Summary Part 2: Concentration of Measure and Functional Methods

Tianpei Xie

Jan. 26th., 2023

Contents

1 Logarithmic Sobolev Inequality

1.1 Functional Form of Logarithmic Sobolev Inequality

- From functional analysis, we have *the Sobolev inequality*,

  Remark (***The Sobolev Inequality***) [Evans, 2010]
  The Sobolev inequality states for smooth function $f : \mathbb{R}^n \to \mathbb{R}$ in *Sobolev space* where $n \geq 3$ and $p = \frac{2n}{n-2} > 2$

  $$\|f\|_p^2 \leq C_n \int_{\mathbb{R}^n} |\nabla f|^2 \, dx.$$

  The inequality is sharp when the constant

  $$C_n := \frac{1}{\pi n(n-2)} \left(\frac{\Gamma(n)}{\Gamma(n/2)} \right)^{2/n}$$

- **Proposition 1.1** *(**Euclidean Logarithmic Sobolev Inequality**).*
  Let $f : \mathbb{R}^n \to \mathbb{R}$ be a smooth function and m be Lebesgue measure on $\mathbb{R}^n$, then

  $$\begin{aligned} Ent_m(f^2) &\leq \frac{n}{2} \log\left(\frac{2}{n\pi e} \mathbb{E}_m \left[\|\nabla f\|_2^2 \right] \right) \quad (1) \\ \Leftrightarrow \int f^2 \log\left(\frac{f^2}{\int f^2 dx} \right) dx &\leq \frac{n}{2} \log\left(\frac{2}{n\pi e} \int |\nabla f|^2 \, dx \right) \end{aligned}$$

- **Definition** (***Logarithmic Sobolev Inequality for General Probability Measure***).
  A probability measure μ on $\mathbb{R}^n$ is said to satisfy the ***logarithmic Sobolev inequality*** for some constant $C > 0$ if for any smooth function f

  $$\mathrm{Ent}_\mu(f^2) \leq C \, \mathbb{E}_\mu \left[\|\nabla f\|_2^2 \right] \quad (2)$$

  holds for any ***continuous differentiable*** function $f : \mathbb{R}^n \to \mathbb{R}$. The left-hand side is called ***the entropy functional***, which is defined as

  $$\begin{aligned} \mathrm{Ent}(f^2) &:= \mathbb{E}_\mu \left[f^2 \log f^2 \right] - \mathbb{E}_\mu \left[f^2 \right] \log \mathbb{E}_\mu \left[f^2 \right] \\ &= \int f^2 \log\left(\frac{f^2}{\int f^2 d\mu} \right) d\mu. \end{aligned}$$

  The right-hand side is defined as

  $$\mathbb{E}_\mu \left[\|\nabla f\|_2^2 \right] = \int \|\nabla f\|_2^2 \, d\mu.$$

  Thus we can rewrite *the logarithmic Sobolev inequality* in *functional form*

  $$\int f^2 \log\left(\frac{f^2}{\int f^2 d\mu} \right) d\mu \leq C \int \|\nabla f\|_2^2 \, d\mu \quad (3)$$

- **Remark** (***Logarithmic Sobolev Inequality***)
  For non-negative function f, we can replace $f \to \sqrt{f}$, so that *the logarithmic Sobolev inequality* becomes

$$\mathrm{Ent}_{\mu}(f) \le C \int \frac{\|\nabla f\|_2^2}{f} d\mu \tag{4}$$

- **Remark** (***Modified Logarithmic Sobolev Inequality via Convex Cost and Duality***)
  For some ***convex non-negative cost*** $c : \mathbb{R}^n \to \mathbb{R}_+$, ***the convex conjugate*** of c (Legendre transform of c) is defined as

$$c^*(x) := \sup_{y} \{\langle x\,,\, y\rangle - c(y)\}$$

  Then we can obtain ***the modified logarithmic Sobolev inequality***

$$\mathrm{Ent}_{\mu}(f) \le \int f^2\, c^* \left(\frac{\nabla f}{f}\right) d\mu \tag{5}$$

1.2 Bernoulli Logarithmic Sobolev Inequality

- **Remark** (*Setting*)
  Consider a ***uniformly distributed binary vector*** $Z = (Z_1\,, \ldots, Z_n)$ on the hypercube $\{-1, +1\}^n$. In other words, the components of X are *independent*, *identically distributed* ***random sign (Rademacher) variables*** with $\mathbb{P}\{Z_i = -1\} = \mathbb{P}\{Z_i = +1\} = 1/2$ (i.e. *symmetric Bernoulli random variables*).

  Let $f : \{-1, +1\}^n \to \mathbb{R}$ be a real-valued function on ***binary hypercube***. $X := f(Z)$ is an induced real-valued random variable. Define $\widetilde{Z}^{(i)} = (Z_1\,, \ldots, Z_{i-1}, Z_i', Z_{i+1}\,, \ldots, Z_n)$ be the sample Z with i-th component replaced by an *independent copy* Z_i'. Since $Z, \widetilde{Z}^{(i)} \in \{-1, +1\}^n$, $\widetilde{Z}^{(i)} = (Z_1\,, \ldots, Z_{i-1}, -Z_i, Z_{i+1}\,, \ldots, Z_n)$, i.e. *the i-th sign is **flipped***. Also denote the i-th *Jackknife sample* as $Z_{(i)} = (Z_1\,, \ldots, Z_{i-1}, Z_{i+1}\,, \ldots, Z_n)$ by *leaving out* the i-th component. $\mathbb{E}_{(-i)}[X] := \mathbb{E}\left[X | Z_{(i)}\right]$.

