Minkowski's Theorem and Its Applications

Zichao Dong

This is a mini talk about Minkowski's theorem in convex geometry and several interesting applications of it. It is based on [1]. We first introduce two proofs of the main theorem.

Theorem. (Minkowski's theorem)

Suppose C to be a symmetric, convex, bounded subset $\mathbb{R}^d$. If $\operatorname{vol}(C) > 2^d$, then C contains at least one lattice point other than 0.

Proof. Take $C' = \frac{1}{2}C = \left\{\frac{1}{2}x : x \in C\right\}$, then $\operatorname{vol}(C') > 1$. Then we claim that

There exists a nonzero integral vector $v \in \mathbb{Z}^d \backslash \{0\}$, such that $v \in C' - C'$.

Let's first show that the claim implies the our theorem:

Suppose $x \in C'$ and $x + v \in C'$, then $-x \in C'$ by symmetry, and hence $\frac{1}{2}v = \frac{1}{2}(-x) + \frac{1}{2}(v+x) \in C'$ by convexity. Thus $v \in C$ follows from $C' = \frac{1}{2}C$, we're done.

Next, we introduce two proofs of the claim:

1) **(Minkowski)**

We show by contradiction. Suppose $(C' - C') \cap \mathbb{Z}^d = \{0\}$, then $C' + u$ and $C' + v$ are always disjoint for different $u, v \in \mathbb{Z}^d$. Suppose D is the diameter of C', then we have

$$\bigcup_{[-N,N]^d} (C' + u) \subset [-(N+D), N+D].$$

Take the volume for both sides, we see that

$$(2N+1)^d \operatorname{vol}(C') \leqslant 2^d (N+D)^d,$$

and hence

$$1 < \operatorname{vol}(C') \leqslant \left(\frac{2(N+D)}{2N+1}\right)^d,$$

Take $N \to \infty$, we see the contradiction. □

2) **(Blichfeldt)**

We prove some general result:

Suppose $A \subset \mathbb{R}^n$ be measurable with $\operatorname{vol}(A) > k$, then we can find some $x \in \mathbb{R}^n$, such that $A + x$ contains at least $k+1$ points from $\mathbb{Z}^n$.

Denote $f(x) = \sum_{y \in \mathbb{Z}^n} 1_{A+x}(y)$ which is definitely measurable, and we may assume without loss of generality that A is bounded, since the Lebesgue measure is inner regular. Then

$$\begin{aligned}
\int_{[0,1]^n} f(x)dx &= \sum_{y\in\mathbb{Z}^n} \int_{[0,1]^n} 1_{A+x}(y)dx \\
&= \sum_{y\in\mathbb{Z}^n} \int_{[0,1]^n} 1_A(y-x)dx \\
&= \sum_{y\in\mathbb{Z}^n} \int_{y-[0,1]^n} 1_A(t)dt \\
&= \int_{\mathbb{R}^n} 1_A(t)dt \\
&= \mathrm{vol}(A) \\
&> k.
\end{aligned}$$

Thus, $\max_{x\in[0,1]^n} f(x) \geqslant k+1$, we're done. □

Now we introduce a quick application to take a glance on the power of Minkowski's theorem.

Theorem. (Dirichlet's approximation)

For any $\alpha \in (0,1)$ and $N \in \mathbb{N}$, there exists a pair of natural numbers m, n with $m \leqslant N$, such that

$$\left|\alpha - \frac{n}{m}\right| < \frac{1}{mN}.$$

Proof. Consider the parallelogram $P = \left\{(x,y) : |x| \leqslant N + \frac{1}{2}, |y - \alpha x| \leqslant \frac{1}{N}\right\} \subset \mathbb{R}^2$. We have P is bounded and symmetric convex with $|P| = \frac{2(2N+1)}{N} > 2^2$, hence by Minkowski's theorem, there exists some $(m,n) \in P \cap \mathbb{Z}^2$. Then $|\alpha m - n| < \frac{1}{N}$, hence $\left|\alpha - \frac{n}{m}\right| < \frac{1}{mN}$ with $m \leqslant N$ (by symmetry of P, we can choose $m > 0$), we're done. □

Before showing some deeper applications of Minkowski's theorem, we need to generalize the concept of lattice at first.

Definition. (General lattice)

Let $z_1, z_2, \cdots, z_d$ be d linearly independent vectors in $\mathbb{R}^d$, we define

$$\Lambda = \Lambda(z_1, z_2, \cdots, z_d) := \{i_1 z_1 + i_2 z_2 + \cdots + i_d z_d : (i_1, i_2, \cdots, i_d) \in \mathbb{Z}^d\}$$

as the lattice with basis $\{z_1, z_2, \cdots, z_d\}$. In particular, we denote $\det \Lambda = |\det Z|$ where Z is the matrix $(z_1, z_2, \cdots, z_d)$ made up of column vectors.

