

# Lecture 5: Concentration of Measure and Isoperimetry

Tianpei Xie

Jan. 19th., 2023

## Contents

<b>1</b>	<b>The Classic Isoperimetry Inequalities</b>	<b>2</b>
1.1	Brunn-Minkowski Inequality . . . . .	2
1.2	The Blowup of Sets and Classical Isoperimetry Theorem . . . . .	5
<b>2</b>	<b>Concentration via Isoperimetry</b>	<b>6</b>
2.1	Levy's Inequalities . . . . .	6
2.2	Isoperimetric Inequalities on the Unit Sphere . . . . .	10
2.3	Gaussian Isoperimetric Inequalities and Concentration of Gaussian Measure . . . . .	13
2.4	Convex Distance Inequality . . . . .	17
2.5	Edge Isoperimetric Inequality on the Binary Hypercube . . . . .	19
2.6	Vertex Isoperimetric Inequality on the Binary Hypercube . . . . .	19

# 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  $\lambda \in [0, 1]$ ,

$$\text{Vol}(\lambda A + (1 - \lambda)B)^{\frac{1}{n}} \geq \lambda \text{Vol}(A)^{\frac{1}{n}} + (1 - \lambda) \text{Vol}(B)^{\frac{1}{n}}. \quad (1)$$

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  $c \geq 0$  then  $\text{Vol}(cA) = c \text{Vol}(A)$ . Thus it suffice to prove

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

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,

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

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  $f, g, h : \mathbb{R}^n \rightarrow [0, \infty)$  be **non-negative measurable functions** such that for all  $x, y \in \mathbb{R}^n$ ,

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

Then

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

**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

$$\begin{aligned}\int_{\mathbb{R}} f(x) dx &= \int_0^1 \text{Vol} \{x : f(x) \geq t\} dt \\ \int_{\mathbb{R}} g(x) dx &= \int_0^1 \text{Vol} \{x : g(x) \geq t\} dt.\end{aligned}$$

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

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

Thus

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

2. For the induction step, assume that the theorem holds for all dimensions  $1, \dots, n - 1$  and let  $f, g, h : \mathbb{R}^n \rightarrow [0, \infty)$ ,  $\lambda \in (0, 1)$  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

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

so by the inductive hypothesis

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

In other words, introducing

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

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  $\lambda \in [0, 1]$ ,

$$\text{Vol}(\lambda A + (1 - \lambda)B) \geq \text{Vol}(A)^\lambda \text{Vol}(B)^{1-\lambda}. \quad (3)$$

**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  $B$ ,

$$\text{Vol}(A + B)^{\frac{1}{n}} \geq \text{Vol}(A)^{\frac{1}{n}} + \text{Vol}(B)^{\frac{1}{n}}$$

since  $\text{Vol}(cA)^{1/n} = c \text{Vol}(A)^{1/n}$  for any  $c \in \mathbb{R}$  and  $A \subset \mathbb{R}^n$ . Also notice that we may assume that  $\text{Vol}(A), \text{Vol}(B) > 0$  because otherwise the inequality holds trivially. Defining  $A' = \text{Vol}(A)^{-\frac{1}{n}} A$  and  $B' = \text{Vol}(B)^{-\frac{1}{n}} B$ , we have  $\text{Vol}(A') = \text{Vol}(B') = 1$ . By *weaker Brunn-Minkowski inequality*, for  $\lambda \in (0, 1)$ ,

$$\text{Vol}(\lambda A' + (1 - \lambda)B') \geq 1.$$

Finally, we apply this *inequality* with the choice

$$\lambda = \frac{\text{Vol}(A)^{\frac{1}{n}}}{\text{Vol}(A)^{\frac{1}{n}} + \text{Vol}(B)^{\frac{1}{n}}}$$

obtaining

$$\begin{aligned} &\text{Vol} \left( \frac{\text{Vol}(A)^{\frac{1}{n}} A'}{\text{Vol}(A)^{\frac{1}{n}} + \text{Vol}(B)^{\frac{1}{n}}} + \frac{\text{Vol}(B)^{\frac{1}{n}} B'}{\text{Vol}(A)^{\frac{1}{n}} + \text{Vol}(B)^{\frac{1}{n}}} \right) \geq 1 \\ \Rightarrow &\text{Vol} \left( \frac{A}{\text{Vol}(A)^{\frac{1}{n}} + \text{Vol}(B)^{\frac{1}{n}}} + \frac{B}{\text{Vol}(A)^{\frac{1}{n}} + \text{Vol}(B)^{\frac{1}{n}}} \right) \geq 1 \\ \Rightarrow &\text{Vol} \left( \frac{A + B}{\text{Vol}(A)^{\frac{1}{n}} + \text{Vol}(B)^{\frac{1}{n}}} \right) \geq 1 \\ \Rightarrow &\frac{\text{Vol}(A + B)}{\left( \text{Vol}(A)^{\frac{1}{n}} + \text{Vol}(B)^{\frac{1}{n}} \right)^n} \geq 1 \end{aligned}$$

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  $\varepsilon$ -neighborhood  $A_\varepsilon$ .

