対数 Sobolev 不等式と Talagrand 不等式の改良について

大阪大学大学院 理学研究科 数学専攻
辻　寛 (Hiroshi TSUJI)

概要

対数 Sobolev 不等式と Talagrand 不等式は相対エントロピーと，Fisher 情報量または最適輸送距離を用いて記述される不等式であり，重みつきリッチ曲率に関する幾何学の研究において，適切な条件のもと成立することが知られている．本講演では，特別な場合であるユークリッド空間上の正規分布に基づくこれらの不等式の改良について考える．とくに，インプットする分散がある意味で大きい場合に最適な形で不等式が改良されることを報告する．本研究は埼玉大学の Neal Bez 氏と大阪大学の中村昌平氏との共同研究に基づく．

1　序文

本予稿では Euclid 空間上の対数 Sobolev 不等式と Talagrand 不等式について議論する．これらは，二つの相異なる確率測度の間の“距離”を測る尺度として用いられている，相対エントロピー，Fisher 情報量，及び最適輸送距離の大小比較を与える不等式であり，幾何解析や確率論などの様々な分野において利用されている．ここでは特に Gauss 測度に基づく対数 Sobolev 不等式と Talagrand 不等式について考察する．以下でこれらの不等式について詳しく紹介する．

一般に $\mathcal{P}(\mathbb{R}^n)$ によって $\mathbb{R}^n$ 上の確率測度全体の集合を表すことにし，$\mathcal{P}^2(\mathbb{R}^n)$ をその中で二次モーメントが有限な確率測度だけを集めた集合とする．ここで確率測度 $\mu \in \mathcal{P}(\mathbb{R}^n)$ の二次モーメントが有限であるとは,

$$\int_{\mathbb{R}^n} |x|^2 \, d\mu(x) < +\infty$$

となることを指す（$|\cdot|$ は通常の Euclid ノルムとしている）．また $\mathbb{R}^n$ 上の標準正規分布を

$$d\gamma(x) = \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}|x|^2} \, dx$$

とし，γ に絶対連続な $\mathcal{P}(\mathbb{R}^n)$，$\mathcal{P}^2(\mathbb{R}^n)$ 内の確率測度をそれぞれ $\mathcal{P}_{\mathrm{ac}}(\mathbb{R}^n;\gamma)$，$\mathcal{P}^2_{\mathrm{ac}}(\mathbb{R}^n;\gamma)$ とあらわすことにする．標準正規分布 γ はすべての上記クラスに属していることを注意しておく．

与えられた確率測度 $\mu \in \mathcal{P}(\mathbb{R}^n)$ に対して，Gauss 測度に基づく μ の相対エントロピーは次で定義される：

$$\mathrm{Ent}_\gamma(\mu) := \begin{cases} \int_{\mathbb{R}^n} \rho \log \rho \, d\gamma & \mu \ll \gamma \text{ and } \rho = \frac{d\mu}{d\gamma} \\ +\infty & \text{otherwise.} \end{cases}$$

また $\rho(\log \rho)_+$ が γ に関して可積分でない場合には, 負の部分の可積分性によらずに $\mathrm{Ent}_\gamma(\mu) := +\infty$ と約束する．このとき Jensen の不等式により，相対エントロピーは常に非負であることが保証され

る．特に相対エントロピーが 0 となるのは μ が標準正規分布 γ に一致するときであり，またそのときに限る．この意味で，相対エントロピーは標準正規分布との近さを測る $\mathcal{P}_{ac}(\mathbb{R}^n;\gamma)$ 上の“距離”とみなされる．しかしながら，相対エントロピーは正確な意味での距離関数ではないことを注意しておく．

また，γ に関して局所リプシッツ連続な確率密度関数を持つ確率測度 $\mu=\rho\gamma\in\mathcal{P}_{\mathrm{ac}}(\mathbb{R}^n;\gamma)$ に対して，Gauss 測度に基づく μ の Fisher 情報量は次で与えられる：

$$I_\gamma(\mu) \coloneqq \int_{\mathbb{R}^n} \rho|\nabla \log \rho|^2\, d\gamma.$$

ここで $\rho^{-1}(0)$ 上での上記の積分は 0 とする．定義から明らかなように，Fisher 情報量も常に非負である．さらに，その量が 0 となるのは $\mu=\gamma$ となるときであり，またそのときに限る．

