CS 58500 – Theoretical Computer Science Toolkit

Lecture 12 (02/26)

Oracles and Reductions

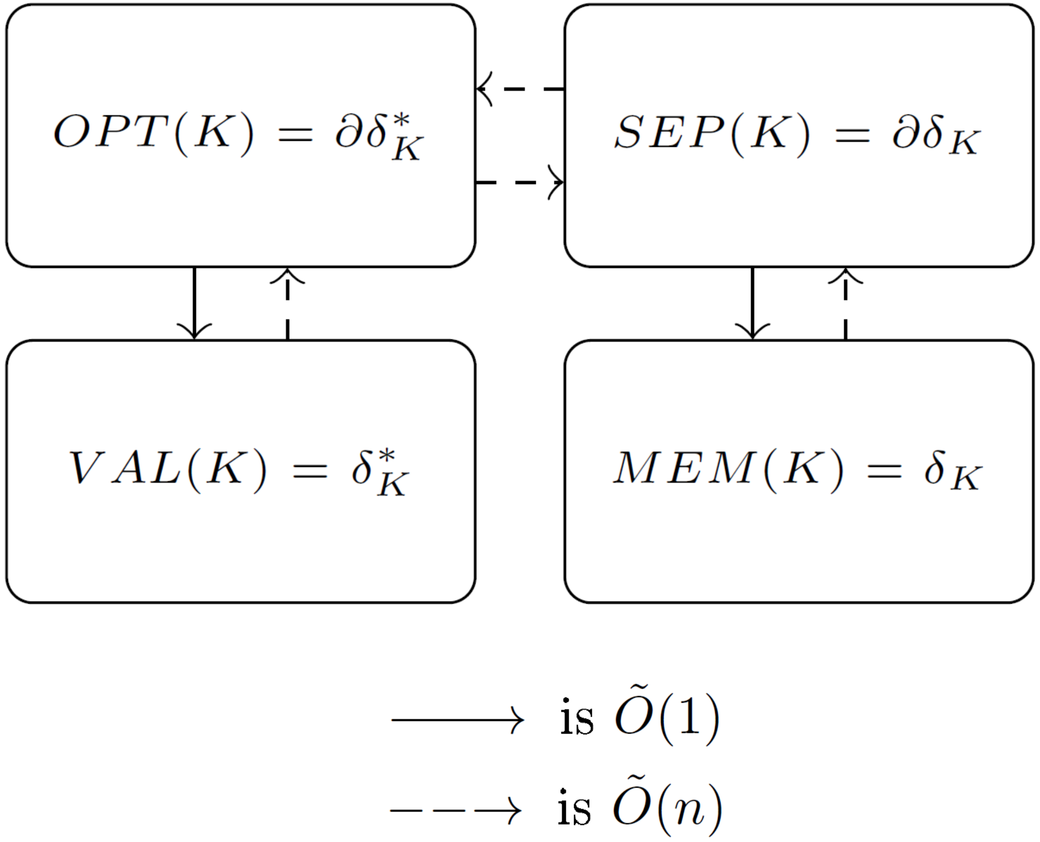
https://ruizhezhang.com/course_spring_2026.html

Today’s Lecture

- Oracles for Convex Sets
- Oracles for Convex Functions
- Separation for Convex Set

