

Article

Intersection Information Based on Common Randomness

Virgil Griffith ^{1,*}, Edwin K. P. Chong ², Ryan G. James ³, Christopher J. Ellison ⁴ and James P. Crutchfield ^{5,6}

¹ Computation and Neural Systems, Caltech, Pasadena, CA 91125, USA

² Dept. of Electrical & Computer Engineering, Colorado State University, Fort Collins, CO 80523, USA; E-Mail: edwin.chong@colostate.edu

³ Department of Computer Science, University of Colorado, Boulder, CO 80309, USA; E-Mail: ryan.james@colorado.edu

⁴ Center for Complexity and Collective Computation, Wisconsin Institute for Discovery, University of Wisconsin-Madison, Madison, WI 53715, USA; E-Mail: cellison@wisc.edu

⁵ Complexity Sciences Center and Physics Dept, University of California Davis, Davis, CA 95616, USA; E-Mail: chaos@cse.ucdavis.edu

⁶ Santa Fe Institute, 1399 Hyde Park Rd, Santa Fe, NM 87501, USA

* Author to whom correspondence should be addressed; E-Mail: i@virgil.gr.

Received: 25 October 2013; in revised form: 27 March 2014 / Accepted: 28 March 2014 /

Published: 4 April 2014

Abstract: The introduction of the partial information decomposition generated a flurry of proposals for defining an intersection information that quantifies how much of “the same information” two or more random variables specify about a target random variable. As of yet, none is wholly satisfactory. A palatable measure of intersection information would provide a principled way to quantify slippery concepts, such as synergy. Here, we introduce an intersection information measure based on the Gács-Körner common random variable that is the first to satisfy the coveted target monotonicity property. Our measure is imperfect, too, and we suggest directions for improvement.

Keywords: intersection information; partial information decomposition; lattice; Gács-Körner; synergy; redundant information

1. Introduction

Partial information decomposition (PID) [1] is an immensely suggestive framework for deepening our understanding of multivariate interactions, particularly our understanding of informational redundancy and synergy. In general, one seeks a decomposition of the mutual information that n predictors X_1, \dots, X_n convey about a target random variable, Y . The intersection information is a function that calculates the information that every predictor conveys about the target random variable; the name draws an analogy with intersections in set theory. An anti-chain lattice of redundant, unique and synergistic partial information is then built from the intersection information.

As an intersection information measure, [1] proposes the quantity:

$$I_{\min}(X_1, \dots, X_n : Y) = \sum_{y \in \mathcal{Y}} \Pr(y) \min_{i \in \{1, \dots, n\}} D_{\text{KL}} \left[\Pr(X_i | y) \parallel \Pr(X_i) \right], \quad (1)$$

where D_{KL} is the Kullback–Leibler divergence. Although I_{\min} is a plausible choice for the intersection information, it has several counterintuitive properties that make it unappealing [2]. In particular, I_{\min} is not sensitive to the possibility that differing predictors, X_i and X_j , can reduce uncertainty about Y in nonequivalent ways. Moreover, the \min operator effectively treats all uncertainty reductions as the same, causing it to overestimate the ideal intersection information. The search for an improved intersection information measure ensued and continued through [3–5], and today, a widely accepted intersection information measure remains undiscovered.

Here, we do not definitively solve this problem, but explore a candidate intersection information based on the so-called common random variable [6]. Whereas Shannon mutual information is relevant to communication channels with arbitrarily small error, the entropy of the common random variable (also known as the zero-error information) is relevant to communication channels without error [7]. We begin by proposing a measure of intersection information for the simpler zero-error information case. This is useful in and of itself, because it provides a template for exploring intersection information measures. Then, we modify our proposal, adapting it to the Shannon mutual information case.

The next section introduces several definitions, some notation and a necessary lemma. We extend and clarify the desired properties for intersection information. Section 3 introduces zero-error information and its intersection information measure. Section 4 uses the same methodology to produce a novel candidate for the Shannon intersection information. Section 5 shows the successes and shortcoming of our candidate intersection information measure using example circuits and diagnoses the shortcoming's origin. Section 6 discusses the negative values of the resulting synergy measure and identifies its origin. Section 7 summarizes our progress towards the ideal intersection information measure and suggests directions for improvement. Appendices are devoted to technical lemmas and their proofs, to which we refer in the main text.

2. Preliminaries

2.1. Informational Partial Order and Equivalence

We assume an underlying probability space on which we define random variables denoted by capital letters (e.g., X , Y and Z). In this paper, we consider only random variables taking values on finite

spaces. Given random variables X and Y , we write $X \preceq Y$ to signify that there exists a measurable function, f , such that $X = f(Y)$ almost surely (*i.e.*, with probability one). In this case, following the terminology in [8], we say that X is informationally poorer than Y ; this induces a partial order on the set of random variables. Similarly, we write $X \succeq Y$ if $Y \preceq X$, in which case we say X is informationally richer than Y .

If X and Y are such that $X \preceq Y$ and $X \succeq Y$, then we write $X \cong Y$. In this case, again following [8], we say that X and Y are informationally equivalent. In other words, $X \cong Y$ if and only if one can relabel the values of X to obtain a random value that is equal to Y almost surely and *vice versa*.

This “information-equivalence” can easily be shown to be an equivalence relation, and it partitions the set of all random variables into disjoint equivalence classes. The \preceq ordering is invariant within these equivalence classes in the following sense. If $X \preceq Y$ and $Y \cong Z$, then $X \preceq Z$. Similarly, if $X \preceq Y$ and $X \cong Z$, then $Z \preceq Y$. Moreover, within each equivalence class, the entropy is invariant, as shown in Section 2.2.