We have the following remarks to state here.

Remark.

- The basis is not unique generally.
- We have $\det \Lambda = \mathrm{vol}(P)$ for $P = \{\alpha_1 z_1 + \alpha_2 z_2 + \cdots + \alpha_d z_d : \alpha_1, \alpha_2, \cdots, \alpha_d \in [0,1]\}$.
- In fact, $\det \Lambda$ does not depend on the choice of the basis.

Now we can introduce the following generalized version of Minkowski's theorem.

Theorem. (Minkowski's theorem for general lattices)

Suppose Λ to be a lattice and C to be a bounded symmetric convex subset in $\mathbb{R}^d$. If $\mathrm{vol}(C) > 2^d \det \Lambda$, then C contains at least a point of Λ different from 0.

There is actually nothing new to prove since we can apply some affine mapping to reduct this general case into the original theorem. Now we may apply the generalized Minkowski's theorem to solve some classical number theory problems.

Theorem. (Two-square theorem)

For prime $p \equiv 1 \pmod 4$, we can always find some $a, b \in \mathbb{Z}$, such that $p = a^2 + b^2$.

Proof. We first claim that $\left(\dfrac{-1}{p}\right) = 1$ for $p \equiv 1 \pmod 4$. If $\left(\dfrac{-1}{p}\right) = -1$, then we have $(p-1)! \equiv (-1)^{\frac{p-1}{2}} \equiv 1 \pmod p$ by pairing each element with their inverse respectively in $\mathbb{Z}/p\mathbb{Z}$, which contradicts Wilson's theorem. So the claim is true.

Now suppose $q^2 \equiv -1 \pmod p$, and take $z_1 = (1, q)$, $z_2 = (0, p)$, and $\Lambda = \Lambda(z_1, z_2)$, then we have $\det \Lambda = p$.

Consider $C = \{(x, y) : x^2 + y^2 < 2p\}$ in $\mathbb{R}^2$, then we have

$$\mathrm{vol}(C) = 2\pi p > 4p = 2^2 \det \Lambda.$$

According to Minkowski's theorem, we see that C contains a point $(a, b) \in \Lambda \backslash \{0\}$. Then $(a, b) = iz_1 + jz_2 = (i, iq + jp) \in \mathbb{R}^2$, which implies

$$a^2 + b^2 = i^2 + (iq + jp)^2 \equiv (q^2 + 1)i^2 \equiv 0 \pmod p.$$

Note that we also have $0 < a^2 + b^2 < 2p$, hence it has to be $a^2 + b^2 = p$, we're done. □

Theorem. (Lagrange's four-square theorem)

For $n \in \mathbb{N}$, we can always find some $a, b, c, d \in \mathbb{Z}$, such that $n = a^2 + b^2 + c^2 + d^2$.

Proof. Suppose the prime factorization of n is $n = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_k^{\alpha_k}$. It is easily seen that to prove the theorem, we only need to deal with the case when $\alpha_1, \alpha_2, \cdots, \alpha_k \in \{0, 1\}$.

We first claim that there exists some $x, y \in \mathbb{Z}$ such that $x^2 + y^2 \equiv -1 \pmod n$. We show for prime $n = p$ ($p = 2$ is trivial hence suppose p to be odd): if $\left(\dfrac{-1}{p}\right) = 1$, then we have $a^2 \equiv -1 \pmod p$ for some a, and by taking $x = a, y = 0$, we're done; if $\left(\dfrac{-1}{p}\right) = -1$, then consider the residue pairs $(0, p-1), (1, p-2), \cdots, (\frac{p-1}{2}, \frac{p-1}{2})$ and since both 0 and $p-1$ are not quadratic residues, there must be a pair consisting of quadratic residues by pigeonhole principle, hence we're done. The general case follows by applying Chinese Remainder theorem to all of the prime factors of n.

Now suppose $A^2 + B^2 \equiv -1 \pmod n$, and take $\Lambda = \{(x, y, z, t) : z \equiv Ax + By \pmod n, t \equiv Bx - Ay \pmod n\}$ in $\mathbb{R}^4$, then we have $\det \Lambda = n^2$.

Consider $C = \{(x, y, z, t) : x^2 + y^2 + z^2 + t^2 < 2n\}$ in $\mathbb{R}^4$, then we have

$$\operatorname{vol}(C) = \frac{\sqrt{\pi}^4}{\Gamma(1 + \frac{4}{2})}(\sqrt{2n})^4 = 2n^2\pi^2 > 2^4 n^2 = 2^4 \det \Lambda.$$

According to Minkowski's theorem, we see that C contains a point $(a, b, c, d) \in \Lambda \backslash \{0\}$. It is easy to check that $a^2 + b^2 + c^2 + d^2 \equiv 0 \pmod{n}$ while $0 < a^2 + b^2 + c^2 + d^2 < 2n$, hence we're done. □

References

[1] Matoušek, J., Lectures on discrete geometry. Graduate Texts in Mathematics, 212. *Springer-Verlag, New York*, 2002.

