

Article

A Partial Information Decomposition for Multivariate Gaussian Systems Based on Information Geometry

Jim W. Kay 

School of Mathematics and Statistics, University of Glasgow, Glasgow G12 8QQ, UK; jim.kay@glasgow.ac.uk

Abstract: There is much interest in the topic of partial information decomposition, both in developing new algorithms and in developing applications. An algorithm, based on standard results from information geometry, was recently proposed by Niu and Quinn (2019). They considered the case of three scalar random variables from an exponential family, including both discrete distributions and a trivariate Gaussian distribution. The purpose of this article is to extend their work to the general case of multivariate Gaussian systems having vector inputs and a vector output. By making use of standard results from information geometry, explicit expressions are derived for the components of the partial information decomposition for this system. These expressions depend on a real-valued parameter which is determined by performing a simple constrained convex optimisation. Furthermore, it is proved that the theoretical properties of non-negativity, self-redundancy, symmetry and monotonicity, which were proposed by Williams and Beer (2010), are valid for the decomposition I_{ig} derived herein. Application of these results to real and simulated data show that the I_{ig} algorithm does produce the results expected when clear expectations are available, although in some scenarios, it can overestimate the level of the synergy and shared information components of the decomposition, and correspondingly underestimate the levels of unique information. Comparisons of the I_{ig} and I_{dep} (Kay and Ince, 2018) methods show that they can both produce very similar results, but interesting differences are provided. The same may be said about comparisons between the I_{ig} and I_{mmi} (Barrett, 2015) methods.

Keywords: partial information decomposition; mutual information; synergy; redundancy; information geometry



Citation: Kay, J.W. A Partial Information Decomposition for Multivariate Gaussian Systems Based on Information Geometry. *Entropy* **2024**, *26*, 542. <https://doi.org/10.3390/e26070542>

Academic Editor: Carlo Cafaro

Received: 22 May 2024
Revised: 17 June 2024
Accepted: 18 June 2024
Published: 25 June 2024



Copyright: © 2024 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Williams and Beer [1] introduced a new method for the decomposition of information in a probabilistic system termed *partial information decomposition* (PID). This allows the joint mutual information between a number of input sources and a target (output) to be decomposed into components which quantify different aspects of the transmitted information in the system. These are the unique information that each source conveys about the target; the shared information that all sources possess about the target; the synergistic information that the sources in combination possess regarding the target. An additional achievement was to prove that the interaction information [2] is actually the difference between the synergy and redundancy in a system. Thus, a positive value for interaction information signifies that there is more synergy than redundancy in the system, while a negative value indicates the opposite. The work by Williams and Beer has led to many new methods for defining a PID, mainly for discrete probabilistic systems [3–13] spawning a variety of applications [14–19].

There has been considerable interest in PID methods for Gaussian systems. The case of static and dynamic Gaussian systems with two scalar sources and a scalar target was considered in [20], which applied the minimum mutual information PID, I_{mmi} . Further insights were developed in [21] regarding synergy. A PID for Gaussian systems based on common surprisal was published in [7]. Barrett's work [20] was extended to multivariate

Gaussian systems with two vector sources and a vector target in [22] using the I_{dep} method which was introduced for discrete systems in [8]. Further work based on the concept of statistical deficiency is reported in [23]. Application of PID for Gaussian systems has been used in a range of applications [18,24–30].

We focus in particular here on the method proposed by Niu and Quinn [3]. They applied standard results from information geometry [31–33] in order to define a PID for three scalar random variables which follow an exponential family distribution, including a trivariate Gaussian distribution.

Here, we extend this work in two ways: (a) we provide general formulae for a PID involving multivariate Gaussian systems which have two vector sources and a vector target by making use of the same standard methods from information geometry as in [3] and (b) we prove that the Williams–Beer properties of non-negativity, self-redundancy, symmetry and monotonicity are valid for this PID. We also provide some illustrations of the resulting algorithm using real and simulated data. The PID developed herein is based on some of the probability models in the same partially ordered lattice on which the I_{dep} algorithm is based. Therefore, we also compare the results obtained with those obtained by using the I_{dep} method. The I_{ig} results are also compared with those obtained using the I_{mmi} algorithm.

2. Methods

2.1. Notation

A generic ‘ p ’ will be used to denote an absolutely continuous probability density function (pdf), with the arguments of the function signifying which distribution is intended. Bold capital letters are used to denote random vectors, with their realised values appearing in bold lowercase—so that $p(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$ denotes the joint pdf of the random vectors, $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3$, while $p(\mathbf{x}_1, \mathbf{x}_3 | \mathbf{x}_2)$ is the conditional pdf of $(\mathbf{X}_1, \mathbf{X}_3)$ given a value for $\mathbf{X}_2$.

We consider the case where random vectors $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3$, of dimensions n_1, n_2, n_3 , respectively, have partitioned mean vectors equal to zero vectors of lengths n_1, n_2, n_3 , respectively, and a conformably partitioned covariance matrix. We stack these random vectors into the random vector $\mathbf{Z}$, so that $\mathbf{Z}$ has dimension $m = n_1 + n_2 + n_3$, and assume that $\mathbf{Z}$ has a positive definite multivariate Gaussian distribution with pdf $p(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$, mean vector $\mathbf{0}$ and covariance matrix given by

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{13} \\ \Sigma_{12}^T & \Sigma_{22} & \Sigma_{23} \\ \Sigma_{13}^T & \Sigma_{23}^T & \Sigma_{33} \end{bmatrix}, \quad (1)$$

where the covariance matrices $\Sigma_{11}, \Sigma_{22}, \Sigma_{33}$ of $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3$, respectively, are of sizes $n_1 \times n_1, n_2 \times n_2, n_3 \times n_3$, and $\Sigma_{12}, \Sigma_{13}, \Sigma_{23}$ are the pairwise cross-covariance matrices between the three vectors $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3$. We also denote the conformably partitioned precision (or concentration) matrix K by

$$K = \begin{bmatrix} K_{11} & K_{12} & K_{13} \\ K_{12}^T & K_{22} & K_{23} \\ K_{13}^T & K_{23}^T & K_{33} \end{bmatrix},$$

where $K = \Sigma^{-1}$. The pdf of $\mathbf{Z}$ is

$$f(\mathbf{z} | K) = \frac{\sqrt{|K|}}{(2\pi)^{\frac{1}{2}m}} \exp\left(-\frac{1}{2}\mathbf{z}^T K \mathbf{z}\right) \equiv \frac{\sqrt{|K|}}{(2\pi)^{\frac{1}{2}m}} \exp\left(-\frac{1}{2} \sum_{i=1}^3 \sum_{j=1}^3 \mathbf{x}_i^T K_{ij} \mathbf{x}_j\right), \quad \mathbf{z} \in \mathbb{R}^m. \quad (2)$$

2.2. Some Information Geometry

We now describe some standard results from information geometry [32,33] as applied to zero-mean, partitioned multivariate Gaussian probability distributions. The fact that

there is no loss of generality in making this zero-mean assumption will be justified by Lemma 1 in Section 3. The multivariate Gaussian pdf defined in (2) may be written in the form

$$\frac{\sqrt{|K|}}{(2\pi)^{\frac{1}{2}m}} \exp\left(-\frac{1}{2}(\mathbf{x}_1^T K_{11} \mathbf{x}_1 + \mathbf{x}_2^T K_{22} \mathbf{x}_2 + \mathbf{x}_3^T K_{33} \mathbf{x}_3 - (\mathbf{x}_2^T K_{12}^T \mathbf{x}_1 + \mathbf{x}_3^T K_{13}^T \mathbf{x}_1 + \mathbf{x}_3^T K_{23}^T \mathbf{x}_2))\right),$$

which may be written in terms of the Frobenius inner product as

$$\exp\left(\left\langle -\frac{1}{2}K_{11}, \mathbf{x}_1 \mathbf{x}_1^T \right\rangle + \left\langle -\frac{1}{2}K_{22}, \mathbf{x}_2 \mathbf{x}_2^T \right\rangle + \left\langle -\frac{1}{2}K_{33}, \mathbf{x}_3 \mathbf{x}_3^T \right\rangle + \left\langle -K_{12}, \mathbf{x}_1 \mathbf{x}_2^T \right\rangle + \left\langle -K_{13}, \mathbf{x}_1 \mathbf{x}_3^T \right\rangle + \left\langle -K_{23}, \mathbf{x}_2 \mathbf{x}_3^T \right\rangle - \psi(\boldsymbol{\theta})\right),$$

where

$$\psi(\boldsymbol{\theta}) = -\frac{1}{2} \log \frac{|K|}{(2\pi)^m}.$$

This is of exponential family form [33] (p. 34) and [34] with natural parameter

$$\boldsymbol{\theta} = \left\{-\frac{1}{2}K_{11}, -\frac{1}{2}K_{22}, -\frac{1}{2}K_{33}, -K_{12}, -K_{13}, -K_{23}\right\},$$

and expectation parameter $\boldsymbol{\eta} = \{\eta_{ij}, i = 1, 2, 3; j = 1, 2, 3; i \leq j\}$, where $\eta_{ij} = \mathbb{E}(\mathbf{x}_i \mathbf{x}_j^T) = \Sigma_{ij}$. We note that there is something of a terminal ambiguity here, since a ‘parameter’ is usually a real number. It is convenient to use the more compact notation provided by matrices since this enables all of the elements of a matrix natural parameter to be set to zero simultaneously.

The exponential family distribution in (2) is a dually flat manifold [31], which we denote by M .

We define the following e -flat submanifolds of M :

$$\begin{aligned} S_7 &= \{p \in M : K_{12} = 0\} \\ S_6 &= \{p \in M : K_{13} = 0\} \\ S_5 &= \{p \in M : K_{23} = 0\} \\ S_4 &= \{p \in M : K_{12} = 0 \text{ and } K_{13} = 0\} \\ S_3 &= \{p \in M : K_{12} = 0 \text{ and } K_{23} = 0\} \\ S_2 &= \{p \in M : K_{13} = 0 \text{ and } K_{23} = 0\} \\ S_1 &= \{p \in M : K_{12} = 0, K_{13} = 0 \text{ and } K_{23} = 0\} \end{aligned}$$

which may be conveniently pictured as the partially ordered lattice in Figure 1. The submanifolds S_5 and S_6 are necessary for the definition of the information-geometric PID [3] and the others will be considered in the sequel. Lattices similar to that in Figure 1 appear in [8,35,36] in relation to information decomposition, and in [37] who consider dually flat manifolds on posets. See also [38], and references therein, for the use of a variety of lattices of models in statistical work.

Hierarchical chains of submanifolds were considered in [31] but here the submanifolds are not all in a hierarchical chain due to the presence of two antichains: $\{S_2, S_3, S_4\}$ and $\{S_5, S_6, S_7\}$. There are, however, several useful chains within the lattice. Of particular relevance here are the chains $\{S_2, S_5, M\}$, $\{S_2, S_6, M\}$ and $\{S_2, M\}$. Application of Amari’s mixed-cut coordinates [31] and calculation of divergences produces measures of mutual information that are of direct relevance in PID (as was noted by [3] for three scalar random variables) in that the equations

$$\begin{aligned}
 I[\mathbf{X}_1; \mathbf{X}_2; \mathbf{X}_3] &= I[\mathbf{X}_1; \mathbf{X}_3] + I[\mathbf{X}_2; \mathbf{X}_3 | \mathbf{X}_1] \\
 &= I[\mathbf{X}_2; \mathbf{X}_3] + I[\mathbf{X}_1; \mathbf{X}_3 | \mathbf{X}_2]
 \end{aligned}$$

are obtained—and they are standard results in information theory based on the chain rule for mutual information [39]. These are nice illustrations of Amari’s method.

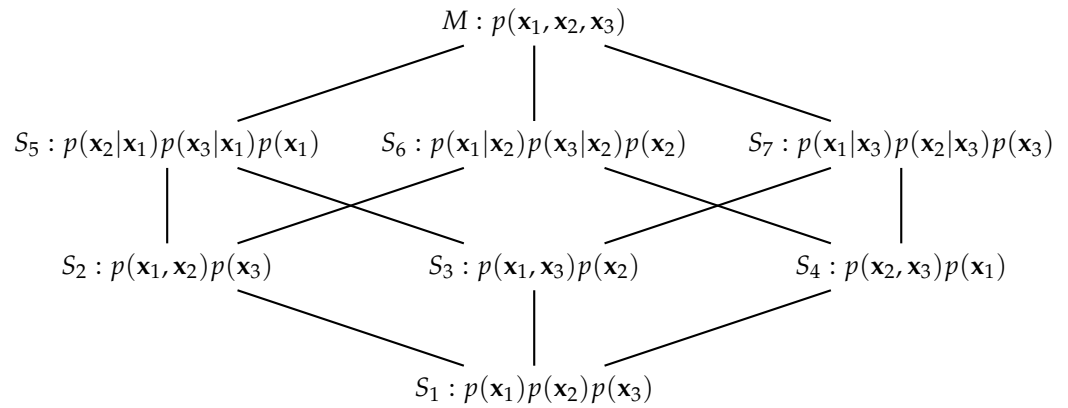


Figure 1. A partially ordered lattice of the manifold M and submanifolds $S_1 - S_7$. The form of the pdf that is shown for each submanifold is that obtained by m -projection of the distribution $p(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$ onto the submanifold.