2.2. Information Lattice

Next, we follow [8] and consider the join and meet operators. These operators were defined for information elements, which are σ -algebras or, equivalently, equivalence classes of random variables. We deviate from [8] slightly and define the join and meet operators for random variables.

Given random variables X and Y , we define $X \vee Y$ (called the join of X and Y) to be an informationally poorest (“smallest” in the sense of the partial order \preceq) random variable, such that $X \preceq X \vee Y$ and $Y \preceq X \vee Y$. In other words, if Z is such that $X \preceq Z$ and $Y \preceq Z$, then $X \vee Y \preceq Z$. Note that $X \vee Y$ is unique only up to equivalence with respect to \cong . In other words, $X \vee Y$ does not define a specific, unique random variable. Nonetheless, standard information-theoretic quantities are invariant over the set of random variables satisfying the condition specified above. For example, the entropy of $X \vee Y$ is invariant over the entire equivalence class of random variables satisfying the condition above. Similarly, the inequality $Z \preceq X \vee Y$ does not depend on the specific random variable chosen, as long as it satisfies the condition above. Note, the pair (X, Y) is an instance of $X \vee Y$.

In a similar vein, given random variables X and Y , we define $X \wedge Y$ (called the meet of X and Y) to be an informationally richest random variable (“largest” in the sense of \succeq), such that $X \wedge Y \preceq X$ and $X \wedge Y \preceq Y$. In other words, if Z is such that $Z \preceq X$ and $Z \preceq Y$, then $Z \preceq X \wedge Y$. Following [6], we also call $X \wedge Y$ the common random variable of X and Y .

An algorithm for computing an instance of the common random variable between two random variables is provided in [7]; it generalizes straightforwardly to n random variables. One can also take intersections of the σ -algebras generated by the random variables that define the meet.

The \vee and \wedge operators satisfy the algebraic properties of a lattice [8]. In particular, the following hold:

- commutative laws: $X \vee Y \cong Y \vee X$ and $X \wedge Y \cong Y \wedge X$;
- associative laws: $X \vee (Y \vee Z) \cong (X \vee Y) \vee Z$ and $X \wedge (Y \wedge Z) \cong (X \wedge Y) \wedge Z$;
- absorption laws: $X \vee (X \wedge Y) \cong X$ and $X \wedge (X \vee Y) \cong X$;
- idempotent laws: $X \vee X \cong X$ and $X \wedge X \cong X$;
- generalized absorption laws: if $X \preceq Y$, then $X \vee Y \cong Y$ and $X \wedge Y \cong X$.

Finally, the partial order \preceq is preserved under \vee and \wedge , *i.e.*, if $X \preceq Y$, then $X \vee Z \preceq Y \vee Z$ and $X \wedge Z \preceq Y \wedge Z$.

Let $H(\cdot)$ represent the entropy function and $H(\cdot|\cdot)$ the conditional entropy. We denote the Shannon mutual information between X and Y by $I(X:Y)$. The following results highlight the invariance and monotonicity of the entropy and conditional entropy functions with respect to \cong and \preceq [8]. Given that $X \preceq Y$ if and only if $X = f(Y)$, these results are familiar in information theory, but are restated here using the current notation:

- (a) If $X \cong Y$, then $H(X) = H(Y)$, $H(X|Z) = H(Y|Z)$, and $H(Z|X) = H(Z|Y)$.
- (b) If $X \preceq Y$, then $H(X) \leq H(Y)$, $H(X|Z) \leq H(Y|Z)$, and $H(Z|X) \geq H(Z|Y)$.
- (c) $X \preceq Y$ if and only if $H(X|Y) = 0$.

2.3. Desired Properties of Intersection Information

We denote $\mathcal{I}(X:Y)$ as a nonnegative measure of information between X and Y . For example, \mathcal{I} could be the Shannon mutual information; *i.e.*, $\mathcal{I}(X:Y) \equiv I(X:Y)$. Alternatively, we could take \mathcal{I} to be the zero-error information. Yet, other possibilities include the Wyner common information [9] or the quantum mutual information [10]. Generally, though, we require that $\mathcal{I}(X:Y) = 0$ if Y is a constant, which is satisfied by both the zero-error and Shannon information.

For a given choice of \mathcal{I} , we seek a function that captures the amount of information about Y that is captured by each of the predictors X_1, \dots, X_n . We say that I_\cap is an intersection information for \mathcal{I} if $I_\cap(X:Y) = \mathcal{I}(X:Y)$. There are currently 11 intuitive properties that we wish the ideal intersection information measure, I_\cap , to satisfy. Some are new (e.g., lower bound (LB), strong monotonicity (M₁), and equivalence-class invariance (Eq)), but most were introduced earlier, in various forms, in [1–5]. They are as follows:

- (GP) Global positivity: $I_\cap(X_1, \dots, X_n:Y) \geq 0$.
- (Eq) Equivalence-class invariance: $I_\cap(X_1, \dots, X_n:Y)$ is invariant under substitution of X_i (for any $i = 1, \dots, n$) or Y by an informationally equivalent random variable.
- (TM) Target monotonicity: If $Y \preceq Z$, then $I_\cap(X_1, \dots, X_n:Y) \leq I_\cap(X_1, \dots, X_n:Z)$.
- (M₀) Weak monotonicity: $I_\cap(X_1, \dots, X_n, W:Y) \leq I_\cap(X_1, \dots, X_n:Y)$ with equality if there exists a $Z \in \{X_1, \dots, X_n\}$ such that $Z \preceq W$.
- (S₀) Weak symmetry: $I_\cap(X_1, \dots, X_n:Y)$ is invariant under reordering of X_1, \dots, X_n .

The next set of properties relate the intersection information to the chosen measure of information between X and Y .