Oracles for Convex Sets

- Membership oracle (MEM):
  - **Input:** $\boldsymbol{y} \in \mathbb{R}^n$
  - **Output:** "$\boldsymbol{y} \in K$" or "$y \notin K$"
- Separation oracle (SEP):
  - **Input:** $\boldsymbol{y} \in \mathbb{R}^n$
  - **Output:** "$\boldsymbol{y} \in K$" or $\boldsymbol{c} \in \mathbb{R}^n$ s.t. $\langle \boldsymbol{c}, \boldsymbol{x} \rangle \leq \langle \boldsymbol{c}, \boldsymbol{y} \rangle \; \forall \boldsymbol{x} \in K$
- Validity oracle (VAL):
  - **Input:** $\boldsymbol{c} \in \mathbb{R}^n$
  - **Output:** "$K = \emptyset$" or $\max_{\boldsymbol{x} \in K} \langle \boldsymbol{c}, \boldsymbol{x} \rangle$
- Optimization oracle (OPT):
  - **Input:** $\boldsymbol{c} \in \mathbb{R}^n$
  - **Output:** "$K = \emptyset$" or $\boldsymbol{y} \in K$ s.t. $\langle \boldsymbol{c}, \boldsymbol{x} \rangle \leq \langle \boldsymbol{c}, \boldsymbol{y} \rangle \; \forall x \in K$



Oracles for Convex Sets

Membership oracle (MEM):

- **Input:** $\boldsymbol{y} \in \mathbb{R}^n$
- **Output:** "$\boldsymbol{y} \in K$" or "$y \notin K$"

Separation oracle (SEP):

- **Input:** $\boldsymbol{y} \in \mathbb{R}^n$
- **Output:** "$\boldsymbol{y} \in K$" or $\boldsymbol{c} \in \mathbb{R}^n$ s.t. $\langle \boldsymbol{c}, \boldsymbol{x} \rangle \leq \langle \boldsymbol{c}, \boldsymbol{y} \rangle \; \forall \boldsymbol{x} \in K$

Validity oracle (VAL):

- **Input:** $\boldsymbol{c} \in \mathbb{R}^n$
- **Output:** "$K = \emptyset$" or $\max_{\boldsymbol{x} \in K} \langle \boldsymbol{c}, \boldsymbol{x} \rangle$

Optimization oracle (OPT):

- **Input:** $\boldsymbol{c} \in \mathbb{R}^n$
- **Output:** "$K = \emptyset$" or $\boldsymbol{y} \in K$ s.t. $\langle \boldsymbol{c}, \boldsymbol{x} \rangle \leq \langle \boldsymbol{c}, \boldsymbol{y} \rangle \; \forall x \in K$

The complexities of these oracles depend on the problem

- If $K = \{\boldsymbol{Ax} \leq \boldsymbol{b}\}$, then SEP/MEM are easy while VAL/OPT need to solve a linear program
- If $K = \mathrm{conv}(\{\boldsymbol{a}_i\})$, then VAL/OPT are easy since we just need to check $\langle \boldsymbol{c}, \boldsymbol{a}_i \rangle$ for each i

Oracles for Convex Sets

Suppose there are n factories such that each factory can use resources $(r_1, \dots, r_k) \in \mathbb{R}_+^k$ to produce goods $(g_1, \dots, g_\ell) \in \mathbb{R}_+^\ell$. We use a vector $\boldsymbol{a} = (r_1, \dots, r_k, g_1, \dots, g_\ell) \in \mathbb{R}_+^{k+\ell}$ to represent it

Factory **Resources** **Goods**

For $\boldsymbol{a}_1, \dots, \boldsymbol{a}_n \in \mathbb{R}_+^{k+\ell}$, we define the Zonotope as

$$K := \left\{ \sum_{i=1}^{n} \lambda_i \boldsymbol{a}_i \in \mathbb{R}^{k+\ell} : 0 \le \lambda_i \le 1 \; \forall i \in [n] \right\}$$

- How to implement MEM, SEP, VAL, OPT for K?

Oracles for Convex Sets

For $\boldsymbol{a}_1, \dots, \boldsymbol{a}_n \in \mathbb{R}_+^{k+\ell}$, we define the Zonotope as

$$K := \left\{ \sum_{i=1}^{n} \lambda_i \boldsymbol{a}_i \in \mathbb{R}^{k+\ell} : 0 \leq \lambda_i \leq 1 \ \forall i \in [n] \right\}$$

- How to implement MEM, SEP, VAL, OPT for K?

