

A Sharp Gaussian Log-Partition Bound for Mutual Information

Alper Ferudun*

18 October 2026

Abstract

Let X have an arbitrary Borel probability law μ on $\mathbb{R}^n$, and let N be an independent standard Gaussian. We prove

$$I(X; X + N) \leq \int_{\mathbb{R}^n} \log \left(\int_{\mathbb{R}^n} e^{\langle z, x \rangle} \mu(dx) \right) \gamma_n(dz).$$

The constant one is optimal, and no moment, centering, or log-concavity assumption is required. The proof uses Caffarelli's contraction theorem in the direction from the Gaussian convolution $\mu * \gamma_n$ to γ_n . Entropy change and Gaussian integration by parts yield a coupling whose correlation is at least twice the channel mutual information; the Gibbs variational inequality then gives the result. Conditional truncation handles arbitrary input laws. This answers the first numbered question in Jingbo Liu's contribution to Oberwolfach Report 14/2022. The separate generalized dual Sudakov question in that report is not addressed.

1 Statement and context

Write γ_n for standard Gaussian measure on $\mathbb{R}^n$ and φ_n for its density. All logarithms are natural. Mutual information is defined by relative entropy:

$$I(U; V) = D(P_{UV} \| P_U \otimes P_V).$$

For a Borel probability measure μ , set

$$M_\mu(z) = \int e^{\langle z, x \rangle} \mu(dx), \quad F(\mu) = \int \log M_\mu(z) \gamma_n(dz).$$

This extended integral is always well-defined. Indeed, if $p = \mu(B_R) > 0$, where $B_R = \{x : |x| \leq R\}$, then

$$\log M_\mu(z) \geq \log p - R|z|. \tag{1}$$

Its negative part is therefore Gaussian-integrable.

Theorem 1. *For every Borel probability measure μ on $\mathbb{R}^n$, if $X \sim \mu$ and $N \sim \gamma_n$ are independent, then*

$$I(X; X + N) \leq F(\mu). \tag{2}$$

The inequality is understood in the extended-real sense. The multiplicative constant one cannot be decreased.

*Mercury Software GmbH. Email: alper@mercurycodelab.com. GitHub: AlperTheKing.

Liu's first question in [1, p. 695] asks for a universal constant in (2). The exponent in that question has no quadratic penalty, and the expectation over the Gaussian lies outside the logarithm. Theorem 1 gives the sharp constant in every finite dimension.

The proof passes through a fixed-marginal coupling. For bounded μ , define

$$\mathcal{E}(\mu) = \sup_{\pi \in \Pi(\mu, \gamma_n)} \left\{ \int \langle x, z \rangle \pi(dx, dz) - D(\pi \| \mu \otimes \gamma_n) \right\}. \quad (3)$$

The conditional Gibbs inequality gives $\mathcal{E}(\mu) \leq F(\mu)$; this fixed-marginal bound also appears in [4, eq. (12)]. We construct a particular coupling proving the stronger intermediate statement

$$I(X; X + N) \leq \mathcal{E}(\mu) \leq F(\mu) \quad \text{for bounded input laws.} \quad (4)$$

The related soft-minoration estimates in [4] and the rate-distortion integral bounds in [5] concern optimized couplings and entropic transport. Chu and Raginsky [6] study soft maxima for finite Gaussian processes. The present argument compares the Gaussian log-partition functional directly with the information of the independent Gaussian observation channel and obtains the sharp constant one. We make no absolute priority claim.

2 A Gaussian integration-by-parts estimate

We will use the usual Gaussian integration-by-parts identity for a Lipschitz vector field. For an inverse transport map it is convenient to retain an inequality valid even in the presence of a singular distributional Hessian.

Lemma 2. *Let $\Phi : \mathbb{R}^n \rightarrow \mathbb{R}$ be a finite convex function and $T = \nabla \Phi$ almost everywhere. Suppose $T \in L^2(\gamma_n; \mathbb{R}^n)$. Let A be the almost-everywhere Hessian of Φ , equivalently the density of the absolutely continuous part of its distributional Hessian. Then $\text{tr } A$ is integrable and*

$$\int \text{tr } A d\gamma_n \leq \int \langle z, T(z) \rangle \gamma_n(dz). \quad (5)$$

Proof. The distributional divergence $m = \text{div } T$ is a nonnegative Radon measure. Its absolutely continuous part is $(\text{tr } A) dz$; see the convex-function framework used in [3, pp. 310, 316]. Choose smooth radial cutoffs $0 \leq \chi_R \leq 1$, equal to one on B_R , supported in B_{2R} , increasing pointwise to one, and satisfying $|\nabla \chi_R| \leq C/R$. Distributional integration by parts gives

$$\begin{aligned} \int \chi_R \varphi_n dm &= - \int \langle \nabla(\chi_R \varphi_n), T \rangle dz \\ &= \int \chi_R(z) \langle z, T(z) \rangle \gamma_n(dz) - \int \langle \nabla \chi_R, T \rangle d\gamma_n. \end{aligned}$$