  Denote the i-th component of ***discrete gradient*** of f as

$$\nabla_i f(z) := \frac{1}{2}\left(f(z) - f(\widetilde{z}^{(i)})\right)$$

  and $\nabla f(z) = (\nabla_1 f(z)\,, \ldots, \nabla_n f(z))$

- **Proposition 1.2** *(**Logarithmic Sobolev Inequality for Rademacher Random Variables**). [Boucheron et al., 2013]*
  *If $f : \{-1, +1\}^n \to \mathbb{R}$ be an arbitrary real-valued function defined on the n-dimensional **binary hypercube** and assume that Z is **uniformly distributed** over $\{-1, +1\}^n$. Then*

$$Ent(f^2) \le \mathcal{E}(f) \tag{6}$$

$$\Leftrightarrow Ent(f^2(Z)) \le 2\mathbb{E}\left[\|\nabla f(Z)\|_2^2\right] \tag{7}$$

- **Remark** (***Logarithmic Sobolev Inequality*** $\Rightarrow$ ***Efron-Stein Inequality***). [Boucheron et al., 2013]

Note that for f non-negative,

$$\mathrm{Var}(f(Z)) \le \mathrm{Ent}(f^2(Z)).$$

Thus *logarithmic Sobolev inequality* (6) implies

$$\mathrm{Var}(f(Z)) \le \mathcal{E}(f)$$

which is *the Efron-Stein inequality.*

- **Corollary 1.3** ***(Logarithmic Sobolev Inequality for Asymmetric Bernoulli Random Variables)****. [Boucheron et al., 2013]*
  If $f : \{-1, +1\}^n \to \mathbb{R}$ be an arbitrary real-valued function and $Z = (Z_1, \ldots, Z_n) \in \{-1, +1\}^n$ with $p = \mathbb{P}\{Z_i = +1\}$. Then

$$Ent(f^2) \le c(p)\mathbb{E}\left[\|\nabla f(Z)\|_2^2\right] \tag{8}$$

  where

$$c(p) = \frac{1}{1-2p}\log\frac{1-p}{p}$$

  Note that $\lim_{p\to 1/2} c(p) = 2.$

1.3 Gaussian Logarithmic Sobolev Inequality

- **Proposition 1.4** ***(Gaussian Logarithmic Sobolev Inequality)****. [Boucheron et al., 2013]*
  Let $f : \mathbb{R}^n \to \mathbb{R}$ be a ***continuous differentiable*** *function and let $Z = (Z_1, \ldots, Z_n)$ be a vector of n* ***independent standard Gaussian*** *random variables. Then*

$$Ent(f^2(Z)) \le 2\mathbb{E}\left[\|\nabla f(Z)\|_2^2\right]. \tag{9}$$

1.4 Modified Logarithmic Sobolev Inequalities

- **Proposition 1.5** ***(A Modified Logarithmic Sobolev Inequalities for Moment Generating Function)*** *[Boucheron et al., 2013]*
  Consider independent random variables $Z_1, \ldots, Z_n$ taking values in $\mathcal{X}$, a real-valued function $f : \mathcal{X}^n \to \mathbb{R}$ and the random variable $X = f(Z_1, \ldots, Z_n)$. Also denote $Z_{(-i)} = (Z_1, \ldots, Z_{i-1}, Z_{i+1}, \ldots, Z_n)$ and $X_{(-i)} = f_i(Z_{(-i)})$ where $f_i : \mathcal{X}^{n-1} \to \mathbb{R}$ is an arbitrary function. Let $\phi(x) = e^x - x - 1$. Then for all $\lambda \in \mathbb{R}$,

$$Ent(e^{\lambda X}) := \mathbb{E}\left[\lambda X e^{\lambda X}\right] - \mathbb{E}\left[e^{\lambda X}\right]\log\mathbb{E}\left[e^{\lambda X}\right] \le \sum_{i=1}^n \mathbb{E}\left[e^{\lambda X}\phi(-\lambda(X - X_{(-i)}))\right] \tag{10}$$

- **Proposition 1.6** ***(Symmetrized Modified Logarithmic Sobolev Inequalities)*** *[Boucheron et al., 2013]*

Consider independent random variables $Z_1, \ldots, Z_n$ taking values in $\mathcal{X}$, a real-valued function $f : \mathcal{X}^n \to \mathbb{R}$ and the random variable $X = f(Z_1, \ldots, Z_n)$. Also denote $\widetilde{X}^{(i)} = f(Z_1, \ldots, Z_{i-1}, Z_i', Z_{i+1}, \ldots, Z_n)$. Let $\phi(x) = e^x - x - 1$. Then for all $\lambda \in \mathbb{R}$,

$$\lambda \mathbb{E}\left[X e^{\lambda X}\right] - \mathbb{E}\left[e^{\lambda X}\right] \log \mathbb{E}\left[e^{\lambda X}\right] \le \sum_{i=1}^{n} \mathbb{E}\left[e^{\lambda X} \phi(-\lambda(X - \widetilde{X}^{(i)}))\right] \tag{11}$$

Moreover, denoting $\tau(x) = x(e^x - 1)$, for all $\lambda \in \mathbb{R}$,

$$\lambda \mathbb{E}\left[X e^{\lambda X}\right] - \mathbb{E}\left[e^{\lambda X}\right] \log \mathbb{E}\left[e^{\lambda X}\right] \le \sum_{i=1}^{n} \mathbb{E}\left[e^{\lambda X} \tau(-\lambda(X - \widetilde{X}^{(i)})_+)\right],$$

$$\lambda \mathbb{E}\left[X e^{\lambda X}\right] - \mathbb{E}\left[e^{\lambda X}\right] \log \mathbb{E}\left[e^{\lambda X}\right] \le \sum_{i=1}^{n} \mathbb{E}\left[e^{\lambda X} \tau(\lambda(\widetilde{X}^{(i)} - X)_-)\right].$$

2 Isoperimetric Inequalities and Concentration of Measure

2.1 Brunn-Minkowski Inequality

- **Definition** (***Minkowski Sum of Sets***)
  Consider sets $A, B \subseteq \mathbb{R}^n$ and define ***the Minkowski sum*** of A and B as the set of all vectors in $\mathbb{R}^n$ formed by sums of elements of A and B:

  $$A + B := \{x + y : x \in A, y \in B\}$$

  Similarly, for $c \in \mathbb{R}$, let $cA = \{cx : x \in A\}$. Denote by $\mathrm{Vol}(A)$ the ***Lebesgue measure*** of a *(measurable) set $A \subset \mathbb{R}^n$.*

