Lecture 5: Concentration of Measure and Isoperimetry

Tianpei Xie

Jan. 19th., 2023

Contents

1 The Classic Isoperimetry Inequalities

1.1 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 $\text{Vol}(A)$ the ***Lebesgue measure*** of a *(measurable) set* $A \subset \mathbb{R}^n$.

- **Theorem 1.1** *(**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 $\lambda \in [0, 1]$,*

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

  Note: a convex body in $\mathbb{R}^n$ is closed and compact set.

  Proof: (***Part 1,*** $n = 1$)
  Note that if $A \subset \mathbb{R}$, and $c \ge 0$ then $\text{Vol}(cA) = c\text{Vol}(A)$. Thus it suffice to prove

  $$\text{Vol}\,(A + B) \ge \text{Vol}(A) + \text{Vol}(B).$$

  To see this, observe that none of the three volumes involved changes if the sets A and B are ***translated*** arbitrarily. Since A, B are compact subsets in $\mathbb{R}$, it is closed and bounded. Let $a = \max\{a' : a' \in A\}$ and $b = \min\{b' : b' \in B\}$. Let $A' = A + \{-a\}$ and $B' = B + \{-b\}$ so that $A' \subset (-\infty, 0]$ and $B' \subset [0, +\infty)$. Also $\text{Vol}(A') = \text{Vol}(A)$ and $\text{Vol}(B') = \text{Vol}(B)$. However,

  $$\begin{aligned} A' \cup B' &\subset A' + B' \\ \Rightarrow \text{Vol}(A') + \text{Vol}(B') &= \text{Vol}(A' \cup B') \le \text{Vol}(A' + B') \end{aligned}$$

  This prove the 1-dimensional case for *the Brunn-Minkowski inequality.* ∎

  To prove $n > 1$ case, we need the following inequalities:

- **Theorem 1.2** *(**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 $x, y \in \mathbb{R}^n$,*

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

  Then

  $$\int_{\mathbb{R}^n} h(x)dx \ge \left(\int_{\mathbb{R}^n} f(x)dx\right)^{\lambda} \left(\int_{\mathbb{R}^n} g(x)dx\right)^{1-\lambda}. \tag{2}$$

  Proof: The proof goes by induction with respect to the dimension n.

1. ($n = 1$ ***case***). Consider measurable non-negative functions f, g, h satisfying the condition of the theorem. By *the monotone convergence theorem*, it suffices to prove the statement for ***bounded functions*** f and g. Without loss of generality, assume that $\sup_{x\in\mathbb{R}^n} f(x) = \sup_{x\in\mathbb{R}^n} g(x) = 1$. Then

$$\int_{\mathbb{R}} f(x)dx = \int_0^1 \mathrm{Vol}\left\{x : f(x) \geq t\right\} dt$$
$$\int_{\mathbb{R}} g(x)dx = \int_0^1 \mathrm{Vol}\left\{x : g(x) \geq t\right\} dt.$$

For any fixed $t \in [0, 1]$, if $f(x) \geq t$ and $g(y) \geq t$, then by the hypothesis of the theorem, $h\left(\lambda x + (1-\lambda)y\right) \geq t$. This implication may be re-written as

$$\lambda\left\{x : f(x) \geq t\right\} + (1-\lambda)\left\{x : g(x) \geq t\right\} \subset \left\{x : h(x) \geq t\right\}.$$

Thus

$$\begin{aligned}
\int_{\mathbb{R}} h(x)dx &= \int_0^\infty \mathrm{Vol}\left\{x : h(x) \geq t\right\} dt \\
&\geq \int_0^1 \mathrm{Vol}\left\{x : h(x) \geq t\right\} dt \\
&\geq \int_0^1 \mathrm{Vol}\left(\lambda\left\{x : f(x) \geq t\right\}\right) + \mathrm{Vol}\left((1-\lambda)\left\{x : g(x) \geq t\right\}\right) dt \\
&\text{(by 1-dimensional \textit{Brunn-Minkowski inequality})} \\
&\geq \lambda\int_0^1 \mathrm{Vol}\left(\left\{x : f(x) \geq t\right\}\right) dt + (1-\lambda)\int_0^1 \mathrm{Vol}\left(\left\{x : g(x) \geq t\right\}\right) dt \\
&= \lambda\int_{\mathbb{R}} f(x)dx + (1-\lambda)\int_{\mathbb{R}} g(x)dx \\
&\geq \left(\int_{\mathbb{R}} f(x)dx\right)^\lambda \left(\int_{\mathbb{R}} g(x)dx\right)^{1-\lambda} \quad \text{(by the \textit{arithmetic-geometric mean inequality})}
\end{aligned}$$

