Comptes Rendus

Mathématique

Pierre Bizeul

Entropy and Information jump for log-concave vectors

Functional analysis, Probability theory / *Analyse fonctionnelle, Probabilités*

Entropy and Information jump for log-concave vectors

Pierre Bizeul[*, a]

[a] Institut de Mathématiques de Jussieu-Paris Rive Gauche, Sorbonne Université, 4 place de Jussieu 75005 Paris, France

E-mail: pbizeul@imj-prg.fr

Abstract. We extend the result of Ball and Nguyen on the jump of entropy under convolution for log-concave random vectors. We show that the result holds for any pair of vectors (not necessarily identically distributed) and that a similar inequality holds for the Fisher information, thus providing a quantitative Blachmann–Stam inequality.

1. Introduction

Let X be a random vector distributed according to a measure μ in $\mathbb{R}^d$, with density f with respect to the Lebesgue measure. We will denote this by $X \sim \mu = f\,d\lambda$. If $\int f|\log f| < +\infty$, we define its entropy by

$$\text{Ent}_L(\mu) = \text{Ent}_L(X) = - \int_{\mathbb{R}^d} f \log f,$$

where the subscript L stands for "Lebesgue".

It should be noted that this entropy can be either positive or negative, and that for any invertible matrix A, $\text{Ent}_L(AX) = \text{Ent}_L(X) + \log(|\det A|)$. Entropy is also translation invariant, and it is classical that, when the covariance matrix is fixed, the Gaussian distribution maximizes entropy. It will be useful to normalize vectors so that they are centered, and have covariance matrix identity. Such a vector, as well as its distribution, is called *isotropic*.

The classical Shannon–Stam inequality asserts that taking a convolution increases entropy: for two iid random vectors X_1 and X_2

$$\text{Ent}_L(X_1) \le \text{Ent}_L\left(\frac{X_1 + X_2}{\sqrt{2}}\right).$$

Moreover, there is equality if and only if X has a Gaussian distribution. Now, one can wonder if that equality case is stable, meaning if the *entropy jump* $\text{Ent}_L(\frac{X_1+X_2}{\sqrt{2}}) - \text{Ent}_L(X)$ is small, does

---

[*] Corresponding author.

it imply that X is almost Gaussian? The general answer is no, as one can convince himself by considering a well chosen double-bump Gaussian [4].

However, when the distribution of X admits a spectral gap, excluding double-bumped type distributions, some positive answers exist. Recall that X is said to have a spectral gap, or equivalently satisfy a Poincaré inequality, if there exists a constant $c > 0$ such that for any smooth enough function f, the variance of $f(X)$ can be controlled in terms of the euclidean norm of $\nabla f(x)$ as follows:

$$\mathrm{var} f(X) \leq c \mathbb{E}\left[\left|\nabla f(X)\right|^2\right].$$

The smallest such constant c will be denoted c_X and called the Poincaré constant of X. Under a spectral gap assumption, it was proven by Ball, Barthe and Naor in [2] that for a one dimensional isotropic random variable X,

$$\mathrm{Ent}_L\left(\frac{X_1 + X_2}{\sqrt{2}}\right) - \mathrm{Ent}_L(X_1) \geq \frac{1}{2\left(1 + c_X\right)}\left(\mathrm{Ent}_L(G) - \mathrm{Ent}_L(X_1)\right),$$

where G is a standard Gaussian.

We can rewrite the right-hand side as a Kullback–Leibler divergence. Recall that, if X is isotropic,

$$\mathrm{Ent}_L(G) - \mathrm{Ent}_L(X) = D(X||G) = \int_{\mathbb{R}^d} f_\gamma \log\left(f_\gamma\right) d\gamma \geq 0,$$

where f_γ is the relative density of X with respect to the Gaussian measure γ, that is $X \sim f_\gamma d\gamma$ and $G \sim \gamma$. In the sequel we shall use the notation $D(X) = D(X||G)$. This is a strong measure of closeness to the Gaussian; for instance the Pinsker–Csiszar–Kullback inequality states that

$$\left(\int_{\mathbb{R}^d}\left|f - g\right|\right)^2 \leq \frac{1}{2}D(X||G),$$

where f and g are the density of X and G, respectively. In 2012, Ball and Nguyen generalized the result to arbitrary dimension, assuming log-concavity of X ([3]). They use a semigroup approach, differentiating twice the entropy along the Ornstein–Uhlenbeck semigroup.