**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 (or,  $t$ -enlargement) of  $A$  is defined by

$$A_t := \{x \in \mathbb{R}^n : d(x, A) < t\} = A + tB$$

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  $\text{Vol}(A)$  its *Lebesgue measure*. The surface area of  $A$  is defined by

$$\text{Vol}(\partial A) = \lim_{t \rightarrow 0} \frac{\text{Vol}(A_t) - \text{Vol}(A)}{t}.$$

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 theorem 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  $\text{Vol}(A) = \text{Vol}(B)$  where  $B := \{x \in \mathbb{R}^n : d(0, x) < 1\}$  is a unit ball. Then for any  $t > 0$ ,

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

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

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

**Proof:** By the Brunn-Minkowski inequality,

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

establishing the first statement. The second follows simply because

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

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

## 2 Concentration via Isoperimetry

### 2.1 Levy's Inequalities

- **Remark** We can generalize the classical isoperimetry 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  $\sigma$ -algebra 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*

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

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  $\sigma$ -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*  $\alpha : [0, \infty) \rightarrow \mathbb{R}_+$  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  $\alpha(t) = \mathbb{P}(A_t^c)$ .

- **Example (Concentration Function of Lebesgue Measure in  $\mathbb{R}^n$  and Isoperimetric Inequality)**

Note that the volume of a  $t$ -ball in  $\mathbb{R}^n$  is

$$\text{Vol}(tB) = \frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2} + 1)} t^n \equiv c_n t^n$$

Thus the radius of ball  $B$  with the same volume of  $A$  is

$$r := \left( \frac{\text{Vol}(A)}{c_n} \right)^{\frac{1}{n}}.$$

*The classical isoperimetric inequality* states that

$$\begin{aligned} \text{Vol}(A_t) &\geq \left( (r + t) \text{Vol}(B)^{1/n} \right)^n \\ \Leftrightarrow \text{Vol}(A_t) &\geq c_n \left( \left( \frac{\text{Vol}(A)}{c_n} \right)^{\frac{1}{n}} + t \right)^n \\ \Leftrightarrow \left( \frac{\text{Vol}(A_t)}{c_n} \right)^{\frac{1}{n}} &\geq \left( \frac{\text{Vol}(A)}{c_n} \right)^{\frac{1}{n}} + t \end{aligned} \tag{6}$$

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

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

- **Theorem 2.1 (Levy's Inequalities)** [Boucheron et al., 2013, Wainwright, 2019]

For any Lipschitz function  $f : \mathcal{X} \rightarrow \mathbb{R}$ ,

$$\begin{aligned} \mathbb{P}\{f(X) \geq \text{Med}(f(X)) + t\} &\leq \alpha_{\mathbb{P}}(t) \\ \mathbb{P}\{f(X) \leq \text{Med}(f(X)) - t\} &\leq \alpha_{\mathbb{P}}(t). \end{aligned} \tag{7}$$

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

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

**Proof:** Consider the set  $A = \{x : f(x) \leq \text{Med}(f(X))\}$ . 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  $y \in A$ ,  $f(y) \leq \text{Med}(f(X))$

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

Equivalently,

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

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

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

- **Theorem 2.2 (Converse of Levy's Inequalities)**[Boucheron et al., 2013, Wainwright, 2019]

If  $\beta : \mathbb{R}_+ \rightarrow [0, 1]$  is a function such that for **every Lipschitz function**  $f : \mathcal{X} \rightarrow \mathbb{R}$

$$\mathbb{P}\{f(X) - \text{Med}(f(X)) \geq t\} \leq \beta(t). \quad (8)$$

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  $\mathbb{P}(A) \geq 1/2$ , 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

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

- **Proposition 2.3 (Levy's Inequalities for Mean)**[Boucheron et al., 2013, Wainwright, 2019]

