Breakdown of a concavity property of mutual information for non-Gaussian channels

Anastasia Kireeva* Jean-Christophe Mourrat†

Abstract

Let S and $\tilde{S}$ be two independent and identically distributed random variables, which we interpret as the signal, and let P_1 and P_2 be two communication channels. We can choose between two measurement scenarios: either we observe S through P_1 and P_2, and also $\tilde{S}$ through P_1 and P_2; or we observe S twice through P_1, and $\tilde{S}$ twice through P_2. In which of these two scenarios do we obtain the most information on the signal $(S,\tilde{S})$? While the first scenario always yields more information when P_1 and P_2 are additive Gaussian channels, we give examples showing that this property does not extend to arbitrary channels. As a consequence of this result, we show that the continuous-time mutual information arising in the setting of community detection on sparse stochastic block models is not concave, even in the limit of large system size. This stands in contrast to the case of models with diverging average degree, and brings additional challenges to the analysis of the asymptotic behavior of this quantity.

1 Introduction

Let P_S be a probability measure with finite support $\mathscr{S}$, and let S be a random variable sampled according to P_S, which we think of as a signal. A *communication channel* over $\mathscr{S}$, or more simply a *channel*, is a family of probability measures $(P(\cdot \mid s))_{s\in\mathscr{S}}$ over $\mathbb{R}^d$ for some integer $d \geqslant 1$, which we view as a conditional probability distribution over $\mathbb{R}^d$ given S. Let $f : \mathscr{S} \to \mathbb{R}^d$, and let W be a standard d-dimensional Gaussian random vector independent of S. The conditional law, given S, of the random variable

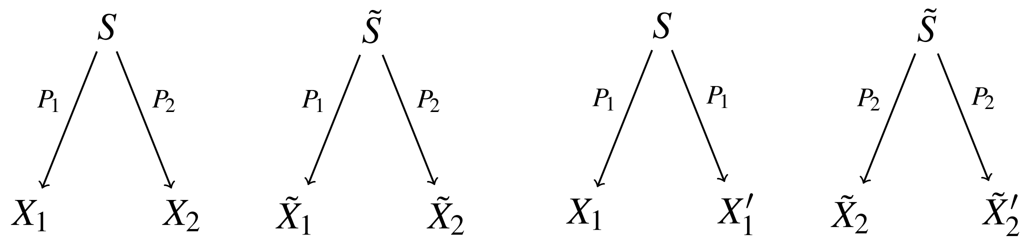
$$X := f(S) + W \tag{1.1}$$

defines a channel. We call any channel that can be constructed in this way a *Gaussian channel*. The information-theoretic quantities studied in this paper are invariant under bijective bimeasurable transformations of the channel output; in particular, there is no loss of generality in assuming the covariance matrix of the noise term in (1.1) to be the identity. For random variables X and Y defined on the same probability space, we denote by $I(X;Y)$ their mutual information, that is,

$$I(X;Y) := \mathbb{E}\left[\log\left(\frac{\mathrm{d}P_{(X,Y)}}{\mathrm{d}P_X \otimes \mathrm{d}P_Y}(X,Y)\right)\right],$$

*Department of Mathematics, ETH Zurich, anastasia.kireeva@ethz.ch

†Department of Mathematics, ENS Lyon and CNRS, jean-christophe.mourrat@ens-lyon.fr



(a) Scenario 1: We observe the signal and its independent copy twice through both channels P_1 and P_2.

(b) Scenario 2: We observe the signal twice through channel P_1 and its independent copy twice through channel P_2.

Figure 1: Does the scenario on the left side give us more information about $(S,\tilde{S})$ than the scenario on the right side?

where $P_{(X,Y)}$, P_X and P_Y are the laws of (X,Y), X and Y respectively.

Let P_1 and P_2 be two channels over $\mathscr{S}$. Conditionally on S, we sample X_1, X_1', X_2 and X_2' independently, with X_1, X_1' sampled according to $P_1(\cdot \mid S)$, and X_2, X_2' sampled according to $P_2(\cdot \mid S)$. We consider the following question.

$$\text{Do we have } \; I(S;(X_1,X_1')) + I(S;(X_2,X_2')) \leqslant 2I(S;(X_1,X_2)) \text{ ?} \tag{Q1}$$

A possibly more intuitive way to ask this question, following the phrasing in the abstract, is displayed in Figure 1, where we denote by $(\tilde{S},\tilde{X}_1,\tilde{X}_1',\tilde{X}_2,\tilde{X}_2')$ an independent copy of (S,X_1,X_1',X_2,X_2'). As will be seen below, the answer to this question is positive whenever P_1 and P_2 are Gaussian channels. However, we will show that the answer to this question is actually negative if P_1 and P_2 can be arbitrary channels. In fact, our counterexamples are even such that

$$\min\left(I(S;(X_1,X_1')), I(S;(X_2,X_2'))\right) > I(S;(X_1,X_2)).$$

While we find this question interesting on its own, we are also motivated by its implications in the context of community detection problems. We consider the setting of the stochastic block model [20, 24, 49, 50], sometimes also called the planted partition model [8, 9, 19] or the inhomogeneous random graph model [7]. In the case of two communities, this model is defined as follows. First, we independently attribute each individual to one of the two possible communities. Next, independently for each pair of individuals, we draw an edge between these two individuals with probability d_{in}/N if the two individuals belong to the same community, or with probability d_{out}/N if the two individuals belong to different communities, where N is the total number of individuals. We are then shown the resulting graph, but not the underlying community structure, which we aim to reconstruct. The choice of scaling for the link probabilities ensures that the average degree of a node remains bounded as N tends to infinity.

This problem has received considerable attention. An early contribution is the very inspiring work of [15], which relies on deep non-rigorous statistical physics arguments. In the case when the individuals are equally likely to belong to one or the other community, it was shown in [34, 37, 39] that one can recover meaningful information on the underlying community structure if and only if $(d_{\text{in}} - d_{\text{out}})^2 > 2(d_{\text{in}} + d_{\text{out}})$; and in this case, there exists an efficient algorithm for doing so.

A more refined question consists in studying the asymptotic behavior of the mutual information between the observed graph and the community structure, in the limit of large N. When $d_{\text{in}} < d_{\text{out}}$, this problem was resolved in [2, 14]. The case when $d_{\text{in}} > d_{\text{out}}$ is more challenging and was only resolved very recently in [51]; we also refer to [1, 27, 38, 40] for earlier work on this. The core of the argument of [51] is to show that there is a unique fixed point to a certain belief-propagation (BP) distributional recursion.

For situations with four or more communities, the problem becomes more complicated, and there exist choices of parameters for which this BP fixed-point equation admits more than one solution [21]. In these cases, a strategy in the spirit of that deployed in [51] therefore cannot be adapted in a straightforward way, and further work is necessary.

An alternative approach to the problem of identifying the asymptotic behavior of the mutual information between the observed graph and the community structure has been initiated in [17, 18]. The gist of the approach is to identify the limit mutual information as the solution to a certain partial differential equation (PDE). This technique allowed for the asymptotic analysis of the mutual information of a very large class of models involving Gaussian channels [10]; see also [11, 12, 13, 41, 42]. Using other approaches, a number of special cases had been solved earlier in [4, 5, 6, 26, 30, 31, 32, 33, 35, 36, 46, 47]. As shown in [3, 16, 30], a Gaussian equivalence property ensures that these results also allow us to identify the asymptotic behavior of the mutual information of the community detection problem in regimes in which the average degree of a node diverges with the system size.

In the approach taken up in [10, 18], one can leverage a certain regularity property of the mutual information to obtain a lower bound on the limit mutual information in terms of the solution to the PDE. This is similar to the results obtained in [43, 44] in the context of spin glasses. In order to show the matching upper bound, a central ingredient of the approach taken up in [10] is the observation that the mutual information is a *concave* function of the signal-to-noise ratios of the various observations considered for the resolution of the problem. For the community detection problem, if the mutual information studied in [18] happened to be concave in its parameters, we would be optimistic that the approach of [10] would be adaptable to this setting, and thus would allow us to obtain the matching upper bound. However, we show here that the mutual information is in fact *not* a concave function of its parameters. We find this surprising given that this concavity property does hold for the problems with Gaussian channels considered in [10] and elsewhere. We derive this breakdown of concavity as a consequence of the fact that the answer to Question Q1 is negative in general. Precisely, we will show that, although the Hessian of the mutual information only contains nonpositive entries, we can witness a breakdown of concavity that scales as $(d_{\text{in}} - d_{\text{out}})^6$ in the regime of small $|d_{\text{in}} - d_{\text{out}}|$. We are also surprised by the relatively high exponent 6 appearing here, suggesting a rather subtle deviation from concavity in the regime of small $|d_{\text{in}} - d_{\text{out}}|$.

