \quad t \in \mathbb{R}. \tag{53}$$

On the other hand, recall the expansion of $\left(1 + \varrho_i^2 t^2\right)^{-\frac{1}{2}}$ given in (22), which yields together with (52) and the application of the natural logarithm the identity

$$\log(\varphi_{\tilde{\nu}}(t)) = \log\left(\left(1 + \varrho_r^2 t^2\right)^{-\frac{r}{2}} \left(\prod_{i=1}^{r-1} \frac{\varrho_r}{\varrho_i}\right)\right)$$
$$+ \sum_{i=1}^{r-1} \log\left(\left(1 + \left(\frac{\varrho_r^2}{\varrho_i^2} - 1\right)\left(1 + \varrho_r^2 t^2\right)^{-1}\right)^{-\frac{1}{2}}\right). \tag{54}$$

Now consider the power series

$$\log(1 + y) = \sum_{\ell=1}^{\infty} \frac{(-1)^{\ell+1}}{\ell} y^\ell, \tag{55}$$

which is absolutely convergent for $|y| < 1$. With the same arguments as in the proof of Theorem 1, in particular due to (26), we can apply the series expansion (55) to the second term on the right-hand side of (54) to obtain the absolutely convergent series representation

$$\log(\varphi_{\tilde{\nu}}(t)) = \log\left(\left(1 + \varrho_r^2 t^2\right)^{-\frac{r}{2}} \left(\prod_{i=1}^{r-1} \frac{\varrho_r}{\varrho_i}\right)\right) + \sum_{\ell=1}^{\infty} \gamma_\ell \left(1 + \varrho_r^2 t^2\right)^{-\ell}, \tag{56}$$

where we have further used the definition of γ_ℓ given in (50). Applying the exponential function to both sides of (56) then yields the following expression for the characteristic function $\varphi_{\tilde{\nu}}$.

$$\varphi_{\tilde{\nu}}(t) = \left(1 + \varrho_r^2 t^2\right)^{-\frac{r}{2}} \left(\prod_{i=1}^{r-1} \frac{\varrho_r}{\varrho_i}\right) \exp\left(\sum_{\ell=1}^{\infty} \gamma_\ell \left(1 + \varrho_r^2 t^2\right)^{-\ell}\right) \tag{57}$$

Comparing (53) and (57) yields the identity

$$\sum_{\ell=0}^{\infty} \delta_\ell \left(1 + \varrho_r^2 t^2\right)^{-\ell} = \exp\left(\sum_{\ell=1}^{\infty} \gamma_\ell \left(1 + \varrho_r^2 t^2\right)^{-\ell}\right). \tag{58}$$

We now define $x = \left(1 + \varrho_r^2 t^2\right)^{-1}$ and take the $(k+1)$-th derivative w. r. t. x on both sides of (58) using the identity

$$\frac{\mathrm{d}^m}{\mathrm{d}x^m}\left(\sum_{\ell=0}^{\infty} a_\ell x^\ell\right) = \frac{\mathrm{d}^m}{\mathrm{d}x^m}\left(\sum_{\ell=1}^{\infty} a_\ell x^\ell\right) = \sum_{\ell=m}^{\infty} \frac{\ell!}{(\ell-m)!} a_\ell x^{\ell-m} \tag{59}$$

for the m-th derivative of a power series $\sum_{\ell=0}^{\infty} a_\ell x^\ell$. For the left-hand side of (58), we obtain

$$\frac{\mathrm{d}^{k+1}}{\mathrm{d}x^{k+1}}\left(\sum_{\ell=0}^{\infty} \delta_\ell x^\ell\right) = \sum_{\ell=k+1}^{\infty} \frac{\ell!}{(\ell-k-1)!} \delta_\ell x^{\ell-k-1}. \tag{60}$$

For the right-hand side of (58), we obtain

$$
\frac{d^{k+1}}{dx^{k+1}}\left(\exp\left(\sum_{\ell=1}^{\infty}\gamma_\ell x^\ell\right)\right) = \sum_{j=1}^{k+1}\binom{k}{j-1}\frac{d^j}{dx^j}\left(\sum_{\ell=1}^{\infty}\gamma_\ell x^\ell\right)\frac{d^{k-j+1}}{dx^{k-j+1}}\left(\exp\left(\sum_{\ell=1}^{\infty}\gamma_\ell x^\ell\right)\right)
$$

$$
= \sum_{j=1}^{k+1}\binom{k}{j-1}\frac{d^j}{dx^j}\left(\sum_{\ell=1}^{\infty}\gamma_\ell x^\ell\right)\frac{d^{k-j+1}}{dx^{k-j+1}}\left(\sum_{\ell=0}^{\infty}\delta_\ell x^\ell\right)
$$

$$
= \sum_{j=1}^{k+1}\binom{k}{j-1}\left(\sum_{\ell=j}^{\infty}\frac{\ell!\gamma_\ell}{(\ell-j)!}x^{\ell-j}\right)\times
$$

$$
\left(\sum_{\ell=k+1-j}^{\infty}\frac{\ell!\delta_\ell}{(\ell-k+j-1)!}x^{\ell-k+j-1}\right), \tag{61}
$$

where we used Lemma 5 and the identities (58) and (59). From the equality

$$
\frac{d^{k+1}}{dx^{k+1}}\left(\sum_{\ell=0}^{\infty}\delta_\ell x^\ell\right) = \frac{d^{k+1}}{dx^{k+1}}\left(\exp\left(\sum_{\ell=1}^{\infty}\gamma_\ell x^\ell\right)\right)
$$

and the evaluation of the right-hand side of (60) and (61), we obtain

$$
(k+1)!\delta_{k+1}\,x^0 + (\dots)\,x^1 + (\dots)\,x^2 \dots
$$

$$
= \left(\sum_{j=1}^{k+1}\binom{k}{j-1}j!\gamma_j(k+1-j)!\delta_{k+1-j}\right)x^0 + (\dots)\,x^1 + (\dots)\,x^2 \dots
$$

Comparing the coefficients for x^0 finally yields

$$
\delta_{k+1} = \frac{1}{(k+1)!}\sum_{j=1}^{k+1}\binom{k}{j-1}j!\,\gamma_j(k+1-j)!\,\delta_{k+1-j}
$$

$$
= \frac{1}{(k+1)!}\sum_{j=1}^{k+1}\frac{k!}{(j-1)!(k+1-j)!}j!\,\gamma_j(k+1-j)!\,\delta_{k+1-j}
$$

$$
= \frac{1}{(k+1)}\sum_{j=1}^{k+1}j\,\gamma_j\,\delta_{k+1-j}.
$$

This completes the proof of Lemma 4. □

5.2. Finite Sum Approximations

The results in the previous Section 5.1 can be used in the following way for efficient numerical calculations. Consider

$$
\hat{f}_{i(\xi;\eta)}(x,n) = \frac{1}{\varrho_r\sqrt{\pi}}\left(\prod_{i=1}^{r-1}\frac{\varrho_r}{\varrho_i}\right)\sum_{k=0}^{n}\delta_k U_k\left(\frac{|x-I(\xi;\eta)|}{\varrho_r}\right), \quad x \in \mathbb{R} \tag{62}
$$

for $n \in \mathbb{N}_0$, i.e., the finite sum approximation of the PDF given in (40). To calculate $\hat{f}_{i(\xi;\eta)}(x,n)$, first calculate $U_0\big(|x-I(\xi;\eta)|/\varrho_r\big)$ and $U_1\big(|x-I(\xi;\eta)|/\varrho_r\big)$ using (37). Then, use the recurrence formulas (42) and (49) to calculate the remaining summands in (62). The great advantage of this approach is that only two evaluations of the modified Bessel function are required and for the rest of the calculations efficient recursive formulas are employed making the numerical computations efficient.