Steiner symmetrisation and Blaschke–Santaló inequality

Yihan Zhang

This presentation is based on [1].

Definition 1. *Let $K \subset \mathbb{R}^n$ be a convex body and $u \in S^{n-1}$. The Steiner symmetral $S_u(K)$ of K along u is defined as the body such that for all $x \in u^{\perp}$,*

$$|K \cap (x + \mathbb{R}u)| = |S_u(K) \cap (x + \mathbb{R}u)|.$$

In other words,

$$S_u(K) \coloneqq \left\{ (x, tu) : x \in \mathrm{proj}_{u^\perp}(K),\ |t| \leqslant \frac{1}{2} |K \cap (x + \mathbb{R}u)| \right\}.$$

Fact 2. *For any K, L convex and $u \in S^{n-1}$, the following properties hold.*

1. *$S_u(K)$ is convex.*
2. *$|S_u(K)| = |K|$.*
3. *$S_u(K) + S_u(L) \subset S_u(K + L)$.*
4. *If $K \subset L$, then $S_u(K) \subset S_u(L)$.*

Proof. 1. Define function

$$\begin{array}{llll} f: & \mathrm{proj}_{u^\perp}(K) & \to & \mathbb{R} \\ & x & \mapsto & |\ell_x| \coloneqq |K \cap (\mathbb{R}u + x)|. \end{array}$$

It suffice to prove f is concave. For any $x, y \in \mathrm{proj}_{u^\perp}(K)$ and $\lambda \in [0,1]$, let $z = \lambda x + (1-\lambda) y$. By convexity of K,

$$\lambda (K \cap (x + \mathbb{R}u)) + (1-\lambda)(K \cap (y + \mathbb{R}u)) \subset K \cap (z + \mathbb{R}u).$$

The fact follows by taking length on both sides and noting that the length of the Minkowski sum of two segments is the sum of their lengths.

2. $|K| = \int_{u^\perp} |K \cap (x + \mathbb{R}u)| \, \mathrm{d}x = \int_{u^\perp} |S_u(K) \cap (x + \mathbb{R}u)| \, \mathrm{d}x = |S_u(K)|$.

3. For any $x = (x', su) \in S_u(K), y = (y', tu) \in S_u(L)$, we want to show $x + y \in S_u(K + L)$. We know $x' \in \mathrm{proj}_{u^\perp}(K), y' \in \mathrm{proj}_{u^\perp}(L)$ and by definition,

$$|s| \leqslant \frac{1}{2}|K \cap (\mathbb{R}u + x)|, \quad |t| \leqslant \frac{1}{2}|L \cap (\mathbb{R}u + y)|.$$

$x + y = (x' + y', (s + t)u)$. It suffices to show

$$|s + t| \leqslant \frac{1}{2}|(K + L) \cap (\mathbb{R}u + x + y)|.$$

Indeed

$$|s + t| \leqslant \frac{1}{2}|K \cap (\mathbb{R}u + x)| + \frac{1}{2}|L \cap (\mathbb{R}u + y)|.$$

Also $|K \cap (\mathbb{R}u + x)| + |L \cap (\mathbb{R}u + y)| \leqslant |(K + L) \cap (\mathbb{R}u + x + y)|$, since again by convexity

$$K \cap (x + \mathbb{R}u) + L \cap (y + \mathbb{R}u) \subset (K + L) \cap (x + y + \mathbb{R}u).$$

4. Obviously, symmetrization preserves inclusion.

□

Proposition 3. *For any $K \subset \mathbb{R}^n$ symmetric and convex, $u \in S^{n-1}$, $|K^\circ| \leqslant |S_u(K)^\circ|$.*

Proof. WLOG, let $u = e_n$. The symmetral of K along u can be written as

$$S_u(K) = \left\{\left(x, \frac{s-t}{2}\right) : (x, s), (x, t) \in K\right\}.$$

And the polar of it can be written as

$$\begin{aligned} S_u(K)^\circ &= \left\{(y, r) : \sup_{(x,s),(x,t) \in K} \left\langle \left(x, \frac{s-t}{2}\right), (y, r)\right\rangle \leqslant 1\right\} \\ &= \left\{(y, r) : \langle x, y\rangle + \frac{s-t}{2} r \leqslant 1, \forall (x, s), (x, t) \in K\right\}. \end{aligned}$$