Had the mutual information been concave in its parameters, we would presumably have been able to represent the solution to the relevant PDE as a saddle-point variational problem, using a version of the Hopf formula (see [10], and also [17] for a proof of a related variational formula under different assumptions). Given that this concavity property is in fact invalid, we tend to think that there will not be any reasonable way to represent the limit mutual information of community detection as a variational problem, unlike the situation with Gaussian channels. In the context of spin glasses, this point is discussed more precisely in [43, Section 6].

The rest of the paper is organized as follows. In Section 2, we show that the answer to Question Q1 is positive for Gaussian channels, and construct counterexamples in general. We pay special attention to the case of Bernoulli channels with very low signal-to-noise ratios, as these examples will be fundamental to subsequent considerations concerning the community detection problem. In Section 3, we focus on Gaussian channels and explore variants of the inequality appearing in Question Q1 that involve more than two channels. In Section 4, we turn to the setting of community detection, for the stochastic block model with two communities. We use the results of Section 2 to show that the mutual information is not a concave function of its parameters, even after we pass to the limit of large system size.

2 Answers to Question Q1

We start by providing a positive answer to Question Q1 in the case of Gaussian channels.

Proposition 2.1 (Mixing Gaussian channels yields more information)**.** *If P_1 and P_2 are Gaussian channels, then the answer to Question Q1 is positive.*

Proof. The proof of Proposition 2.1 is based on remarkable identities involving derivatives of the mutual information with respect to the signal-to-noise ratio. In particular, the first-order derivative of the mutual information is half of the minimal mean-square error, as was explained in [22] and extended to the matrix case in [29, 45, 48]. Here we will rely on the calculation of second-order derivatives of the mutual information, which already appeared in [23, 29, 45].

By definition of Gaussian channels, for each $i \in \{1,2\}$, there exists a mapping $f_i : \mathscr{S} \to \mathbb{R}^{d_i}$ such that the channel P_i can be represented as

$$S \mapsto f_i(S) + W_i,$$

where W_1, W_2 are independent standard Gaussians, independent of S, of dimension d_1 and d_2 respectively. For every $i \in \{1,2\}$ and $t_i \geqslant 0$, we define

$$X_i(t_i) := \sqrt{t_i} f_i(S) + W_i,$$

as well as

$$\mathscr{I}(t_1,t_2) := I(S;(X_1(t_1),X_2(t_2))).$$

Since the mapping $s \mapsto (s, f_1(s), f_2(s))$ is injective, we have

$$\mathscr{I}(t_1,t_2) = I((S, f_1(S), f_2(S)); (X_1(t_1), X_2(t_2))).$$

We can therefore replace the signal S by $(S, f_1(S), f_2(S))$ if desired, and apply [29, Theorem 3] or [45, Theorem 5] with H chosen to be the identity matrix and P chosen to be a diagonal matrix with d_1 entries at $\sqrt{t_1}$ and d_2 entries at $\sqrt{t_2}$. The conclusion of these theorems is that the function $\mathscr{I}$ is jointly concave in (t_1, t_2). In particular,

$$\mathscr{I}(2,0) + \mathscr{I}(0,2) \leqslant 2\mathscr{I}(1,1).$$

Recalling that

$$\mathscr{I}(1,1) = I(S;(X_1(1), X_2(1))) = I(S;(X_1, X_2)),$$

Proposition 2.1 will be proved once we verify that

$$\mathscr{I}(2,0) = I(S;(X_1, X_1')) \quad \text{and} \quad \mathscr{I}(0,2) = I(S;(X_2, X_2')). \tag{2.1}$$

We fix $i \in \{1,2\}$, let W_i' be a d_i-dimensional standard Gaussian independent of (S, W_i), and use it to represent X_i' as

$$X_i' = f_i(S) + W_i'.$$

We define

$$Z_i := \frac{X_i + X_i'}{\sqrt{2}} = \sqrt{2t} f_i(S) + \frac{W_i + W_i'}{\sqrt{2}},$$

and

$$D_i := X_i - X_i' = W_i - W_i'.$$

Using that the the map $(x,y) \mapsto ((x+y)/\sqrt{2}, x-y)$ is bijective and the chain rule, we can write

$$I(S;(X_i, X_i')) = I(S;(Z_i, D_i)) = I(S;D_i) + I(S;Z_i \mid D_i).$$

The random variables S and D_i being independent, the first term on the right side of this identity vanishes. We also observe that the pair (W_i, W_i') is independent of S, and moreover, the Gaussian random variables $W_i + W_i'$ and $W_i - W_i'$ are independent. This implies that the random variables $(S, W_i + W_i', W_i - W_i')$ are independent, and thus that D_i is independent of the pair (S, Z_i). The previous display therefore simplifies into

$$I(S;(X_i, X_i')) = I(S;Z_i).$$

Since the pairs $(S, X_i(2))$ and (S, Z_i) have the same law, this is (2.1). $\square$

We now turn to showing that Proposition 2.1 does not generalize to non-Gaussian channels. Before doing so, we record a simple observation allowing to simplify the question somewhat.

Lemma 2.2. *Let S be a random variable with finite support $\mathscr{S}$, let P_1, P_2 be two communication channels over $\mathscr{S}$, and conditionally on S, let (X_1, X_1', X_2, X_2') be independent random variables, with X_1, X_1' sampled according to $P_1(\cdot \mid S)$ and X_2, X_2' sampled according to $P_2(\cdot \mid S)$. For every $i, j \in \{1,2\}$, we have*

$$I(S;(X_i;X_j')) = I(S;X_i) + I(S;X_j) - I(X_i;X_j'). \tag{2.2}$$

Proof. By the chain rule,

$$\begin{aligned} I(S;(X_i,X'_j)) &= I(S;X_i) + I(S;X'_j \mid X_i) \\ &= I(S;X_i) + I(X'_j;(S,X_i)) - I(X_i;X'_j) \\ &= I(S;X_i) + I(S;X'_j) + I(X_i;X'_j \mid S) - I(X_i;X'_j). \end{aligned}$$

Conditionally on S, the random variables X_i and X'_j are independent. It thus follows that $I(X_i,X'_j \mid S) = 0$, and we obtain (2.2). □

A direct consequence of Lemma 2.2 is that

$$2I(S;(X_1,X_2)) - I(S;(X_1,X'_1)) - I(S;(X_2,X'_2)) = I(X_1,X'_1) + I(X_2,X'_2) - 2I(X_1,X_2),$$

And in particular, Question Q1 can be rephrased as:

$$\text{Do we have } \; 2I(X_1,X_2) \leqslant I(X_1,X'_1) + I(X_2,X'_2) \text{ ?} \tag{Q2}$$

For every $p \in [0,1]$, we write $\mathrm{Ber}(p) := p\delta_1 + (1-p)\delta_0$ for the law of a Bernoulli random variable of parameter p. For our counterexamples, we assume that S is a $\mathrm{Ber}(1/2)$ random variable, and we consider channels of the following form, for different choices of $p_0, p_1, q_0, q_1 \in [0,1]$:

$$P_1(\cdot \mid s) = \mathrm{Ber}(p_s) \quad \text{and} \quad P_2(\cdot \mid s) = \mathrm{Ber}(q_s) \qquad (s \in \{0,1\}).$$

Already for the choice of $p_0 = 1/2$, $p_1 = 0$, $q_0 = 0$, $q_1 = 1/2$, we find that

$$I(X_1;X_2) = \frac{5}{2}\log(2) - \frac{3}{2}\log(3) \simeq 0.0849,$$

while

$$I(X_1;X'_1) = I(X_2;X'_2) = \log(2) + \frac{5}{8}\log(5) - \frac{3}{2}\log(3) \simeq 0.0511.$$

In particular, this leads to a counterexample to the inequalities appearing in Questions Q1 and Q2. In fact, for Bernoulli channels with $p_0 = q_1$ and $p_1 = q_0$, we observe numerically that there are large regions of values of p_0, p_1 where the inequality in Question Q2 does not hold—in fact, probably all values with $p_0 \neq p_1$, see Figure 2. Perhaps surprisingly in view of Proposition 2.1 and its proof, the next proposition shows that the inequalities in Questions Q1 and Q2 can be violated even in regimes of small signal-to-noise ratio. This class of examples will be particularly relevant in the context of community detection discussed later.