Membership oracle (MEM):

- **Input:** $\boldsymbol{y} \in \mathbb{R}^n$
- **Output:** "$\boldsymbol{y} \in K$" or "$y \notin K$"

Separation oracle (SEP):

- **Input:** $\boldsymbol{y} \in \mathbb{R}^n$
- **Output:** "$\boldsymbol{y} \in K$" or $\boldsymbol{c} \in \mathbb{R}^n$ s.t. $\langle \boldsymbol{c}, \boldsymbol{x} \rangle \leq \langle \boldsymbol{c}, \boldsymbol{y} \rangle \ \forall \boldsymbol{x} \in K$

Validity oracle (VAL):

- **Input:** $\boldsymbol{c} \in \mathbb{R}^n$
- **Output:** "$K = \emptyset$" or $\max_{\boldsymbol{x} \in K} \langle \boldsymbol{c}, \boldsymbol{x} \rangle$

Optimization oracle (OPT):

- **Input:** $\boldsymbol{c} \in \mathbb{R}^n$
- **Output:** "$K = \emptyset$" or $\boldsymbol{y} \in K$ s.t. $\langle \boldsymbol{c}, \boldsymbol{x} \rangle \leq \langle \boldsymbol{c}, \boldsymbol{y} \rangle \ \forall x \in K$

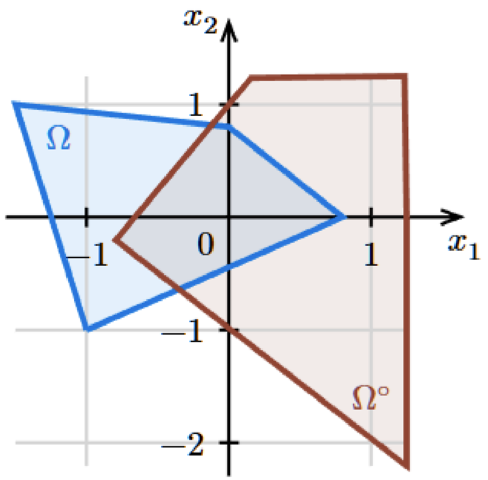
Detour: Duals of Convex Sets

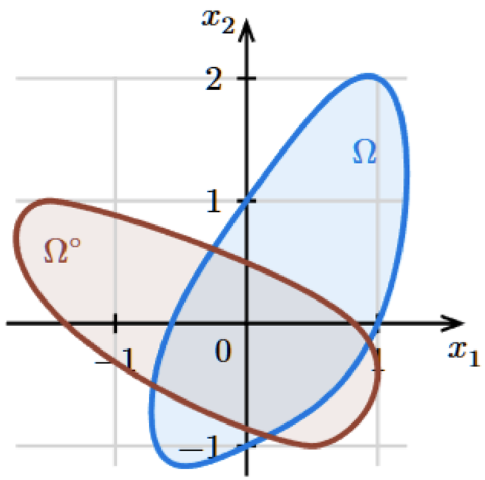
The polar of a convex set K is defined as

$$K^\circ := \{\boldsymbol{y} \in \mathbb{R}^n : \langle \boldsymbol{x}, \boldsymbol{y} \rangle \leq 1 \ \forall \boldsymbol{x} \in K\}$$

Theorem. Let $K \subseteq \mathbb{R}^n$ be convex and compact, and $B(0,r) \subseteq K \subseteq B(0,R)$. Then

- K° is convex and compact, and $B(0,1/R) \subseteq K^\circ \subseteq B(0,1/r)$
- $(K^\circ)^\circ = K$