For a set $A \in \mathbb{R}^n$, define its section along e_n at r as $A(r) := \left\{x \in \mathbb{R}^{n-1} : (x, r) \in A\right\}$. Now let's compute

$$\begin{aligned} \frac{1}{2}(K^\circ(r) + K^\circ(-r)) &= \left\{\frac{y+z}{2} : \forall (x, s), (w, t) \in K, \langle x, y\rangle + sr \leqslant 1, \langle w, z\rangle - tr \leqslant 1\right\} \\ &\subset \left\{\frac{y+z}{2} : \forall (x, s), (x, t) \in K, \langle x, y\rangle + sr \leqslant 1, \langle x, z\rangle - tr \leqslant 1\right\} \\ &\subset \left\{\frac{y+z}{2} : \forall (x, s), (x, t) \in K, \left\langle x, \frac{y+z}{2}\right\rangle + \frac{s-t}{2} r \leqslant 1\right\} \end{aligned}$$

$$\begin{aligned}&=\left\{v\colon \forall\left(x,s\right),\left(x,t\right)\in K,\left\langle x,v\right\rangle+\frac{s-t}{2}r\leqslant 1\right\}\\&=S_u\left(K\right)^\circ\left(r\right).\end{aligned}$$

Note that K° is symmetric since K is symmetric. Indeed, for any $y\in K^\circ$, $h_K\left(-y\right)=h_{-K}\left(y\right)=h_K\left(y\right)\leqslant 1$, i.e., $-y\in K^\circ$. Hence

$$\begin{aligned}K^\circ\left(r\right)&=\left\{x\in\mathbb{R}^{n-1}\colon\ \left(x,-r\right)\in K^\circ\right\}\\&=\left\{x\in\mathbb{R}^{n-1}\colon\ \left(-x,r\right)\in K^\circ\right\}\\&=\left\{-x\in\mathbb{R}^{n-1}\colon\ \left(x,r\right)\in K^\circ\right\}\\&=-K^\circ\left(r\right).\end{aligned}$$

By Brunn–Minkowski,

$$\begin{aligned}\left|S_u\left(K\right)^\circ\left(r\right)\right|\geqslant&\left|\frac{1}{2}\left(K^\circ\left(r\right)+K^\circ\left(-r\right)\right)\right|\\\geqslant&\sqrt{\left|K^\circ\left(r\right)\right|}\sqrt{\left|K^\circ\left(-r\right)\right|}\\=&\left|K^\circ\left(r\right)\right|.\end{aligned}$$

Finally the proposition follows by integrating the sections along r,

$$\left|S_u\left(K\right)^\circ\right|=\int_{\mathbb{R}}\left|S_u\left(K\right)^\circ\left(r\right)\right|\,\mathrm{d}r\geqslant\int_{\mathbb{R}}\left|K^\circ\left(r\right)\right|\,\mathrm{d}r=\left|K^\circ\right|.$$

□

Theorem 4. *For any convex K such that $|K|=|B_2^n|$, there exists $\{u_j\}\subset S^{n-1}$, $K_j=S_{u_j}\left(K_{j-1}\right)$, such that $K_j\to B_2^n$ in Hausdorff distance*[1]*.*

We need the following lemma which is the geometric analog of the fact that every bounded sequence has a convergent subsequence.

Lemma 5 (Blaschke's selection theorem)**.** *Any sequence of convex bodies $\{K_j\}$ such that $K_j\subset RB_2^n$ for all j has a convergent (in Hausdorff distance) subsequence.*

Proof of the theorem. If $K\subset rB_2^n$, then $K_j\subset rB_2^n$ for all j since symmetrization preserves inclusion. Let $R_0\coloneqq\inf_j\inf\left\{r\colon K_j\subset rB_2^n\right\}$ be the infimum of the circumradii of the sequence of convex bodies. Take a subsequence of $\{K_j\}$ with circumradii converging to R_0. By Blaschke's selection theorem, further take a subsequence of the subsequence such that $K_{i_j}\to L$ for some L. Note that R_0 is the circumradius of L. It's left to show that actually $L=R_0B_2^n$. Suppose not, then ∂L misses a cap $C=\partial\left(R_0B_2^n\right)\cap\left\{\left\langle x,u\right\rangle\geqslant R_0-\epsilon\right\}$ for some $u\in S^{n-1}$ and $\epsilon>0$. We can cover $\partial\left(R_0B_2^n\right)$ using reflected images of C wrt a finite sequence of hyperplanes $H_1,\cdots,H_k$.

[1] The hausdorff distance between two convex bodies K and L is defined as $d_H\left(K,L\right)\coloneqq\inf\left\{\delta>0\colon K\subset L+\delta B_2^n, L\subset K+\delta B_2^n\right\}$.