Similarly, consider

$$\hat{F}_{\mathrm{i}(\xi;\eta)}(x,n) = \begin{cases} \dfrac{1}{2} - \hat{\mathrm{V}}(I(\xi;\eta) - x, n) & \text{if} \quad x \le I(\xi;\eta) \\[2mm] \dfrac{1}{2} + \hat{\mathrm{V}}(x - I(\xi;\eta), n) & \text{if} \quad x > I(\xi;\eta) \end{cases}, \tag{63}$$

with
$$\hat{\mathrm{V}}(z,n) = \left(\prod_{i=1}^{r-1} \frac{\varrho_r}{\varrho_i}\right) \sum_{k=0}^{n} \delta_k \mathrm{D}_k(z), \quad z \ge 0, \tag{64}$$

for $n \in \mathbb{N}_0$, i.e., the finite sum approximation of the alternative representation of the CDF of the information density, where $\hat{\mathrm{V}}(z,n)$ is the finite sum approximation of the function $\mathrm{V}(\cdot)$ given in (41). To calculate $\hat{F}_{\mathrm{i}(\xi;\eta)}(x,n)$, first calculate $\mathrm{D}_0(z)$, $\mathrm{U}_0(z/\varrho_r)$, and $\mathrm{U}_1(z/\varrho_r)$ for $z = I(\xi;\eta) - x$ or $z = x - I(\xi;\eta)$ using (37) and (38). Then, use the recurrence formulas (42), (46), and (49) to calculate the remaining summands in (64). This approach requires only three evaluations of the modified Bessel and Struve L function resulting in efficient numerical calculations also for the CDF of the information density.

The following theorem provides suitable bounds to evaluate and control the error related to the introduced finite sum approximations.

Theorem 4 (Bounds of the approximation error for the alternative representation of PDF and CDF). *For the finite sum approximations in* (62)–(64) *of the alternative representation of the PDF and CDF of the information density as given in Proposition 2, we have for $n \in \mathbb{N}$ summands the error bounds*

$$\left|f_{\mathrm{i}(\xi;\eta)}(x) - \hat{f}_{\mathrm{i}(\xi;\eta)}(x,n)\right| \le \frac{\Gamma\left(\frac{r-1}{2} + n\right)}{2\varrho_r\sqrt{\pi}\,\Gamma\left(\frac{r}{2} + n\right)} \left(1 - \left(\prod_{i=1}^{r-1} \frac{\varrho_r}{\varrho_i}\right) \sum_{k=0}^{n} \delta_k\right), \quad x \in \mathbb{R} \tag{65}$$

and

$$\left|\mathrm{V}(z) - \hat{\mathrm{V}}(z,n)\right| \le \frac{1}{2}\left(1 - \left(\prod_{i=1}^{r-1} \frac{\varrho_r}{\varrho_i}\right) \sum_{k=0}^{n} \delta_k\right), \quad z \ge 0. \tag{66}$$

Proof. From the special case where all canonical correlations are equal, we can conclude from the CDF given in Corollary 1 that the function

$$z \mapsto z\left[\mathrm{K}_\alpha(z)\mathrm{L}_{\alpha-1}(z) + \mathrm{K}_{\alpha-1}(z)\mathrm{L}_\alpha(z)\right], \quad z \ge 0, \tag{67}$$

is monotonically increasing for all $\alpha = (j-1)/2$, $j \in \mathbb{N}$ and that further

$$\lim_{z\to\infty} z\left[\mathrm{K}_\alpha(z)\mathrm{L}_{\alpha-1}(z) + \mathrm{K}_{\alpha-1}(z)\mathrm{L}_\alpha(z)\right] = 1 \tag{68}$$

holds. Using (68), we obtain from (4)

$$\lim_{z\to\infty} 2\mathrm{V}(z) = \sum_{k_1=0}^{\infty} \sum_{k_2=0}^{\infty} \cdots \sum_{k_{r-1}=0}^{\infty} \left[\prod_{i=1}^{r-1} \frac{\varrho_r}{\varrho_i} \frac{(2k_i)!}{(k_i!)^2 4^{k_i}}\left(1 - \frac{\varrho_r^2}{\varrho_i^2}\right)^{k_i}\right]$$

by exchanging the limit and the summation, which is justified by the monotone convergence theorem. Due to the properties of the CDF, we have $\lim_{z\to\infty} 2\mathrm{V}(z) = 1$, which implies

$$\left(\prod_{i=1}^{r-1} \frac{\varrho_r}{\varrho_i}\right) \sum_{k=0}^{\infty} \delta_k = \sum_{k_2=0}^{\infty} \cdots \sum_{k_{r-1}=0}^{\infty} \left[\prod_{i=1}^{r-1} \frac{\varrho_r}{\varrho_i} \frac{(2k_i)!}{(k_i!)^2 4^{k_i}}\left(1 - \frac{\varrho_r^2}{\varrho_i^2}\right)^{k_i}\right] = 1, \tag{69}$$

where the first equality follows from the definition of the coefficient δ_k in (39).

We now obtain with (41) and (64)

$$\left|V(z) - \hat{V}(z,n)\right| = \left(\prod_{i=1}^{r-1} \frac{\varrho_r}{\varrho_i}\right) \sum_{k=n+1}^{\infty} \delta_k D_k(z) \leq \left(\prod_{i=1}^{r-1} \frac{\varrho_r}{\varrho_i}\right) \sum_{k=n+1}^{\infty} \frac{1}{2}\delta_k.$$

The inequality follows from the definition of the function $D_k(\cdot)$ in (38), the monotonicity of the function in (67), and from (68). Then, (66) follows from (69).

