CPSC 536C: Algorithms for Convex Optimization	2025-26 Spring
Lecture: Efficient Convex Optimization with Membership Oracles	
Lecturer: Arqam Patel	Scribe: Yixian Huang

General Reference: Lee et al. [2018]

1 Motivation

The problem we are considering is to minimise a convex function over a convex set:

$$\min_{x \in K} f(x).$$

Classical cutting-plane methods require a **SEParation oracle** for K: given a point outside K, find a hyperplane that separates the point from K. But what if we only have a weaker **MEMbership oracle** that can decide whether a point belongs to K? This lecture shows how to build a separation oracle from a membership oracle, with some small residuals.

Given: Suppose we know a point $x_0 \in K$ and numbers $0 < r < R$ for the convex set K such that.

$$B(x_0, r) \subseteq K \subseteq B(x_0, R)$$

We have the access to oracle MEM(K) that determines if y is in K in the form:

MEM($y \in \mathbb{R}^n$, $\delta > 0$)
Output: with probability $1 - \delta$,

If $y \in B(K, \delta)$ return TRUE,

Else if $y \notin B(K, -\delta)$ return FALSE,

where $B(K, -\delta)$ and $B(K, \delta)$ are defined below.

Definition 1. *We define the "ball" around K to be the set of points at most δ distance from K:*

$$B_p(K, \delta) := \{x \in \mathbb{R}^n : \exists y \in K \ s.t. \ \|x - y\|_p \leq \delta\}.$$

By similar ideas, we also let

$$B_p(K, -\delta) := \{x \in \mathbb{R}^n : B_p(x, \delta) \subseteq K\}.$$

Goal: We want to implement a SEP(K, x) to find a separating hyperplane between K and x in the form:

SEP: Input: $y \in \mathbb{R}^n, \delta > 0$
Output: with prob $1 - \delta$:

If $y \in B(K, \delta)$ output TRUE

Else $c \in \mathbb{R}^n$ s.t. $\langle c, x\rangle \leq \langle c, y\rangle + \delta \quad \forall x \in B(K, -\delta)$

The output c defines a half-space that contains most of K and puts y on the other side. For simplicity, we write $\tilde{O}(f(n))$ for $O(f(n))$ up to polylogarithmic factors, i.e. $\tilde{O}(f(n)) = O(f(n) \log^k(n))$ for any $k \in \mathbb{N}$.

2 Main Results

We want to finally show the following result:

Theorem 2 (Lemma 13 in Lee et al. [2018])**.** *The Algorithm 1 bellow gives a half space containing K with probability $1-\delta$.*

Algorithm 1 Separate$_{\delta,P}$(K,x)

$B(x_0, r) \subset K \subset B(x_0, R)$.
if MEM(K,x) **then**
    return True
else if $x \notin B(0, R)$
    return $y : 0 \geq \langle y - x, x \rangle$
end if
Use the subgradient algorithm (Algorithm 2) to obtain an approximate subgradient $\tilde{g}$ of h_x at 0.
return the half-space

$$\left\{ y : \frac{50}{\delta} n^{7/6} R^{2/3} \varepsilon^{1/3} \kappa \geq \langle \tilde{g}, y - x \rangle \right\}$$

.

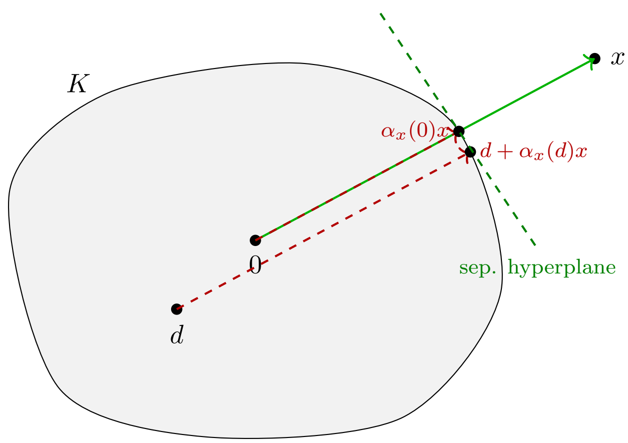

Figure 1: Intuition for the depth function h_x and the separating half-space.

