The Information Decomposition and Information Delta: a Unified Approach to Disentangling Non-Pairwise Information

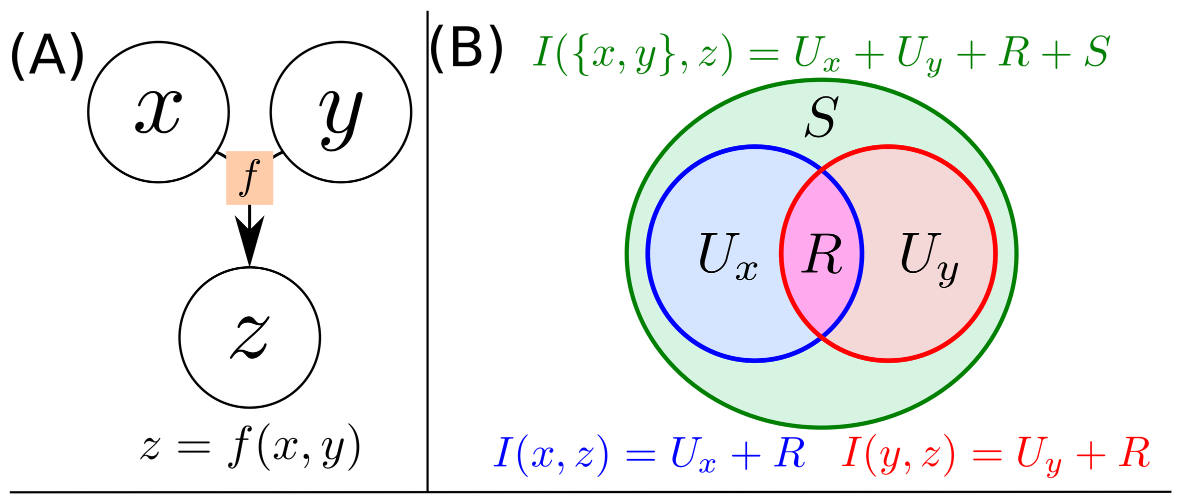
James M. Kunert-Graf [1] Nikita Sakhanenko [1] David J. Galas [1]

[1]Pacific Northwest Research Institute

Abstract

Neurons must integrate information from multiple inputs. Information theory provides robust measures of multivariable interdependence, but traditionally says little about the nature of the interaction. Popular in computational neuroscience, the Information Decomposition of Williams and Beer decomposes the mutual information into unique, redundant, and synergistic components [1, 2]. Unfortunately, there is no universally accepted method for its computation. Independently, the quantitative genetics community has developed the Information Delta measures for detecting non-pairwise interactions in genetic data [4, 5, 6]. This has been exhaustively characterized for discrete variables (common in genetics), yielding a geometric interpretation of how functions map onto delta-space. We show that the PID and Delta frameworks are largely equivalent, and results from each are relevant to open questions in the other. For example, we find that the PID approach of [3] addresses the problem of linkage disequilibrium in genetic data. We map the probability space defined by Bertschinger et al. onto a well-defined plane in the space of delta measures, and find that the corresponding optimization problem is solved by the previously-characterized function coordinates. This geometric mapping can thereby both side-step an expensive optimization and characterize the functional relationships between neurons.

Partial Information Decomposition



(C)

(i)

x	y	z
0	0	0
0	1	0
1	0	1
1	1	1

$z = x$ Fully Unique

(ii)

x	y	z
0	0	0
1	1	1

$z = x = y$ Fully Redundant

(iii)

x	y	z
0	0	0
0	1	1
1	0	1
1	1	0

$z = x \oplus y$ Fully Synergistic

Figure 1. **(A)** Let x, y, z be neurons with discrete states (e.g. resting, spiking). x and y determine z via function f. **(B)** The Information Decomposition (adapted from [2]) . The mutual information is decomposed into unique informations U_x and U_y, redundant information R, and synergistic information S. The system of equations is underdetermined. **(C)** Example functions: (i) For $z = x$, only x contains information about z (i.e. only U_x is nonzero). (ii) For $z = x = y$, the information encoded by x and y is fully redundant. (iii) when z is given by the XOR function, the encoded information is fully synergistic.