Similarly, we obtain with (40) and (62)

$$\left|f_{i(\xi;\eta)}(x) - \hat{f}_{i(\xi;\eta)}(x,n)\right| = \frac{1}{\varrho_r\sqrt{\pi}}\left(\prod_{i=1}^{r-1}\frac{\varrho_r}{\varrho_i}\right)\sum_{k=n+1}^{\infty}\delta_k U_k\left(\frac{|x - I(\xi;\eta)|}{\varrho_r}\right)$$

$$\leq \frac{1}{\varrho_r\sqrt{\pi}}\left(\prod_{i=1}^{r-1}\frac{\varrho_r}{\varrho_i}\right)\sum_{k=n+1}^{\infty}\delta_k\frac{\Gamma\left(\frac{r-1}{2}+n\right)}{2\,\Gamma\left(\frac{r}{2}+n\right)}.$$

The inequality follows from the definition of the function $U_k(\cdot)$, Proposition 1, and the decreasing monotonicity of $\Gamma(\frac{r-1}{2}+k)/\Gamma(\frac{r}{2}+k)$ w.r.t. $k \in \mathbb{N}_0$. Then, (65) follows from (69). □

Remark 1. *Note that the bound in* (65) *can be further simplified using the inequality* $\Gamma(\alpha)/\Gamma(\alpha + 1/2) \leq \sqrt{\pi}$. *Further note that the derived error bounds are uniform in the sense that they only depend on the parameters of the given Gaussian distribution and the number of summands considered. As can be seen from* (69), *the bounds converge to zero as the number of summands jointly increase.*

Remark 2 (Relation to Bell polynomials). *Interestingly, the coefficient δ_k can be expressed for all $k \in \mathbb{N}$ in the following form*

$$\delta_k = \frac{B_k(\gamma_1, 2\gamma_2, 6\gamma_3, \ldots, k!\gamma_k)}{k!},$$

where γ_j is defined in (50), *and B_k denotes the complete Bell polynomial of order k* [28], *Sec. 3.3. Even though this is an interesting connection to the Bell polynomials, which provides an explicit formula of δ_k, the recursive formula given in Lemma 4 is more efficient for numerical calculations.*

6. Numerical Examples and Illustrations

We illustrate the results of this paper with some examples, which all can be verified with the Python implementation publicly available on GITLAB [26].

Equal canonical correlations. First, we consider the special case of Corollary 1 when all canonical correlations are equal. The PDF and CDF given by (6) and (7) are illustrated in Figures 1 and 2 in centered form, i.e., shifted by $I(\xi;\eta)$, for $r \in \{1,2,3,4,5\}$ and equal canonical correlations $\varrho_i = 0.9, i = 1, \ldots, r$. In Figures 3 and 4, a fixed number of $r = 5$ equal canonical correlations $\varrho_i \in \{0.1, 0.2, 0.5, 0.7, 0.9\}, i = 1, \ldots, r$ is considered. When all canonical correlations are equal, then, due to the central limit theorem, the distribution of the information density $i(\xi;\eta)$ converges to a Gaussian distribution as $r \to \infty$. Figures 5 and 6 show for $r \in \{5, 10, 20, 40\}$ and equal canonical correlations $\varrho_i = 0.2, i = 1, 2, \ldots, r$ the PDF and CDF of the information density together with corresponding Gaussian approximations. The approximations are obtained by considering Gaussian distributions, which have the same variance as the information density $i(\xi;\eta)$. Recall that the variance of the information density is given by (11), i.e., by the sum of the squared canonical correlations. The illustrations show that only for a high number of equal canonical correlations the distribution of the information density becomes approximately Gaussian.
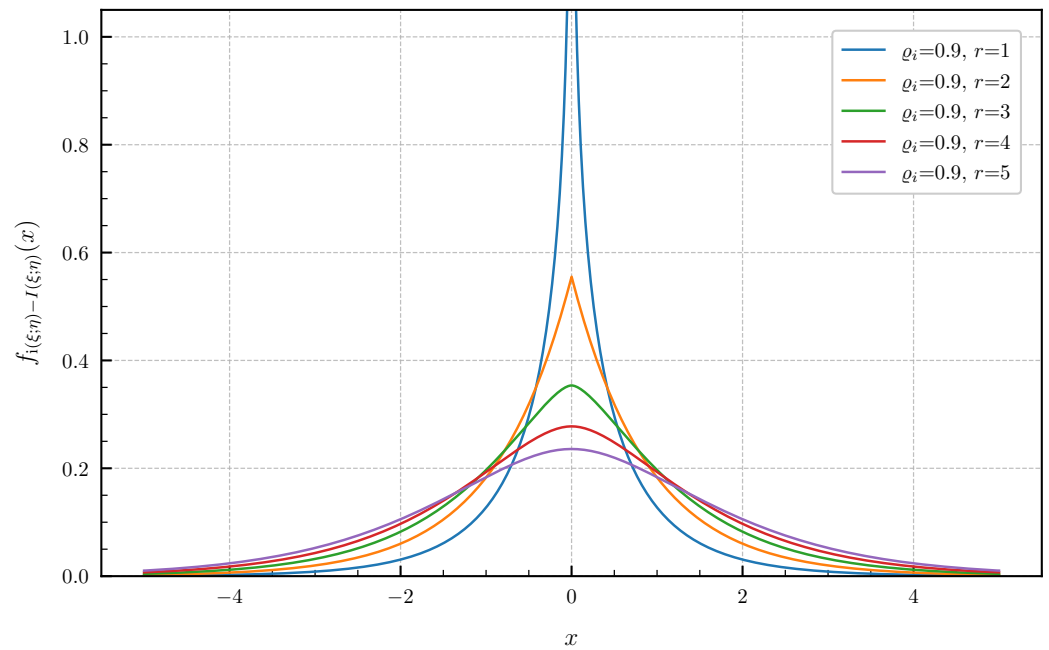
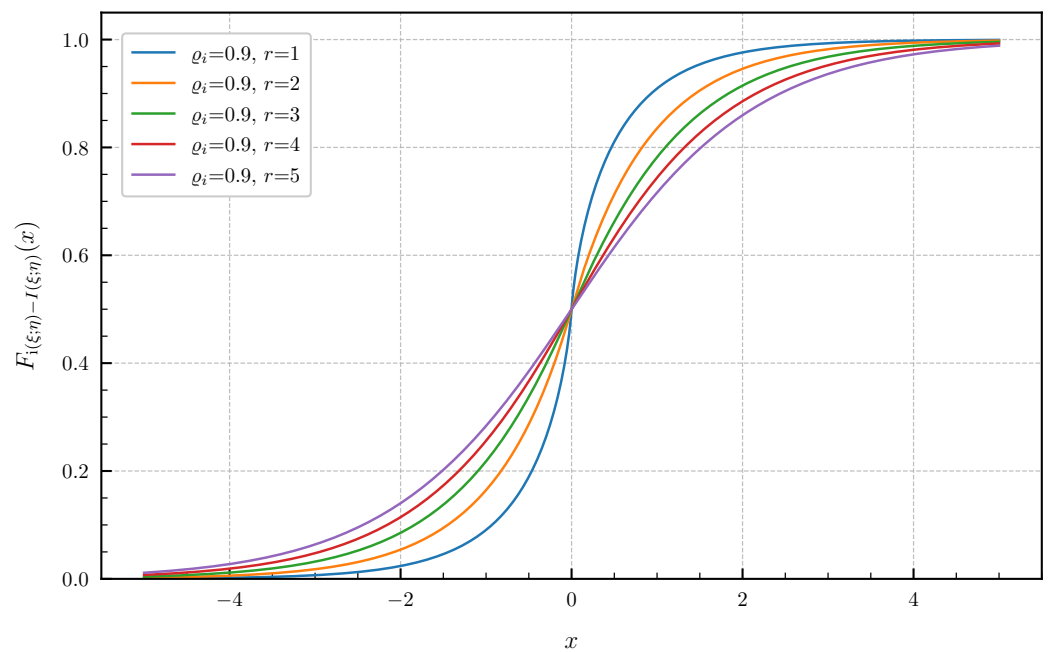**Figure 1.** PDF $f_{\mathrm{i}(\xi;\eta)-I(\xi;\eta)}$ for $r \in \{1,2,3,4,5\}$ equal canonical correlations $\varrho_i = 0.9$.