We now consider m -projections from the pdf $p \in M$ to each of the submanifolds, $S_1 - S_7$ [31]. It is easy to find the pdf in each submanifold that is closest to the given pdf in M , p , in terms of Kullback–Leibler (KL) divergence [40], ([Ch. 4]). They are given in Figure 1. We know [34,40] that setting a block of the inverse covariance for a multivariate Gaussian distribution to zero expresses a conditional dependence between the variables involved. For example, consider S_5 . On this submanifold $K_{23} = 0$ and so $\mathbf{X}_2$ and $\mathbf{X}_3$ are conditionally independent given a value for $\mathbf{X}_1$. Therefore, this pdf, which we denote by p_5 , has the form $p_5(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) = p(\mathbf{x}_2 | \mathbf{x}_1) p(\mathbf{x}_3 | \mathbf{x}_1) p(\mathbf{x}_1)$. On submanifold S_2 , there are two conditional independences and so $\mathbf{X}_3$ and the pair $(\mathbf{X}_1, \mathbf{X}_2)$ are independent and the closest pdf in S_2 to the pdf p has the form $p_2(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) = p(\mathbf{x}_1, \mathbf{x}_2) p(\mathbf{x}_3)$.

The probability distributions defined by these information projections could also have been obtained by the method of maximum entropy, subject to constraints on model interactions [31], and they were obtained in this manner in [22] by making use of Gaussian graphical models [34,40].

We now mention important results from information geometry which are crucial for defining a PID [3]. Consider the pdfs p_5, p_6, p belonging to the submanifolds S_5, S_6 , and to manifold M , and the e -geodesic passing through p_5 and p_6 . Then, any pdf on this e -geodesic path is also a zero-mean multivariate Gaussian pdf [41], ([Ch. 1]). We denote such a pdf by p_t . It has covariance matrix Σ_t , defined by

$$\Sigma_t^{-1} = (1 - t)\Sigma_5^{-1} + t\Sigma_6^{-1} \quad (3)$$

provided that Σ_t^{-1} is positive definite. We consider also an m -geodesic from p to p_t . Then, by standard results [31,33], this m -geodesic meets the e -geodesic through p_5 and p_6 at a unique pdf p_{t^*} such that generalized Pythagorean relationships hold in terms of the KL divergence:

$$D(p || p_5) = D(p || p_{t^*}) + D(p_{t^*} || p_5) \quad (4)$$

$$D(p || p_6) = D(p || p_{t^*}) + D(p_{t^*} || p_6) \quad (5)$$

The pdf $p(t^*)$ minimizes the KL divergence between the pdf p in M and the pdf p_t which lies on the e -geodesic which passes through pdfs p_5 and p_6 .

2.3. The Partial Information Decomposition

Williams and Beer [1] introduce a framework called the partial information decomposition (PID) which decomposes the joint mutual information between a target and a set of multiple predictor variables into a series of terms reflecting information which is shared, unique or synergistically available within and between subsets of predictors. The joint mutual information, conditional mutual information and bivariate mutual information are defined as follows.

$$\begin{aligned} I[\mathbf{X}_1, \mathbf{X}_2; \mathbf{X}_3] &= \iiint p(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) \log \frac{p(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)}{p(\mathbf{x}_1, \mathbf{x}_2)p(\mathbf{x}_3)} d\mathbf{x}_1 d\mathbf{x}_2 d\mathbf{x}_3, \\ I[\mathbf{X}_1; \mathbf{X}_3 | \mathbf{X}_2] &= \iiint p(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) \log \frac{p(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)p(\mathbf{x}_2)}{p(\mathbf{x}_1, \mathbf{x}_2)p(\mathbf{x}_2, \mathbf{x}_3)} d\mathbf{x}_1 d\mathbf{x}_2 d\mathbf{x}_3, \\ I[\mathbf{X}_1; \mathbf{X}_3] &= \iint p(\mathbf{x}_1, \mathbf{x}_3) \log \frac{p(\mathbf{x}_1, \mathbf{x}_3)}{p(\mathbf{x}_1)p(\mathbf{x}_3)} d\mathbf{x}_1 d\mathbf{x}_3, \end{aligned}$$

Here, we focus on the case of two vector sources, $\mathbf{X}_1, \mathbf{X}_2$, and a vector target $\mathbf{X}_3$. Adapting the notation of [42], we express the joint mutual information in four terms as follows:

$\text{Unq1} \equiv I_{\text{unq}}[\mathbf{X}_1; \mathbf{X}_3 \mathbf{X}_2]$	denotes the unique information that $\mathbf{X}_1$ conveys about $\mathbf{X}_3$;
$\text{Unq2} \equiv I_{\text{unq}}[\mathbf{X}_2; \mathbf{X}_3 \mathbf{X}_1]$	is the unique information that $\mathbf{X}_2$ conveys about $\mathbf{X}_3$;
$\text{Shd} \equiv I_{\text{shd}}[\mathbf{X}_1, \mathbf{X}_2; \mathbf{X}_3]$	gives the common (or redundant or shared) information that both $\mathbf{X}_1$ and $\mathbf{X}_2$ have about $\mathbf{X}_3$;
$\text{Syn} \equiv I_{\text{syn}}[\mathbf{X}_1, \mathbf{X}_2; \mathbf{X}_3]$	is the synergy or information that the joint vector $(\mathbf{X}_1, \mathbf{X}_2)$ has about $\mathbf{X}_3$ that cannot be obtained by observing $\mathbf{X}_1$ and $\mathbf{X}_2$ separately.

It is possible to make deductions about a PID by using the following four equations which give a link between the components of a PID and certain classical Shannon measures of mutual information. The following are from ([42] Equations (4) and (5)), with amended notation; see also [1].

$$I[\mathbf{X}_1; \mathbf{X}_3] = \text{Unq1} + \text{Shd}, \quad (6)$$

$$I[\mathbf{X}_2; \mathbf{X}_3] = \text{Unq2} + \text{Shd}, \quad (7)$$

$$I[\mathbf{X}_1; \mathbf{X}_3 | \mathbf{X}_2] = \text{Unq1} + \text{Syn}, \quad (8)$$

$$I[\mathbf{X}_2; \mathbf{X}_3 | \mathbf{X}_1] = \text{Unq2} + \text{Syn}. \quad (9)$$

Also, the joint mutual information may be written as

$$I[\mathbf{X}_1, \mathbf{X}_2; \mathbf{X}_3] = \text{Syn} + \text{Unq1} + \text{Unq2} + \text{Shd}. \quad (10)$$

The Equations (6)–(9) are of rank 3 and so it is necessary to provide a value for any one of the components, and then the remaining terms can be easily calculated. The initial formulation of [1] was based on quantifying the shared information and deriving the other quantities, but others have focussed on quantifying unique information or synergy directly [4,5,8]. Also, the following form [16] of the interaction information [2] will be useful. It was shown [1] to be equal to the difference in $\text{Syn} - \text{Shd}$.

$$II[\mathbf{X}_1; \mathbf{X}_2; \mathbf{X}_3] = I[\mathbf{X}_1; \mathbf{X}_2 | \mathbf{X}_3] - I[\mathbf{X}_1; \mathbf{X}_2]. \quad (11)$$

3. Results

3.1. A PID for Gaussian Vector Sources and a Gaussian Vector Target

We now apply the results from the previous two sections in order to derive a partial information decomposition by making use of the method defined in [3]. The following lemma will confirm that without any loss of generality, we may assume, for all of the multivariate normal distributions considered herein, that the mean vector can be taken

to be $\mathbf{0}$ and the covariance matrix of $\mathbf{Z}$, defined on $\mathbb{R}^m$, where $m = n_1 + n_2 + n_3$, can have the form

$$\Sigma = \begin{bmatrix} I_{n_1} & P & Q \\ P^T & I_{n_2} & R \\ Q^T & R^T & I_{n_3} \end{bmatrix}, \quad (12)$$

where the matrices P, Q, R are of size $n_1 \times n_2, n_1 \times n_3, n_2 \times n_3$, respectively, and are the cross-covariance (correlation) matrices between the three pairings of the three random vectors $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3$, and so

$$\mathbb{E}(\mathbf{X}_1 \mathbf{X}_2^T) = P, \quad \mathbb{E}(\mathbf{X}_1 \mathbf{X}_3^T) = Q, \quad \mathbb{E}(\mathbf{X}_2 \mathbf{X}_3^T) = R.$$

The calculation of the partial information coefficients will involve the computation of KL divergences [43] between two multivariate Gaussian distributions associated with two submanifolds in the lattice, defined in Figure 1; see Lemma 1, with proof in Appendix C. These probability distributions will have two features in common: they each have the same partitioned mean vector and also the same variance–covariance matrices for the random vectors $\mathbf{X}_1, \mathbf{X}_2$ and $\mathbf{X}_3$, but different cross covariance matrices for each pair of the random vectors $\mathbf{X}_1, \mathbf{X}_2$ and $\mathbf{X}_3$.

Lemma 1. Consider two multivariate Gaussian pdfs, f_1 and f_2 , which have the same partitioned mean vector, $\boldsymbol{\mu} = [\boldsymbol{\mu}_1^T, \boldsymbol{\mu}_2^T, \boldsymbol{\mu}_3^T]^T$, and conformably partitioned $m \times m$ covariance matrices

$$\Phi = \begin{bmatrix} \Sigma_{11} & \Phi_{12} & \Phi_{13} \\ \Sigma_{12}^T & \Sigma_{22} & \Phi_{23} \\ \Phi_{13}^T & \Phi_{23}^T & \Sigma_{33} \end{bmatrix} \quad \text{and} \quad \Lambda = \begin{bmatrix} \Sigma_{11} & \Lambda_{12} & \Lambda_{13} \\ \Lambda_{12}^T & \Sigma_{22} & \Lambda_{23} \\ \Lambda_{13}^T & \Lambda_{23}^T & \Sigma_{33} \end{bmatrix} \quad (13)$$

respectively, where the diagonal blocks Σ_{11}, Σ_{22} , and Σ_{33} are square.

Then, the Kullback–Liebler divergence $D(f_1 || f_2)$ does not depend on the mean vector $\boldsymbol{\mu}$, nor does it depend directly on the variance–covariance matrices $\Sigma_{11}, \Sigma_{22}, \Sigma_{33}$. The divergence is equal to

$$D(f_1 || f_2) = \frac{1}{2} \left(\log \frac{|\Lambda_1|}{|\Phi_1|} - m + \text{Tr}(\Phi_1^{-1} \Lambda_1) \right), \quad (14)$$

where

$$\Phi_1 = \begin{bmatrix} I_{n_1} & P_{12} & P_{13} \\ P_{12}^T & I_{n_2} & P_{23} \\ P_{13}^T & P_{23}^T & I_{n_3} \end{bmatrix} \quad \text{and} \quad \Lambda_1 = \begin{bmatrix} I_{n_1} & L_{12} & L_{13} \\ L_{12}^T & I_{n_2} & L_{23} \\ L_{13}^T & L_{23}^T & I_{n_3} \end{bmatrix}$$

with

$$P_{ij} = \Sigma_{ii}^{-\frac{1}{2}} \Phi_{ij} \Sigma_{jj}^{-\frac{1}{2}}, \quad L_{ij} = \Sigma_{ii}^{-\frac{1}{2}} \Lambda_{ij} \Sigma_{jj}^{-\frac{1}{2}}, \quad (i, j = 1, 2, 3; i \leq j),$$

which are the respective cross-correlation matrices among $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3$. The KL divergence depends only on these cross-correlation matrices.

3.2. Covariance Matrices

Table 1 gives the covariance matrices corresponding to each of the projected distributions $p_1 - p_7$ on the submanifolds. It is known from Gaussian graphical models [34,40] that the probability distributions associated with submanifolds S_5 and S_6 are defined by setting $K_{23} = 0$ and $K_{13} = 0$, respectively, in the precision matrix K . These conditions were shown in [22] to be equivalent to the equations $R = P^T Q$ and $Q = PR$, respectively. From Table 1, we see that the covariance matrices for pdfs p_5 and p_6 have the following form.

$$\Sigma_5 = \begin{bmatrix} I_{n_1} & P & Q \\ P^T & I_{n_2} & P^T Q \\ Q^T & Q^T P & I_{n_3} \end{bmatrix}, \quad \text{and} \quad \Sigma_6 = \begin{bmatrix} I_{n_1} & P & PR \\ P^T & I_{n_2} & R \\ R^T P^T & R^T & I_{n_3} \end{bmatrix}. \quad (15)$$

The following lemma, which is proved in Appendix D, gives some useful results on determinants that will be used in the sequel.