Detour: Duals of Convex Sets

Lemma. Let K be a polytope with $\mathbf{0} \in \operatorname{int}(K)$

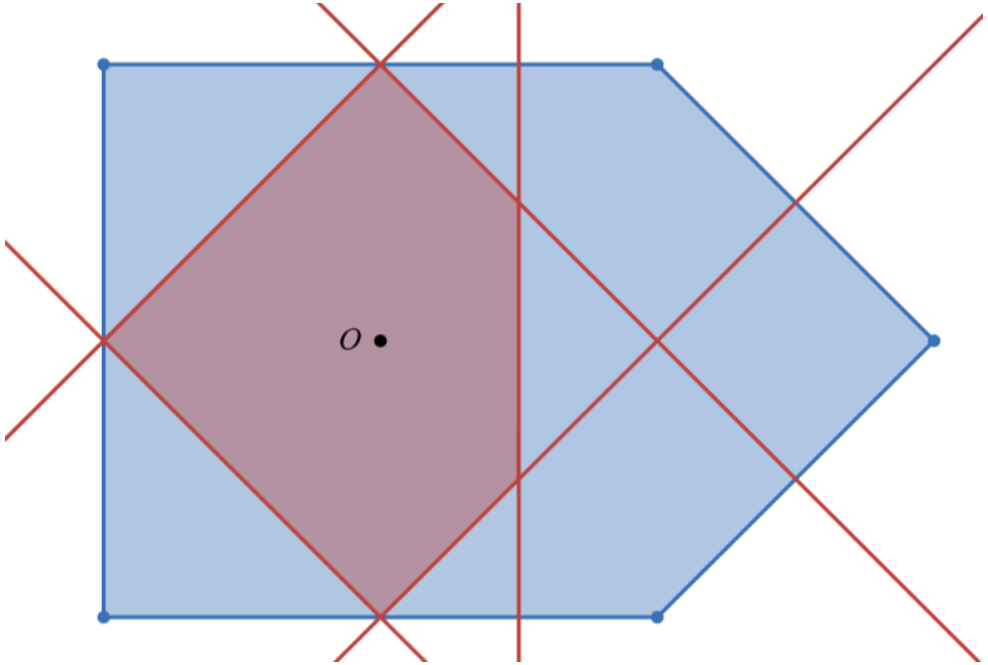
1. If $K = \operatorname{conv}(\{\boldsymbol{a}_1, \dots, \boldsymbol{a}_N\})$, then $K^\circ = \{\boldsymbol{y} : \langle \boldsymbol{a}_1, \boldsymbol{y} \rangle \leq 1, \dots, \langle \boldsymbol{a}_N, \boldsymbol{y} \rangle \leq 1\}$
2. If $K = \{\boldsymbol{x} : \langle \boldsymbol{a}_1, \boldsymbol{x} \rangle \leq 1, \dots, \langle \boldsymbol{a}_N, \boldsymbol{x} \rangle \leq 1\}$, then $K^\circ = \operatorname{conv}(\{\boldsymbol{a}_1, \dots, \boldsymbol{a}_N\})$

Proof of 2.