Figure 2. CDF $F_{\mathrm{i}(\xi;\eta)-I(\xi;\eta)}$ for $r \in \{1,2,3,4,5\}$ equal canonical correlations $\varrho_i = 0.9$.
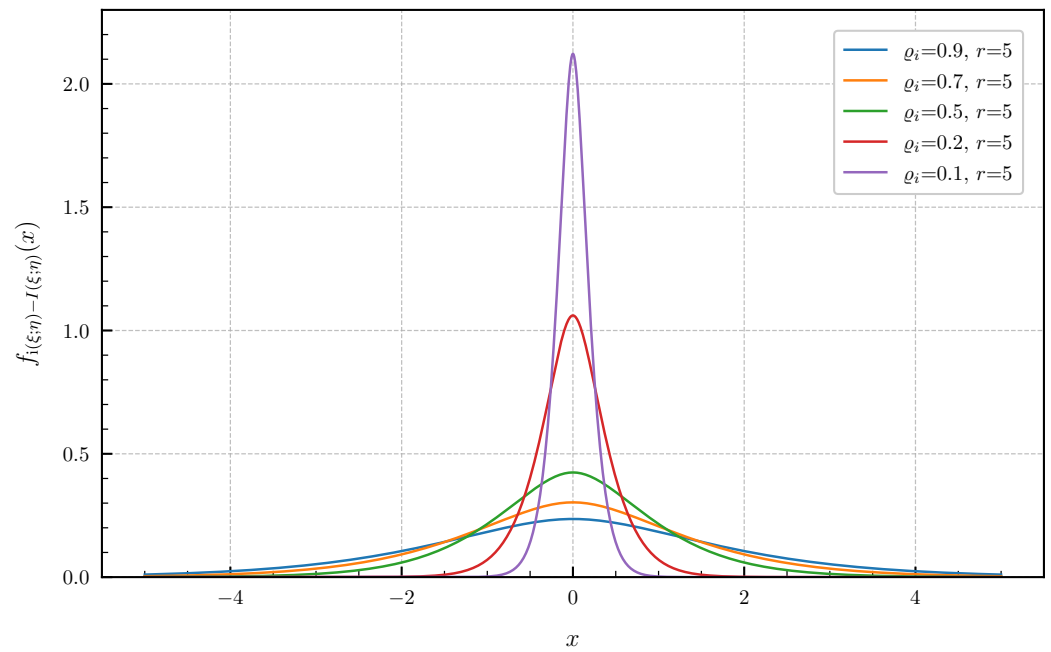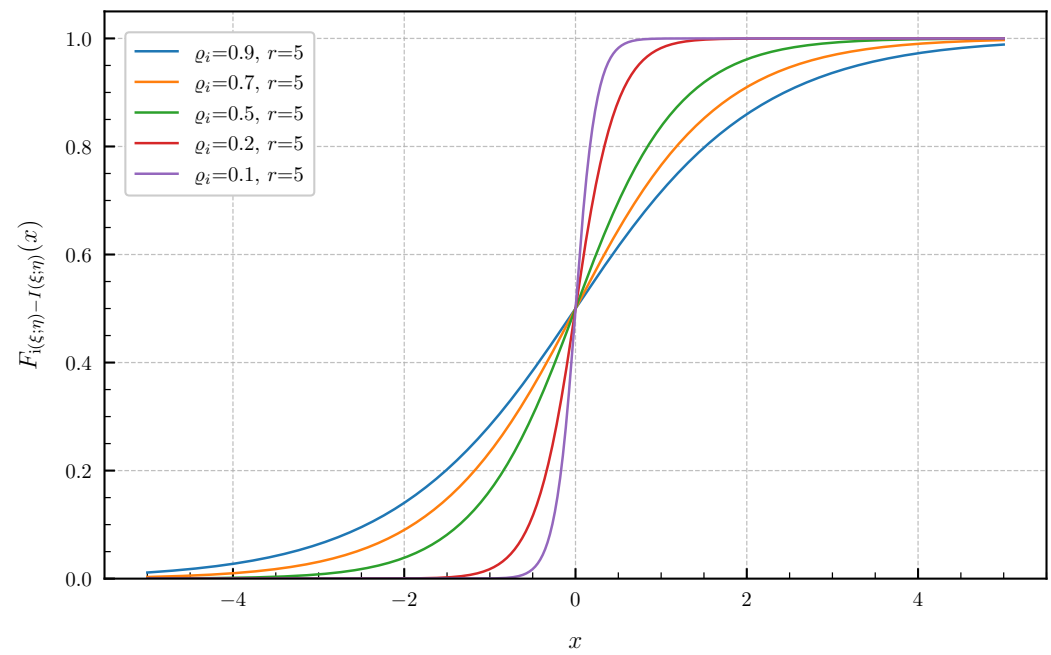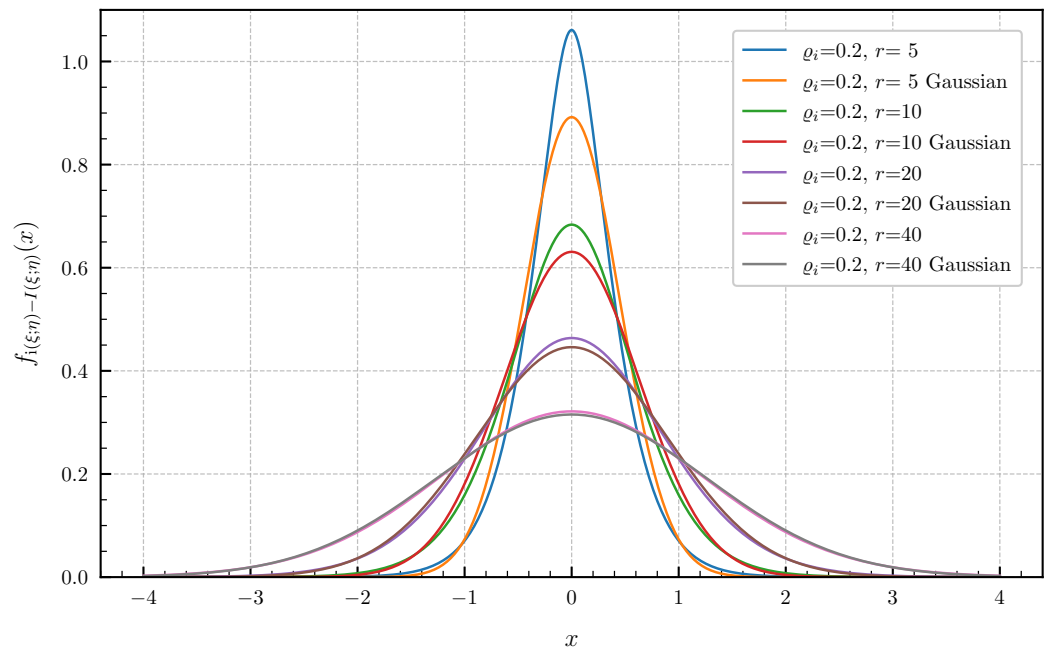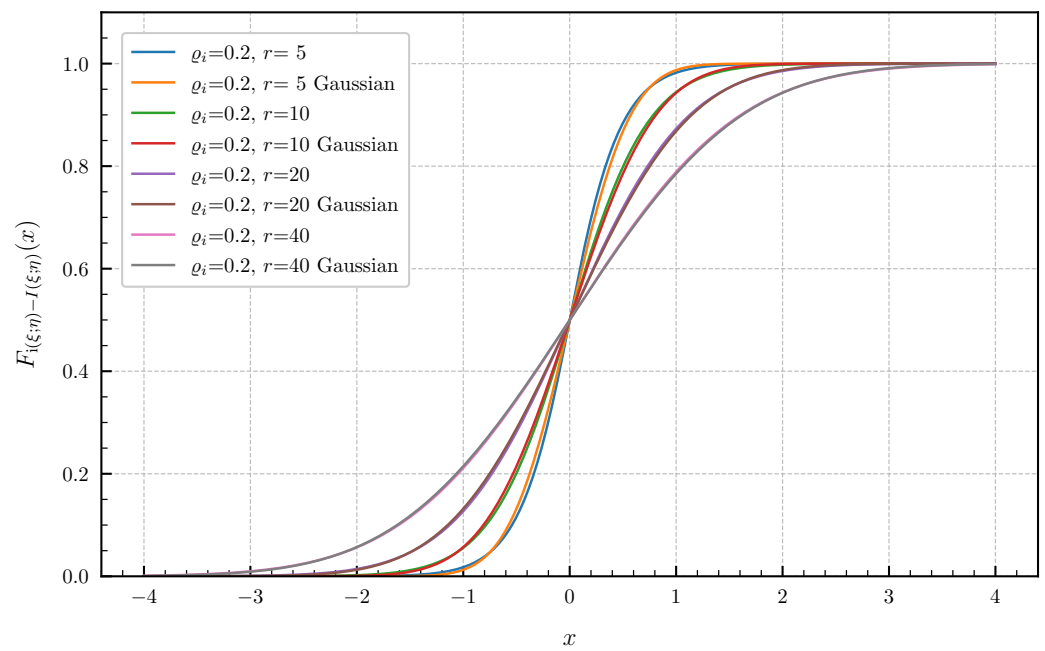
Figure 3. PDF $f_{\mathrm{i}(\xi;\eta)-I(\xi;\eta)}$ for $r = 5$ equal canonical correlations $\varrho_i \in \{0.1, 0.2, 0.5, 0.7, 0.9\}$.