Proposition 2.3. *Assume that S is a $\mathrm{Ber}(1/2)$ random variable, that we sample X_1, X'_1 according to $P_1(\cdot \mid S)$ and X_2, X'_2 according to $P_2(\cdot \mid S)$, where P_1 and P_2 are such that, for some $p_0, p_1, q_0, q_1 \geqslant 0$ and $\varepsilon > 0$,*

$$P_1(\cdot \mid s) = \mathrm{Ber}(\varepsilon p_s) \quad \textit{and} \quad P_2(\cdot \mid s) = \mathrm{Ber}(\varepsilon q_s) \qquad (s \in \{0,1\}).$$

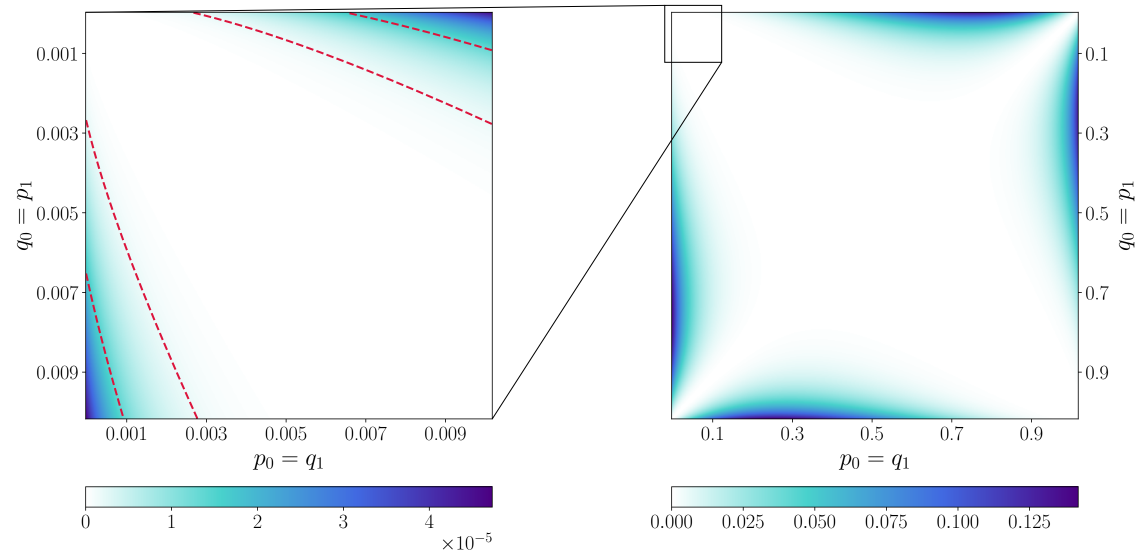


Figure 2: Value of the $2I(X_1,X_2) - I(X_1,X_1') - I(X_2,X_2')$ for the setting of Bernoulli channels (see description in the text). The value is represented by color, the larger values correspond to darker color. Left: the regime of small p_0, p_1. Red dashed lines are contour lines of $(p_0-p_1)^6/(p_0+p_1)^4$. Right: general $p_0, p_1 \in [0,1]$.

If $p_0 = q_1$, $p_1 = q_0$, then

$$2I(X_1,X_2) - I(X_1,X_1') - I(X_2,X_2') \geqslant \frac{\varepsilon^2(p_0-p_1)^6}{6(p_0+p_1)^4} + o(\varepsilon^2) \qquad (\varepsilon \to 0).$$

In particular, the inequalities in Questions Q1 and Q2 are false whenever $p_0 \neq p_1$ and $\varepsilon > 0$ is sufficiently small.

Proof. For unconstrained values of $p_0, p_1, q_0, q_1 \geqslant 0$, the definition of mutual information yields that

$$\begin{aligned} I(X_1,X_2) = \log &\left(\frac{\frac{1}{2}(1-\varepsilon p_0)(1-\varepsilon q_0) + \frac{1}{2}(1-\varepsilon p_1)(1-\varepsilon q_1)}{\frac{1}{4}(2-\varepsilon p_0 - \varepsilon p_1)(2-\varepsilon q_0 - \varepsilon q_1)} \right) \\ &\cdot \frac{1}{2}\left((1-\varepsilon p_0)(1-\varepsilon q_0) + (1-\varepsilon p_1)(1-\varepsilon q_1)\right) \\ &+ \log\left(\frac{\frac{1}{2}\varepsilon p_0(1-\varepsilon q_0) + \frac{1}{2}\varepsilon p_1(1-\varepsilon q_1)}{\frac{1}{4}\varepsilon(p_0+p_1)(2-\varepsilon q_0-\varepsilon q_1)} \right) \cdot \frac{1}{2}\varepsilon(p_0(1-\varepsilon q_0) + p_1(1-\varepsilon q_1)) \\ &+ \left(\frac{\frac{1}{2}\varepsilon q_0(1-\varepsilon p_0) + \frac{1}{2}\varepsilon q_1(1-\varepsilon p_1)}{\frac{1}{4}\varepsilon(2-\varepsilon p_0-\varepsilon p_1)(q_0+q_1)} \right) \cdot \frac{1}{2}\varepsilon(q_0(1-\varepsilon p_0) + q_1(1-\varepsilon p_1)) \\ &+ \log\left(\frac{\frac{1}{2}\varepsilon^2 p_0 q_0 + \frac{1}{2}\varepsilon^2 p_1 q_1}{\frac{1}{4}\varepsilon^2(p_0+p_1)(q_0+q_1)} \right) \cdot \frac{1}{2}\varepsilon^2(p_0 q_0 + p_1 q_1) \\ =: I_1 + I_2 + I_3 + I_4. \end{aligned}$$

The Taylor expansions of these terms are given by

$$\begin{aligned}
I_1 &= \frac{1}{4}\varepsilon^2(p_0-p_1)(q_0-q_1)+o(\varepsilon^2),\\
I_2 = I_3 &= -\frac{1}{4}\varepsilon^2(p_0-p_1)(q_0-q_1)+o(\varepsilon^2),\\
I_4 &= \frac{1}{2}\varepsilon^2(p_0q_0+p_1q_1)\log\left(\frac{2\,(p_0q_0+p_1q_1)}{(p_0+p_1)\,(q_0+q_1)}\right)+o(\varepsilon^2).
\end{aligned}$$

Hence we obtain

$$\begin{aligned}
I(X_1,X_2) = &-\frac{1}{4}\varepsilon^2(p_0-p_1)(q_0-q_1)\\
&+\frac{1}{2}\varepsilon^2\log\left(1+\frac{(p_0-p_1)(q_0-q_1)}{(p_0+p_1)(q_0+q_1)}\right)(p_0q_0+p_1q_1)+o(\varepsilon^2).
\end{aligned}$$

In the case when $p_0=q_0$ and $p_1=q_1$, we get

$$\begin{aligned}
I(X_1,X_1') &= I(X_2,X_2')\\
&= -\frac{1}{4}\varepsilon^2(p_0-p_1)^2+\frac{1}{2}\log\left(1+\frac{(p_0-p_1)^2}{(p_0+p_1)^2}\right)(p_0^2+p_1^2)+o(\varepsilon^2),
\end{aligned}$$

while in the considered case when $p_0=q_1$ and $p_1=q_0$,

$$I(X_1,X_2) = \frac{1}{4}\varepsilon^2(p_0-p_1)^2+p_0p_1\log\left(1-\frac{(p_0-p_1)^2}{(p_0+p_1)^2}\right)+o(\varepsilon^2).$$

Rescaling the mutual information by ε^2, we obtain that

$$\begin{aligned}
&\frac{1}{\varepsilon^2}\left(2I(X_1,X_2)-I(X_1,X_1')-I(X_2,X_2')\right)\\
&\quad= (p_0-p_1)^2+2p_0p_1\log\left(1-\frac{(p_0-p_1)^2}{(p_0+p_1)^2}\right)\\
&\qquad-(p_0^2+p_1^2)\log\left(1+\frac{(p_0-p_1)^2}{(p_0+p_1)^2}\right)+o(1)
\end{aligned}\tag{2.3}$$

Denoting $t := {}^{(p_0-p_1)^2}/_{(p_0+p_1)^2}\in[0,1]$, we can rewrite the above identity as

$$\begin{aligned}
&\frac{1}{\varepsilon^2(p_0+p_1)^2}\left(2I(X_1,X_2)-I(X_1,X_1')-I(X_2,X_2')\right)+o(1)\\
&\quad= t+\frac{1-t}{2}\log(1-t)-\frac{1+t}{2}\log(1+t) =: g(t).
\end{aligned}$$