The Fisher information of a random vector with smooth density f is

$$I_L(X) = \int \frac{\left|\nabla f\right|^2}{f} = 4 \int \left|\nabla\left(\sqrt{f}\right)\right|^2,$$

whenever those integrals are well defined. As for the entropy, for a fixed covariance matrix, the Gaussian is extremal; in this case, it has the smallest information. Information is classically the derivative of the entropy along the semi-group. In the spirit of the Shannon–Stam Inequality, the Blachman–Stamn inequality asserts that taking a convolution decreases the information:

$$I_L(X_1) \geq I_L\left(\frac{X_1 + X_2}{\sqrt{2}}\right).$$

As before, we can define a relative information, notably to the Gaussian measure $d\gamma$. If $X \sim f_\gamma d\gamma$ is a random vector with smooth density, we will denote its relative information to $d\gamma$ by

$$I(X||G) = I_\gamma(X) = \int \frac{\left|\nabla f_\gamma\right|^2}{f_\gamma} d\gamma.$$

When X is isotropic, integrating by parts yields:

$$I(X||G) = I_L(X) - d = I_L(X) - I_L(G).$$

Consequently, for a measure μ on $\mathbb{R}^d$, we write $D(\mu||\gamma) = D(X||G)$, $I_L(\mu) = I_L(X)$ and $I(\mu||\gamma) = I(X||G)$ where $X \sim \mu$ is a random vector distributed according to μ

In this note, we use the same strategy as in [3], but improve their result in two directions. First we generalize it to non identically distributed pairs of random vectors. For two measures μ and ν define

$$\delta_{E,\lambda}(\mu,\nu) = \text{Ent}_L(X_\lambda) - (1-\lambda)\text{Ent}_L(X_0) - \lambda\text{Ent}_L(X_1),$$

where X_0 and X_1 are independent random vectors distributed according to μ and ν respectively, and $X_\lambda = \sqrt{1-\lambda}X_0 + \sqrt{\lambda}X_1$. The Shannon–Stam inequality asserts that $\delta_\lambda(\mu,\nu) \geq 0$ and this quantity is precisely the deficit in the Shannon–Stam inequality.

Theorem 1 (Quantitative Shannon–Stam). *Let μ,ν be two log-concave isotropic measures with Poincaré constant respectively c_0 and c_1, and $\lambda \in [0,1]$. Then,*

$$\delta_{E,\lambda}(\mu,\nu) \geq \frac{\lambda(1-\lambda)}{4\max(c_0,c_1)}\left(D\left(\mu||\gamma\right) + D\left(\nu||\gamma\right)\right).$$

This should be compared with a recent result of Eldan and Mikulincer ([5, Theorem 3]). They get a more general result, allowing μ and ν to have different covariance matrices, but in the case where μ and ν have the same covariance matrix, they get a worst dependence on the Poincaré constant.

Secondly, we get a same kind of inequality for the information, yielding a stability result for the Blachman-Stam inequality. Define this time the information deficit of a pair of measures by

$$\delta_{I,\lambda}(\mu,\nu) = (1-\lambda)I_L(X_0) + \lambda I_L(X_1) - I_L(X_\lambda)$$

where X_0 and X_1 are independent random vectors distributed according to μ and ν respectively, and $X_\lambda = \sqrt{1-\lambda}X_0 + \sqrt{\lambda}X_1$

Theorem 2 (Quantitative Blachman–Stam). *Let μ,ν be two log-concave isotropic measures with Poincaré constant respectively c_0 and c_1, and $\lambda \in [0,1]$. Then,*

$$\delta_{I,\lambda}(\mu,\nu) \geq \frac{\lambda(1-\lambda)}{4\max(c_0,c_1)}\left(I(\mu||d\gamma) + I(\nu||d\gamma)\right).$$

In the sequel, quantities computed with respect to the Lebesgue measure have a subscript "L" while quantities that are computed with respect to the Gaussian measure have none.

