

Fast and exact search for the partition with minimal information loss

Shohei Hidaka¹, Masafumi Oizumi^{2,3},

1 Japan Advanced Institute of Science and Technology, Nomi-shi, Ishikawa, Japan

2 Araya Inc., Minato-ku, Tokyo, Japan

3 RIKEN Brain Science Institute, Wako-shi, Saitama, Japan

*shhidaka@jaist.ac.jp, oizumi@araya.org

Abstract

In analysis of multi-component complex systems, such as neural systems, identifying groups of units that share similar functionality will aid understanding of the underlying structures of the system. To find such a grouping, it is useful to evaluate to what extent the units of the system are separable. Separability or inseparability can be evaluated by quantifying how much information would be lost if the system were partitioned into subsystems, and the interactions between the subsystems were hypothetically removed. A system of two independent subsystems are completely separable without any loss of information while a system of strongly interacted subsystems cannot be separated without a large loss of information. Among all the possible partitions of a system, the partition that minimizes the loss of information, called the Minimum Information Partition (MIP), can be considered as the optimal partition for characterizing the underlying structures of the system. Although the MIP would reveal novel characteristics of the neural system, an exhaustive search for the MIP is numerically intractable due to the combinatorial explosion of possible partitions. Here, we propose a computationally efficient search to precisely identify the MIP among all possible partitions by exploiting the *submodularity* of the measure of information loss. Submodularity is a mathematical property of set functions which is analogous to convexity in continuous functions. Mutual information is one such submodular information loss function, and is a natural choice for measuring the degree of statistical dependence between paired sets of random variables. By using mutual information as a loss function, we show that the search for MIP can be performed in a practical order of computational time for a reasonably large system ($N = 100 \sim 1000$). We also demonstrate that MIP search allows for the detection of underlying global structures in a network of nonlinear oscillators.

Author summary

To understand the complex network of the brain, it is important to find groups of neurons that share similar functionality. If structures in the network were obvious and clearly defined, finding natural groupings would be easy. Finding such networks is always challenging, however, because neurons interact with each other in a complicated manner, both structurally and functionally. One way of solving the problem is grouping neurons so that the loss of information from the original network is minimal when the interactions between the groups are hypothetically removed.

Although the idea of finding such a critical partition in real neural data is attractive, an exhaustive search for it is computationally intractable because the number of possible partitions increases explosively as the number of neurons increases. Here, to resolve this combinatorial problem, we propose an efficient algorithm which utilizes the mathematical concept *submodularity*. This algorithm enables identification of the optimal partition in a reasonably large network within a practical amount of time. Our method represents a new analysis tool to untangle the complex network of the brain and reveal its underlying structures.

1 Introduction

The brain can be envisaged as a multi-component dynamical system, in which each of individual components interact with each other. One of the goals of system neuroscience is to identify a group of neural units (neurons, brain area, and so on) that share similar functionality [1–4].

Approaches to identify such functional groups can be classified as “external” or “internal”. In the external approach, responses to external stimuli are measured under the assumption that a group of neurons share similar functionality if their responses are similar. A vast majority of studies in neuroscience have indeed used the external approach, by associating the neural function with an external input to identify groups of neurons or brain areas with similar functionality [5].

On the other hand, the internal approach measures internal interactions between neural units under the assumption that neurons with similar functionality are connected with each other. The attempts to measure internal interactions have rapidly grown following recent advancements in simultaneous recording techniques [6–8]. It is undoubtedly important to elucidate how neurons or brain areas interact with each other in order to understand various brain computations. In particular, to understand awareness or consciousness, the internal approach is preferable to the external approach because consciousness cannot be essentially associated with any external signals. For example, dreaming is a phenomenon which clearly tells us that consciousness occurs regardless of any external signals [9].

Consistent with this, the Integrated Information Theory (IIT) of consciousness is constructed to focus on the internal interactions in a neural system [10–12]. IIT states that the prerequisite for consciousness is integration of information realized by the internal interactions of neurons in the brain. IIT proposes to quantify the degree of information integration by an information theoretic measure, “integrated information” and hypothesizes that integrated information should be related to the level of consciousness [13].

Conceptually, integrated information quantifies the degree of interaction between parts or, equivalently, the amount of information loss caused by splitting a system into parts [10, 11, 14]. By using a measure of information loss such as integrated information, we can find functional groups of neural units using the criterion of “minimal information loss”. For example, consider the system consisting of 4 neurons shown in Fig. 1(a). The two neurons on each of the left and right sides are connected with each other but do not connect with those on the opposite side. The natural inclination is to partition the system into left (orange) and right (blue) subsystems, as shown in Fig. 1(a). This critical partition can be identified by searching for that partition where information loss is minimal or integrated information is minimized. In IIT, the partition with minimal information loss is called the “Minimum Information Partition (MIP)” [10]. In fact, if a system is partitioned with MIP as in the example system, integrated information is 0 because there is no information loss. If the system is partitioned in a different way than MIP, as shown in Fig. 1(b), integrated

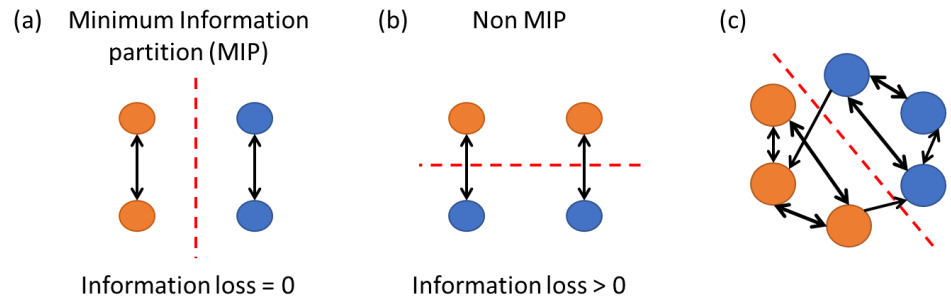


Fig 1. (a) Minimum information partition (MIP). (b) Another possible partition, which differs from the MIP. (c) MIP in a general network where there is no clear-cut partition.

information is non-zero because there are connections between the top (orange) and the bottom (blue) subsystems. This is not the optimal grouping of the system from the viewpoint of information loss.

With the MIP approach, we can quantify “global” integration in a whole system rather than “local” integration in subsystems. The system in Fig. 1 is considered to be a system in which information is locally integrated within each left or right subsystem but is not globally integrated between the left and right subsystems. Integrated information with the MIP shown in Fig. 1(a) quantifies global integration, which is zero and integrated information with the partition shown in Fig. 1(b) quantifies local integration, which is non-zero. IIT considers that a system’s global integration determines consciousness as a whole system. Aside from its relevance to consciousness, MIP could also be used to identify functional groups of neural units from the viewpoint of global information loss.

In a general case in which there is no obvious clear-cut partition (Fig. 1(c)), an exhaustive search for the MIP would take an exceptionally large computational time which increases exponentially with the number of units. This computational difficulty hinders the use of MIP-based characterization of a system, although the theoretical idea is attractive to research into consciousness and other fields of neuroscience, as well as to network science in general.

In this study, we show that the computational cost of searching for the MIP can be reduced to the polynomial order of the system size, if a measure of integrated information is a submodular function. We utilize one of the submodular optimization, the Queyranne’s algorithm [15], and show that the exponentially large computational time is drastically reduced to $O(N^3)$ where N is the number of units. The algorithm proposed in this study is an exact search for the MIP, unlike previous studies which found only the approximate MIP [16, 17]. This algorithm makes it feasible to find MIP-based functional groups in real neural data such as multi-unit recordings, EEG, ECoG, etc., which typically consist of ~ 100 channels.