We observe that $g(0)=0$, or equivalently, this difference is zero when $p_0=p_1$. We compute the derivative of g with respect to t and obtain that

$$g'(t) = -\frac{1}{2}(\log(1-t)+\log(1+t)) = -\frac{1}{2}\log(1-t^2)\geqslant\frac{t^2}{2},$$

so $g(t) \geqslant \frac{t^3}{6}$ for every $t \in [0,1]$. Substituting back t yields that

$$2I(X_1,X_2) - I(X_1,X_1') - I(X_2,X_2') \geqslant \frac{\varepsilon^2(p_0-p_1)^6}{6(p_0+p_1)^4} + o(\varepsilon^2),$$

as desired. $\square$

3 Positive semidefinite kernels in the Gaussian case

For arbitrary random variables $(Z_i)_{1\leqslant i\leqslant n}$, one may wonder whether $(I(Z_i;Z_j))_{1\leqslant i,j\leqslant n}$ is a positive semidefinite matrix; a negative answer to this question was provided in [25]. In our setting, consider multiple channels $(P_1,\ldots,P_n)$ over $\mathscr{S}$, and conditionally on S, denote by $(X_i,X_i')_{1\leqslant i\leqslant n}$ conditionally independent random variables with X_i, X_i' distributed according to $P_i(\cdot \mid S)$. One could ask:

$$\text{Is the matrix } (I(X_i;X_j'))_{1\leqslant i,j\leqslant n} \text{ positive semidefinite?} \tag{Q3}$$

We find that this is a natural question on its own; as will be seen below, it also arises naturally in the study of the continuous-time mutual information discussed below in relation with the problem of community detection. If the answer to Question Q3 were positive, then it would mean that the mapping $(P_i,P_j)\mapsto I(X_i;X_j')$ defines a positive semidefinite kernel over the space of channels. Notice that Question Q2 can be rephrased as

$$\text{Do we have } \begin{pmatrix}1\\-1\end{pmatrix}\cdot\begin{pmatrix}I(X_1;X_1') & I(X_1;X_2')\\ I(X_2;X_1') & I(X_2;X_2')\end{pmatrix}\begin{pmatrix}1\\-1\end{pmatrix} \geqslant 0\ ?$$

Since we identified examples for which the inequality in Question Q2 is violated, it follows that the answer to Question Q3 is also negative in general. We do not know whether the answer to Question Q3 is positive for Gaussian channels. Roughly speaking, the next proposition states that the answer to Question Q3 is positive for Gaussian channels in the low signal-to-noise regime.

Proposition 3.1 (psd kernel for Gaussian channels)**.** *Let $n\geqslant 1$ be an integer. For every $i\in\{1,\ldots,n\}$, let $f_i:\mathscr{S}\to\mathbb{R}^{d_i}$, and let $(W_i,W_i')_{1\leqslant i\leqslant n}$ be independent standard Gaussian random vectors, independent of the signal S, with W_i and W_i' of dimension d_i. For every $i\in\{1,\ldots,n\}$ and $t\geqslant 0$, we define*

$$X_i(t) := \sqrt{t} f_i(S) + W_i \quad \text{and} \quad X_i'(t) := \sqrt{t} f_i(S) + W_i'.$$

For every $i,j\in\{1,\ldots,n\}$, we have

$$\lim_{t\to 0} t^{-2} I(X_i(t);X_j'(t)) = \left|\mathbb{E}\left[(f_i(S)-\mathbb{E}[f_i(S)])(f_j(S)-\mathbb{E}[f_j(S)])^*\right]\right|^2, \tag{3.1}$$

where the superscript * denotes the transpose operator, and the norm $|\cdot|$ over matrices is the Frobenius norm. Moreover, the matrix

$$\left(\left|\mathbb{E}\left[(f_i(S)-\mathbb{E}[f_i(S)])(f_j(S)-\mathbb{E}[f_j(S)])^*\right]\right|^2\right)_{1\leqslant i,j\leqslant n} \tag{3.2}$$

is positive semidefinite.

Proof. The proof is again based on the fundamental identities derived in [22, 29, 45]. In order to ligthen the notation, we define, for every $i \in \{1, \ldots, n\}$ and $s \in \mathscr{S}$,

$$\overline{f}_i(s) := f_i(s) - \mathbb{E}[f_i(S)].$$

Recalling that we assume the state space $\mathscr{S}$ of the signal S to be finite, one can check that the mapping $t \mapsto I(S; X_a(t))$ is infinitely differentiable. The I-MMSE relation from [22] yields that

$$\partial_t I(S; X_i(t))_{|t=0} = \frac{1}{2} \mathbb{E}\left[\left| \overline{f}_i(S) \right|^2 \right], \tag{3.3}$$

while [45, Theorem 5] or the proof of [29, Theorem 3] imply that

$$\partial_t^2 I(S; X_i(t))_{|t=0} = \frac{1}{2} \left| \mathbb{E}\left[\overline{f}_i(S) \overline{f}_i(S)^* \right] \right|^2. \tag{3.4}$$

Since the choice of f_i is arbitrary, the identities (3.3) and (3.4) also imply that

$$\partial_t I(S; (X_i(t), X_j(t)))_{|t=0} = \frac{1}{2} \mathbb{E}\left[\left| \overline{f}_i(S) \right|^2 \right] + \frac{1}{2} \mathbb{E}\left[\left| \overline{f}_j(S) \right|^2 \right], \tag{3.5}$$

and

$$\begin{aligned} &\partial_t^2 I(S; (X_i(t), X_j(t)))_{|t=0} \\ &\qquad = \frac{1}{2} \left| \mathbb{E}\left[\overline{f}_i(S) \overline{f}_i(S)^* \right] \right|^2 + \frac{1}{2} \left| \mathbb{E}\left[\overline{f}_j(S) \overline{f}_j(S)^* \right] \right|^2 + \left| \mathbb{E}\left[\overline{f}_i(S) \overline{f}_j(S)^* \right] \right|^2. \end{aligned} \tag{3.6}$$

By Lemma 2.2, we have that

$$I(X_i(t); X_j'(t)) = I(S; (X_i(t); X_j'(t))) - I(S; X_i(t)) - I(S; X_j(t)).$$

A Taylor expansion near $t = 0$ of this identity, combined with the expressions of the derivatives obtained above, therefore yields (3.1). To see that the matrix in (3.2) is positive semidefinite, let us denote by $\tilde{S}$ an independent copy of the random variable S. Writing $\cdot$ for the entrywise scalar product between vectors or matrices, we have for every $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$ that

$$\begin{aligned} \sum_{i,j=1}^n \alpha_i \alpha_j \left| \mathbb{E}\left[\overline{f}_i(S) \overline{f}_j(S)^* \right] \right|^2 &= \sum_{i,j=1}^n \alpha_i \alpha_j \mathbb{E}\left[\left(\overline{f}_i(S) \overline{f}_j(S)^* \right) \cdot \left(\overline{f}_i(\tilde{S}) \overline{f}_j(\tilde{S})^* \right) \right] \\ &= \sum_{i,j=1}^n \alpha_i \alpha_j \mathbb{E}\left[\left(\overline{f}_i(S) \cdot \overline{f}_i(\tilde{S}) \right) \left(\overline{f}_j(S) \cdot \overline{f}_j(\tilde{S}) \right) \right] \\ &= \mathbb{E}\left[\left(\sum_{i=1}^n \alpha_i \overline{f}_i(S) \cdot \overline{f}_i(\tilde{S}) \right)^2 \right] \geqslant 0. \end{aligned}$$

This completes the proof of Proposition 3.1. □

4 Consequences for community detection

Our initial motivation for exploring questions such as Q1 comes from the study of the mutual information of a problem of community detection in the stochastic block model. In the notation of [18], we specialize to the choice of parameters $p = \frac{1}{2}$, $t = 0$, $\mu = t_1\delta_1 + t_2\delta_{-1}$, with $t_1, t_2 \geqslant 0$, so that the mutual information considered there simplifies and matches the assumptions of Proposition 2.3, as we explain now. First, we sample S as a Bernoulli random variable with parameter $1/2$ (this is one coordinate of σ^* in the notation of [18], except that we reparametrize this random variable taking values $\{-1,1\}$ into S taking values in $\{0,1\}$ for notational consistency). Conditionally on S, we let $(X_1^{(\ell)}, X_2^{(\ell)})_{\ell \geqslant 1}$ be independent random variables, with $X_1^{(\ell)}$ sampled according to P_1 and $X_2^{(\ell)}$ sampled according to P_2, where the channels P_1 and P_2 are defined by

$$P_1(\cdot \mid s) = \mathrm{Ber}(p_s/N) \quad \text{and} \quad P_2(\cdot \mid s) = \mathrm{Ber}(q_s/N) \qquad (s \in \{0,1\}), \tag{4.1}$$

and $p_0, p_1, q_0, q_1 \in [0,\infty)$ are such that $p_0 = q_1$ and $p_1 = q_0$ (in the notation of [18], we have $p_1 = q_0 = c + \Delta$, and $p_0 = q_1 = c - \Delta$, with the identification that $\sigma^* = 1$ and -1 correspond to $S = 1$ and 0 respectively). While we will not always say it explicitly, we always understand that N is taken sufficiently large that the quantities p_s/N and q_s/N appearing in (4.1) belong to the interval $[0,1]$. Finally, we let $\Pi^{(1)}_{Nt_1}$ and $\Pi^{(2)}_{Nt_2}$ be two independent Poisson random variables of parameters Nt_1 and Nt_2 respectively, independent of the all other random variables. With all these choices, and using the Poisson coloring theorem (see for instance [28, Chapter 5]) we get that the mutual information studied in [18] simplifies into

$$\mathscr{I}_N(t_1,t_2) := I\left(S; \left((X_1^{(\ell)})_{\ell \leqslant \Pi^{(1)}_{Nt_1}}, (X_2^{(\ell)})_{\ell \leqslant \Pi^{(2)}_{Nt_2}}\right)\right).$$

Although this is not apparent in the notation, we emphasize that the laws of $X_1^{(\ell)}$ and $X_2^{(\ell)}$ depend on N. As shown in [18, Lemma 3.1], the function $\mathscr{I}_N$ converges pointwise to a limit, which we denote by $\mathscr{I}_\infty$.