If  $\beta : \mathbb{R}_+ \rightarrow [0, 1]$  is a function such that for **every Lipschitz function**  $f : \mathcal{X} \rightarrow \mathbb{R}$

$$\mathbb{P}\{f(X) - \mathbb{E}[f(X)] \geq t\} \leq \beta(t). \quad (9)$$

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  $Z_1, \dots, Z_n$  taking their values in a (measurable) set  $\mathcal{X}$  and denote the vector of these variables by  $Z = (Z_1, \dots, Z_n)$  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  $t > 0$ ,

$$\mathbb{P} \left\{ d_H(x, A) \geq \sqrt{\frac{n}{2} \log \frac{1}{\mathbb{P}(A)}} + t \right\} \leq \exp \left( -\frac{2t^2}{n} \right) \quad (10)$$

**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

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

where  $\tilde{x}^{(i)} = (x_1, \dots, x_{i-1}, y_i, x_{i+1}, \dots, x_n)$ , 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

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

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

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

- **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**:

$$\mathbb{P}(A)\mathbb{P}\{d_H(x, A) \geq t\} \leq \min\{\mathbb{P}(A), \mathbb{P}\{d_H(x, A) \geq t\}\} \leq \exp\left(-\frac{t^2}{2n}\right) \quad (11)$$

- **Remark (Concentration of Measure)**

To interpret the result in (10), 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 less than  $O(\sqrt{n})$  is **extremely large**. 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**.

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

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**.

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

## 2.2 Isoperimetric Inequalities on the Unit Sphere

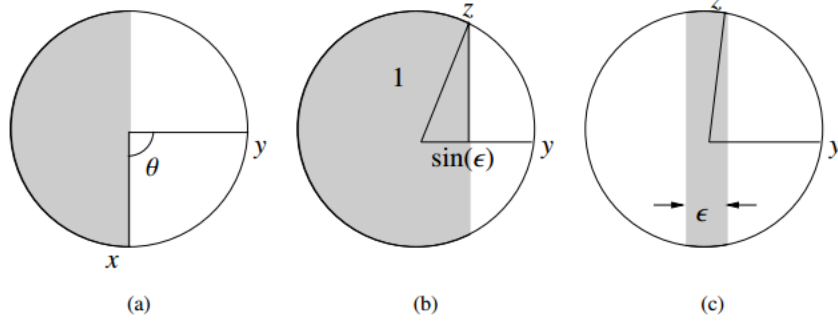
- **Definition (Spherical Cap and its  $t$ -Blowup)**

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

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

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

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



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

**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  $\sigma$  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

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

- **Remark** Define a *metric*  $\rho$  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  $t \in [0, 1]$ ,

$$\alpha_{\mathbb{S}^{n-1}}(t) \leq c \exp\left(-\frac{nt^2}{2}\right) \quad (12)$$

for some constant  $c$ .

**Proof:** Consider spherical cap

$$C(y, 0) := \{x \in \mathbb{S}^{n-1} : \langle x, y \rangle \geq 0\}$$

and its  $t$ -blowup

$$C(y, t) := \{x \in \mathbb{S}^{n-1} : \langle x, y \rangle \geq t\}.$$

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

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

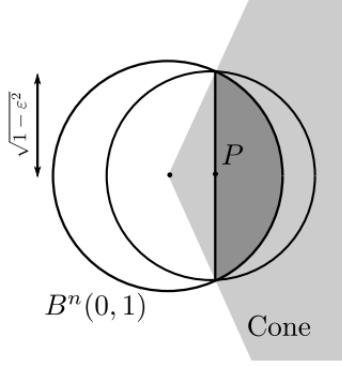


Figure 2: Small  $\varepsilon$ .

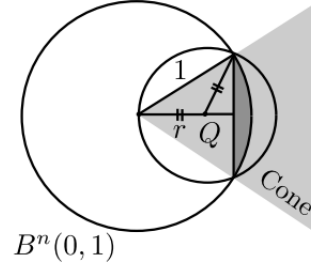


Figure 3: Large  $\varepsilon$ . By the congruence  $\frac{1/2}{r} = \frac{\varepsilon}{1}$ .

Figure 3: proof for upper bound of area of spherical cap (left) for small  $t$  (right) for large  $t$

Note that  $\mathbb{P}(C(y, 0)) \leq 1/2$ . In order to bound the concentration function from above, consider for small  $t \in [0, 1/\sqrt{2}]$ ,

$$\begin{aligned} \alpha_{\mathbb{S}^{n-1}}(t) = \mathbb{P}(C(y, t)) &= \frac{\text{Vol}(B^n(0, 1) \cap \text{Cone})}{\text{Vol}(B^n(0, 1))} \\ &\leq \frac{\text{Vol}(B^n(P, \sqrt{1-t^2}))}{\text{Vol}(B^n(0, 1))} \\ &= (\sqrt{1-t^2})^n \\ &\leq \exp\left(-\frac{nt^2}{2}\right) \end{aligned}$$