The paper is organized as follows. In the Section 2, we formulate the search for the MIP, and show that mutual information is one of the submodular functions, and that we can treat it as a measure of information loss for a bi-partition. In Section 3, we report on numerical case studies which demonstrate the computational time of this MIP search for analysis of a system-wise correlation and also demonstrate its use for analysis of a nonlinear system. In Section 4, we discuss the potential use of the submodular search for other measures which are not exactly submodular.

2 Methods

2.1 Submodular function

For a ground set $\mathcal{X} = \{x_1, \dots, x_N\}$ of N elements and any pair of subsets $X, Y \in 2^{\mathcal{X}}$, if a set function $f : 2^{\mathcal{X}} \mapsto \mathbb{R}$ holds the inequality

$$f(X) + f(Y) \geq f(X \cup Y) + f(X \cap Y), \quad (1)$$

we call it *submodular* (See [18] for a review of submodularity). Equivalently, for $X \subseteq Y \subseteq Z$ and $z \in Z \setminus Y$, a submodular set function f holds

$$f(X \cup z) - f(X) \geq f(Y \cup z) - f(Y). \quad (2)$$

If $-f(X)$ is submodular, we call it *supermodular*. Submodularity in discrete functions can be considered as an analogue of convexity in continuous functions. It has been shown that the minimization of submodular functions can be solved in polynomial order of computational time, circumventing the combinatorial explosion. In this study, we utilize submodular optimization to find the partition with minimal information loss (Minimum Information Partition (MIP)).

2.2 Minimum Information Partition (MIP)

We analyze a system with $N \in \mathbb{N}$ distinct components. Assume that each of the N components is a random variable, and denote the random variable of the i^{th} components by x_i for $1 \leq i \leq N$. Denote the set of indices $\mathcal{N} := \{1, 2, \dots, N\}$ and the set of the N variables by $V = \{x_1, \dots, x_N\}$. For the sake of simplicity, we consider bipartition of the whole system V for the explanatory purpose. We will deal with a general k -partition in Section 2.5. V is divided into two parts M and \overline{M} where M is a non-empty subset of the whole system V , $M \subset V$ and \overline{M} is the complement of M , i.e., $\overline{M} = V \setminus M$. Note that bipartition $(V \setminus M, M)$ of a fixed set V is uniquely determined by specifying only one part, M , because the other part is determined as the complement of M . Minimum Information Partition (MIP), M_{MIP} , is defined as the subset that minimizes the information loss caused by partition, indicated by a non-empty subset $M \subset V$,

$$M_{\text{MIP}}(V) := \arg \min_{M \subset V, M \neq \emptyset} f(M), \quad (3)$$

where $f(M)$ is the information loss caused by a bipartition specified by the subset M . More precisely, “MIP” defined in Eq. 3 should be called “Minimum Information Bipartition (MIB)” because only bi-partition is taken into consideration. However, as we will show in Section 2.5, the proposed method is not restricted only to a bi-partition and can be extended to a general k -partition. To simplify terminology, we only use the term “MIP” through out the paper even when only bi-partition is considered. IIT proposes several measures for information loss, which it terms “integrated information” [10, 11, 14]. In terms of integrated information, MIP is defined as the partition where integrated information between the parts is minimized.

The number of possible bi-partitions for the system size N is $2^{N-1} - 1$, which grows exponentially as a function of the system size N . Thus, for even a modestly large number N of variables ($N \sim 40$), exhaustively searching all bi-partitions is computationally intractable.

2.3 Information loss function

In this study, we use the mutual information between the two parts M and \overline{M} as an information loss function,

$$f(M) := I(M; \overline{M}), \quad (4)$$

$$= H(M) + H(\overline{M}) - H(M, \overline{M}), \quad (5)$$

where $H(X)$ is the Shannon entropy [19, 20] of a random variable X ,

$$H(X) := - \sum_{x \in X} P(x) \log P(x).$$

As we will show in the next section, the mutual information is a submodular function. The mutual information (Eq. 5) is expressed as the KL-divergence between $P(V)$ and the partitioned probability distribution $Q(V) = P(M)P(\overline{M})$ where the two parts M and \overline{M} are forced to be independent,

$$I(M; \overline{M}) = D_{KL}(P(X) || P(M)P(\overline{M})). \quad (6)$$

The Kullback-Leibler divergence measures the difference between the probability distributions and can be interpreted as the information loss when $Q(V)$ is used to approximate $P(V)$ [21]. Thus, the mutual information between M and \overline{M} (Eq. 6) can be interpreted as information loss when the probability distribution $P(V)$ is approximated with $Q(V)$ under the assumption that M and \overline{M} are independent [14]. In the first version of IIT (IIT 1.0), the mutual information was used as a measure of integrated information [10]. The difference between the mutual information considered in this study and that in IIT 1.0 is that the maximum entropy distribution is used as a probability distribution of one part M or \overline{M} . Although different measures that take account of the dynamical aspects of a system are proposed in the later versions of IIT [11, 12], we do not consider them in this study (see Discussion). Also, note that in IIT, the measures of integrated information are based on a perturbational approach, which attempts to quantify actual causation by perturbing a system into all possible states [10–12, 22]. The perturbational approach requires the full knowledge of the physical mechanisms of a system, i.e., how the system behaves in response to all possible perturbations. Aiming at an empirical application of our method, we only consider the measure (Eq. 6) based on an observational distribution that can be estimated from empirical data.

2.4 Submodularity of the loss functions

We will show that the mutual information (Eq. 5) is submodular. To do so, we use the submodularity of entropy. The entropy $H(X)$ is submodular because for $X \subset Y \subset Z$ and $z \in Z \setminus Y$,

$$\begin{aligned} H(X \cup \{z\}) - H(Y \cup \{z\}) &= -H(Y \setminus X | X \cup \{z\}) \\ &\geq -H(Y \setminus X | X) \\ &= -H(Y) + H(X), \end{aligned}$$

which satisfies the condition of submodularity (Eq. 2).

By straightforward calculation, we can find that the following identity holds for the loss function $f(M) = I(M; \overline{M})$.

$$\begin{aligned} f(A \cup B) + f(A \cap B) - f(A) - f(B) &= H(A \cup B) + H(A \cap B) - H(A) - H(B) \\ &\quad + H(\overline{A} \cup \overline{B}) + H(\overline{A} \cap \overline{B}) - H(\overline{A}) - H(\overline{B}). \end{aligned} \quad (7)$$

Thus, from the submodularity of the entropy, the following inequality holds,

$$f(A \cup B) + f(A \cap B) - f(A) - f(B) \leq 0, \quad (8)$$

which shows that $f(M) = I(M; \overline{M})$ is submodular.

A submodular system (V, f) is said to be *symmetric* if $f(M) = f(V \setminus M)$ for any subset $M \subseteq V$. It is easy to see that the mutual information is a symmetric submodular function from Eq. 5. When a submodular function is symmetric, the minimization of submodular function can be solved more efficiently. Applying Queyranne's algorithm [15] (See also Appendix in Section 4), we can precisely identify the bi-partition with the minimum information loss in computational time $O(N^3)$.