Lemma 2. *The determinants of the matrices Σ_5, Σ_6 are given by*

$$|\Sigma_5| = |I_{n_2} - P^T P| |I_{n_3} - Q^T Q| \quad \text{and} \quad |\Sigma_6| = |I_{n_2} - P^T P| |I_{n_3} - R^T R|.$$

Also,

$$\Sigma_5 = \Sigma_6 \iff Q = 0 \text{ and } R = 0 \iff I(\mathbf{X}_1, \mathbf{X}_2; \mathbf{X}_3) = 0.$$

Table 1. Submanifold probability distributions with corresponding covariance matrices (modified from [22]).

Submanifold	Σ_i	Submanifold	Σ_i
$S_1 : p(\mathbf{x}_1)p(\mathbf{x}_2)p(\mathbf{x}_3)$	$\begin{bmatrix} I_{n_1} & 0 & 0 \\ 0 & I_{n_2} & 0 \\ 0 & 0 & I_{n_3} \end{bmatrix}$	$S_5 : p(\mathbf{x}_2 \mathbf{x}_1)p(\mathbf{x}_3 \mathbf{x}_1)p(\mathbf{x}_1)$	$\begin{bmatrix} I_{n_1} & P & Q \\ P^T & I_{n_2} & P^T Q \\ Q^T & Q^T P & I_{n_3} \end{bmatrix}$
$S_2 : p(\mathbf{x}_1, \mathbf{x}_2)p(\mathbf{x}_3)$	$\begin{bmatrix} I_{n_1} & P & 0 \\ P^T & I_{n_2} & 0 \\ 0 & 0 & I_{n_3} \end{bmatrix}$	$S_6 : p(\mathbf{x}_1 \mathbf{x}_2)p(\mathbf{x}_3 \mathbf{x}_2)p(\mathbf{x}_2)$	$\begin{bmatrix} I_{n_1} & P & PR \\ P^T & I_{n_2} & R \\ R^T P^T & R^T & I_{n_3} \end{bmatrix}$
$S_3 : p(\mathbf{x}_1, \mathbf{x}_3)p(\mathbf{x}_2)$	$\begin{bmatrix} I_{n_1} & 0 & Q \\ 0 & I_{n_2} & 0 \\ Q^T & 0 & I_{n_3} \end{bmatrix}$	$S_7 : p(\mathbf{x}_1 \mathbf{x}_3)p(\mathbf{x}_2 \mathbf{x}_3)p(\mathbf{x}_3)$	$\begin{bmatrix} I_{n_1} & QR^T & Q \\ RQ^T & I_{n_2} & R \\ Q^T & R^T & I_{n_3} \end{bmatrix}$
$S_4 : p(\mathbf{x}_2, \mathbf{x}_3)p(\mathbf{x}_1)$	$\begin{bmatrix} I_{n_1} & 0 & 0 \\ 0 & I_{n_2} & R \\ 0 & R^T & I_{n_3} \end{bmatrix}$	$M : p(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$	$\begin{bmatrix} I_{n_1} & P & Q \\ P^T & I_{n_2} & R \\ Q^T & R^T & I_{n_3} \end{bmatrix}$

3.3. Feasible Values for the Parameter t

From (3), the m -projection from manifold M to the e -geodesic passing through the pdfs p_5 and p_6 meets in general at pdf p_t which has covariance matrix Σ_t defined by

$$\Sigma_t^{-1} = (1-t)\Sigma_5^{-1} + t\Sigma_6^{-1},$$

and Σ_t must be positive definite. Therefore, when finding the optimal pdf p_{t^*} , we require to constrain the values of the parameter t to be such that Σ_t is positive definite. We define the set of feasible values for t as

$$F = \{t \in \mathbb{R} : \Sigma_t \text{ is positive definite}\}. \quad (16)$$

F is a closed interval in $\mathbb{R}$ of the form $[-a, 1+b]$, where $a, b > 0$. The interior of F —the open interval $]-a, 1+b[$ —is an open convex set. To enable the derivation of explicit results, it is useful to define the matrix Σ'_t by

$$\Sigma_t'^{-1} = (1-t)\Sigma_6 + t\Sigma_5. \quad (17)$$

We also require a feasible value t^* for t when working with the matrix Σ'_t , and so we define the set G of feasible values as follows

$$G = \{t \in \mathbb{R} : \Sigma'_t \text{ is positive definite}\}. \quad (18)$$

It turns out that the sets of feasible values F, G are actually the same set, as stated in the following lemma, which is proved in Appendix E, and this fact allows us to infer that Σ'_t is positive definite when Σ_t is.

Lemma 3. *If the parameter t belongs to the closed interval $[0,1]$, then the matrices Σ_t and Σ'_t are both positive definite. Also, the two feasible sets F and G defined above in (16) and (18) are equal.*

3.4. A Convex Optimisation Problem

The optimal value t^* of the parameter t is defined by

$$t^* = \arg \min_{t \in F} D(p||p_t) \quad (19)$$

The following lemma, with proof in Appendix F, provides details of the optimisation required to find t^* .

Lemma 4. For $t \in F^\circ$, we define the real valued function g by $g(t) = D(p||p_t)$. Then,

$$\text{Tr}(\Sigma_t^{-1}\Sigma) = m, \quad (20)$$

and

$$g(t) = \frac{1}{2} \log \frac{|I_{n_2} - P^T P|^2 |I_{n_3} - Q^T Q| |I_{n_3} - R^T R|}{|\Sigma|} - \frac{1}{2} \log |(1-t)\Sigma_6 + t\Sigma_5| \quad (21)$$

Provided that the joint mutual information is positive, the minimization of g subject to the constraint $t \in F^\circ$, an open convex set, is a strictly convex problem, and the optimal value t^* is unique.

The minimum value of g is equal to

$$\frac{1}{2} \log \frac{|I_{n_2} - P^T P| |I_{n_3} - Q^T Q| |I_{n_3} - R^T R|}{|\Sigma|} + \frac{1}{2} \log \frac{1}{d(t^*)}, \quad (22)$$

where the determinant $d(t)$ is defined by

$$d(t) = |I_{n_3} - (1-t)^2 Q^T Q - t(1-t)(R^T P^T Q + Q^T P R) - t^2 R^T R|, \quad (t \in F) \quad (23)$$

Alternatively, the minimum could occur at either endpoint of F .

We now define the PID components.

3.5. Definition of the PID Components

Following the proposal in [3], we define the synergy of the system to be

$$\text{Syn} = D(p||p_t^*) \quad (24)$$

and by Lemma 1 and (20) the expression for the synergy is

$$\text{Syn} = \frac{1}{2} \log \frac{|\Sigma_{t^*}|}{|\Sigma|} - \frac{1}{2} m + \frac{1}{2} \text{Tr}(\Sigma_{t^*}^{-1}\Sigma) = \frac{1}{2} \log \frac{|\Sigma_{t^*}|}{|\Sigma|}. \quad (25)$$

Before defining the other PID terms, we require the following lemma, with proof in Appendix G.

Lemma 5. The trace terms required in the definitions of the unique information are both equal to m :

$$\text{Tr}(\Sigma_5^{-1}\Sigma_{t^*}) = m \quad \text{and} \quad \text{Tr}(\Sigma_6^{-1}\Sigma_{t^*}) = m. \quad (26)$$

From (4), we know that

$$D(p||p_5) = D(p||p_t^*) + D(p_t^*||p_5) \quad (27)$$

and we define the unique information in the system that is due to source $\mathbf{X}_2$ to be

$$\text{Unq}_2 = D(p_t^*||p_5) \quad (28)$$

as in [3]. By (5), we also have that

$$D(p||p_6) = D(p||p_t^*) + D(p_t^*||p_6) \quad (29)$$

and we define the unique information in the system that is due to source $\mathbf{X}_1$ to be

$$\text{Unq1} = D(p_t^*||p_6). \quad (30)$$

as in [3]. Finding the optimal point, t^* , of minimisation of the KL divergence $D(p||p_t)$, and the orthogonality provided by the generalised Pythagorean theorems, define a clear connection between the geometry of the tangent space to manifold M and the definition of the information-geometric PID developed herein.

By using two of the defining equations of a PID (6) and (7), there are two possible expressions for the shared information, Shd, in the system:

$$I(\mathbf{X}_1; \mathbf{X}_3) - \text{Unq1} \quad \text{or} \quad I(\mathbf{X}_2; \mathbf{X}_3) - \text{Unq2}. \quad (31)$$

Using the result in Lemma 1, we may write the unique information terms as follows. The unique information provided by $\mathbf{X}_2$ is defined to be

$$\text{Unq2} = D(p_{t^*}||p_5) = \frac{1}{2} \log \frac{|\Sigma_5|}{|\Sigma_{t^*}|} - \frac{1}{2}m + \frac{1}{2} \text{Tr}(\Sigma_5^{-1}\Sigma_{t^*}) = \frac{1}{2} \log \frac{|\Sigma_5|}{|\Sigma_{t^*}|}.$$

by Lemma 5.

The unique information provided by $\mathbf{X}_1$ is defined to be

$$\text{Unq1} = D(p_{t^*}||p_6) = \frac{1}{2} \log \frac{|\Sigma_6|}{|\Sigma_{t^*}|} - \frac{1}{2}m + \frac{1}{2} \text{Tr}(\Sigma_6^{-1}\Sigma_{t^*}) = \frac{1}{2} \log \frac{|\Sigma_6|}{|\Sigma_{t^*}|}.$$

by Lemma 5.

3.6. The I_{ig} PID

Explicit expressions for the PID components are given in Proposition 1, with proof in Appendix H.

Proposition 1. *The partial information decomposition I_{ig} for the zero-mean multivariate Gaussian system defined in (12) has the following components.*

$$\begin{aligned} \text{Syn} &= \frac{1}{2} \log \frac{|I_{n_2} - P^T P| |I_{n_3} - Q^T Q| |I_{n_3} - R^T R|}{|\Sigma|} + \frac{1}{2} \log \frac{1}{d(t^*)} \\ \text{Unq1} &= \frac{1}{2} \log \frac{1}{|I_{n_3} - Q^T Q|} - \frac{1}{2} \log \frac{1}{d(t^*)} \\ \text{Unq2} &= \frac{1}{2} \log \frac{1}{|I_{n_3} - R^T R|} - \frac{1}{2} \log \frac{1}{d(t^*)} \\ \text{Shd} &= \frac{1}{2} \log \frac{1}{d(t^*)}. \end{aligned}$$

where the determinant $d(t)$ is defined by

$$d(t) = |I_{n_3} - (1-t)^2 Q^T Q - t(1-t)(R^T P^T Q + Q^T P R) - t^2 R^T R|, \quad (t \in F) \quad (32)$$

and F is the interval of real values of t for which Σ_t is positive definite.

The two possible expressions for the shared information in (31) are equal.

Theoretical properties of the I_{ig} PID are presented in Proposition 2, with proof in Appendix I.2.

Proposition 2. *The PID defined in Proposition 1 possesses the Williams–Beer properties of non-negativity, self-redundancy, symmetry and monotonicity.*

3.7. Some Examples and Illustrations

Example 1. Prediction of calcium contents.

This dataset was considered in [22]. The I_{ig} PID developed here, along with the I_{dep} PID [22] and I_{mmi} PID [20], was applied using data on 73 women involving one set of predictors $\mathbf{X}_1$ (Age, Weight, Height), another set of two predictors $\mathbf{X}_2$ (diameter of os calcis, diameter of radius and ulna), and target $\mathbf{X}_3$ (calcium content of heel and forearm). The following results were obtained.