Proposition 4.1 (Breakdown of concavity of mutual information)**.** *For every $N \in \mathbb{N} \cup \{\infty\}$, the entries of the Hessian of the mapping $(t_1,t_2) \mapsto \mathscr{I}_N(t_1,t_2)$ are nonpositive. However, in the regime of finite N going to infinity, we have*

$$\left(\partial_{t_1}^2 \mathscr{I}_N + \partial_{t_2}^2 \mathscr{I}_N - 2\partial_{t_1}\partial_{t_2}\mathscr{I}_N\right)(0,0) \geqslant \frac{(p_0-p_1)^6}{6(p_0+p_1)^4} + o(1) \tag{4.2}$$

as well as

$$\left(\partial_{t_1}^2 \mathscr{I}_\infty + \partial_{t_2}^2 \mathscr{I}_\infty - 2\partial_{t_1}\partial_{t_2}\mathscr{I}_\infty\right)(0,0) \geqslant \frac{(p_0-p_1)^6}{6(p_0+p_1)^4}. \tag{4.3}$$

In particular, for every sufficiently large $N \in \mathbb{N} \cup \{\infty\}$, the mapping $(t_1,t_2) \mapsto \mathscr{I}_N(t_1,t_2)$ is not concave.

Proof. We decompose the proof into four steps.

Step 1. In this step, we derive convenient representations for the second derivatives of $\mathscr{I}_N$, for finite N. For every $t \geqslant 0$ and $L \in \mathbb{Z}_+$, we denote

$$\pi(t,L) := e^{-t}\frac{t^L}{L!}.$$

With the understanding that $\pi(t,-1) = 0$, we have the identity

$$\partial_t \pi(t,L) = \pi(t,L-1) - \pi(t,L). \tag{4.4}$$

In order to lighten the calculations, we also introduce the shorthand notation

$$I_N(L_1,L_2) = I\left(S;\left((X_1^{(\ell)})_{\ell \leqslant L_1},(X_2^{(\ell)})_{\ell \leqslant L_2}\right)\right).$$

We start by observing that

$$\mathscr{I}_N(t_1,t_2) = \sum_{L_1,L_2=0}^{+\infty} \pi(Nt_1,L_1)\pi(Nt_2,L_2)I_N(L_1,L_2).$$

The identity (4.4) yields that

$$\partial_{t_1}\mathscr{I}_N(t_1,t_2) = N\sum_{L_1,L_2=0}^{+\infty} \pi(Nt_1,L_1)\pi(Nt_2,L_2)\left(I_N(L_1+1,L_2) - I_N(L_1,L_2)\right),$$

and thus

$$\partial_{t_1}^2\mathscr{I}_N(t_1,t_2) = N^2\sum_{L_1,L_2=0}^{+\infty} \pi(Nt_1,L_1)\pi(Nt_2,L_2) \\ \left(I_N(L_1+2,L_2) - 2I_N(L_1+1,L_2) + I_N(L_1,L_2)\right). \tag{4.5}$$

A similar expression can be obtained for $\partial_{t_2}^2\mathscr{I}_N$, with the finite-difference operation acting on the variable L_2 in place of L_1. The cross-derivative takes the form

$$\partial_{t_1}\partial_{t_2}\mathscr{I}_N(t_1,t_2) = N^2\sum_{L_1,L_2=0}^{+\infty} \pi(Nt_1,L_1)\pi(Nt_2,L_2) \\ \left(I_N(L_1+1,L_2+1) - I_N(L_1+1,L_2) - I_N(L_1,L_2+1) + I_N(L_1,L_2)\right). \tag{4.6}$$

Step 2. In this step, we show that the entries of the Hessian of $\mathscr{I}_N$ are nonpositive. Since this property can be understood in a weak sense, or in terms of the signs of certain finite differences, it suffices to show its validity for finite N. From the expressions of the second derivatives obtained in the previous step, we see that it suffices to show that, for every $N \in \mathbb{N}, L_1, L_2 \in \mathbb{Z}_+$,

$$I_N(L_1+2,L_2) - 2I_N(L_1+1,L_2) + I_N(L_1,L_2) \leqslant 0, \tag{4.7}$$

$$I_N(L_1,L_2+2)-2I_N(L_1,L_2+1)+I_N(L_1,L_2)\leqslant 0, \tag{4.8}$$

and

$$I_N(L_1+1,L_2+1)-I_N(L_1+1,L_2)-I_N(L_1,L_2+1)+I_N(L_1,L_2)\leqslant 0. \tag{4.9}$$

We only show the validity of (4.9), the arguments for (4.7) and (4.8) being similar. In order to lighten the notation, we write

$$Z := \left((X_1^{(\ell)})_{\ell\leqslant L_1},(X_2^{(\ell)})_{\ell\leqslant L_2}\right).$$

By the chain rule for mutual information, we have

$$\begin{aligned}I_N(L_1+1,L_2+1) &= I\left(S;\left((X_1^{(\ell)})_{\ell\leqslant L_1+1},(X_2^{(\ell)})_{\ell\leqslant L_2+1}\right)\right)\\ &= I\left(S;\left(X_1^{(L_1+1)},X_2^{(L_2+1)}\right)\mid Z\right)+I(S;Z),\end{aligned}$$

and similarly,

$$I_N(L_1+1,L_2)=I\left(S;X_1^{(L_1+1)}\mid Z\right)+I(S;Z),$$

and

$$I_N(L_1,L_2+1)=I\left(S;X_2^{(L_2+1)}\mid Z\right)+I(S;Z).$$

Showing (4.9) is thus equivalent to showing that

$$I\left(S;\left(X_1^{(L_1+1)},X_2^{(L_2+1)}\right)\mid Z\right)-I\left(S;X_1^{(L_1+1)}\mid Z\right)-I\left(S;X_2^{(L_2+1)}\mid Z\right)\leqslant 0. \tag{4.10}$$

We use again the chain rule of mutual information to write

$$I\left(S;\left(X_1^{(L_1+1)},X_2^{(L_2+1)}\right)\mid Z\right)=I\left(S;X_1^{(L_1+1)}\mid Z\right)+I\left(S;X_2^{(L_2+1)}\mid X_1^{(L_1+1)},Z\right).$$

The last term of the identity above can be rewritten as

$$\begin{aligned}&I\left(X_2^{(L_2+1)};\left(S,X_1^{(L_1+1)}\right)\mid Z\right)-I\left(X_2^{(L_2+1)};X_1^{(L_1+1)}\mid Z\right)\\ &\quad=I\left(S;X_2^{(L_2+1)}\mid Z\right)+I\left(X_1^{(L_1+1)};X_2^{(L_2+1)}\mid S,Z\right)-I\left(X_1^{(L_1+1)};X_2^{(L_2+1)}\mid Z\right).\end{aligned} \tag{4.11}$$

Conditionally on S, the random variables $(X_1^{(L_1+1)},X_2^{(L_2+1)},Z)$ are independent, and thus the second term on the right side of (4.11) is zero. Combining these identities, we obtain that the left side of (4.10) equals $-I\left(X_1^{(L_1+1)};X_2^{(L_2+1)}\mid Z\right)$, which is indeed nonpositive.