- (LB) Lower bound: If $Q \preceq X_i$ for all $i = 1, \dots, n$, then $I_\cap(X_1, \dots, X_n:Y) \geq \mathcal{I}(Q:Y)$. Note that $X_1 \wedge \dots \wedge X_n$ is a valid choice for Q . Furthermore, given that we require $I_\cap(X:Y) = \mathcal{I}(X:Y)$, it follows that (M₀) implies (LB).
- (Id) Identity: $I_\cap(X, Y:X \vee Y) = \mathcal{I}(X:Y)$.
- (LP₀) Weak local positivity: For $n = 2$ predictors, the derived “partial information” defined in [1] and described in Section 5 are nonnegative. If both (GP) and (M₀) are satisfied, as well as $I_\cap(X_1, X_2:Y) \geq \mathcal{I}(X_1:Y) + \mathcal{I}(X_2:Y) - \mathcal{I}(X_1 \vee X_2:Y)$, then (LP₀) is satisfied.

Finally, we have the “strong” properties:

- (M₁) Strong monotonicity: $I_{\cap}(X_1, \dots, X_n, W : Y) \leq I_{\cap}(X_1, \dots, X_n : Y)$ with equality if there exists $Z \in \{X_1, \dots, X_n, Y\}$ such that $Z \preceq W$.
- (S₁) Strong symmetry: $I_{\cap}(X_1, \dots, X_n : Y)$ is invariant under reordering of X_1, \dots, X_n, Y .
- (LP₁) Strong local positivity: For all n , the derived “partial information” defined in [1] is nonnegative.

Properties (LB), (M₁) and (Eq) are introduced for the first time here. However, (Eq) is satisfied by most information-theoretic quantities and is implicitly assumed by others. Though absent from our list, it is worthwhile to also consider continuity and chain rule properties, in analogy with the mutual information [4,11].

3. Candidate Intersection Information for Zero-Error Information

3.1. Zero-Error Information

Introduced in [7], the zero-error information, or Gács–Körner common information, is a stricter variant of Shannon mutual information. Whereas the mutual information, $I(A : B)$, quantifies the magnitude of information A conveys about B with an arbitrarily small error $\epsilon > 0$, the zero-error information, denoted $I^0(A : B)$, quantifies the magnitude of information A conveys about B with exactly zero error, *i.e.*, $\epsilon = 0$. The zero-error information between A and B equals the entropy of the common random variable $A \wedge B$,

$$I^0(A : B) \equiv H(A \wedge B).$$

Zero-error information has several notable properties, but the most salient is that it is nonnegative and bounded by the mutual information,

$$0 \leq I^0(A : B) \leq I(A : B).$$

3.2. Intersection Information for Zero-Error Information

For the zero-error information case (*i.e.*, $\mathcal{I} = I^0$), we propose the zero-error intersection information $I_{\lambda}^0(X_1, \dots, X_n : Y)$ as the maximum zero-error information, $I^0(Q : Y)$, that a random variable, Q , conveys about Y , subject to Q being a function of each predictor X_1, \dots, X_n :

$$I_{\lambda}^0(X_1, \dots, X_n : Y) \equiv \max_{\Pr(Q|Y)} I^0(Q : Y) \tag{2}$$

subject to $Q \preceq X_i \forall i \in \{1, \dots, n\}$.

In Lemma 7 of Appendix 7.3, it is shown that the common random variable across all predictors is the maximizing Q . This simplifies Equation (2) to:

$$I_{\lambda}^0(X_1, \dots, X_n : Y) = I^0(X_1 \wedge \dots \wedge X_n : Y) = H(X_1 \wedge \dots \wedge X_n \wedge Y). \tag{3}$$

Most importantly, the zero-error information $I_{\lambda}^0(X_1, \dots, X_n : Y)$ satisfies nine of the 11 desired properties from Section 2.3, leaving only (LP₀) and (LP₁) unsatisfied. See Lemmas 1, 2, and 3 in Appendix 7.3 for details.

4. Candidate Intersection Information for Shannon Information

In the last section, we defined an intersection information for zero-error information that satisfies the vast majority of the desired properties. This is a solid start, but an intersection information for Shannon mutual information remains the goal. Towards this end, we use the same method as in Equation (2), leading to I_λ , our candidate intersection information measure for Shannon mutual information:

$$I_\lambda(X_1, \dots, X_n : Y) \equiv \max_{Pr(Q|Y)} I(Q : Y) \tag{4}$$

subject to $Q \preceq X_i \forall i \in \{1, \dots, n\}$.

In Lemma 8 of Appendix 7.3, it is shown that Equation (4) simplifies to:

$$I_\lambda(X_1, \dots, X_n : Y) = I(X_1 \wedge \dots \wedge X_n : Y) . \tag{5}$$

Unfortunately, I_λ does not satisfy as many of the desired properties as I_λ^0 . However, our candidate, I_λ , still satisfies seven of the 11 properties; most importantly, the coveted (TM) that, until now, had not been satisfied by any proposed measure. See Lemmas 4, 5 and 6 in Appendix 7.3 for details. Table 1 lists the desired properties satisfied by I_{\min} , I_λ and I_λ^0 . For reference, we also include I_{red} , the proposed measure from [3].

Table 1. The I_\cap desired properties that each measure satisfies. (The appendices provide proofs for I_λ and I_λ^0 .)

