# Generation of Synergistic Random Variables

# THESIS

Submitted in partial fulfillment of the requirements of
BITS F422T Thesis
by

Yashee Sinha
ID No.: 2019B5A70652G

Under the supervision of
Dr. Rick Quax, University of Amsterdam

Birla Institute of Technology and Science, Pilani

16th May 2024

# Abstract

In the present literature, there is no good measure for synergistic information. With the increase in research interests in complex systems, it is highly relevant to find a new measure for synergy. One of the most promising frameworks proposed by Rick et al. in 2017 assumes the existence of random variables defined as *synergistic random variables* (SRVs). This report outlines the properties of such an SRV and proposes novel algorithms to generate arbitrary SRVs using analytical solutions of the equations. Moreover, it provides a framework to optimize the SRVs and outlines the several degrees of freedom that exists in the generation of such SRVs.

# Acknowledgements

I thank my advisor Dr. Rick Quax for helping and guiding me throughout the process of working on this project. I thank the open source contributors to the Python NumPy library which played a huge role in achieving the numerical results in this project. Finally, I thank my family for supporting me throughout this venture.

# Contents

**1 Introduction** **4**
1.1 Problem statement . . . . . 5

**2 Existence of Solutions** **8**
2.1 Consistency conditions . . . . . 9

**3 Algorithms** **10**
3.1 Numerical approach . . . . . 10
3.2 Analytical approach . . . . . 13
3.2.1 Constraint II . . . . . 13
3.2.2 Constraint I . . . . . 13
3.2.3 Encoding high mutual info . . . . . 14
3.3 Comparisions between the numerical and analytical approach . . . . . 15

**4 Discussion and Conclusions** **17**

# Chapter 1

# Introduction

Shanon's entropy gives a quantitative measure for the information contained in a specified random variable[1]. The information measures used in Shanon's framework obey the simple axioms of non-negativity, continuity, monotonicity, and additivity. The entropy function, then, is the average information:

$$H(X) = \sum_{x \in X} -p(X = x) \ln(p(X = x)) \tag{1.1}$$

Similar arguments lead us to the idea of mutual information measures which quantify the information shared between two or more variables. The mutual information defined between two sets of variables, for example, is defined as:

$$\begin{aligned} I(X : Y) &= \sum_{y \in Y} \sum_{x \in X} p(X = x, Y = y) \ln \frac{p(X = x, Y = y)}{p(X = x)p(Y = y)} && (1.2) \\ &= \sum_{y \in Y} -p(Y = y) \ln p(Y = y) && (1.3) \\ &\quad + \sum_{y \in Y} \sum_{x \in X} p(X = x, Y = y) \ln p(X = x|Y = y) && (1.4) \\ &= H(Y) - H(X|Y) && (1.5) \end{aligned}$$

Mutual information manages to capture the pairwise correlation that exists between the two sets. However, a generalization of this concept to higher number of sets is not available. As research interests in the field of complex systems keep rising, the search for such multivariate measures of mutual information become more relevant in studying complex dynamical systems ranging from molecular biology to social phenomena and even entanglement[2, 3].

One of the most important higher-order information measure is given by the synergistic information [4, 5, 6]. To explain what synergy means we will use the classic example of an XOR gate described in Table 1.1. The idea of synergy encapsulates the fact that simultaneously both the inputs $X$ and $Y$ fully determine $S$, however individually observing any of those variables does not improve our prediction of $S$. In terms of mutual information we can equivalently say:

$$I(S : X) = 0 \tag{1.6}$$

$$I(S : Y) = 0 \tag{1.7}$$

$$I(S : X, Y) = 1 \tag{1.8}$$

This describes what synergistic information between three variables looks like and can be easily generalized to a set of $N$ random variables, $X = \{X_1, X_2, \dots X_N\}$. There are numerous measures for synergistic information proposed in the literature today. The most intuitive method given by Whole Minus Sum (WMS) simply subtracts the sum of pairwise ('individual') mutual information quantities from the total mutual information, i.e., $I(S : X) - \sum_i I(S : X_i)$[7]. However, the inputs could themselves depend on each other which produces negative measures of information in this framework which is undesirable. On the other hand, one of the most promising systematic approach to quantify synergistic information is through the Partial Information Decomposition method [5]. In this framework, one defines in addition the concept of an intersection information or the redundant information between two variables. Then, $I(S : X) = \text{Synergistic Info}(S : X) + \sum_i \text{Unique Info}(X_i) - \text{Redundant Info}(X)$. However, a good measure for this intersection information is still elusive.

| X | Y | S = X ⊕ Y |
|---|---|---|
| 0 | 0 | 0 |
| 0 | 1 | 1 |
| 1 | 0 | 1 |
| 1 | 1 | 0 |

Table 1.1: Truth table for XOR gate

Rick et al. proposed a new framework for quantifying the synergistic information that attempts to find an orthogonal decomposition of the set of random variables using certain synergistic random variables[2]. The authors define a *synergistic random variable*(SRV), $S$, associated to random variables $X \equiv \{X_i\}_i$ as:

$$I(S : X) > 0, \tag{1.9}$$

$$\forall i, I(S : X_i) = 0 \tag{1.10}$$

Our aim in this report is to find a method to generate arbitrary SRVs for any given inputs. We will also only deal with the cases where the inputs are independent of each other. As we will discuss in the later sections, we would also like the mutual information, $I(S : X)$, to be high.