Intuition The core construction is a one-dimensional "depth" function in the direction of x. For any point $d \in K$, let

$$\alpha_x(d) := \sup\{ \alpha \geq 0 \mid d + \alpha x \in K \}$$

be the largest step we can take from d in the direction x while staying inside K. Then define

$$h_x(d) := -\alpha_x(d) \, \|x\|_2,$$

a non-positive function on K that measures how "deep" d is from the boundary. Since h_x is convex and L-Lipschitz as shown below, a subgradient g of h_x at 0 is bounded and satisfies $\langle g, y \rangle \leq h_x(y) - h_x(0)$ for all y. Hence $h_x(y) \leq 0$ on K gives the inequality that separates x from K. However, it remains as a question that how we evaluate the gradient of h_x at 0 using membership oracles, which will be discussed in Lemma 5 next section.

In the following, we will first prove the convexity and Lipschitzness of $h_x(d)$ in Lemma 3 and 4, then give the intuition for the subgradient algorithm in Lemma 5, and finally show the main result in Section 4.

3 Technical Lemmas

Lemma 3 (Lemma 11 in Lee et al. [2018])**.** *$h_x(d)$ is a convex function on K.*

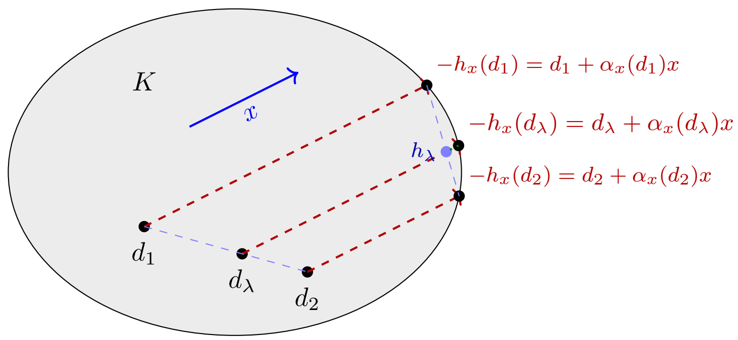

Figure 2: Graph for Lemma 3: The function $-h_x(d)$ is concave. d_λ is a linear combination of d_1 and d_2. Then h_λ, the linear combination of $h_x(d_1)$ and $h_x(d_2)$, is on the line segment between d_λ and $-h_x(d_\lambda)$.

Proof. Let $d_1, d_2 \in K$, and $\lambda \in [0,1]$. Since

$$d_1 + \alpha_x(d_1)x \in K,$$
$$d_2 + \alpha_x(d_2)x \in K,$$

We have

$$[\lambda d_1 + (1-\lambda)d_2] + [\lambda \alpha_x(d_1) + (1-\lambda)\alpha_x(d_2)]x \in K$$

Denote $d = (\lambda d_1 + (1-\lambda)d_2) \in K$ and $\beta = \lambda\alpha_x(d_1) + (1-\lambda)\alpha_x(d_2)$, then by the definition of α_x, $d + \beta x \in K$ yields

$$\alpha_x(d) = \sup\{\alpha : d + \alpha x \in K\} \geq \beta.$$

Hence the definition of d and β gives

$$\alpha_x(\lambda d_1 + (1-\lambda)d_2) \geq \lambda\alpha_x(d_1) + (1-\lambda)\alpha_x(d_2)$$