2. For the induction step, assume that the theorem holds for all dimensions $1, \ldots, n-1$ and let $f, g, h : \mathbb{R}^n \to [0, \infty)$, $\lambda \in (0, 1)$ be such that they satisfy the assumption of the theorem. Now let $x, y \in \mathbb{R}^{n-1}$ and $a, b \in \mathbb{R}$. Then

$$h\left(\lambda\left(x, a\right) + (1-\lambda)\left(y, b\right)\right) \geq f\left((x, a)\right)^\lambda g((y, b))^{1-\lambda},$$

so by the inductive hypothesis

$$\int_{\mathbb{R}^{n-1}} h\left((x, \lambda a + (1-\lambda)b)\right) dx \geq \left(\int_{\mathbb{R}^{n-1}} f\left((x, a)\right) dx\right)^\lambda \left(\int_{\mathbb{R}^{n-1}} g((x, b))dx\right)^{1-\lambda}$$

In other words, introducing

$$F(a) := \int_{\mathbb{R}^{n-1}} f\left((x, a)\right) dx, \quad G(b) := \int_{\mathbb{R}^{n-1}} g((x, b))dx$$
$$H((\lambda a + (1-\lambda)b)) := \int_{\mathbb{R}^{n-1}} h\left((x, \lambda a + (1-\lambda)b)\right) dx.$$

We have

$$H((\lambda a + (1-\lambda)b)) \geq (F(a))^{\lambda} (G(b))^{1-\lambda},$$

so by *Fubini's theorem* and the one-dimensional inequality, we have

$$\begin{aligned}\int_{\mathbb{R}^n} h(x)dx = \int_{\mathbb{R}} H(a)da &\geq \left(\int_{\mathbb{R}} F(a)da\right)^{\lambda} \left(\int_{\mathbb{R}} G(a)da\right)^{1-\lambda} \\ &= \left(\int_{\mathbb{R}^n} f(x)dx\right)^{\lambda} \left(\int_{\mathbb{R}^n} g(x)dx\right)^{1-\lambda}. \quad \blacksquare\end{aligned}$$

- **Corollary 1.3** *(**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 $\lambda \in [0,1]$,*
$$Vol(\lambda A + (1-\lambda)B) \geq Vol(A)^{\lambda} Vol(B)^{1-\lambda}. \tag{3}$$
Proof: We apply *the Prékopa-Leindler inequality* with $f(x) = \mathbb{1}\{x \in A\}$, $g(x) = \mathbb{1}\{x \in B\}$ and $h(x) = \mathbb{1}\{x \in \lambda A + (1-\lambda)B\}$. We see that
$$h(\lambda x + (1-\lambda)y) = \mathbb{1}\{\lambda x + (1-\lambda)y \in \lambda A + (1-\lambda)B\} \geq \mathbb{1}\{x \in A, y \in B\} = f(x)^{\lambda} g(y)^{1-\lambda}.$$
Thus the hypothesis of *the Prékopa-Leindler inequality* holds. $\blacksquare$