	Property	I_{\min}	I_{red}	I_λ	I_λ^0
(GP)	Global Positivity	✓	✓	✓	✓
(Eq)	Equivalence-Class Invariance	✓	✓	✓	✓
(TM)	Target Monotonicity			✓	✓
(M ₀)	Weak Monotonicity	✓		✓	✓
(S ₀)	Weak Symmetry	✓	✓	✓	✓
(LB)	Lower bound	✓	✓	✓	✓
(Id)	Identity		✓		✓
(LP ₀)	Weak Local Positivity	✓	✓		
(M ₁)	Strong Monotonicity				✓
(S ₁)	Strong Symmetry				✓
(LP ₁)	Strong Local Positivity	✓			

Lemma 9 in Appendix 7.3 allows a comparison of the three subject intersection information measures:

$$0 \leq I_\lambda^0(X_1, \dots, X_n : Y) \leq I_\lambda(X_1, \dots, X_n : Y) \leq I_{\min}(X_1, \dots, X_n : Y) . \tag{6}$$

Despite not satisfying (LP₀), I_λ remains an important stepping-stone towards the ideal Shannon I_\cap . First, I_λ captures what is inarguably redundant information (the common random variable); this makes

I_{λ} necessarily a lower bound on any reasonable redundancy measure. Second, it is the first proposal to satisfy target monotonicity. Lastly, I_{λ} is the first measure to reach intuitive answers in many canonical situations, while also being generalizable to an arbitrary number of inputs.

5. Three Examples Comparing I_{\min} and I_{λ}

Example UNQ illustrates how I_{\min} gives undesirable (some claim fatally so [2]) decompositions of redundant and synergistic information. Examples UNQ and RDNXOR illustrate I_{λ} 's successes and example IMPERFECTRDN illustrates I_{λ} 's paramount deficiency. For each, we give the joint distribution $\Pr(x_1, x_2, y)$, a diagram and the decomposition derived from setting I_{\min} or I_{λ} as the I_{\cap} measure. At each lattice junction, the left number is the I_{\cap} value of that node, and the number in parentheses is the I_{∂} value (this is the same notation used in [4]). Readers unfamiliar with the $n = 2$ partial information lattice should consult [1], but in short, I_{∂} measures the magnitude of “new” information at this node in the lattice beyond the nodes lower in the lattice. Specifically, the mutual information between the pair, $X_1 \vee X_2$ and Y , decomposes into four terms:

$$\mathcal{I}(X_1 \vee X_2 : Y) = I_{\partial}(X_1, X_2 : Y) + I_{\partial}(X_1 : Y) + I_{\partial}(X_2 : Y) + I_{\partial}(X_1 \vee X_2 : Y) .$$

In order, the terms are given by the redundant information that X_1 and X_2 both provide to Y , the unique information that X_1 provides to Y , the unique information that X_2 provides to Y and finally, the synergistic information that X_1 and X_2 jointly convey about Y . Each of these quantities can be written in terms of standard mutual information and the intersection information, I_{\cap} , as follows:

Redundant	$I_{\partial}(X_1, X_2 : Y)$	=	$I_{\cap}(X_1, X_2 : Y)$
Unique	$I_{\partial}(X_1 : Y)$	=	$\mathcal{I}(X_1 : Y) - I_{\cap}(X_1, X_2 : Y)$
Unique	$I_{\partial}(X_2 : Y)$	=	$\mathcal{I}(X_2 : Y) - I_{\cap}(X_1, X_2 : Y)$
Synergetic	$I_{\partial}(X_1 \vee X_2 : Y)$	=	$\mathcal{I}(X_1 \vee X_2 : Y) - \mathcal{I}(X_1 : Y) - \mathcal{I}(X_2 : Y) + I_{\cap}(X_1, X_2 : Y)$

These quantities occupy the bottom, left, right and top nodes in the lattice diagrams, respectively. Except for IMPERFECTRDN, measures I_{λ} and I_{\cap}^0 reach the same decomposition for all presented examples.

5.1. Example UNQ (Figure 1)

The desired decomposition for example UNQ is two bits of unique information; X_1 uniquely specifies one bit of Y , and X_2 uniquely specifies the other bit of Y . The chief criticism of I_{\min} in [2] was that I_{\min} calculated one bit of redundancy and one bit of synergy for UNQ (Figure 1c). We see that unlike I_{\min} , I_{λ} satisfyingly arrives at two bits of unique information. This is easily seen by the inequality,

$$0 \leq I_{\lambda}(X_1, X_2 : Y) \leq H(X_1 \wedge X_2) \leq I(X_1 : X_2) = 0 \text{ bits} . \tag{7}$$

Therefore, as $I(X_1 : X_2) = 0$, we have $I_{\lambda}(X_1, X_2 : Y) = 0$ bits leading to $I_{\partial}(X_1 : Y) = 1$ bit and $I_{\partial}(X_2 : Y) = 1$ bit (Figure 1d).

Figure 1. Example UNQ. This is the canonical example of unique information. X_1 and X_2 each uniquely specify a single bit of Y . This is the simplest example, where I_{\min} calculates an undesirable decomposition (c) of one bit of redundancy and one bit of synergy. I_{λ} and I_{λ}^0 each calculate the desired decomposition (d). (a) Distribution and information quantities; (b) circuit diagram; (c) I_{\min} ; (d) I_{λ} and I_{λ}^0 .

X_1	X_2	Y	
a	b	ab	$1/4$
a	B	aB	$1/4$
A	b	Ab	$1/4$
A	B	AB	$1/4$

$$I(X_1 \vee X_2 : Y) = 2$$

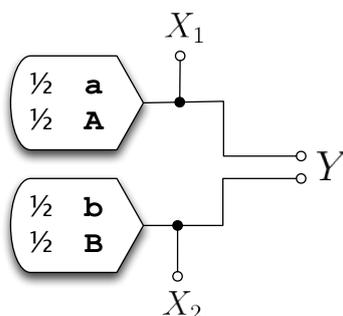
$$I(X_1 : Y) = 1$$

$$I(X_2 : Y) = 1$$

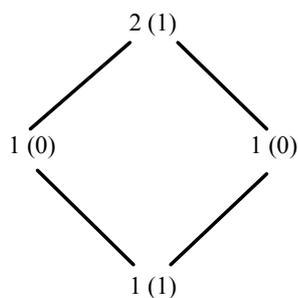
$$I_{\min}(X_1, X_2 : Y) = 1$$

$$I_{\lambda}(X_1, X_2 : Y) = 0$$

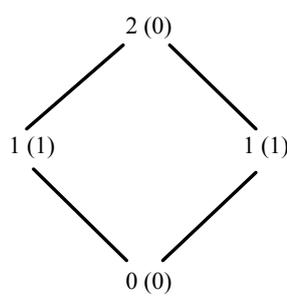
(a)