For  $t \in [1/\sqrt{2}, 1)$ , it is enough to consider a different auxiliary ball which includes the set  $\text{Cone} \cap B^n(0, 1)$ . We obtain

$$\begin{aligned} \alpha_{\mathbb{S}^{n-1}}(t) = \mathbb{P}(C(y, t)) &\leq \frac{\text{Vol}(B^n(Q, r))}{\text{Vol}(B^n(0, 1))} \\ &= r^n = \left(\frac{1}{2t}\right)^n \\ &\leq \exp\left(-\frac{nt^2}{2}\right) \end{aligned}$$

where the last inequality is from  $e^{x^2/2} \leq 2x$  for  $x \in [1/\sqrt{2}, 1]$ . Due to convexity, this is only to be checked at the boundary of our interval  $[1/\sqrt{2}, 1]$ ,  $\blacksquare$

- 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

$$\mathbb{P}\{|f(Z) - \text{Med}(f(Z))| \geq t\} \leq \sqrt{2\pi} \exp\left(-\frac{nt^2}{2}\right) \quad (13)$$

Moreover, replacing median by the mean, we have

$$\mathbb{P}\{|f(Z) - \mathbb{E}[f(Z)]| \geq t\} \leq 2\sqrt{2\pi} \exp\left(-\frac{nt^2}{8}\right) \quad (14)$$

- **Exercise 2.8 (The Blow-Up Phenomenon)**

Let  $A$  be a subset of the sphere  $\sqrt{n}\mathbb{S}^{n-1}$  such that

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

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

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

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

- **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  $\varphi(x) := \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} = (\Phi(x))'$  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

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

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 2.9** (*Basic Property of the Gaussian Isoperimetric Function*) [Boucheron et al., 2013]

The Gaussian isoperimetric function  $\gamma$  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  $(0, 1)$ .

**Proof:** 1. See that

$$\begin{aligned}
\varphi'(x) &= \frac{1}{\sqrt{2\pi}}(-x)e^{-\frac{x^2}{2}} = (-x)\varphi(x) \\
\varphi''(x) &= (x^2 - 1)\varphi(x)
\end{aligned}$$

Thus

$$\gamma(x)' = (\varphi(\Phi^{-1}(x)))' = \frac{d\varphi}{dy}(\Phi^{-1}(x)) \frac{d\Phi^{-1}}{dx}(x) = (-\Phi^{-1}(x)) (\Phi^{-1}(x))' \gamma(x) = -\Phi^{-1}(x),$$

since  $(\Phi^{-1}(x))' \gamma(x) = 1$ , we have the result.

2.

$$\begin{aligned}
\gamma''(x) &= (\gamma'(x))' = -(\Phi^{-1}(x))' = -\frac{1}{\gamma(x)} \\
\gamma(x)\gamma''(x) &= -1
\end{aligned}$$

3. Since  $\gamma > 0$  for  $x \in (0, 1)$ ,

$$\gamma''(x) = -\frac{1}{\gamma(x)} < 0.$$

Thus  $\gamma(x)$  is concave function in  $(0, 1)$ . ■

- **Lemma 2.10** (*Asymptotic Behavior of Gaussian Isoperimetric Function*) [Boucheron et al., 2013]

For all  $x \in [0, 1/2]$ ,

$$x\sqrt{\frac{1}{2}\log\frac{1}{x}} \leq \gamma(x) \leq x\sqrt{2\log\frac{1}{x}}.$$

Moreover,

$$\lim_{x \rightarrow 0} \frac{\gamma(x)}{x \sqrt{2 \log \frac{1}{x}}} = 1$$

- **Proposition 2.11 (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  $f : \{-1, 1\}^n \rightarrow [0, 1]$ ,

$$\gamma(\mathbb{E}[f(Z)]) \leq \mathbb{E} \left[ \sqrt{\gamma(f(Z))^2 + \|\nabla f(Z)\|_2^2} \right] \quad (15)$$

- **Proposition 2.12 (Bobkov's Gaussian Inequality)** [Boucheron et al., 2013]

Let  $Z := (Z_1, \dots, Z_n)$  be a vector of **independent standard Gaussian** random variables. Let  $f : \mathbb{R}^n \rightarrow [0, 1]$  be a differentiable function with gradient  $\nabla f$ . Then

$$\gamma(\mathbb{E}[f(X)]) \leq \mathbb{E} \left[ \sqrt{\gamma(f(X))^2 + \|\nabla f(X)\|_2^2} \right] \quad (16)$$

where  $\gamma = \varphi \circ \Phi^{-1}$  is the **Gaussian isoperimetric function**.