- **Proof:** ($n > 1$ ***case for Brunn-Minkowski Inequality***). First observe that it suffices to prove that for all *nonempty compact sets* A and B,
$$\mathrm{Vol}(A+B)^{\frac{1}{n}} \geq \mathrm{Vol}(A)^{\frac{1}{n}} + \mathrm{Vol}(B)^{\frac{1}{n}}$$
since $\mathrm{Vol}(cA)^{1/n} = c\mathrm{Vol}(A)^{1/n}$ for any $c \in \mathbb{R}$ and $A \subset \mathbb{R}^n$. Also notice that we may assume that $\mathrm{Vol}(A), \mathrm{Vol}(B) > 0$ because otherwise the inequality holds trivially. Defining $A' = \mathrm{Vol}(A)^{-\frac{1}{n}} A$ and $B' = \mathrm{Vol}(B)^{-\frac{1}{n}} B$, we have $\mathrm{Vol}(A') = \mathrm{Vol}(B') = 1$. By *weaker Brunn-Minkowski inequality*, for $\lambda \in (0,1)$,
$$\mathrm{Vol}\left(\lambda A' + (1-\lambda)B'\right) \geq 1.$$
Finally, we apply this *inequality* with the choice
$$\lambda = \frac{\mathrm{Vol}(A)^{\frac{1}{n}}}{\mathrm{Vol}(A)^{\frac{1}{n}} + \mathrm{Vol}(B)^{\frac{1}{n}}}$$
obtaining
$$\begin{aligned} \mathrm{Vol}\left(\frac{\mathrm{Vol}(A)^{\frac{1}{n}} A'}{\mathrm{Vol}(A)^{\frac{1}{n}} + \mathrm{Vol}(B)^{\frac{1}{n}}} + \frac{\mathrm{Vol}(B)^{\frac{1}{n}} B'}{\mathrm{Vol}(A)^{\frac{1}{n}} + \mathrm{Vol}(B)^{\frac{1}{n}}}\right) &\geq 1 \\ \Rightarrow \mathrm{Vol}\left(\frac{A}{\mathrm{Vol}(A)^{\frac{1}{n}} + \mathrm{Vol}(B)^{\frac{1}{n}}} + \frac{B}{\mathrm{Vol}(A)^{\frac{1}{n}} + \mathrm{Vol}(B)^{\frac{1}{n}}}\right) &\geq 1 \\ \Rightarrow \mathrm{Vol}\left(\frac{A+B}{\mathrm{Vol}(A)^{\frac{1}{n}} + \mathrm{Vol}(B)^{\frac{1}{n}}}\right) &\geq 1 \\ \Rightarrow \frac{\mathrm{Vol}(A+B)}{\left(\mathrm{Vol}(A)^{\frac{1}{n}} + \mathrm{Vol}(B)^{\frac{1}{n}}\right)^n} &\geq 1 \end{aligned}$$
which proves the theorem. $\blacksquare$

Figure 5.1 Isoperimetric inequality in $\mathbb{R}^n$ states that among all sets A of given volume, the Euclidean balls minimize the volume of the ε-neighborhood A_ε.

Figure 1: Isoperimetry in $\mathbb{R}^n$ [Vershynin, 2018]

1.2 The Blowup of Sets and Classical Isoperimetry Theorem

- **Definition** (***Blowup of Sets***)
  For any $t > 0$, and any (measurable) sets $A \subset \mathbb{R}^n$, ***the t-blowup*** *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 1.4** ***(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 $t > 0$,

$$Vol(A_t) \geq Vol(B_t) \tag{4}$$

  Moreover, if $Vol(\partial A)$ exists, then

$$Vol(\partial A) \geq Vol(\partial B). \tag{5}$$

  Proof: By *the Brunn-Minkowski inequality,*

$$\begin{aligned}\mathrm{Vol}(A_t)^{1/n} = \mathrm{Vol}(A + tB)^{1/n} &\geq \mathrm{Vol}(A)^{1/n} + t\mathrm{Vol}(B)^{1/n} \\ &= (1 + t)\mathrm{Vol}(B)^{1/n} \\ &= \mathrm{Vol}(B_t)^{1/n},\end{aligned}$$

establishing the first statement. The second follows simply because

$$\mathrm{Vol}(A_t) - \mathrm{Vol}(A) \geq \mathrm{Vol}(B)((1+t)^n - 1) \geq nt\mathrm{Vol}(B)$$

where $(1+t)^n \geq 1 + nt$ for $t \geq 0$. Thus $\mathrm{Vol}(\partial A) \geq n\mathrm{Vol}(B)$. The isoperimetric theorem now follows from the fact that $\mathrm{Vol}(\partial B) = n\mathrm{Vol}(B)$. ∎

