CSE 599: Interplay between Convex Optimization and Geometry **Winter 2018**

Lecture 5: John Ellipsoid

Lecturer: Yin Tat Lee

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

The algorithm at the end is unpublished. Feel free to contact me for collaboration.

5.1 John Ellipsoid

In the last lecture, we discussed that any convex set is very close to an ellipsoid in probabilistic sense. More precisely, after renormalization by covariance matrix, we have $\|x\|_2 = \sqrt{n} \pm \Theta(1)$ with high probability. In this lecture, we will talk about how convex set is close to an ellipsoid in a strict sense. If the convex set is isotropic, it is close to a sphere as follows:

Theorem 5.1.1. *Let K be a convex body in $\mathbb{R}^n$ in isotropic position. Then,*

$$\sqrt{\frac{n+1}{n}} B_n \subseteq K \subseteq \sqrt{n(n+1)} B_n.$$

Roughly speaking, this says that any convex set can be approximated by an ellipsoid by a n factor. This result has a lot of applications. Although the bound is tight, making a body isotropic is pretty time-consuming. In fact, making a body isotropic is the current bottleneck for obtaining faster algorithm for sampling in convex sets. Currently, it can only be done in $O^*(n^4)$ membership oracle plus $O^*(n^5)$ total time.

Problem 5.1.2. Find a faster algorithm to approximate the covariance matrix of a convex set.

In this lecture, we consider another popular position of a convex set called John position and its corresponding ellipsoid is called John ellipsoid.

Definition 5.1.3. Given a convex set K. The John ellipsoid $J(K)$ of K is the maximum volume ellipsoid inside K. We call K is in John ellipsoid position if $J(K)$ is an unit ball centered at 0.

This ellipsoid has few properties that are useful for computational purposes. The first one is that the volume of John ellipsoid shrinks by constant factor when the body is cut through the center of John ellipsoid.

Theorem 5.1.4. *For any convex set K, let x be the center of $J(K)$ and let H be any half space containing x. Then,*

$$\mathrm{vol}(J(K \cap H^c)) \leq 0.87 \mathrm{vol}(J(K)).$$

The second one is its rounding property.

Theorem 5.1.5. *For any convex set K, $J(K) \subset K \subset nJ(K)$ and for any symmetric convex set K, $J(K) \subset K \subset \sqrt{n}J(K)$.*

In particular, this shows that any normed space can be approximated by ℓ_2 space by $\sqrt{n}$ factor.

Corollary 5.1.6. *For any norm $\|\cdot\|$, we can find a matrix A such that*

$$\|Ax\|_2 \leq \|Ax\| \leq \sqrt{n} \|Ax\|_2$$

for all x.

5.1.1 Existence and Uniqueness of John Ellipsoid

We first show that John ellipsoid can be computed by using convex programming.

Theorem 5.1.7. *Given any convex set K. Define $a_i^\top x \leq 1$ be the separating hyperplanes of K, namely $K = \bigcap_{i\in I}\{x : a_i^\top x \leq 1\}$. Then, $J(K)$ uniquely exists and is given by $J(K) = \{x : \left\|G^{-1}(x-v)\right\|_2 \leq 1\}$ where G is the maximizer of the problem*

$$\max_{G\succeq 0, v\in\mathbb{R}^n} \log\det G \quad \textit{subjects to} \quad \|Ga_i\|_2 \leq 1 - a_i^\top v \textit{ for all } i \in I. \tag{5.1}$$

Proof. Let us represent an ellipsoid by $E = \{x : \left\|G^{-1}(x-v)\right\|_2 \leq 1\}$. Note that $E \subset K$ if and only if

$$\begin{aligned} 1 \geq \max_{x\in E} a_i^\top x &= \max_{\|G^{-1}(x-v)\|_2\leq 1} a_i^\top (x-v) + a_i^\top v \\ &= \max_{\|y\|_2\leq 1} (G_i a_i)^\top y + a_i^\top v \\ &= \|G_i a_i\|_2 + a_i^\top v. \end{aligned}$$