2.5 Recursive search for k -partition

As introduced above, Queyranne's algorithm can find the minimum information bi-partition. We can extend the algorithm to the exact search for a general k -partition with minimal information loss although it is more computationally costly. In what follows, we specifically explain 3-partition case for simplicity, but the argument is applicable to any k -partition in a form of mathematical induction.

Denote by the set of all k -partitions $(M_0, M_1, \dots, M_{k-1})$ of V by $P_{k,V}$. For the set $P_{k,V}$ and $2 \leq k \leq |V|$, define the function $g : P_{k,V} \mapsto \mathbb{R}$ by

$$g((M_0, M_1, \dots, M_k)) := \sum_{i=1}^k H(M_i) - H(V),$$

and it gives a natural extension of the information loss function to the k -partition. For the special case $k = 2$, the reader would find this corresponds with the function $g((M_0, M_1)) = I(M_0; V \setminus M_0)$, mainly discussed in this study. This extension of the mutual information is known as total correlation [23] or multi-information [24]

For $k = 3$, we wish to find the 3-partition (M_0, M_1, M_2) minimizing the function $g((M_0, M_1, M_2))$. Let us simplify the notation by defining $f_V(M_0) := g((M_0, M_1))$. The 3-partition minimization problem can be reduced to the minimization of the symmetric submodular function of a non-empty subset $U \subset V$ defined by

$$g_{3,V}(U) := f_V(U) + h_{2,V}(U),$$

where

$$h_{2,V}(U) := \begin{cases} \min \begin{cases} \min_{(M_0, M_1) \in P_{2,V \setminus U}} f_{V \setminus U}(M_0) \\ \min_{(M_0, M_1) \in P_{2,U}} f_U(M_0) \end{cases} & (\min(|U|, |V \setminus U|) > 1) \\ \min_{(M_0, M_1) \in P_{2,V \setminus U}} f_{V \setminus U}(M_0) & (|U| = 1) \\ \min_{(M_0, M_1) \in P_{2,U}} f_U(M_0) & (|V \setminus U| = 1) \end{cases}.$$

The proof for the submodularity of this function $g_{3,V}$ is given by Lemma 1 in Supporting Information 2. The desired 3-partition is (\hat{U}, M_0, M_1) such that

$$\hat{U} = \arg \min_{\emptyset \subset U \subset V} g_{3,V}(U), \quad (M_0, M_1) = \arg \min_{(M_0, M_1) \in V \setminus \hat{U}} f_{V \setminus \hat{U}}(M_0),$$

or $(V \setminus \hat{U}, M_0, M_1)$ such that

$$\hat{U} = \arg \min_{\emptyset \subset U \subset V} g_{3,V}(U), \quad (M_0, M_1) = \arg \min_{(M_0, M_1) \in \hat{U}} f_{\hat{U}}(M_0).$$

Applying this argument recursively, the k -partition minimizer is used to find the $(k+1)$ -partition minimizing the function $g_{k+1,V}$. By applying the Queyranne's algorithm to each of these recursive steps, this algorithm needs $O(N^{3(k-1)})$ times of function calls.

2.6 Sequential MIP search

As shown in the previous section, the exact search for MIP in the case of a general k -partition is computationally costly. Moreover, for the moment, it is unclear which criterion we should employ to choose the k -partition solution over $(k - 1)$ -partition one. In practical application, thus, one may want to apply the bi-partition procedure sequentially, which is much faster than the exact minimum k -partition, and easy to decide the partition size – partitioning until no more set found to make a further partition. Thus, we implement *sequential MIP search*, in which each of subsets of the MIP found in a given set of variables is further partitioned at each recursive step until it satisfies certain condition (e.g., a given set size limit, a number of maximum number of steps, and so on). In Study 2 (Section 3.2), we demonstrate a sequential MIP search until it reaches the step with no further possible partition. Note that the sequential MIP search introduced here does not coincide with the exact minimum k -partition in general and is considered to be a crude approximation. However, it would be useful for a practical application because of the low computational costs.

3 Numerical Studies

To demonstrate this search for the bi-partition with the minimal loss of information, we report here several case studies with artificial datasets. Throughout these case studies, we assume that the data is distributed normally. Under this assumption, we obtain the simple closed form

$$f(M) = I(M; \overline{M}) = \log_2 |\Sigma_M| + \log_2 |\Sigma_{\overline{M}}| - \log_2 |\Sigma_X|, \quad (9)$$

where Σ_X is the covariance matrix of the data, $\Sigma_M, \Sigma_{\overline{M}}$ is the covariance matrix of the variables in the subsets M and \overline{M} , and $|\Sigma|$ denotes the determinant of the matrix Σ . The computation of $|\Sigma_X|$ can be omitted because $|\Sigma_X|$ is constant across every step in the search and has no effect on the minimization of $I(M; \overline{M})$.

3.1 Study 1: Computational time

In the first case study, we compare the practical computational time of the submodular search with that of the exhaustive search. We artificially generated a dataset consisting of 10,000 points normally distributed over N dimensional space for $N = 2, 3, \dots, 400$. Each dimension is treated as an element in the set. The exhaustive search is performed up to $N = 16$, but could not run in a reasonable time for the dataset with $N = 17$ or larger due to limitations of the computational resource. Up to $N = 16$, we confirmed that the submodular search found the correct MIPs indicated by the exhaustive search.

Figure 2 (a) shows the semi-logarithm plot of the computation time of the two searches. The empirical computation time of the exhaustive search was closely along the line, $\log_2 T = 0.891N - 12.304$. This indicated that the exhaustive search took an exponentially large computational time $\approx O(2^N)$, which fits with the number of possible bi-partitions. Figure 2 (b) shows exactly the same results as the double-logarithm plot. In this plot, the computational time of the Queyranne's search was closely along the line $\log_2 T = 3.210 \log_2 N - 18.722$, which indicated that the Queyranne's search took cubic time $\approx O(N^3)$, as expected from the theory. With $N = 1000$, the Queyranne's search takes 9738 seconds of running time. The computational advantage of the Queyranne's search over the exhaustive search is obviously substantial. For example, even with a modest number of elements, say $N = 40$, the computational time of the exhaustive search is estimated to be $1.07 \times 10^7 \text{sec} \approx 123 \text{days}$ while that of the Queyranne's search is only 1sec.