(b)



(c)



(d)

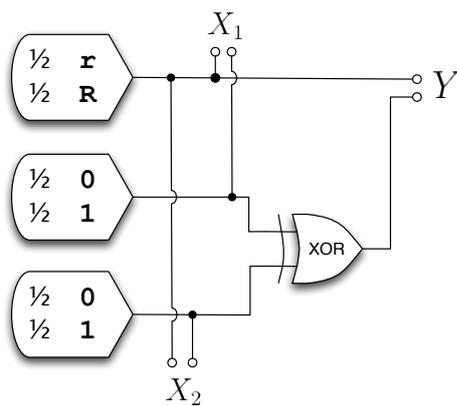
5.2. Example RDNXOR (Figure 2)

In [2], RDNXOR was an example where I_{\min} shined by reaching the desired decomposition of one bit of redundancy and one bit of synergy. We see that I_{λ} finds this same answer. I_{λ} extracts the common random variable within X_1 and X_2 —the r/R bit—and calculates the mutual information between the common random variable and Y to arrive at $I_{\lambda}(X_1, X_2 : Y) = 1$ bit.

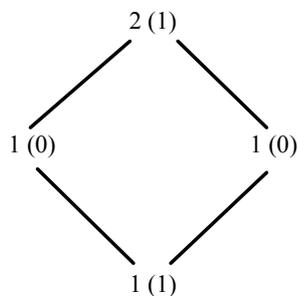
Figure 2. Example RDNXOR. This is the canonical example of redundancy and synergy coexisting. I_{\min} and I_{λ} each reach the desired decomposition of one bit of redundancy and one bit of synergy. This is the simplest example demonstrating I_{λ} and I_{λ}^0 correctly extracting the embedded redundant bit within X_1 and X_2 . **(a)** Distribution and information quantities; **(b)** circuit diagram; **(c)** I_{\min} ; **(d)** I_{λ} and I_{λ}^0 .

X_1	X_2	Y		
r0	r0	r0	$\frac{1}{8}$	$I(X_1 \vee X_2 : Y) = 2$
r0	r1	r1	$\frac{1}{8}$	
r1	r0	r1	$\frac{1}{8}$	
r1	r1	r0	$\frac{1}{8}$	
R0	R0	R0	$\frac{1}{8}$	$I(X_1 : Y) = 1$
R0	R1	R1	$\frac{1}{8}$	$I(X_2 : Y) = 1$
R1	R0	R1	$\frac{1}{8}$	$I_{\min}(X_1, X_2 : Y) = 1$
R1	R1	R0	$\frac{1}{8}$	$I_{\lambda}(X_1, X_2 : Y) = 1$

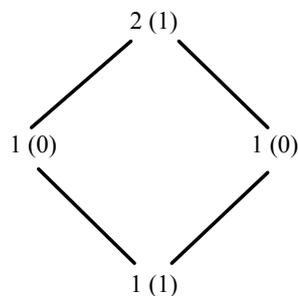
(a)



(b)



(c)



(d)

5.3. Example IMPERFECTRDN (Figure 3)

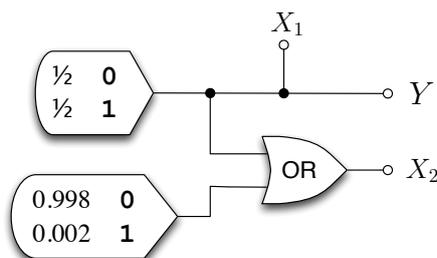
IMPERFECTRDN highlights the foremost shortcoming of I_λ : It does not detect “imperfect” or “lossy” correlations between X_1 and X_2 . Given (LP_0) , we can determine the desired decomposition analytically. First, $I(X_1 \vee X_2 : Y) = I(X_1 : Y) = 1$ bit, and thus, $I(X_2 : Y | X_1) = 0$ bits. Since the conditional mutual information is the sum of the synergy $I_\partial(X_1, X_2 : Y)$ and unique information $I_\partial(X_2 : Y)$, both quantities must also be zero. Then, the redundant information $I_\partial(X_1, X_2 : Y) = I(X_2 : Y) - I_\partial(X_2 : Y) = I(X_2 : Y) = 0.99$ bits. Having determined three of the partial informations, we compute the final unique information: $I_\partial(X_1 : Y) = I(X_1 : Y) - 0.99 = 0.01$ bits.

Figure 3. Example IMPERFECTRDN. I_λ is blind to the noisy correlation between X_1 and X_2 and calculates zero redundant information. An ideal I_\cap measure would detect that all of the information X_2 specifies about Y is also specified by X_1 to calculate $I_\cap(X_1, X_2 : Y) = 0.99$ bits. (a) Distribution and information quantities; (b) circuit diagram; (c) I_{\min} ; (d) I_λ ; (e) I_λ^0 .

