Lecture 16: Barriers

Lecturer: Yin Tat Lee

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

Recall the last lecture, we introduced ν-self concordant barrier and proved the following result:

Theorem 16.0.6. *Given a ν-self concordant barrier ϕ on K (namely, $\phi(x) \to +\infty$ as $x \to \partial K$, $D^3\phi(x)[h,h,h] \leq 2(D^2\phi(x)[h,h])^{3/2}$ and $(\nabla\phi(x))^\top(\nabla^2\phi(x))^{-1}(\nabla\phi(x)) \leq \nu$), we can solve*

$$\min_{x\in K} c^\top x$$

in $O^*(\sqrt{\nu})$ iterations and each iteration involves computing $\left(\nabla^2\phi(x)\right)^{-1}(y+\nabla\phi(x))$ for some vector y.

In this lecture, we will visit some common ways to construct barrier for convex set in practice and an existence result. We note that there are other applications of self-concordant barrier.

16.1 Logarithmic barrier

For polytope $K = \{Ax \geq b\}$, we can simply think it as $\bigcap_i \{a_i^\top x \geq b_i\}$. For each constraint individually, the optimal barrier is $-\ln(a_i^\top x - b_i)$. Therefore, one popular barrier to use for K is the logarithm barrier

$$-\sum_{i=1}^m \ln(a_i^\top x - b_i).$$

Since each barrier has self-concordance 1, the sum has self-concordance m. However, as an exercise, we have a direct proof here:

Lemma 16.1.1. *$\phi(x) = -\sum_{i=1}^m \ln(a_i^\top x - b_i)$ has self-concordance m.*

Proof. (I have deliberately skipped steps in this proof and you should convince yourself every step). Note that

$$\nabla\phi(x) = -A^\top \frac{1}{s}$$

where $s_i = a_i^\top x - b_i$ and each row of A is $a_i^\top$. Then, we have

$$\nabla^2\phi(x) = A^\top S^{-2} A.$$

Hence, we have that

$$\nabla\phi(x)(\nabla^2\phi(x))^{-1}\nabla\phi(x) \leq 1^\top SA(A^\top S^{-2}A)^{-1}A^\top S1 \leq 1^\top 1 = m$$

where we used that $SA(A^\top S^{-2}A)^{-1}A^\top S$ is an orthogonal projection matrix.

For the self concordance, note that

$$D\phi(x)[h,h,h] = -2\sum_i \left(\frac{a_i^\top h}{s_i}\right)^3 \leq 2\left(\sum_i \left(\frac{a_i^\top h}{s_i}\right)^2\right)^{3/2} = 2(D\phi(x)[h,h])^{3/2}.$$

□

More generally, suppose we want to solve the problem

$$\min_x \sum_i f_i(a_i^\top x - b_i)$$

where f_i are convex functions on scalars. Then we can rewritten it as

$$\min_{s,t} \sum_i t_i \text{ subjects to } \bigcap_i \{(s_i,t_i) : f_i(a_i^\top x - b_i) \leq t_i\}$$

Note that $\{(s_i,t_i) : f_i(s_i) \leq t_i\}$ is a 2 dimensional convex function and hence we can find a barrier $\phi_i(x)$ (say by the existence proof in the next section). Then, we can simply use the barrier

$$\sum_i \phi_i(a_i^\top x - b_i).$$

16.2 Entropic barrier

For any convex set in $\mathbb{R}^n$, Nesterov and Nemirovski constructed a barrier, called universal barrier, with self-concordance $O(n)$ [90]. Later, Hildebrand constructed n-self concordant barriers for any convex cone, called canonical barrier [89]. This implies a $n+1$ for convex sets. The first one can be computed in polynomial time but very expensive. The second seems to need even exponential space to compute. In this lecture, we will discuss the entropic barrier introduced by [88]. They proved the self-concordance is $n + O(\sqrt{n})$. For simplicity, we will only show $O(n)$. The reason I discuss the last one is that ~~Sébastien is my good collaberator~~ the barrier itself is very natural despite the proof is slightly more complicated than the universal barrier.

