Lecture 4: Marginal of Convex Set

Lecturer: Yin Tat Lee

Disclaimer: *Please tell me any mistake you noticed.*

Some of the materials are copied from the survey written by Santosh Vempala and me.

4.1 Marginal of Convex Set

After discussed the applications of cutting plane methods, we go back to study a particular cutting plane method, the center of gravity method. This is currently the best cutting plane method in terms of the convergence rate. (I suspect this is the best in the worst case for the general setting.) Recall that in the cutting plane framework, we maintain a convex set $E^{(k)}$ that contains the minimizer x^* of f. Each iteration of this algorithm, we compute the gradient of f at the center of gravity of $E^{(k)}$. The convexity of f shows that $x^* \in H^{(k)} \cap E^{(k)}$ for some half-space $H^{(k)}$ containing the center of gravity. The algorithm continues by setting $E^{(k+1)} = H^{(k)} \cap E^{(k)}$. To analyze the performance, we simply need to bound how much volume we cut each iteration. More precisely, given a convex set K with $\mathbb{E}_{x\sim K} x = 0$, what is

$$\mathbb{P}_{x\sim K}(x^\top \theta \geq 0).$$

In general, we call $p_\theta(t) = \frac{d}{dt}\mathbb{P}_{x\sim K}(x^\top\theta \leq t)$ is the marginal of the convex set. In this lecture, we focus to study the properties of p_θ.

4.1.1 Log-concave distribution

The marginal of convex set satisfies a similar "convexity" as the convex set itself.

Definition 4.1.1. We call a distribution $p(x)$ on $\mathbb{R}^n$ is logconcave if $p(x) = e^{-f(x)}$ for some convex function. Equivalently, $p(\lambda x + (1-\lambda)y) \geq p(x)^\lambda p(y)^{1-\lambda}$ for all $x, y \in \mathbb{R}^n$ and $0 \leq \lambda \leq 1$.

Note that the uniform measure of any convex set K, $p(x) = \begin{cases} \frac{1}{\mathrm{vol} K} & \text{if } x \in K \\ 0 & \text{else} \end{cases}$, is logconcave. Many common probability distributions are also logconcave e.g., Gaussian, exponential, logistic and gamma distributions. An alternative definition of logconcave involves the measures of all sets.

Lemma 4.1.2. *A distribution p is logconcave if and only if $p(\lambda A + (1-\lambda)B) \geq p(A)^\lambda p(B)^{1-\lambda}$ for all compact subsets A and B where $p(S) = \int_S p(x)\,dx$.*

The following lemma shows that the marginal of convex set is logconcave.

Lemma 4.1.3 (Dinghas; Prékopa; Leindler)**.** *The product, minimum, and convolution of two logconcave functions is also logconcave; in particular, any linear transformation or marginal of a logconcave density is logconcave.*

Conversely, all log-concave distributions can be approximated by some (one or higher dimensional) marginal of convex sets. To study log-concave distributions, it is often convenience to renormalize the distribution as follows:

Definition 4.1.4. We call a distribution p is isotropic if $\mathbb{E}_{x\sim p}=0$ and $\mathbb{E}_{x\sim p}xx^\top = I$.

For any distribution with mean μ and covariance A, both are well-defined, we can apply the transformation $A^{-\frac{1}{2}}(x-\mu)$ to get an isotropic distribution.

Lemma 4.1.5. *For any isotropic log-concave distribution on $\mathbb{R}$, $p(x)=e^{-O(|x|)}$ and $p(0)=\Theta(1)$.*

Remark. The proof is just an illustration of the log-concavity. There are more elegant proofs.

Proof. If $p(x)<\frac{1}{8}$ for all $|x|\leq 2$, then we would have more than half of the mass outside $|x|\leq 2$. This contradicts to $\mathbb{E}x^2\leq 1$. Therefore, we have some $|x|\leq 2$ such that $p(x)\geq\frac{1}{8}$.

Next, we note that if $p(18)\geq\frac{1}{16}$, then all t between x and 18 has density larger than $\frac{1}{16}$ and this gives too much mass. Therefore, we have $p(18)\leq\frac{1}{16}$ and $p(-18)\leq\frac{1}{16}$ (for the same reason).

Since p drops from at least $\frac{1}{8}$ to at most $\frac{1}{16}$ between x and 18, the log-concavity shows that it continues to drop at the same or larger rate afterward. Hence, we have that $p(x)=e^{-O(|x|)}$ for all $|x|\geq 18$.