Indeed, to cover $x \in \partial \left(R_0 B_2^n\right)$, we can take $H_x = \left(x - x_0\right)^{\perp}$, where x_0 is the center of C. Then by compactness, take a finite sub-covering. If L misses C, then $S_{H_x^{\perp}}\left(L\right)$ misses both C and its relfected image wrt H_x. Then $L_0 \coloneqq S_{H_k^{\perp}} \circ S_{H_{k-1}^{\perp}} \circ \cdots \circ S_{H_1^{\perp}}\left(L\right)$ misses the whole $\partial \left(R_0 B_2^n\right)$. Let's say $L_0 \subset \left(R_0 - \epsilon\right) B_2^n$ for some $\epsilon > 0$. Suppose $B_2^n \subset t L_0$ for some $t > 0$. Take $L' \in \left\{K_{i_j}\right\}$ such that $d_H\left(L', L\right) \leqslant \epsilon' = \left(\frac{R_0 - \epsilon/2}{R_0 - \epsilon} - 1\right)\frac{1}{t}$. Then

$$\begin{aligned} L' \subset & L_0 + \epsilon' B_2^n \\ \subset & L_0 + \epsilon' t L_0 \\ = & \left(1 + \epsilon' t\right) L_0 \\ \subset & \left(1 + \epsilon' t\right)\left(R_0 - \epsilon\right) B_2^n \\ = & \left(R_0 - \epsilon/2\right) B_2^n, \end{aligned}$$

which contradicts the definition of R_0. □

Theorem 6 (Blaschke–Santaló)**.** *For any symmetric convex body K, $\left|K\right|\left|K^{\circ}\right| \leqslant \left|B_2^n\right|^2$.*

Remark 7. *The LHS of the inequality is $GL\left(n, \mathbb{R}\right)$-invariant, i.e., for any $A \in GL\left(n, \mathbb{R}\right)$,*

$$\left|AK\right|\left|\left(AK\right)^{\circ}\right| = \det\left(A\right)\left|K\right|\left|\left(A^T\right)^{-1} K^{\circ}\right| = \det\left(A\right)\left|K\right|\det\left(A\right)^{-1}\left|K^{\circ}\right| = \left|K\right|\left|K^{\circ}\right|.$$

Proof. By Steiner symmetrization, for any L with $\left|L\right| = \left|B_2^n\right|$, there exists $\left\{u_j\right\} \subset S^{n-1}$, $K_0 = L, K_j = S_{u_j}\left(K_{j-1}\right)$, such that $K_j \to B_2^n$. Take $L = \left(\frac{\left|B_2^n\right|}{\left|K\right|}\right)^{1/n} K$. Note that $\left|L\right| = \left|B_2^n\right|$. By proposition,

$$\left|L^{\circ}\right| = \left|K_0^{\circ}\right| \leqslant \left|K_1^{\circ}\right| \leqslant \cdots \leqslant \left|\left(B_2^n\right)^{\circ}\right| = \left|B_2^n\right|.$$

This finishes the proof by noting that

$$\left|L^{\circ}\right| = \left|\left(\left(\frac{\left|B_2^n\right|}{\left|K\right|}\right)^{1/n} K\right)^{\circ}\right| = \left|\left(\frac{\left|K\right|}{\left|B_2^n\right|}\right)^{1/n} K^{\circ}\right| = \frac{\left|K\right|}{\left|B_2^n\right|}\left|K^{\circ}\right|.$$

□

We finish this note with a conjecture.

Conjecture 8. *For any symmetric convex body K, $\left|K\right|\left|K^{\circ}\right| \geqslant \left|B_1^n\right|\left|B_\infty^n\right|$.*

References

[1] Artstein-Avidan, S., Giannopoulos, A., Milman, V. D., Asymptotic geometric analysis. Part I. Mathematical Surveys and Monographs, 202. *American Mathematical Society, Providence, RI*, 2015.

Pinelis' Inequality

Joshua Siktar

We would like to prove the following slight generalisation of Pinelis' inequality from [1].

Theorem. *Let $\epsilon_1, \epsilon_2, \ldots$ be i.i.d. symmetric ± 1-valued random variables and let g be a standard Gaussian random variable (mean zero, variance one). For every reals $x, a_1, \ldots, a_n$ and positive t we have*

$$\mathbb{P}\left(|x + a_1\epsilon_1 + \ldots + a_n\epsilon_n| > t\right) \leq 25\mathbb{P}\left(|x + (a_1^2 + \ldots + a_n^2)^{1/2} g| > t\right).$$

For $x = 0$ the above reduces to Pinelis' theorem from [1]. We shall repeat the inductive arguent of Bobkov, Götze and Houdré from [2]. We shall need their lemma.