## 1.1 Problem statement

Let us properly formulate a specific instance of the problem statement that we will focus on. There are two random variables, $X_1$ and $X_2$, which will be referred to as the inputs. Each of these random variables has three possible states in their individual sample space, such that $p(X_1 = i) = p_i$ and $p(X_2 = j) = q_j$. The conditional probability distributions of the synergistic random variable is denoted by the convention used in Table 1.2. Specifically, $p(S = k|X_1 = i, X_2 = j) = S^k_{ij}$.

To satisfy Eq. (1.10), we must find appropriate $S^k_{ij}$ that satisfy:

$$p(S = k|X_1 = i) = p(S = k|X_2 = j) \tag{1.11}$$

$$= p(S = k) \quad \forall i, j, k \tag{1.12}$$

| $X_1 \backslash X_2$ | $q_1$ | $q_2$ | $q_3$ |
|---|---|---|---|
| $p_1$ | $(S_{11}^1, S_{11}^2, S_{11}^3)$ | $(S_{12}^1, S_{12}^2, S_{12}^3)$ | $(S_{13}^1, S_{13}^2, S_{13}^3)$ |
| $p_2$ | $(S_{21}^1, S_{21}^2, S_{21}^3)$ | $(S_{22}^1, S_{22}^2, S_{22}^3)$ | $(S_{23}^1, S_{23}^2, S_{23}^3)$ |
| $p_3$ | $(S_{31}^1, S_{31}^2, S_{31}^3)$ | $(S_{32}^1, S_{32}^2, S_{32}^3)$ | $(S_{33}^1, S_{33}^2, S_{33}^3)$ |

Table 1.2: Conditional probabilities of a random variable $S$. Here: (a) $p(X_1 = i) = p_i$, (b) $p(X_2 = j) = q_j$, and (c) $p(S = k|X_1 = i, X_2 = j) = S_{ij}^k$

A trivial solution is simply $S_{ij}^k = \frac{1}{3}\ \forall\ i, j, k$. However, the $I(S : X) = 0$ for this trivial case and that violates the Eq. (1.9). In general, we would like to also be able to generate arbitrary number of SRVs.

To solve this problem of finding $S_{ij}^k$s we shift our perspective and reformulate the problem in the terms of matrices. Let us define a matrix, $\mathbf{S}^{(k)}$, whose elements are given by:

$$[\mathbf{S}^{(k)}]_{ij} = S_{ij}^k \tag{1.13}$$

If we denote the probabilities of the input as column vectors given by $\boldsymbol{p} = \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix}$ and $\boldsymbol{q} = \begin{bmatrix} q_1 \\ q_2 \\ q_3 \end{bmatrix}$, the probability of the marginal of $S$ as $p(S = k) = [\boldsymbol{b}]_k = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}_k$, and denote the vector full of ones to be $\mathbf{1}_{3\times 1} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}_{3\times 1}$, then we can reformulate our problem given in Eq.(1.11) as:

$$\boldsymbol{p}^T \mathbf{S}^{(k)} = b_k \mathbf{1}^T \tag{1.14}$$

$$\mathbf{S}^{(k)} \boldsymbol{q} = b_k \mathbf{1} \tag{1.15}$$

This is possible because the inputs are independent of each other. The SRV must also simultaneously obey the following constraints since each of these denote a probability tuple:

**a**. $0 \leq S_{ij}^{(k)} \leq 1 \quad \forall i, j, k$

**b**. $\sum_k S_{ij}^{(k)} = 1 \quad \forall i, j$

In other words, there are two specific properties that the probability distribution of any SRV must satisfy given by Fig. 1.1.

Note that here we are left with two choices for any algorithm that generates an SRV. One can either satisfy the condition **a** in Fig. 1.1 and thereafter adjust till the conditions of **b** are met. Or alternatively, one can satisfy **a** first and proceed on finding the appropriate SRVs. In this report we will detail two different algorithms that produce solutions to the SRV. We find that the former method provides a more analytical view of SRV generation while the latter method is more numerical.