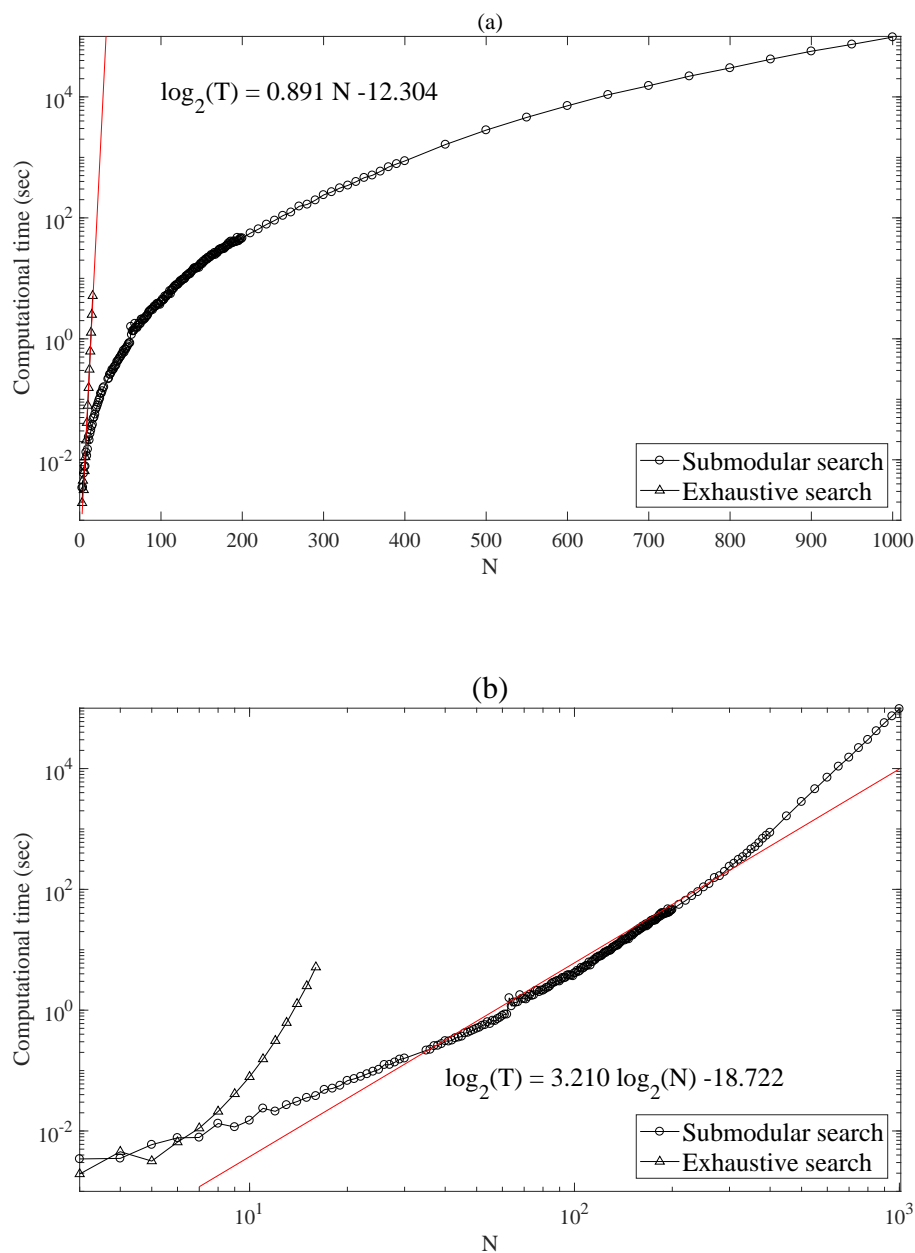


Fig 2. (a) The semi-logarithm plot and (b) double-logarithm of the computation time for the two searches.

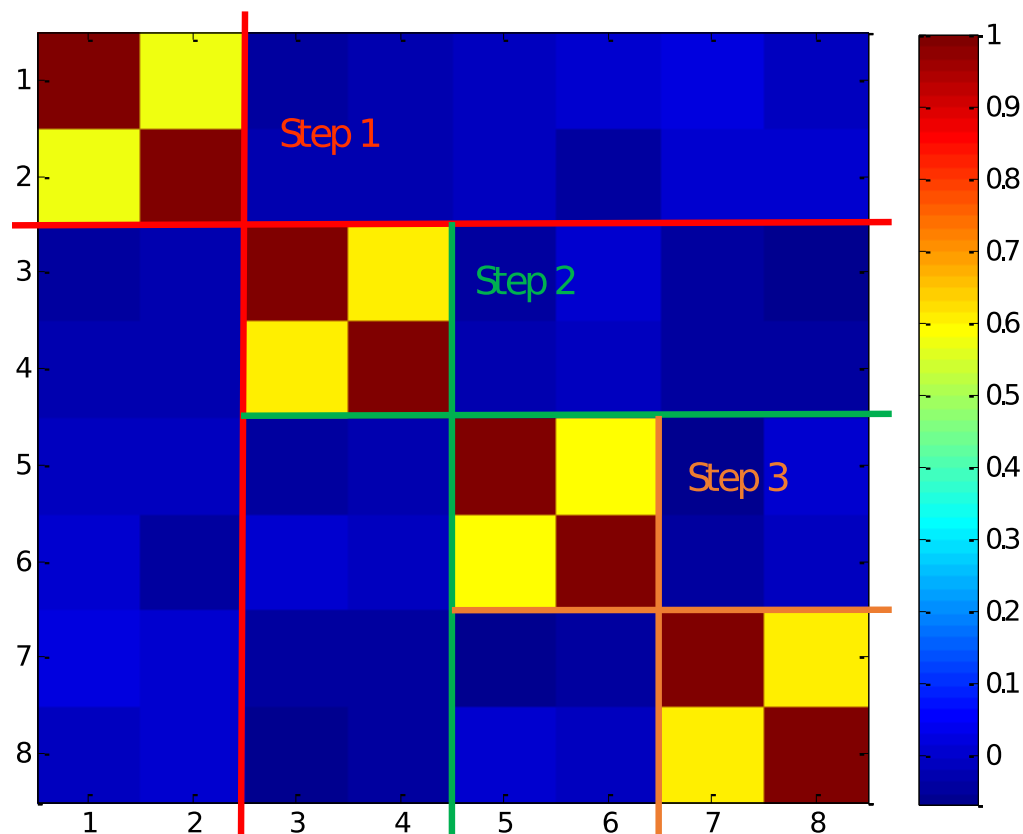


Fig 3. Correlation matrix which reflects the toy problem with four correlated pairs in a system with eight components. The colored line shows partitioning at each step.

3.2 Study 2: Sequential MIP search

As a demonstration of sequential MIP search (Section 2.6), we consider a set of 8 random variables with the correlation matrix shown in Figure 3. There are four pairs, variables 1 and 2, 3 and 4, 5 and 6, and 7 and 8, with positive correlations, while any other pairs of variables show nearly zero correlation. This correlation matrix is constructed as follows: For each pair of correlated variables (such as the Variable 1 and 2), a set of 1000 samples (X, Y) is drawn using the independent bi-variate normal distribution with the mean $(0, 0)$ and the variance $(1, 0.5)$, and the coordinate (X, Y) is rotated by $\pi/4$ to (X', Y') . This rotation was applied so that it lets the generated uncorrelated normally-distributed samples be correlated. The eight dimensional dataset analyzed is constructed by concatenating the four pairs of variables, each of which is constructed in this way.

By applying the first MIP search, the system is partitioned into the pair of Variables 1 and 2, and the rest. As the pair of variables has no further meaningful partition, the other set of six variables is further partitioned. At this second step, the remaining six variables are partitioned into the pair of variables 3 and 4 and the rest. By recursively applying this procedure to the subset with a larger number of variables, we obtain the four subsets $(1, 2)$, $(3, 4)$, $(5, 6)$, and $(7, 8)$. As expected from Figure 3, the sequential partitioning successfully re-discovers the underlying four groups of elements from the dataset.

3.3 Study 3: Nonlinear dynamical systems

In Study 3, we demonstrate how MIP changes depending on underlying network structures. For this purpose, we chose a nonlinear dynamical system in which multiple nonlinear components are chained on a line. Specifically, we construct a series of variants of the Coupled Map Lattice (CML) [25]. Kaneko [25] analyzed the CML in which each component is a logistic map and interacts with the one or two other nearest components on a line, and showed the emergence of multiple types of dynamics in the CML. In this model, each component is treated as a nonlinear oscillator, and the degree of interaction between other oscillators can be manipulated parametrically. By manipulating the degree of interaction, we can continuously change the global structure of the CMLs from one coherent chain to two separable chains. We apply the MIP search for the CMLs with different interaction parameters, and test whether the MIP captures this underlying global structure of the network.