以上の量を用いて，対数 Sobolev 不等式は次のように記述される．

定理 1.1. 任意の $\mu=\rho\gamma\in\mathcal{P}_{\mathrm{ac}}(\mathbb{R}^n;\gamma)$ に対して，ρ が局所リプシッツ連続であれば，

$$\mathrm{Ent}_\gamma(\mu) \le \frac{1}{2} I_\gamma(\mu) \tag{1.1}$$

が成り立つ．また等号成立は μ が標準正規分布 γ，またはそれを平行移動した測度に一致するときであり，またそのときに限る．

対数 Sobolev 不等式は最初に Gross[4] によって与えられた．また等号成立条件については Carlen[2] によって調べられている．対数 Sobolev 不等式の重要な点は，その不等式内に空間次元があらわれない点である．この利点は無限次元空間上での解析を行う際に非常に有効であり，確率論などの文脈において広く利用されている．

次に Talagrand 不等式を紹介する．そのために，まず最適輸送理論の文脈で知られている最適輸送距離を導入する．一般に二つの確率測度 $\mu,\nu\in\mathcal{P}^2(\mathbb{R}^n)$ に対して，μ と ν の間の L^2-Wasserstein 距離は次で与えられる：

$$W_2(\mu,\nu) \coloneqq \left(\inf_{\pi\in\Pi(\mu,\nu)} \int_{\mathbb{R}^n\times\mathbb{R}^n} |x-y|^2\, d\pi(x,y) \right)^{\frac{1}{2}}.$$

ここで，$\Pi(\mu,\nu)$ は μ と ν のカップリング全体の集合であり，確率測度 $\pi\in\mathcal{P}(\mathbb{R}^n\times\mathbb{R}^n)$ が μ と ν のカップリングであるとは，

$$\pi(A\times\mathbb{R}^n)=\mu(A), \quad \pi(\mathbb{R}^n\times A)=\nu(A)$$

が任意の Borel 集合 $A\subset\mathbb{R}^n$ に対して成り立つことを指す．上記で与えられた最適輸送距離は，その名が表すように，一方の分布をもう一方の分布に輸送する際にかかるコストを最小化した量となっている．任意の $\mu,\nu\in\mathcal{P}^2(\mathbb{R}^n)$ に対して常に $W_2(\mu,\nu)<+\infty$ であり，W_2 は $\mathcal{P}^2(\mathbb{R}^n)$ 上の距離関数となっている．特に $(\mathcal{P}^2(\mathbb{R}^n),W_2)$ は L^2-Wasserstein 空間と呼ばれる．最適輸送距離に関するより詳しい性質や，関連する話題などについては [7], [8] などに詳しい．

以上の量を用いて Talagrand 不等式は次のように記述される．

定理 1.2. 任意の $\mu \in \mathcal{P}^2(\mathbb{R}^n)$ に対して，

$$\frac{1}{2}W_2(\mu,\gamma)^2 \le \mathrm{Ent}_\gamma(\mu) \tag{1.2}$$

が成り立つ．等号成立は μ が標準正規分布 γ，またはそれを平行移動した測度に一致するときであり，またそのときに限る．

この不等式ははじめ Talagrand[6] によって発見され，測度の集中を容易に導くことができる不等式として考案された．この不等式の利点もまた，空間次元が不等式内にあらわれない点であることを注意しておく．

注意 1.3. これまで述べてきた対数 Sobolev 不等式と Talagrand 不等式は標準正規分布に基づくものであったが，同様の不等式は正規分布とは限らない測度に対しても考えることができる．より一般に，上述した不等式たちは重みつきの測度を備えたリーマン多様体や測度距離空間等の枠組みでも意味を成す．このような空間の上において，いつ対数 Sobolev 不等式や Talagrand 不等式が成り立つかという問題が自然に考えられるが，この問題に対する一つの回答として曲率次元条件と呼ばれる枠組みが存在する．この条件は，荒く言えばリッチ曲率の下限と次元の上限を与えるものである．より詳細な定義や，これらの条件の下での不等式の構成などについては例えば [8] に詳しい．しかしながら，本予稿ではこれらの枠組みには立ち入らない．