The last term tends to zero, since $\mathbb{E}|T(Z)| < \infty$. The first term tends to $\mathbb{E}\langle Z, T(Z) \rangle$ by Cauchy-Schwarz and dominated convergence. Monotone convergence on the left shows that $\int \varphi_n dm$ is finite and equals this limit. Discarding the nonnegative singular part proves (5). $\square$

3 The transport coupling for bounded inputs

Assume in this section that $|X| \leq R$ almost surely, and put $Y = X + N$. Its positive smooth density is

$$q(y) = \varphi_n(y) \int e^{\langle y, x \rangle - |x|^2/2} \mu(dx).$$

Differentiation of this bounded-input mixture gives

$$\nabla^2(-\log q)(y) = \text{Id} - \text{Cov}(X \mid Y = y) \preceq \text{Id}. \quad (6)$$

No log-concavity of q is asserted.

Apply Caffarelli's contraction theorem from $q(y) dy$ to γ_n . In the formulation of [2, Theorem 1], an upper Hessian bound βId for the source potential and a lower Hessian bound αId for the target potential give a $\sqrt{\beta/\alpha}$ -Lipschitz Brenier map. Here both constants are one by (6). Both measures have positive smooth full-support densities and finite second moments, so the stated hypotheses apply. We obtain a 1-Lipschitz map $S = \nabla \Psi$ with $S_{\#}(q dy) = \gamma_n$.

Set $Z = S(Y)$. The inverse Brenier map $T = \nabla \Psi^*$ satisfies $T(Z) = Y$ almost surely. This follows from the inverse relation for conjugate potentials and uniqueness of the quadratic transport maps. The inverse potential is finite on $\mathbb{R}^n$: its effective domain is convex and carries full Gaussian measure. Data processing in both directions gives

$$I(X; Z) = I(X; Y). \quad (7)$$

At almost every relevant pair $z = S(y)$, write

$$B(y) = DS(y), \quad A(z) = DT(z) = B(T(z))^{-1}.$$

The positive-density change-of-variables equation is

$$q(y) = \varphi_n(S(y)) \det B(y). \quad (8)$$

These are standard Brenier change-of-variables identities; see [3, p. 310]. For the present smooth positive densities with convex full support, local transport regularity also allows the inverse derivative identity to be read in the classical sense. The almost-everywhere form suffices. Convexity, contraction, and (8) imply

$$0 \prec B(Y) \preceq \text{Id}, \quad A(Z) \succeq \text{Id} \quad \text{almost surely}. \quad (9)$$

Bounded input gives quadratic bounds on $-\log q$ and a finite second moment for Y . Thus all entropy terms in the logarithm of (8) are integrable. Consequently,

$$\begin{aligned} I(X; Y) &= h(Y) - h(N) \\ &= -\mathbb{E} \log \det B(Y) = \mathbb{E} \log \det A(Z). \end{aligned} \quad (10)$$

In particular this information is finite. It is also bounded by $\mathbb{E}|X|^2/2$, by comparison of each translated Gaussian channel law with γ_n .

Lemma 3. *The coupling just constructed satisfies*

$$\mathbb{E}\langle Z, X \rangle \geq 2I(X; X + N). \quad (11)$$

Proof. Gaussian integration by parts in N , conditional on X , is legitimate for the Lipschitz map S and gives

$$\mathbb{E}\langle Z, N \rangle = \mathbb{E} \text{tr} B(Y). \quad (12)$$

Also $T \in L^2(\gamma_n)$, since $T(Z) = Y$. Lemma 2 therefore yields

$$\mathbb{E}\langle Z, Y \rangle = \mathbb{E}\langle Z, T(Z) \rangle \geq \mathbb{E} \text{tr} A(Z).$$

Subtracting (12), with $B(Y) = A(Z)^{-1}$, we obtain

$$\mathbb{E}\langle Z, X \rangle \geq \mathbb{E} \operatorname{tr}(A(Z) - A(Z)^{-1}). \quad (13)$$

For every $\lambda \geq 1$,

$$\lambda - \lambda^{-1} \geq 2 \log \lambda. \quad (14)$$

Indeed the difference is zero at one and has derivative $(\lambda - 1)^2/\lambda^2$. Apply (14) to the eigenvalues in (9), sum, and use (10). This proves (11). $\square$

Proof of Theorem 1 for bounded input. For almost every z , apply the Gibbs variational inequality with reference law μ , conditional law $P_{X|Z=z}$, and test function $x \mapsto \langle z, x \rangle$:

$$\log M_\mu(z) \geq \mathbb{E}[\langle z, X \rangle \mid Z = z] - D(P_{X|Z=z} \parallel \mu).$$