Multiplying $-\|x\|$ on both sides and using that $h_x(d) = -\|x\|\alpha_x(d)$, we have

$$h_x(\lambda d_1 + (1-\lambda)d_2) \leq \lambda h_x(d_1) + (1-\lambda)h_x(d_2),$$

which proves the convexity of h_x. ∎

Lemma 4 (Lemma 12 in Lee et al. [2018])**.** *h_x is $L = \frac{R+\delta}{r-\delta}$-Lipschitz on $B_2(0,\delta)$ for all $\delta < r$.*

Proof. Let $d_1, d_2 \in B(0,\delta)$. Without losing of generality, we assume $\alpha_x(d_1) \geq \alpha_x(d_2)$, then we have

$$|h_x(d_1) - h_x(d_2)| = \|x\|_2[\alpha_x(d_1) - \alpha_x(d_2)],$$

and we therefore want to show that

$$\|x\|_2[\alpha_x(d_1) - \alpha_x(d_2)] \leq L\|d_1 - d_2\|.$$

Case 1: $\|d_2 - d_1\|_2 \leq r - \delta$.

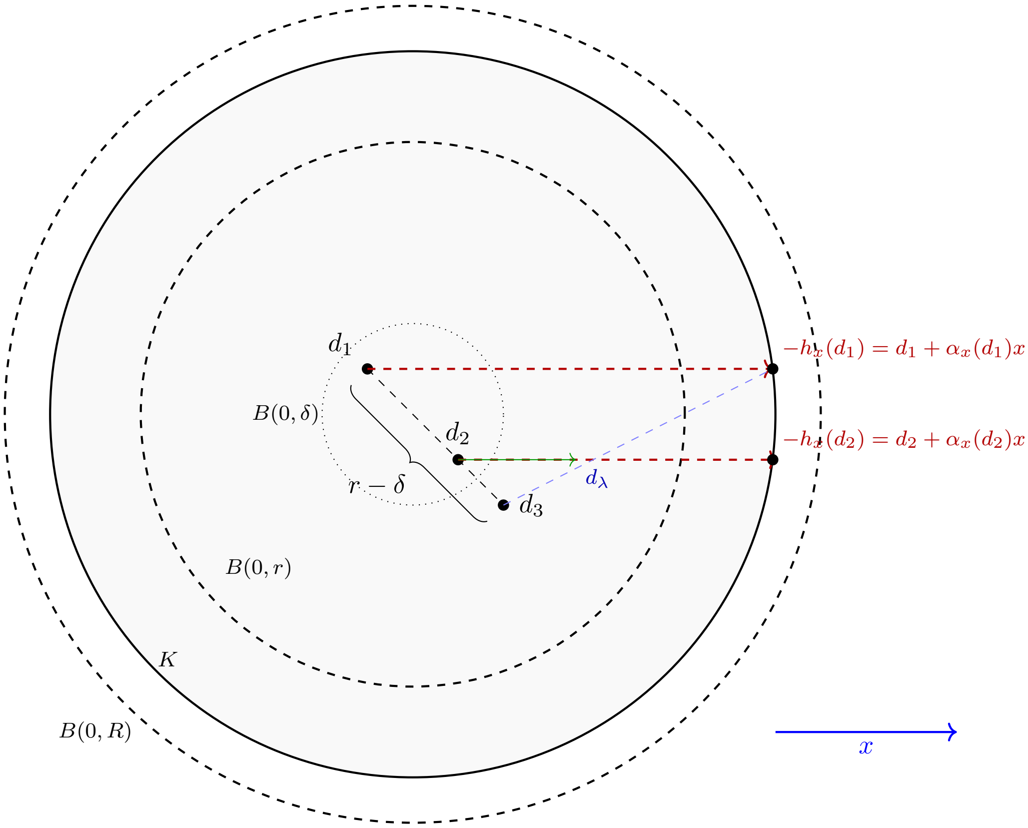

Figure 3: Lemma 4, Case 1: Construct d_3 by extending d_1d_2 to the length of $r-\delta$ and denote $\lambda = \|d_1d_2\|/(r-\delta)$. Suppose $(-h_x(d_1))d_3$ intersects $(-h_x(d_2))d_2$ at d_λ, then $d_2d_\lambda = (1-\lambda)\alpha_x(d_1)\|x\|$ by similar triangles. We only need to show that $\|x\|(\alpha_x(d_1)-\alpha_x(d_2)) \leq \lambda\alpha_x(d_1)\|x\|$, equivalently $(1-\lambda)\alpha_x(d_1)\|x\| \leq \alpha_x(d_2)\|x\|$, which could be concluded by $\|d_2d_\lambda\| \leq \|d_2(-h_x(d_2))$.