- **Theorem 2.13 (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

$$\liminf_{t \rightarrow 0} \frac{\mathbb{P}(A_t) - \mathbb{P}(A)}{t} \geq \gamma(\mathbb{P}(A)), \quad (17)$$

where  $A_t := \{x : d(x, A) < t\}$  be the  $t$ -blowup of  $A$ . Moreover, if  $A$  is a **half-space** defined by  $A := \{x \in \mathbb{R}^n : x_1 \leq z\}$ , then

$$\liminf_{t \rightarrow 0} \frac{\mathbb{P}(A_t) - \mathbb{P}(A)}{t} = \gamma(\mathbb{P}(A)) = \varphi(z), \quad (18)$$

where  $\gamma := \varphi \circ \Phi^{-1}$  is the **Gaussian isoperimetric function**.

- **Proposition 2.14 (Differentiability 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 2.15 (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  $t \geq 0$ ,

$$\begin{aligned} \mathbb{P}(A_t) &\geq \Phi(\Phi^{-1}(\mathbb{P}(A)) + t) . \\ \Leftrightarrow \Phi^{-1}(\mathbb{P}(A_t)) &\geq \Phi^{-1}(\mathbb{P}(A)) + t \end{aligned} \quad (19)$$

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

**Proof:** We call a Borel set  $A \subset \mathbb{R}^n$  **smooth** if  $\mathbb{P}(A_t)$  is a differentiable function of  $t$  on  $(0, \infty)$ .

1. Observe that if  $A$  is *smooth*, then

$$\begin{aligned} \frac{d}{dt} \Phi^{-1}(\mathbb{P}(A_t)) &= [(\Phi^{-1})'(\mathbb{P}(A_t))] \frac{d}{dt} \mathbb{P}(A_t) \\ &= \frac{1}{\gamma(\mathbb{P}(A_t))} \frac{d}{dt} \mathbb{P}(A_t) \\ &\geq \frac{1}{\frac{d}{dt} \mathbb{P}(A_t)} \left( \frac{d}{dt} \mathbb{P}(A_t) \right) = 1 \end{aligned}$$

The last inequality is due to *the Gaussian isoperimetric inequality*

$$\frac{d}{dt} \mathbb{P}(A_t) \geq \liminf_{s \rightarrow 0} \frac{\mathbb{P}(A_{t+s}) - \mathbb{P}(A_t)}{s} \geq \gamma(\mathbb{P}(A_t)).$$

Therefore, by integration

$$\begin{aligned} \Phi^{-1}(\mathbb{P}(A_t)) &= \Phi^{-1}(\mathbb{P}(A)) + \int_0^t \frac{d}{ds} \Phi^{-1}(\mathbb{P}(A_s)) ds \\ &\geq \Phi^{-1}(\mathbb{P}(A)) + \int_0^t ds = \Phi^{-1}(\mathbb{P}(A)) + t. \end{aligned}$$

Hence, the theorem holds for all smooth sets. The remaining work is to extend this to all *Borel sets*.

2. Note first that if  $\mathbb{P}(A) = 0$ , the theorem is automatically satisfied and therefore we may focus on Borel sets  $A$  with *positive probability*. By Proposition 2.14, the concentration property holds for **any finite union of open balls**.
3. Now let  $A$  be *any Borel set* with  $\mathbb{P}(A) > 0$ . Let  $0 < \epsilon < t$ . Then by **Vitali's covering theorem**, there exists a *countable* collection of *disjoint open balls*  $\{B_1, B_2, \dots\}$ , all intersecting  $A$  and *diameter at most*  $\epsilon$ , such that  $\mathbb{P}(A - \bigcup_{n=1}^{\infty} B_n) = 0$ . But then

$$\begin{aligned} \mathbb{P}(A_t) &\geq \mathbb{P}\left(\bigcup_{n=1}^{\infty} (B_n)_{t-\epsilon}\right) \\ &= \lim_{n \rightarrow \infty} \mathbb{P}\left(\bigcup_{i=1}^n (B_i)_{t-\epsilon}\right) \\ &\geq \lim_{n \rightarrow \infty} \Phi\left(\Phi^{-1}\left(\mathbb{P}\left(\bigcup_{i=1}^n (B_i)_{t-\epsilon}\right)\right) + t - \epsilon\right) \\ &= \Phi\left(\Phi^{-1}\left(\mathbb{P}\left(\bigcup_{i=1}^{\infty} (B_i)_{t-\epsilon}\right)\right) + t - \epsilon\right) \\ &\geq \Phi\left(\Phi^{-1}(\mathbb{P}(A)) + t - \epsilon\right) \end{aligned}$$