2 Concentration via Isoperimetry

2.1 Levy's Inequalities

- **Remark** We can generalize the classical isoperimety problem to a probability space $(\mathcal{X}, \mathcal{B}[\mathcal{X}], \mathbb{P})$ where $\mathcal{X}$ is a *metric space* with metric d, $\mathcal{B}[\mathcal{X}]$ is the Borel σ-algrebra and $\mathbb{P}$ is a probability measure on $\mathcal{B}[\mathcal{X}]$. Let $B := \{x \in \mathbb{R}^n : d(0, x) < 1\}$. The classical isoperimetry problem aims at finding the set $A^* \subset \mathcal{X}$ that ***minimizes the surface area***

  $$\mathbb{P}(\partial A) = \lim_{t \to 0} \frac{\mathbb{P}(A_t) - \mathbb{P}(A)}{t}$$

  This is equivalent to find subset A in $\mathcal{X}$ with ***minimal t-blowup*** for given p, and for all $t > 0$

  $$A^* := \inf_{A \subset \mathcal{X} : \mathbb{P}(A) \geq p} \mathbb{P}(A_t), \quad \forall t > 0$$

  where

  $$A_t = A + tB = \{x \in \mathcal{X} : \exists y \in A \text{ s.t. } d(x, y) < t\} = \left\{x \in \mathcal{X} : \inf_{y \in A} d(x, y) := d(x, A) < t\right\}.$$

  We write the definition formally.

- **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 $\mathbb{P}[X \in A] \geq p$ for which *the measure*

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

  is ***maximal***.

- **Remark** (***Isoperimetric Inequalities***)
  Even though the exact solution is only known in a few special cases, useful *bounds* for $\mathbb{P}[d(X, A) \geq t]$ 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

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

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 $\alpha(t) = \mathbb{P}(A_t^c)$.

- **Theorem 2.1** *(**Levy's Inequalities**)[Boucheron et al., 2013, Wainwright, 2019]*
For any Lipschitz function $f : \mathcal{X} \to \mathbb{R}$,

$$\begin{aligned} \mathbb{P}\{f(X) \geq \mathit{Med}(f(X)) + t\} &\leq \alpha_{\mathbb{P}}(t) \\ \mathbb{P}\{f(X) \leq \mathit{Med}(f(X)) - t\} &\leq \alpha_{\mathbb{P}}(t). \end{aligned} \tag{6}$$

*where $\mathit{Med}(f(X))$ is **the median** of $f(X)$, i.e.*

$$\mathbb{P}\{f(X) \leq \mathit{Med}(f(X)\} \geq \frac{1}{2}, \quad \textit{and} \quad \mathbb{P}\{f(X) \geq \mathit{Med}(f(X)\} \geq \frac{1}{2}.$$

Proof: Consider the set $A = \{x : f(x) \leq \mathrm{Med}(f(X))\}$. By the definition of a *median*, $\mathbb{P}(A) \geq \frac{1}{2}$. On the other hand, by *the Lipschitz property* of f, for any $x, y \in \mathcal{X}$,

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

So for all $y \in A$, $f(y) \leq \mathrm{Med}(f(X))$

$$\begin{aligned} f(x) - \mathrm{Med}(f(X)) &\leq f(x) - f(y) \leq d(x, y) \\ \Rightarrow f(x) - \mathrm{Med}(f(X)) &\leq \inf_{y \in A} d(x, y) := d(x, A). \end{aligned}$$

Equivalently,

$$\begin{aligned} A_t := \{x \in \mathcal{X} : d(x, A) < t\} &\subseteq \{x \in \mathcal{X} : f(x) < \mathrm{Med}(f(X)) + t\} \\ \mathbb{P}(A_t^c) &\geq \mathbb{P}\{f(X) \geq \mathrm{Med}(f(X)) + t\} \end{aligned}$$

The first inequality now follows from the definition of the concentration function. The second inequality follows from the first by considering f. ∎