Specifically, the CML is defined as follows. Let us write the logistic map with a parameter a by $f_a(x) := 1 - ax^2$. Let $x_{i,t} \in [0, 1]$ be a real number indicating the i^{th} variable at t^{th} time step for $i = 1, 2, \dots, N, t = 0, 1, \dots, T$. For each $i = 1, \dots, N$, the initial state of the variable $x_{i,0}$ is set to a random number drawn from the uniform distribution on $[0, 1]$. For $t > 0$, we set the variables with the lateral connection parameter $\epsilon \in [0, 1]$ by

$$\begin{aligned} x_{1,t} &= (1 - \epsilon)f_a(x_{1,t-1}) + \epsilon f_a(x_{2,t-1}), \\ x_{N,t} &= (1 - \epsilon)f_a(x_{N,t-1}) + \epsilon f_a(x_{N-1,t-1}), \\ x_{i,t} &= (1 - \epsilon)f_a(x_{i,t-1}) + \frac{\epsilon}{2}(f_a(x_{i-1,t-1}) + f_a(x_{i+1,t-1})) \text{ for } 1 < i < N. \end{aligned}$$

According to the previous study [25], the *defect turbulence* pattern in the spatio-temporal evolution in $(x)_{i,t}$ is observed with the parameter $a = 1.8950$ and $\epsilon = 0.1$. In this study, we additionally introduce the “connection” parameter between the variables $i = 20, 21$ among $N = 30$ variables (Figure 4(a)). Namely, with the connection parameter δ , we redefine variables 19, 20, 21 and 22 by

$$\begin{aligned} x_{19,t} &:= (1 - \epsilon)f_a(x_{19,t-1}) + \frac{\epsilon}{2}f_a(x_{18,t-1}) + (1 - \delta)\epsilon f_a(x_{20,t-1}) \\ x_{20,t} &:= (1 - \epsilon)f_a(x_{20,t-1}) + \frac{\epsilon}{2}f_a(x_{19,t-1}) + \delta\epsilon f_a(x_{21,t-1}) \\ x_{21,t} &:= (1 - \epsilon)f_a(x_{21,t-1}) + \frac{\epsilon}{2}f_a(x_{22,t-1}) + \delta\epsilon f_a(x_{20,t-1}) \\ x_{22,t} &:= (1 - \epsilon)f_a(x_{22,t-1}) + \frac{\epsilon}{2}f_a(x_{23,t-1}) + (1 - \delta)\epsilon f_a(x_{21,t-1}). \end{aligned}$$

With the connection parameter $\delta = 1/2$, this model is identical to the original CML above, and with $\delta = 0$, it is equivalent to the two independent CMLs of (x_1, \dots, x_{20}) and (x_{21}, \dots, x_{30}) , as it has no interaction between variable 20 and 21.

Given a sufficiently small connection parameter $0 \leq \delta < 1/2$, we expect that the MIP would separate the system into the subsets $\{1, 2, \dots, 20\}$ and $\{21, 22, \dots, 30\}$ because the degree of the interaction between units 20 and 21 is the smallest. On the other hand, if the system is fully connected, which happens when $\delta = 1/2$, we expect that the MIP would separates the system into the subsets $\{1, 2, \dots, 15\}$ and $\{16, 17, \dots, 30\}$ (in the middle of 30 units), due to the symmetry of connectivity on the line. The correlation matrices for different connection parameters δ from 0 to 1/2 are shown in Figure 4. In each matrix of 4 (b), (c) and (d), the crossed white lines show the expected separation between variables 20 and 21 at which the parameter δ is manipulated. We found the block-wise correlation patterns in the matrix with $\delta = 0$, as expected (a typically found partition is depicted in the black

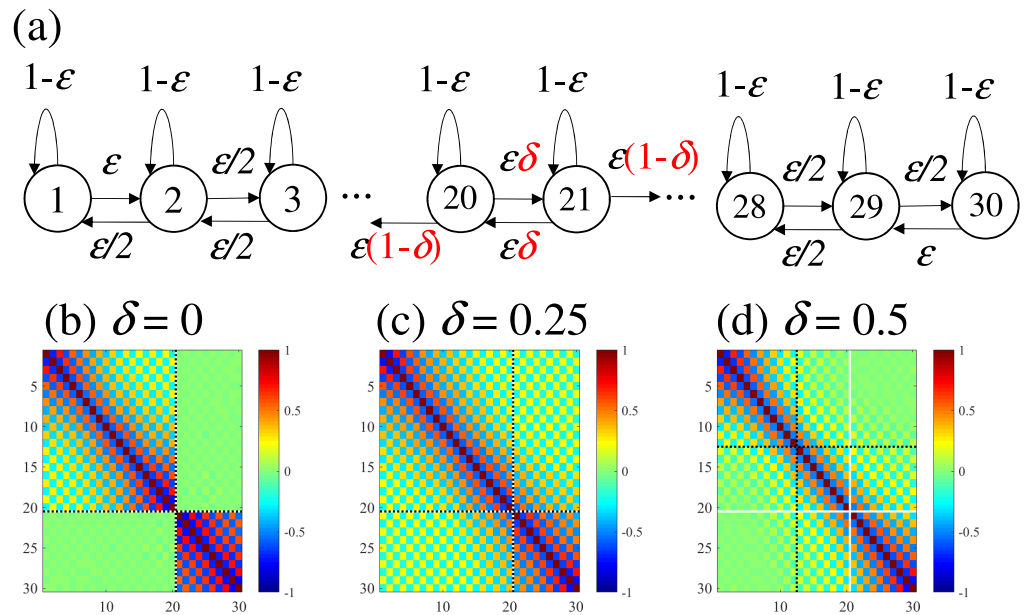


Fig 4. (a) The Coupled Map Lattice model with connection parameter δ indicating the connectivity between variables 20 and 21. The correlation matrices with (b) $\delta = 0$ (disconnected), (c) $\delta = 0.25$ (half-connected), and (d) $\delta = 0.5$ (fully connected). The white crossing lines show the expected partition as the ground truth at which the parameter δ is manipulated, and the black dotted crossing lines show the MIP typically found for each particular parameter.

dotted line in Figure 4 (b), (c) and (d)); similar but less clear patterns with $\delta = 0.25$; and no clear block-wise patterns with $\delta = 1/2$.

To summarize, this case study confirmed our theoretical expectation that the MIP captures the block-wise informational components; namely that the partition probability is a decreasing function of the connection parameter δ (Figure 5). This means that the MIP search detects the weakest underlying connection between 20 and 21, and successfully separates it into the two subsets, if the connections between 20 and 21 are weak.

4 Discussion

In this paper, we proposed a fast and exact algorithm which finds the weakest link of a network at which the network can be partitioned with the minimal information loss (MIP). Since searching for the MIP has the problem of combinatorial explosion, we employed Queyranne's algorithm for a submodular function. We first showed that the mutual information is a symmetric submodular function. Then, we used it as an information loss function to which Queyranne's algorithm can be applied. Our numerical case studies demonstrate the utility of the MIP search for a set of locally interacting nonlinear oscillators. This demonstration opens the general use of the MIP search for system neuroscience as well as other fields.