Since the log of the volume of E is proportional to $\log\det G$, the problem of finding an John ellipsoid can be written as

$$\max_{G\succeq 0, v\in\mathbb{R}^n} \log\det G \quad \text{subjects to} \quad \|Ga_i\|_2 \leq 1 - a_i^\top v \text{ for all } i \in I.$$

Since the constraint set is convex and $-\log\det G$ is strictly convex (proved in the next lemma), $J(K)$ uniquely exists. □

Definition 5.1.8. We call a function $f : \mathbb{R}^n \to \mathbb{R}$ is strictly convex if $\nabla^2 f(x) \succ 0$ for all $x \in \text{dom} f$.

To compute the Hessian of f, we need some calculus notations:

Definition 5.1.9. For any continuously k-th differentiable function $f : \mathbb{R}^n \to \mathbb{R}^m$, we define the directional derivative of f on direction h is

$$Df(x)[h] = \frac{d}{dt} f(x+th)\bigg|_{t=0}.$$

Similarly, the k-th directional derivative of f on directions h_i is

$$D^k f(x)[h_1, h_2, \cdots, h_k] = \frac{d^k}{dt_1 dt_2 \cdots dt_k} f(x + \sum_{i=1}^k t_i h_i)\bigg|_{t_i=0}.$$

Now, we have some fun exercises of matrix calculus. We will continue using matrix calculus in later part of the course and you should be comfortable with the matrix calculus in general.

Lemma 5.1.10. *Let X be a symmetric matrix. We have that*

- *For $f(X) = X^{-1}$, $Df(X)[H] = X^{-1}HX^{-1}$ and $D^2 f(X)[H_1, H_2] = X^{-1}H_1X^{-1}H_2X^{-1} + X^{-1}H_2X^{-1}H_1X^{-1}$.*
- *For $f(X) = \text{tr}(g(X))$ for any nice enough scalar function g, $Df(X)[H] = \text{tr}(g'(X)H)$.*
- *For $f(X) = \log\det X$, $Df(X)[H] = \text{tr}(X^{-1}H)$.*

Proof. For $f(X) = X^{-1}$, we note that $Xf(X) = I$. Hence, we have

$$0 = Hf(X) + X \cdot Df(X)[H] = HX^{-1} + X \cdot Df(X)[H].$$

Now, we apply $Df(X)[H] = X^{-1}HX^{-1}$ repeatedly and get $D^2 f(X)[H_1, H_2]$.

For $f(X) = \text{tr}(g(X))$, suppose there is a sequence of polynomial $g_l(x)$ such that $g_l(x)$ and $g_l'(x)$ converges uniformly to $g(x)$ and $g'(X)$ on an open set containing the eigenvalues of X. It is easy to check that $D\text{tr}(g_l(X))[H] = \text{tr}(g_l'(X)H)$. Taking limit, we have $Df(X)[H] = \text{tr}(g'(X)H)$.

For $f(X) = \log\det X$, we note that $\log\det X = \text{tr}\log X$. Hence, we have $Df(X)[H] = \text{tr}(X^{-1}H)$. □

Now, we can prove that $-\log\det X$ is strictly convex.

Lemma 5.1.11. *Let $f(X) = -\log\det X$. For any $X \succ 0$, we have that*

$$D^2 f(X)[H,H] = \left\| X^{-\frac{1}{2}} H X^{-\frac{1}{2}} \right\|_F^2 > 0$$

for any non-zero H.