- **Remark** For L-Lipschitz function f, the inequality becomes

$$\mathbb{P}\{f(X) - \mathrm{Med}(f(X)) \geq t\} \leq \alpha\left(\frac{t}{L}\right), \quad \mathbb{P}\{f(X) - \mathrm{Med}(f(X)) \leq -t\} \leq \alpha\left(\frac{t}{L}\right).$$

- **Theorem 2.2** *(**Converse of Levy's Inequalities**)[Boucheron et al., 2013, Wainwright, 2019]*
*If $\beta : \mathbb{R}_+ \to [0, 1]$ is a function such that for **every Lipschitz function** $f : \mathcal{X} \to \mathbb{R}$*

$$\mathbb{P}\{f(X) - \mathit{Med}(f(X)) \geq t\} \leq \beta(t). \tag{7}$$

then $\beta(t) \geq \alpha_{\mathbb{P}}(t)$.

Proof: Note that for any $A \subset \mathcal{X}$, the function f_A defined by $f_A(x) = d(x, A)$ is *Lipschitz* since

$$|f_A(x) - f_A(y)| = |d(x, A) - d(y, A)| \leq d(x, y).$$

Also, if $P(A) \geq 1/2$, then 0 is a median of $f_A(X)$, since

$$\mathbb{P}\{f_A(x) \leq 0\} = \mathbb{P}\{d(X, A) \leq 0\} = \mathbb{P}(A) \geq \frac{1}{2}.$$

Therefore

$$\alpha(t) := \sup_{A \subset \mathcal{X}: \mathbb{P}(A) \geq 1/2} \mathbb{P}\{f_A(x) - \mathrm{Med}(f_A(X)) \geq t\} \leq \beta(t). \quad \blacksquare$$

- **Proposition 2.3** ***(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*** $f : \mathcal{X} \to \mathbb{R}$

  $$\mathbb{P}\{f(X) - \mathbb{E}[f(X)] \geq t\} \leq \beta(t). \tag{8}$$

  then $\beta(t) \geq \alpha_{\mathbb{P}}(t/2)$*.*

- **Remark** (***Isoperimetric Inequalities*** $\Leftrightarrow$ ***Concentration of Lipschitz Functions***)
  The first result points out that *isoperimetric inequalities* (more precisely, ***upper bounds*** *for* ***the concentration function***) imply *concentration of Lipschitz functions.*

  The converse shows that *concentration of Lipschitz functions* implies an *isoperimetric inequality.* In other word, *among all upper bounds* of $\mathbb{P}(A_t^c)$ for fixed A_t,

- **Corollary 2.4** ***(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* $t > 0$*,*

  $$\mathbb{P}\left\{d_H(x, A) \geq \sqrt{\frac{n}{2} \log \frac{1}{\mathbb{P}(A)}} + t\right\} \leq \exp\left(-\frac{2t^2}{n}\right) \tag{9}$$

  Proof: As we shown in previous proof, $f_A(x) = d_H(x, A)$ is a Lipschitz function with respect to Hamming distance d_H. It follows from the definition that

  $$\sup_{x \in \mathcal{X}^n, y_i \in \mathcal{X}} \left| f_A(x) - f_A(\widetilde{x}^{(i)}) \right| \leq d_H(x, \widetilde{x}^{(i)}) = 1$$

  where $\widetilde{x}^{(i)} = (x_1, \ldots, x_{i-1}, y_i, x_{i+1}, \ldots, x_n)$, so f_A has the bounded difference property. By bounded difference inequality,

  $$\mathbb{P}\{\mathbb{E}[f_A(Z)] - f_A(Z) \geq t\} \leq \exp\left(-\frac{2t^2}{n}\right).$$

  Taking $t = \mathbb{E}[f_A(Z)] = \mathbb{E}[d_H(Z, A)]$, the left-hand side becomes $\mathbb{P}\{f_A(Z) \leq 0\} = \mathbb{P}\{d_H(Z, A) \leq 0\} = \mathbb{P}(A)$. Then the inequality becomes

  $$\mathbb{P}(A) \leq \exp\left(-\frac{2}{n}\left(\mathbb{E}[d_H(Z, A)]\right)^2\right)$$
  $$\Rightarrow \mathbb{E}[d_H(Z, A)] \leq \sqrt{\frac{n}{2} \log \frac{1}{\mathbb{P}(A)}}.$$