Let us consider $d_3 := d_1 + \frac{1}{\lambda}(d_2 - d_1)$ where $\lambda = \frac{\|d_2-d_1\|_2}{r-\delta} \in [0,1]$. Then we have

$$\lambda d_3 + (1-\lambda)[d_1 + \alpha_x(d_1)x] = d_2 + (1-\lambda)\alpha_x(d_1)x. \tag{3.1}$$

Besides, we also have

$$\|d_3\|_2 \leq \|d_1\|_2 + \frac{1}{\lambda}\|d_2 - d_1\|_2 \leq \delta + \frac{1}{\lambda}\|d_2 - d_1\|_2 = r,$$

where the first inequality comes from the Pythagorean Theorem, the second comes from $d_1 \in B_2(0,\delta)$, and the last comes from the definition of λ. Hence we have $d_3 \in B(0,r) \subseteq K$. Meanwhile, the definition of α_x yields $(d_1 + \alpha_x(d_1)x) \in K$. Thus, their convex combination

$$\lambda d_3 + (1-\lambda)[d_1 + \alpha_x(d_1)x] \in K.$$

Thus, by (3.1) and the definition of α_x, we get

$$\alpha_x(d_2) = \sup\{\alpha : d_2 + \alpha x\} \geq (1-\lambda)\alpha_x(d_1) = \left(1 - \frac{\|d_2 - d_1\|_2}{r-\delta}\right)\alpha_x(d_1).$$

Therefore, we have

$$\|x\|_2[\alpha_x(d_1) - \alpha_x(d_2)] \leq \frac{\|d_2 - d_1\|_2}{r-\delta}\alpha_x(d_1)\|x\|_2.$$

Now we need to control $\alpha_x(d_1)\|x\|_2$. $(d_1 + \alpha_x(d_1)x) \in K \subseteq B(0, R)$ and $\|d_1\| \le \delta$ give

$$\begin{aligned}\alpha_x(d_1)\|x\| &\le \|d_1 + \alpha_x(d_1)x - d_1\| \\ &\le \|d_1 + \alpha_x(d_1)x\| + \|d_1\| \\ &\le R + \delta,\end{aligned}$$

which concludes the result.