Proof. By the calculus rules mentioned above, we have that $Df(X)[H] = -\mathrm{tr}(X^{-1}H)$ and that $D^2 f(X)[H,H] = \mathrm{tr}(X^{-1}HX^{-1}H) = \left\| X^{-\frac{1}{2}} H X^{-\frac{1}{2}} \right\|_F^2$. □

Now, let us explain why we should expect the volume of John ellipsoid decreases by a constant factor after a cut (Theorem 5.1.4). Without loss of generality, we can assume K is in John position and hence $J(K)$ is the unit ball. Let H be the half space in Theorem 5.1.4 and let $J(K \cap H^c) = \{x : \ (x-v)^\top A(x-v)\}$. Since the unit sphere is the maximizer of the problem (5.1), the Hessian of $\log\det(X)$ shows that

$$\log\det A \leq \log\det I - \frac{\|A - I\|_F^2}{(1 + \|A - I\|_F)^2}.$$

Therefore, either the volume of $J(K \cap H^c)$ is a constant factor smaller than $J(K)$ (then we are done), or we have that $\|A - I\|_F^2 \leq 0.001$. In the later case, A is very close to a sphere. Since $J(K \cap H^c)$ does not contain 0, it implies that the center between $J(K)$ and $J(K \cap H^c)$ is constant far away. Using this, one can construct an ellipsoid with volume larger than the unit sphere and hence it draws a contradiction.

5.1.2 Structure of John Ellipsoid for symmetric body

The classical way to compute John ellipsoids is via semidefinite programming or (5.1) [33]. This makes people believe that computing John ellipsoid is a computational intensive task. To get a faster algorithm, we need to have a more structural understanding of John ellipsoid. We first study the structure of John ellipsoid for symmetric convex set (i.e $K = -K$). For notation simplicity, we assume the body is a polytope.

Lemma 5.1.12. *Let $K = \{x \in \mathbb{R}^n : \ |(Ax)_i| \leq 1\}$ a polytope with nonempty interior where $A \in \mathbb{R}^{m \times n}$. Then, $J(K)$ is given by $\{x : \ x^\top A^\top W A x \leq 1\}$ with a diagonal matrix $W \succ 0$. Furthermore, the weight satisfies $\sum_i w_i = n$ and is given by the convex problem*

$$\max_{w_i \geq 0} \log\det \sum_i w_i a_i a_i^\top - \sum_i w_i.$$

In particular, the weight w is optimal if and only if

$$a_i^\top (\sum_i w_i a_i a_i^\top)^{-1} a_i = 1 \text{ if } w_i \neq 0,$$
$$a_i^\top (\sum_i w_i a_i a_i^\top)^{-1} a_i < 1 \text{ otherwises.}$$

Remark. The matrix $A^\top W A$ is unique, but the weight itself may not be unique. Geometrically, the first condition for the weight requires the ellipsoid touch the boundary of K for every active constraint a_i (i.e. $w_i \neq 0$) and the second condition (comes from $w_i \geq 0$) requires that the ellipsoid inside the body K.

Proof. Since $K = -K$, the center of $J(K)$ is 0. The John ellipsoid problem can be simplify to

$$\max_{G \succeq 0} \log\det G \quad \text{subjects to} \quad \|Ga_i\|_2 \leq 1 \text{ for all } i \in [m]$$

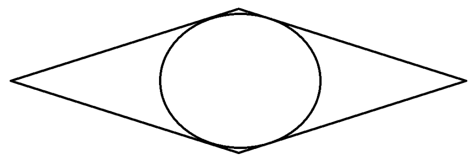

Figure 5.1: $J(K) \subset K \subset \sqrt{n} \cdot J(K)$

where the John ellipsoid is represented by $\{x : \left\|G^{-1}x\right\|_2 \leq 1\}$. Let $M = G^2$, we continue to simplify the equation as follows:

$$\begin{aligned}\max_{a_i^\top M a_i \leq 1, M \succeq 0} \log\det M &= \max_{M \succeq 0} \min_{w_i \geq 0} \log\det M + \sum_i w_i (1 - a_i^\top M a_i) \\ &= \min_{w_i \geq 0} \max_{M \succeq 0} \sum_i w_i + \log\det M - \mathrm{tr}(M \cdot \sum_i w_i a_i a_i^\top).\end{aligned}$$