The partial information decomposition is defined by [1, 2]:

$$I(\{x, y\}, z) = U_x + U_y + R + S \quad (1)$$
$$I(x, z) = U_x + R \quad (2)$$
$$I(y, z) = U_y + R \quad (3)$$

Note that this set of equations is underdetermined. An additional definition (such as provided by [3]) is needed to compute these values.

Differential Interaction Information

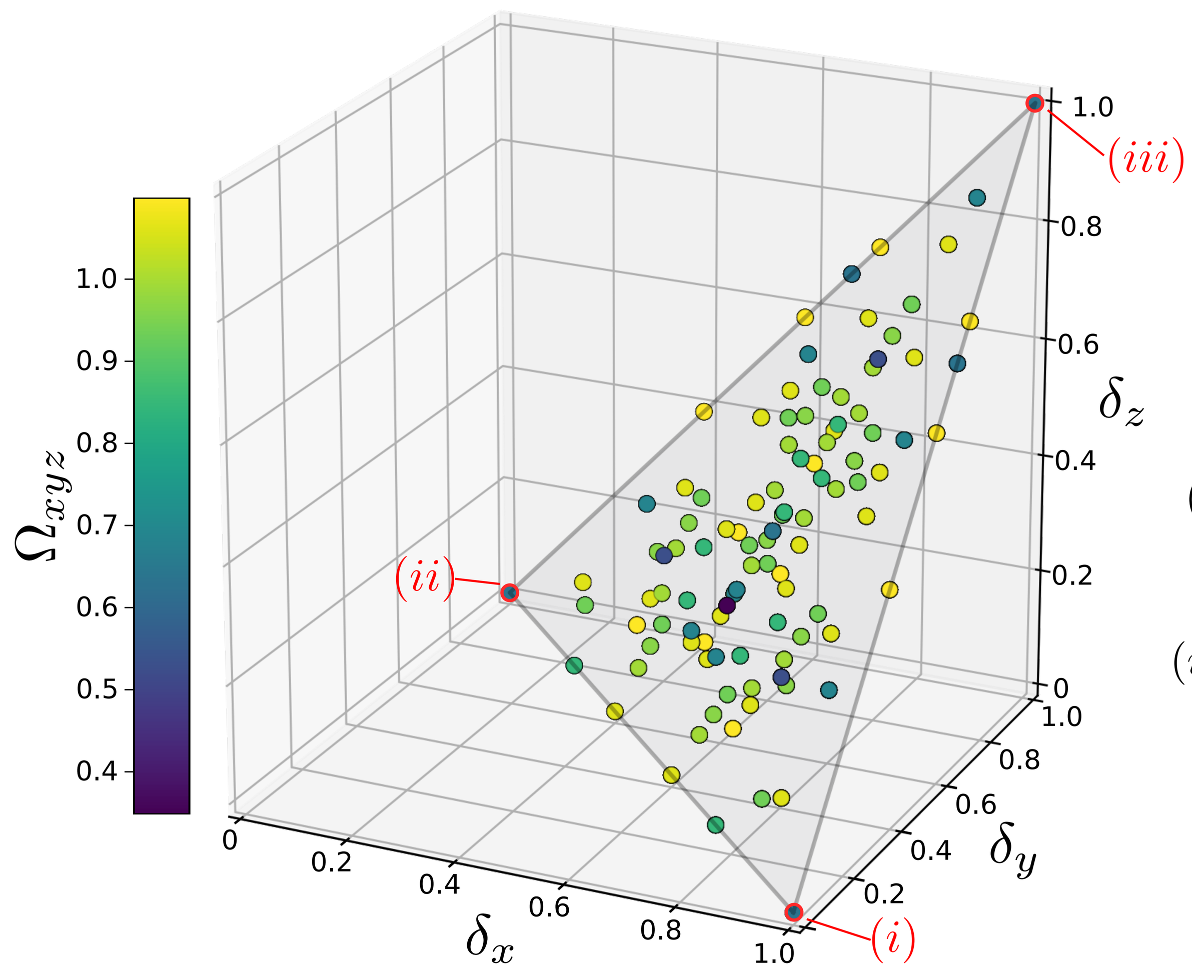
Consider a set of variables ν_n (i.e. $\nu_n = \{x, y, z\}$). The "information delta" is given by [4, 5]:

$$\Delta_z(\nu_n) = I(\nu_n) - I(\nu_n \setminus \{z\}) = -I(\nu_n \setminus \{z\}|z) \quad (4)$$