2. A lemma of Ball–Nguyen

Let X be a random vector with smooth density $f = e^{-\psi}$ with respect to the *Lebesgue* measure. We define $\sigma_L(X)$ to be the random matrix $\sigma_L(X) := \nabla^2(\psi)(X)$ and, we denote

$$K_L(X) = \mathbb{E}\left[\|\sigma_L(X)\|^2\right],$$

where $\|.\|$ denotes the Hilbert–Schmidt norm on matrices and the subscripts L yet again stands for "Lebesgue", which we take temporarily as the reference measure. Understanding this quantitiy will prove to be important later on, as it will appear in the second derivative of the entropy along the Ornstein–Uhlenbeck semigroup.

We recall a lemma of Ball–Nguyen [3], for which we provide a simple proof.

Lemma 3 ((Ball-Nguyen)). *Let X be a random vector in $\mathbb{R}^d$ with smooth density, $E \subset \mathbb{R}^d$ be any subspace, p_E be the orthogonal projection onto E and $X_E = p_E(X)$. Then*

$$\sigma_L(X_E) \leq p_E\mathbb{E}\left[\sigma_L(X) \,|\, X_E\right]p_E^*, \quad a.s,$$

for the partial order on symmetric matrices.

Proof. Let $\psi_E : E \to \mathbb{R}$ be such that $X_E \sim e^{-\psi_E(x)}dx$. We have for $x \in E$

$$\psi_E(x) = -\ln \int_{E^\perp} e^{-\psi(x,y)}dy.$$

Then, setting $dv_x = \frac{e^{-\psi(x,y)}}{\int_{E^\perp} e^{-\psi(x,y)}dy}dy$, a straightforward computation shows that:

$$\forall\, x \in E, \quad \nabla^2 \psi_E(x) = \int_{E^\perp} \nabla^2_{xx}\psi\, dv_x - \mathrm{cov}_{dv_x}\left(\nabla_x\psi\right).$$

In particular,

$$\forall\, x \in E, \quad \nabla^2\psi_E(x) \le \int_{E^\perp}\nabla^2_{xx}\psi\, dv_x,$$

which is the desired result.   □

If X_0 and X_1 are two independent random vectors in $\mathbb{R}^d$ and $\lambda \in [0,1]$, applying the previous lemma to the random vector (X_0, X_1) and the projection $p(x,y) = \sqrt{1-\lambda}x + \sqrt{\lambda}y$ yields:

Lemma 4. *For any independent random vectors X_0, X_1 in $\mathbb{R}^d$, with smooth densities, and any $\lambda \in [0,1]$*

$$\sigma_L(X_\lambda) \le \mathbb{E}\left[(1-\lambda)\sigma_L(X_0) + \lambda\sigma_L(X_1)\,|\,X_\lambda\right] \quad a.s,$$

where $X_\lambda = \sqrt{1-\lambda}X_0 + \sqrt{\lambda}X_1$.

If X_0 and X_1 are log-concave and independent, then so is X_λ, by Prékopa's theorem. Thus $\sigma_L(X_0), \sigma_L(X_1)$ and $\sigma(X_\lambda)$ are positive matrices, so as in Ball–Nguyen's article, the inequality above translates to an inequality on their norm.