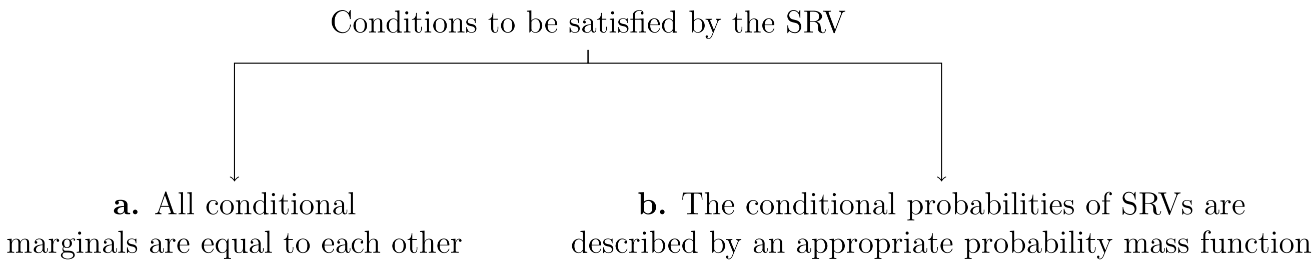


Figure 1.1: The properties that the probability distribution of all SRVs must satisfy.

# Chapter 2

# Existence of Solutions

Before we attempt to solve Eq. (1.15), we must transform the system of equations into the usual $\mathbf{A}\boldsymbol{x} = \boldsymbol{b}$ form. We will utilize a vectorization operator given by[8]:

$$\mathbf{vec}(\mathbf{A}) = [a_{11}, \ldots a_{m1}, a_{12}, \ldots a_{m2} \ldots a_{1n} \ldots a_{mn}]^T \tag{2.1}$$

where $a_{ij}$ are the $i,j$-th elements of the matrix $\mathbf{A}$. For a vector $\mathbf{A}$ of size $k \times l$, and matrix $\mathbf{B}$ of size $l \times m$, this operator holds the following properties:

$$\mathbf{vec}(\mathbf{AB}) = (\mathbb{1}_m \otimes A)\mathbf{vec}(B) \tag{2.2}$$
$$= (B^T \otimes \mathbb{1}_k)\mathbf{vec}(A) \tag{2.3}$$

where $\mathbb{1}_m$ denotes the identity matrix of size $m \times m$. On applying this vectorization operator on both sides of Eq. (1.14), we get:

$$\mathbf{vec}(\boldsymbol{p}^T\mathbf{S}^{(k)}) = \mathbf{vec}(b_k\mathbf{1}^T) \tag{2.4}$$
$$\implies (\mathbb{1}_3 \otimes \boldsymbol{p}^T)\mathbf{vec}(\mathbf{S}^{(k)}) = b_k\mathbf{1} \tag{2.5}$$

Similarly for Eq. (1.15),

$$\mathbf{vec}(\mathbf{S}^{(k)}\boldsymbol{q}) = \mathbf{vec}(b_k\mathbf{1}) \tag{2.6}$$
$$\implies (\boldsymbol{q}^T \otimes \mathbb{1}_3)\mathbf{vec}(\mathbf{S}^{(k)}) = b_k\mathbf{1} \tag{2.7}$$

Since this denotes a rectangular matrix multiplication anyway, we can freely stack the two equations above vertically to produce:

$$\begin{pmatrix} \boldsymbol{q}^T \otimes \mathbb{1} \\ \mathbb{1} \otimes \boldsymbol{p}^T \end{pmatrix}_{6\times 9} \mathbf{vec}(\mathbf{S}^{(k)})_{9\times 1} = b_k\mathbf{1}_{6\times 1} \tag{2.8}$$
$$\implies \mathbf{M}\boldsymbol{v}^{(k)} = b_k\mathbf{1} \tag{2.9}$$
$$\implies \mathbf{M}\boldsymbol{v}^{(k)} = \mathbf{1}_{b_k} \tag{2.10}$$

where we have defined $\mathbf{1}_{b_k} = b_k\mathbf{1}$ for our convenience. $\mathbf{M}$ will be referred to as the master matrix of this problem. And our goal is to find vectors, $\boldsymbol{v}^{(k)}$, that satisfy the Eq. (2.10) along with the constraints:

1. $0 \leq v_i^{(k)} \leq 1$
2. $\sum_k \boldsymbol{v}^{(k)} = \mathbf{1}$

## 2.1 Consistency conditions

We must first analyze if the Eq. (2.10) is consistent and has any solutions at all. We start by a simple gaussian elimination technique:

$$\mathbf{M} = \left(\begin{array}{ccccccccc|c} q_1 & 0 & 0 & q_2 & 0 & 0 & q_3 & 0 & 0 & b_k \\ 0 & q_1 & 0 & 0 & q_2 & 0 & 0 & q_3 & 0 & b_k \\ 0 & 0 & q_1 & 0 & 0 & q_2 & 0 & 0 & q_3 & b_k \\ p_1 & p_2 & p_3 & 0 & 0 & 0 & 0 & 0 & 0 & b_k \\ 0 & 0 & 0 & p_1 & p_2 & p_3 & 0 & 0 & 0 & b_k \\ 0 & 0 & 0 & 0 & 0 & 0 & p_1 & p_2 & p_3 & b_k \end{array}\right) \tag{2.11}$$