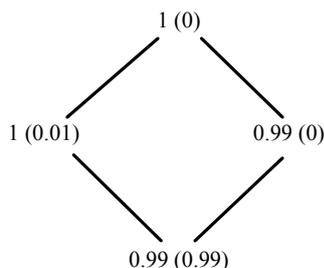
X_1	X_2	Y	
0	0	0	0.499
0	1	0	0.001
1	1	1	0.500

$$\begin{aligned}
 I(X_1 \vee X_2 : Y) &= 1 \\
 I(X_1 : Y) &= 1 \\
 I(X_2 : Y) &= 0.99 \\
 I_{\min}(X_1, X_2 : Y) &= 0.99 \\
 I_\lambda(X_1, X_2 : Y) &= 0
 \end{aligned}$$

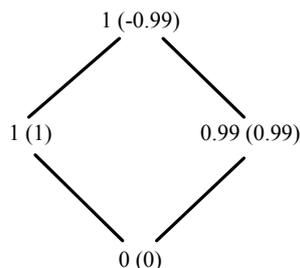
(a)



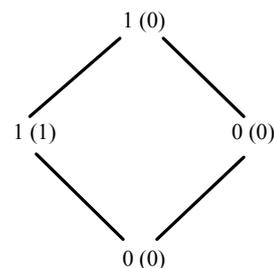
(b)



(c)



(d)



(e)

How well do I_{\min} and I_{λ} match the desired decomposition of IMPERFECTRDN? We see that I_{\min} calculates the desired decomposition (Figure 3c); however, I_{λ} does not (Figure 3d). Instead, I_{λ} calculates zero redundant information, that $I_{\cap}(X_1, X_2 : Y) = 0$ bits. This unpleasant answer arises from $\Pr(X_1 = 0, X_2 = 1) > 0$. If this were zero, then both I_{λ} and I_{\min} reach the desired one bit of redundant information. Due to the nature of the common random variable, I_{λ} only sees the “deterministic” correlations between X_1 and X_2 ; add even an iota of noise between X_1 and X_2 , and I_{λ} plummets to zero. This highlights the fact that I_{λ} is not continuous: an arbitrarily small change in the probability distribution can result in a discontinuous jump in the value of I_{λ} . As with traditional information measures, such as the entropy and the mutual information, it may be desirable to have an I_{\cap} measure that is continuous over the simplex.

To summarize, IMPERFECTRDN shows that when there are additional “imperfect” correlations between A and B , *i.e.*, $I(A : B | A \wedge B) > 0$, I_{λ} sometimes underestimates the ideal $I_{\cap}(A, B : Y)$.

6. Negative Synergy

In IMPERFECTRDN, we saw I_{λ} calculate a synergy of -0.99 bits (Figure 3d). What does this mean? Could negative synergy be a “real” property of Shannon information? When $n = 2$, it is fairly easy to diagnose the cause of negative synergy from the equation for $I_{\theta}(X_1 \curlyvee X_2 : Y)$ in Equation (7). Given (GP), negative synergy occurs if and only if,

$$I(X_1 \curlyvee X_2 : Y) < I(X_1 : Y) + I(X_2 : Y) - I_{\cap}(X_1, X_2 : Y) = I_{\cup}(X_1, X_2 : Y), \tag{8}$$

where I_{\cup} is dual to I_{\cap} and related by the inclusion-exclusion principle. For arbitrary n , this is $I_{\cup}(X_1, \dots, X_n : Y) \equiv \sum_{S \subseteq \{X_1, \dots, X_n\}} (-1)^{|S|+1} I_{\cap}(S_1, \dots, S_{|S|} : Y)$. The intuition behind I_{\cup} is that it represents the aggregate information contributed by the sources, X_1, \dots, X_n , without considering synergies or double-counting redundancies.

From Equation (8), we see that negative synergy occurs when I_{\cap} is small, probably too small. Equivalently, negative synergy occurs when the joint random variable conveys less about Y than the sources, X_1 and X_2 , convey separately; mathematically, when $I(X_1 \curlyvee X_2 : Y) < I_{\cup}(X_1, X_2 : Y)$. On the face of it, this sounds strange. No structure “disappears” after X_1 and X_2 are combined by the \curlyvee operator. By the definition of \curlyvee , there are always functions f_1 and f_2 , such that $X_1 \cong f_1(Z)$ and $X_2 \cong f_2(Z)$. Therefore, if your favorite I_{\cap} measure does not satisfy (LP₀), it is too strict.

This means that our measure, I_{λ}^0 , does not account for the full zero-information overlap between $I^0(X_1 : Y)$ and $I^0(X_2 : Y)$. This is shown in the example, SUBTLE (Figure 4), where I_{λ}^0 calculates a synergy of -0.252 bits. Defining a zero-error, I_{\cap} , that satisfies (LP₀) is a matter of ongoing research.

Figure 4. Example SUBTLE. In this example, both I_λ and I_λ^0 calculate a synergy of -0.252 bits of synergy. What kind of redundancy must be captured for a nonnegative decomposition for this example? **(a)** Distribution and information quantities; **(b)** circuit diagram; **(c)** I_{\min} ; **(d)** I_λ and I_λ^0 .

X_1	X_2	Y	
0	0	00	$\frac{1}{3}$
0	1	01	$\frac{1}{3}$
1	1	11	$\frac{1}{3}$

$$I(X_1 \vee X_2 : Y) = 1.585$$

$$I(X_1 : Y) = 0.918$$

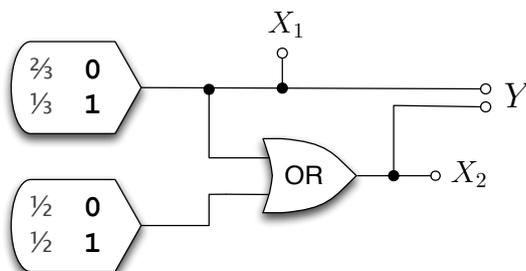
$$I(X_2 : Y) = 0.918$$

$$I(X_1 : X_2) = 0.252$$

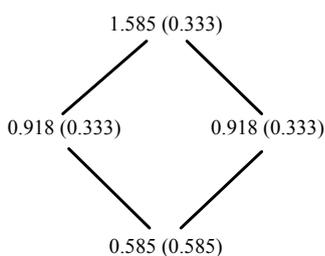
$$I_{\min}(X_1, X_2 : Y) = 0.585$$

$$I_\lambda(X_1, X_2 : Y) = 0.0$$

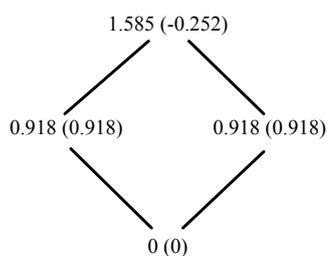
(a)



(b)



(c)



(d)

7. Conclusions and Path Forward

We made incremental progress on several fronts towards the ideal Shannon I_\cap .