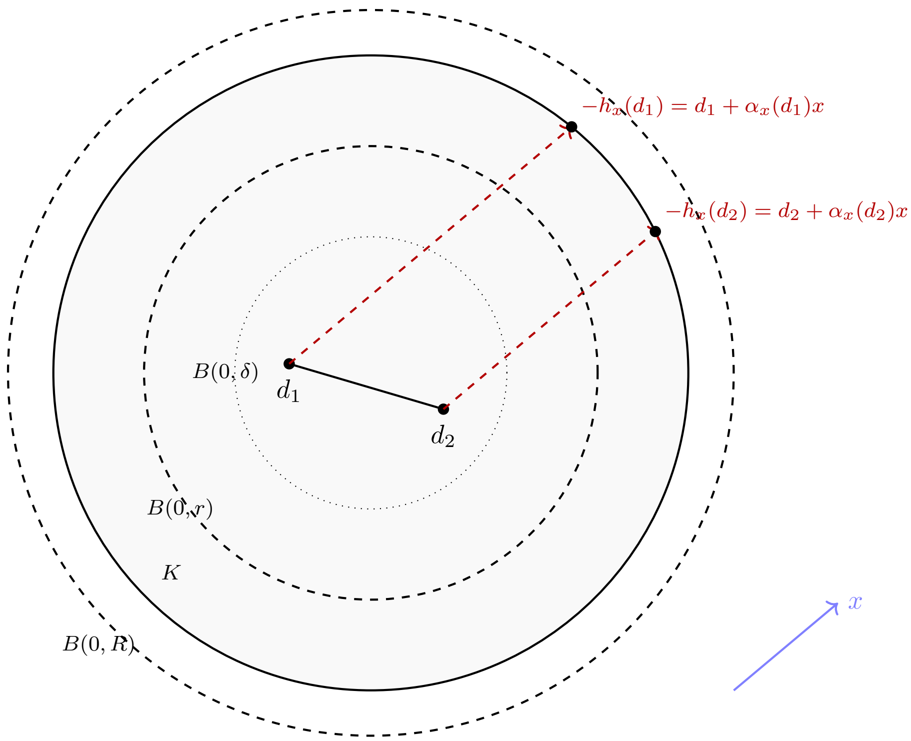

Figure 4: Lemma 4, Case 2: when $\|d_1 - d_2\| \ge r - \delta$, the Lipschitz bound follows directly from $\|h_x(d_1) - h_x(d_2)\| \le R + \delta$.

Case 2: $\|d_2 - d_1\| \ge r - \delta$. Since $d_1, d_2 \in B(0, \delta)$, the distance from d_i to $B(0, R)$ is less than or equals to $\delta + R$ for $i = 1, 2$. Hence we have $h_x(d_1), h_x(d_2) \in [-R - \delta, 0]$.

$$\begin{aligned}|h_x(d_1) - h_x(d_2)| &\le (R + \delta) \cdot 1 \\ &\le (R + \delta)\frac{\|d_2 - d_1\|_2}{r - \delta},\end{aligned}$$

which gives the result. ■

With the Lipschitzness of the target function, we can now approximate the subgradient near the desire point.

Lemma 5 (Lemma 10 in Lee et al. [2018])**.** *Given $r_1, L > 0$. Let f be a convex function on $B_\infty(x, 2r_1)$. Suppose that $\|\partial f(z)\|_\infty \le L$ for any $z \in B_\infty(x, 2r_1)$. Also, assume that we can compute function f with ε additive error with $\varepsilon \le r_1\sqrt{n}L$. Let $\tilde{g}$ be an approximate subgradient function of f at x given by Algorithm 2 defined below. Then, there is a random variable $\zeta \ge 0$ with $\mathbb{E}\zeta \le 3\sqrt{\frac{L\varepsilon}{r_1}}n^{5/4}$ such that*

$$f(q) \ge f(x) + \langle \tilde{g}, q - x\rangle - \zeta\|q - x\|_\infty - 4nr_1L \quad \textit{\textbf{for all}}\ q \in K.$$

Algorithm 2 `ApproxSubGrad`(f, x, r_1, ε) (Algorithm 2: SeparateConvexFunc in Lee et al. [2018])

Require: $r_1 > 0$, $\|\partial f(z)\|_\infty \le L$ for any $z \in B_\infty(x, 2r_1)$.
Set $r_2 = \sqrt{\frac{\varepsilon r_1}{\sqrt{n}L}}$.
Sample $y \in B_\infty(x, r_1)$ and $z \in B_\infty(y, r_2)$ independently and uniformly at random.
for $i = 1, 2, \cdots, n$ **do**
    Let α_i and β_i denote the end points of the interval $B_\infty(y, r_2) \cap \{z + se_i : s \in \mathbb{R}\}$.
    Set $\tilde{g}_i = \frac{f(\beta_i) - f(\alpha_i)}{2r_2}$ where we compute f with ε additive error.
end for

Output $\tilde{g}$ as the approximate subgradient of f at x.