Next, to prove that $p(0)=\Omega(1)$, we note that $p(x)\leq p(0)e^{\frac{p'(0)}{p(0)}x}$. If $p(0)$ is tiny, then one side has too little mass. Combining it with the fact $\mathbb{E}x^2\geq 1$, this contradicts to the fact $\mathbb{E}x=0$.

Finally, to prove that $p(x)=O(1)$ for all x, we note that if not, the density must drop really fast to make sure the total mass is 1. But the logconcavity shows that it must continue to drop and hence contradicts to the condition $\mathbb{E}x^2\geq 1$. □

4.1.2 Center of Gravity Method

The following theorem shows that the volume of the set decreases by $1-\frac{1}{e}$ factor. For computation purpose, it is important to establish a stable version of the Theorem that does not require an exact center of gravity as follows:

Theorem 4.1.6 (Grunbaum Theorem)**.** *Let p be an isotropic logconcave distribution. For any θ and t, we have*

$$\mathbb{P}_{x\sim p}(x^\top\theta\geq t)\geq\frac{1}{e}-O(t).$$

Proof. By taking the marginal with respect to the direction θ, we can assume the distribution is one dimensional. Let $P(t)=\mathbb{P}_{x\sim p}(x^\top\theta\geq t)$. Note that $P(t)$ is the convolution of p and $1_{(-\infty,0]}$. Hence, it is logconcave (Lemma 4.1.3). Without loss of generality, we can assume $P(-M)=0$ and $P(M)=1$. Since $\mathbb{E}_{x\sim p}x=0$, we have that

$$\int_{-M}^{M} tP'(t)=0$$

Integration by parts gives that $\int_{-M}^{M}P(t)dt=M$. Note that $P(t)$ is increasing logconcave, if $P(0)$ is too small, it would make $\int_{-M}^{M}P(t)dt$ too small. Precisely, since P is logconcave, we have that $P(t)\leq P(0)e^{\alpha t}$ for some α. Hence,

$$M=\int_{-M}^{M}P(t)dt\leq\int_{-\infty}^{\frac{1}{\alpha}}P(0)e^{\alpha t}dt+\int_{1/\alpha}^{M}1dt=\frac{eP(0)}{\alpha}+M-\frac{1}{\alpha}.$$

This shows that $P(0)\geq\frac{1}{e}$.

Next, Lemma 4.1.5 shows that $\max_x p(x)=O(1)$. Therefore, P is $O(1)$-Lipschitz and

$$P(t)\geq P(0)-O(t)\geq\frac{1}{e}-O(t).$$

□

Currently, the only known way to compute center of gravity is by taking an empirical average of random samples. For general convex sets, it requires $O(n)$ many samples [22]. The following lemma shows that $O(\log^2 m)$ samples suffice for polytopes.

Lemma 4.1.7. *Let p be a logconcave distribution with $\mathbb{E}_{x\sim p}x = 0$. Let z be N random samples from p. Then, for any vector a, the probability that "$a^\top x \geq a^\top z$ cuts off at least $\frac{1}{3}$ of K" is at least $1 - e^{-\Omega(\sqrt{N})}$.*

Proof. Without loss of generality, we assume that $\mathbb{E}_{x\sim p}a^\top x = 0$ and $\mathbb{E}_{x\sim p}(a^\top x)^2 = 1$. Then, we have that $\mathbb{E}(a^\top z) = 0$ and $\mathbb{E}(a^\top z)^2 = \frac{1}{N}$. Since z is the convolution of p, $\widetilde{z} \stackrel{\text{def}}{=} a^\top z$ follows a logconcave distribution with $\mathbb{E}\widetilde{z}^2 \leq \frac{1}{N}$. Lemma 4.1.5 shows that, for any constant $c > 0$, we have that

$$\mathbb{P}(|a^\top z| > c) \leq \exp(-\Omega(\sqrt{N}))$$

The result follows from Theorem 4.1.6. □

4.2 Random Marginal of Convex Set

The rest of the lecture is devoted to understand the marginals of convex set on a random direction. One can show by calculations that the marginal of a sphere is a very good approximation of Gaussian distribution. It turns out that this is true for a random marginal of any convex set. Unfortunately, the precise understanding of such phenomenon is very much an open problem.

4.2.1 Localization Lemma

One of the key hammer to understand logconcave distribution is the localization lemma [**?**].

Definition 4.2.1. A point x is an extreme point of a set K if x does not lie in any open line segment joining two points of K. Let $\text{Ext}(K)$ be the set of extreme points of K.

Here, we state the modern version of the localization lemma by [23].