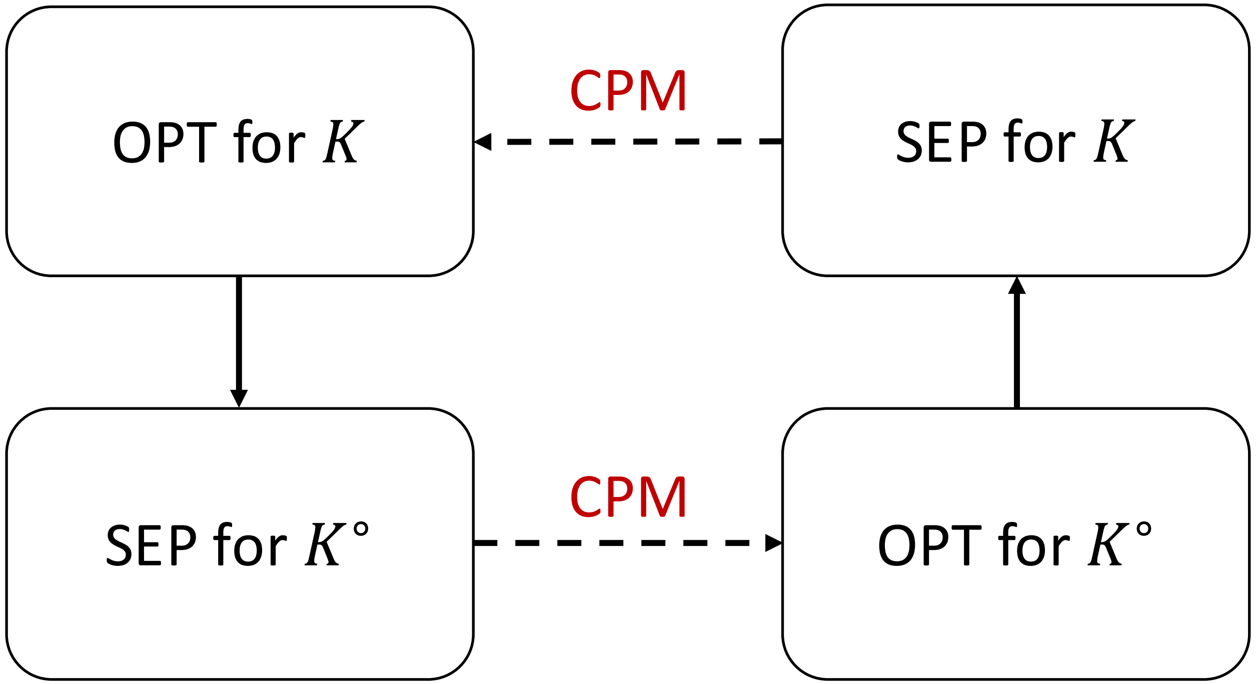
- Let $\boldsymbol{\lambda} \in \mathbb{R}^N_{\geq 0}$ with $\lambda_1 + \cdots + \lambda_N = 1$
- $\left\langle \sum_{i=1}^N \lambda_i \boldsymbol{a}_i , \boldsymbol{x} \right\rangle \leq 1$ for all $x \in K$
- Thus, $C \coloneqq \operatorname{conv}(\{\boldsymbol{a}_i\}) \subseteq K^\circ$
- Suppose $\boldsymbol{y} \notin C$. Then $\exists$ separating hyperplane:

$$\langle \boldsymbol{a}, \boldsymbol{g} \rangle < \beta < \langle \boldsymbol{y}, \boldsymbol{g} \rangle \;\; \forall \boldsymbol{a} \in C$$

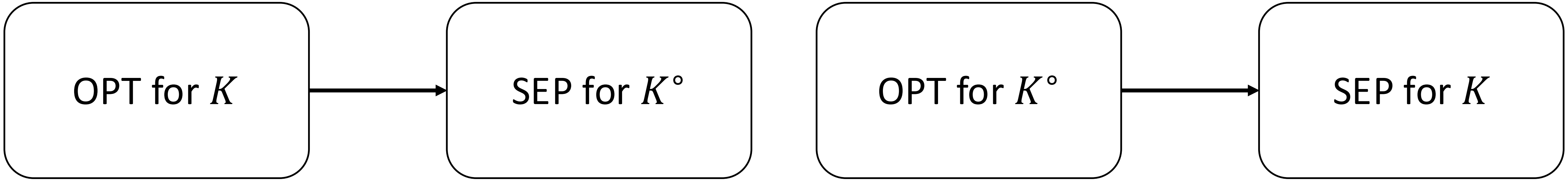
- We can scale $\boldsymbol{g}$ so that $\langle \boldsymbol{a}_i, \boldsymbol{g} \rangle < 1 < \langle \boldsymbol{y}, \boldsymbol{g} \rangle \;\; \forall i \in [N]$
- Thus, $\boldsymbol{g} \in K$ and $\boldsymbol{y} \notin K^\circ$