Step 3. In this step, we show the validity of (4.2), and thus deduce the non-concavity of $\mathscr{I}_N$ for every N sufficiently large and finite. Using the expressions for the second derivative obtained in (4.5) and (4.6), we can write

$$\begin{aligned}&\begin{pmatrix}1\\-1\end{pmatrix}\cdot\begin{pmatrix}\partial_{t_1}^2\mathscr{I}_N(0,0) & \partial_{t_1}\partial_{t_2}\mathscr{I}_N(0,0)\\ \partial_{t_1}\partial_{t_2}\mathscr{I}_N(0,0) & \partial_{t_2}^2\mathscr{I}_N(0,0)\end{pmatrix}\begin{pmatrix}1\\-1\end{pmatrix}\\ &\quad=N^2\left[I(S;(X_1^{(1)},X_1^{(2)}))+I(S;(X_2^{(1)},X_2^{(2)}))-2I(S;(X_1^{(1)},X_2^{(1)}))\right]\\ &\quad=N^2\left[2I(X_1^{(1)};X_2^{(1)})-I(X_1^{(1)};X_1^{(2)})-I(X_2^{(1)};X_2^{(2)})\right],\end{aligned}$$

where we used Lemma 2.2 in the last step. Proposition 2.3 ensures that, for finite N going to infinity, we have

$$N^2\left[2I(X_1^{(1)};X_2^{(1)})-I(X_1^{(1)};X_1^{(2)})-I(X_2^{(1)};X_2^{(2)})\right]\geqslant\frac{(p_0-p_1)^6}{6(p_0+p_1)^4}+o(1),$$

which gives the desired result.

Step 4. In this last step, we show the validity of (4.3). Instead of trying to justify that the second derivatives of $\mathscr{I}_N$ converge to those of $\mathscr{I}_\infty$, we simply borrow from [18] an explicit expression for $\mathscr{I}_\infty$, and observe that it satisfies (4.3) by calculating its derivatives. We recall that $p_1=q_0$ corresponds to $c+\Delta$ in the notation of [18], while $p_0=q_1$ corresponds to $c-\Delta$ in the notation of [18]. The statement of [18, Lemma 3.1] involves two Poisson point processes, denoted by Π_+ and Π_- there, and which in our present context can be represented as $\Pi^{(1)}_{p_1t_1}\delta_1+\Pi^{(2)}_{p_0t_2}\delta_{-1}$ and $\Pi^{(1)}_{p_0t_1}\delta_1+\Pi^{(2)}_{p_1t_2}\delta_{-1}$ respectively. The quantity $\mu[-1,1]\mathbb{E}x_1$ appearing in [18, Lemma 3.1] translates into t_1-t_2 in our context. The function that is denoted by $\psi(\mu)$ in the notation of [18, Lemma 3.1] becomes, in our current setting, the function

$$\begin{aligned}\psi(t_1,t_2):=&-(t_1+t_2)\frac{p_1+p_0}{2}\\&+\frac{1}{2}\mathbb{E}\log\left[\frac{1}{2}e^{-\frac{(t_1-t_2)(p_1-p_0)}{2}}p_1^{\Pi^{(1)}_{p_1t_1}}p_0^{\Pi^{(2)}_{p_0t_2}}+\frac{1}{2}e^{\frac{(t_1-t_2)(p_1-p_0)}{2}}p_0^{\Pi^{(1)}_{p_1t_1}}p_1^{\Pi^{(2)}_{p_0t_2}}\right]\\&+\frac{1}{2}\mathbb{E}\log\left[\frac{1}{2}e^{-\frac{(t_1-t_2)(p_1-p_0)}{2}}p_1^{\Pi^{(1)}_{p_0t_1}}p_0^{\Pi^{(2)}_{p_1t_2}}+\frac{1}{2}e^{\frac{(t_1-t_2)(p_1-p_0)}{2}}p_0^{\Pi^{(1)}_{p_0t_1}}p_1^{\Pi^{(2)}_{p_1t_2}}\right].\end{aligned}$$

Arguing as for [18, (1.16)-(1.17)], one can check that the mutual information $\mathscr{I}_N(t_1,t_2)$ is obtained as a simple (and convergent as $N\to\infty$) linear term in (t_1,t_2), minus a function, denoted by $\psi_N(\mu)$ in the notation of [18, Lemma 3.1], that converges to $\psi(t_1,t_2)$. In order to show that the mapping $(t_1,t_2)\mapsto\mathscr{I}_\infty(t_1,t_2)$ is not concave, it thus suffices to show that the mapping

$$\begin{aligned}(t_1,t_2)\mapsto\phi(t_1,t_2):=&\frac{1}{2}\mathbb{E}\log\left[\frac{1}{2}e^{-\frac{(t_1-t_2)(p_1-p_0)}{2}}p_1^{\Pi^{(1)}_{p_1t_1}}p_0^{\Pi^{(2)}_{p_0t_2}}+\frac{1}{2}e^{\frac{(t_1-t_2)(p_1-p_0)}{2}}p_0^{\Pi^{(1)}_{p_1t_1}}p_1^{\Pi^{(2)}_{p_0t_2}}\right]\\&+\frac{1}{2}\mathbb{E}\log\left[\frac{1}{2}e^{-\frac{(t_1-t_2)(p_1-p_0)}{2}}p_1^{\Pi^{(1)}_{p_0t_1}}p_0^{\Pi^{(2)}_{p_1t_2}}+\frac{1}{2}e^{\frac{(t_1-t_2)(p_1-p_0)}{2}}p_0^{\Pi^{(1)}_{p_0t_1}}p_1^{\Pi^{(2)}_{p_1t_2}}\right]\end{aligned}$$

is not convex. For every $s\in\mathbb{R}$ and integers $L_1,L_2\geqslant 0$, we denote

$$J(s,L_1,L_2):=\log\left[\frac{1}{2}e^{-\frac{s(p_1-p_0)}{2}}p_1^{L_1}p_0^{L_2}+\frac{1}{2}e^{\frac{s(p_1-p_0)}{2}}p_0^{L_1}p_1^{L_2}\right],$$

and observe that

$$\phi(t_1,t_2)=\frac{1}{2}\sum_{L_1,L_2=0}^{+\infty}(\pi(p_1t_1,L_1)\pi(p_0t_2,L_2)+\pi(p_0t_1,L_1)\pi(p_1t_2,L_2))J(t_1-t_2,L_1,L_2).$$

Using the identity (4.4), we write

$$\begin{aligned}\partial_{t_1}\phi(t_1,t_2) = \frac{1}{2}\sum_{L_1,L_2=0}^{+\infty} & J(t_1-t_2,L_1,L_2)\Big[\left(\pi(p_1t_1,L_1-1)-\pi(p_1t_1,L_1)\right)p_1\pi(p_0t_2,L_2) \\ & \quad +\left(\pi(p_0t_1,L_1-1)-\pi(p_0t_1,L_1)\right)p_0\pi(p_1t_2,L_2)\Big] \\ & +\partial_{t_1}J(t_1-t_2,L_1,L_2)\Big[\pi(p_1t_1,L_1)\pi(p_0t_2,L_2)+\pi(p_0t_1,L_1)\pi(p_1t_2,L_2)\Big],\end{aligned}$$

and

$$\begin{aligned}\partial_{t_1}^2\phi(t_1,t_2) = \frac{1}{2}\sum_{L_1,L_2=0}^{+\infty} & J(t_1-t_2,L_1,L_2) \\ & \Big[\left(\pi(p_1t_1,L_1-2)-2\pi(p_1t_1,L_1-1)+\pi(p_1t_1,L_1)\right)p_1^2\pi(p_0t_2,L_2) \\ & \quad +\left(\pi(p_0t_1,L_1-2)-2\pi(p_0t_1,L_1-1)+\pi(p_0t_1,L_1)\right)p_0^2\pi(p_1t_2,L_2)\Big] \\ & +\partial_{t_1}J(t_1-t_2,L_1,L_2)\Big[\left(\pi(p_1t_1,L_1-1)-\pi(p_1t_1,L_1)\right)p_1\pi(p_0t_2,L_2) \\ & \quad +\left(\pi(p_0t_1,L_1-1)-\pi(p_0t_1,L_1)\right)p_0\pi(p_1t_2,L_2)\Big] \\ & +\partial_{t_1}^2J(t_1-t_2,L_1,L_2)\Big[\pi(p_1t_1,L_1)\pi(p_0t_2,L_2)+\pi(p_0t_1,L_1)\pi(p_1t_2,L_2)\Big].\end{aligned}$$