Lemma 5. *For any log-concave independent random vectors X_0, X_1 in $\mathbb{R}^d$, with smooth densities, and any $\lambda \in [0,1]$*

$$(1-\lambda)K_L(X_0) + \lambda K_L(X_1) - K_L(X_\lambda) \ge \lambda(1-\lambda)\mathbb{E}\left[\|\sigma_L(X_1) - \sigma_L(X_0)\|^2\right].$$

Remark 6. In particular, we have:

$$K_L(X_\lambda) \le (1-\lambda)K_L(X_0) + \lambda K_L(X_1),$$

which can be seen as a second-order Blachman–Stam inequality.

Proof. As explained, the matrices being positive, the inequality in Lemma 4 implies that:

$$\|\sigma_L(X_\lambda)\| \le \|\mathbb{E}\left[(1-\lambda)\sigma_L(X_0) + \lambda\sigma_L(X_1)\,|\,X_\lambda\right]\|.$$

Taking the expectation of the square and using Jensen's inequality then implies:

$$\begin{aligned}
K_L(X_\lambda) &\le \mathbb{E}\left[\|(1-\lambda)\sigma_L(X_0) + \lambda\sigma_L(X_1)\|^2\right]\\
&= (1-\lambda)^2 K_L(X_0) + \lambda^2 K_L(X_1) + 2\lambda(1-\lambda)\mathbb{E}\left[\langle\sigma_L(X_0), \sigma_L(X_1)\rangle\right]\\
&= (1-\lambda)K_L(X_0) + \lambda K_L(X_1) - \lambda(1-\lambda)\mathbb{E}\left[\|\sigma_L(X_0) - \sigma_L(X_1)\|^2\right]
\end{aligned}$$

since X_0 and X_1 are independent.   □

Now we want to translate this result to the Gaussian setting. Assuming that X has density $f_\gamma = e^{-\varphi}$ with respect to the Gaussian measure, we similarly introduce:

$$\begin{aligned}
\sigma(X) &:= \sigma_\gamma(X) := \nabla^2(\varphi)(X)\\
K(X) &:= K_\gamma(X) = \mathbb{E}\left[\|\sigma(X)\|^2\right]\\
I(X) &= I_\gamma(X) = \mathbb{E}\left[\left\|\nabla\varphi(X)\right\|^2\right].
\end{aligned}$$

Note that the definition of I is consistent with the one given in the introduction. Then Lemma 5 becomes:

Lemma 7. *For any log-concave independent isotropic random vectors X_0, X_1 in $\mathbb{R}^d$, with smooth densities, and any $\lambda \in [0, 1]$, we have*

$$(1 - \lambda)M(X_0) + \lambda M(X_1) - M(X_\lambda) \geq \lambda(1 - \lambda)\mathbb{E}\left[\|\sigma(X_1) - \sigma(X_0)\|^2\right], \tag{1}$$

where $M(X) = K(X) + 2I(X)$.

Proof. Let X be an isotropic log-concave random vector with density $e^{-\varphi}$ with respect to the Gaussian measure, and $e^{-\psi}$ with respect to the Lebesgue measure. By definition we have:

$$\sigma(X) = \sigma_L(X) - Id.$$

Hence,

$$K(X) = K_L(X) - 2\mathbb{E}\left[\text{tr}\left(\sigma_L(X)\right)\right] + n.$$

Now, by integration by parts:

$$\begin{aligned}
\mathbb{E}\left[\text{tr}\left(\sigma_L(X)\right)\right] &= \int_{\mathbb{R}^n} \text{div}\left(\nabla\psi\right)(x)e^{-\psi(x)}dx \\
&= \int_{\mathbb{R}^n} \nabla\psi(x) \cdot \nabla\psi(x)e^{-\psi(x)}dx \\
&= \mathbb{E}\left[\|\nabla\psi(X)\|^2\right] = \mathbb{E}\left[\|\nabla\varphi(X) + X\|^2\right] = I(X) + n.
\end{aligned}$$

The lemma follows.   □

The next lemma provides a lower bound for the right-hand side in the inequality (1).