The proposed method can be utilized in Integrated Information Theory (IIT). There are several variants of integrated information [10–12, 14, 16, 26–28]. The mutual information was used as a measure of integrated information in the earliest version of IIT [10], but different measures which take account of dynamical aspects of a network were proposed in the later versions [11, 12]. To apply the proposed method, we first

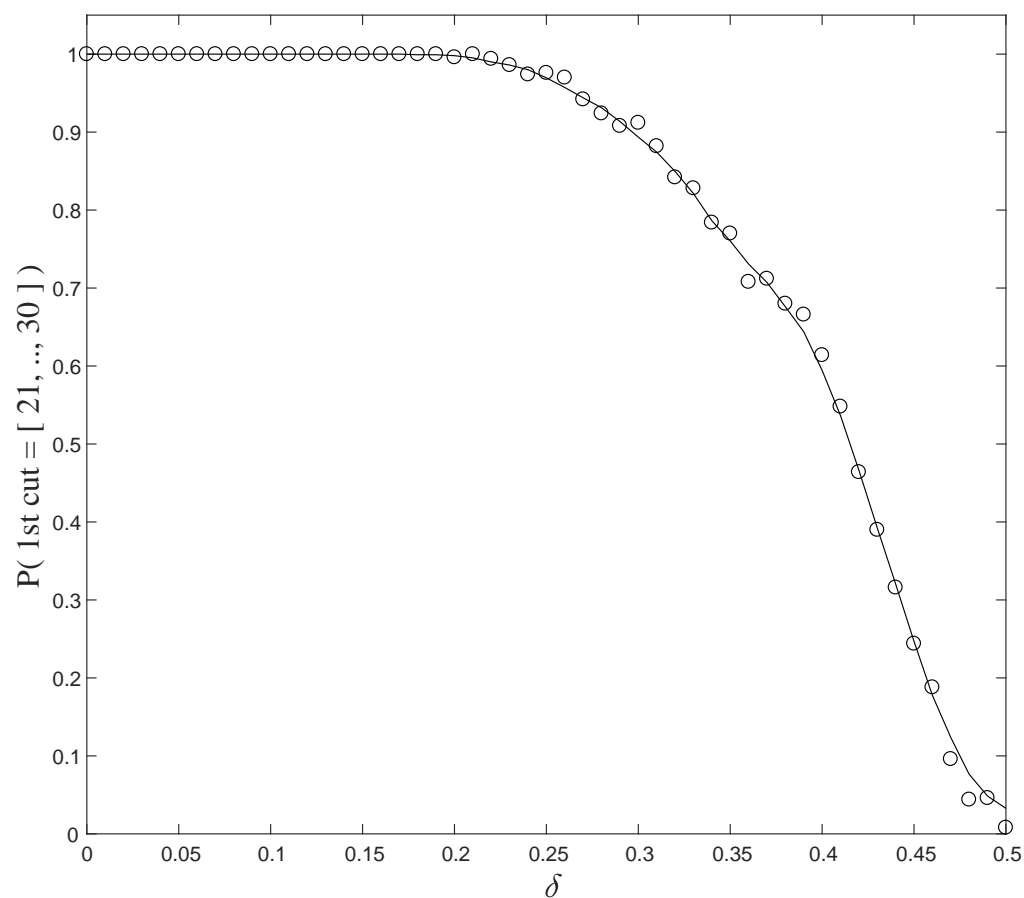


Fig 5. The probability of the subset with the smaller number of elements in the MIP is $\{21, 22, \dots, 30\}$ is plotted as a function of the connection parameter $\delta \in [0, 1/2]$. Each circle shows the sample probability of 500 independent simulations, and the solid line shows the moving average of the probability.

need to assess whether the other measures of integrated information are submodular or not. Even when the measures are not strictly submodular, the proposed algorithm may provide a good approximation of the MIP. An important future work is to assess the submodularity of the measures of integrated information, and also the goodness of the proposed algorithm as an approximation.

In the present study, we limit our focus to bi-partition to simplify the problem. There are some cases, however, for which tri-partition or general k -partition is considered appropriate. In general, the sequential bi-partitioning procedure does not necessarily result in the same partitioning for a given set of elements. Thus, a proper optimization procedure that is efficiently computable would in future be formulated in the form of recursive use of submodular optimization.

Supporting information: Queyranne's algorithm

In this appendix, we briefly describe Queyranne's algorithm [15]. Suppose we have a submodular system (V, g) where V is a given set of elements and g is a submodular function defined for the power set of V . Let $f(U) := g(U) + g(V \setminus U)$ for every subset $U \subseteq V$ be a symmetric function constructed with the submodular function g .

Queyranne's algorithm is used to search the subset U which minimizes the symmetric submodular function $f(U)$. For example, in this study, we consider the case that $f(U)$ is mutual information (Eq. 5) and $g(U)$ is entropy. For example, in this study, we consider the case that $f(U) = 2I(U; V \setminus U)$, identified up to a constant multiplier, and $g(U) = I(U; V \setminus U)$ are both mutual information (Eq. 5).

In the algorithm proposed in [15], the key observation is that a special ordered pair (t, u) , called a *pendent pair*, can be identified for an arbitrary subset $U \subseteq V$ in $O(N^2)$ time. Identification of a pendent pair (t, u) of the set V reduces the search space because for the desired subset U minimizing $f(U)$, either case (1) $U = \{u\}$ or (2) $U \supseteq \{t, u\}$ holds. Thus, by keeping case (1) as a candidate for the minimal partition, we can further refine case (2), in which we define a new ground set V' where the elements $\{t, u\}$ are treated as an inseparable unit element u' . By using the new merged element u' , V' is defined as

$$V' := V \setminus \{t, u\} \cup \{u'\}.$$

After this procedure is applied once, the effective number of elements is reduced to $N - 1$. By applying this procedure recursively to search the set V' with $N - 1$ elements, we would obtain another candidate for the minimal partition and a candidate set V'' with $N - 2$ elements for further search. Thus, by finding the pendent pair for the given set V at each step recursively, we obtain $N - 1$ candidates for the minimal partition, and then find the minimal one from among them. In summary, this recursive computation takes $O(N^3)$ time because it requires the construction of a series of pendent pairs in $O(N^2)$, and $N - 1$ pendent pairs are needed to construct for minimization.

Next we illustrate the construction of a pendant pair. An ordered pair (t, u) of elements of V is called a pendent pair for (V, g) , if $f(u)$ takes the minimum in all subsets of V which separate t from u , or equivalently

$$f(u) = \min\{f(U) \mid U \subset V, t \notin U \text{ and } u \in U\}.$$

There is at least one pendent pair for any symmetric submodular function. Further, a pendent pair can be constructed specifically for an element $x \in V$ as follows. For an element $x \in V$, let us write $v_1 := x$, $W_0 = \emptyset$, and $W_1 = \{v_1\}$. For $i > 1$,

$$v_i := \arg \min_{u \in V \setminus W_{i-1}} g(W_{i-1} \cup \{u\}) - g(\{u\}),$$

and $W_i := W_{i-1} \cup \{v_i\}$. For a set V of the size $N = |V|$, the (v_{N-1}, v_N) is a pendent pair. This construction of a pendent pair needs $O(N^2)$ times of evaluation of the function f . Importantly, for all $y \in V \setminus W_i$ and all $x \subseteq W_{i-1}$ in the series $(W_i)_{i=1}^N$ constructed by the procedure above for the submodular system (V, f) , the following inequality holds

$$g(W_i) + g(y) \leq g(W_i \setminus X) + g(X + y).$$

See [15] for the proof of this inequality. By putting $i = N - 1$ in the inequality, we can see that the partition $(v_N, V \setminus \{v_N\})$ gives the minimum among all subsets separating v_N from v_{N-1} .