Similar calculations yield

$$\begin{aligned}\partial_{t_1}\partial_{t_2}\phi(t_1,t_2) = \frac{1}{2}\sum_{L_1,L_2=0}^{+\infty} & J(t_1-t_2,L_1,L_2) \\ & p_0p_1\Big[\left(\pi(p_1t_1,L_1-1)-\pi(p_1t_1,L_1)\right)\left(\pi(p_0t_2,L_2-1)-\pi(p_0t_2,L_2)\right) \\ & \quad +\left(\pi(p_0t_1,L_1-1)-\pi(p_0t_1,L_1)\right)\left(\pi(p_1t_2,L_2-1)-\pi(p_1t_2,L_2)\right)\Big] \\ & +\partial_{t_1}J(t_1-t_2,L_1,L_2)\Big[p_0\pi(p_1t_1,L_1)\left(\pi(p_0t_2,L_2-1)-\pi(p_0t_2,L_2)\right) \\ & \quad +p_1\pi(p_0t_1,L_1)\left(\pi(p_1t_2,L_2-1)-\pi(p_1t_2,L_2)\right)\Big] \\ & +\partial_{t_2}J(t_1-t_2,L_1,L_2)\Big[p_1\left(\pi(p_1t_1,L_1-1)-\pi(p_1t_1,L_1)\right)\pi(p_0t_2,L_2) \\ & \quad +p_0\left(\pi(p_0t_1,L_1-1)-\pi(p_0t_1,L_1)\right)\pi(p_1t_2,L_2)\Big] \\ & +\partial_{t_1}\partial_{t_2}J(t_1-t_2,L_1,L_2)\Big[\pi(p_1t_1,L_1)\pi(p_0t_2,L_2)+\pi(p_0t_1,L_1)\pi(p_1t_2,L_2)\Big].\end{aligned}$$

We get a similar expression for $\partial_{t_2}^2\phi(t_1,t_2)$ as for $\partial_{t_1}^2\phi(t_1,t_2)$:

$$\begin{aligned}\partial_{t_2}^2\phi(t_1,t_2) = \frac{1}{2}\sum_{L_1,L_2=0}^{+\infty} & J(t_1-t_2,L_1,L_2) \\ & \Big[p_0^2\pi(p_0t_1,L_1)\left(\pi(p_0t_2,L_2-2)-2\pi(p_0t_2,L_2-1)+\pi(p_0t_2,L_2)\right) \\ & \qquad + p_1^2\pi(p_0t_1,L_1)\left(\pi(p_1t_2,L_2-2)-2\pi(p_1t_2,L_2-1)+\pi(p_1t_2,L_2)\right)\Big] \\ & + \partial_{t_2}J(t_1-t_2,L_1,L_2)\Big[p_0\pi(p_1t_1,L_1)\left(\pi(p_0t_2,L_2-1)-\pi(p_0t_2,L_2)\right) \\ & \qquad + p_1\pi(p_0t_1,L_1)\left(\pi(p_1t_2,L_2-1)-\pi(p_1t_2,L_2)\right)\Big] \\ & + \partial_{t_2}^2J(t_1-t_2,L_1,L_2)\Big[\pi(p_1t_1,L_1)\pi(p_0t_2,L_2)+\pi(p_0t_1,L_1)\pi(p_1t_2,L_2)\Big].\end{aligned}$$

We are interested in the value of $2\partial_{t_1}\partial_{t_2}\phi(t_1,t_2)-\partial_{t_1}^2\phi(t_1,t_2)-\partial_{t_2}^2\phi(t_1,t_2)$ at $(t_1,t_2)=(0,0)$. Hence, we use the Taylor expansion of $J(t_1-t_2,L_1,L_2)$ for arbitrary $L_1,L_2\in\mathbb{Z}_+$ at $t_1=t_2=0$ to get expressions for derivatives up to the second order

$$\begin{aligned}J(t_1-t_2,N_1,N_2) = \log\left(\frac{p_1^{N_1}p_0^{N_2}+p_0^{N_1}p_1^{N_2}}{2}\right) - (t_1-t_2)\frac{p_1-p_0}{2}\frac{p_1^{N_1}p_0^{N_2}-p_0^{N_1}p_1^{N_2}}{p_1^{N_1}p_0^{N_2}+p_0^{N_1}p_1^{N_2}} \\ + \frac{1}{2}(t_1-t_2)^2(p_1-p_0)^2\frac{p_1^{N_1+N_2}p_0^{N_1+N_2}}{(p_1^{N_1}p_0^{N_2}+p_0^{N_1}p_1^{N_2})^2}+O((t_1-t_2)^3). \quad (4.12)\end{aligned}$$

Further, note that $\pi(0,L)=0$ for all $L\geqslant 1$ and $\pi(0,0)=1$. With this observation, we can write the second derivatives of ϕ as sum of a few terms.

$$\begin{aligned}&2\partial_{t_1}\partial_{t_2}\phi(0,0)-\partial_{t_1}^2\phi(0,0)-\partial_{t_2}^2\phi(0,0) \\ &\quad = \frac{1}{2}\Big(-(p_0-p_1)^2J(0,0,0)+2(p_0-p_1)^2J(0,1,0)+2(p_0-p_1)^2J(0,0,1) \\ &\qquad + 4p_0p_1J(0,1,1)-(p_0^2+p_1^2)J(0,2,0)-(p_0^2+p_1^2)J(0,0,2)\Big) \\ &\qquad - (p_0+p_1)\partial_{t_1}J(0,1,0)+(p_0+p_1)\partial_{t_1}J(0,0,1) \\ &\qquad + (p_0+p_1)\partial_{t_2}J(0,1,0)-(p_0+p_1)\partial_{t_2}J(0,0,1) \\ &\qquad + 2\partial_{t_1}\partial_{t_2}J(0,0,0)-\partial_{t_1}^2J(0,0,0)-\partial_{t_2}^2J(0,0,0).\end{aligned} \quad (4.13)$$

From (4.12) we get

$$2\partial_{t_1}\partial_{t_2}J(0,0,0)-\partial_{t_1}^2J(0,0,0)-\partial_{t_2}^2J(0,0,0)=-(p_0-p_1)^2.$$

Using that $\partial_{t_1}J(0,1,0)=-\partial_{t_1}J(0,0,1)=-{(p_0-p_1)^2}/{2(p_0+p_1)}$, we obtain that

$$-(p_0+p_1)\partial_{t_1}J(0,1,0)+(p_0+p_1)\partial_{t_1}J(0,0,1)=(p_0-p_1)^2.$$

Similarly, $(p_0+p_1)\partial_{t_2}J(0,1,0)-(p_0+p_1)\partial_{t_2}J(0,0,1)=(p_0-p_1)^2$. It remains to compute the first terms of (4.13) that do not contain derivatives of J.

$$\begin{aligned}
&-(p_0-p_1)^2J(0,0,0)+2(p_0-p_1)^2J(0,1,0)+2(p_0-p_1)^2J(0,0,1)\\
&\qquad +4p_0p_1J(0,1,1)-(p_0^2+p_1^2)J(0,2,0)-(p_0^2+p_1^2)J(0,0,2)\\
&\quad =4(p_0-p_1)^2\log\left(\frac{p_0+p_1}{2}\right)+4p_0p_1\log(p_0p_1)-2(p_0+p_1)^2\log\left(\frac{p_0^2+p_1^2}{2}\right).
\end{aligned} \tag{4.14}$$

Combining all together and rearranging terms in (4.14), we get

$$\begin{aligned}
&2\partial_{t_1}\partial_{t_2}\phi(0,0)-\partial_{t_1}^2\phi(0,0)-\partial_{t_2}^2\phi(0,0)=\\
&\qquad (p_0-p_1)^2+2p_0p_1\log\left(1-\frac{(p_0-p_1)^2}{(p_0+p_1)^2}\right)-(p_0^2+p_1^2)\log\left(1+\frac{(p_0-p_1)^2}{(p_0+p_1)^2}\right).
\end{aligned}$$

The right-hand side of the above equality coincides with the right-hand side of (2.3), with the $o(1)$ term taken out. As was shown in the proof of Proposition 2.3, this expression is lower-bounded by $(p_0-p_1)^6/6(p_0+p_1)^4$. □

References

[1] E. Abbe, E. Cornacchia, Y. Gu, and Y. Polyanskiy. Stochastic block model entropy and broadcasting on trees with survey. In *Proceedings of Thirty Fourth Conference on Learning Theory*, volume 134, pages 1–25. PMLR, 2021.

[2] E. Abbe and A. Montanari. Conditional random fields, planted constraint satisfaction, and entropy concentration. *Theory of Computing*, 11:413–443, 12 2015.

[3] J. Barbier, C. L. Chan, and N. Macris. Mutual information for the stochastic block model by the adaptive interpolation method. In *2019 IEEE International Symposium on Information Theory*, page 405–409. IEEE Press, 2019.

[4] J. Barbier, M. Dia, N. Macris, F. Krzakala, T. Lesieur, and L. Zdeborová. Mutual information for symmetric rank-one matrix estimation: A proof of the replica formula. In *Advances in Neural Information Processing Systems (NIPS)*, volume 29, pages 424–432, 2016.