- **Theorem 2.1** *(**The Prékopa-Leindler Inequality**). [Boucheron et al., 2013, Wainwright, 2019]*
  *Let $\lambda \in (0,1)$, and let $f, g, h : \mathbb{R}^n \to [0, \infty)$ be **non-negative measurable functions** such that for all $x, y \in \mathbb{R}^n$,*

  $$h\left(\lambda x + (1-\lambda) y\right) \ge f(x)^{\lambda} g(y)^{1-\lambda}.$$

  Then

  $$\int_{\mathbb{R}^n} h(x) dx \ge \left(\int_{\mathbb{R}^n} f(x) dx\right)^{\lambda} \left(\int_{\mathbb{R}^n} g(x) dx\right)^{1-\lambda}. \tag{12}$$

- **Corollary 2.2** *(**Weaker Brunn-Minkowski Inequality**) [Boucheron et al., 2013, Wainwright, 2019]*
  *Let $A, B \subset \mathbb{R}^n$ be **non-empty compact sets**. Then for all $\lambda \in [0,1]$,*

  $$Vol\left(\lambda A + (1-\lambda) B\right) \ge Vol(A)^{\lambda}\, Vol(B)^{1-\lambda}. \tag{13}$$

- **Theorem 2.3** *(**Brunn-Minkowski Inequality**) [Boucheron et al., 2013, Vershynin, 2018, Wainwright, 2019]*
  *Let $A, B \subset \mathbb{R}^n$ be **non-empty compact sets**. Then for all $\lambda \in [0,1]$,*

  $$Vol\left(\lambda A + (1-\lambda) B\right)^{\frac{1}{n}} \ge \lambda\, Vol(A)^{\frac{1}{n}} + (1-\lambda)\, Vol(B)^{\frac{1}{n}}. \tag{14}$$

2.2 Classical Isoperimetric Problem on Euclidean Space $\mathbb{R}^n$

- **Definition** (***Blowup of Sets***)
  For any $t > 0$, and any (measurable) sets $A \subset \mathbb{R}^n$, ***the t-blowup (or, t-enlargement)*** *of* A is defined by
  $$A_t := \{x \in \mathbb{R}^n : d(x, A) < t\} = A + t\,B$$
  where $B = \{x \in \mathbb{R}^n : d(0, x) < 1\}$ is an *open unit ball* and $d(x, A) = \inf_{y \in A} d(x, y)$.

- **Definition** (***Surface Area of Sets***)
  let $A \subset \mathbb{R}^n$ be a measurable set and denote by $\mathrm{Vol}(A)$ its *Lebesgue measure. The* ***surface area*** *of* A is defined by
  $$\mathrm{Vol}(\partial A) = \lim_{t \to 0} \frac{\mathrm{Vol}(A_t) - \mathrm{Vol}(A)}{t}.$$
  provided that the limit exists. Here A_t denotes *the t-blowup* of A.

- **Remark** (***Isoperimetry Theorem***)
  The classical isoperimetric theorem in $\mathbb{R}^n$ states that, among all sets with ***a given volume***, ***the Euclidean unit ball minimizes the surface area***. This theoerm can be formally stated as below:

- **Theorem 2.4** ***(Isoperimetry Theorem)*** *[Boucheron et al., 2013, Vershynin, 2018, Wainwright, 2019]*
  Let $A \subset \mathbb{R}^n$ *be such that Vol(A) = Vol(B) where* $B := \{x \in \mathbb{R}^n : d(0, x) < 1\}$ *is an unit ball. Then for any* $t > 0$,
  $$Vol(A_t) \geq Vol(B_t) \tag{15}$$
  Moreover, if $Vol(\partial A)$ *exists, then*
  $$Vol(\partial A) \geq Vol(\partial B). \tag{16}$$

- **Example** (***Concentration of Lebesgue Measure in*** $\mathbb{R}^n$ ***and Isoperimetric Inequality***)
  Note that the volume of a *t-ball in* $\mathbb{R}^n$ is
  $$\mathrm{Vol}(tB) = \frac{\pi^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2} + 1\right)} t^n \equiv c_n t^n$$
  Thus the radius of ball B with the same volume of A is
  $$r := \left(\frac{\mathrm{Vol}(A)}{c_n}\right)^{\frac{1}{n}}.$$
  The classical isoperimetric inequality states that
  $$\mathrm{Vol}(A_t)) \geq \left((r + t)\mathrm{Vol}(B)^{1/n}\right)^n$$
  $$\Leftrightarrow \mathrm{Vol}(A_t) \geq c_n \left(\left(\frac{\mathrm{Vol}(A)}{c_n}\right)^{\frac{1}{n}} + t\right)^n$$
  $$\Leftrightarrow \left(\frac{\mathrm{Vol}(A_t)}{c_n}\right)^{\frac{1}{n}} \geq \left(\frac{\mathrm{Vol}(A)}{c_n}\right)^{\frac{1}{n}} + t \tag{17}$$

- **Definition** (***Isoperimetric Function of Probability Measure***)
  Define ***the isoperimetric function*** of *the Lebesgue measure space* $(\mathbb{R}^n, \mu)$ as

$$\lambda(u) := \left(\frac{u}{c_n}\right)^{\frac{1}{n}}$$

  so *the classical isoperimetric inequality* is equivalent to *the concentration of Lebesgue measure*

$$\lambda\left(\mu(A_t)\right) \geq \lambda\left(\mu(A)\right) + t.$$

2.3 Isoperimetric Problem on Unit Sphere

- **Definition** (***Spherical Cap and its t-Blowup***)
  Let $\mathbb{S}^{n-1} := \{x \in \mathbb{R}^n : \|x\| = 1\}$ be the $(n-1)$-*dimensional* ***unit sphere***. The ***intersection*** of a ***half-space*** and $\mathbb{S}^{n-1}$ is called a ***spherical cap***. In particular, for some $y \in \mathbb{R}^n$, denote the associated spherical cap as

$$H_y := \left\{x \in \mathbb{S}^{n-1} : \langle x\,,\, y\rangle \leq 0\right\}$$