By definition of the pendent pair, one of the following two cases, case 1 or 2, holds for a given pendent pair (t, u) .

1. The set $\{u\}$ is a solution of the minimization problem.
2. Some set $U \supseteq \{t, u\}$ is a solution of the minimization problem.

In the first case, the algorithm reports it. In the second case, the algorithm constructs another submodular system (V', f) , in which a new element is defined by merging the pendent pair $u' = \{t, u\}$ and $V' = V \setminus \{t, u\} \cup u'$. The new system (V', f) with the merged pair is also submodular, and thus the same argument for the pendent pair can apply recursively.

Supporting information 2: Extension to k -partition algorithm

The Queyranne's algorithm works on minimization of $g(U) = f(U) + f(V \setminus U)$ with respect to non-empty set $U \subset V$ or $g((U, V \setminus U)) = f(U) + f(V \setminus U)$ over bi-partition $(U, V \setminus U)$ with an arbitrary submodular set function f . Here we show a recursive method extending this symmetric submodular search over a set of bi-partitions to that of 3-partitions. The following argument will be easily extended to that of k -partition.

First let us denote the set of k -partitions of a given set V by

$$P_{k,V} := \left\{ (M_0, M_1, \dots, M_{k-1}) \mid \bigcup_i M_i = V, M_i \cap M_j = \emptyset \text{ for any } i \neq j \text{ and, } M_i \neq \emptyset \text{ for every } i \right\}$$

For a submodular system (V, f) of a given underlying set V and a submodular set function $f : 2^V \mapsto \mathbb{R}$, we consider minimization of function $g : P_{3,V} \mapsto \mathbb{R}$ of the form

$$g((M_0, M_1, M_2)) = \sum_{i=0}^2 f(M_i) + c, \quad (10)$$

where $c \in \mathbb{R}$ is a constant. This is an extension of the bi-partition function $g((U, V \setminus U)) = f(U) + f(V \setminus U)$ to 3-partition function. In this section, we provide an algorithm to minimize this k -partition function by employing Queyranne's algorithm.

By defining $f(M) := H(M)$ for $M \subseteq V$, (V, f) is a submodular system, and the information loss function is written with a constant $c = -f(V)$ by

$$g((M_0, M_1, M_2)) = \sum_{i=0}^2 f(M_i) - f(V).$$

For the special case $k = 2$, $g((M_0, M_1)) = I(M_0; M_1)$, this is identical to the minimal loss of information introduced in this study.

Our argument below does not depend on any specific form of a particular submodular function f , as long as the objective function takes the form in (10). The basic idea is to reduce the original objective function $g : P_{3,V} \mapsto \mathbb{R}$ to a set function $g_{3,V} : 2^V \mapsto \mathbb{R}$ by recursively defining $g_{2,U}$ for the remaining two subsets in a given bi-partition. As our goal is to minimize $g_{3,V}$, such reduction can be written specifically for non-empty $U \subset V$ by

$$g_{3,V}(U) := f_V(U) + h_{2,V}(U), \quad (11)$$

where for any $\emptyset \subset U_1 \subset U_2$, $f_{U_2}(U_1) := f(U_1) + f(U_2 \setminus U_1)$ and $h_{1,U_2}(U_1) := 0$ and

$$h_{2,U_2}(U_1) := \begin{cases} \min \begin{cases} \min_{\emptyset \subset U' \subset U_2 \setminus U_1} g_{2,U_2 \setminus U_1}(U') & (\min(|U_1|, |U_2 \setminus U_1|) > 1) \\ \min_{\emptyset \subset U' \subset U_1} g_{2,U_1}(U') \end{cases} \\ \min_{\emptyset \subset U' \subset U_2 \setminus U_1} g_{2,U_2 \setminus U_1}(U') & (|U_1| = 1) \\ \min_{\emptyset \subset U' \subset U_1} g_{2,U}(U') & (|U_2 \setminus U_1| = 1) \end{cases}. \quad (12)$$

This function (11) can be interpreted as recursive bi-partitioning across multiple stages: The first partition $(U, V \setminus U)$ of the set V is made on f_V , and the second partition (U', \overline{U}') of either U or $V \setminus U$ on h_{k-1} , and so forth. For $|U| = 1$ or $|V \setminus U| = 1$, there is only one set for which the second partition can be made, otherwise smaller one of either $f_U(M_0)$ or $f_{V \setminus U}(M_0)$ has the solution. For $k = 2$, $g_{2,V}(U) = f_V(U)$, and minimization of $f_V(U)$ over the set of bi-partitions of V can be computed by the Queyranne's algorithm.

If this function $g_{3,V}$ is symmetric and submodular, we can apply the Queyranne's algorithm to this function at every recursive step above. Then, the minimum of $g_{3,V}(\hat{U})$ is identical to $g((\hat{U}, M_0, M_1))$ with the 3-partition is $(\hat{U}, \hat{M}_0, \hat{M}_1)$ such that

$$\hat{U} = \arg \min_{\emptyset \subset U \subset V} g_{3,V}(U) \text{ and } (\hat{M}_0, \hat{M}_1) = \arg \min_{(M_0, M_1) \in P_{2,V \setminus \hat{U}}} h_2(M_0, M_1)$$

or $(V \setminus \hat{U}, \hat{M}_0, \hat{M}_1)$ such that

$$\hat{U} = \arg \min_{\emptyset \subset U \subset V} g_{3,V}(U) \text{ and } (\hat{M}_0, \hat{M}_1) = \arg \min_{(M_0, M_1) \in P_{2,\hat{U}}} h_2(M_0, M_1).$$

As $g_{k,V}$ is obviously symmetric by definition, our main question now is whether it is submodular. The lemma following states that it is submodular.

Lemma 1. *For a given submodular system (V, f) , the function $g_{3,V} : 2^V \mapsto \mathbb{R}$ is submodular.*

The proof of Lemma 1 needs Lemma 2 which is stated after the proof.

Proof. For a pair of set $X, Y \subseteq V$ and $k > 2$, consider the difference

$$\Delta := g_{3,V}(X) + g_{3,V}(Y) - g_{3,V}(X \cup Y) - g_{3,V}(X \cap Y),$$

and

$$\Delta_h := h_{3,V}(X) + h_{3,V}(Y) - h_{3,V}(X \cup Y) - h_{3,V}(X \cap Y).$$

As $\Delta - \Delta_h \geq 0$ due to submodularity of the function f , $\Delta_h \geq 0$ implies $\Delta \geq 0$ for any k . We show $\Delta_h \geq 0$ as follows. Now we will show that $\Delta_h \geq 0$, by using the fact that any function $g_{2,U}$ for $U \subseteq V$ is submodular.