Figure 4. CDF $F_{\mathrm{i}(\xi;\eta)-I(\xi;\eta)}$ for $r = 5$ equal canonical correlations $\varrho_i \in \{0.1, 0.2, 0.5, 0.7, 0.9\}$.

Figure 5. PDF $f_{i(\xi;\eta)-I(\xi;\eta)}$ for $r \in \{5, 10, 20, 40\}$ equal canonical correlations $\varrho_i = 0.2$ vs. Gaussian approximation.

Figure 6. CDF $F_{i(\xi;\eta)-I(\xi;\eta)}$ for $r \in \{5, 10, 20, 40\}$ equal canonical correlations $\varrho_i = 0.2$ vs. Gaussian approximation.

Different canonical correlations. To illustrate the case with different canonical correlations, let us consider two more examples.

(i) First, assume that the random vectors $\xi = (\xi_1, \xi_2, \ldots, \xi_p)$ and $\eta = (\eta_1, \eta_2, \ldots, \eta_q)$ have equal dimensions, i.e., $p = q$, and are related by

$$(\eta_1, \eta_2, \ldots, \eta_p) = (\xi_1 + \zeta_1, \xi_2 + \zeta_2, \ldots, \xi_p + \zeta_p),$$

where $\xi = (\xi_1, \xi_2, \ldots, \xi_p)$ and $\zeta = (\zeta_1, \zeta_2, \ldots, \zeta_p)$ are zero mean Gaussian random vectors, independent of each other and with covariance matrices

$$R_\xi = \left(\rho^{|i-j|} \right)_{i,j=1}^p \qquad \text{and} \qquad R_\zeta = \sigma_z^2 \mathbb{I}_p,$$

for parameters $0 < |\rho| < 1$ and $\sigma_z^2 > 0$, where $\mathbb{I}_p$ denotes the identity matrix of dimension $p \times p$. The covariance matrix of the Gaussian random vector $(\xi_1, \xi_2, \ldots, \xi_p, \eta_1, \eta_2, \ldots, \eta_p)$ is the basis of the canonical correlation analysis and is given by

$$\begin{pmatrix} R_\xi & R_{\xi\eta} \\ R_{\xi\eta} & R_\eta \end{pmatrix} = \begin{pmatrix} R_\xi & R_\xi \\ R_\xi & R_\xi + R_\zeta \end{pmatrix}.$$

The specified situation corresponds to a discrete-time additive noise channel, where a stationary first-order Markov-Gaussian input process is corrupted by a stationary additive white Gaussian noise process. In this setting, a block of p consecutive input and output symbols is considered.

For given parameter values ρ and σ_z^2, the canonical correlations can be calculated numerically with the method described in Section 3. However, the example at hand even allows the derivation of explicit formulas for the canonical correlations. Evaluating the approach in Section 3 analytically yields

$$\varrho_i(\rho, \sigma_z^2) = \sqrt{\frac{\lambda_i}{\lambda_i + \sigma_z^2}} \quad \text{with} \quad \lambda_i = \frac{1 - \rho^2}{1 - 2\rho\cos(\theta_i) + \rho^2}, \quad i = 1, 2, \ldots, r = p, \qquad (70)$$

where $\theta_1, \theta_2, \ldots, \theta_r$ are the zeros of the function

$$g(\theta) = \sin\left((r+1)\theta\right) - 2\rho\sin\left(r\theta\right) + \rho^2 \sin\left((r-1)\theta\right), \qquad \theta \in (0, \pi).$$

In this representation, $\lambda_1, \lambda_2, \ldots, \lambda_r$ denote the eigenvalues of the covariance matrix $R_\xi = (\rho^{|i-j|})_{i,j=1}^p$ derived in [29], Sec. 5.3.