Intuition We have a convex function f (in our case $f = h_x$) that is Lipschitz, and we can only obtain noisy function values by binary search on EVALuation oracles. The algorithm first picks a random point y near the origin (inside a small ball of radius $r_1, B(x, r_1)$), then approximates the partial gradients at y by taking differences on a segment in a direction of an infinite ball near y. More formally speaking:

We further randomly choose a point z in a smaller ball with radius $r_2, B(y, r_2)$. For each coordinate i, look at the line through z in direction e_i; the intersection of this line with the ball $B_\infty(y, r_2)$ is a segment $[\alpha_i, \beta_i]$. Estimate the i-th partial derivative as

$$\tilde{g}_i = \frac{f(\beta_i) - f(\alpha_i)}{2r_2},$$

using noisy function evaluations.

Omitting many technical details, the key outcome is:

$$f(q) \ \ge \ f(0) + \langle \tilde{g}, q \rangle - \zeta \|q\|_\infty - O(nr_1 L),$$

where ζ is a random variable with small mean. If we ignore the $O(\cdot)$ terms and the random ζ, we have the standard supporting hyperplane inequality for a convex function.

Sketch of Proof. By the convexity of f, we have for all q in K:

$$\begin{aligned} f(q) &\ge f(z) + \langle \nabla f(z), q - z \rangle && (3.2)\\ &= f(z) + \langle \tilde{g}, q - x \rangle + \langle \nabla f(z) - \tilde{g}, q - x \rangle + \langle \nabla f(z), x - z \rangle. && (3.3) \end{aligned}$$

Now we have $f(z) \ge f(x) - L\|z - x\|$ by L-Lipschitz, $\langle \nabla f(z), x - z \rangle$ lower bounded by $-\|\nabla f(z)\|_\infty \|x - z\|_1$ thus also $O(L\|z - x\|)$ by Lipschitz. They are both $O(nr_1 L)$ because z is selected from the small ball near x. Since $\langle \nabla f(z) - \tilde{g}, q - x \rangle \ge -\|\nabla f(z) - \tilde{g}\|_1 \|q - x\|_\infty$, it remains to show that the random variable $\|\nabla f(z) - \tilde{g}\|_1$ has a small mean, which is proven by Lemma 6 below. ■

Lemma 6. *For any $0 < r_2 \le r_1$ and twice differentiable convex function f defined on $B_\infty(x, r_1 + r_2)$ with $\|\nabla f(z)\|_\infty \le L$ for any $z \in B_\infty(x, r_1 + r_2)$ we have*

$$\mathbb{E}_{y \in B_\infty(x, r_1)} \mathbb{E}_{z \in B_\infty(y, r_2)} \|\nabla f(z) - g(y)\|_1 \le n^{3/2} \frac{r_2}{r_1} L$$

where $g(y)$ is the average of ∇f over $B_\infty(y, r_2)$.

Sketch of Proof. The idea is that $\|\nabla f(z) - \tilde{g}\|_1$ is bounded by the second derivative of f at z. By Stoke's Theorem, given the expectation integrates the second derivative, the term is controlled by integration of $\langle \nabla f, n \rangle$ over the boundary, where n is the normal vector. Therefore, $\mathbb{E}[\|\nabla f(z) - \tilde{g}\|_1]$ is bounded. ■

4 Proof of Main Results

Now we begin the proof of the main theorem.

Proof of Theorem 2. The proof of some technical lemmas are in the next section. When $x \notin B(0, R)$, Algorithm 1 will output SEParated for $B(0, R)$. Thus in the rest of our proof, we may assume $x \in B(0, R)$ but $x \notin B(K, -\varepsilon)$.