$$\xrightarrow[\text{Reduction}]{\text{Row}} \left(\begin{array}{ccccccccc|c} q_1 & 0 & 0 & q_2 & 0 & 0 & q_3 & 0 & 0 & b_k \\ 0 & q_1 & 0 & 0 & q_2 & 0 & 0 & q_3 & 0 & b_k \\ 0 & 0 & q_1 & 0 & 0 & q_2 & 0 & 0 & q_3 & b_k \\ 0 & 0 & 0 & -\frac{p_1 q_2}{q_1} & -\frac{p_2 q_2}{q_1} & -\frac{p_3 q_2}{q_1} & -\frac{p_3 q_3}{q_1} & -\frac{p_3 q_3}{q_1} & -\frac{p_3 q_3}{q_1} & (1 - \frac{p_1+p_2+p_3}{q_1}) b_k \\ 0 & 0 & 0 & 0 & 0 & 0 & -\frac{p_3 q_3}{q_2} & -\frac{p_3 q_3}{q_2} & -\frac{p_3 q_3}{q_2} & \frac{q_1+q_2-(p_1+p_2+p_3)}{q_2} b_k \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{q_1+q_2+q_3-(p_1+p_2+p_3)}{q_3} b_k \end{array}\right) \tag{2.12}$$

Hence, this equation is consistent **if and only if** $\sum_k q_k = \sum_k p_k$. Since the vectors $\boldsymbol{p}$ and $\boldsymbol{q}$ represent probabilities, we know that: $\sum_k q_k = \sum_k p_k = 1$. In addition, we also notice that the master-matrix, $\mathbf{M}$, is not full rank. So we conclude that, in general, there exist infinite solutions for Eq. (2.10).

Note, however, that we have assumed here that at least two elements of the vector $\boldsymbol{q}$ are non-zero. We will continue making this assumption since we are interested in random variables with 2 or more states. The case where $X_2$ has only 1 state is trivial and can be solved without the need of these methods. Moreover, Sec. 3.2 describes a more general method that no longer needs to make these assumptions.

# Chapter 3

# Algorithms

## 3.1 Numerical approach

Using, Eq. (2.12), the solution to the master matrix equation is then given by the Table 3.1 when transformed back to the matrix form. where $r_k, s_k, t_k, v_k$ are free

| $X_1 \backslash X_2$ | $q_1$ | $q_2$ | $q_3$ |
|---|---|---|---|
| $p_1$ | $\frac{1}{p_1 q_1}\begin{pmatrix} b_k(p_1+q_1-1) \\ +q_2(p_2 r_k + p_3 s_k) \\ +q_3(p_2 t_k + p_3 v_k) \end{pmatrix}$ | $\frac{1}{p_1}(b_k - p_2 r_k - p_3 s_k)$ | $\frac{1}{p_1}(b_k - p_2 t_k - p_3 v_k)$ |
| $p_2$ | $\frac{1}{q_1}(b_k - q_2 r_k - q_3 t_k)$ | $r_k$ | $t_k$ |
| $p_3$ | $\frac{1}{q_1}(b_k - q_2 s_k - q_3 v_k)$ | $s_k$ | $v_k$ |

Table 3.1: $\mathbf{S}^{(k)}$; the $k$-th elements of the SRV, $S$.

parameters and must be chosen in a way such that the constraints below Eq. (2.10) are satisfied.

There are 4 free parameters, $(r_k, s_k, t_k, v_k)$ for each $k \in \{1, 2\}$ in $\mathbf{S}^{(k)}$. (Note, $\mathbf{S}^{(3)}$ is entirely determined by $1 - \mathbf{S}^{(1)} - \mathbf{S}^{(2)}$). In addition, we are also free to choose the marginal probability vector, $\boldsymbol{b}$, of $S$, which provides us with two extra free parameters. Hence, any generation of SRV has 10 degrees of freedom.

Before we detail the algorithm, we must make note of one final fact. We aim to find the highest synergistic information between the SRV, $S$, and the inputs, $X_1$ and $X_2$. This implies that the conditional probabilities should ideally be close to one of the unit vectors given by $\{(0,0,1), (0,1,0), (1,0,0)\}$. One such solution is given by Table 3.2.