Note that the optimality condition for $\max_{M \succeq 0} \log\det M - \mathrm{tr}(M \cdot C)$ is given by $M = C^{-1}$. Hence, we have

$$\begin{aligned}\max_{a_i^\top M a_i \leq 1, M \succeq 0} \log\det M &= \min_{w_i \geq 0} \sum_i w_i - \log\det \sum_i w_i a_i a_i^\top - n \\ &= n + \max_{w_i \geq 0} \log\det \sum_i w_i a_i a_i^\top - \sum w_i.\end{aligned}$$

The optimality condition for the later problem is given by

$$a_i^\top (\sum_i w_i a_i a_i^\top)^{-1} a_i = 1 - \lambda_i$$

where $\lambda_i w_i = 0$. For the total weight, we have that

$$\begin{aligned}\sum_{i=1}^m w_i &= \sum_{i=1}^m w_i a_i^\top (\sum_i w_i a_i a_i^\top)^{-1} a_i \\ &= \mathrm{tr}(WA(A^\top WA)^{-1} A^\top) \\ &= \mathrm{tr}((A^\top WA)^{-1} A^\top WA) \\ &= n\end{aligned}$$

where we used that K is full dimension and so is $A^\top WA$. $\square$

Now, we show that John ellipsoid is a $\sqrt{n}$ rounding of symmetric convex set.

Lemma 5.1.13. *Given a symmetric convex set K. We have that $J(K) \subset K \subset \sqrt{n} \cdot J(K)$.*

Proof. This lemma is easier to prove by a picture (see Figure 5.1) . Let $K = \{x \in \mathbb{R}^n : |(Ax)_i| \leq 1\}$. Lemma 5.1.12 shows that $J(E) = \{x : x^\top A^\top WAx \leq 1\}$ for some diagonal W with $\mathrm{tr}W = n$. For any $x \in K$, we have that

$$x^\top A^\top WAx \leq \sum_i w_i = n.$$

Hence, we have $K \subset \sqrt{n} J(K)$. $\square$

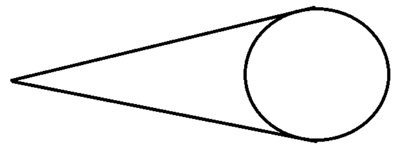

Figure 5.2: $J(K) \subset K \subset n \cdot J(K)$

5.1.3 Structure of John Ellipsoid for asymmetric body

Now, we study the property of John ellipsoid for asymmetric body.

Lemma 5.1.14. *Let $K = \{x \in \mathbb{R}^n : \ Ax \geq b\}$ a polytope with nonempty interior where $A \in \mathbb{R}^{m \times n}$. Let $V(x) = -\log \mathrm{vol}(K \cap (x - K))$. Then, we have that*

$$V(x) = C_n + \frac{1}{2} \max_{w_i \geq 0} \log \det \sum_i w_i \frac{a_i a_i^\top}{s_i^2(x)} - \sum w_i$$

where $s_i(x) = a_i^\top x - b_i$ and C_n is some constant depending on n. Furthermore, V is convex and that x is the minimizer of V if and only if

$$\sum_i w_i \frac{a_i}{s_i(x)} = 0.$$

Proof. Note that $K \cap (x - K) = \left\{y : \ \left|a_i^\top (y - x)\right| \leq s_i\right\}$. Using this and Lemma 5.1.12 gives the formula of $V(x)$. For the optimality condition for x, we calculate ∇V as follows:

$$\begin{aligned} DV(x)[h] &= \mathrm{tr}\left[(\sum_i w_i \frac{a_i a_i^\top}{s_i^2(x)})^{-1} (-2 \sum_i w_i \frac{a_i a_i^\top}{s_i^3(x)} a_i^\top h) \right] \\ &= -2 \sum_i \frac{w_i}{s_i(x)} (a_i^\top h) (\frac{a_i^\top}{s_i(x)} (\sum_i w_i \frac{a_i a_i^\top}{s_i^2(x)})^{-1} \frac{a_i}{s_i(x)}) \\ &= -2 \sum_i \frac{w_i}{s_i(x)} (a_i^\top h) \end{aligned}$$