Theorem 4.2.2 (Localization Lemma)**.** *Let K be a compact convex set in $\mathbb{R}^n$ and f be an upper semi-continuous function. Let P_f be the set of logconcave distributions μ supported on K satisfying $\int f d\mu \geq 0$. The set of extreme points of $\text{conv}P_f$ is exactly:*

1. *the Dirac measure at points x such that $f(x) \geq 0$, or*
2. *the distribution v satisfies*
   - (a) *density function is of the form $e^{c^\top x + b}$ for some vector c and scalar b,*
   - (b) *support equals to a segment $[a, b] \subset K$,*
   - (c) *$\int f dv = 0$,*
   - (d) *$\int_a^x f dv > 0$ for $x \in (a, b)$ or $\int_x^b f dv > 0$ for $x \in (a, b)$.*

Corollary 4.2.3. *Under the same assumption in Theorem 4.2.2. For any upper semi-continuous convex function Φ, we have that*

$$\sup_{\mu \in P_f} \Phi(\mu)$$

attained by some extreme points of $\text{conv}P_f$, which are classified by Theorem 4.2.2.

In some sense, localization lemma gives us the power to do case analysis on many nontrivial statements about logconcave distributions. Whenever localization lemma can be used, the result will be tight in some sense because it is a direct reduction.

Exercise 4.2.4. Use localization lemma to show the reverse holder inequality for logconcave distributions

$$\left(\mathbb{E}_{x\sim p}\left\|x\right\|_2^k\right)^{1/k} = O(k)\cdot \mathbb{E}_{x\sim p}\left\|x\right\|_2 .$$

4.2.2 Concentration

The phenomenon of concentration of measure appears everywhere. Most of a Euclidean unit ball in $\mathbb{R}^n$ lies within distance $O(\frac{1}{n})$ of its boundary, and also within distance $O(\frac{1}{\sqrt{n}})$ of any central hyperplane. Most of a Gaussian lies in an annulus of thickness $O(1)$. For any subset of the sphere of measure $\frac{1}{2}$, the measure of points at distance at least $\sqrt{\frac{\log n}{n}}$ from the set is a vanishing fraction. These concentration phenomena are closely related to isoperimetry.

Definition 4.2.5. The boundary measure of this subset is

$$p(\partial S) = \inf_{\varepsilon\rightarrow 0^+}\frac{p(S+\varepsilon B_2)-p(S)}{\varepsilon}$$

where εB_2 is the unit ball with radius ε. The isoperimetric constant of p is

$$\psi_p = \inf_{S\subseteq\mathbb{R}^n}\frac{p(\partial S)}{\min\left\{p(S), p(\mathbb{R}^n\setminus S)\right\}}.$$

It is pretty easy to see that large isoperimetry implies good concentration.

Lemma 4.2.6. *Given a distribution p in $\mathbb{R}^n$ with isoperimetric constant ψ_p and a 1-Lipschitz function f on $\mathbb{R}^n$. Then, we have that*

$$\mathbb{P}_{x\sim p}\left(|f(x)-\mathrm{med}_{x\sim p}f(x)|\geq t\right)\leq 2\exp(-\psi_p t).$$

where $\mathrm{med}_{x\sim p}f(x)$ *is the median of f for x sampled from p.*

Proof. Let $m \stackrel{\mathrm{def}}{=} \mathrm{med}_{x\sim p}$, $E_t \stackrel{\mathrm{def}}{=} \{x:\ f(x)\geq m+t\}$, $\alpha_t \stackrel{\mathrm{def}}{=} \log\frac{1}{p(E_t)}$. By the definition of median, we have $\alpha_0 = \log 2$. Also, we have that

$$\frac{d\alpha_t}{dt} = -\frac{1}{p(E_t)}\frac{dp(E_t)}{dt}. \tag{4.1}$$

Since f is 1-Lipschitz, for any x such that $\left\|x-y\right\|_2\leq h$ and $y\in E_t$, we have that

$$g(x)\geq m+t-h.$$

Hence, we have $E_t + hB_2 \subset E_{t-h}$ and hence

$$-\frac{dp(E_t)}{dt} = \lim_{h\rightarrow 0}\frac{p(E_{t-h})-p(E_t)}{h}\geq \lim_{h\rightarrow 0}\frac{p(E_t+hB_2)-p(E_t)}{h} = p(\partial E_t). \tag{4.2}$$

Combining (4.1) and (4.2) gives that

$$\frac{d\alpha_t}{dt}\geq\frac{p(\partial E_t)}{p(E_t)}\geq\psi_p$$

for $t\geq 0$ because $p(E_t)\leq p(E_0)\leq\frac{1}{2}$. Therefore, we have that $\alpha_t\geq\psi_p t$ and hence $p(E_t)\leq\exp(-\psi_p t)$ for $t\geq 0$.