As numerical examples Figure 7 and 8 show, the approximated PDF $\hat{f}_{i(\xi;\eta)-I(\xi;\eta)}(\cdot, n)$ and CDF $\hat{F}_{i(\xi;\eta)-I(\xi;\eta)}(\cdot, n)$ for $p = r \in \{5, 10, 20, 40\}$ and the parameter values $\rho = 0.9$ and $\sigma_z^2 = 10$ using the finite sums (62) and (64). The bounds of the approximation error given in Theorem 4 are chosen $< 1 \times 10^{-3}$ to obtain a high precision of the plotted curves. The number n of summands required in (62) and (64) to achieve these error bounds for $r \in \{5, 10, 20, 40\}$ is equal to $n \in \{217, 333, 462, 649\}$ for the PDF and $n \in \{282, 444, 618, 847\}$ for the CDF. For this example, the distribution of the information density $i(\xi; \eta)$ converges to a Gaussian distribution as $r \to \infty$. However, Figures 7 and 8 show that, even for $r = 40$, there is still a significant gap between the exact distribution and the corresponding Gaussian approximation.
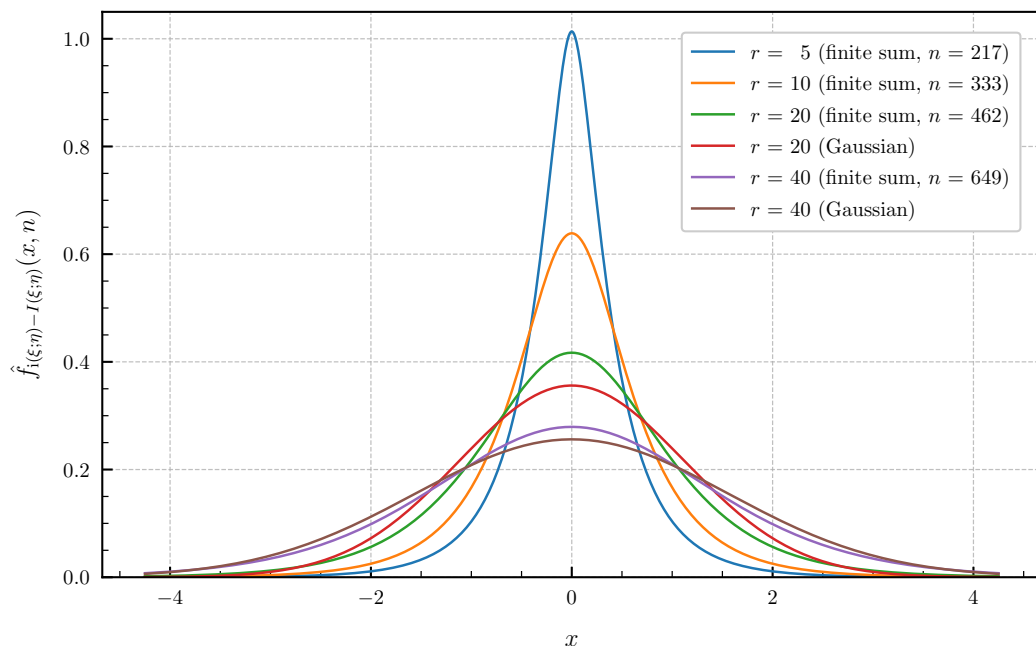
Figure 7. Approximated PDF $\hat{f}_{\mathrm{i}(\xi;\eta)-I(\xi;\eta)}(\cdot,n)$ for $r \in \{5, 10, 20, 40\}$ canonical correlations $\varrho_i(\rho, \sigma_z^2)$ given in (70) for $\rho = 0.9$ and $\sigma_z^2 = 10$ (approximation error $< 1 \times 10^{-3}$) vs. Gaussian approximation ($r \in \{20, 40\}$).
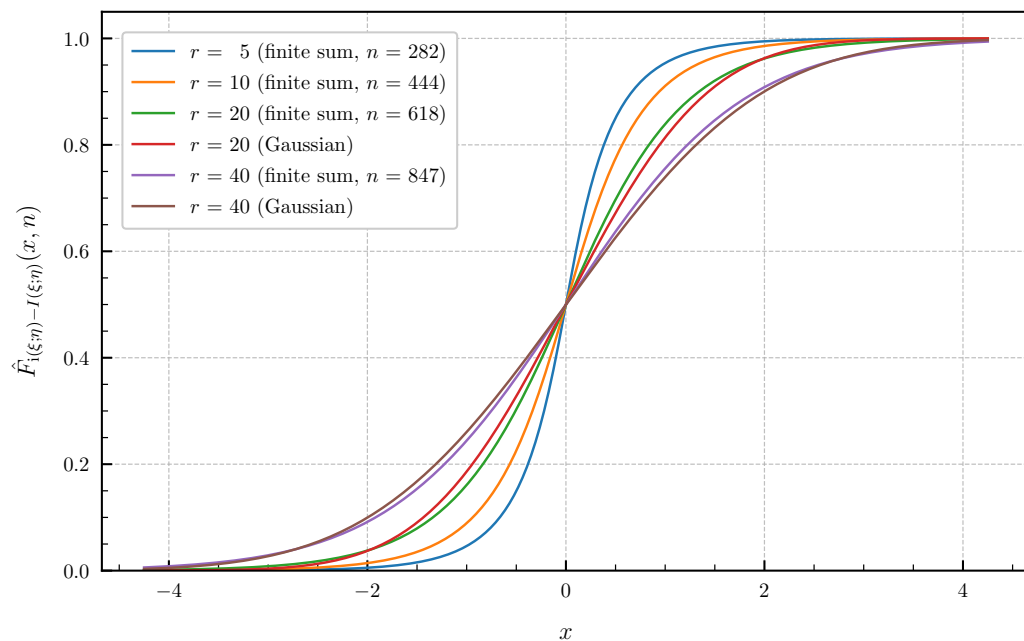
Figure 8. Approximated CDF $\hat{F}_{\mathrm{i}(\xi;\eta)-I(\xi;\eta)}(\cdot,n)$ for $r \in \{5, 10, 20, 40\}$ canonical correlations $\varrho_i(\rho, \sigma_z^2)$ given in (70) for $\rho = 0.9$ and $\sigma_z^2 = 10$ (approximation error $< 1 \times 10^{-3}$) vs. Gaussian approximation ($r \in \{20, 40\}$).