The family of the exponential distribution on K

$$p_\theta(y) \stackrel{\text{def}}{=} \frac{e^{\theta^\top y} 1_{y\in K}}{\int_K e^{\theta^\top z} dz}$$

is (one of) the most natural family of distribution on K. As we see from localization lemma, these are in some sense extreme distributions among logconcave distributions. For any $x \in K$, there is an unique $\theta(x)$ such that

$$\mathbb{E}_{y\sim p_{\theta(x)}} y = x.$$

When x is closer to the boundary of K, the distribution $p_{\theta(x)}$ need to be more concentrated to make the mean equals to x. Hence, these $p_{\theta(x)}$ has smaller entropy. The entropic barrier is simply the negative entropy of $p_{\theta(x)}$:

$$\phi(x) = \int p_{\theta(x)}(y) \log p_{\theta(x)}(y) dy.$$

To check this barrier makes sense, we note that $\theta(x) = -\frac{1}{x}$ for $K = [0,\infty)$ and hence

$$p_{\theta(x)}(y) = \frac{1}{x} e^{-\frac{y}{x}} 1_{x\geq 0}.$$

Hence, we have

$$\phi_{[0,\infty)}(x) = \int_0^\infty \frac{1}{x} e^{-\frac{y}{x}} \log(\frac{1}{x} e^{-\frac{y}{x}}) dy = -\log(x) - 1$$

which is exactly the logarithmic barrier! Roughly speaking, this barrier must blows up like $-\log$ on the boundary because of the entropy.

16.2.1 Derivatives of the entropic barrier

First, we need to compute the derivatives of the entropic barrier. There is a better proof in [88]. However, I decide to give a straightforward but long proof of this and the purpose is just to have again some exercise on chain rule. The proof is long but just a simple calculation. The best way to read this lemma is to try it yourself and read the proof only if you failed.

Lemma 16.2.1. *We have that* $\nabla\phi(x) = \theta(x)$, $\nabla^2\phi(x) = Cov(x)^{-1}$ *and*

$$D^3\phi(x)[h,h,h] = -\mathbb{E}_{y\sim p_{\theta(x)}}\left((y-x)^\top (Cov(x))^{-1}h\right)^3$$

where $Cov(x) = \mathbb{E}_{y\sim p_{\theta(x)}}(y-x)(y-x)^\top$.

Proof. By chain rule, the directional derivative of $\phi(x)$ on the direction h is

$$\begin{aligned} D\phi(x)[h] &= \int_K Dp_{\theta(x)}(y)[h]\log p_{\theta(x)}(y)dy + \int_K p_{\theta(x)}(y)\frac{Dp_{\theta(x)}(y)[h]}{p_{\theta(x)}(y)}dy \\ &= \int_K Dp_{\theta(x)}(y)[h]\log p_{\theta(x)}(y)dy \end{aligned}$$

Using the formula $p_{\theta(x)}(y) = \frac{e^{\theta(x)^\top y}}{\int e^{\theta(x)^\top z}dz}$, we can simplify the formula as

$$\begin{aligned} D\phi(x)[h] &= \int_K Dp_{\theta(x)}(y)[h]\log\frac{e^{\theta(x)^\top y}}{\int e^{\theta(x)^\top z}dz}dy \\ &= \int_K Dp_{\theta(x)}(y)[h](\theta(x)^\top y)dy - \int_K Dp_{\theta(x)}(y)[h]dy\cdot\log\int_K e^{\theta(x)^\top z}dz \\ &= \int_K Dp_{\theta(x)}(y)[h](\theta(x)^\top y)dy. \end{aligned} \tag{16.1}$$

Now, we calculate $Dp_{\theta(x)}(y)$:

$$\begin{aligned} Dp_{\theta(x)}(y)[h] &= \frac{e^{\theta(x)^\top y}}{\int e^{\theta(x)^\top z}dz}D\theta(x)[h]^\top y - \frac{e^{\theta(x)^\top y}\cdot\int_K e^{\theta(x)^\top z}D\theta(x)[h]^\top z dz}{\int_K e^{\theta(x)^\top z}dz\cdot\int_K e^{\theta(x)^\top z}dz} \\ &= \frac{e^{\theta(x)^\top y}}{\int_K e^{\theta(x)^\top z}dz}D\theta(x)[h]^\top y - \frac{e^{\theta(x)^\top y}\cdot D\theta(x)[h]^\top\int_K e^{\theta(x)^\top z}zdz}{\int_K e^{\theta(x)^\top z}dz\cdot\int_K e^{\theta(x)^\top z}dz} \\ &= \frac{e^{\theta(x)^\top y}}{\int_K e^{\theta(x)^\top z}dz}D\theta(x)[h]^\top y - \frac{e^{\theta(x)^\top y}\cdot D\theta(x)[h]^\top x}{\int_K e^{\theta(x)^\top z}dz} \\ &= p_{\theta(x)}(y)\cdot D\theta(x)[h]^\top(y-x). \end{aligned} \tag{16.2}$$