Then, by using the bounded difference inequality again, we obtain

$$\mathbb{P}\left\{d_H(Z,A) \geq \sqrt{\frac{n}{2}\log\frac{1}{\mathbb{P}(A)}} + t\right\} \leq \mathbb{P}\left\{d_H(Z,A) \geq \mathbb{E}\left[d_H(Z,A)\right] + t\right\} \leq \exp\left(-\frac{2t^2}{n}\right). \qquad \blacksquare$$

- **Remark** (***Concentration of Measure***)
To interpret the result in (9), we see that on the left-hand side we have the measure of the set of points whose Hamming distance is at least $t+\sqrt{\frac{n}{2}\log\frac{1}{\mathbb{P}(A)}}$ away from A. This inequality means that for A with ***small measure*** $\mathbb{P}(A)$, the measure of points whose ***Hamming distance*** from A is *more than* $O\left(\sqrt{n}\right)$ is ***extremely small***.

  In other words, ***product measure*** *on Hamming metric space are* ***concentrated*** *on* ***extremely small sets***. This phenomenon is called "***concentration of measure***".

- **Example** (***Bounded Difference Property*** $\Leftrightarrow$ ***Lipschitz Condition w.r.t. Hamming Distance***)
Note that any function with ***bounded difference property*** is ***Lipschitz function*** with respect to ***Hamming distance***.

$$\begin{aligned}
&\sup_{x\in\mathcal{X}^n, y_i\in\mathcal{X}} \left|f(x_1,\ldots,x_n) - f(x_1,\ldots,x_{i-1},y_i,x_{i+1},\ldots,x_n)\right| \\
&\leq c_i = c_i\, d_H((x_1,\ldots,x_n),(x_1,\ldots,x_{i-1},y_i,x_{i+1},\ldots,x_n)), \quad 1 \leq i \leq n \\
\Rightarrow |f(x)-f(y)| &= \left|\sum_{i=1}^{n}(f(x_{(i-1)}) - f(x_{(i)}))\right| \\
&\leq \sum_{i=1}^{n}\left|f(x_{(i-1)}) - f(x_{(i)})\right| \\
&\leq \sum_{i=1}^{n} c_i \mathbb{1}\left\{x_{(i-1)}[i] \neq x_{(i)}[i]\right\} \\
&= d_{H,c}(x,y)
\end{aligned}$$

  where $x_{(i)}$ is replicate of $x_{(i-1)}$ except for i-th component, which is replaced by y_i. Note that $x_{(0)} = x$ and $x_{(n)} = y$. Therefore, *the bounded difference inequality* can be seen as *an isoperimetry inequality* for *Lipschitz function with respect to Hamming distance.*

$$\mathbb{P}\left\{f(Z) - \mathbb{E}\left[f(Z)\right] \geq t\right\} \leq \exp\left(-\frac{2t^2}{n}\right)$$

2.2 Isoperimetric Inequalities on the 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

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

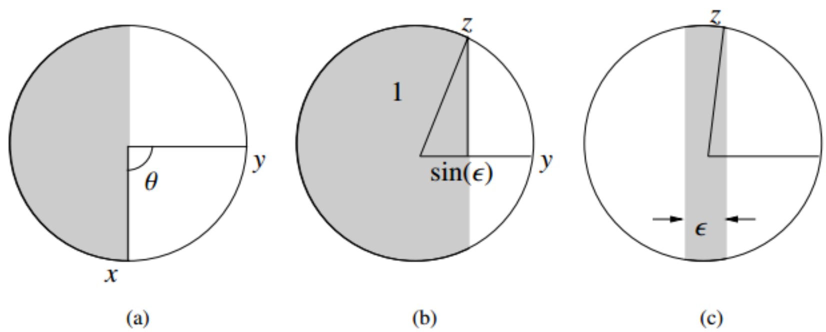

Figure 3.1 (a) Idealized illustration of the sphere $\mathbb{S}^{n-1}$. Any vector $y \in \mathbb{S}^{n-1}$ defines a hemisphere $H_y = \{x \in \mathbb{S}^{n-1} \mid \langle x, y \rangle \le 0\}$, corresponding to those vectors whose angle $\theta = \arccos \langle x, y \rangle$ with y is at least $\pi/2$ radians. (b) The ϵ-enlargement of the hemisphere H_y. (c) A central slice $T_y(\epsilon)$ of the sphere of width ϵ.