where the last equation follows from the optimality condition $\frac{a_i^\top}{s_i(x)} (\sum_i w_i \frac{a_i a_i^\top}{s_i^2(x)})^{-1} \frac{a_i}{s_i(x)} = 1$ (Lemma 5.1.12). Hence, we have that

$$\nabla V(x) = -2 \sum_i \frac{w_i}{s_i(x)} a_i.$$

Hence, x is the minimizer of V if and only if the right-hand side is 0.

We skip the calculation that V is convex until we teach self-concordance barrier function. $\square$

Finally, we show the rounding property of John ellipsoid for asymmetric body.

Lemma 5.1.15. *Given a convex set K. We have that $J(K) \subset K \subset n \cdot J(K)$.*

Proof. Again, this lemma is easier to prove by a picture (see Figure 5.2) . But this proof sometimes is more convenient when we deal with approximate John ellipsoids.

Without loss of generality, the center of $J(K) = 0$ and that $s_i = 1$ for all i. Hence, $K = \{y : \ Ay \geq -1\}$. Fix $y \in K$. The optimality condition of the center shows that $\sum w_i a_i = 0$ and in particular,

$$\sum w_i (Ay)_i = 0.$$

Next, we note that $|Ay|_i \leq \sqrt{\sum w_i (Ay)_i^2}$ because that $J(K) \subset \{|Ay| \leq 1\}$. Combining all these, we have that

$$\begin{aligned} y^\top A^\top W A y &= \sum_i w_i (Ay)_i^2 \\ &= \sum_i w_i (Ay+1)_i^2 - \sum w_i \\ &\leq \left(\sum_i w_i (Ay+1)_i\right) \max |Ay+1|_i - \sum w_i \\ &= \sum_i w_i \max_i |Ay|_i \\ &\leq \sum w_i \cdot \sqrt{y^\top A^\top W A y}. \end{aligned}$$

This shows that $\sqrt{y^\top A^\top W A y} \leq \sum w_i = n$. □

5.1.4 Algorithms

The current fastest algorithm to compute John ellipsoid exactly is via semi-definite programming, and the faster SDP algorithm is based on the cutting plane methods and the dual reduction we mentioned before. This is hilarious if we plan to use John ellipsoid as a cutting plane method. (In the next lecture, we will see that this may not be a problem in real word.)

However, if we only need to have an ellipsoid with similar rounding property, then it is easier. If the convex set is a polytope, the running time to compute such ellipsoid is roughly the same as solving a linear program [31]. This is the best possible because finding John ellipsoid involves finding a point in the convex set and that is equivalent to solving a linear program if the convex set is a polytope.

For symmetric convex set, the problem becomes very easy. The current fastest algorithm is an unpublished result by Micheal Cohen and me. Recall that the weight of John ellipsoid satisfies the equation

$$a_i^\top (\sum_i w_i a_i a_i^\top)^{-1} a_i = 1.$$

The algorithm in simply the following:

1. Start with $w_i^{(1)} = \frac{n}{m}$ for all i.
2. For $k = 1, \cdots, T$, compute $w_i^{(k+1)} = w_i^{(k)} a_i^\top (\sum_i w_i^{(k)} a_i a_i^\top)^{-1} a_i$.
3. Output: $\frac{1}{T} \sum_{k=1}^{T} w_i^{(k)}$.

The key observation is the following:

Exercise 5.1.16. For any matrix $A \in \mathbb{R}^{m \times n}$, the function $\phi_i(w) = \log(a_i^\top (A^\top W A)^{-1} a_i)$ is convex.