Lemma. *Let g be a standard Gaussian random variable. For every $u \geq \sqrt{3}$ and $\alpha \in (0, \pi/2)$ we have*

$$\mathbb{P}\left(g > \frac{u - \sin\alpha}{\cos\alpha}\right) + \mathbb{P}\left(g > \frac{u + \sin\alpha}{\cos\alpha}\right) \leq 2\mathbb{P}\left(g > u\right).$$

Proof of the theorem. We proceed by induction on n. Suppose $n = 1$. We can assume that $a_1 = 1$. We need to check whether

$$\mathbb{P}\left(|x + \epsilon_1| > t\right) \leq 25\mathbb{P}\left(|x + g| > t\right).$$

This holds trivially for $t > \max\{|x-1|, |x+1|\}$. For $t \leq \max\{|x-1|, |x+1|\}$ we have

$$\min\{t - x, t + x\} = t - |x| \leq \max\{|x-1| - |x|, |x+1| - |x|\} \leq 1,$$

hence

$$\begin{aligned}\mathbb{P}\left(|x + g| > t\right) = \mathbb{P}\left(g > t - x\right) + \mathbb{P}\left(g > t + x\right) &\geq \mathbb{P}\left(g \geq \min\{t - x, t + x\}\right) \\ &\geq \mathbb{P}\left(g > 1\right) > 1/25.\end{aligned}$$

Suppose $n \geq 2$. Assume without loss of generality that x and the a_i are positive and $a_1^2 + \ldots + a_n^2 = 1$. Since $\mathbb{P}\left(g > \sqrt{3}\right) > 1/25$ and $\mathbb{P}\left(|x+g| > t\right) = \mathbb{P}\left(g > t - x\right) + \mathbb{P}\left(g > t + x\right)$, we can also assume that $t \pm x \geq \sqrt{3}$. We have

$$\begin{aligned}\mathbb{P}\left(|x + a_1\epsilon_1 + \ldots + a_n\epsilon_n| > t\right) = &\frac{1}{2}\mathbb{P}\left(|(x + a_1) + a_2\epsilon_2 + \ldots + a_n\epsilon_n| > t\right) \\ &+ \frac{1}{2}\mathbb{P}\left(|(x - a_1) + a_2\epsilon_2 + \ldots + a_n\epsilon_n| > t\right),\end{aligned}$$

so by the inductive assumption it is enough to show that

$$\frac{1}{2}\mathbb{P}\left(|x + a_1 + a'g| > t\right) + \frac{1}{2}\mathbb{P}\left(|x - a_1 + a'g| > t\right) \leq \mathbb{P}\left(|x + g| > t\right),$$

where $a' = (a_2^2 + \ldots + a_n^2)^{1/2}$. Because $a_1^2 + a'^2 = 1$, we can set $a_1 = \sin\alpha$ and $a' = \cos\alpha$ for some $\alpha \in (0, \pi/2)$. Then the above inequality becomes

$$\begin{aligned}&\mathbb{P}\left(g > \frac{t - x - \sin\alpha}{\cos\alpha}\right) + \mathbb{P}\left(g > \frac{t + x + \sin\alpha}{\cos\alpha}\right)\\&+ \mathbb{P}\left(g > \frac{t - x + \sin\alpha}{\cos\alpha}\right) + \mathbb{P}\left(g > \frac{t + x - \sin\alpha}{\cos\alpha}\right) \le 2\mathbb{P}\left(g > t - x\right) + 2\mathbb{P}\left(g > t + x\right).\end{aligned}$$

Putting $u = t - x$, $v = t + x$, we have $u, v \ge \sqrt{3}$ and ask whether

$$\begin{aligned}&\mathbb{P}\left(g > \frac{u - \sin\alpha}{\cos\alpha}\right) + \mathbb{P}\left(g > \frac{u + \sin\alpha}{\cos\alpha}\right)\\&+ \mathbb{P}\left(g > \frac{v + \sin\alpha}{\cos\alpha}\right) + \mathbb{P}\left(g > \frac{v - \sin\alpha}{\cos\alpha}\right) \le 2\mathbb{P}\left(g > u\right) + 2\mathbb{P}\left(g > v\right).\end{aligned}$$

This follows from the lemma.

References

[1] Bobkov, S., Gö tze, F., Houdré, Ch., On Gaussian and Bernoulli covariance representations. *Bernoulli* 7 (2001), no. 3, 439–451.

[2] Pinelis, I., Extremal probabilistic problems and Hotelling's T^2 test under a symmetry condition. *Ann. Statist.* 22 (1994), no. 1, 357–368.

Pinsker's Inequality

Kai Wen Wang

The finite and coin flip examples are adapted from [2]. The general case is adapted from [1].