(ii) As a second example with different canonical correlations, let us consider the sequence $\{\varrho_1(T), \varrho_2(T), \ldots, \varrho_r(T)\}$ with

$$\varrho_i(T) = \sqrt{\frac{T^2}{T^2 + \pi\left(i - \frac{1}{2}\right)^2}}, \qquad i = 1, 2, \ldots, r. \tag{71}$$

These canonical correlations are related to the information density of a continuous-time additive white Gaussian noise channel confined to a finite time interval $[0, T]$ with a Brownian motion as input signal (see, e.g., Huffmann [30], Sec. 8.1 for more details). Figures 9 and 10 show the approximated PDF $\hat{f}_{\mathrm{i}(\xi;\eta) - I(\xi;\eta)}(\cdot, n)$ and CDF $\hat{F}_{\mathrm{i}(\xi;\eta) - I(\xi;\eta)}(\cdot, n)$ for $r \in \{2, 5, 10, 15\}$ and $T = 1$ using the finite sums (62) and (64). The bounds of the approximation error given in Theorem 4 are chosen $< 1 \times 10^{-2}$ such that there are no differences visible in the plotted curves by further lowering the approximation error. The number n of summands required in (62) and (64) to achieve these error bounds for $r \in \{2, 5, 10, 15\}$ is equal to $n \in \{15, 141, 638, 1688\}$ for the PDF and $n \in \{20, 196, 886, 2071\}$ for the CDF. Choosing r larger than 15 for the canonical correlations (71) with $T = 1$ does not result in visible changes of the PDF and CDF compared to $r = 15$. This demonstrates, together with Figures 9 and 10, that a Gaussian approximation is not valid for this example, even if $r \to \infty$.
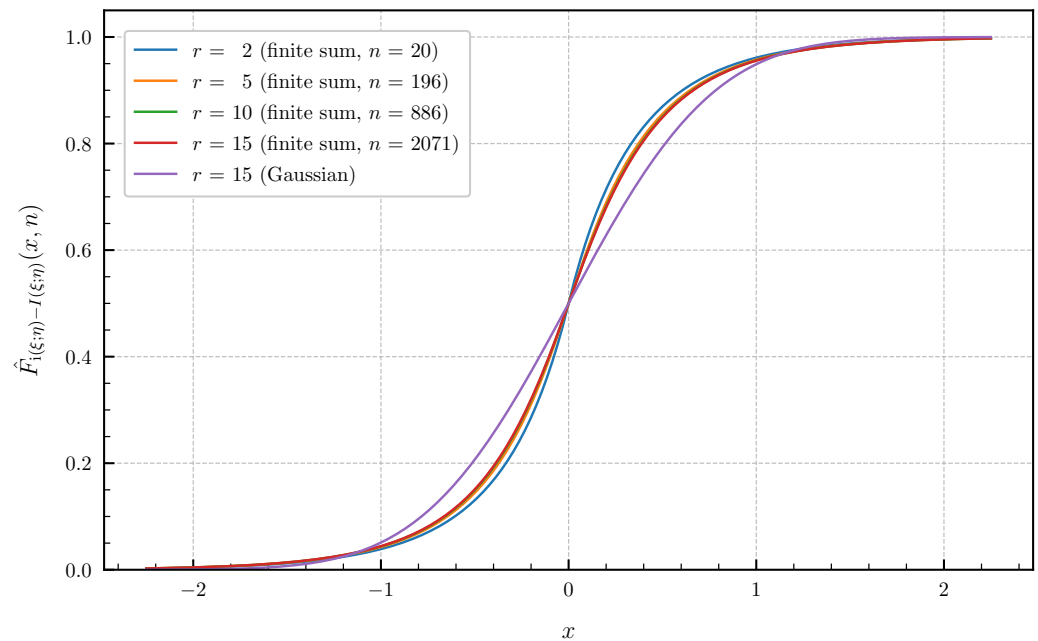
Indeed, from [8], Th. 9.6.1 and the comment above Eq. (9.6.45) in [8], one can conclude that, whenever the canonical correlations satisfy

$$\lim_{r \to \infty} \sum_{i=1}^{r} \varrho_i^2 < \infty,$$

then the distribution of the information density is *not* Gaussian.
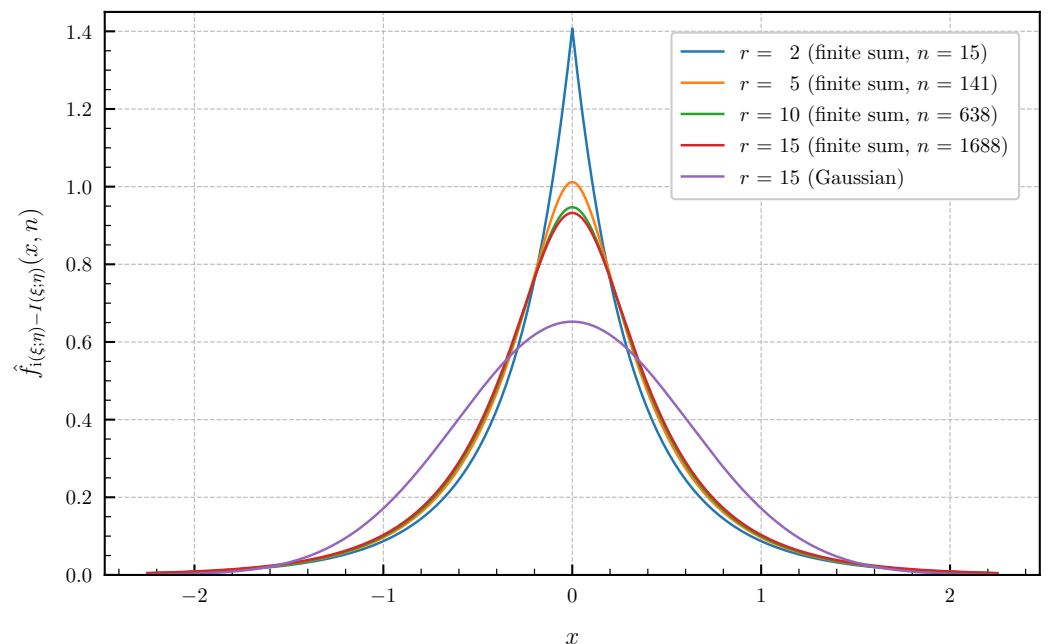
Figure 9. Approximated PDF $\hat{f}_{\mathrm{i}(\xi;\eta) - I(\xi;\eta)}(\cdot, n)$ for $r \in \{2, 5, 10, 15\}$ canonical correlations $\varrho_i(T)$ given in (71) for $T = 1$ (approximation error $< 1 \times 10^{-2}$) vs. Gaussian approximation ($r = 15$).

Figure 10. Approximated CDF $\hat{F}_{i(\xi;\eta)-I(\xi;\eta)}(\cdot, n)$ for $r \in \{2, 5, 10, 15\}$ canonical correlations $\varrho_i(T)$ given in (71) for $T = 1$ (approximation error $< 1 \times 10^{-2}$) vs. Gaussian approximation ($r = 15$).