The argument is completed by taking  $\epsilon$  to 0. ■



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

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

- **Exercise 2.16 (From Isoperimetry to Concentration)** [Boucheron et al., 2013]  
Assume that a probability distribution  $\mathbb{P}$  on  $\mathbb{R}^n$  satisfies, for all Borel sets  $A \subset \mathbb{R}^n$ ,

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

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

$$\mathbb{P}(A_t) \geq F(F^{-1}(\mathbb{P}(A)) + ct).$$

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

**Theorem 2.17 (Gaussian Concentration Inequality, Sharp Bound)** [Boucheron et al., 2013, Wainwright, 2019]  
Let  $Z = (Z_1, \dots, Z_n)$  be a vector of  $n$  independent standard normal random variables. Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  denote an  $L$ -Lipschitz function. Then, for all  $t > 0$ ,

$$\mathbb{P}\{f(Z) - \text{Med}(f(Z)) \geq t\} \leq 1 - \Phi\left(\frac{t}{L}\right). \quad (20)$$

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

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

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

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  $t > 0$ .

## 2.4 Convex Distance Inequality

- **Definition (Weighted Hamming Distance)**

Given  $\alpha = (\alpha_1, \dots, \alpha_n)$ , where  $\alpha_i \geq 0$ , the weighed Hamming distance between  $x, y \in \mathcal{X}^n$  is defined as

$$d_\alpha(x, y) = \sum_{i=1}^n \alpha_i \mathbb{1}\{x_i \neq y_i\}.$$

- **Remark (*Measure Concentration in Weighted Hamming Distance Space*)**

Similar to the inequality (10), for *metric measure space*  $\mathcal{X}^n$  with respect to *weighted Hamming distance*, we have the measure concentration inequality for  $A \subset \mathcal{X}^n$

$$\mathbb{P} \left\{ d_\alpha(x, A) \geq \sqrt{\frac{\|\alpha\|_2}{2} \log \frac{1}{\mathbb{P}(A)}} + t \right\} \leq \exp \left( -\frac{2t^2}{\|\alpha\|_2} \right)$$

where  $\|\alpha\|_2 = \sqrt{\sum_{i=1}^n \alpha_i^2}$ . Assume  $\|\alpha\|_2 = 1$

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

Following the same argument, we can find *an equivalent form* as in (11)

$$\sup_{\alpha \in \mathbb{R}_+^n : \|\alpha\|_2=1} \mathbb{P}(A) \mathbb{P} \{d_\alpha(x, A) \geq t\} \leq \sup_{\alpha \in \mathbb{R}_+^n : \|\alpha\|_2=1} \min \{ \mathbb{P}(A), \mathbb{P} \{d_\alpha(x, A) \geq t\} \} \leq \exp \left( -\frac{t^2}{2} \right)$$

A key contribution for ***convex distance inequality*** is that the above inequality remains true even if the ***supremum*** is taken ***within the probability***; i.e.

$$\mathbb{P}(A) \mathbb{P} \left\{ \sup_{\alpha \in \mathbb{R}_+^n : \|\alpha\|_2=1} d_\alpha(x, A) \geq t \right\} \leq \exp \left( -\frac{t^2}{4} \right).$$

- **Definition (*Convex Distance*)**

For any  $x = (x_1, \dots, x_n) \in \mathcal{X}^n$ , ***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 2.18 (*Convex Distance Inequality*)** [Boucheron et al., 2013]

For any subset  $A \subset \mathcal{X}^n$  and  $t > 0$ ,

$$\mathbb{P}(A) \mathbb{P} \{d_T(X, A) \geq t\} = \mathbb{P}(A) \mathbb{P} \left\{ \sup_{\alpha \in \mathbb{R}_+^n : \|\alpha\|_2=1} d_\alpha(X, A) \geq t \right\} \leq \exp \left( -\frac{t^2}{4} \right). \quad (21)$$

- 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 2.19 (*Convex Distance vs. Distance to Convex Set*)** [Boucheron et al., 2013]

Let  $A \subset [0, 1]^n$  be a ***convex set*** and let  $x = (x_1, \dots, x_n) \in [0, 1]^n$ . Then

$$d(x, A) := \inf_{y \in A} \|x - y\|_2 \leq d_T(x, A). \quad (22)$$

- **Theorem 2.20 (*Concentration of Quasi-Convex Lipschitz Functions*)** [Boucheron et al., 2013]

Let  $Z := (Z_1, \dots, Z_n)$  be independent random variables taking values in the interval  $[0, 1]$  and let  $f : [0, 1]^n \rightarrow \mathbb{R}$  be a **quasi-convex function**; that is