  With some simple geometry, it can be shown that its *t-blowup* corresponds to the set

$$H_y^t := \left\{x \in \mathbb{S}^{n-1} : \langle x\,,\, y\rangle < \sin(t)\right\}$$

- **Theorem 2.5** *(**Isoperimetry Theorem on Unit Sphere**) [Boucheron et al., 2013, Vershynin, 2018, Wainwright, 2019]*
  *Let A be a subset of the sphere $\mathbb{S}^{n-1}$, and let σ denote the **normalized area** on that sphere. Let $t > 0$. Then, among all sets $A \subset \mathbb{S}^{n-1}$ with given area $\sigma(A)$, the **spherical caps minimize the area of the neighborhood** $\sigma(A_t)$, where*

$$A_t := \left\{x \in \mathbb{S}^{n-1} : \exists y \in A \text{ such that } \|x - y\| < t\right\}$$

- **Remark** Define a *metric* ρ on sphere $\mathbb{S}^{n-1}$ as

$$\rho(x, y) := \arccos(\langle x\,,\, y\rangle)$$

  Thus $(\mathbb{S}^{n-1}, \rho)$ is a ***metric space***. Let $\mathbb{P}$ be uniform distribution on $\mathbb{S}^{n-1}$ so that $((\mathbb{S}^{n-1}, \rho), \mathbb{P})$ is a probability space.

- **Proposition 2.6** *(**Isoperimetric Inequalities for Uniform Distribution over Sphere**) [Boucheron et al., 2013, Vershynin, 2018, Wainwright, 2019]*
  *Let $\mathbb{S}^{n-1} := \{x \in \mathbb{R}^n : \|x\| = 1\}$ be the $(n-1)$-dimensional **unit sphere**. For any $t \in [0, 1]$,*

$$\alpha_{\mathbb{S}^{n-1}}(t) \leq c \exp\left(-\frac{nt^2}{2}\right) \tag{18}$$

  for some constant c.

- By Levy's inequality, we have the following proposition

Proposition 2.7 ***(Lipschitz Function on*** $\mathbb{S}^{n-1}$*) [Wainwright, 2019]*
For any 1*-Lipschitz function* f *defined on the sphere* $\mathbb{S}^{n-1}$*, we have the two-sided bound*

$$\mathbb{P}\left\{|f(Z) - Med(f(Z))| \geq t\right\} \leq \sqrt{2\pi} \exp\left(-\frac{nt^2}{2}\right) \tag{19}$$

Moreover, replacing median by the mean, we have

$$\mathbb{P}\left\{|f(Z) - \mathbb{E}\left[f(Z)\right]| \geq t\right\} \leq 2\sqrt{2\pi} \exp\left(-\frac{nt^2}{8}\right) \tag{20}$$

- **Exercise 2.8** ***(The Blow-Up Phenomenon)***
Let A *be a subset of the sphere* $\sqrt{n}\mathbb{S}^{n-1}$ *such that*

$$\mathbb{P}\left(A\right) > 2\exp\left(-cs^2\right) \text{ for some } s > 0;$$

1. *Prove that* $\mathbb{P}(A_s) > 1/2$.
2. *Deduce from this that for any* $t \geq s$,

$$\mathbb{P}\left(A_{2t}\right) > 1 - 2\exp\left(-ct^2\right).$$

Here $c > 0$ *is the absolute constant in upper bound of concentration function.*

2.4 Concentration via Isoperimetric Inequalities

- **Definition** (***Isoperimetry Problem***) [Boucheron et al., 2013]
Given a ***metric space*** $\mathcal{X}$ with corresponding *distance* d, consider ***the measure space*** formed by $\mathcal{X}$, *the* σ*-algebra* of all ***Borel sets*** *of* $\mathcal{X}$, and a probability measure $\mathbb{P}$. Let X be a *random variable* taking values in $\mathcal{X}$, distributed according to $\mathbb{P}$.

  The isoperimetric problem in this case is the following: given $p \in (0,1)$ and $t > 0$, ***determine the sets*** A with $\mathbb{P}\left[X \in A\right] \geq p$ for which *the measure*

$$\mathbb{P}\left[d(X, A) \geq t\right]$$

  is ***maximal***.

- **Remark** (***Isoperimetric Inequalities***)
Even though the exact solution is only known in a few special cases, useful *bounds* for $\mathbb{P}\left[d(X, A) \geq t\right]$ can be derived under remarkably general circumstances. *Such bounds are usually referred to as* ***isoperimetric inequalities***.

- **Definition** (***Concentration Function***) [Boucheron et al., 2013, Wainwright, 2019]
The concentration function $\alpha : [0, \infty) \to \mathbb{R}_+$ associated with ***metric measure space*** $((\mathcal{X}, d), \mathbb{P})$ is given by

$$\alpha_{\mathbb{P},(\mathcal{X},d)}(t) := \sup_{A \subset \mathcal{X}:\, \mathbb{P}(A) \geq \frac{1}{2}} \mathbb{P}\left[d(X, A) \geq t\right] = \sup_{A \subset \mathcal{X}:\, \mathbb{P}(A) \geq \frac{1}{2}} \mathbb{P}\left(A_t^c\right)$$

  where $A_t := A + tB = \{x \in \mathcal{X} : d(x, A) < t\}$ is *the* t*-blowup* of $A \subset \mathcal{X}$. We simply denote it as $\alpha(t)$.

  Thus the optimal A^* for isoperimetry problem is the one that attains the $\alpha(t) = \mathbb{P}\left(A_t^c\right)$.