7. Summary of Contributions

We derived series representations of the PDF and CDF of the information density for arbitrary Gaussian random vectors as well as a general formula for the central moments using canonical correlation analysis. We provided simplified and closed-form expressions for important special cases, in particular when all canonical correlations are equal, and derived recurrence formulas and uniform error bounds for finite sum approximations of the general series representations. These approximations and recurrence formulas are suitable for efficient and arbitrarily accurate numerical calculations, where the approximation error can be easily controlled with the derived error bounds. Moreover, we provided examples showing the (in)validity of approximating the information density with a Gaussian random variable.

Author Contributions: J.E.W.H. and M.M. conceived this work, performed the analysis, validated the results, and wrote the manuscript. All authors have read and agreed to this version of the manuscript.

Data Availability Statement: An implementation in Python allowing efficient numerical calculations related to the main results of the paper is publicly available on GitLab: https://gitlab.com/infth/information-density (accessed on 24 June 2022).

References

1. Han, T.S.; Verdú, S. Approximation Theory of Output Statistics. *IEEE Trans. Inf. Theory* **1993**, *39*, 752–772. [CrossRef]
2. Han, T.S. *Information-Spectrum Methods in Information Theory*; Springer: Berlin/Heidelberg, Germany, 2003.
3. Shannon, C.E. Probability of Error for Optimal Codes in a Gaussian Channel. *Bell Syst. Tech. J.* **1959**, *38*, 611–659. [CrossRef]
4. Dobrushin, R.L. Mathematical Problems in the Shannon Theory of Optimal Coding of Information. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*; Volume 1: Contributions to the Theory of Statistics; University of California Press: Berkeley, CA, USA, 1961; pp. 211–252.

5. Strassen, V. Asymptotische Abschätzungen in Shannons Informationstheorie. In *Transactions of the Third Prague Conference on Information Theory, Statistical Decision Functions, Random Processes (Held 1962)*; Czechoslovak Academy of Sciences: Prague, Czech Republic, 1964; pp. 689–723.

6. Polyanskiy, Y.; Poor, H.V.; Verdú, S. Channel Coding Rate in the Finite Blocklength Regime. *IEEE Trans. Inf. Theory* **2010**, *56*, 2307–2359. [CrossRef]

7. Durisi, G.; Koch, T.; Popovski, P. Toward Massive, Ultrareliable, and Low-Latency Wireless Communication With Short Packets. *Proc. IEEE* **2016**, *104*, 1711–1726. [CrossRef]

8. Pinsker, M.S. *Information and Information Stability of Random Variables and Processes*; Holden-Day: San Francisco, CA, USA, 1964.

9. Olver, F.W.J.; Lozier, D.W.; Boisvert, R.F.; Clark, C.W. (Eds.) *NIST Handbook of Mathematical Functions*; Cambridge University Press: Cambridge, UK, 2010.

10. Mathai, A.M. Storage Capacity of a Dam With Gamma Type Inputs. *Ann. Inst. Stat. Math.* **1982**, *34*, 591–597. [CrossRef]

11. Grad, A.; Solomon, H. Distribution of Quadratic Forms and Some Applications. *Ann. Math. Stat.* **1955**, *26*, 464–477. [CrossRef]

12. Kotz, S.; Johnson, N.L.; Boyd, D.W. Series Representations of Distributions of Quadratic Forms in Normal Variables. I. Central Case. *Ann. Math. Stat.* **1967**, *38*, 823–837. [CrossRef]

13. Huffmann, J.E.W.; Mittelbach, M. On the Distribution of the Information Density of Gaussian Random Vectors: Explicit Formulas and Tight Approximations. *Entropy* **2021**, *24*, 924.

14. Simon, M.K. *Probability Distributions Involving Gaussian Random Variables: A Handbook for Engineers and Scientists*; Springer: Berlin/Heidelberg, Germany, 2006.

15. Laneman, J.N. On the Distribution of Mutual Information. In Proceedings of the Workshop Information Theory and Its Applications (ITA), San Diego, CA, USA, 13 February 2006.

16. Wu, P.; Jindal, N. Coding Versus ARQ in Fading Channels: How Reliable Should the PHY Be? *IEEE Trans. Commun.* **2011**, *59*, 3363–3374. [CrossRef]

17. Buckingham, D.; Valenti, M.C. The Information-Outage Probability of Finite-Length Codes Over AWGN Channels. In Proceedings of the 42nd Annual Conference on Information Sciences and Systems (CISS), Princeton, NJ, USA, 19–21 March 2008.

18. Hotelling, H. Relations Between Two Sets of Variates. *Biometrika* **1936**, *28*, 321–377. [CrossRef]

19. Gelfand, I.M.; Yaglom, A.M. Calculation of the Amount of Information About a Random Function Contained in Another Such Function. In *AMS Translations, Series 2*; AMS: Providence, RI, USA, 1959; Volume 12, pp. 199–246.

20. Härdle, W.K.; Simar, L. *Applied Multivariate Statistical Analysis*, 4th ed.; Springer: Berlin/Heidelberg, Germany, 2015.

21. Koch, I. *Analysis of Multivariate and High-Dimensional Data*; Cambridge University Press: Cambridge, UK, 2014.

22. Timm, N.H. *Applied Multivariate Analysis*; Springer: Berlin/Heidelberg, Germany, 2002.

23. Ibragimov, I.A.; Rozanov, Y.A. On the Connection Between Two Characteristics of Dependence of Gaussian Random Vectors. *Theory Probab. Appl.* **1970**, *15*, 295–299. [CrossRef]

24. Gradshteyn, I.S.; Ryzhik, I.M. *Table of Integrals, Series, and Products*, 7th ed.; Elsevier: Amsterdam, The Netherlands, 2007.

25. Prudnikov, A.P.; Brychov, Y.A.; Marichev, O.I. *Integrals and Series, Volume 2: Special Functions*; Gordon and Breach Science: New York, NY, USA, 1986.

26. Huffmann, J.E.W.; Mittelbach, M. Efficient Python Implementation to Numerically Calculate PDF, CDF, and Moments of the Information Density of Gaussian Random Vectors. Source Code Provided on GitLab, 2021. Available online: https://gitlab.com/infth/information-density (accessed on 24 June 2022).

27. Moschopoulos, P.G. The Distribution of the Sum of Independent Gamma Random Variables. *Ann. Inst. Stat. Math.* **1985**, *37*, 541–544. [CrossRef]

28. Comtet, L. *Advanced Combinatorics: The Art of Finite and Infinite Expansions*, Revised and Enlarged ed.; D. Reidel Publishing Company: Dordrecht, The Netherlands, 1974.

29. Grenander, U.; Szegö, G. *Toeplitz Forms and Their Applications*; University of California Press: Berkeley, CA, USA, 1958.

30. Huffmann, J.E.W. Canonical Correlation and the Calculation of Information Measures for Infinite-Dimensional Distributions. Diploma Thesis, Department of Electrical Engineering and Information Technology, Technische Universität Dresden, Dresden, Germany, 2021. Available online: https://nbn-resolving.org/urn:nbn:de:bsz:14-qucosa2-742541 (accessed on 24 June 2022).