Obviously, such a random variable is an SRV only if $\boldsymbol{q}^T = \boldsymbol{p}^T = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. However, we will utilize this fact to choose the first few free parameters in our algorithm.

| $X_1 \backslash X_2$ | $q_1$ | $q_2$ | $q_3$ |
|---|---|---|---|
| $p_1$ | $(0,0,1)$ | $(1,0,0)$ | $(0,1,0)$ |
| $p_2$ | $(0,1,0)$ | $(0,0,1)$ | $(1,0,0)$ |
| $p_3$ | $(1,0,0)$ | $(0,1,0)$ | $(0,0,1)$ |

Table 3.2: An example of a random variable with the highest possible mutual info. Each cell represents the conditional probability: $p(S = k, X_1 = i, X_2 = j)$.

We will expect $(r, s, t, v) \in Q = \{(1, 0, 0, 0), (0, 1, 1, 0), (0, 0, 0, 1), \ldots\}$. Using this, the numerical algorithm is described in Algorithm 1.

Due to step 21, this method is not guaranteed to provide a synergistic random variable and can produce SRVs such that $I(S : X_i) \neq 0$. But as we will see in Sec. 3.3, it works very well for most cases.

**Algorithm 1** A numerical algorithm to produce SRV

1: Choose two vectors from $Q = \{(1,0,0,0),(0,1,1,0),(0,0,0,1),\dots\}$, that denote $(r_1,s_1,t_1,v_1),(r_2,s_2,t_2,v_2)$.
2: Calculate $\boldsymbol{l} = (S^1_{31}, S^2_{31}, 1 - S^1_{31} - S^2_{31})$ using Table 3.1.
3: Choose $(b_1, b_2, 1 - b_1 - b_2)$ such that $0 \le l_i \le 1$, $\forall i$.
4: **for** $(i,j) \in \{(1,1),(1,2),(1,3),(2,1)\}$ **do**
5:     Calculate $\boldsymbol{l} = (S^1_{ij}, S^2_{ij}, 1 - S^1_{ij} - S^2_{ij})$ using Table 3.1.
6:     **if** not $(\forall k, 0 \le l_k \le 1)$ **then**
7:         **if** $\exists$ non-zero free parameters in $i$-th row **then**
8:             **while** $l_k < 0$ **do**
9:                 Let the lowest valued free parameter be $S^k_{ij'}$.
10:                 $S^k_{ij'} = \max\{S^k_{ij'} + l_k \cdot \dfrac{p_j}{p_{j'}}, 0\}$
11:                 Re-evaluate $l_k$
12:             **end while**
13:         **else if** $\exists$ non-zero free parameters in $j$-th column **then**
14:             **while** $l_k < 0$ **do**
15:                 Let the lowest valued free parameter be $S^k_{i'j}$.
16:                 $S^k_{i'j} = \max\{S^k_{i'j} + l_k \cdot \dfrac{q_i}{q_{i'}}, 0\}$
17:                 Re-evaluate $l_k$
18:             **end while**
19:         **else**
20:             Let $S^k_{mn}$ be the lowest non-zero value in the $i$th row or $j$-th column.
21:             $S^k_{mn} = \begin{cases} S^k_{mn} + l_k \cdot \frac{q_j}{q_m} & \text{if } n = j \\ S^k_{mn} + l_k \cdot \frac{p_i}{p_n} & \text{if } m = i \end{cases}$
22:             Set $(i,j) = (m,n)$.
23:             Go to step 5
24:         **end if**
25:     **end if**
26: **end for**

## 3.2 Analytical approach

Eq. (2.10) is a very well known problem of finding solutions to simultaneous linear equations. However, we must keep in mind the fact that $\mathbf{M}$ is a rectangular matrix and hence doesn't have a matrix inverse. Instead, we must use the Moore-Penrose inverse of the matrix[9].

The Moore-Penrose inverse (or equivalently, the pseudo-inverse) of a matrix $\mathbf{M}$ is denoted by $\mathbf{M}^g$, and has the properties:

1. $\mathbf{M}\mathbf{M}^g\mathbf{M} = \mathbf{M}$
2. $\mathbf{M}^g\mathbf{M}\mathbf{M}^g = \mathbf{M}^g$

It is also known that for a given system, $\mathbf{M}\boldsymbol{v} = \boldsymbol{c}$, has a solution iff $\mathbf{M}\mathbf{M}^g\boldsymbol{c} = \boldsymbol{c}$ and the solution is given by the form[10]:

$$\boldsymbol{v} = \mathbf{M}^g\boldsymbol{c} + (\mathbb{1} - \mathbf{M}^g\mathbf{M})\,\boldsymbol{w} \tag{3.1}$$

where $\boldsymbol{w}$ is an arbitrary vector.

Since we have already proved that Eq. (2.10) is consistent in Sec. (2.1), we know that a solution must exist. Hence, the solution must be of the form:

$$\boldsymbol{v}^{(k)} = \mathbf{M}^g\mathbf{1}_{b_k} + (\mathbb{1} - \mathbf{M}^g\mathbf{M})\,\boldsymbol{w} \tag{3.2}$$

where $k \in \{1, 2, 3\}$.