Oracles for Convex Sets



Oracles for Convex Sets



- For the input $\boldsymbol{y} \in \mathbb{R}^n$, query the optimization oracle OPT for K
- If $\max_{\boldsymbol{x} \in K} \langle \boldsymbol{y}, \boldsymbol{x} \rangle \leq 1$, then $\boldsymbol{y} \in K^\circ$
- Otherwise, let $\boldsymbol{x}^\star \coloneqq \arg\max_{\boldsymbol{x} \in K} \langle \boldsymbol{y}, \boldsymbol{x} \rangle$
- Then, return $\boldsymbol{x}^\star$ since $\langle \boldsymbol{x}^\star, \boldsymbol{w} \rangle \leq 1$ for all $\boldsymbol{w} \in K^\circ$ and $\langle \boldsymbol{x}^\star, \boldsymbol{y} \rangle > 1$
- Since $(K^\circ)^\circ = K$, we can also use OPT for K° to implement SEP for K

Today's Lecture

- Oracles for Convex Sets
- Oracles for Convex Functions
- Separation for Convex Set

Oracles for Convex Functions

- Evaluation oracle (EVAL):
  - **Input:** $\boldsymbol{y} \in \mathbb{R}^n$
  - **Output:** $f(\boldsymbol{y})$
- Subgradient oracle (GRAD):
  - **Input:** $\boldsymbol{y} \in \mathbb{R}^n$
  - **Output:** $f(\boldsymbol{y})$ and $\boldsymbol{g} \in \mathbb{R}^n$ such that $f(\boldsymbol{x}) \geq f(\boldsymbol{y}) + \langle \boldsymbol{g}, \boldsymbol{x} - \boldsymbol{y} \rangle \ \forall \boldsymbol{x} \in \mathbb{R}^n$

For any function f, its convex conjugate is defined as

$$f^*(\boldsymbol{y}) \coloneqq \sup_{\boldsymbol{x} \in \mathbb{R}^n} \langle \boldsymbol{x}, \boldsymbol{y} \rangle - f(\boldsymbol{x})$$

- $f^*(\boldsymbol{y})$ is convex
- $\delta_K^*(\boldsymbol{c}) = \sup_{\boldsymbol{x} \in K} \langle \boldsymbol{c}, \boldsymbol{x} \rangle$, i.e., VAL for K = EVAL for δ_K^*

Oracles for Convex Functions

Lemma. For any convex, closed, and proper f, its convex conjugate f^* is convex, closed, and proper. Moreover, for any $\boldsymbol{y} \in \mathbb{R}^n$, $\boldsymbol{x} \in \partial f^*(\boldsymbol{y})$ iff $x \in \arg\max_{\boldsymbol{x}} \langle \boldsymbol{y}, \boldsymbol{x} \rangle - f(\boldsymbol{x})$