Lemma 8. *For any log-concave independent isotropic random vectors X_0, X_1 in $\mathbb{R}^d$ with smooth densities we have*

$$(1 - \lambda)M(X_0) + \lambda M(X_1) - M(X_\lambda) \geq \frac{\lambda(1 - \lambda)}{2\max(c_0, c_1)}\left(I(X_0) + K(X_0) + I(X_1) + K(X_1)\right),$$

where c_0, c_1 are the Poincaré constants of X_0 and X_1, respectively.

Proof. We denote by c_0 the Poincaré constant of X_0. We condition on X_1 and apply the Poincaré inequality for X_0 to the function $\nabla\varphi_0(X_0) - \nabla^2\varphi_1(X_1)X_0$ which is centered and we use the fact that X_0 and X_1 are isotropic:

$$\begin{aligned}
\mathbb{E}\left[\|\nabla^2\varphi_0(X_0) - \nabla^2\varphi_1(X_1)\|^2\right] &\geq \frac{1}{c_0}\mathbb{E}\left[\|\nabla\varphi_0(X_0) - \nabla^2\varphi_1(X_1)X_0\|^2\right] \\
&= \frac{1}{c_0}\mathbb{E}\left[\|\nabla\varphi_0(X_0)\|^2 + \|\nabla^2\varphi_1(X_1)\|\right] \\
&= \frac{1}{c_0}(I(X_0) + K(X_1)).
\end{aligned}$$

By symmetry we get:

$$\mathbb{E}\left[\|\sigma_\gamma(X_1) - \sigma_\gamma(X_0)\|^2\right] \geq \frac{1}{2\max(c_0, c_1)}\left(I(X_0) + K(X_0) + I(X_1) + K(X_1)\right).$$

Plugging this into Lemma 7 concludes the proof.   □

3. The Ornstein–Uhlenbeck process and proof of the theorems

Let X be a random vector in $\mathbb{R}^d$ with density f with respect to the *Gaussian* measure γ. Let L_γ be the diffusion operator defined by $L_\gamma f(x) = \Delta f(x) - \nabla f(x) \cdot x$. The differential equation associated to L_γ is the modified heat equation:

$$\frac{\partial f_t}{\partial t} = L_\gamma f_t \, ; \quad f_0 = f.$$

Its solution f_t is the relative density of the random vector $X_t = e^{-t} X + \sqrt{1 - e^{-2t}} G$, where G is a standard Gaussian independent of X, with respect to the Gaussian measure. From this description of X_t it is clear that X_t has a $\mathscr{C}^\infty$ density, with integrability properties as good as f, and that the process commutes with convolutions, in the sense that, with the notations of the previous sections, for all independent X_0, X_1 random vectors, and any $\lambda \in [0,1]$, we have

$$(X_\lambda)_t = (X_t)_\lambda \quad \text{in law.} \tag{2}$$

It is also useful to note that if X is such that $c_X \geq 1$, in particular if X is isotropic, then for all $t \geq 0$:

$$c_{X_t} \leq c_X.$$

Indeed, if X and Y are independent random vectors satisfying a Poincaré inequality and $\lambda \in [0,1]$, then using the conditional variance formula, a few computations show that

$$c_{\sqrt{\lambda} X + \sqrt{1-\lambda} Y} \leq \lambda c_X + (1-\lambda) c_Y,$$

which in our case yields $c_{X_t} \leq e^{-2t} c_X + (1 - e^{-2t}) \times 1 \leq c_X$.

We denote by $(P_t)_{t \geq 0}$ the semi-group defined by $P_t(f) = f_t$. The following computations are standard (see [1]): if X has finite entropy, then for $t > 0$

$$\frac{\partial}{\partial t} \operatorname{Ent}(X_t) = I(X_t), \quad \frac{\partial I(X_t)}{\partial t} = -2I(X_t) - 2K(X_t) = -M(X_t) - K(X_t).$$

As a consequence, a linear inequality on the information can be integrated along the semi-group to get the same inequality for the entropy. We also get that $\frac{\partial I(X_t)}{\partial t} \leq -2I(X_t)$ which implies that $I(X_t) \leq e^{-2t} I(X_0)$. Moreover, the control role of L comes from the observation that

$$e^{2t} \frac{\partial \left(e^{-2t} I(X_t) \right)}{\partial t} = -2M(X_t). \tag{3}$$

Now we are in position to prove Theorem 2, from which Theorem 1 will be an immediate corollary.