- **Theorem 2.9** ***(Levy's Inequalities)****[Boucheron et al., 2013, Wainwright, 2019]*
  For any Lipschitz function $f : \mathcal{X} \to \mathbb{R}$,
  $$\begin{aligned} \mathbb{P}\{f(X) \geq Med(f(X)) + t\} \leq \alpha_{\mathbb{P}}(t) \\ \mathbb{P}\{f(X) \leq Med(f(X)) - t\} \leq \alpha_{\mathbb{P}}(t). \end{aligned} \tag{21}$$
  where $Med(f(X))$ is ***the median*** *of $f(X)$, i.e.*
  $$\mathbb{P}\{f(X) \leq Med(f(X)\} \geq \frac{1}{2}, \quad \text{and } \mathbb{P}\{f(X) \geq Med(f(X)\} \geq \frac{1}{2}.$$
  Conversely*, if $\beta : \mathbb{R}_+ \to [0,1]$ is a function such that for* ***every Lipschitz function*** *$f : \mathcal{X} \to \mathbb{R}$*
  $$\mathbb{P}\{f(X) - Med(f(X)) \geq t\} \leq \beta(t). \tag{22}$$
  then *$\beta(t) \geq \alpha_{\mathbb{P}}(t)$.*

- **Corollary 2.10** ***(Concentration of Measure on Hamming Metric Space)*** *[Boucheron et al., 2013]*
  Consider independent random variables $Z_1, \ldots, Z_n$ *taking their values in a* (measurable) set $\mathcal{X}$ *and denote the vector of these variables by $Z = (Z_1, \ldots, Z_n)$ taking its value in $\mathcal{X}^n$. For an arbitrary (measurable) set $A \subset \mathcal{X}^n$, we write $\mathbb{P}(A) = \mathbb{P}(Z \in A)$. The* ***Hamming distance*** *$d_H(x, y)$ between the vectors $x, y \in \mathcal{X}^n$ is defined as* ***the number of coordinates*** *in which x and y* ***differ****. Then for any $t > 0$,*
  $$\mathbb{P}\left\{d_H(x, A) \geq \sqrt{\frac{n}{2}\log\frac{1}{\mathbb{P}(A)}} + t\right\} \leq \exp\left(-\frac{2t^2}{n}\right) \tag{23}$$

- **Remark** (***Equivalent Form***)
  From above isoperimetric inequality,
  $$\mathbb{P}\left\{d_H(x, A) \geq \sqrt{\frac{n}{2}\log\frac{1}{\mathbb{P}(A)}} + t\right\} \leq \exp\left(-\frac{2t^2}{n}\right)$$
  Denote $u := \sqrt{\frac{n}{2}\log\frac{1}{\mathbb{P}(A)}}$. By change of variable, for any $t \geq u$,
  $$\mathbb{P}\{d_H(x, A) \geq t\} \leq \exp\left(-\frac{2\,(t-u)^2}{n}\right).$$
  On the one hand, if $t \leq 2u = \sqrt{-2n \log \mathbb{P}(A)}$, then $\mathbb{P}(A) \leq \exp(-t^2/(2n))$. On the other hand, since $(t-u)^2 \geq t^2/4$ for $t \geq 2u = \sqrt{-2n \log \mathbb{P}(A)}$. the inequality above implies $\mathbb{P}\{d_H(x, A) \geq t\} \leq \exp(-t^2/(2n))$. Thus, for all $t > 0$, we have ***the concentration of measure in Hamming metric space***:
  $$\mathbb{P}(A)\mathbb{P}\{d_H(x, A) \geq t\} \leq \min\{\mathbb{P}(A), \mathbb{P}\{d_H(x, A) \geq t\}\} \leq \exp\left(-\frac{t^2}{2n}\right) \tag{24}$$

- **Proposition 2.11** ***(Levy's Inequalities for Mean)****[Boucheron et al., 2013, Wainwright, 2019]*
  If $\beta : \mathbb{R}_+ \to [0,1]$ is a function such that for ***every Lipschitz function*** *$f : \mathcal{X} \to \mathbb{R}$*
  $$\mathbb{P}\{f(X) - \mathbb{E}[f(X)] \geq t\} \leq \beta(t). \tag{25}$$
  then $\beta(t) \geq \alpha_{\mathbb{P}}(t/2)$.

2.5 Convex Distance Inequality

- **Definition** (***Weighted Hamming Distance***)
  Given $\alpha = (\alpha_1, \ldots, \alpha_n)$, where $\alpha_i \geq 0$, ***the weighed Hamming distance*** between $x, y \in \mathcal{X}^n$ is defined as
  $$d_\alpha(x, y) = \sum_{i=1}^{n} \alpha_i \mathbb{1}\{x_i \neq y_i\}.$$

- **Definition** (***Convex Distance***)
  For any $x = (x_1, \ldots, x_n) \in \mathcal{X}^n$, ***the convex distance*** of x from the set A by
  $$d_T(x, A) := \sup_{\alpha \in \mathbb{R}_+^n : \|\alpha\|_2 = 1} d_\alpha(x, A)$$

- **Theorem 2.12** *(**Convex Distance Inequality**) [Boucheron et al., 2013]*
  For any subset $A \subset \mathcal{X}^n$ and $t > 0$,
  $$\mathbb{P}(A)\mathbb{P}\{d_T(X, A) \geq t\} \leq \exp\left(-\frac{t^2}{4}\right). \tag{26}$$
  $$\Leftrightarrow \mathbb{P}(A)\mathbb{P}\left\{\sup_{\alpha \in \mathbb{R}_+^n : \|\alpha\|_2 = 1} \inf_{y \in A} \sum_{i=1}^{n} \alpha_i \mathbb{1}\{x_i \neq y_i\} \geq t\right\} \leq \exp\left(-\frac{t^2}{4}\right).$$

2.6 Concentration of Convex Lipschitz Functions

- **Theorem 2.13** *(**Concentration of Separately Convex Lipschitz Functions**) [Boucheron et al., 2013]*
  *Let $Z := (Z_1, \ldots, Z_n)$ be independent random variables taking values in the interval $[0, 1]$ and let $f : [0, 1]^n \to \mathbb{R}$ be a **separately convex function** (i.e. f is convex in each coordinate while the others are fixed) such that*
  $$|f(x) - f(y)| \leq \|x - y\| \quad \text{for all } x, y \in [0, 1]^n.$$
  Then $X = f(Z_1, \ldots, Z_n)$ satisfies, for all $t > 0$,
  $$\mathbb{P}\{f(Z) - \mathbb{E}[f(Z)] \geq t\} \leq \exp\left(-\frac{t^2}{2}\right). \tag{27}$$