本予稿では (1.1) と (1.2) を適切な条件下で改良することを考える．その目的のために, 次の deficit と呼ばれる量を導入することが好ましい．与えられた $\mu \in \mathcal{P}_{\mathrm{ac}}(\mathbb{R}^n;\gamma)$ に対して，対数 Sobolev 不等式の deficit を

$$\delta_{\mathrm{LSI}}(\mu) \coloneqq \frac{1}{2}I_\gamma(\mu) - \mathrm{Ent}_\gamma(\mu)$$

によって定める．同様に $\mu \in \mathcal{P}^2(\mathbb{R}^n)$ に対して，Talagrand 不等式の deficit は

$$\delta_{\mathrm{Tal}}(\mu) \coloneqq \mathrm{Ent}_\gamma(\mu) - \frac{1}{2}W_2(\mu,\gamma)^2$$

によって与えられる．いずれの場合においても，$\infty - \infty = \infty$ と約束する．(1.1), (1.2) より，これらの deficit は常に非負であることがわかる．次の節以降では，これらの deficit の非自明な下界評価について議論する．

2　分散が小さい場合の改良

この節では，我々の結果の動機となる先行研究についてご紹介したい．そのためには次の共分散行列の概念が必要となる．確率測度 $\mu \in \mathcal{P}^2(\mathbb{R}^n)$ が与えられたとき，μ の共分散行列は次で与えられる：

$$\mathrm{cov}(\mu) \coloneqq \left(\int_{\mathbb{R}^n} x_i x_j \, d\mu(x) - \left(\int_{\mathbb{R}^n} x_i \, d\mu(x)\right)\left(\int_{\mathbb{R}^n} x_j \, d\mu(x)\right)\right)_{i,j=1,2,\ldots,n}.$$

共分散行列の直感的意味を理解するために次の Gauss 測度の例を考える．

今，$\beta > 0$ に対して，Gauss 測度 γ_β を

$$d\gamma_\beta(x) \coloneqq \frac{1}{(2\pi\beta)^{n/2}} e^{-\frac{1}{2\beta}|x|^2} \, dx$$

と定める．$\beta=1$ のときは $\gamma_1=\gamma$ であることを注意しておく．この γ_β に対して，その共分散行列は

$$\mathrm{cov}(\gamma_\beta)=\beta I_n \tag{2.1}$$

となる．ここで I_n は n 次単位行列を指す．例えば，β が十分大きいとき，$\mathrm{cov}(\gamma_\beta)$ は大きくなるが，このときの γ_β のグラフは平たくつぶれたような山のグラフとなる．一方で，β が十分小さいとき，$\mathrm{cov}(\gamma_\beta)$ も小さくなるが，このときの γ_β のグラフは原点まわりに鋭い山を形成するようなフラフとなる．このように，共分散行列の値の大きさ（正確にはその固有値）は，その分布がどの程度散らばっているかを定量的に測る指標として理解される．

以上で導入した共分散行列を用いて，各不等式の deficit に関する次の結果が知られている．

定理 2.1 ([3], [5])**.** 定数 $\beta>0$ は $0<\beta\le 1$ を満たすとし，$\mu\in\mathcal{P}^2_{\mathrm{ac}}(\mathbb{R}^n;\gamma)$ とする．このとき，μ が $\mathrm{cov}(\mu)\le\beta I_n$ *1 をみたすならば，

$$\delta_{\mathrm{LSI}}(\mu)\ge\delta_{\mathrm{LSI}}(\gamma_\beta)=\frac{n}{2}(\log\beta-1+\frac{1}{\beta}), \tag{2.2}$$

$$\delta_{\mathrm{Tal}}(\mu)\ge\frac{n(2(1-\beta)+(\beta+1)\log\beta)}{2(\beta-1)} \tag{2.3}$$

が成り立つ．

(2.2) は Eldan–Lehec–Shenfeld[3] らによって与えられ，(2.3) は Mikulincer[5] によって示された．定数 β が $\beta\neq1$ のとき，(2.2), (2.3) の各右辺は真に正である．特に，(2.1) より (2.2) はシャープな改良となっている．一方で，$\beta\neq1$ のとき，

$$\delta_{\mathrm{Tal}}(\gamma_\beta)=n(\sqrt{\beta}-\frac{1}{2}\log\beta-1)>\frac{n(2(1-\beta)+(\beta+1)\log\beta)}{2(\beta-1)}$$