Let us denote the minimizer of $g_{3,X}$ by

$$[X] := \arg \min_{\emptyset \subset W \subset X} g_{3,X}(W).$$

For all three cases of the function h (12), we consider the case below as the lower bound for Δ_h , and the other cases follow the essentially same argument:

$$\Delta_h \geq g_{3,X}([X]) + g_{3,Y}([Y]) - g_{3,X \cup Y}([X \cup Y]) - g_{3,X \cap Y}([X \cap Y])$$

By definition, $g_{3,X}([X]) \leq g_{3,X}(W)$ for any non-empty $W \subset X$, and thus for any non-empty $W \subset X \cup Y$ and $Z \subset X \cap Y$, we have

$$\Delta_{h,3} \geq g_{3,X}([X]) + g_{3,Y}([Y]) - g_{3,X \cup Y}(W) - g_{3,X \cap Y}(Z).$$

By Lemma 2, $\Delta_h \geq 0$ due to submodularity of $g_{2,U}$ for any $U \subseteq V$. □

The last line in the proof of Lemma 1 is shown as follows.

Lemma 2. *For an arbitrary submodular system (V, f) , $f(\emptyset) = 0$ and $\emptyset \subset W \subset Z \subseteq V$, write*

$$f_Z(W) := f(W) + f(Z \setminus W).$$

For any given pair of partitions $(W, \overline{W}) \in P_{2,X \cup Y}$ and $(Z, \overline{Z}) \in P_{2,X \cap Y}$, there is some pair $(A, \overline{A}) \in P_{2,X}$ and $(B, \overline{B}) \in P_{2,Y}$ which holds the inequality

$$f_X(A) + f_Y(B) \geq f_{X \cup Y}(W) + f_{X \cap Y}(Z).$$

Vice versa, for any given pair of partitions $(A, \overline{A}) \in P_{2,X}$ and $(B, \overline{B}) \in P_{2,Y}$, there is some pair $(W, \overline{W}) \in P_{2,X \cup Y}$ and $(Z, \overline{Z}) \in P_{2,X \cap Y}$ which holds the inequality above.

This lemma states about a kind of submodularity on the bi-partition function f_X by existence of some pair W and Z for any pair A and B or vice versa. The proof of Lemma 2 is beyond the scope of this section, and will be reported elsewhere.

The Lemma 1 states that Queyranne's algorithm can be used to minimize the symmetric submodular set function $g_{m,U}(U')$ with respect to non-empty $U' \subset U$ at each step of the function $g_{k,V}(U)$. This observation further applies recursively for a k -partition problem, which can be reduced to the minimization of $(k-1)$ -partition problem by defining an appropriate function. As minimization of a k -partition function includes the minimization of $(k-1)$ -partition function, the computational time of this recursive algorithm for the minimal k -partition of n elements would take the order $O(n^{3(k-1)})$.

Acknowledgments

We thank Ryota Kanai for his comments and discussions on earlier versions of this manuscript. This work was partially supported by CREST, Japan Science and Technology Agency, and by the JSPS KAKENHI Grant-in-Aid for Scientific Research on Innovative Areas JP 16H01609 and for Scientific Research B (Generative Research Fields) JP 15KT0013.

References

1. Humphries MD. Spike-train communities: finding groups of similar spike trains. *Journal of Neuroscience*. 2011;31(6):2321–2336.
2. Lopes-dos Santos V, Ribeiro S, Tort ABL. Detecting cell assemblies in large neuronal populations. *Journal of Neuroscience Methods*. 2013;220(2):149–166. doi:10.1016/j.jneumeth.2013.04.010.

3. Carrillo-Reid L, Yang W, Bando Y, Peterka DS, Yuste R. Imprinting and recalling cortical ensembles. *Science*. 2016;353(6300):691–694.
4. Romano SA, Pérez-Schuster V, Jouary A, Boulanger-Weill J, Candeo A, Pietri T, et al. An integrated calcium imaging processing toolbox for the analysis of neuronal population dynamics. *PLOS Computational Biology*. 2017;13(6):e1005526.
5. Hubel DH, Wiesel TN. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*. 1962;160(1):106–154. doi:10.1113/jphysiol.1962.sp006837.
6. Harris KD, Csicsvari J, Hirase H, Dragoi G, Buzsáki G. Organization of cell assemblies in the hippocampus. *Nature*. 2003;424(6948):552–556.
7. Schneidman E, Berry MJ, Segev R, Bialek W. Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature*. 2006;440(7087):1007–1012.
8. Stevenson IH, Kording KP. How advances in neural recording affect data analysis. *Nature neuroscience*. 2011;14(2):139–142.
9. Nir Y, Tononi G. Dreaming and the brain: from phenomenology to neurophysiology. *Trends in cognitive sciences*. 2010;14(2):88–100.
10. Tononi G. An information integration theory of consciousness. *BMC Neurosci*. 2004;5:42. doi:10.1186/1471-2202-5-42.
11. Balduzzi D, Tononi G. Integrated information in discrete dynamical systems: motivation and theoretical framework. *PLoS Comput Biol*. 2008;4(6):e1000091. doi:10.1371/journal.pcbi.1000091.
12. Oizumi M, Albantakis L, Tononi G. From the phenomenology to the mechanisms of consciousness: integrated information theory 3.0. *PLoS Comput Biol*. 2014;10(5):e1003588. doi:10.1371/journal.pcbi.1003588.
13. Tononi G, Boly M, Massimini M, Koch C. Integrated information theory: from consciousness to its physical substrate. *Nature Reviews Neuroscience*. 2016;17(7).
14. Oizumi M, Tsuchiya N, Amari Si. Unified framework for information integration based on information geometry. *Proceedings of the National Academy of Sciences*. 2016;113(51):14817–14822.
15. Queyranne M. Minimizing symmetric submodular functions. *Mathematical Programming*. 1998;82(1-2):3–12.
16. Tegmark M. Improved measures of integrated information. *PLoS Computational Biology*. 2016;12(11):e1005123.
17. Toker D, Sommer F. Moving Past the Minimum Information Partition: How To Quickly and Accurately Calculate Integrated Information. *arXiv preprint arXiv:160501096*. 2016;.
18. Iwata S. Submodular function minimization. *Mathematical Programming*. 2008;112(1):45–64.

19. Cover TM, Thomas JA. Elements of Information Theory. 99th ed. Wiley-Interscience; 1991. Available from: <http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/04>
20. Shannon CE. A Mathematical Theory of Communication. The Bell System Technical Journal. 1948;27:379–423, 623–656.
21. Burnham KP, Anderson DR. Model selection and multimodel inference: a practical information-theoretic approach. Springer Science & Business Media; 2003.
22. Pearl J. Causality. Cambridge university press; 2009.
23. Watanabe S. Information theoretical analysis of multivariate correlation. IBM J Res Dev. 1960;4(1):66–82. doi:10.1147/rd.41.0066.
24. Studený J M & Vejnarová. In: Jordan MI, editor. The multiinformation function as a tool for measuring stochastic dependence. Cambridge, MA: MIT Press; 1999.
25. Kaneko K. Overview of coupled map lattices. Chaos: An Interdisciplinary Journal of Nonlinear Science. 1992;2(3):279–282.
26. Barrett AB, Seth AK. Practical measures of integrated information for time-series data. PLoS Comput Biol. 2011;7(1):e1001052. doi:10.1371/journal.pcbi.1001052.
27. Ay N. Information geometry on complexity and stochastic interaction. Entropy. 2015;17(4):2432–2458. doi:10.3390/e17042432.
28. Oizumi M, Amari S, Yanagawa T, Fujii N, Tsuchiya N. Measuring integrated information from the decoding perspective. PLoS Comput Biol. 2016;12(1):e1004654. doi:10.1371/journal.pcbi.1004654.