PID	t^*	Unq1	Unq2	Shd	Syn
I_{ig}	0.2408	0.3581	0.0304	0.0728	0.1904
I_{dep}		0.4077	0.0800	0.0232	0.1408
I_{mmi}		0.3277	0	0.1032	0.2209

A plot of the ‘synergy’ function $g(t)$ is shown in Figure 2a. All three PIDs indicate the presence of synergy and a large component of unique information due to the variables in $\mathbf{X}_1$. The I_{ig} PID suggests the transmission of more of the joint mutual information as shared and synergistic information and correspondingly less unique information due to either source vector than does the I_{dep} PID. This is true also for the results from the I_{mmi} PID, but it has higher values for synergistic and shared information and a lower value for Unq1 than those produced by the I_{ig} PID. It was shown in [22] that pdf p in manifold M provides a better fit to these data than any of the submanifold distributions. This pdf contains pairwise cross-correlation between the vectors $\mathbf{X}_1$ and $\mathbf{X}_3$, and between $\mathbf{X}_2$ and $\mathbf{X}_3$. Hence, it is no surprise to find that a relatively large Unq1 component. One might also anticipate a large value for Unq2. That this is not the case is explained, at least partly, by the presence of unique information asymmetry, in that the mutual information between $\mathbf{X}_1$ and $\mathbf{X}_3$ (0.4309) is much larger than that between $\mathbf{X}_2$ and $\mathbf{X}_3$ (0.1032) and also bearing in mind the constraints imposed by (6)–(10).

The PIDs were also computed with the same $\mathbf{X}_1$ and $\mathbf{X}_3$ but taking $\mathbf{X}_2$ to be another set of four predictors (surface area, strength of forearm, strength of leg, area of os calcis). The following results were obtained.

PID	t^*	Unq1	Unq2	Shd	Syn
I_{ig}	0.0027	0.3522	0.0000	0.0787	0.0186
I_{dep}		0.3708	0.0186	0.0601	0
I_{mmi}		0.3522	0	0.0787	0.0186

A plot of the ‘synergy’ function $g(t)$ is shown in Figure 2b. In this case, the PIDs obtained from all three methods are very similar, with the main component being unique information due to the variables in $\mathbf{X}_1$. The PIDs indicate almost zero synergy and almost zero unique information due to the variables in $\mathbf{X}_2$. In [22], it was shown that the best of the pdfs is p_5 associated with submanifold S_5 . If this model were to hold exactly, then a PID must have Syn and Unq2 components that are equal to zero. Therefore, all three PIDs perform very well here, and the fact that the Unq1 component is much larger than the Shd component is due to unique information asymmetry, since the mutual information between $\mathbf{X}_2$ and $\mathbf{X}_3$ is only 0.0787. In this dataset, the I_{ig} PID suggests the transmission just a little more of the joint mutual information as shared and synergistic information and correspondingly less unique information due to either source vector than does the I_{dep} PID. The I_{ig} and I_{mmi} PIDs produce identical results (to 4 d.p.).

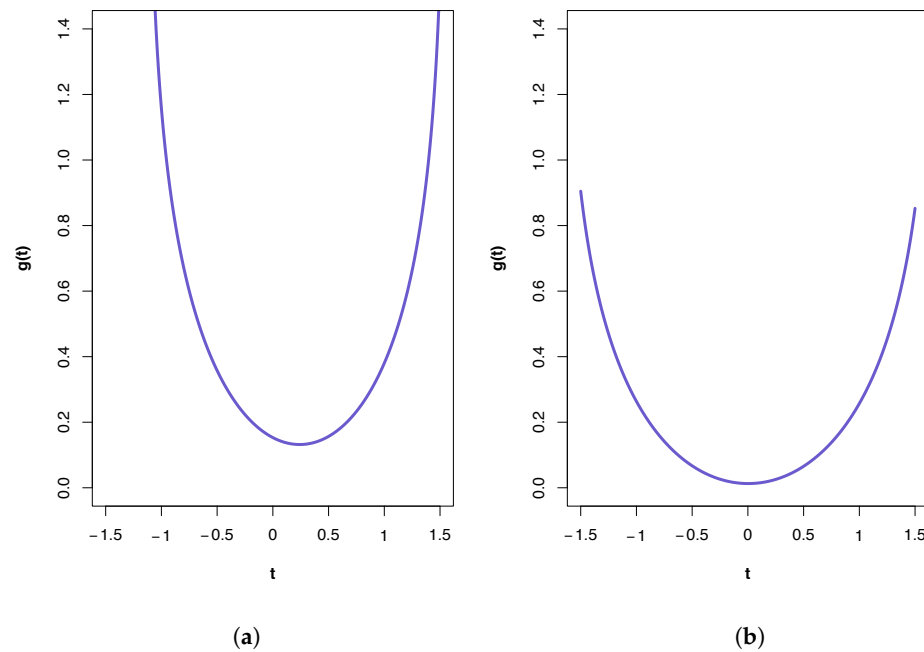


Figure 2. Plots of the ‘synergy’ function $g(t)$ for the two calcium datasets. **(a)** First calcium dataset. The feasible range for t is $(-1.13, 1.56)$, with $t^* = 0.2408$. **(b)** Second calcium dataset. The feasible range for t is $(-1.67, 1.69)$, with $t^* = 0.0027$.

When working with real or simulated data, it is important to use the correct covariance matrix. In order to use the results given in Proposition 1, it is essential that the input covariance matrix has the structure of Σ , as given in (12). Further detail is provided in Appendix J.

Example 2. *PID expectations and exact results.*

Since there is no way to know the true PID for any given dataset it is useful to consider situations under which some values of the PID components can be predicted, and this approach has been used in developments of the topic. Here, we consider such expectations provided by the pdfs associated with the submanifolds $S_3 - S_7$, defined in Figure 1. In submanifold S_3 , the source $\mathbf{X}_2$ is independent of both the other source $\mathbf{X}_1$ and the target $\mathbf{X}_3$. Hence, we expect only unique information due to source $\mathbf{X}_1$ to be transmitted. Submanifold S_4 is similar but we expect only unique information due to source $\mathbf{X}_2$ to be transmitted. In manifold S_5 , $\mathbf{X}_2$ and $\mathbf{X}_3$ are conditionally independent given a value for $\mathbf{X}_1$. Hence, from (9), we expect the Unq2 and Syn components to be zero. Similarly, for S_6 , we expect the Unq1 and Syn components to be equal to zero, by (8). On submanifold S_7 , the sources $\mathbf{X}_1, \mathbf{X}_2$ are conditionally independent given a value for the target $\mathbf{X}_3$ (which does not mean that the sources are marginally independent). Since the target $\mathbf{X}_3$ interacts with both source vectors, one might expect some shared information as well as unique information from both sources, and also perhaps some synergy. Here, from (11), the interaction information must be negative or zero, and so we can expect to see transmission of more shared information than synergy.

We will examine these expectations by using the following multivariate Gaussian distribution (which was used in [22]). The matrices P, Q, R are given an equi-cross-correlation structure in which all the entries are equal within each matrix:

$$P = p\mathbf{1}_{n_1}\mathbf{1}_{n_2}^T, \quad Q = q\mathbf{1}_{n_1}\mathbf{1}_{n_3}^T, \quad R = r\mathbf{1}_{n_2}\mathbf{1}_{n_3}^T, \quad (33)$$

where p, q, r denote here the constant cross correlations within each matrix and $\mathbf{1}_n$ denotes an n -dimensional vector whose entries are each equal to unity.

The values of (p, q, r) are taken to be $(-0.15, 0.15, 0.15)$, with $n_1 = 3, n_2 = 4, n_3 = 3$. Covariance matrices for pdfs $p_3 - p_7$ were computed using the results in Table 1. Thus, we have the exact covariance matrices which can be fed into the I_{ig} , I_{dep} and I_{mmi} algorithms. The PID results are displayed in Table 2.

From Table 2, we see that all three PIDs meet the expectations exactly for pdfs $p_3 - p_6$, with only unique information transmitted when the pdfs p_3, p_4 , are true, respectively, and zero unique for the relevant component and zero synergy when the models p_5, p_6 are true, respectively. When model p_8 is the true model, we find that the I_{ig} and I_{dep} PIDs produce virtually identical results: the joint mutual information is transmitted almost entirely as synergistic information. The I_{mmi} PID is slightly different, with less unique information transmitted about the variables in $\mathbf{X}_2$, and more shared and synergistic information transmitted than with the other two PIDs. The PIDs produce very different results for pdf p_7 , although, as expected, they do express more shared information than synergy. When this model is satisfied, I_{dep} sets the synergy to 0, even if there is no compelling reason to support this. This curiosity is mentioned and illustrated in [22]. On the other hand, the I_{ig} PID suggests that each of the four components contributes to the transmission of the joint mutual information, with unique information due to $\mathbf{X}_2$ and shared information making more of a contribution than the other two components. The I_{mmi} PID transmits a higher percentage of the joint information as shared and synergistic information, and a smaller percentage due to the variables in $\mathbf{X}_2$, than is found with I_{ig} ; these differences are much stronger when comparison is made with the corresponding I_{dep} components. As with model p , it appears that the setting of the Unq1 component in I_{ig} to zero has been translated into its percentage being subtracted from the Unq2 component and added to both the Shd and Syn components in I_{ig} to produce I_{mmi} .

Table 2. PID results for exact pdfs, reported as a percentage of the joint mutual information.

pdf	PID	Unq1	Unq2	Shd	Syn
p	I_{ig}	4.3	6.6	1.5	87.7
	I_{dep}	4.3	6.6	1.4	87.6
	I_{mmi}	0.0	2.3	5.8	91.9
p_7	I_{ig}	15.0	33.2	31.7	20.1
	I_{dep}	35.1	53.3	11.6	0.0
	I_{mmi}	0.0	18.2	46.7	35.1
p_6	I_{ig}	0.0	75.9	24.1	0.0
	I_{dep}	0.0	75.9	24.1	0.0
	I_{mmi}	0.0	75.9	24.1	0.0
p_5	I_{ig}	75.2	0.0	24.8	0.0
	I_{dep}	75.2	0.0	24.8	0.0
	I_{mmi}	75.2	0.0	24.8	0.0
p_4	I_{ig}	0.0	100.0	0.0	0.0
	I_{dep}	0.0	100.0	0.0	0.0
	I_{mmi}	0.0	100.0	0.0	0.0
p_3	I_{ig}	100.0	0.0	0.0	0.0
	I_{dep}	100.0	0.0	0.0	0.0
	I_{mmi}	100.0	0.0	0.0	0.0

Example 3. Some simulations.

Taking the same values of p, q, r and n_1, n_2, n_3 as in the previous example, a small simulation study was conducted. From each of the pdfs, $p_3 - p_7, p$, a simple random sample of size 1000 was generated from the 10-dimensional distribution, a covariance

matrix estimated from the data and the I_{ig} , I_{dep} and I_{mmi} algorithms were applied. This procedure was repeated 1000 times. In order to make the PID results from the sample of 1000 datasets comparable each PID was normalized by dividing each of its components by the joint mutual information; see (10). A summary of the results is provided in Table 3. We focus here on the comparison of I_{ig} and I_{dep} , and also I_{ig} and I_{mmi} , since I_{dep} has been compared with I_{mmi} for Gaussian systems [22].

Table 3. PID results for simulated datasets from the pdfs, $p_3 - p_7$, p , reported as median (in bold) and range of the sample of percentages of the joint mutual information, apart from t^* which gives the median and range of the actual values obtained using the I_{ig} algorithm.

pdf	PID	t^*	Unq1	Unq2	Shd	Syn
p	I_{ig}	0.55 (0.44, 0.63)	4.4 (2.6, 6.6)	6.7 (3.7, 9.0)	1.6 (1.1, 2.1)	87.3 (85.4, 89.2)
	I_{dep}		4.5 (2.8, 6.8)	6.8 (3.9, 9.3)	1.4 (1.0, 1.9)	87.2 (85.3, 89.0)
	I_{mmi}		0.0 (0.0, 2.9)	2.3 (3.2, 6.2)	6.0 (3.9, 8.0)	91.7 (89.2, 93.6)
p_7	I_{ig}	0.59 (0.47, 0.71)	15.1 (7.6, 26.5)	33.0 (19.9, 48.0)	31.3 (23.4, 37.3)	20.2 (15.2, 24.9)
	I_{dep}		34.3 (23.6, 46.4)	52.3 (38.9, 64.5)	11.8 (7.8, 20.3)	0.2 (0.0, 8.6)
	I_{mmi}		0.0 (0.0, 11.4)	18.0 (0.0, 45.9)	46.4 (30.4, 57.7)	35.6 (21.3, 45.9)
p_6	I_{ig}	0.97 (0.82, 1.06)	0.1 (0.0, 2.4)	72.5 (55.6, 84.7)	25.0 (11.4, 38.4)	2.4 (0.2, 8.1)
	I_{dep}		2.5 (0.2, 10.4)	74.9 (60.4, 88.6)	22.6 (7.6, 35.3)	0.0 (0.0, 0.0)
	I_{mmi}		0.0 (0.0, 0.0)	72.3 (45.7, 83.6)	25.1 (13.5, 45.4)	2.6 (0.3, 8.9)
p_5	I_{ig}	0.06 (-0.06, 0.20)	68.0 (50.0, 81.7)	0.3 (0.0, 2.9)	27.4 (12.5, 43.3)	4.3 (0.9, 12.1)
	I_{dep}		72.3 (54.9, 87.3)	4.7 (0.9, 13.3)	22.8 (6.8, 38.4)	0.0 (0.0, 0.0)
	I_{mmi}		67.5 (45.7, 82.6)	0.0 (0.0, 0.0)	27.9 (13.9, 45.5)	4.7 (0.9, 14.2)
p_4	I_{ig}	0.97 (0.91, 1.0)	0.06 (0.0, 0.7)	95.0 (83.0, 99.5)	2.5 (1.9, 7.8)	2.4 (0.3, 8.3)
	I_{dep}		2.2 (0.2, 8.4)	97.2 (90.7, 99.7)	0.3 (0.0, 2.6)	0.2 (0.0, 2.7)
	I_{mmi}		0.0 (0.0, 0.0)	94.9 (83.1, 99.3)	2.5 (0.1, 9.4)	2.6 (0.4, 8.6)
p_3	I_{ig}	0.05 (0.01, 0.14)	90.9 (75.7, 98.4)	0.2 (0.0, 1.8)	4.5 (0.7, 11.5)	4.5 (0.9, 11.5)
	I_{dep}		94.8 (86.2, 99.1)	4.2 (0.7, 12.6)	0.3 (0.0, 3.1)	0.3 (0.0, 3.5)
	I_{mmi}		90.6 (76.3, 97.6)	0.0 (0.0, 0.0)	4.6 (1.0, 11.9)	4.8 (1.2, 13.0)

I_{ig} vs. I_{dep}

For pdf p , the I_{ig} and I_{dep} PIDs produce very similar results in terms of both median and range, and the median results are very close indeed to the corresponding exact values in Table 2. For pdf p_7 , the differences between the PID components found in Table 2 persist here although each PID, respectively, produces median values of their components that are close to the exact results in Table 2. For the other four pdfs, there are some small but interesting differences among the results produced by the two PID methods.