[5] J. Barbier and N. Macris. The adaptive interpolation method: a simple scheme to prove replica formulas in bayesian inference. *Probability Theory and Related Fields*, 174(3-4):1133–1185, 2019.

[6] J. Barbier, N. Macris, and L. Miolane. The layered structure of tensor estimation and its mutual information. In *2017 55th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 1056–1063. IEEE, 2017.

[7] B. Bollobás, S. Janson, and O. Riordan. The phase transition in inhomogeneous random graphs. *Random Struct. Algorithms*, 31(1):3–122, 2007.

[8] R. B. Boppana. Eigenvalues and graph bisection: An average-case analysis. In *28th Annual Symposium on Foundations of Computer Science*, pages 280–285, 1987.

[9] T. N. Bui, S. Chaudhuri, F. T. Leighton, and M. Sipser. Graph bisection algorithms with good average case behavior. *Combinatorica*, 7(2):171–191, 1987.

[10] H. Chen, J.-C. Mourrat, and J. Xia. Statistical inference of finite-rank tensors. *Annales Henri Lebesgue*, 5:1161–1189, 2022.

[11] H.-B. Chen. Hamilton-Jacobi equations for nonsymmetric matrix inference. *Ann. Appl. Probab.*, 32(4):2540–2567, 2022.

[12] H.-B. Chen and J. Xia. Limiting free energy of multi-layer generalized linear models. *Preprint arXiv:2108.12615*, 2021.

[13] H.-B. Chen and J. Xia. Hamilton-Jacobi equations for inference of matrix tensor products. *Ann. Inst. Henri Poincaré Probab. Stat.*, 58(2):755–793, 2022.

[14] A. Coja-Oghlan, F. Krzakala, W. Perkins, and L. Zdeborová. Information-theoretic thresholds from the cavity method. *Advances in Mathematics*, 333:694–795, 2018.

[15] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová. Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Physical Review E*, 84:066106, 2011.

[16] Y. Deshpande, E. Abbe, and A. Montanari. Asymptotic mutual information for the balanced binary stochastic block model. *Information and Inference: A Journal of the IMA*, 6(2):125–170, 2016.

[17] T. Dominguez and J.-C. Mourrat. Infinite-dimensional Hamilton-Jacobi equations for statistical inference on sparse graphs. *Preprint, arXiv:2209.04516*, 2022.

[18] T. Dominguez and J.-C. Mourrat. Mutual information for the sparse stochastic block model. *Preprint, arXiv:2209.04513*, 2022.

[19] M. Dyer and A. Frieze. The solution of some random NP-hard problems in polynomial expected time. *Journal of Algorithms*, 10(4):451–489, 1989.

[20] S. E. Fienberg, M. M. Meyer, and S. S. Wasserman. Statistical analysis of multiple sociometric relations. *Journal of the American Statistical Association*, 80(389):51–67, 1985.

[21] Y. Gu and Y. Polyanskiy. Uniqueness of BP fixed point for the potts model and applications to community detection. *Preprint, arXiv:2303.14688*, 2023.

[22] D. Guo, S. Shamai, and S. Verdú. Mutual information and minimum mean-square error in Gaussian channels. *IEEE Transactions on Information Theory*, 51(4):1261–1282, 2005.

[23] D. Guo, Y. Wu, S. S. Shitz, and S. Verdú. Estimation in Gaussian noise: Properties of the minimum mean-square error. *IEEE Transactions on Information Theory*, 57(4):2371–2385, 2011.

[24] P. W. Holland, K. B. Laskey, and S. Leinhardt. Stochastic blockmodels: First steps. *Social Networks*, 5(2):109–137, 1983.

[25] S. K. Jakobsen. Mutual information matrices are not always positive semidefinite. *IEEE Transactions on information theory*, 60(5):2694–2696, 2014.

[26] J. Kadmon and S. Ganguli. Statistical mechanics of low-rank tensor decomposition. In *Advances in Neural Information Processing Systems*, pages 8201–8212, 2018.

[27] V. Kanade, E. Mossel, and T. Schramm. Global and local information in clustering labeled block models. *IEEE Transactions on Information Theory*, 62(10):5906–5917, 2016.

[28] J. Kingman. *Poisson Processes*. Oxford Studies in Probability. Clarendon Press, 1992.

[29] M. Lamarca. Linear precoding for mutual information maximization in MIMO systems. In *2009 6th International Symposium on Wireless Communication Systems*, pages 26–30. IEEE, 2009.

[30] M. Lelarge and L. Miolane. Fundamental limits of symmetric low-rank matrix estimation. *Probability Theory and Related Fields*, 173(3):859–929, 2019.

[31] T. Lesieur, L. Miolane, M. Lelarge, F. Krzakala, and L. Zdeborová. Statistical and computational phase transitions in spiked tensor estimation. In *2017 IEEE International Symposium on Information Theory (ISIT)*, pages 511–515. IEEE, 2017.

[32] C. Luneau, J. Barbier, and N. Macris. Mutual information for low-rank even-order symmetric tensor estimation. *Information and Inference: A Journal of the IMA*, 10(4):1167–1207, 2021.

[33] C. Luneau, N. Macris, and J. Barbier. High-dimensional rank-one nonsymmetric matrix decomposition: the spherical case. In *2020 IEEE International Symposium on Information Theory (ISIT)*, pages 2646–2651. IEEE, 2020.

[34] L. Massoulié. Community detection thresholds and the weak ramanujan property. In *Proceedings of the Forty-Sixth Annual ACM Symposium on Theory of Computing*, page 694–703. Association for Computing Machinery, 2014.

[35] V. Mayya and G. Reeves. Mutual information in community detection with covariate information and correlated networks. In *2019 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 602–607. IEEE, 2019.

[36] L. Miolane. Fundamental limits of low-rank matrix estimation: the non-symmetric case. *Preprint, arXiv:1702.00473*, 2017.

[37] E. Mossel, J. Neeman, and A. Sly. Reconstruction and estimation in the planted partition model. *Probab. Theory Related Fields*, 162(3-4):431–461, 2015.

[38] E. Mossel, J. Neeman, and A. Sly. Belief propagation, robust reconstruction and optimal recovery of block models. *The Annals of Applied Probability*, 26(4):2211–2256, 2016.

[39] E. Mossel, J. Neeman, and A. Sly. A proof of the block model threshold conjecture. *Combinatorica*, 38(3):665–708, 2018.

[40] E. Mossel and J. Xu. Local algorithms for block models with side information. In *Proceedings of the 2016 ACM Conference on Innovations in Theoretical Computer Science*, pages 71–80, 2016.

[41] J.-C. Mourrat. Hamilton–Jacobi equations for finite-rank matrix inference. *The Annals of Applied Probability*, 30(5):2234–2260, 2020.

[42] J.-C. Mourrat. Hamilton–Jacobi equations for mean-field disordered systems. *Annales Henri Lebesgue*, 4:453–484, 2021.

[43] J.-C. Mourrat. Nonconvex interactions in mean-field spin glasses. *Probability and Mathematical Physics*, 2(2):61–119, 2021.

[44] J.-C. Mourrat. Free energy upper bound for mean-field vector spin glasses. *Ann. Inst. Henri Poincaré Probab. Stat.*, to appear.

[45] M. Payaró and D. P. Palomar. Hessian and concavity of mutual information, differential entropy, and entropy power in linear vector Gaussian channels. *IEEE Transactions on Information Theory*, 55(8):3613–3628, 2009.

[46] G. Reeves. Information-theoretic limits for the matrix tensor product. *IEEE Journal on Selected Areas in Information Theory*, 1(3):777–798, 2020.

[47] G. Reeves, V. Mayya, and A. Volfovsky. The geometry of community detection via the mmse matrix. In *2019 IEEE International Symposium on Information Theory (ISIT)*, pages 400–404. IEEE, 2019.

[48] G. Reeves, H. D. Pfister, and A. Dytso. Mutual information as a function of matrix SNR for linear Gaussian channels. In *2018 IEEE International Symposium on Information Theory (ISIT)*, pages 1754–1758. IEEE, 2018.

[49] Y. J. Wang and G. Y. Wong. Stochastic blockmodels for directed graphs. *Journal of the American Statistical Association*, 82(397):8–19, 1987.

[50] H. C. White, S. A. Boorman, and R. L. Breiger. Social structure from multiple networks. i. blockmodels of roles and positions. *American Journal of Sociology*, 81(4):730–780, 1976.

[51] Q. Yu and Y. Polyanskiy. Ising model on locally tree-like graphs: Uniqueness of solutions to cavity equations. *Preprint, arXiv:2211.15242*, 2022.