These measures can be normalized by the multi-information $\Omega_{xyz} = H_x + H_y + H_z - H_{xyz}$:

$$\delta_x = \Delta_x / \Omega_{xyz} \quad (5)$$

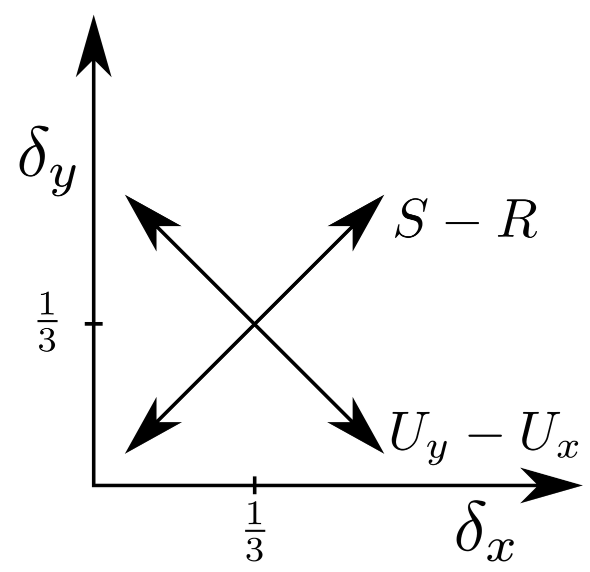
If $z = f(x, y)$ and x, y are i.i.d., then all functions $f(x, y)$ map onto a highly-structured plane[6]. The delta-coordinates of a function clearly encode similar information as the PID components.



	Delta Coordinate	Function
(i)	$\vec{\delta} = (1, 0, 0)$	$z = y$
(ii)	$\vec{\delta} = (0, 1, 0)$	$z = x$
(iii)	$\vec{\delta} = (1, 1, 1)$	$z = x \oplus y$

Figure 2. **(A)** A geometric interpretation of the Information Deltas, as developed in [6]. **(A)** Consider neurons with three discrete states (e.g. resting, spiking, refractory). There are 19,683 possible functions $f(x, y)$ which map onto 105 unique points within a highly-structured plane in δ-space. **(B)** Sample functions and their mappings onto δ-space. Functions with a full pairwise dependence on x or y map to oppose lower corners, whereas the fully-synergistic XOR is mapped to the uppermost corner.

Relationship between the Frameworks



We prove that the delta-coordinates and PID components are directly related as follows:

$$S - R = \frac{\Omega_{xyz}}{2}(\delta_x + \delta_y + \delta_z - 1) \quad (6)$$
$$U_x - U_y = \Omega_{xyz}(\delta_y - \delta_x) \quad (7)$$

The delta-space encodes the differences between PID components intuitively (functions in the upper corner being highly synergistic, with the lower corners being highly pairwise).

The optimization of Q in $\vec{\delta}$-space

One solution to the PID problem was provided by Bertschinger et al.[3], who defined:

$$U_x = \min_{q \in Q} I(y, z|x) = \max_{q \in Q} \Delta_x \quad (8)$$

This is well-motivated but requires an optimization over the set of probability distributions Q.

Let Ψ be the set of all joint probability distributions of x,y and z. Then we define Q as the set of all distributions with equal marginal probability distributions $p(x, z)$ and $p(y, z)$:

$$Q = \{q \in \Psi | q(x, z) = p(x, z) \wedge q(y, z) = p(y, z)\} \quad (9)$$

We map Q onto delta-space and find that it is constrained to a plane, and that the optimization is solved by the point where x, y are i.i.d. Coincidentally, these are the delta-coordinates which have been fully characterized by Sakhanenko et al.[6].