- **Remark** For $Z_i \in [a_i, b_i]$, and f being L-Lipschitz function, we have
  $$\mathbb{P}\{f(Z) - \mathbb{E}[f(Z)] \geq t\} \leq \exp\left(-\frac{t^2}{2L^2 \sum_{i=1}^{n} (b_i - a_i)^2}\right)$$

- With convex distance inequality, we can improve *the concentration bound* for *convex Lipschitz functions*. First, we relate convex distance with the minimal distance to convex set

  Lemma 2.14 *(**Convex Distance vs. Distance to Convex Set**) [Boucheron et al., 2013]*
  *Let $A \subset [0, 1]^n$ be a **convex set** and let $x = (x_1, \ldots, x_n) \in [0, 1]^n$. Then*
  $$d(x, A) := \inf_{y \in A} \|x - y\|_2 \leq d_T(x, A). \tag{28}$$

- **Theorem 2.15** *(**Concentration of Convex Lipschitz Functions**) [Boucheron et al., 2013]*
  *Let $Z := (Z_1, \ldots, Z_n)$ be independent random variables taking values in the interval $[0, 1]$ and let $f : [0, 1]^n \to \mathbb{R}$ be a **quasi-convex function**; that is*

  $$\{z : f(z) \le s\} \text{ is convex set for all } s \in \mathbb{R}.$$

  Moreover, f is Lipschitz function satisfying

  $$|f(x) - f(y)| \le \|x - y\| \quad \text{for all } x, y \in [0, 1]^n.$$

  Then $X = f(Z_1, \ldots, Z_n)$ satisfies, for all $t > 0$,

  $$\mathbb{P}\{f(Z) \ge Med(f(Z)) + t\} \le 2\exp\left(-\frac{t^2}{4}\right), \tag{29}$$
  $$\mathbb{P}\{f(Z) \le Med(f(Z)) - t\} \le 2\exp\left(-\frac{t^2}{4}\right).$$

- **Remark** This result can be generalized to ***convex Lipschitz function*** *of* ***sub-Gaussian random variables***. Note that in above theorems, $Z_i \in [0, 1]$ is ***sub-Gaussian*** as well.

3 Concentration of Gaussian Measure

3.1 Gaussian Isoperimetric Theorem and Gaussian Concentration Theorem

- **Remark** (***Gaussian Isoperimetric Problem***)
  The Gaussian isoperimetric problem is to determine which (Borel) sets A have ***minimal Gaussian boundary measure*** among all sets in $\mathbb{R}^n$ with a *given probability p*.

  The Gaussian isoperimetric theorem states the beautiful fact that ***the extremal sets*** are ***linear half-spaces*** *in all dimensions* and *for all p*.

- **Definition** (***Gaussian Isoperimetric Function***)
  Denote *the cumulative distribution function of standard Normal distribution*:

  $$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt := \int_{-\infty}^{x} \varphi(t)dt$$

  where $\varphi(x) := \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} = (\Phi(x))'$ is *the probability density function* of standard normal distribution. $\Phi^{-1}(x)$ is *the quantile function of normal distribution.*

  Define ***the Gaussian isoperimetric function*** as

  $$\gamma(x) := \varphi\left(\Phi^{-1}(x)\right), \quad x \in (0, 1).$$

  Also we define $\gamma(0) = \gamma(1) = 0$.

- **Remark** Note that

  $$\begin{aligned} x &= \Phi(\Phi^{-1}(x)) \\ \Rightarrow 1 &= \varphi(\Phi^{-1}(x))(\Phi^{-1}(x))' = \gamma(x)(\Phi^{-1}(x))' \\ \Leftrightarrow 1/\gamma(x) &= (\Phi^{-1}(x))'. \end{aligned}$$

The quantity $1/\gamma(x) = (\Phi^{-1}(x))'$ is known as ***quantile-density function*** *of normal distribution.*

- **Proposition 3.1** *(**Basic Property of the Gaussian Isoperimetric Function**) [Boucheron et al., 2013]*
  The Gaussian isoperimetric function γ satisfies:

  1.
  $$\gamma'(x) = -\Phi^{-1}(x), \quad \text{for all } x \in (0,1),$$

  2.
  $$\gamma(x)\gamma''(x) = -1, \quad \text{for all } x \in (0,1),$$

  3. $(\gamma')^2$ *is convex over* $(0,1)$.

- **Lemma 3.2** *(**Asymptotic Behavior of Gaussian Isoperimetric Function**) [Boucheron et al., 2013]*
  For all $x \in [0, 1/2]$,
  $$x\sqrt{\frac{1}{2}\log\frac{1}{x}} \le \gamma(x) \le x\sqrt{2\log\frac{1}{x}}.$$
  Moreover,
  $$\lim_{x\to 0}\frac{\gamma(x)}{x\sqrt{2\log\frac{1}{x}}} = 1$$

- **Proposition 3.3** *(**Bobkov's Gaussian Inequality**) [Boucheron et al., 2013]*
  Let $Z := (Z_1, \ldots, Z_n)$ be a vector of ***independent standard Gaussian*** random variables. *Let $f : \mathbb{R}^n \to [0,1]$ be a differentiable function with gradient ∇f. Then*
  $$\gamma\left(\mathbb{E}\left[f(X)\right]\right) \le \mathbb{E}\left[\sqrt{\gamma(f(X))^2 + \|\nabla f(X)\|_2^2}\right] \tag{30}$$
  where $\gamma = \varphi \circ \Phi^{-1}$ is ***the Gaussian isoperimetric function***.