For simplicity, we denote $\kappa = \frac{R}{r}$. By Lemma 3, h_x is a convex function. Moreover, given $S = r/2$, Lemma 4 yields h_x is a L-Lipschitz function with

$$L = \frac{R + r/2}{r/2} < \frac{3R/2}{r/2} = 3\kappa.$$

Step 1: Approximate subgradient inequality. We want to apply Lemma 5 to the function $f = h_x$ over the domain K. Given the Lipschitz continuity, the lemma requires a small radius r_1 such that $B_\infty(0, 2r_1) \subset B_2(0, S)$. We will choose r_1 later (in terms of ε). The lemma then supplies a random vector $\tilde{g}$ and a random variable $\xi \geq 0$ with

$$\mathbb{E}[\xi] \;\leq\; 3\sqrt{\frac{L\varepsilon}{r_1}}\, n^{5/4},$$

such that for every $y \in K$,

$$h_x(y) \;\geq\; h_x(0) + \langle \tilde{g}, y \rangle - \xi \|y\|_\infty - Cnr_1L, \tag{4.1}$$

where C is a constant (the proof gives $C = 24$, but we keep the constant for clarity).

To approximate a seperating hyperplane, we need to show $\langle \tilde{g}, x \rangle \geq \langle \tilde{g}, y \rangle + C'$ for some constant C' independent of y. Given $h_x(y) \leq 0$, it remains to control $-h_x(0) - \langle \tilde{g}, x \rangle$.

Step 2: Lower bound $\langle \tilde{g}, x \rangle$ and $h_x(0)$. Since $\left\|\frac{x}{\kappa}\right\| = r \cdot \frac{\|x\|}{R} \leq r$, we have $\frac{-x}{\kappa} \in B(0, r) \subset K$ in the domain of h_x, therefore we can derive $h_x\left(\frac{-x}{\kappa}\right)$ from definition:

$$\begin{aligned}
h_x\left(\frac{-x}{\kappa}\right) &= -\|x\|_2 \sup\left\{\alpha : -\frac{x}{\kappa} + \alpha x \in K\right\} \\
&= -\|x\|_2 \sup\left\{\beta + \frac{1}{\kappa} : 0 + \beta x \in K\right\} \quad \textbf{by letting } \alpha - \frac{1}{\kappa} = \beta \\
&= -\|x\|_2 \left[\alpha_x(0) + \frac{1}{\kappa}\right] \\
&= h_x(0) - \frac{\|x\|_2}{\kappa}.
\end{aligned}$$

On the other hand, letting $y = -x/\kappa$ in (4.1) yields

$$h_x\left(\frac{-x}{\kappa}\right) \geq h_x(0) + \left\langle \tilde{g}, \frac{-x}{\kappa} \right\rangle - \frac{1}{\kappa}\xi\|x\|_\infty - 12nr_1\kappa$$

Thus we can get

$$\langle \tilde{g}, x \rangle \geq \|x\|_2 - \xi\|x\|_\infty - 12nr_1\kappa^2 \tag{4.2}$$

Moreover, by Lemma 7, we obtain

$$h_x(0) \geq -\|x\|_2 - \varepsilon\kappa. \tag{4.3}$$

Final Step: Derive separation inequality. Combining (4.1), (4.2), and (4.3) gives

$$\begin{aligned} h_x(y) &\geq \langle \tilde{g}, y-x\rangle - \xi\|y\|_\infty - \xi\|x\|_\infty - 12nr_1\kappa - 12nr_1\kappa^2 - \varepsilon\kappa \\ &\geq \langle \tilde{g}, y-x\rangle - 2\xi R - 24nr_1\kappa^2 - \varepsilon\kappa. \end{aligned}$$

for any $y \in K$. Then $h_x(y) \leq 0$ yields

$$\langle \tilde{g}, y-x\rangle \leq 2\xi R + 24nr_1\kappa^2 + \varepsilon\kappa,$$

where we define $\tilde{\xi}$ to be the right hand side. Thus, we have $\tilde{\xi} \geq \langle \tilde{g}, y-x\rangle$ where $\tilde{\xi}$ is a random variable independent of y. Now we can choose $r_1 = n^{1/6}R^{2/3}\varepsilon^{1/3}\kappa^{-1}$ and use $\mathbb{E}\xi \leq 3\sqrt{\frac{12\kappa\varepsilon}{r_1}}n^{5/4}$ in Lemma 5 to have