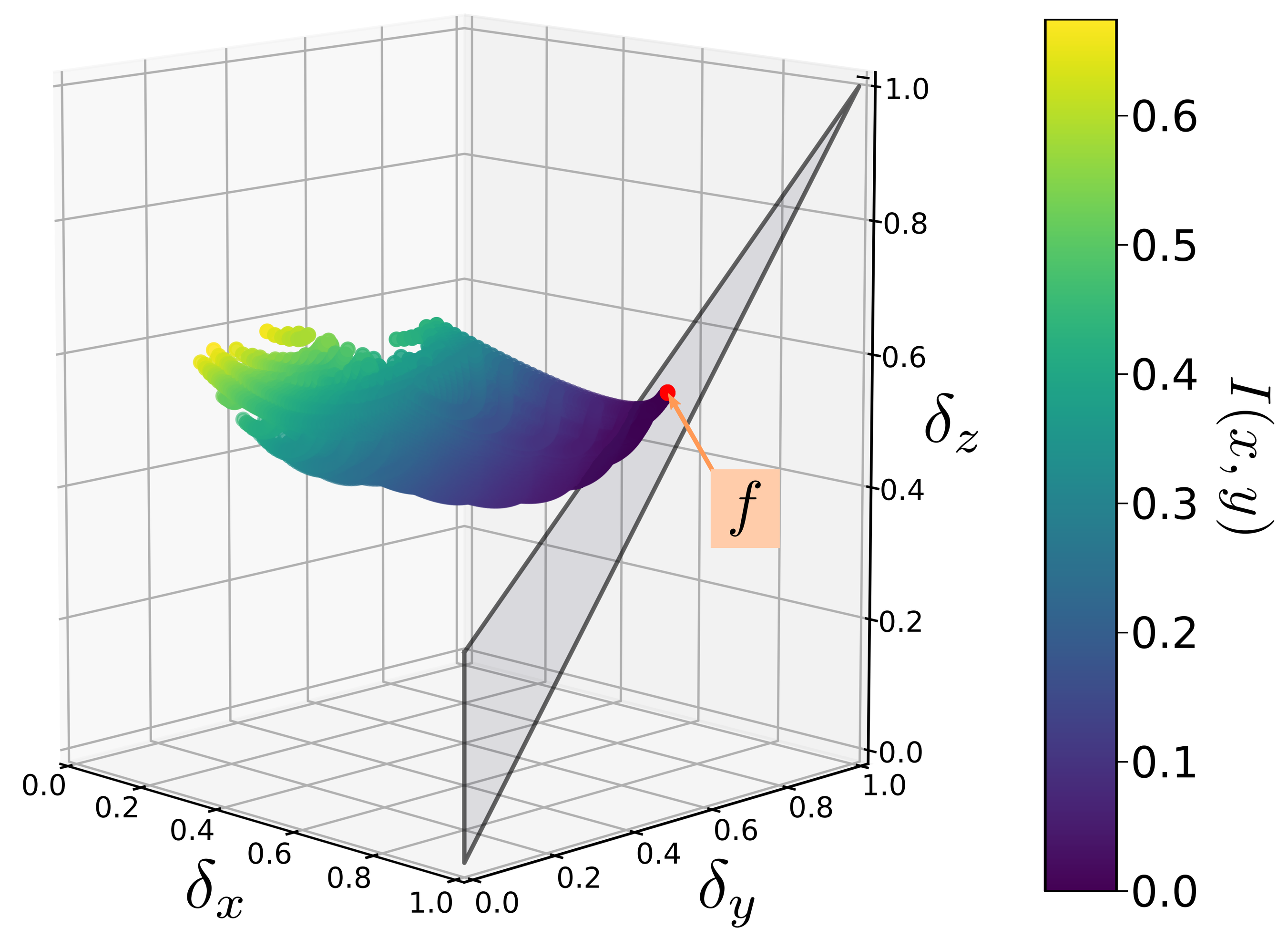
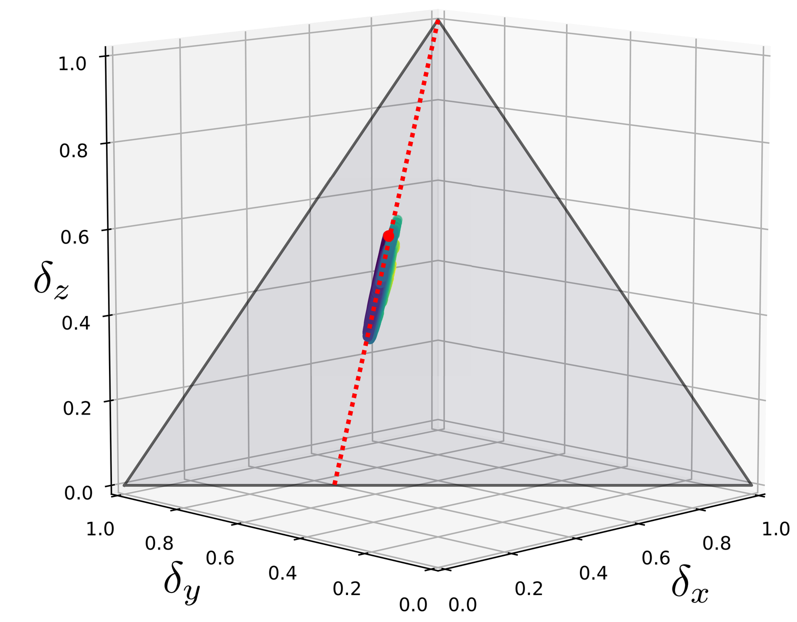