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

Moreover,  $f$  is Lipschitz function satisfying

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

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

$$\begin{aligned} \mathbb{P}\{f(Z) \geq \text{Med}(f(Z)) + t\} &\leq 2 \exp\left(-\frac{t^2}{4}\right), \\ \mathbb{P}\{f(Z) \leq \text{Med}(f(Z)) - t\} &\leq 2 \exp\left(-\frac{t^2}{4}\right). \end{aligned} \tag{23}$$

**Proof:** For some  $s \in \mathbb{R}$ , define the set  $A_s = \{z : f(z) \leq s\} \subset [0, 1]^n$ . Because of *quasi-convexity*,  $A_s$  is *convex*. By the Lipschitz property and Lemma 2.19, for all  $z \in [0, 1]^n$ ,

$$f(z) \leq s + d(z, A) \leq s + d_T(z, A).$$

So by *convex distance inequality*,

$$\mathbb{P}\{f(Z) \leq s\} \mathbb{P}\{f(Z) \geq s + t\} \leq \exp\left(-\frac{t^2}{4}\right)$$

Take  $s = \text{Med}(f(Z))$  to get the *upper tail inequality* and  $s = \text{Med}(f(Z)) - t$  to get the *lower tail inequality*. ■

## 2.5 Edge Isoperimetric Inequality on the Binary Hypercube

## 2.6 Vertex Isoperimetric Inequality on the Binary Hypercube

- **Definition (Vertex Boundary of Graph)** [Boucheron et al., 2013]  
Consider a graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  and let  $A \subset \mathcal{V}$  be a set of its vertices. The vertex boundary of  $A$  is defined as the set of those vertices, **not in**  $A$ , which are **connected** to some vertex in  $\mathcal{V}$  by an edge. We denote the vertex boundary of  $A$  by  $\partial V(A)$ .
- **Definition (Vertex Isoperimetric Problem of Graph)** [Boucheron et al., 2013]  
The vertex isoperimetric problem in a graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  is to determine the sets  $\mathcal{A} \subset \mathcal{V}$  of a given cardinality whose vertex boundary contains a **minimal number of vertices**.
- **Remark (Binary Hypercube as Nearest Neighbor Graph with Hamming Distance)**  
Consider the graph as binary hypercube  $\{-1, +1\}^n$  in which two vertices are connected by an edge if and only if their **Hamming distance equals 1**. Define the norm as the Hamming distance to  $-1^n = (-1, \dots, -1)$

$$\|x\|_H := \sum_{i=1}^n \mathbb{1}\{x_i = 1\} = d_H(x, -1^n)$$

- **Definition (*Simplicial Order*)**

We define the so-called ***simplicial order*** of the elements of the binary hypercube. We say that  $x = (x_1, \dots, x_n) \in \{-1, 1\}^n$  ***precedes***  $y = (y_1, \dots, y_n) \in \{-1, 1\}^n$  in the ***simplicial order*** if either  $\|x\|_H < \|y\|_H$  (where  $\|x\|_H := \sum_{i=1}^n \mathbb{1}\{x_i = 1\} = d_H(x, -1^n)$ ) or  $\|x\|_H = \|y\|_H$  and  $x_i = 1$  and  $y_i = -1$  for the smallest  $i$  for which  $x_i \neq y_i$ . That is

$$x \prec y \\ \Leftrightarrow \{(x, y) : \|x\|_H < \|y\|_H \vee (\|x\|_H = \|y\|_H \wedge (x_i = 1 \wedge y_i = -1, \text{ where } i = \min \{k : x_k \neq y_k\}))\}$$

In other words, the vector with ***less*** 1's ***precedes*** the vector with more 1's. If the number of 1's are the same, then the first 1's on the leftmost position is preferred.

- **Example** For  $n = 3$ ,  $(-1, -1, -1) \prec (1, -1, -1) \prec (-1, 1, -1) \prec (-1, -1, 1) \prec (1, 1, -1) \prec (1, -1, 1) \prec (-1, 1, 1) \prec (1, 1, 1)$

**Theorem 2.21 (*Harp's Vertex Isoperimetric Theorem*)** [Boucheron et al., 2013]  
For  $N = 1, \dots, 2^n$ , let  $S_N$  denote the set of ***first***  $N$  ***elements*** of  $\{-1, +1\}^n$  in the ***simplicial order***. For any subset  $A \subset \{-1, +1\}^n$ , where  $|A| = N$ ,

$$|\partial V(A)| \geq |\partial V(S_N)|$$

- **Remark** Note that if  $N = \sum_{i=0}^k \binom{n}{i}$ , for  $k = 0, \dots, n$ , then

$$S_N = \{x \in \{-1, +1\}^n : d_H(x, -1^n) \leq k\} = B_H(-1^n, k)$$

In other words,  $S_N$  is a ***Hamming ball*** centered at the vector  $-1^n = (-1, \dots, -1)$ .