However we also must make sure that each tuple $(v_i^{(1)}, v_i^{(2)}, v_i^{(3)})$ is a set of probability, i.e., it must satisfy the constraints detailed below Eq. (2.10).

### 3.2.1 Constraint II

For the second constraint, we note that if:

$$\begin{aligned}
\boldsymbol{v}^{(3)} &= \mathbf{1} - \boldsymbol{v}^{(1)} - \boldsymbol{v}^{(2)} && (3.3)\\
\implies \mathbf{M}\boldsymbol{v}^{(3)} &= \mathbf{M}\mathbf{1} - \mathbf{M}\boldsymbol{v}^{(1)} - \mathbf{M}\boldsymbol{v}^{(2)} && (3.4)\\
\implies \mathbf{M}\boldsymbol{v}^{(3)} &= \mathbf{1} - b_1\mathbf{1} - b_2\mathbf{1} && (3.5)\\
&= (1 - b_1 - b_2)\mathbf{1} && (3.6)\\
&= b_3\mathbf{1} && (3.7)
\end{aligned}$$

where we have used the fact that $\mathbf{M}\mathbf{1} = \mathbf{1}$ by simple substitution. Hence, $\boldsymbol{v}^{(3)} = \mathbf{1} - \boldsymbol{v}^{(1)} - \boldsymbol{v}^{(2)}$ is a valid solution to the master equation. And hence, the second constraint is satisfied.

### 3.2.2 Constraint I

To satisfy the first constraint ($0 \leq v_i^{(k)} \leq 1$) we need to identify valid operations on the $\boldsymbol{v}^{(k)}$ that preserves the property of Eq. (2.10). Since, this is a linear system, the following holds:

1. If $\boldsymbol{v}^{(k)}$ is a solution of Eq. (2.10) for certain $b_k$, then $\mathbf{M}\boldsymbol{v}^{(k')} = \mathbf{M}(\boldsymbol{v}^{(k)} - a) = (b_k - a)\mathbf{1}$.
2. If $\boldsymbol{v}^{(k)}$ is a solution of Eq. (2.10) for certain $b_k$, then $\mathbf{M}\boldsymbol{v}^{(k')} = \dfrac{1}{d}\mathbf{M}(\boldsymbol{v}^{(k)}) = \dfrac{b_k}{d}\mathbf{1}$.

Hence, given any arbitrary solution from Eq. (3.2), one can always find an appropriate value of $d$ and $a$ such that $\boldsymbol{v}^{(1)}, \boldsymbol{v}^{(2)}, \boldsymbol{v}^{(3)}$ all obey the first constraint.

### 3.2.3 Encoding high mutual info

As discussed in Sec. 3.1, we aim to generate SRVs with high mutual information. Ideally, such an SRV will look like Table 3.2.

From Sec. 3.1 we also know that this problem has 10 degrees of freedom. For each $k \in \{1, 2\}$, there exist 5 free parameters. If $\mathbf{A}[:, U]$ denotes the matrix formed with taking the columns of $\mathbf{A}$ enumerated in the set $U$, then it is found that the matrix given by $\mathbf{M}[:, [1, 3, 4, 5, 7]]$ has full rank. We will not provide the proof here but the results found from running Algorithm 2 validates this fact. This implies that if $\mathbf{M}'$ is defined as:

$$[\mathbf{M}']_{ij} = \begin{cases} [\mathbf{M}]_{ij} & \text{if } j \in \{1, 3, 4, 5, 7\} \\ 0 & \text{if } j \in \{2, 6, 8, 9\} \end{cases} \tag{3.8}$$

then $\boldsymbol{v}^{(k)}$ is a valid solution and $[\boldsymbol{v}^{(k)\prime}]_{2,6,8,9} = [\mathbf{M}'^{g}\mathbf{1}_{b_k}]_{2,6,8,9} = 0$.

In addition, if $b_k = q_m$, then

$$[\mathbf{M}']_{ij} = \begin{cases} 0 & \text{if } j \in \{2, 6, 7, 8, 9\} \\ [\mathbf{M}]_{ij} & \text{otherwise} \end{cases} \tag{3.9}$$

$$[\boldsymbol{v}^{(k)\prime}]_i = \begin{cases} \mathbf{M}'^{g}(\mathbf{1}_{p_m} - \mathbf{M}[:, 7 + m - 1]) & \text{if } i \neq 7 + m - 1 \\ 1 & \text{if } i = 7 + m - 1 \end{cases} \tag{3.10}$$

is also a valid solution.

In doing so we have also set the last three elements of vector $\boldsymbol{v}^{(k)}$ to be one of the unit vectors $\in \{(0, 0, 1), (0, 1, 0), (1, 0, 0)\}$ which corresponds to the ideal case in Table 3.2.