7.1. Desired Properties

We have expanded, tightened and grounded the desired properties for I_\cap . Particularly,

- **(LB)** highlights an uncontentious, yet tighter lower bound on I_\cap than **(GP)**.
- Inspired by $I_\cap(X_1 : Y) = \mathcal{I}(X_1 : Y)$ and **(M₀)** synergistically implying **(LB)**, we introduced **(M₁)** as a desired property.

- What was before an implicit assumption, we introduced (Eq) to better ground one's thinking.

7.2. A New Measure

Based on the Gács–Körner common random variable, we introduced a new Shannon I_{\cap} measure. Our measure, I_{λ} , is theoretically principled and the first to satisfy (TM). A point to keep in mind is that our intersection information is zero whenever the distribution $\text{Pr}(x_1, x_2, y)$ has full support; this dependence on structural zeros is inherited from the common random variable.

7.3. How to Improve

We identified where I_{λ} fails; it does not detect “imperfect” correlations between X_1 and X_2 . One next step is to develop a less stringent I_{\cap} measure that satisfies (LP₀) for IMPERFECTRDN, while still satisfying (TM). Satisfying continuity would also be a good next step.

Contrary to our initial expectation, SUBTLE, showed that I_{λ}^0 does not satisfy (LP₀). This matches a result from [4], which shows that (LP₀), (S₁), (M₀) and (Id) cannot all be simultaneously satisfied, and it suggests that I_{λ}^0 is too strict. Therefore, what kind of zero-error informational overlap is I_{λ}^0 not capturing? The answer is of paramount importance. The next step is to formalize what exactly is required for a zero-error I_{\cap} to satisfy (LP₀). From SUBTLE, we can likewise see that within zero-error information, (Id) and (LP₀) are incompatible.

Acknowledgments

Virgil Griffith thanks Tracey Ho, and Edwin K. P. Chong thanks Hua Li for valuable discussions. While intrepidly pulling back the veil of ignorance, Virgil Griffith was funded by a Department of Energy Computational Science Graduate Fellowship; Edwin K. P. Chong was funded by Colorado State University's Information Science & Technology Center; Ryan G. James and James P. Crutchfield were funded by Army Research Office grant W911NF-12-1-0234; Christopher J. Ellison was funded by a subaward from the Santa Fe Institute under a grant from the John Templeton Foundation.

Author Contributions

Each of the authors contributed to the design, analysis, and writing of the study.

Conflicts of Interest

The authors declare no conflicts of interest.

Appendix

By and large, most of these proofs follow directly from the lattice properties and also from the invariance and monotonicity properties with respect to \cong and \preceq .

A. Properties of I_λ^0

Lemma 1. $I_\lambda^0(X_1, \dots, X_n : Y)$ satisfies (GP), (Eq), (TM), (M₀), and (S₀), but not (LP₀).

Proof. (GP) follows immediately from the nonnegativity of the entropy. (Eq) follows from the invariance of entropy within the equivalence classes induced by \cong . (TM) follows from the monotonicity of the entropy with respect to \preceq . (M₀) also follows from the monotonicity of the entropy, but now applied to $\wedge_i X_i \wedge W \wedge Y \preceq \wedge_i X_i \wedge Y$. If there exists some j , such that $X_j \preceq W$, then generalized absorption says that $\wedge_i X_i \wedge W \wedge Y \cong \wedge_i X_i \wedge Y$, and thus, we have the equality condition. (S₀) is a consequence of the commutativity of the \wedge operator. To see that (LP₀) is not satisfied by the I_λ^0 , we point to the example, SUBTLE (Figure 4), which has negative synergy. One can also rewrite (LP₀) as the supermodularity law for common information, which is known to be false in general. (See [8], Section 5.4.) \square

Lemma 2. $I_\lambda^0(X_1, \dots, X_n : Y)$ satisfies (LB), (SR), and (Id).

Proof. For (LB), note that $Q \preceq X_1 \wedge \dots \wedge X_n$ for any Q obeying $Q \preceq X_i$ for $i = 1, \dots, n$. Then, apply the monotonicity of the entropy. (SR) is trivially true given Lemma 7 and the definition of zero-error information. Finally, (Id) follows from the absorption law and the invariance of the entropy. \square

Lemma 3. $I_\lambda^0(X_1, \dots, X_n : Y)$ satisfies (M₁) and (S₁), but not (LP₁).

Proof. (M₁) follows using the absorption and monotonicity of the entropy in nearly the same way that (M₀) does. (S₁) follows from commutativity, and (LP₁) is false, because (LP₀) is false. \square

B. Properties of I_λ

The proofs here are nearly identical to those used for I_λ^0 .

Lemma 4. $I_\lambda(X_1, \dots, X_n : Y)$ satisfies (GP), (Eq), (TM), (M₀), and (S₀), but not (LP₀).