The I_{ig} method has higher median values for synergy and shared information than for the unique information, when compared against the corresponding exact values in Table 2. In particular, the values of unique information given by I_{ig} are much lower than expected for pdfs p_3, p_4, p_6 , and the levels of synergy are larger than expected particularly for pdfs p_3 and p_5 . On the other hand, the I_{dep} PID tends to have larger values for the unique information, and lower values for synergy, especially for datasets generated from pdfs p_3, p_4 and p_5 . For models $p_3 - p_6$, I_{dep} has median values of synergy that are closer to the corresponding exact values than those produced by I_{ig} . The suggestion that the I_{ig} method can produce more synergy and shared information than the I_{dep} method, given the same dataset, is supported by the fact that for all the pdfs and all 6000 datasets considered, the I_{ig} method produced greater levels of synergy and shared information and smaller values of the unique information in *every* dataset. This raises a question of whether such a finding is generally the case and whether there is this type of a systematic difference between the methods. In the case of scalar variables, it is easy to derive general analytic formulae for the I_{ig} PID components and such a systematic difference is present in this case.

I_{ig} vs. I_{mmi}

The I_{ig} and I_{mmi} PIDs produce similar results for the datasets generated from pdf p , although the I_{mmi} PID suggests the transmission of more shared and synergistic information and less unique information than does I_{ig} . For pdf p_7 , the differences between the PID results are much more dramatic, with the I_{mmi} PID allocating an additional 15% of the joint mutual information to be shared and the synergistic information, and correspondingly 15% less of the unique information. Both methods produce almost identical summary statistics on the datasets generated from pdfs $p_3 - p_6$. Since the same patterns are present for all four distributions, we discuss the results for pdf p_5 as an exemplar and compare them with the corresponding exact values in Table 2. The results for component Unq1 show that both methods produce an underestimate of approximately 7%, on average, of the joint mutual information. The median values of Unq2 are close to those expected. The underestimates on the Unq1 component are coupled with overestimates, on average, for the shared and synergistic components; they are 2.6% and 4.3%, respectively, with the I_{ig} method, and 3.1% and 4.7%, respectively, with I_{mmi} .

As to be expected with percentage data, the variation in results for each component tends to be larger for values that are not extreme and much smaller for the extreme values. Also, the optimal values of t^* are shown in Table 3. They were all found to be in the range $[0, 1]$, except for 202 of the datasets generated from pdf p_5 or p_6 .

4. Discussion

For the case of multivariate Gaussian systems with two vector inputs and a vector output, results have been derived using standard theorems from information geometry in order to develop simple, almost exact formulae for the I_{ig} PID, thus extending the scope of the work of [3] on scalar inputs and output. The formulae require one parameter to be determined by a simple, constrained convex optimisation. In addition, it has been proved that this I_{ig} PID algorithm satisfies the desirable theoretical properties of non-negativity, self-redundancy, symmetry and monotonicity, first postulated by Williams and Beer [1]. These results strengthen the confidence that one might have in using the I_{ig} method to separate the joint mutual information in a multivariate Gaussian system into shared, unique and synergistic components. The examples demonstrate that the I_{ig} method is simple to use and a small simulation study reveals that it is fairly robust, although in some of the scenarios considered the I_{ig} method produced more synergy and shared information than expected, and correspondingly less unique information; in some other scenarios, it performed as expected. Comparison of the I_{ig} and I_{dep} algorithms reveal that they can produce exactly the same, or very similar, results in some scenarios, but in other situations, it is clear that the I_{ig} method tends to have larger levels of shared information and synergy, and correspondingly, lower levels of unique information when compared with the results from the I_{dep} method.

For datasets generated from pdfs $p_3 - p_6$, the PIDs produced using the I_{ig} and I_{mmi} methods are, on average, very similar indeed, and both methods overestimate synergy and shared information and underestimate unique information. The extent of these biases, as a percentage of the joint mutual information, is fairly small, on average, when pdf p_4 or p_6 is the true pdf, but larger, on average, when p_3 or p_5 is the true pdf. When pdf p_7 or p is the true pdf, the I_{mmi} algorithm produces even more shared and synergistic information than obtained with the I_{ig} method. This effect is particularly dramatic in the case of p_7 , where on average with I_{mmi} 82% of the joint mutual information is transmitted as shared or synergistic information, as compared with 51.5% for I_{ig} . It appears that the fact that the I_{mmi} method forces one of the unique informations to be zero leads to an underestimation of the other unique information and an overestimate of both the shared and synergistic information, especially when p_7 or p is the true pdf and both unique information are expected to be non-zero.

While some numerical support is presented here for the hypothesis that there might be a systematic difference in this type between the I_{ig} and I_{dep} methods further research would be required to investigate this possibility. Also, the I_{ig} developed here is a bivariate PID and it would be of interest to explore whether the method could be extended to deal with more than two source vectors.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The data used in Example 1 is available from <https://github.com/JWKay/PID> (accessed on 17 June 2024).

Conflicts of Interest: The author declares no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

d.p.	decimal places
I_{dep}	Idep PID for Gaussian systems
I_{ig}	Information-geometric PID
I_{mmi}	Immi for Gaussian systems
KL	Kullback–Leibler
pdf	probability density function
PID	Partial information decomposition

Appendix A. Two Matrix Lemmas from [22]

Lemma A1. Suppose that a symmetric matrix M is partitioned as

$$M = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix},$$

where A and C are symmetric and square. Then,

- (i) The matrix M is positive definite if and only if A and $C - B^T A^{-1} B$ are positive definite.
- (ii) The matrix M is positive definite if and only if C and $A - B C^{-1} B^T$ are positive definite.
- (iii) $|M| = |A| |C - B^T A^{-1} B|$.

Lemma A2. When the covariance matrix Σ in (12) is positive definite then the following matrices are also positive definite, and hence nonsingular:

$$I_{n_2} - P^T P, \quad I_{n_1} - P P^T, \quad I_{n_3} - R^T R, \quad I_{n_2} - R R^T, \quad I_{n_3} - Q^T Q, \quad I_{n_1} - Q Q^T.$$

Also, the determinant of each of these matrices is positive and bounded above by unity, and it is equal to unity if, and only if, the matrix involved is the zero matrix. Furthermore,

$$\begin{vmatrix} I_{n_1} & P \\ P^T & I_{n_2} \end{vmatrix} = |I_{n_2} - P^T P|.$$

Appendix B. Some Formulae from [22]

The following equations provide the relevant information-theoretic terms with a slight change of notation: Σ_Z is now Σ , the indices on the vectors run from 1 to 3 rather than from 0 to 2, and $\mathbf{Y}$ is replaced by $\mathbf{X}_3$.

$$I(\mathbf{X}_1; \mathbf{X}_3) = \frac{1}{2} \log \frac{1}{|I_{n_3} - Q^T Q|}, \quad (\text{A1})$$

$$I(\mathbf{X}_2; \mathbf{X}_3) = \frac{1}{2} \log \frac{1}{|I_{n_3} - R^T R|}, \quad (\text{A2})$$

$$I(\mathbf{X}_1; \mathbf{X}_3 | \mathbf{X}_2) = \frac{1}{2} \log \frac{|I_{n_2} - P^T P| |I_{n_3} - R^T R|}{|\Sigma|}, \quad (\text{A3})$$

$$I(\mathbf{X}_2; \mathbf{X}_3 | \mathbf{X}_1) = \frac{1}{2} \log \frac{|I_{n_2} - P^T P| |I_{n_3} - Q^T Q|}{|\Sigma|} \quad (\text{A4})$$

$$I(\mathbf{X}_1, \mathbf{X}_2; \mathbf{X}_3) = \frac{1}{2} \log \frac{|I_{n_2} - P^T P|}{|\Sigma|} \quad (\text{A5})$$

$$II(\mathbf{X}_1; \mathbf{X}_2; \mathbf{X}_3) = I(\mathbf{X}_1, \mathbf{X}_2; \mathbf{X}_3) - I(\mathbf{X}_1; \mathbf{X}_3) - I(\mathbf{X}_2; \mathbf{X}_3) \quad (\text{A6})$$

$$= \frac{1}{2} \log \frac{|I_{n_2} - P^T P| |I_{n_3} - Q^T Q| |I_{n_3} - R^T R|}{|\Sigma|} \quad (\text{A7})$$

Appendix C. Proof of Lemma 1

From [43] (p. 189), the KL divergence is

$$D(f_1 || f_2) = \frac{1}{2} \left(\log \frac{|\Lambda|}{|\Phi|} - m + \text{Tr}(\Phi^{-1} \Lambda) \right)$$

since the mean vectors are equal. Each of the diagonal blocks in (13) is positive definite, by Lemma A1, and so possesses a positive definite square root [44] (p. 472). We define the positive definite block diagonal matrix

$$D = \text{diag} \left\{ \Sigma_{11}^{-\frac{1}{2}}, \Sigma_{22}^{-\frac{1}{2}}, \Sigma_{33}^{-\frac{1}{2}} \right\}$$

Hence, we may write

$$\Phi = D \begin{bmatrix} I_{n_1} & P_{12} & P_{13} \\ P_{12}^T & I_{n_2} & P_{23} \\ P_{13}^T & P_{23}^T & I_{n_3} \end{bmatrix} D \quad \text{and} \quad \Lambda = D \begin{bmatrix} I_{n_1} & L_{12} & L_{13} \\ L_{12}^T & I_{n_2} & L_{23} \\ L_{13}^T & L_{23}^T & I_{n_3} \end{bmatrix} D \quad (\text{A8})$$

where

$$P_{12} = \Sigma_{11}^{-\frac{1}{2}} \Sigma_{12} \Sigma_{22}^{-\frac{1}{2}}, \quad P_{13} = \Sigma_{11}^{-\frac{1}{2}} \Sigma_{13} \Sigma_{33}^{-\frac{1}{2}}, \quad P_{23} = \Sigma_{22}^{-\frac{1}{2}} \Sigma_{23} \Sigma_{33}^{-\frac{1}{2}}.$$

and

$$L_{12} = \Sigma_{11}^{-\frac{1}{2}} \Lambda_{12} \Sigma_{22}^{-\frac{1}{2}}, \quad L_{13} = \Sigma_{11}^{-\frac{1}{2}} \Lambda_{13} \Sigma_{33}^{-\frac{1}{2}}, \quad L_{23} = \Sigma_{22}^{-\frac{1}{2}} \Lambda_{23} \Sigma_{33}^{-\frac{1}{2}}.$$

Therefore, using standard properties of determinants, we have from

$$|\Phi| = |\Sigma_{11}||\Sigma_{22}||\Sigma_{33}|\begin{vmatrix} I_{n_1} & P_{12} & P_{13} \\ P_{12}^T & I_{n_2} & P_{23} \\ P_{13}^T & P_{23}^T & I_{n_3} \end{vmatrix} = |\Sigma_{11}||\Sigma_{22}||\Sigma_{33}||\Phi_1|.$$

Use of a similar argument provides a similar expression for the determinant of Λ :

$$|\Lambda| = |\Sigma_{11}||\Sigma_{22}||\Sigma_{33}||\Lambda_1|.$$

It follows that

$$\frac{|\Lambda|}{|\Phi|} = \frac{|\Lambda_1|}{|\Phi_1|}.$$

We now consider the trace term. We have from (A8) that

$$\Phi = D\Phi_1D, \quad \text{and} \quad \Lambda = D\Lambda_1D$$

and so

$$\Phi^{-1}\Lambda = D^{-1}\Phi_1^{-1}D^{-1}D\Lambda_1D = D^{-1}\Phi_1^{-1}\Lambda D.$$

Since $\text{Tr}(ABC) = \text{Tr}(CBA)$, for any three conformable matrices, the required result follows.