Moreover, we can also project the vectors given in the Table 3.2, i.e.,

$$R = \{(1, 0, 0, 0, 1, 0, 0, 0, 1), (0, 0, 1, 1, 0, 0, 0, 1, 0), \dots\}$$

using Eq. (3.2) by:

$$\boldsymbol{v}^{(k)\prime} = \boldsymbol{v}^{(k)\prime} + (\mathbb{1} - \mathbf{M}^{g}\mathbf{M})[\boldsymbol{w} - \boldsymbol{v}^{(k)\prime}] \tag{3.11}$$

where $\boldsymbol{w} \in R$. This is usually a good choice of projection and helps to increase the mutual information in the SRV.

Hence the algorithm is finally given by Algorithm 2.

**Algorithm 2** An analytical algorithm to produce SRV

Set **sum** $= (0, 0, \dots)^T$.
**for** $k = 1 \to 3$ **do**
    Evaluate the master matrix, $\mathbf{M}$, given by Eq. (2.10).
    Evaluate the matrix, $\mathbf{M}'$, using Eq. (3.9).
    Evaluate the solution, $\boldsymbol{v}^{(k)\prime}$, using Eq.(3.10) with $m = k$.
    Choose a $\boldsymbol{w} \in R$.
    Project $\boldsymbol{w}$ on to the kernel using Eq. (3.11) to update $\boldsymbol{v}^{(k)\prime}$.
    Choose $a, d$ such that $0 \leq \dfrac{([\boldsymbol{v}^{(k)\prime}]_i - a)}{d} \leq 1 - [\mathbf{sum}]_i, \forall i$
    $[\boldsymbol{v}^{(k)\prime}]_i = \dfrac{([\boldsymbol{v}^{(k)\prime}]_i - a)}{d}, \forall i$
    $\mathbf{sum} = \mathbf{sum} + \boldsymbol{v}^{(k)\prime}$
**end for**

## 3.3 Comparisions between the numerical and analytical approach

We produced 50 samples of uniformly distributed input probabilities, $\boldsymbol{p}$, and $\boldsymbol{q}$. We then ran the two algorithms provided in this section for each of the samples. The plot of the mutual information given by $I(S : X)$ is given in Fig. 3.1. Since, the numerical approach is not guaranteed to produce an SRV in all cases, the absolute sum of all the individual mutual informations given by $\sum_{i=1}^{2} |I(S : X_i)|$ is also plotted alongside them.

It is easily visible from the Fig. 3.1 that the Algorithm 1 tends to provide SRVs with higher mutual information. However, the algorithm can fail to produce a synergistic random variable in many cases. On the other hand, Algorithm 2 provides SRVs that have relatively lower mutual information, but is also a reliable method that is guaranteed to produce an SRV.

Moreover, the choice of $\boldsymbol{w}, a$ and $d$ in the Algorithm 2 is undetermined and provides us with a way to optimize the solution further whereas Algorithm 1 only provides one solution. Importantly, we also know that all solutions of the Eq. (1.11) is given by Eq. (3.2) and hence any SRV produced by the Algorithm 1 can be found using Algorithm 2 through an appropriate choice of $\boldsymbol{w}$ even though the opposite is not true.