1 Finite Case

Let P and Q be distributions on a finite universe U with densities p and q respectively. Probability spaces are $(U, \mathcal{F}, P)$, $(U, \mathcal{F}, Q)$.
The **KL-divergence** of P with Q is

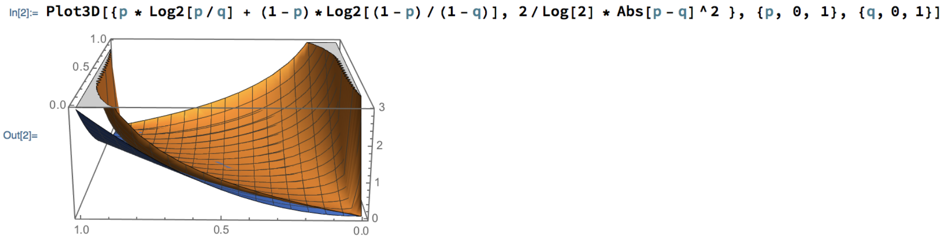
$$D(P||Q) = \sum_{x \in U} p(x) \lg(\frac{p(x)}{q(x)})$$

Remark: $D(P||Q) \geq 0$ with equality iff $P = Q$.

The **total-variation distance** between P and Q is

$$||P - Q||_{TV} = \frac{1}{2}||P - Q||_1 = \frac{1}{2} \sum_{x \in U} |p(x) - q(x)|$$

1.1 U = {0,1}



```
In[2]:= Plot3D[{p * Log2[p / q] + (1 - p) * Log2[(1 - p) / (1 - q)], 2 / Log[2] * Abs[p - q] ^ 2 }, {p, 0, 1}, {q, 0, 1}]
```

Figure 1: Picture that shows $p \lg(p/q) + (1 - p) \lg((1 - p)/(1 - q)) \geq \frac{2}{\ln(2)}|p - q|^2$.

Alternatively, set $f(p, q) = p \lg(p/q) + (1 - p) \lg((1 - p)/(1 - q)) - \frac{2}{\ln(2)}|p - q|^2$ and show that it is zero at $p = q$ and concave up, with minimum at $p = q$.

$$\frac{\delta f}{\delta q} = \frac{p - q}{\ln(2)} \left[4 - \frac{1}{q(1 - q)} \right]$$

1.2 Any finite universe

Idea is to reduce any case to the special $U = \{0, 1\}$ case. Introduce a set $A = \{p \geq q\}$ and $P_A = Ber(\sum_{x \in A} p(x)), Q_A = Ber(\sum_{x \in A} q(x))$. Then, we actually have

$$||P - Q||_1 = ||P_A - Q_A||_1$$

So,

$$\begin{aligned} D(P||Q) &= D(P(X,1_A)||Q(X,1_A)) \\ &= D(P(1_A)||Q(1_A)) + D(P(X|1_A)||Q(X|1_A)) \\ &\geq D(P(1_A)||Q(1_A)) \\ &= D(P_A||Q_A) \\ &\geq \frac{1}{2\ln(2)}||P_A - Q_A||_1^2 \\ &= \frac{1}{2\ln(2)}||P-Q||_1^2 \end{aligned}$$

as desired.
The first equality, though intuitively true, can be shown; if $Z = f(X)$ then $D(P(X)||Q(X)) = D(P(X,Z)||Q(X,Z))$ since $P(X = x, f(X) = y) = P(X = x, f(x) = y) = P(X = x)1_{f(x)=y}$. Therefore, expanding this out gives the desired result.

2 General Case

Let μ and ν be two probability measures on a measurable space (X, F). Then the total-variation distance between μ and ν is defined to be

$$||\mu - \nu||_{TV} = \sup_{A\in F} |\mu(A) - \nu(A)|$$

The relative entropy of ν w.r.t. μ is defined as

$$H(\nu|\mu) = \begin{cases} \int f \ln f d\mu, & \text{if } f = \frac{d\nu}{d\mu} \text{ exists} \\ \infty, & \text{otherwise} \end{cases}$$

Pinsker's inequality states that the total-variation distance is upper bounded by the relative entropy:

$$||\mu - \nu||_{TV}^2 \leq \frac{1}{2}H(\nu|\mu) \tag{1}$$

2.1 Proof of Pinsker's inequality

Assuming f exists, the total variation distance can be written as

$$||\mu - \nu||_{TV} = \frac{1}{2}\int_X |1 - f|d\mu$$

($\leq$) Split into positive and negative parts

$$\begin{aligned} 1 - f &= (1-f)_+ - (1-f)_- \\ |1 - f| &= (1-f)_+ + (1-f)_- \\ \int_X 1 - f &= 0 \end{aligned}$$

By integrating, we see that $\int_X (1-f)_+ d\mu = \int_X (1-f)_- d\mu = \frac{1}{2}\int_X |1-f|d\mu$.
For any $A \in F$,

$$|\mu(A) - \nu(A)| = |\int_A (1-f)d\mu| \leq \max\{\int_A (1-f)_+ d\mu, \int_A (1-f)_- d\mu\} \leq \frac{1}{2}\int_X |1-f|d\mu$$

Taking supremum on both sides over A yields the result.
($\geq$) We witness a single A for equality. Let $A = \{1 \geq f\}$. Then,

$$|\mu(A) - \nu(A)| = \int_A (1-f)_+ d\mu = \int_X (1-f)_+ d\mu = \frac{1}{2}\int_X |1-f| d\mu$$