であることも注意しておく．特に (2.3) はシャープな改良ではないと考えられる．

定理 2.1 を受けて，次の問題が自然に考えられる：共分散行列が大きい場合には同様の結果が成り立つか．より正確には，$\beta\ge1$ と $\mu\in\mathcal{P}^2_{\mathrm{ac}}(\mathbb{R}^n;\gamma)$ が与えられたとき，μ が $\mathrm{cov}(\mu)\ge\beta I_n$ を満たすならば，(2.2), (2.3) は成り立つか．実際，$\beta>1$ が与えられた場合，任意の $\beta'\ge\beta$ に対して

$$\delta_{\mathrm{LSI}}(\gamma_{\beta'})\ge\delta_{\mathrm{LSI}}(\gamma_\beta),$$
$$\delta_{\mathrm{Tal}}(\gamma_{\beta'})\ge\delta_{\mathrm{Tal}}(\gamma_\beta)>\frac{n(2(1-\beta)+(\beta+1)\log\beta)}{2(\beta-1)}$$

が成り立つことが確認できるため，上の問題が正しいことが十分期待される．しかしながら，実際には上記の問題は否定的であることが [3], [5] において言及されている．実際，かれらは各々の研究において具体的な反例を与えた．ここでは δ_{LSI} に対する [3] での反例を紹介する．[5] では，δ_{Tal} に対して同様の反例を構成している．

今，各 $k\in\mathbb{N}$ に対して，$\mathbb{R}$ 上の確率測度 γ^k を中心が k^2，分散が 1 となる Gauss 測度とし，$\mathbb{R}$ 上の確率測度 μ^k を，

$$\mu^k:=(1-\frac{1}{k})\gamma+\frac{1}{k}\gamma^k$$

*1 一般に対称行列 $A,B\in\mathbb{R}^{n\times n}$ が $A\ge B$ を満たすとは，任意の $x\in\mathbb{R}^n$ に対して，$\langle Ax,x\rangle\ge\langle Bx,x\rangle$ となることを指す．

として定める．このとき，$(\mu^k)_{k\in\mathbb{N}}$ は $k\to+\infty$ において，

$$\mathrm{cov}(\mu^k)\to+\infty$$

だが，

$$\delta_{\mathrm{LSI}}(\mu^k)\to 0$$

となることが確認できる．特に，この例は上記で述べた問題に対する反例を与えている．

しかしながら，上述したような反例がある一方で，Gauss 測度 $(\gamma_{\beta'})_{\beta'\geq\beta}$ に対して上述した問題は肯定的でもある．次節ではこのような事実を受けて，どのような枠組みであれば“共分散行列が大きい”場合にも同様の結論が得られるかについて考える．

3　分散が大きい場合の改良

前節で述べたように，共分散行列が大きい場合に deficit の改良を与えるためには，共分散行列に制限を与えるだけでは不十分であることがわかった．この節では共分散行列を考える代わりに，次のような概念を考える．

定義 3.1. 定数 $\beta>0$ と C^2 級関数 $f\colon\mathbb{R}^n\to(0,\infty)$ を固定する．このとき，f が

- β-semi-log-convex であるとは，
$$\nabla^2\log f\geq-\frac{1}{\beta}I_n.$$
- β-semi-log-concave であるとは，
$$\nabla^2\log f\leq-\frac{1}{\beta}I_n.$$
- β-semi-log-subharmonic であるとは，
$$\Delta\log f\geq-\frac{n}{\beta}.$$

ここで，トレースを考えることにより，

$$\beta\text{-semi-log-convex}\quad\Rightarrow\quad\beta\text{-semi-log-subharmonic}$$

が直ちに従うことを注意しておく．上で導入された概念は次の命題を通して，共分散行列と関連していることがわかる．

命題 3.2. 定数 $\beta>0$ と C^2 級関数 $f\colon\mathbb{R}^n\to(0,\infty)$ を固定し，f を確率密度関数にもつ確率測度 $d\mu(x)\coloneqq f(x)\,dx$ は $\mu\in\mathcal{P}^2(\mathbb{R}^n)$ であるとする．このとき，