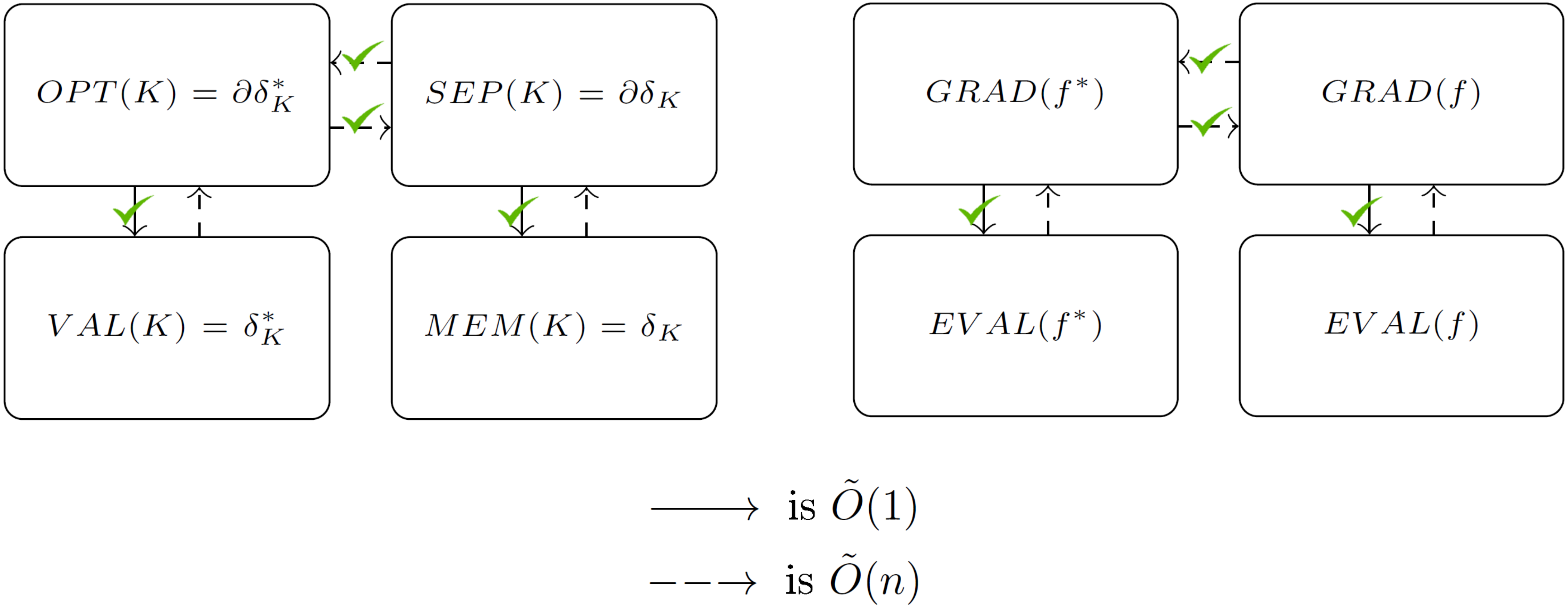
Proof sketch.

- Since $\boldsymbol{x} \in \partial f^*(\boldsymbol{y})$, $f^*(\boldsymbol{w}) \geq f^*(\boldsymbol{y}) + \langle \boldsymbol{x}, \boldsymbol{w} - \boldsymbol{y} \rangle$ for all $\boldsymbol{w} \in \mathbb{R}^n$
- $\langle \boldsymbol{y}, \boldsymbol{x} \rangle - f^*(\boldsymbol{y}) \geq \langle \boldsymbol{w}, \boldsymbol{x} \rangle - f^*(\boldsymbol{w})$ for all $\boldsymbol{w} \in \mathbb{R}^n$
- $\langle \boldsymbol{y}, \boldsymbol{x} \rangle - f^*(\boldsymbol{y}) \geq f^{**}(\boldsymbol{x}) = f(\boldsymbol{x})$
- That is, $f^*(\boldsymbol{y}) \leq \langle \boldsymbol{y}, \boldsymbol{x} \rangle - f(\boldsymbol{x})$, which implies that $\boldsymbol{x} \in \arg\max_{\boldsymbol{x}} \langle \boldsymbol{y}, \boldsymbol{x} \rangle - f(\boldsymbol{x})$

Fact. For closed, proper, and convex f, $f^{**} = f$

- ➢ $\partial \delta_K^*(\boldsymbol{c}) = \arg\max_{\boldsymbol{x} \in K} \langle \boldsymbol{c}, \boldsymbol{x} \rangle$, i.e., GRAD for $\delta_K^* =$ OPT for K
- ➢ GRAD for f^* can be implemented with $\tilde{O}(n)$ queries to GRAD for f (via cutting plane method), and vice versa