$$\begin{aligned} \mathbb{E}\tilde{\xi} &\leq 6\sqrt{\frac{12\kappa\varepsilon}{r_1}}n^{5/4}R + 24nr_1\kappa^2 + \varepsilon\kappa \\ &\leq 45n^{7/6}R^{2/3}\varepsilon^{1/3}\kappa + \varepsilon\kappa \\ &\leq 50n^{7/6}R^{2/3}\varepsilon^{1/3}\kappa. \end{aligned}$$

Then the Algorithm 1 would output the half-space

$$\left\{ y : \frac{50}{\delta}n^{7/6}R^{2/3}\varepsilon^{1/3}\kappa \geq \langle \tilde{g}, y-x\rangle \right\}$$

with probability, by Markov's inequality,

$$1 - \mathbb{P}\left(\langle \tilde{g}, y-x\rangle \geq \frac{50}{\delta}n^{7/6}R^{2/3}\varepsilon^{1/3}\kappa\right) \geq 1 - \mathbb{P}\left(\tilde{\xi} \geq \frac{50}{\delta}n^{7/6}R^{2/3}\varepsilon^{1/3}\kappa\right) \geq 1-\delta.$$

■

Lemma 7. *$h_x(0) \geq -\|x\|_2 - \varepsilon\kappa$.* [1]

Proof. For $h_x(0)$, we want to control $\alpha_x(0)$, i.e. we want to find some point out of K to cover $\alpha_x(0)$. In the following, we show that for all $x \notin B(K, -\varepsilon)$, $x/(1-\frac{\varepsilon}{r}) \notin K$.

Suppose for contradition, there exists $u \in K$ such that $\left(1-\frac{\varepsilon}{r}\right)u = x$. Then for any $\|z - (1-\frac{\varepsilon}{r})u\| \leq \varepsilon$, we denote $z := \left(1-\frac{\varepsilon}{r}\right)u + \frac{\varepsilon}{r}v$, where $\|v\| \leq r$. $B(0,r) \subset K$, we have

$$v \in B(0,r) \subset K,$$

that is to say, z is a convex combination of u and v, both in K. Hence z is in K too. As a result, we get

$$B(x,\varepsilon) = B\left(\left(1-\frac{\varepsilon}{r}\right)y, \varepsilon\right) \subset K.$$

which means $x \in B(K, -\varepsilon)$, contradition. Given $x/(1-\frac{\varepsilon}{r}) \notin K$, we have

$$0 < \alpha_x(0) \leq \left(1-\frac{\varepsilon}{r}\right).$$

Rearranging yields

$$1 - \frac{1}{\alpha_x(0)} \leq \frac{\varepsilon}{r}$$

Multiplying R on both sides gives

$$R\left(\frac{\alpha_x(0)-1}{\alpha_x(0)}\right) \leq \varepsilon\kappa.$$

[1] The authors made a typo when proving the bound but it doesn't affect the main result.

For all α such that $\alpha x \in K \subseteq B(0,R)$, given $\|\alpha\|\|x\| \leq R$, we have $\alpha_x(0)\|x\| = \|\alpha_x(0)\|\|x\| \leq R$ by the definition of $\alpha_x(0)$. Hence we can further extend the inequality as

$$\|x\| \left(\alpha_x(0) - 1\right) \leq \varepsilon\kappa.$$

Rearranging gives

$$h_x(0) = -\alpha_x(0)x \geq -\|x\|_2 - \varepsilon\kappa.$$

■

References

Yin Tat Lee, Aaron Sidford, and Santosh S Vempala. Efficient convex optimization with membership oracles. In *Conference On Learning Theory*, pages 1292–1294. PMLR, 2018. 1, 2, 3, 5, 6