- $$\nabla^2\log f\geq-\frac{1}{\beta}I_n\quad\Rightarrow\quad\mathrm{cov}(\mu)\geq\beta I_n.$$
- $$\nabla^2\log f\leq-\frac{1}{\beta}I_n\quad\Rightarrow\quad\mathrm{cov}(\mu)\leq\beta I_n.$$

- $$\Delta \log f \geq -\frac{n}{\beta} \quad \Rightarrow \quad \mathrm{tr}\,\mathrm{cov}(\mu) \geq n\beta.$$

これらの性質は特に，β-semi-log-convexity または β-semi-log-subharmonicity は分散が大きい場合に対応しており，β-semi-log-concavity は分散が小さい場合に対応していることを意味している．

これらの概念に基づいて，次の結果が確認できることが本予稿の主結果である．

定理 3.3 ([1])**.** 定数 $\beta > 0$ と $\mu \in \mathcal{P}^2_{\mathrm{ac}}(\mathbb{R}^n;\gamma)$ を固定し，$f := d\mu/dx$ は正値 C^2 級関数とする．このとき，μ と β が，$\beta \geq 1$ かつ f が *β-semi-log-subharmonic*，または，$\beta \leq 1$ かつ f が *β-semi-log-concave*，のいずれかを満たすならば，(2.2) が成り立つ．

定理 3.4 ([1])**.** 定数 $\beta > 0$ と $\mu \in \mathcal{P}^2_{\mathrm{ac}}(\mathbb{R}^n;\gamma)$ を固定し，$f := d\mu/dx$ は正値 C^2 級関数とする．このとき，μ と β が，$\beta \geq 1$ かつ f が *β-semi-log-convex* と $\nabla^2 \log f \leq 0$，または，$\beta \leq 1$ かつ f が *β-semi-log-concave*，のいずれかを満たすならば，次が成り立つ：

$$\delta_{\mathrm{Tal}}(\mu) \geq \delta_{\mathrm{Tal}}(\gamma_\beta) = n(\sqrt{\beta} - \frac{1}{2}\log \beta - 1). \tag{3.1}$$

命題 3.2 より，$\beta \geq 1$ かつ f が β-semi-log-subharmonic，または β-semi-log-convex を満たす場合は，μ の共分散行列が大きいことを意味しているため，2 節において述べた問題に対するある種の回答を与えている．とくに，定理 3.3 は Eldan–Lehec–Shenfeld[3] らによる対数 Sobolev 不等式の改良と対をなす改良であるとみなせる．一方で，$\beta \leq 1$ かつ f が β-semi-log-concave を満たす場合は，μ の共分散行列が小さい場合を意味しているため，対数 Sobolev 不等式に対する新しい改良は与えていない．一方で，Talagrand 不等式の deficit に関してはその限りではなく，Mikulincer[5] による deficit の評価よりもよい評価を提示している．

参考文献

[1] N. Bez, S. Nakamura, H. Tsuji, *Stability of hypercontractivity, the logarithmic Sobolev inequality, and Talagrand's cost inequality*, arXiv:2201.12478.

[2] E. Carlen, *Supper additivity of Fisher's information and logarithmic Sobolev inequalities*, J. Funct. Anal. **101** (1991), 194–211.

[3] R. Eldan, J. Lehec, Y. Shenfeld, *Stability of the logarithmic Sobolev inequality via the Föllmer process*, Ann. Inst. Henri Poincaré Probab. Stat. **56** (2020), 2253–2269.

[4] L. Gross, *Logarithmic Sobolev inequalities*, Amer. J. Math. **97** (1975), 1061–1083.

[5] D. Mikulincer, *Stability of Talagrand's Gaussian transport-entropy inequality via the Föllmer process*, Israel J. Math. **242** (2021), 215–241.

[6] M. Talagrand, *Transportation cost for Gaussian and other product measures*, Geom. Funct. Anal. **6** (1996), no. 3, 587–600.

[7] C. Villani, *Topics in optimal transportation*, Graduate Studies in Mathematics, 58, American Mathematical Society, Providence, RI, 2003.

[8] C. Villani, *Optimal transport, old and new*, Springer-Verlag, Berlin, 2009.