The proof for the case $t\leq 0$ is same. □

For logconcave distribution, it is also true that good concentration implies large isoperimetry [28]. For many particular distributions such as Gaussian distribution, we know precisely the isoperimetric constant. But it is still an active area for general logconcave distributions. For a decade, the best bound was given by the localization lemma. We first need a popular corollary of KLS lemma:

Corollary 4.2.7. *Let f_1, f_2 be two upper semi-continuous non-negative functions on $\mathbb{R}^n$ and f_3, f_4 be two semi-continuous non-negative functions on $\mathbb{R}^n$. Let $\alpha, \beta > 0$. Suppose that $f_1^\alpha f_2^\beta \leq f_3^\alpha f_4^\beta$ and that for every $a, b \in \mathbb{R}^n$, for every exponential distribution v supported on $[a,b]$ (here exponential distribution means $v(x) = e^{c^\top x + d} 1_{[a,b]}$ for some vector x and scalar d),*

$$\left(\int f_1 dv\right)^\alpha \left(\int f_2 dv\right)^\beta \leq \left(\int f_3 dv\right)^\alpha \left(\int f_4 dv\right)^\beta$$

Then, for any logconcave distribution μ on $\mathbb{R}^n$,

$$\left(\int f_1 d\mu\right)^\alpha \left(\int f_2 d\mu\right)^\beta \leq \left(\int f_3 d\mu\right)^\alpha \left(\int f_4 d\mu\right)^\beta .$$

Proof. Rewriting the inequality, it suffices to prove that

$$\frac{\int f_1 d\mu}{\int f_3 d\mu} \leq \left(\frac{\int f_4 d\mu}{\int f_2 d\mu}\right)^t$$

where $t = \alpha/\beta$. Let μ be the given logconcave distribution. Consider P be the set of distribution v satisfying $\int \left(f_1 - \frac{\int f_1 d\mu}{\int f_3 d\mu} f_3\right) dv \geq 0$, equivalently

$$\frac{\int f_1 d\mu}{\int f_3 d\mu} \leq \frac{\int f_1 dv}{\int f_3 dv}.$$

Let $\Phi(v) = \left(\frac{\int f_1 d\mu}{\int f_3 d\mu}\right)^t \int f_2 dv - \int f_4 dv$. Note that $\mu \in P$. Therefore, localization lemma shows that there is a delta measure or an exponential distribution v such that $v \in P$ and $\Phi(\mu) \leq \Phi(v)$. Therefore, we have that

$$\begin{aligned}\left(\frac{\int f_1 d\mu}{\int f_3 d\mu}\right)^t \int f_2 d\mu - \int f_4 d\mu &\leq \left(\frac{\int f_1 d\mu}{\int f_3 d\mu}\right)^t \int f_2 dv - \int f_4 dv \\ &\leq \left(\frac{\int f_1 dv}{\int f_3 dv}\right)^t \int f_2 dv - \int f_4 dv \\ &\leq 0\end{aligned}$$

where the last line follows from the assumption. □

Theorem 4.2.8 ([24])**.** *For any logconcave distribution p,*

$$\psi_p \gtrsim \frac{1}{\mathbb{E}_{x\sim p} \|x - \mu\|_2}$$

where $\mu = \mathbb{E}_{x\sim p} x$. In particular, for any logconcave distribution p with covariance matrix A, we have that

$$\psi_p \gtrsim \frac{1}{\sqrt{\mathrm{tr} A}}.$$

Proof of Theorem 4.2.8. The difficult case is $\mu = 0$. Note that

$$\psi_p = \inf_p \frac{p(\partial S)}{\min(p(S), p(S^c))} \geq \inf_{p, h\to 0^+} \frac{p((S + hB_2)\backslash S)}{h \cdot p(S) p(S^c)}.$$

Hence, it suffices to come up with the largest λ such that

$$\mathbb{E}\left\|x\right\|_2 \cdot p((S+hB_2)\backslash S) \geq \lambda \cdot p(S) \cdot p(S^c).$$

Writing it differently, we have

$$\int \left\|x\right\|_2 dp \cdot \int 1_{(S+hB_2)\backslash S} dp \geq \lambda \cdot \int 1_S dp \cdot \int 1_{S^c} dp.$$

By Corollary 4.2.7, it suffices to find the largest λ among delta measure and exponential distribution. The rest is just a case analysis and one would get $\lambda = \Omega(h)$.