Recall that $\mathbb{E}_{y\sim p_{\theta(x)}}y = x$. Hence, the covariance matrix of $p_{\theta(x)}$ is given by $\mathrm{Cov}(x) = \mathbb{E}_{y\sim p_{\theta(x)}}(y-x)(y-x)^\top$. Using this notation and putting (16.2) into (16.1) gives that

$$\begin{aligned} D\phi(x)[h] &= \int_K p_{\theta(x)}(y)\cdot D\theta(x)[h]^\top(y-x)\cdot(\theta(x)^\top y)dy \\ &= D\theta(x)[h]^\top\mathbb{E}_{y\sim p_{\theta(x)}}((y-x)y^\top)\theta(x) \\ &= D\theta(x)[h]^\top\mathbb{E}_{y\sim p_{\theta(x)}}((y-x)(y-x)^\top)\theta(x) \\ &= D\theta(x)[h]^\top\mathrm{Cov}(x)\cdot\theta(x). \end{aligned} \tag{16.3}$$

The trick to compute $D\theta(x)[h]$ is to take the derivative of the definition of θ, which is $\mathbb{E}_{y\sim p_{\theta(x)}} y = x$, on the both sides. This gives

$$\begin{aligned} h &= D\mathbb{E}_{y\sim p_{\theta(x)}} y[h] \\ &= \int_K p_{\theta(x)}(y)(D\theta(x)[h]^\top (y-x)) \cdot y dy \\ &= \left(\mathbb{E}_{y\sim p_{\theta(x)}} y \cdot (y-x)^\top\right) D\theta(x)[h] \\ &= \mathrm{Cov}(x) \cdot D\theta(x)[h]. \end{aligned}$$

Therefore, we have

$$D\theta(x)[h] = (\mathrm{Cov}(x))^{-1} h. \tag{16.4}$$

Putting (16.4) into (16.3) gives that

$$D\phi(x)[h] = h^\top \theta(x).$$

Equivalently, this shows $\nabla \phi(x) = \theta(x)$.

Using (16.4), we have

$$D^2\phi(x)[h,h] = h^\top (\mathrm{Cov}(x))^{-1} h$$

Equivalently, this shows $\nabla^2 \phi(x) = (\mathrm{Cov}(x))^{-1}$.

Finally, we have

$$D^3\phi(x)[h,h,h] = -h^\top (\mathrm{Cov}(x))^{-1} D\mathrm{Cov}(x)[h] (\mathrm{Cov}(x))^{-1} h.$$

where

$$\begin{aligned} D\mathrm{Cov}(x)[h] &= \int_K (y-x)(y-x)^\top p_{\theta(x)}(y) dy \\ &= \int_K (y-x)(y-x)^\top p_{\theta(x)}(y) \cdot D\theta(x)[h]^\top (y-x) dy. \end{aligned}$$

Hence, we have

$$D^3\phi(x)[h,h,h] = -\mathbb{E}_y \left((y-x)^\top (\mathrm{Cov}(x))^{-1} h\right)^3.$$

□

16.2.2 Self-concordance

Lemma 16.2.2. *We have that* $D^3\phi(x)[h,h,h] = O(1)\left(D^2\phi(x)[h,h]\right)^{3/2}$.

Proof. Note that $D^3\phi(x)[h,h,h] = -\mathbb{E}_{y\sim p_\theta}\left((y-x)^\top (\mathrm{Cov}(x))^{-1} h\right)^3$. Note that any logconcave distribution p with mean x, we have

$$\left|\mathbb{E}_{y\sim p}((y-x)^\top h)^k\right| = k^{O(k)} \left(\mathbb{E}_{y\sim p}((y-x)^\top h)^2\right)^{k/2}.$$

(We asked a similar question in the assignment and it can be proved by localization lemma.) Therefore, we have that

$$\begin{aligned} D^3\phi(x)[h,h,h] &= O(1)\left(\mathbb{E}_{y\sim p_\theta}\left((y-x)^\top (\mathrm{Cov}(x))^{-1} h\right)^2\right)^{3/2} \\ &= O(1)\left(\mathbb{E}_{y\sim p_\theta} h(\mathrm{Cov}(x))^{-1}(y-x)(y-x)^\top (\mathrm{Cov}(x))^{-1} h\right)^{3/2} \\ &= O(1)\left(D^2\phi(x)[h,h]\right)^{3/2}. \end{aligned}$$

□

The proof of the following lemma is a little bit rough. I ignored some small details that I expect you can fill in.

Lemma 16.2.3. *We have that* $\nabla\phi(x)^\top(\nabla^2\phi(x))^{-1}\nabla\phi(x) = O(n)$.