Now set $u = f - 1$ so that $H(\nu|\mu) = \int f \ln(f) d\mu = \int (1+u)\ln(1+u) - u d\mu$. The $-u$ makes things easier, even though $\int u d\mu = 0$.
Define $\phi(x) = (1+x)\ln(1+x) - x$. We have $\phi'(x) = \ln(1+x), \phi''(x) = \frac{1}{1+x}$.
Thus,

$$\begin{aligned} \phi(t) &= \int_0^t \phi'(x) dx \\ &= -\int_0^t (t-x)'\phi'(x) dx \\ &= \int_0^t (t-x)\phi''(x) dx \\ &= \int_0^t \frac{t-x}{1+x} dx \\ &= t^2 \int_0^1 \frac{1-s}{1+ts} ds \end{aligned}$$

with the last step from $x = ts$.
From this, we have $H(\nu|\mu) = \int_{X\times[0,1]} u^2 \frac{1-s}{1+us} d\mu ds$.
Since

$$||\mu - \nu||_{TV} = \frac{1}{2}\int_X |u| d\mu = \int_{X\times[0,1]} |u|(1-s) d\mu ds$$

By Cauchy-Swartz,

$$\begin{aligned} ||\mu - \nu||_{TV}^2 &= (\int_{X\times[0,1]} |u|(1-s) d\mu ds)^2 \\ &\leq (\int_{X\times[0,1]} |u|^2 \frac{1-s}{1+us} d\mu ds) \cdot (\int_{X\times[0,1]} (1-s)(1+us) d\mu ds) \\ &= \frac{1}{2} H(\mu|\nu) \end{aligned}$$

as desired.

Remark: You may be wondering why the total-variation distance looks a bit different from before, and why we used relative entropy instead of KL-divergence. But if we set $P = \mu, Q = \nu$ and assume that $U = X$ is finite, then using probability densities we see they're equivalent.

3 Example of coin flips

Lemma: Let P, Q be any distributions on U. Let $f : U \to [0, B]$. Then

$$|\mathbb{E}_P[f] - \mathbb{E}_Q[f]| \leq B||P - Q||_{TV}$$

Suppose you have two coins:

$$\texttt{Coin P} = \begin{cases} H & w.p. \ \ \frac{1}{2} - \epsilon \\ T & w.p. \ \ \frac{1}{2} + \epsilon \end{cases} \qquad \texttt{Coin Q} = \begin{cases} H & w.p. \ \ \frac{1}{2} \\ T & w.p. \ \ \frac{1}{2} \end{cases}$$

We want a classifier $f : \{0,1\}^m \to \{0,1\}$ that takes an input of flips and returns 1 if it thinks the coin is P and 0 otherwise. Let's say we want f to predict correctly with probability $\frac{9}{10}$. Equivalently,

$$\mathbb{E}_{x \sim P^m}[f(x)] = \Pr_{x \sim P^m}[f(x) = 1] \geq \frac{9}{10} \text{ and } \mathbb{E}_{x \sim Q^m}[f(x)] = 1 - \Pr_{x \sim Q^m}[f(x) = 0] \leq \frac{1}{10}$$

So we derive a lower bound for $||P^m - Q^m||_1 \geq \frac{8}{5}$ as

$$\frac{4}{5} \leq |\mathbb{E}_{x \sim P^m}[f(x)] - \mathbb{E}_{x \sim Q^m}[f(x)]| \leq \frac{1}{2}||P^m - Q^m||_1$$

Now, we find an upper bound for $||P^m - Q^m||_1$ using Pinsker's inequality

$$m \cdot D(P||Q) = D(P^m||Q^m) \geq \frac{1}{2\ln(2)}||P^m - Q^m||_1^2 \geq \frac{64}{25\ln(2)}$$

We can calculate the KL-divergence explicitly and find an upper bound

$$\begin{aligned} D(P||Q) &= \frac{1}{2}\lg((1-2\epsilon)(1+2\epsilon)) + \epsilon\lg(\frac{1+2\epsilon}{1-2\epsilon}) \\ &\leq \frac{\epsilon}{\ln(2)}\ln(1 + \frac{4\epsilon}{1-2\epsilon}) \\ &\leq \frac{8\epsilon^2}{\ln(2)} \end{aligned}$$

when $\epsilon \leq \frac{1}{4}$.
Simplifying, we get that $m \geq \frac{4}{25\epsilon^2}$.

References

[1] Nathael Gozlan and Christian Léonard. Transport inequalities. a survey. *arXiv preprint arXiv:1003.3852*, 2010.

[2] Madhur Tulsiani. Information and coding theory - autumn 2017. `http://ttic.uchicago.edu/~madhurt/courses/infotheory2017/index.html`.