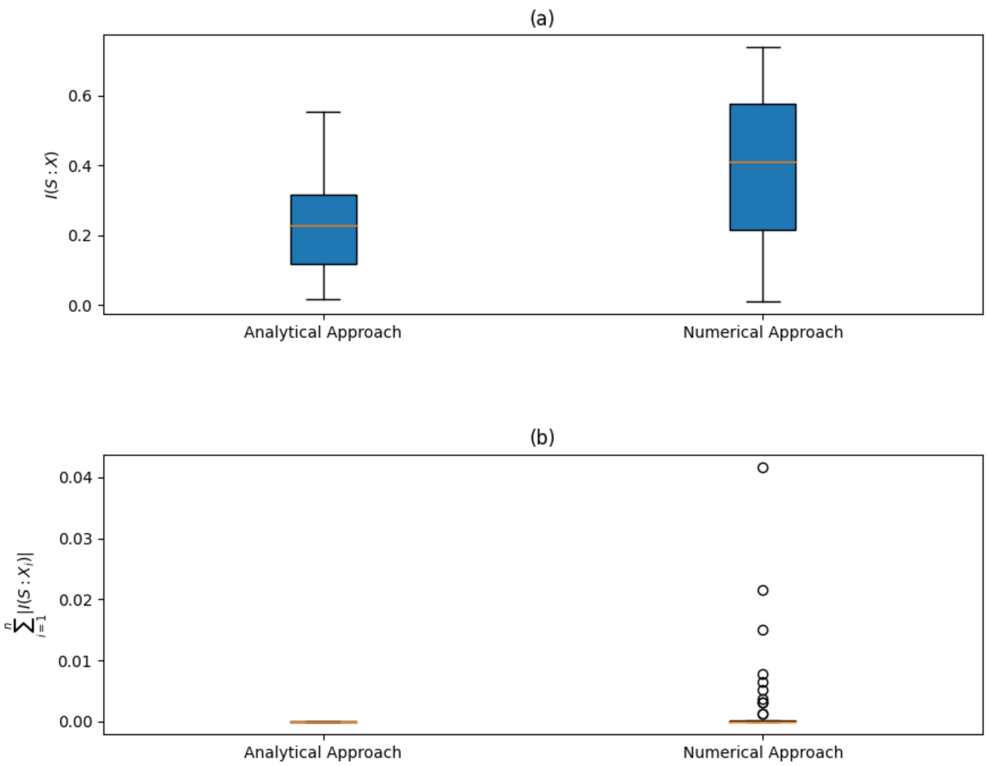


Figure 3.1: Performance of Algorithm 1 vs Algorithm 2; (a) The mutual information of SRVs produced by the two algorithms for 50 random vectors, $\boldsymbol{p}, \boldsymbol{q}$, sampled from a uniform distribution. (b) The absolute sum of all individual information given by $\sum_i |I(S : X_i)|$ where $S$ denotes the correspoding SRVs.

# Chapter 4

# Discussion and Conclusions

The algorithms outlined in this report can analytically generate arbitrary SRVs as required for any given independent inputs $X_1$ and $X_2$. We find that the two approaches have a trade-off between them. Whereas the numerical approach generates SRVs with higher mutual information but can also fail to produce a truly synergistic variable, the analytical approach always generates an SRV but they tend to have a lower mutual information than the former. In numerous applications of synergy, where one calculates measures like Whole Minus Sum, it might be more important to produce true SRVs and hence the analytical approach could be more desirable. On the other hand SRVs with higher mutual information could be more beneficial in the orthogonalization process proposed by Rick et al. Even such, the analytical solution to this problem is complete, and hence includes the results produced by Algorithm 1 too in an appropriate choice for the arbitrary vector $\boldsymbol{w}$ in Eq. (3.2). In the future, it could be possible to determine the appropriate $\boldsymbol{w}$ that maximizes the mutual information in the SRV.

Even though this report only deals with random variables with three states, it is easily generalizable to any number of states since the matrix equations don't change. Moreover, a tensor description of the problem statement can help further generalize the solution for more number of inputs. It would be possible to use higher order Singular Value Decomposition of matrices to calculate pseudo-inverses for such tensor equations. Moreover, the analytical approach can also be used to calculate bounds to the synergistic information in terms of the inputs and the marginal probabilities used. It can also prove to be useful in calculating the gradient of the mutual information in the future. Further research is required to find the constraints that the marginal probabilities must follow to produce a better SRV in Algorithm 1. One hopes to find a way to use both the analytical and numerical approach to find the most optimal SRV for any given input distributions.

# Bibliography


[1] Shannon C.E. and W. Weaver. “A mathematical theory of connection”. In: *Mathematical Theory of Communication* (1963).

[2] Quax R, Har-Shemesh O, and Sloot PMA. “Quantifying Synergistic Information Using Intermediate Stochastic Variables”. In: *Entropy* 19.2 (12017), p. 85. DOI: https://doi.org/10.3390/e19020085.

[3] Časlav Brukner, Marek Zukowski, and Anton Zeilinger. “The essence of entanglement”. In: *arXiv* (2001). eprint: quant-ph/0106119. URL: https://arxiv.org/abs/quant-ph/0106119.

[4] Nicholas Timme et al. “Synergy, redundancy, and multivariate information measures: an experimentalist’s perspective”. In: *Journal of Computational Neuroscience* 36 (2014), pp. 119–140.

[5] Peter L. Williams and Randall D. Beer. “Nonnegative decomposition of multivariate information”. In: *arXiv preprint arXiv:1004.2515* (2010).

[6] Elad Schneidman, William Bialek, and Michael J. Berry. “Synergy, redundancy, and independence in population codes”. In: *Journal of Neuroscience* 23 (2003), pp. 11539–11553.

[7] Virgil Griffith and Christof Koch. “Quantifying synergistic mutual information”. In: *Guided Self-Organization: Inception*. Springer, 2014, pp. 159–190.

[8] Jan R. Magnus and Heinz Neudecker. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Chichester, UK: John Wiley & Sons, 1988. ISBN: 978-0471915164.

[9] R. Penrose. “A Generalized Inverse for Matrices”. In: *Proceedings of the Cambridge Philosophical Society* 51 (1955), pp. 406–13. DOI: https://doi.org/10.1017%2FS0305004100030401.

[10] James M. “The generalised inverse.” In: *The Mathematical Gazette.* 62.420 (1978), pp. 109–114. DOI: doi:10.2307/3617665.