Proof. We have that

$$\nabla\phi(x)^\top(\nabla^2\phi(x))^{-1}\nabla\phi(x) = \theta(x)^\top \mathrm{Cov}(x)\theta(x) = \|\theta\|_2^2 \cdot \mathrm{Var}(y_\theta)$$

where $y_\theta = \langle y, \theta/\|\theta\|_2\rangle$ with $y \sim p_\theta$. Let u be the log-density of y_θ and v be the log-marginal of the uniform measure on K in the direction $\theta/\|\theta\|_2$. Note that

$$u(t) = v(t) + t\,\|\theta\|_2 + \mathrm{cst}.$$

Since v is the log-marginal of uniform measure of convex set, Brunn-Minkowski shows that $v''(t) \leq -\frac{1}{n}(v'(t))^2$ and so

$$u''(t) \leq -\frac{1}{n}(u'(t) - \|\theta\|_2)^2$$

which implies that for any $|u'(t)| \leq \frac{\|\theta\|_2}{2}$, we have

$$u''(t) \leq -\frac{\|\theta\|_2^2}{4n}$$

and hence

$$u(t) \leq -\frac{|t-t_0|^2}{8n/\|\theta\|_2^2} + u(t_0).$$

Using this and Lemma 4.1.5, we have that the variance of u is $O(n/\|\theta\|_2^2)$. Therefore,

$$\nabla\phi(x)^\top(\nabla^2\phi(x))^{-1}\nabla\phi(x) = O(n).$$

□

Problem 16.2.4. Construct a n-self concordance for every convex set. (or prove it does not exist for some convex set.)

16.2.3 Property of central path

Recall that the central path is given by

$$x_t \stackrel{\text{def}}{=} \arg\min_x tc^\top x + \phi(x).$$

For the entropic barrier, we have that

$$x_t = \arg\min_x tc^\top x + \int_K p_{\theta(x)}(y)\log p_{\theta(x)}(y)dy.$$

Since $\nabla\phi(x) = \theta(x)$, we have that $\theta(x) = -tc$. Now using the definition of $\theta(x)$, x_t is the mean of the distribution

$$\frac{1}{Z_t}e^{-tc^\top y}1_{y\in K}.$$

This is exactly the simulated annealing algorithm.

16.3 Log-determinant barrier

Finally, I want to emphasize that some convex set has self-concordance less than the dimension of the problem. These sets often are very symmetric.

Lemma 16.3.1. *The function $\phi(X) = -\log\det X$ is a n self-concordant barrier on $\mathcal{P}_n = \{X \in \mathbb{R}^{n\times n}$ with $X = X^\top, X \succeq 0\}$.*

Proof. Clearly, $\phi(X) \to +\infty$ as $X \to \partial\mathcal{P}_n$. Next, we note that

$$\begin{aligned} D\phi(X)[H] &= -\mathrm{tr}(X^{-1}H), \\ D^2\phi(X)[H,H] &= \mathrm{tr}(X^{-1}HX^{-1}H), \\ D^3\phi(X)[H,H,H] &= -2\mathrm{tr}(X^{-1}HX^{-1}HX^{-1}H). \end{aligned}$$

Let $\widetilde{H} = X^{-\frac{1}{2}}HX^{-\frac{1}{2}}$. Then, we have

$$D^3\phi(X)[H,H,H] = -2\mathrm{tr}(\widetilde{H}^3) \leq 2\mathrm{tr}(\widetilde{H}^2)^{3/2} = 2D^2\phi(X)[H,H].$$

Also, we have that

$$D\phi(X)[H] = -\mathrm{tr}\widetilde{H} \leq \sqrt{n\cdot\mathrm{tr}\widetilde{H}^2} = \sqrt{n}\sqrt{D^2\phi(X)[H,H]}$$

which implies that

$$\begin{aligned} \nabla\phi(x)\left(\nabla^2\phi(X)\right)^{-1}\nabla\phi(x) &\leq \sqrt{n}\sqrt{D^2\phi(X)[(\nabla^2\phi(X))^{-1}\nabla\phi(x), (\nabla^2\phi(X))^{-1}\nabla\phi(x)]} \\ &= \sqrt{n}\sqrt{\nabla\phi(x)\,(\nabla^2\phi(X))^{-1}\,\nabla\phi(x)}. \end{aligned}$$

Hence, we have $\nabla\phi(x)\left(\nabla^2\phi(X)\right)^{-1}\nabla\phi(x) \leq n$. □

References

[88] Sébastien Bubeck and Ronen Eldan. The entropic barrier: a simple and optimal universal self-concordant barrier. In *Proceedings of The 28th Conference on Learning Theory, COLT 2015, Paris, France, July 3-6, 2015*, page 279, 2015.

[89] Roland Hildebrand. Canonical barriers on convex cones. *Mathematics of operations research*, 39(3):841–850, 2014.

[90] Yurii Nesterov and Arkadii Nemirovskii. *Interior-point polynomial algorithms in convex programming.* SIAM, 1994.