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

Anastasia Kireeva [1] Jean-Christophe Mourrat [2]

[1]ETH Zurich [2]ENS Lyon

Problem setting

Let S and $\tilde{S}$ be two iid random variables, and let P_1 and P_2 be two communication channels. We can choose between two measurement scenarios:

- we observe S through P_1 and P_2, and also $\tilde{S}$ through P_1 and P_2;
- 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 $(S, \tilde{S})$?

(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.

Mutual information

For random variables X and Y defined on the same probability space, we denote by $I(X;Y)$ their *mutual information*, that is,

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

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

Let $S \sim P_S$, where P_S is a probability measure with finite support $\mathscr{S}$. We define a *communication channel* P over $\mathscr{S}$ as a family of probability measures $(P(\cdot \mid s))_{s \in \mathscr{S}}$ over $\mathbb{R}^d$. Let P_1 and P_2 be two channels over $\mathscr{S}$.

Conditionally on S, we sample independently $X_1, X_1' \sim P_1(\cdot \mid S)$, and $X_2, X_2' \sim P_2(\cdot \mid S)$. We consider the following question.

Do we have $I(S;(X_1,X_1')) + I(S;(X_2,X_2')) \leq 2I(S;(X_1,X_2))$, or, equivalently, $I(X_1,X_1') + I(X_2,X_2') \geq 2I(X_1,X_2)$? (Q1)

Mixing Gaussian channels yields more information

Gaussian channel is defined by law of a random variable

$$X := f(S) + W, \quad \text{where } f : \mathscr{S} \to \mathbb{R}^d, \quad W \sim N(0, I_d)$$

and W is independent of S.

If P_1 and P_2 are Gaussian channels, then the answer to Question Q1 is **positive**.

Class of counterexamples to Question Q1

Let $S \sim \mathrm{Ber}(1/2)$, $X_1, X_1' \sim P_1(\cdot \mid S)$ and $X_2, X_2' \sim P_2(\cdot \mid S)$, where $P_1(\cdot \mid s) = \mathrm{Ber}(\varepsilon p_s)$ and $P_2(\cdot \mid s) = \mathrm{Ber}(\varepsilon q_s)$ for $s \in \{0,1\}$ and some $p_0, p_1, q_0, q_1 \geq 0$ and $\varepsilon > 0$.

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

$$2I(X_1,X_2) - I(X_1,X_1') - I(X_2,X_2') \geq \frac{\varepsilon^2 (p_0 - p_1)^6}{6(p_0+p_1)^4} + o(\varepsilon^2) \qquad (\varepsilon \to 0). \qquad (1)$$

In particular, this implies that whenever $p_0 \neq p_1$ and $\varepsilon > 0$ is sufficiently small the answer to Question Q1 is **negative**.

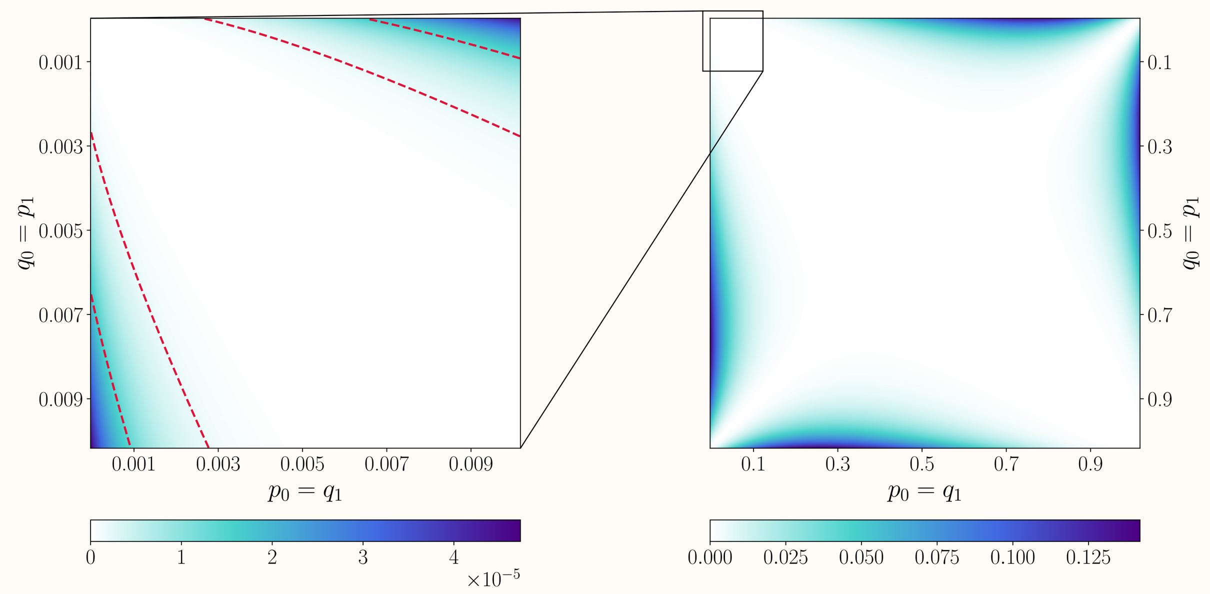


Figure 2. Value of the $2I(X_1,X_2) - I(X_1,X_1') - I(X_2,X_2')$. The larger values correspond to darker color. **Left:** the regime of small p_0, p_1. Red dashed lines are countour lines of $(p_0-p_1)^6/(p_0+p_1)^4$. **Right:** general $p_0, p_1 \in [0,1]$.