- **Definition (*t-Blowup of Set A in Binary Hypercube*)**

For any  $A \subset \{-1, +1\}^n$  and  $x \in \{-1, +1\}^n$ , let  $d_H(x, A) = \min_{y \in A} d_H(x, y)$  be the ***Hamming distance*** of  $x$  to the set  $A$ . Also, denote by

$$A_t := \{x \in \{-1, +1\}^n : d_H(x, A) < t\}$$

the ***t-blowup*** of the set  $A$ , that is, the set of points whose Hamming distance from  $A$  is at most  $t$ .

- **Remark (*Harp's Vertex Isoperimetric Theorem*  $\Leftrightarrow$  *Classical Isoperimetric Theorem in*  $\{-1, 1\}^n$ )**

The fact that among all sets with a ***given volume***, ***balls minimize the surface area*** is in close analogy with the ***classical isoperimetric theorem***.

Observe that if  $S_N$  is a ***Hamming ball*** with radius  $k$ , i.e.  $N = \sum_{i=0}^k \binom{n}{i}$ , then  $S_N \cup \partial V(S_N)$  is the ***Hamming ball*** of radius  $k + 1$ . This implies that for any set  $A \subset \{-1, +1\}^n$  with  $|A| \geq \sum_{i=0}^k \binom{n}{i}$ , we have

$$|A \cup \partial V(A)| \geq \sum_{i=0}^{k+1} \binom{n}{i}.$$

By iterating this argument, we obtain the following simple ***consequence of Harper's theorem***.

**Corollary 2.22** (*Isoperimetric Inequality in Binary Hypercube*) [Boucheron et al., 2013]

Let  $A \subset \{-1, +1\}^n$  such that  $|A| \geq \sum_{i=0}^k \binom{n}{i}$ . Then for any  $t = 1, 2, \dots, n - k + 1$ ,

$$|A_t| \geq \sum_{i=0}^{k+1-t} \binom{n}{i}. \quad (24)$$

In particular, if  $|A|/2^n \geq 1/2$  then we may take  $k = \lfloor n/2 \rfloor$  in the corollary above and

$$\frac{|A_t|}{2^n} \geq \mathbb{P}\{X < \mathbb{E}[X] + t\} \geq 1 - \exp\left(-\frac{2t^2}{n}\right) \quad (25)$$

where  $X \sim \text{Ber}(1/2)$  is a **symmetric Bernoulli** random variable taking values in  $\{-1, +1\}$  with  $\mathbb{P}\{X = 1\} = \mathbb{P}\{X = -1\} = 1/2$ . The last inequality comes from Hoeffding's inequality.

- **Remark** (*Concentration of Uniform Measure on Binary Hypercube*)

Consider any set  $A$  containing at least half of the points of  $\{-1, +1\}^n$ . According to the corollary above, the fraction of those points which **cannot be obtained** by changing **at most**  $c\sqrt{n}$  bits of some point in  $A$  is **at most**  $e^{2c^2}$ . In other words, **an immense majority** of the points in  $\{-1, +1\}^n$  is **within Hamming distance of the order of  $\sqrt{n}$  of  $A$**

- **Definition** (*Monotone Set*)

A set  $A \subset \{-1, +1\}^n$  is **monotone** if  $\mathbb{1}\{x \in A\} \geq \mathbb{1}\{y \in A\}$  for all  $x = (x_1, \dots, x_n)$  and  $y = (y_1, \dots, y_n)$  in  $\{-1, +1\}^n$  such that  $x_i \geq y_i$  **for all**  $i$ .

- **Theorem 2.23** (*Isoperimetric Theorem for Bernoulli Distribution*) [Boucheron et al., 2013]

Let  $k \in \{0, \dots, n\}$  and let  $S = \{x \in \{-1, +1\}^n : \|x\| < t\}$  be a **Hamming ball of radius  $k$** .  $\mathbb{P}$  is Bernoulli distribution on  $\{-1, 1\}^n$  with parameter  $p$  such that

$$\mathbb{P}(x) := p^{\|x\|} (1-p)^{1-\|x\|}.$$

If  $A \subset \{-1, +1\}^n$  is a **monotone set** such that  $\mathbb{P}(A) \geq \mathbb{P}(S)$  then

$$\mathbb{P}(\partial V(A)) \geq \mathbb{P}(\partial V(S)).$$

## 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.