- **Theorem 3.4** *(**Gaussian Isoperimetric Theorem**) [Boucheron et al., 2013] [Boucheron et al., 2013, Vershynin, 2018, Wainwright, 2019]*
  Let $\mathbb{P}$ be the ***standard Gaussian measure*** *on $\mathbb{R}^n$ and let $A \subset \mathbb{R}^n$ be a Borel set. Then*
  $$\liminf_{t\to 0}\frac{\mathbb{P}(A_t) - \mathbb{P}(A)}{t} \ge \gamma\left(\mathbb{P}(A)\right), \tag{31}$$
  where $A_t := \{x : d(x, A) < t\}$ be the t-blowup of A. Moreover, if A is a ***half-space*** *defined by $A := \{x \in \mathbb{R}^n : x_1 \le z\}$, then*
  $$\liminf_{t\to 0}\frac{\mathbb{P}(A_t) - \mathbb{P}(A)}{t} = \gamma\left(\mathbb{P}(A)\right) = \varphi(z), \tag{32}$$
  where $\gamma := \varphi \circ \Phi^{-1}$ is ***the Gaussian isoperimetric function***.

- **Proposition 3.5** ***(Differentiablity of Measure of*** t***-Blowup)*** *[Boucheron et al., 2013]*
  If A *is a* ***finite union*** *of* ***open*** *balls in* $\mathbb{R}^n$*, then* $\mathbb{P}(A_t)$ *is a* ***differentiable*** *function of* $t > 0$*.*

- Next we describe ***an equivalent version*** of ***the Gaussian isoperimetric theorem*** in the manner of *measure concentration*:

  Theorem 3.6 ***(Gaussian Concentration Theorem)*** *[Boucheron et al., 2013] [Boucheron et al., 2013, Vershynin, 2018, Wainwright, 2019]*
  Let $\mathbb{P}$ *be the* ***standard Gaussian measure*** *on* $\mathbb{R}^n$ *and let* $A \subset \mathbb{R}^n$ *be a Borel set. Then for all* $t \geq 0$*,*

$$\begin{aligned} \mathbb{P}(A_t) &\geq \Phi\left(\Phi^{-1}\left(\mathbb{P}(A)\right) + t\right). \\ \Leftrightarrow \Phi^{-1}(\mathbb{P}(A_t)) &\geq \Phi^{-1}\left(\mathbb{P}(A)\right) + t \end{aligned} \tag{33}$$

  Equality holds if A *is a* ***half-space****.*

- **Remark** (***Gaussian Concentration Theorem*** $\equiv$ ***Gaussian Isoperimetric Theorem***)
  The Gaussian concentration theorem is equivalent to *the Gaussian isoperimetric theorem* since

$$\begin{aligned} \liminf_{t \to 0} \frac{\mathbb{P}(A_t) - \mathbb{P}(A)}{t} &\geq \liminf_{t \to 0} \frac{\Phi\left(\Phi^{-1}\left(\mathbb{P}(A)\right) + t\right) - \Phi\left(\Phi^{-1}\left(\mathbb{P}(A)\right)\right)}{t} \\ &= \Phi'(\Phi^{-1}(\mathbb{P}(A))) \\ &= \varphi(\Phi^{-1}(\mathbb{P}(A))) \\ &= \gamma(\mathbb{P}(A)). \end{aligned}$$

- **Exercise 3.7** ***(From Isoperimetry to Concentration)*** *[Boucheron et al., 2013]*
  Assume that a probability distribution $\mathbb{P}$ *on* $\mathbb{R}^n$ *satisfies, for all Borel sets* $A \subset \mathbb{R}^n$*,*

$$\liminf_{t \to 0} \frac{\mathbb{P}(A_t) - \mathbb{P}(A)}{t} \geq c\, f\left(F^{-1}(\mathbb{P}(A))\right),$$

  where $c \in (0, 1]$ *is a constant,* F *is a continously differentiable distribution function and* $f = F'$ *is its derivative. Prove that for all Borel set* A *and all* $t \geq 0$*,*

$$\mathbb{P}(A_t) \geq F\left(F^{-1}(\mathbb{P}(A)) + c\,t\right).$$

3.2 Lipschitz Functions of Gaussian Variables

- **Theorem 3.8** ***(Rademacher Theorem)****.*
  If $f : U \to \mathbb{R}$ *is a L-Lipschitz function where* $U \subseteq \mathbb{R}^n$*, then* f *is* ***differentiable almost everywhere*** *in* U *and* ***the essential supremum*** *of the* ***norm*** *of its* ***derivative*** *is* ***bounded*** *by its* ***Lipschitz constant****.*

- **Theorem 3.9** ***(Lipschitz Functions of Gaussian Variables)*** *[Boucheron et al., 2013]*
  Let $Z = (Z_1, \ldots, Z_n)$ *be a vector of* n ***independent standard normal*** *random variables. Let* $f : \mathbb{R}^n \to \mathbb{R}$ *denote an* L***-Lipschitz function****, that is, there exists a constant* $L > 0$ *such that for all* $x, y \in \mathbb{R}^n$*,*

$$|f(x) - f(y)| \leq L\,\|x - y\|.$$

Then, for all $\lambda \in \mathbb{R}$,

$$\psi_{f(Z)-\mathbb{E}[f(Z)]}(\lambda) := \log \mathbb{E}\left[e^{\lambda(f(Z)-\mathbb{E}[f(Z)])}\right] \le \frac{L^2\lambda^2}{2} \tag{34}$$

- **Theorem 3.10** ***(Gaussian Concentration Inequality) (The Tsirelson-Ibragimov-Sudakov Inequality)*** *[Boucheron et al., 2013, Wainwright, 2019]*
  Let $Z = (Z_1, \ldots, Z_n)$ be a vector of n ***independent standard normal*** *random variables. Let $f : \mathbb{R}^n \to \mathbb{R}$ denote an L-**Lipschitz function**. Then, for all $t > 0$,*

$$\mathbb{P}\left\{f(Z) - \mathbb{E}\left[f(Z)\right] \ge t\right\} \le \exp\left(-\frac{t^2}{2L^2}\right). \tag{35}$$

- As a direct consequence of *the Gaussian isoperimetric inequality*, we have the improved result for *Gaussian concentration inequality*:

  Theorem 3.11 ***(Gaussian Concentration Inequality, Sharp Bound)*** *[Boucheron et al., 2013, Wainwright, 2019]*
  Let $Z = (Z_1, \ldots, Z_n)$ be a vector of n ***independent standard normal*** *random variables. Let $f : \mathbb{R}^n \to \mathbb{R}$ denote an L-**Lipschitz function**. Then, for all $t > 0$,*

$$\mathbb{P}\left\{f(Z) - Med(f(Z)) \ge t\right\} \le 1 - \Phi\left(\frac{t}{L}\right). \tag{36}$$

  where $\Phi(t)$ is the cumulative distribution function of standard normal random variable.