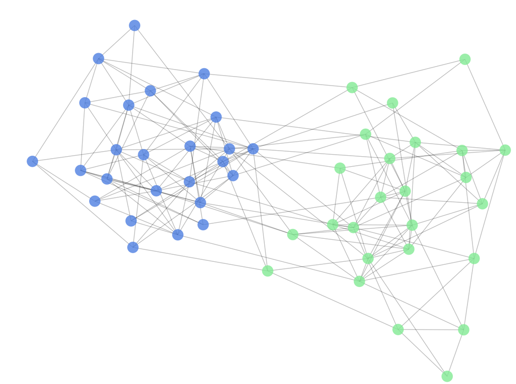
Mutual information in Stochastic Block Model

Stochastic Block Model (SBM)

Let $G_N = (V,E)$ be a random graph on N vertices. Each vertex is independently assigned to a community ± 1, and we denote the assignment vector by $\sigma_N \in \{\pm 1\}^N$.

Edges are sampled independently as follows.

$$\mathbb{P}((u,v) \in E) = \begin{cases} a_N & \text{if } \sigma_u = \sigma_v \\ b_N & \text{otherwise.} \end{cases}$$



We consider symmetric SBM with **two communities** in **sparse assortative** regime, i.e. the edge probabilities scale as $a_N = a/N, b_n = b/N$, for some constant a, b, and $a > b$.

Mutual information in SBM

The mutual information is given by

$$I(G_N, \sigma) = \mathbb{E} \log \frac{\mathbb{P}(G_N|\sigma)}{\mathbb{P}(G_N)}$$

and quantifies the information about the hidden assignment vector σ that we can recover after observing random graph G.

Hamilton-Jacobi equations

A recent method to identify the asymptotic value of the mutual information of a mean-field disordered system is through the solution to a certain *Hamilton-Jacobi equation*.

For SBM with two communities this approach has been initiated in [2, 3].

Lower bound

Theorem (informal, [3]) The lower bound on the limit of free energy can be obtained through *the unique viscosity solution of certain Hamilton-Jacobi equation*.

Upper bound

The central ingredient in showing the matching upper bound in other settings (e.g. [1]) is **concavity** of the *continuous* mutual information.

In particular, the concavity of mutual information in the considered setting would imply the negative semidefiniteness of the Hessian. However, (1) implies that

$$\begin{pmatrix} 1 \\ -1 \end{pmatrix} \cdot \begin{pmatrix} \partial_{t_1}^2 \mathscr{I}_N(0,0) & \partial_{t_1}\partial_{t_2}\mathscr{I}_N(0,0) \\ \partial_{t_1}\partial_{t_2}\mathscr{I}_N(0,0) & \partial_{t_2}^2 \mathscr{I}_N(0,0) \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \geq 0,$$

where $\mathscr{I}_N(t_1,t_2)$ is the continuous mutual information defined below. Consequently, the Hessian is not NSD.

Theorem Let G_N be an SBM with assignment vector $\sigma = 2S - 1$, where $S \sim \mathrm{Ber}(1/2)$.

Conditionally on σ, we sample independent r.v. $X_1^{(\ell)} \sim P_1$ and $X_2^{(\ell)} \sim P_2$, where

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

and $p_0, p_1, q_0, q_1 \geq 0$ are such that $p_0 = q_1 = a$ and $p_1 = q_0 = b$.

Then the mutual information satisfies

$$I(G_N, \sigma) = \mathscr{I}_N(0,0),$$

where

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

and $\Pi^{(1)}_{Nt_1} \sim \mathrm{Poi}(Nt_1)$, $\Pi^{(2)}_{Nt_2} \sim \mathrm{Poi}(Nt_2)$, independent of the all other random variables.

With this choice of parameters, **the mapping** $(t_1,t_2) \mapsto \mathscr{I}_N(t_1,t_2)$ **is not concave for every sufficiently large** $N \in \mathbb{N} \cup \{\infty\}$.

References

[1] Hongbin Chen, Jean-Christophe Mourrat, and Jiaming Xia. Statistical inference of finite-rank tensors. *Annales Henri Lebesgue*, 5:1161–1189, 2022.

[2] Tomas Dominguez and Jean-Christophe Mourrat. Infinite-dimensional Hamilton-Jacobi equations for statistical inference on sparse graphs. *Preprint, arXiv:2209.04516*, 2022.

[3] Tomas Dominguez and Jean-Christophe Mourrat. Mutual information for the sparse stochastic block model. *Preprint, arXiv:2209.04513*, 2022.