Let X_0 and X_1 be two isotropic random log-concave vectors in $\mathbb{R}^d$ and $\lambda \in [0,1]$, and $X_\lambda = \sqrt{1-\lambda} X_0 + \sqrt{\lambda} X_1$. Denote by $(X_0)_t$, $(X_1)_t$ and $(X_\lambda)_t$ their evolution along the Ornstein–Uhlenbeck semi-group. Further define:

$$I_0(t) = I((X_0)_t), \quad I_1(t) = I((X_1)_t) \quad , \quad I_\lambda(t) = I((X_\lambda)_t),$$

and similarly, define $K_0(t), K_1(t), K_\lambda(t), M_0(t), M_1(t), M_\lambda(t)$; using (3), the commutation property (2), the observation that the Poincaré constants only decrease along the semi-group and Lemma 8, we get the following:

Lemma 9. *With the previous notations, for all $t \geq 0$,*

$$-\frac{\partial}{\partial t} \left((1-\lambda) e^{-2t} I_0(t) + \lambda e^{-2t} I_1(t) - e^{-2t} I_\lambda(t) \right)$$

$$\geq \frac{\lambda(1-\lambda) e^{-2t}}{\max(c_0(t), c_1(t))} \left(I_0(t) + K_0(t) + I_1(t) + K_1(t) \right)$$

$$\geq -\frac{\lambda(1-\lambda)}{2\max(c_0, c_1)} e^{-2t} \frac{\partial}{\partial t} \left(I_0(t) + I_1(t) \right).$$

Proofs of Theorem 1 and Theorem 2. Integrating the inequality of Lemma 9 from 0 to ∞, we get:

$$(1-\lambda) I(X_0) + \lambda I(X_1) - I(X_\lambda) \geq -\frac{\lambda(1-\lambda)}{2\max(c_0, c_1)} \int_0^{+\infty} e^{-2t} \frac{d}{dt} \left(I_0(t) + I_1(t) \right) dt$$

$$= \frac{\lambda(1-\lambda)}{2\max(c_0, c_1)} \left(I(X_0) + I(X_1) - 2 \int_0^\infty e^{-2t} (I_0(t) + I_1(t)) \, dt \right)$$

$$\geq \frac{\lambda(1-\lambda)}{4\max(C_0, C_1)} \left(I(X_0) + I(X_1) \right),$$

where in the last inequality, we used the fact that $I_1(t) \le e^{-2t} I(X_1)$ and $I_0(t) \le e^{-2t} I(X_0)$. This proves Theorem 2.

Integrating Theorem 2 along the semi-group yields Theorem 1. $\qquad\qquad\square$

References

[1] D. Bakry, I. Gentil, M. Ledoux, *Analysis and geometry of Markov diffusion operators*, Grundlehren der Mathematischen Wissenschaften, vol. 348, Springer, 2014.

[2] K. Ball, F. Barthe, A. Naor, "Entropy jumps in the presence of a spectral gap", *Duke Math. J.* **119** (2003), no. 1, p. 41-63.

[3] K. Ball, V. H. Nguyen, "Entropy jumps for isotropic log-concave random vectors and spectral gap", *Stud. Math.* **213** (2012), no. 1, p. 81-96.

[4] T. A. Courtade, M. Fathi, A. Pananjady, "Quantitative stability of the entropy power inequality", *IEEE Trans. Inf. Theory* **64** (2018), no. 8, p. 5691-5703.

[5] R. Eldan, D. Mikulincer, "Stability of the Shannon-Stam inequality via the Föllmer process", *Probab. Theory Relat. Fields* **177** (2020), no. 3-4, p. 891-922.