- **Remark** Note that by ***Gordon's inequality***

$$1 - \Phi(t) \le \left(\frac{1}{\sqrt{2\pi}}\right)\frac{1}{t}e^{-\frac{t^2}{2}} = \frac{1}{t}\varphi(t)$$

  The Gaussian concentration inequality fails to capture the corrective factor t^{-1}. The inequality above cannot be improved in general as for $f(x) = n^{-1/2}\sum_{i=1}^{n} x_i$, equality is achieved for all $t > 0$.

3.3 Gaussian Logarithmic Sobolev Inequality

- **Proposition 3.12** ***(Gaussian Logarithmic Sobolev Inequality)****. [Boucheron et al., 2013]*
  Let $f : \mathbb{R}^n \to \mathbb{R}$ be a ***continuous differentiable*** *function and let $Z = (Z_1, \ldots, Z_n)$ be a vector of n* ***independent standard Gaussian*** *random variables. Then*

$$Ent(f^2(Z)) \le 2\mathbb{E}\left[\|\nabla f(Z)\|_2^2\right]. \tag{37}$$

3.4 Gaussian Transportation Inequality

- **Theorem 3.13** ***(Talagrand's Gaussian Transportation Inequality)*** *[Boucheron et al., 2013]*

Let $\mathbb{P}$ be be the standard Gaussian probability measure on $\mathbb{R}^n$, and let $\mathbb{Q}$ be a probability measure absolutely continuous with respect to $\mathbb{P}$. Define two random vectors $X = (X_1, \ldots, X_n), Y = (Y_1, \ldots, Y_n)$ in $\mathcal{X}^n$ with distribution $\mathbb{P}$ and $\mathbb{Q}$ respectively. Then

$$\mathcal{W}_{2,d}(\mathbb{Q}, \mathbb{P}) := \sqrt{\min_{\gamma \in \Pi(\mathbb{Q},\mathbb{P})} \sum_{i=1}^{n} \mathbb{E}_{\gamma}\left[(X_i - Y_i)^2\right]} \leq \sqrt{2\mathbb{KL}\left(\mathbb{Q} \,\|\, \mathbb{P}\right)} \tag{38}$$

$$\Leftrightarrow \min_{\gamma \in \Pi(\mathbb{Q},\mathbb{P})} \sum_{i=1}^{n} \mathbb{E}_{\gamma}\left[(X_i - Y_i)^2\right] \leq 2\mathbb{KL}\left(\mathbb{Q} \,\|\, \mathbb{P}\right)$$

3.5 Gaussian Hypercontractivity

3.6 Suprema of Gaussian Process

- **Definition** (***Gaussian Process***)
  Let T be a *metric space. A stochastic process* $(X_t)_{t\in T}$ is a ***Gaussian process*** *indexed by T* if for *any finite collection* $\{t_1, \ldots, t_n\} \subset T$, the vector $(X_{t_1}, \ldots, X_{t_n})$ has a *jointly Gaussian distribution.*

  In addition, we assume that T is ***totally bounded*** (i.e. for every $t > 0$ it can be covered by *finitely many balls* of radius t) and that the *Gaussian process* is ***almost surely continuous***, that is, with probability 1, X_t is a *continuous function* of t.

- **Theorem 3.14** *(**Concentration of Suprema of Gaussian Process**) [Boucheron et al., 2013, Vershynin, 2018, Wainwright, 2019, Giné and Nickl, 2021]*
  *Let $(X_t)_{t\in T}$ be an **almost surely continuous centered Gaussian process** indexed by a **totally bounded** set T. If*

  $$\sigma^2 := \sup_{t\in T} \mathbb{E}\left[X_t^2\right],$$

  then $Z = \sup_{t\in T} X_t$ satisfies $Var(Z) \leq \sigma^2$, and for all $u > 0$,

  $$\mathbb{P}\left\{Z - \mathbb{E}\left[Z\right] \geq u\right\} \leq \exp\left(-\frac{u^2}{2\sigma^2}\right) \tag{39}$$

  and

  $$\mathbb{P}\left\{\mathbb{E}\left[Z\right] - Z \geq u\right\} \leq \exp\left(-\frac{u^2}{2\sigma^2}\right) \tag{40}$$

4 Concentration of Bernoulli Measure on the Binary Hypercube

4.1 Edge Isoperimetric Inequality on the Binary Hypercube

4.2 Bobkov's Inequality

- **Proposition 4.1** *(**Bobkov's Inequality**) [Boucheron et al., 2013]*
  Suppose Z is uniformly distributed over $\{-1, 1\}^n$. Then for all $n \geq 1$ and for all functions

$f : \{-1,1\}^n \to [0,1]$,

$$\gamma\left(\mathbb{E}\left[f(Z)\right]\right) \le \mathbb{E}\left[\sqrt{\gamma(f(Z))^2 + \|\nabla f(Z)\|_2^2}\right] \tag{41}$$

4.3 Vertex Isoperimetric Inequality on the Binary Hypercube

4.4 Hypercontractivity: The Bonami-Beckner Inequality

4.5 Influence Function

4.6 Monotone Sets

4.7 Threshold Phenomena

References

Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press, 2013.

Lawrence C Evans. *Partial differential equations*, volume 19. American Mathematical Soc., 2010.

Evarist Giné and Richard Nickl. *Mathematical foundations of infinite-dimensional statistical models*. Cambridge university press, 2021.

Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.

Martin J Wainwright. *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge University Press, 2019.