Appendix D. Proof of Lemma 2

Making use of Lemma A1(iii), we may write the determinant of the covariance matrix Σ as

$$|\Sigma| = \begin{vmatrix} I_{n_1} & P & Q \\ P^T & I_{n_2} & R \\ Q^T & R^T & I_{n_3} \end{vmatrix} = |A||I_{n_3} - [Q^T \quad R^T]A^{-1}\begin{bmatrix} Q \\ R \end{bmatrix}| \quad (\text{A9})$$

where

$$A = \begin{bmatrix} I_{n_1} & P \\ P^T & I_{n_2} \end{bmatrix} \quad (\text{A10})$$

Since

$$A^{-1} = \begin{bmatrix} I_{n_1} + P(I_{n_2} - P^TP)^{-1}P^T & -P(I_{n_2} - P^TP)^{-1} \\ -(I_{n_2} - P^TP)^{-1}P^T & (I_{n_2} - P^TP)^{-1} \end{bmatrix} \quad (\text{A11})$$

it can be shown that

$$[Q \quad R]^T A^{-1} \begin{bmatrix} Q \\ R \end{bmatrix} = Q^T Q + (P^T Q - R)^T (I_{n_2} - P^T P)^{-1} (P^T Q - R)$$

From Lemma A1(iii), we have that

$$|A| = |I_{n_2} - P^T P| \quad (\text{A12})$$

and putting these results together gives

$$|\Sigma| = |I_{n_2} - P^T P| |I_{n_3} - Q^T Q - (P^T Q - R)^T (I_{n_2} - P^T P)^{-1} (P^T Q - R)|. \quad (\text{A13})$$

We now apply this result to obtain expressions for the determinants of Σ_5 and Σ_6 , which are defined in Table 1. In Σ_5 , R is replaced by $P^T Q$ and making this substitution in (A13) gives

$$|\Sigma_5| = |I_{n_2} - P^T P| |I_{n_2} - Q^T Q|. \quad (\text{A14})$$

Similarly, replacing Q by PR in (A13), gives the determinant of Σ_6 , after some manipulation, as

$$|\Sigma_6| = |I_{n_2} - P^T P| |I_{n_2} - R^T R|. \quad (\text{A15})$$

For the final result, we have that

$$\Sigma_5 - \Sigma_6 = \begin{bmatrix} 0 & 0 & Q - PR \\ 0 & 0 & P^T Q - R \\ Q^T - R^T P^T & Q^T P - R^T & 0 \end{bmatrix}$$

from which it is clear that $Q = 0, R = 0$ implies that $\Sigma_5 = \Sigma_6$. Conversely,

$$\begin{aligned} \Sigma_5 = \Sigma_6 &\implies Q = PR \text{ and } R = P^T Q \\ &\implies (I_{n_2} - P^T P)R = 0 \text{ and } (I_{n_1} - PP^T)Q = 0 \\ &\implies Q = 0 \text{ and } R = 0 \end{aligned}$$

by Lemma A2. For the last equivalence, we note by (A13) that $Q = 0$ and $R = 0$ implies that $I(\mathbf{X}_1, \mathbf{X}_2; \mathbf{X}_3) = 0$. Conversely, by (A4) and (A13) and the non-negativity of mutual information, $I(\mathbf{X}_1, \mathbf{X}_2; \mathbf{X}_3) = 0$ implies that $I(\mathbf{X}_1; \mathbf{X}_3) = 0$ and $I(\mathbf{X}_2; \mathbf{X}_3) = 0$, which imply from (A1) and (A2) and Lemma A2 that $Q = 0$ and $R = 0$.

Appendix E. Proof of Lemma 3

Σ_t and Σ'_t are each positive definite if and only their inverses are positive definite. We work with the inverses. When $t = 0$ or $t = 1$ clearly both inverses are positive definite, since the inverses of Σ_5 and Σ_6 are positive definite. Suppose that $0 < t < 1$ and consider for Σ_t^{-1} , and for any $\mathbf{x} \in \mathbb{R}^m$, the quadratic form

$$\mathbf{x}^T ((1-t)\Sigma_5^{-1} + t\Sigma_6^{-1})\mathbf{x} = (1-t)\mathbf{x}^T \Sigma_5^{-1}\mathbf{x} + t\mathbf{x}^T \Sigma_6^{-1}\mathbf{x}.$$

This term is non-negative for every $\mathbf{x} \in \mathbb{R}^m$ since the inverses of Σ_5 and Σ_6 are positive definite and $0 < t < 1$. Clearly, for the same reasons, this quadratic form is equal to zero if and only if $\mathbf{x}$ is the zero vector in $\mathbb{R}^m$, hence the result. A similar argument shows that $\Sigma_t'^{-1}$ is positive definite when $0 < t < 1$.

For $t \in F$, we know that the matrix $\Sigma_t^{-1} = (1-t)\Sigma_5^{-1} + t\Sigma_6^{-1}$ is positive definite. The matrix $\Sigma_t'^{-1} = (1-t)\Sigma_6 + t\Sigma_5$ may be written in two ways as

$$(1-t)\Sigma_6 + t\Sigma_5 = \Sigma_6((1-t)\Sigma_5^{-1} + t\Sigma_6^{-1})\Sigma_5 = \Sigma_5((1-t)\Sigma_5^{-1} + t\Sigma_6^{-1})\Sigma_6.$$

which is a product of three positive definite matrices. Since the last two expressions are transposes of each other it follows that this product is symmetric. Hence, from a result by E. P. Wigner [45], we deduce that $\Sigma_t'^{-1}$ is positive definite, so that $t \in G$. A similar argument shows that $t \in G \implies t \in F$. Thus, we have proved that the feasible sets F and G are equal.

Appendix F. Proof of Lemma 4

For the first part, we consider (3):

$$\Sigma_t^{-1}\Sigma = (1-t)\Sigma_5^{-1}\Sigma + t\Sigma_6^{-1}\Sigma \quad (\text{A16})$$

and apply the trace operator, which is linear, to obtain

$$\text{Tr}(\Sigma_t^{-1}\Sigma) = (1-t)\text{Tr}(\Sigma_5^{-1}\Sigma) + t\text{Tr}(\Sigma_6^{-1}\Sigma). \quad (\text{A17})$$

Consider $\text{Tr}(\Sigma_5^{-1}\Sigma)$. We may write this as

$$\text{Tr}(\Sigma_5^{-1}\Sigma) = \text{Tr}(\Sigma_5^{-1}(\Sigma_5 + (\Sigma - \Sigma_5)) = \text{Tr}(I_m) + \text{Tr}(\Sigma_5^{-1}(\Sigma - \Sigma_5)) \quad (\text{A18})$$

Now, from (12) and the form of Σ_5 in Table 1, we have that

$$\Sigma - \Sigma_5 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & R - P^T Q \\ 0 & R^T - Q^T P & 0 \end{bmatrix}.$$

The pdf p_5 is defined by the constraint $K_{23} = 0$ in the inverse covariance matrix K . Hence, Σ_5^{-1} has the form

$$\Sigma_5^{-1} = \begin{bmatrix} * & * & * \\ * & * & 0 \\ * & 0 & * \end{bmatrix}$$

Performing the multiplication of these block matrices shows that the diagonal blocks of $\Sigma_5^{-1}(\Sigma - \Sigma_5)$ are all equal to a zero matrix, and so have trace equal to zero. Hence, the trace of this matrix is equal to zero. Since $\text{Tr}(I_m) = m$, it follows from (A18) that this matrix has trace equal to m .

Now consider the trace of $\Sigma_6^{-1}\Sigma$. From (12) and the form of Σ_6 in Table 1, we have that

$$\Sigma - \Sigma_6 = \begin{bmatrix} 0 & 0 & Q - PR \\ 0 & 0 & 0 \\ Q^T - R^T P^T & 0 & 0 \end{bmatrix}.$$

The pdf p_6 is defined by the constraint $K_{13} = 0$ in the inverse covariance matrix K . Hence, Σ_6^{-1} has the form

$$\Sigma_6^{-1} = \begin{bmatrix} * & * & 0 \\ * & * & * \\ 0 & * & * \end{bmatrix}$$

By adopting a similar argument to that given above for Σ_5 , it follows that $\text{Tr}(\Sigma_6^{-1}\Sigma) = m$. The required result follows from (A17). Now for the second part of the proof, we use Lemma 1 and the trace result just derived to write

$$D(p||p_t) = \frac{1}{2} \log \frac{|\Sigma_t|}{|\Sigma|} - \frac{1}{2}m + \frac{1}{2} \text{Tr}(\Sigma_t^{-1}\Sigma) = \frac{1}{2} \log \frac{|\Sigma_t|}{|\Sigma|}. \quad (\text{A19})$$

Σ_t^{-1} may be written as

$$\Sigma_t^{-1} = \Sigma_5^{-1}((1-t)\Sigma_6 + t\Sigma_5)\Sigma_6^{-1} \quad (\text{A20})$$

and so, by Lemma 2, we have

$$|\Sigma_t| = \frac{|\Sigma_5||\Sigma_6|}{|(1-t)\Sigma_6 + t\Sigma_5|} = \frac{|I_{n_2} - P^T P|^2 |I_{n_3} - Q^T Q| |I_{n_3} - R^T R|}{|(1-t)\Sigma_6 + t\Sigma_5|} \quad (\text{A21})$$

Hence

$$g(t) = \frac{1}{2} \log \frac{|I_{n_2} - P^T P|^2 |I_{n_3} - Q^T Q| |I_{n_3} - R^T R|}{|\Sigma|} - \frac{1}{2} \log |(1-t)\Sigma_6 + t\Sigma_5| \quad (\text{A22})$$

We wish to minimize $g(t)$ with respect to t under the constraint that $t \in F^\circ$. We set $H(t) = (1-t)\Sigma_6 + t\Sigma_5$, which is positive definite by Lemma 3, and apply Jacobi's formula and the chain rule. Differentiating (A22) with respect to t we obtain

$$\begin{aligned} g'(t) &= -\frac{1}{2} \frac{1}{|H(t)|} \times |H(t)| \times \text{Tr}(H(t)^{-1}(\Sigma_5 - \Sigma_6)) \\ &= -\frac{1}{2} \text{Tr}(H(t)^{-1}(\Sigma_5 - \Sigma_6)) \end{aligned}$$

Further differentiation yields

$$\begin{aligned} g''(t) &= -\frac{1}{2} \text{Tr}(H(t)^{-1}(\Sigma_5 - \Sigma_6)H(t)^{-1}(\Sigma_5 - \Sigma_6)) \\ &= -\frac{1}{2} \text{Tr}([H(t)^{-1}(\Sigma_5 - \Sigma_6)]^2) = -\frac{1}{2} \sum_{i=1}^m \lambda_i^2, \end{aligned}$$

where the λ_i are the eigenvalues of the matrix $H(t)^{-1}(\Sigma_5 - \Sigma_6)$. Since $H(t)$ is positive definite it possesses a positive definite square root $K(t)$, say. Since the eigenvalues of $H(t)^{-1}(\Sigma_5 - \Sigma_6)$ are the same as those of $K(t)(\Sigma_5 - \Sigma_6)K(t)$, which is symmetric and so has real eigenvalues, and since the joint mutual information is positive (and so by Lemma 2 $\Sigma_5 \neq \Sigma_6$), it follows that at least one of the eigenvalues must be non-zero. Hence, $g''(t) < 0, \forall t \in F^\circ$. Therefore, $g(t)$ is strictly convex on the convex set F° and the minimum is unique, and given by the value t^* which satisfies the necessary condition $g'(t) = 0$. (This t^* turns out to be exactly the value of t that is required for the two forms of shared information to be equal.) Of course, the minimum could occur at an endpoint of the interval F , in which case the minimum value of $g(t)$ is the value taken at the endpoint.