Figure 3. An example mapping of a Bertschinger set Q (defined in [3]) to δ-space. A set Q consists of all probability distributions $p(x, y, z)$ which share the same marginal distributions $p(x, z)$ and $p(y, z)$. Each Q maps onto a simple plane in δ-space, and the solution to the optimization problem in [3] (which solves the Information Decomposition) occurs at a point lying in the previously-characterized δ-plane defined in [6] (as plotted in Figure 2). If the experimental δ-coordinates were seen to lie in this subspace, we could immediately identify both the decomposition values U_x, U_y, R, S as well as the corresponding class of functions, without explicitly solving the optimization problem



In addition to developing a method to computationally perform this mapping, we prove the following analytically:

1. In any set Q of distributions with equal marginal distributions $p(x, z)$ and $p(y, z)$, the delta-coordinates $(\delta_x, \delta_y, \delta_z)$ will be restricted to a plane. This is true for any alphabet size.
2. For sets Q which include a function, the solution to the Bertschinger optimization occurs at the delta-coordinates of said function (as defined in [6]).

This suggests the following procedure for computing the PID:

1. Given data for a set of variables $\{x, y, z\}$, calculate the $\vec{\delta}$-coordinates.
2. Calculate the nearest Q-plane and its associated function class f.
3. Take the value of Δ_x for the function class f and compute the PID. Using the definition of [3], $U_x = \Delta_x|_f$ and therefore the computation is trivial.

This not only sidesteps the optimization, but additionally provides a candidate function class f which describes the exact functional relationship between the neurons.

References

[1] Paul L Williams and Randall D Beer. Nonnegative decomposition of multivariate information. *arXiv preprint arXiv:1004.2515*, 2010.

[2] Joseph T Lizier, Nils Bertschinger, Jürgen Jost, and Michael Wibral. Information decomposition of target effects from multi-source interactions: Perspectives on previous, current and future work. *Entropy*, 20(4):307, 2018.

[3] Nils Bertschinger, Johannes Rauh, Eckehard Olbrich, Jürgen Jost, and Nihat Ay. Quantifying unique information. *Entropy*, 16(4):2161–2183, 2014.

[4] David J Galas, Nikita A Sakhanenko, Alexander Skupin, and Tomasz Ignac. Describing the complexity of systems: Multivariable "set complexity" and the information basis of systems biology. *Journal of Computational Biology*, 21(2):118–140, 2014.

[5] Nikita A Sakhanenko and David J Galas. Biological data analysis as an information theory problem: Multivariable dependence measures and the shadows algorithm. *Journal of Computational Biology*, 22(11):1005–1024, 2015.

[6] NA Sakhanenko, JM Kunert-Graf, and DJ Galas. The information content of discrete functions and their application in genetic data analysis. *J. Comp. Biol.*, 24(12):1153–1178, 2017.