Proof. (GP) follows from the nonnegativity of mutual information. (Eq) follows from the invariance of entropy. (TM) follows from the data processing inequality. (M₀) follows from applying the monotonicity of the mutual information $I(Y : \cdot)$ to $\wedge_i X_i \wedge W \preceq \wedge_i X_i$. If there exists some j , such that $X_j \preceq W$, then generalized absorption says that $\wedge_i X_i \wedge W \cong \wedge_i X_i$, and thus, we have the equality condition. (S₀) follows from commutativity, and a counterexample for (LP₀) is given by IMPERFECTRDN (Figure 3). \square

Lemma 5. $I_\lambda(X_1, \dots, X_n : Y)$ satisfies (LB) and (SR), but not (Id).

Proof. For (LB), note that $Q \preceq X_1 \wedge \dots \wedge X_n$ for any Q obeying $Q \preceq X_i$ for $i = 1, \dots, n$. Then, apply the monotonicity of the mutual information to $I(Y : \cdot)$. (SR) is trivially true given Lemma 8. Finally, (Id) does not hold, since $X \wedge Y \preceq X \vee Y$, and thus, $I_\lambda(X, Y : Y \wedge Y) = H(X \wedge Y)$. \square

Lemma 6. $I_\lambda(X_1, \dots, X_n : Y)$ does not satisfy (M₁), (S₁), or (LP₁).

Proof. (M₁) is false due to a counterexample provided by IMPERFECTRDN (Figure 3), where $I_\lambda(X_1 : Y) = 0.99$ bits and $I_\lambda(X_1, Y : Y) = 0$ bits. (S₁) is false, since $I_\lambda(X, X : Y) \neq I_\lambda(X, Y : X)$. Finally, (LP₁) is false, due to (LP₀) being false. \square

C. Miscellaneous Results

Lemma 7. Simplification of I_{\wedge}^0 .

$$\begin{aligned} I_{\wedge}^0(X_1, \dots, X_n : Y) &\equiv \max_{\Pr(Q|Y)} I^0(Q : Y) \quad \text{subject to } Q \preceq X_i \forall i \in \{1, \dots, n\} \\ &= H(X_1 \wedge \dots \wedge X_n \wedge Y) \end{aligned}$$

Proof. Recall that $I^0(Q : Y) \equiv H(Q \wedge Y)$, and note that $\wedge_i X_i$ is a valid choice for Q . By definition, $\wedge_i X_i$ is the richest possible Q , and so, monotonicity with respect to \preceq then guarantees that $H(\wedge_i X_i \wedge Y) \geq H(Q \wedge Y)$. \square

Lemma 8. Simplification of I_{\wedge} .

$$\begin{aligned} I_{\wedge}(X_1, \dots, X_n : Y) &\equiv \max_{\Pr(Q|Y)} I(Q : Y) \quad \text{subject to } Q \preceq X_i \forall i \in \{1, \dots, n\} \\ &= I(X_1 \wedge \dots \wedge X_n : Y) \end{aligned}$$

Proof. Note that $\wedge_i X_i$ is a valid choice for Q . By definition, $\wedge_i X_i$ is the richest possible Q , and so, monotonicity with respect to \preceq then guarantees that $I(Q : Y) \leq I(\wedge_i X_i : Y)$. \square

Lemma 9. $I_{\wedge}(X_1, \dots, X_n : Y) \leq I_{\min}(X_1, \dots, X_n : Y)$

Proof. We need only show that $I(\wedge_i X_i : Y) \leq I_{\min}(X_1, \dots, X_n : Y)$. This can be restated in terms of the specific information: $I(\wedge_i X_i : y) \leq \min_i I(X_i : y)$ for each y . Since the specific information increases monotonically on the lattice (cf. Section 2.2 or [8]), it follows that $I(\wedge_i X_i : y) \leq I(X_j : y)$ for any j . \square

References

1. Williams, P.L.; Beer, R.D. Nonnegative Decomposition of Multivariate Information. **2010**, arXiv:1004.2515 [cs.IT].
2. Griffith, V.; Koch, C. Quantifying Synergistic Mutual Information. In *Guided Self-Organization: Inception*; Prokopenko, M., Ed.; Springer: Berlin, Germany, 2014.
3. Harder, M.; Salge, C.; Polani, D. Bivariate measure of redundant information. *Phys. Rev. E* **2013**, *87*, 012130.
4. Bertschinger, N.; Rauh, J.; Olbrich, E.; Jost, J. Shared Information—New Insights and Problems in Decomposing Information in Complex Systems. In Proceedings of the European Conference on Complex Systems 2012, Brussels, Belgium, 3–7 September 2012; Gilbert, T., Kirkilionis, M., Nicolis, G., Eds.; Springer Proceedings in Complexity; Springer: Berlin, Germany, 2013; pp. 251–269.
5. Lizier, J.; Flecker, B.; Williams, P. Towards a synergy-based approach to measuring information modification. In Proceedings of the 2013 IEEE Symposium on Artificial Life (ALIFE), Singapore, 16–17 April 2013; pp. 43–51.

6. Gács, P.; Körner, J. Common information is far less than mutual information. *Probl. Control Inform. Theor.* **1973**, *2*, 149–162.
7. Wolf, S.; Wullschleger, J. Zero-error information and applications in cryptography. *Proc. IEEE Inform. Theor. Workshop* **2004**, *4*, 1–6.
8. Li, H.; Chong, E.K.P. On a Connection between Information and Group Lattices. *Entropy* **2011**, *13*, 683–708.
9. Wyner, A.D. The common information of two dependent random variables. *IEEE Trans. Inform. Theor.* **1975**, *21*, 163–179.
10. Cerf, N.J.; Adami, C. Negative Entropy and Information in Quantum Mechanics. *Phys. Rev. Lett.* **1997**, *79*, 5194–5197.
11. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*; John Wiley: New York, NY, USA, 1991.

© 2014 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).