Finally, we write

$$|(1-t)\Sigma_6 + t\Sigma_5| = \begin{vmatrix} A & B(t) \\ B(t)^T & I_{n_3} \end{vmatrix} = |A| |I_{n_3} - B(t)^T A^{-1} B(t)|, \quad (\text{A23})$$

where A is defined in (A10), with inverse given by (A11), and

$$B(t) = \begin{bmatrix} (1-t)Q + tPR \\ (1-t)P^T Q + tR \end{bmatrix}$$

Some matrix calculation then gives

$$A^{-1}B(t) = \begin{bmatrix} (1-t)Q \\ tR \end{bmatrix}$$

and also

$$B(t)^T A^{-1} B(t) = (1-t)^2 Q^T Q + t(1-t)(R^T P^T Q + Q^T P R) + t^2 R^T R. \quad (\text{A24})$$

Now from (A12) and (A23) it follows that $|(1-t)\Sigma_6 + t\Sigma_5|$ is equal to

$$|I_{n_2} - P^T P| |I_{n_3} - (1-t)^2 Q^T Q - t(1-t)(R^T P^T Q + Q^T P R) - t^2 R^T R| \quad (\text{A25})$$

Making use of this equation and also (A22), and replacing t by the optimal t^* , we have that the minimum of $g(t)$, when $t^* \in F$, is

$$g(t^*) = \frac{1}{2} \log \frac{|I_{n_2} - P^T P| |I_{n_3} - Q^T Q| |I_{n_3} - R^T R|}{|\Sigma|} + \frac{1}{2} \log \frac{1}{d(t^*)}, \quad (\text{A26})$$

where

$$d(t) = |I_{n_3} - (1-t)^2 Q^T Q - t(1-t)(R^T P^T Q + Q^T P R) - t^2 R^T R|. \quad (\text{A27})$$

Appendix G. Proof of Lemma 5

From (4) we know that

$$D(p||p_5) = D(p||p_{t^*}) + D(p_{t^*}||p_5). \quad (\text{A28})$$

Then, by Lemmas 1 and 4,

$$D(p||p_5) = \frac{1}{2} \log \frac{|\Sigma_5|}{|\Sigma|} - \frac{1}{2}m + \frac{1}{2} \text{Tr}(\Sigma_5^{-1}\Sigma) = \frac{1}{2} \log \frac{|\Sigma_5|}{|\Sigma|}.$$

Also,

$$D(p_{t^*}||p_5) = \frac{1}{2} \log \frac{|\Sigma_5|}{|\Sigma_{t^*}|} - \frac{1}{2}m + \frac{1}{2} \text{Tr}(\Sigma_5^{-1}\Sigma_{t^*})$$

and

$$D(p||p_{t^*}) = \frac{1}{2} \log \frac{|\Sigma_{t^*}|}{|\Sigma|}$$

by (20) with $t = t^*$. By substituting these expressions into (A28) it follows that $\text{Tr}(\Sigma_5^{-1}\Sigma_{t^*}) = m$. The proof that $\text{Tr}(\Sigma_6^{-1}\Sigma_{t^*}) = m$ is obtained by using a very similar argument, starting with (5).

Appendix H. Proof of Proposition 1

The formula for synergy is the minimum value of $g(t)$ as provided in (A26). From (30), Lemmas 1 and 5, we have that

$$\begin{aligned} \text{Unq1} &= \frac{1}{2} \log \frac{|\Sigma_6|}{|\Sigma_{t^*}|} \\ &= \frac{1}{2} \log \frac{|\Sigma_5||\Sigma_6|}{|(1-t)\Sigma_6 + t\Sigma_5|}, \quad \text{by (A21),} \\ &= \frac{1}{2} \log \frac{|I_{n_2} - P^T P| d(t^*)}{|I_{n_2} - P^T P| |I_{n_3} - Q^T Q|}, \quad \text{by (A25) and (A27) and Lemma 2,} \\ &= \frac{1}{2} \log \frac{1}{|I_{n_3} - Q^T Q|} - \frac{1}{2} \log \frac{1}{d(t^*)} \end{aligned} \quad (\text{A29})$$

By making use of a similar calculation we find that

$$\text{Unq2} = \frac{1}{2} \log \frac{|\Sigma_5|}{|\Sigma_{t^*}|} = \frac{1}{2} \log \frac{1}{|I_{n_3} - R^T R|} - \frac{1}{2} \log \frac{1}{d(t^*)} \quad (\text{A30})$$

The two expressions for the shared information are

$$I(\mathbf{X}_1; \mathbf{X}_3) - \text{Unq1} \quad \text{or} \quad I(\mathbf{X}_2; \mathbf{X}_3) - \text{Unq2}.$$

By using (6), (7), (A29) and (A30) it is easy to see that both expressions for the shared information, Shd, are equal to $\frac{1}{2} \log \frac{1}{d(t^*)}$.

Appendix I. Proof of Proposition 2

Appendix I.1. Non-Negativity

Since the Syn, Unq1 and Unq2 components in the I_{ig} PID have been defined to be KL divergences they are non-negative. It remains to show that Shd is non-negative. From Appendix H, we see that this is the case if and only if $0 < d(t^*) \leq 1$. From (A23), (A24) and (A27) we have, with $t = t^*$, that

$$|(1-t^*)\Sigma_6 + t^*\Sigma_5| = \left| \begin{array}{cc} A & B(t^*) \\ B(t^*)^T & I_{n_3} \end{array} \right| = |A| |I_{n_3} - B(t^*)^T A^{-1} B(t^*)|,$$

and that

$$d(t^*) = |I_{n_3} - B(t^*)^T A^{-1} B(t^*)|$$

Since $t^* \in F$, we know by Lemma 3 that $t^* \in G$, and so this matrix is positive definite. It follows from Lemma A1(iii) that $I_{n_3} - B(t^*)^T A^{-1} B(t^*)$ is positive definite and from

Lemma A1(i) that A is positive definite, as is A^{-1} . Therefore, A^{-1} possesses a (symmetric) positive definite square root, C say, and we may write

$$I_{n_3} - B(t^*)^T A^{-1} B(t^*) = I_{n_3} - X^T X, \quad \text{with } X = CB(t^*).$$

Then the matrix $X^T X$ is positive semi-definite and so has non-negative eigenvalues, $\lambda_1, \lambda_2, \dots, \lambda_{n_3}$. We denote the eigenvalues of $I_{n_3} - X^T X$ as $\{1 - \lambda_i : i = 1, 2, \dots, n_3\}$. Since $I_{n_3} - X^T X$ is positive definite we know that $1 - \lambda_i > 0$ for $i = 1, 2, \dots, n_3$. It follows that $0 < 1 - \lambda_i \leq 1$ for $i = 1, 2, \dots, n_3$. Since the determinant of a square matrix is the product of its eigenvalues we have that

$$d(t^*) \equiv |I_{n_3} - X^T X| = \prod_{i=1}^{n_3} (1 - \lambda_i),$$

and so $0 < d(t^*) \leq 1$, hence the result.

Appendix I.2. Self-Redundancy

The property of self-redundancy considers the case of $\mathbf{X}_1 = \mathbf{X}_2$, i.e., both sources are the same, and it requires us to show that

$$\text{Shd} \equiv I_{\text{shd}}(\mathbf{X}_1, \mathbf{X}_1; \mathbf{X}_3) = I[\mathbf{X}_1; \mathbf{X}_3].$$

When the sources are the same we have $n_1 = n_2$, $Q = R$ and $P = I_{n_1}$, which results in a singular covariance matrix. Therefore, we take $P = (1 - \epsilon)I_{n_1}$, for very small ϵ such that $0 < \epsilon < 1$, and let $\epsilon \rightarrow 0+$. Using this information in (A27) we have that

$$\begin{aligned} d(t^*) &= |I_{n_3} - (1 - t^*)^2 Q^T Q - (1 - \epsilon)t^*(1 - t^*)Q^T Q - t^{*2}Q^T Q| \\ &= |I_{n_3} - Q^T Q + \epsilon t^*(1 - t^*)Q^T Q| \\ &\rightarrow |I_{n_3} - Q^T Q|, \quad \text{as } \epsilon \rightarrow 0+. \end{aligned}$$

Therefore $\text{Shd} \rightarrow \frac{1}{2} \log \frac{1}{|I_{n_3} - Q^T Q|}$, which is equal to $I[\mathbf{X}_1; \mathbf{X}_3]$ by (A1).

Out of interest, it seems worthwhile to check the limits for the classical information measures given in (A3)–(A7). From (A3) and (A13), after some cancellation, we have that

$$\begin{aligned} I[\mathbf{X}_1; \mathbf{X}_3 | \mathbf{X}_2] &= \frac{1}{2} \log \frac{|I_{n_3} - Q^T Q|}{|I_{n_3} - Q^T Q - (P^T Q - R)^T (I_{n_2} - P^T P)^{-1} (P^T Q - R)|} \\ &= \frac{1}{2} \log \frac{|I_{n_3} - Q^T Q|}{|I_{n_3} - Q^T Q + \frac{\epsilon}{2-\epsilon} Q^T Q|} \\ &\rightarrow 0, \quad \text{as } \epsilon \rightarrow 0+. \end{aligned}$$

Also, since $Q = R$,

$$\begin{aligned} I[\mathbf{X}_2; \mathbf{X}_3 | \mathbf{X}_1] &= \frac{1}{2} \log \frac{|I_{n_3} - R^T R|}{|I_{n_3} - R^T R + \frac{\epsilon}{2-\epsilon} R^T R|} \\ &\rightarrow 0, \quad \text{as } \epsilon \rightarrow 0+. \end{aligned}$$

Similarly, for the joint mutual information and the interaction information:

$$I[\mathbf{X}_1, \mathbf{X}_2; \mathbf{X}_3] \rightarrow \frac{1}{2} \log \frac{1}{|I_{n_3} - Q^T Q|},$$

and

$$I[\mathbf{X}_1; \mathbf{X}_2; \mathbf{X}_3] \rightarrow -\frac{1}{2} \log \frac{1}{|I_{n_3} - Q^T Q|},$$

as expected, since $\epsilon \rightarrow 0+$.

Appendix I.3. Symmetry

To validate the symmetry property we require to prove that $I_{shd}[\mathbf{X}_1, \mathbf{X}_2; \mathbf{X}_3]$ is equal to $I_{shd}[\mathbf{X}_2, \mathbf{X}_1; \mathbf{X}_3]$. Swapping $\mathbf{X}_1$ and $\mathbf{X}_2$ means that the sources are now in order $\mathbf{X}_2, \mathbf{X}_1, \mathbf{X}_3$ and the covariance matrix in (12) becomes

$$\begin{bmatrix} I_{n_2} & P^T & R \\ P & I_{n_1} & Q \\ R^T & Q^T & I_{n_3} \end{bmatrix},$$

The switching of $\mathbf{X}_1$ and $\mathbf{X}_2$ means also that the probability distributions on S_5 and S_6 swap, since

$$\begin{aligned} S_5 : & \quad p(\mathbf{x}_2|\mathbf{x}_1)p(\mathbf{x}_3|\mathbf{x}_1)p(\mathbf{x}_1) \longrightarrow p(\mathbf{x}_1|\mathbf{x}_2)p(\mathbf{x}_3|\mathbf{x}_2)p(\mathbf{x}_2) \\ S_6 : & \quad p(\mathbf{x}_1|\mathbf{x}_2)p(\mathbf{x}_3|\mathbf{x}_2)p(\mathbf{x}_2) \longrightarrow p(\mathbf{x}_2|\mathbf{x}_1)p(\mathbf{x}_3|\mathbf{x}_1)p(\mathbf{x}_1) \end{aligned}$$

and the corresponding covariance matrices are

$$\Sigma'_5 = \begin{bmatrix} I_{n_2} & P^T & R \\ P & I_{n_1} & PR \\ R^T & R^T P^T & I_{n_3} \end{bmatrix} \quad \text{and} \quad \Sigma'_6 = \begin{bmatrix} I_{n_2} & P^T & P^T Q \\ P & I_{n_1} & Q \\ Q^T P^T & Q^T & I_{n_3} \end{bmatrix}$$

We now apply a similar argument to that used in the proof of Lemma 4.

$$|(1-t)\Sigma'_6 + t\Sigma'_5| = \left| \begin{bmatrix} A' & B(t)' \\ B(t)'^T & I_{n_3} \end{bmatrix} \right| = |A| |I_{n_3} - B(t)'^T A'^{-1} B(t)'|,$$

where A' and A'^{-1} are

$$A' = \begin{bmatrix} I_{n_2} & P^T \\ P & I_{n_1} \end{bmatrix}$$

$$A'^{-1} = \begin{bmatrix} I_{n_2} + P^T(I_{n_2} - PP^T)^{-1}P & -P^T(I_{n_2} - PP^T)^{-1} \\ -(I_{n_2} - PP^T)^{-1}P & (I_{n_2} - PP^T)^{-1} \end{bmatrix}$$

and

$$B'(t) = \begin{bmatrix} (1-t)P^T Q + tR \\ (1-t)Q + tPR \end{bmatrix}.$$

Some matrix calculation then gives

$$A'^{-1}B'(t) = \begin{bmatrix} tR \\ (1-t)Q \end{bmatrix}$$

and also

$$B(t)'^T A'^{-1} B'(t) = (1-t)^2 Q^T Q + t(1-t)(R^T P^T Q + Q^T PR) + t^2 R^T R.$$

Hence,

$$d'(t) = |I_{n_3} - (1-t)^2 Q^T Q - t(1-t)(R^T P^T Q + Q^T PR) - t^2 R^T R|$$

which is identical to the expression of $d(t)$ obtained in (A27). It follows that $d(t^*) = d'(t^*)$ and that the shared information is unchanged by swapping $\mathbf{X}_1$ and $\mathbf{X}_2$.