All averaged terms are finite because X is bounded and Z is standard Gaussian. Integrating and using (7) and Lemma 3 gives

$$F(\mu) \geq \mathbb{E}\langle Z, X \rangle - I(X; Z) \geq 2I(X; Y) - I(X; Y) = I(X; Y).$$

The same coupling proves the first inequality in (4). $\square$

4 General input laws and sharpness

Let now μ be arbitrary. Choose R_0 with $\mu(B_{R_0}) > 0$, and for $R \geq R_0$ put

$$p_R = \mu(B_R), \quad \mu_R = \frac{\mu|_{B_R}}{p_R}, \quad L_R(z) = \log \int_{B_R} e^{\langle z, x \rangle} \mu(dx).$$

The functions L_R increase pointwise to $\log M_\mu$, with common integrable lower bound $\log \mu(B_{R_0}) - R_0|z|$. Monotone convergence after subtracting this lower bound, and $p_R \rightarrow 1$, imply

$$F(\mu_R) = \mathbb{E} L_R(Z) - \log p_R \longrightarrow F(\mu). \quad (15)$$

Take $X_R \sim \mu_R$, independently of N . The joint law of $(X_R, X_R + N)$ converges weakly to that of $(X, X + N)$; the product of its marginals converges as well. Joint lower semicontinuity of relative entropy thus gives

$$\begin{aligned} I(X; X + N) &\leq \liminf_{R \rightarrow \infty} I(X_R; X_R + N) \\ &\leq \lim_{R \rightarrow \infty} F(\mu_R) = F(\mu). \end{aligned}$$

This proves the general case, including when the right side is infinite, without an entropy-continuity or moment assumption.

For sharpness, take $X \sim N(0, t\operatorname{Id})$, $t > 0$. Then

$$F(\mu) = \frac{nt}{2}, \quad I(X; X + N) = \frac{n}{2} \log(1 + t),$$

and the ratio of the latter to the former is $\log(1 + t)/t \rightarrow 1$ as $t \downarrow 0$. This completes the proof of Theorem 1.

Corollary 4. *For any $s \geq 0$,*

$$I(X; \sqrt{s}X + N) \leq \int \log \left(\int e^{\sqrt{s}\langle z, x \rangle} \mu(dx) \right) \gamma_n(dz).$$

Proof. Apply Theorem 1 to $\sqrt{s}X$. For $s > 0$, invertibility of this scaling preserves the mutual information with the observation. For $s = 0$, both sides are zero. $\square$

5 Scope and reproducibility

Theorem 1 resolves the first numbered question on page 695 of [1]. The second question, concerning generalized dual Sudakov minoration for log-concave measures, is separate and is not resolved by this paper. In particular, a dataset row bundling the two questions should not be marked fully solved.

The proof is analytic. Supplementary scripts examine 1,415 scalar finite laws by floating-point Gauss–Hermite quadrature and check the entropy, correlation, and integration-by-parts identities for six scalar laws on two numerical meshes. They provide consistency checks only, not certified error enclosures or a substitute for the proof. An additional derivative comparison explored in the scripts is not claimed here.

AI assistance was used in literature exploration, derivation, drafting, and computational checking; the author is responsible for the mathematical claims and the final text. This preprint has not undergone independent peer review or formal verification.

References

- [1] J. Liu, *Minoration via Mixed Volumes and Cover’s Problem for General Channels*, in *New Mathematical Techniques in Information Theory*, Oberwolfach Report 14/2022, pp. 694–695. doi:10.4171/owr/2022/14.
- [2] S. Chewi and A.-A. Pooladian, *An entropic generalization of Caffarelli’s contraction theorem via covariance inequalities*, *Comptes Rendus Mathématique* **361** (2023), 1471–1482. doi:10.5802/crmath.486.
- [3] D. Cordero-Erausquin, B. Nazaret, and C. Villani, *A mass-transportation approach to sharp Sobolev and Gagliardo–Nirenberg inequalities*, *Advances in Mathematics* **182** (2004), 307–332. doi:10.1016/S0001-8708(03)00080-X.
- [4] J. Liu, *From Soft-Minoration to Information-Constrained Optimal Transport and Spiked Tensor Models*, 2023 IEEE International Symposium on Information Theory (ISIT), pp. 666–671. arXiv:2305.08063.
- [5] J. Liu, *Two-Sided Bounds for Entropic Optimal Transport via a Rate-Distortion Integral*, 2026. arXiv:2604.14061v1.
- [6] Y. Chu and M. Raginsky, *Talagrand Meets Talagrand: Upper and Lower Bounds on Expected Soft Maxima of Gaussian Processes with Finite Index Sets*, 2025, revised 2026. arXiv:2502.06709v2.