The second part follows from the calculation that $\mathrm{tr}A = \mathbb{E}\left\|x-\mu\right\|_2^2 \geq (\mathbb{E}\left\|x-\mu\right\|_2)^2$. □

Although the isoperimetric constant lower bound is tight for some distribution, it is not a constant approximation for general logconcave distributions. In the course of the study of algorithms for computing the volume, in 1995, Kannan, Lovász and Simonovits conjectured the following tight characterization of isoperimetric constant for logconcave distributions.

Conjecture 4.2.9 (KLS Conjecture)**.** *For any logconcave distribution p with covariance A,*

$$\frac{1}{\sqrt{\lambda_{\max}(A)}} \gtrsim \psi_p \gtrsim \frac{1}{\sqrt{\lambda_{\max}(A)}}.$$

In particular, $\psi_p = \Theta(1)$ for isotropic logconcave distribution p.

We note that the left-hand side is known and can be derived by Lemma 4.1.5. For the right-hand side, the current best technique is via the stochastic localization technique, a stronger technique to reduce problem to one dimensional.

Theorem 4.2.10 ([25])**.** *For any logconcave distribution p with covariance A, we have $\psi_p \gtrsim \frac{1}{(\mathrm{tr}(A^2))^{1/4}}$. In particular, $\psi_p = \Omega(n^{-1/4})$ for isotropic logconcave distribution p.*

Combining this theorem and Lemma 4.2.6, we have the following concentration inequality. One can think it as a generalization of Lemma 4.1.5.

Theorem 4.2.11. *For isotropic logconcave distribution p and any $t > 0$,*

$$\mathbb{P}_{x\sim p}(\left|\|x\| - \sqrt{n}\right| \geq n^{1/4}t) \leq e^{-O(t)}.$$

This shows that any isotropic logconcave distribution concentrated to a shell of radius $\sqrt{n}$ with width $n^{1/4}$. Using this, one can easily prove that a random marginal of a convex set is close to a Gaussian distribution.

The KLS conjecture is still wide open even. See [26] for the latest development on this. The conjecture is even open for very simple convex set like the following:

Problem 4.2.12. We call a convex set K is unconditional if $(x_1, x_2, \cdots, x_n) \in K$ implies $(\pm x_1, \pm x_2, \cdots, \pm x_n) \in K$ for any sign patterns. Does KLS conjecture hold for this class of convex sets?

I end this lecture with a remark that there is a variant of Theorem 4.2.8 in the Riemannian setting.

Theorem 4.2.13 ([27])**.** *If $K \subset (M, g)$ is a locally convex bounded domain with smooth boundary, diameter D and $Ric_g \geq 0$, then the Poincaré constant is at least $\frac{\pi^2}{4D^2}$, i.e., for any g with $\int g = 0$, we have that*

$$\int \left|\nabla g(x)\right|^2 dx \geq \frac{\pi^2}{4D^2} \int g(x)^2 dx.$$

Problem 4.2.14. How to formulate the KLS conjecture for manifolds?

References

[22] Dimitris Bertsimas and Santosh Vempala. Solving convex programs by random walks. *Journal of the ACM (JACM)*, 51(4):540–556, 2004.

[23] Matthieu Fradelizi and Olivier Guédon. The extreme points of subsets of s-concave probabilities and a geometric localization theorem. *Discrete & Computational Geometry*, 31(2):327–335, 2004.

[24] Ravi Kannan, László Lovász, and Miklós Simonovits. Isoperimetric problems for convex bodies and a localization lemma. *Discrete & Computational Geometry*, 13(1):541–559, 1995.

[25] Yin Tat Lee and Santosh Vempala. Eldan's stochastic localization and the kls hyperplane conjecture: An improved lower bound for expansion. In *Foundations of Computer Science (FOCS), 2017 IEEE 58th Annual Symposium on*. IEEE, 2017.

[26] Yin Tat Lee and Santosh S Vempala. Stochastic localization+ stieltjes barrier= tight bound for log-sobolev. *arXiv preprint arXiv:1712.01791*, 2017.

[27] Peter Li and Shing Tung Yau. Estimates of eigenvalues of a compact riemannian manifold. *Geometry of the Laplace operator*, 36:205–239, 1980.

[28] Emanuel Milman. On the role of convexity in isoperimetry, spectral gap and concentration. *Inventiones mathematicae*, 177(1):1–43, 2009.