Figure 2: spherical cap and t-blowup. [Wainwright, 2019]

- **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 $t > 0$,*

$$\alpha_{\mathbb{S}^{n-1}}(t) \le \sqrt{\frac{\pi}{2}} \exp\left(-\frac{nt^2}{2} \right) \tag{10}$$

  Proof: Consider spherical cap

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

  and its *t-blowup*

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

  According to *the isoperimetry theorem on unit sphere*, the concentration function for uniform distribution over $\mathbb{S}^{n-1}$

$$\alpha_{\mathbb{S}^{n-1}}(t) = 1 - \mathbb{P}(H_y^t).$$

Note that $\mathbb{P}(H_y) \geq 1/2$, so H_y is a feasible set.

In order to bound the concentration function from above, we see that $\sin(t) \geq t/2$ for $t \in (0, \pi/2]$. Then the t-blowup H_y^t must contain set

$$\widetilde{H}_y^t := \left\{ x \in \mathbb{S}^{n-1} : \langle x\,,\, y \rangle < \frac{t}{2} \right\},$$

hence $\mathbb{P}(\widetilde{H}_y^t) \leq \mathbb{P}(H_y^t)$. By geometric calculation, for all $t \in (0, \sqrt{2})$, we have

$$\mathbb{P}(\widetilde{H}_y^t) \geq 1 - \left[1 - \left(\frac{t}{2} \right)^2 \right]^{n/2} \geq 1 - \exp\left(-\frac{n\,t^2}{8} \right)$$

where the last inequality is due to $(1 - x) \leq e^{-x}$. Thus

$$\alpha_{\mathbb{S}^{n-1}}(t) = 1 - \mathbb{P}(H_y^t) \leq \exp\left(-\frac{n\,t^2}{8} \right).$$

A similar but more careful approach to bounding $\mathbb{P}(H_y)$ can be used to establish the sharper upper bound

$$\alpha_{\mathbb{S}^{n-1}}(t) \leq \sqrt{\frac{\pi}{2}} \exp\left(-\frac{nt^2}{2} \right). \quad \blacksquare$$

- 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

  $$\mathbb{P}\left\{ |f(Z) - Med(f(Z))| \geq t \right\} \leq \sqrt{2\pi} \exp\left(-\frac{nt^2}{2} \right) \tag{11}$$

  Moreover, replacing median by the mean, we have

  $$\mathbb{P}\left\{ |f(Z) - \mathbb{E}\left[f(Z) \right]| \geq t \right\} \leq 2\sqrt{2\pi} \exp\left(-\frac{nt^2}{8} \right) \tag{12}$$

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

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

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

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

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

- **Remark** (***Zero-One Law for Independent Variables***) [Vershynin, 2018]
  The blow-up phenomenon we just saw may be quite *counter-intuitive* at first sight. How can an exponentially small set A undergo such a dramatic transition to an exponentially large set A_{2t} under such a small perturbation $2t$? (Remember that t can be much smaller than the radius $\sqrt{n}$ of the sphere.)

  However perplexing this may seem, this is *a typical phenomenon in* ***high dimensions***. It is *reminiscent* of ***zero-one laws*** in *probability theory*, which basically state that *events that are determined by many random variables* tend to have *probabilities either zero or one*.

2.3 Gaussian Isoperimetric Inequalities and Concentration of Gaussian Measure

2.4 Edge Isoperimetric Inequality on the Binary Hypercube

2.5 Vertex Isoperimetric Inequality on the Binary Hypercube

2.6 Convex Distance Inequality

References

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

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.