Appendix I.4. Monotonicity on the Redundancy Lattice

We use the term ‘redundancy’ here for convenience rather than ‘shared information’, since both terms mean exactly the same thing. A redundancy lattice is defined in [1]. When there are two sources and a target, there are four terms of interest that are usually denoted by

$\{\{1\}\{2\}, \{1\}, \{2\}, \{12\}\}$, which are the terms in the redundancy lattice. For monotonicity it is required that the redundancy values for these four terms satisfy the inequalities $\{1\}\{2\} \leq \{1\} \leq \{12\}$ and that $\{1\}\{2\} \leq \{2\} \leq \{12\}$.

The redundancy value for term $\{12\}$ is the self-redundancy $I_{shd}[(\mathbf{X}_1, \mathbf{X}_2), (\mathbf{X}_1, \mathbf{X}_2); \mathbf{X}_3]$ which is, by the self-redundancy property, equal to the joint mutual information $I[\mathbf{X}_1, \mathbf{X}_2; \mathbf{X}_3]$. Similarly, the redundancy values for the terms $\{1\}$ and $\{2\}$ are, by self-redundancy, the mutual information $I[\mathbf{X}_1; \mathbf{X}_3]$ and $I[\mathbf{X}_2; \mathbf{X}_3]$, respectively. The final term $\{1\}\{2\}$ is equal to the shared information defined in Proposition 1.

Since the PID defined in Proposition 1 possesses the non-negativity property it follows from (6)–(10) that the redundancy measure is monotonic on the redundancy lattice.

Appendix J. Computation

Code was written using R [46] in RStudio [46] to compute the I_{ig} PID. This code together with R code to compute the I_{dep} and I_{mmi} PIDs is available from <https://github.com/JWKay/PID> (accessed on 17 June 2024). The I_{ig} code first checks whether the input covariance matrix is positive-definite. If so, the feasible region F is computed by defining a function whose value indicates whether or not the matrix Σ_t is positive definite, and then applying the uniroot root-finding algorithm. The constrained optimisation is performed by using the base R function `optim`. The code produces a plot of $g(t)$ and returns the numerical results. Details of the pre-processing employed to make use of the I_{ig} formulae presented in Proposition 1 are the same as used with the I_{dep} PID and are available from Appendix D [22].

References

- Williams, P.L.; Beer, R.D. Nonnegative Decomposition of Multivariate Information. *arXiv* **2010**, arXiv:1004.2515.
- McGill, W.J. Multivariate Information Transmission. *Psychometrika* **1954**, *19*, 97–116. [CrossRef]
- Niu, X.; Quinn, C.J. A measure of Synergy, Redundancy, and Unique Information using Information Geometry. In Proceedings of the 2019 IEEE International Symposium on Information Theory (ISIT), Paris, France, 7–12 July 2019.
- Griffith, V.; Koch, C. Quantifying synergistic mutual information. In *Guided Self-Organization: Inception. Emergence, Complexity and Computation*; Springer: Berlin/Heidelberg, Germany, 2014; Volume 9, pp. 159–190.
- Bertschinger, N.; Rauh, J.; Olbrich, E.; Jost, J.; Ay, N. Quantifying Unique Information. *Entropy* **2014**, *16*, 2161–2183. [CrossRef]
- Harder, M.; Salge, C.; Polani, D. Bivariate measure of redundant information. *Phys. Rev. E* **2013**, *87*, 012130. [CrossRef] [PubMed]
- Ince, R.A.A. Measuring multivariate redundant information with pointwise common change in surprisal. *Entropy* **2017**, *19*, 318. [CrossRef]
- James, R.G.; Emenheiser, J.; Crutchfield, J.P. Unique Information via Dependency Constraints. *J. Phys. Math. Theor.* **2018**, *52*, 014002. [CrossRef]
- Finn, C.; Lizier, J.T. Pointwise Information Decomposition using the Specificity and Ambiguity Lattices. *arXiv* **2018**, arXiv:1801.09010.
- Lizier, J.T.; Bertschinger, N.; Jost, J.; Wibral, M. Information Decomposition of Target Effects from Multi-Source Interactions: Perspectives on Previous, Current and Future Work. *Entropy* **2018**, *20*, 307. [CrossRef]
- Makkeh, A.; Gutknecht, A.J.; Wibral, M. Introducing a differentiable measure of pointwise shared information. *Phys. Rev.* **2021**, *103*, 032149 [CrossRef]
- Gutknecht, A.J.; Wibral, M.; Makkeh, A. Bits and pieces: Understanding information decomposition from part-whole relationships and formal logic. *Proc. R. Soc. A* **2021**, *477*, 0110. [CrossRef]
- Kolchinsky, A. A Novel Approach to the Partial Information Decomposition. *Entropy* **2022**, *24*, 403. [CrossRef] [PubMed]
- Wibral, M.; Finn, C.; Wollstadt, P.; Lizier, J.T.; Priesemann, V. Quantifying Information Modification in Developing Neural Networks via Partial Information Decomposition. *Entropy* **2017**, *19*, 494. [CrossRef]
- Finn, C.; Lizier, J.T. Quantifying Information Modification in Cellular Automata Using Pointwise Partial Information Decomposition. In *Artificial Life Conference Proceedings*; MIT Press: Cambridge, MA, USA, 2018; pp. 386–387.
- Timme, N.; Alford, W.; Flecker, B.; Beggs, J.M. Synergy, redundancy, and multivariate information measures: An experimentalist's perspective. *J. Comput. Neurosci.* **2014**, *36*, 119–140. [CrossRef] [PubMed]
- Sherrill, S.P.; Timme, N.M.; Beggs, J.M.; Newman, E.L. Partial information decomposition reveals that synergistic neural integration is greater downstream of recurrent information flow in organotypic cortical cultures. *PLoS Comput. Biol.* **2021**, *17*, e1009196. [CrossRef] [PubMed]
- Pinto, H.; Pernice, R.; Silva, M.E.; Javorka, M.; Faes, L.; Rocha, A.P. Multiscale partial information decomposition of dynamical processes with short and long-range correlations: Theory and application to cardiovascular control. *Physiol. Meas.* **2022**, *43*, 085004. [CrossRef]

19. Ince, R.A.A.; Giordano, B.L.; Kayser, C.; Rousselet, G.A.; Gross, J.; Schyns, P.G. A Statistical Framework for Neuroimaging Data Analysis Based on Mutual Information Estimated via a Gaussian Copula. *Hum. Brain Mapp.* **2017**, *38*, 1541–1573. [[CrossRef](#)] [[PubMed](#)]
20. Barrett, A.B. An exploration of synergistic and redundant information sharing in static and dynamical Gaussian systems. *Phys. Rev. E* **2015**, *91*, 052802. [[CrossRef](#)] [[PubMed](#)]
21. Olbrich, E.; Bertschinger, N.; Rauh, J. Information decomposition and synergy. *Entropy* **2015**, *17*, 3501–3517. [[CrossRef](#)]
22. Kay, J.W.; Ince, R.A.A. Exact partial information decompositions for Gaussian systems based on dependency constraints. *Entropy* **2018**, *20*, 240. [[CrossRef](#)]
23. Venkatesh, P.; Schamberg, G. Partial Information Decomposition via Deficiency for Multivariate Gaussians. In Proceedings of the IEEE International Symposium on Information Theory (ISIT), Espoo, Finland, 26 June–1 July 2022; pp. 2892–2897.
24. Niu, X.; Quinn, C.J. Synergy and Redundancy Duality Between Gaussian Multiple Access and Broadcast Channels. In Proceedings of the 2020 International Symposium on Information Theory and Its Applications (ISITA), Kapolei, HI, USA, 24–27 October 2020; pp. 6–10.
25. Faes, F.; Marinazzo, D.; Stramaglia, S. Multiscale Information Decomposition: Exact Computation for Multivariate Gaussian Processes. *Entropy* **2017**, *19*, 408. [[CrossRef](#)]
26. Stramaglia, S.; Wu, G.-R.; Pellicoro, M.; Marinazzo, D. Expanding the transfer entropy to identify information circuits in complex systems. *Phys. Rev. E* **2012**, *86*, 066211. [[CrossRef](#)] [[PubMed](#)]
27. Daube, C.; Ince, R.A.A.; Gross, J. Simple acoustic features can explain phoneme-based predictions of cortical responses to speech. *Curr. Biol.* **2019**, *29*, 1924–1937. [[CrossRef](#)] [[PubMed](#)]
28. Park, H.; Ince, R.A.A.; Schyns, P.G.; Thut, G.; Gross, J. Representational interactions during audiovisual speech entrainment: Redundancy in left posterior superior temporal gyrus and synergy in left motor cortex. *PLoS Biol.* **2018**, *16*, e2006558. [[CrossRef](#)]
29. Schulz, J.M.; Kay, J.W.; Bischofberger, J.; Larkum, M.E. GABA_B Receptor-Mediated Regulation of Dendro-Somatic Synergy in Layer 5 Pyramidal Neurons. *Front. Cell. Neurosci.* **2021**, *15*, 718413. [[CrossRef](#)] [[PubMed](#)]
30. Newman, E.L.; Varley, T.F.; Parakkattu, V.K.; Sherrill, S.P.; Beggs, J.M. Revealing the Dynamics of Neural Information Processing with Multivariate Information Decomposition. *Entropy* **2022**, *24*, 930. [[CrossRef](#)] [[PubMed](#)]
31. Amari, S.-I. Information geometry on hierarchy of probability distributions. *IEEE Trans. Inf. Theory* **2001**, *47*, 1701–1711. [[CrossRef](#)]
32. Amari, S.-I. *Information Geometry and Its Applications*; Springer: Tokyo, Japan, 2020.
33. Amari, S.-I.; Nagaoka, H. *Methods of Information Geometry*; American Mathematical Society: Providence, Rhode Island, 2000.
34. Lauritzen, S.L. *Graphical Models*; Oxford University Press: Oxford, UK, 1996.
35. Zwick, M. An overview of reconstructability analysis. *Kybernetes* **2004**, *33*, 877–905. [[CrossRef](#)]
36. Ay, N.; Polani, D.; Virgo, N. Information Decomposition Based on Cooperative Game Theory. *Kybernetika* **2020**, *56*, 979–1014. [[CrossRef](#)]
37. Sugiyama, M.; Nakahara, H.; Tsuda, K. Information Decomposition on Structured Space. In Proceedings of the IEEE International Symposium on Information Theory (ISIT), Barcelona, Spain, 10–15 July 2016; pp. 575–579.
38. Bailey, R.A. Hasse diagrams as a visual aid for linear models and analysis of variance. *Commun. Stat. Theory Methods* **2021**, *50*, 5034–5067. [[CrossRef](#)]
39. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*; Wiley-Interscience: New York, NY, USA, 1991.
40. Whittaker, J. *Graphical Models in Applied Multivariate Statistics*; Wiley: Chichester, UK, 2008.
41. Eguchi, E.; Komori, O. *Minimum Divergence Methods in Statistical Machine Learning*; Springer: Tokyo, Japan, 2022.
42. Wibral, M.; Priesemann, V.; Kay, J.W.; Lizier, J.T.; Phillips, W.A. Partial information decomposition as a unified approach to the specification of neural goal functions. *Brain Cogn.* **2017**, *112*, 25–38. [[CrossRef](#)]
43. Kullback, S. *Information Theory and Statistics*; Courier Corporation: Gloucester, MA, USA, 1978.
44. Horn, R.A.; Johnson, C.R. *Matrix Analysis*; Cambridge University Press: New York, NY, USA, 1985.
45. Wigner, E.P. On weakly positive matrices. *Can. J. Math.* **1963**, *15*, 313–318. [[CrossRef](#)]
46. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2021. Available online: <https://www.R-project.org/> (accessed on 20 May 2024).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.