Theorem 5.1.17. *We have that $\sum_{i=1}^m w_i = n$ and that $a_i^\top (\sum_i w_i a_i a_i^\top)^{-1} a_i \leq (\frac{m}{n})^{\frac{1}{T}}$ for all i. In particular, we have*

$$a_i^\top (\sum_i w_i a_i a_i^\top)^{-1} a_i \leq 1 + O(\varepsilon)$$

when $T \geq \frac{\log(\frac{m}{n})}{\varepsilon}$.

Proof. Let $\overline{v} = \frac{1}{T}\sum_{k=1}^{T} w_i^{(k)}$. The fact $\sum_{i=1}^{m} w_i = n$ follows from $\sum_{i=1}^{m} w_i^{(k)} a_i^\top (\sum_i w_i^{(k)} a_i a_i^\top)^{-1} a_i = n$. Note that

$$\begin{aligned}\log(a_i^\top(A^\top \overline{V} A)^{-1} a_i) &\leq \frac{1}{T}\sum_{k=1}^{T} \log(a_i^\top (A^\top V^{(k)} A)^{-1} a_i)\\ &= \frac{1}{T}\log(w_i^{(T+1)}/w_i^{(1)})\\ &\leq \frac{1}{T}\log(\frac{m}{n}).\end{aligned}$$

$\square$

Exercise 5.1.18. Prove that the ellipsoid in this algorithm gives a $\sqrt{n}(\frac{m}{n})^{\frac{1}{T}}$-rounding.

To implement this algorithm, one still need some dimension reduction techniques we will cover later. But for completeness, we include the algorithm here. Compare some existing solvers I can find online [34], this solver seems easier and faster especially for large scale problems.

Exercise 5.1.19. Make the program faster and compare it against other existing solvers [29, 34, 30, 32].

Listing 1: Approximate John ellipsoid for symmetric polytope

```
% Find the maximum volume ellipsoid for the symmetric body -1 < Ax < 1
% the output ellipsoid is {x: x' E x <= 1}
function [w,E] = mve_sym(A, iter)
m = size(A,1); n = size(A,2); JL_dim = 5;
w = ones(m,1) * (n/m);
w_sum = w;
for i = 1:iter
    B = spdiags(sqrt(w),0,m,m) * A;
    w = sum((B * ((B'*B)\(B' * randn(m,JL_dim)))).^2,2) / JL_dim;
    w_sum = w_sum + w;
end
w = w_sum/(iter+1);
E = A' * spdiags(w,0,m,m) * A;
```

Although this algorithm is only defined for symmetric body, there is an explicit reduction for computing John ellipsoid from asymmetric convex set to symmetric convex set [29] and it is natural to ask how good does the algorithm perform even for asymmetric body.

We end the lecture with this open problem.

Problem 5.1.20. Can you define a notation of approximate John ellipsoid that is easy to compute while satisfies a property similar to Theorem 5.1.4?

References

[29] Leonid G Khachiyan. Rounding of polytopes in the real number model of computation. *Mathematics of Operations Research*, 21(2):307–320, 1996.

[30] Piyush Kumar and E Alper Yildirim. Minimum-volume enclosing ellipsoids and core sets. *Journal of Optimization Theory and Applications*, 126(1):1–21, 2005.

[31] Yin Tat Lee and Aaron Sidford. Path finding methods for linear programming: Solving linear programs in o(sqrt(rank)) iterations and faster algorithms for maximum flow. In *Foundations of Computer Science (FOCS), 2014 IEEE 55th Annual Symposium on*, pages 424–433. IEEE, 2014.

[32] Yin Tat Lee and Aaron Sidford. Efficient inverse maintenance and faster algorithms for linear programming. In *Foundations of Computer Science (FOCS), 2015 IEEE 56th Annual Symposium on*, pages 230–249. IEEE, 2015.

[33] Yu Nesterov. Introductory lectures on convex programming volume i: Basic course. *Lecture notes*, 1998.

[34] Yin Zhang and Liyan Gao. On numerical solution of the maximum volume ellipsoid problem. *SIAM Journal on Optimization*, 14(1):53–76, 2003.