Oracles for Convex Sets and Convex Functions



Gradient Estimation

If f is in $\mathcal{C}^1$, then the gradient can be approximated by a finite difference with sufficiently small h:

$$\frac{\partial f}{\partial x_i} = \frac{f(\boldsymbol{x} + h\boldsymbol{e}_i) - f(\boldsymbol{x})}{h} + \mathcal{O}(h)$$

- And it takes $n + 1$ queries to EVAL to estimate $\nabla f(\boldsymbol{x})$
- However, convex function may not be differentiable

Lemma. For any L-Lipschitz, convex function $f: B_2^n \to \mathbb{R}$, $\nabla^2 f(\boldsymbol{x})$ exists almost everywhere and $\mathbb{E}_{\boldsymbol{x} \sim B_2^n}[\|\nabla^2 f(\boldsymbol{x})\|_F] \leq nL$

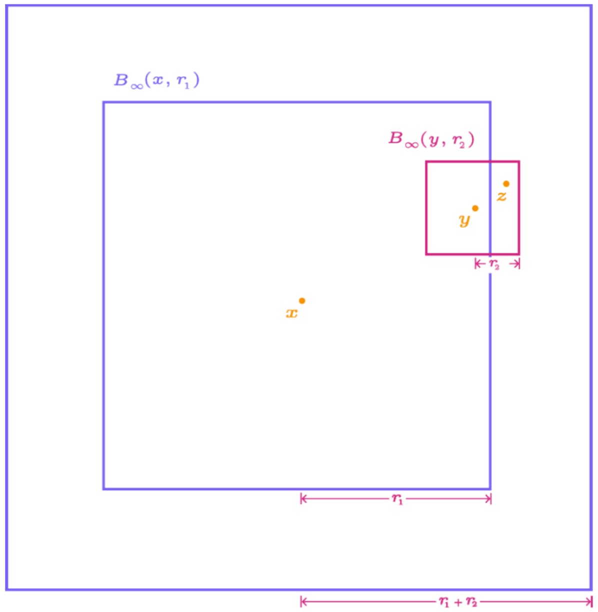
- Thus, we can evaluate gradient by randomly perturb $\boldsymbol{x}$ and apply the finite difference

Gradient Estimation

Lemma. Let $B_\infty(\boldsymbol{x}, r) \coloneqq \{\boldsymbol{y} : \|\boldsymbol{y} - \boldsymbol{x}\|_\infty \le r\}$. For any $0 < r_2 \le r_1$ and any $f : B_\infty(\boldsymbol{0}, r_1 + r_2) \to \mathbb{R}$, suppose $\|\nabla f(\boldsymbol{z})\|_\infty \le L$ for any $\boldsymbol{z}$. Then, we have

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

where $g(\boldsymbol{y}) = \mathbb{E}_{\boldsymbol{z} \sim B_\infty(\boldsymbol{y}, r_2)} [\nabla f(\boldsymbol{z})]$



Proof.

- Define a smoothed function h:

$$h(\boldsymbol{y}) \coloneqq (2r_2)^{-n} \cdot \left(f \star \mathbf{1}_{B_\infty(\mathbf{0},r_2)}\right)(\boldsymbol{y}) = \mathbb{E}_{\boldsymbol{z}\sim B_\infty(\boldsymbol{y},r_2)}[f(\boldsymbol{z})]$$

- We have $\nabla h(\boldsymbol{y}) = g(\boldsymbol{y})$ and $\|\nabla h(\boldsymbol{y})\|_\infty \le L$
- By the divergence theorem (or integration by parts),

$$\int_{B_\infty(\boldsymbol{x},r_1)} \Delta h(\boldsymbol{y})\mathrm{d}\boldsymbol{y} = \int_{\partial B_\infty(\boldsymbol{x},r_1)} \langle \nabla h(\boldsymbol{y}), \boldsymbol{n}(\boldsymbol{y})\rangle \mathrm{d}\boldsymbol{y}$$

- Thus, $\mathbb{E}_{\boldsymbol{y}\sim B_\infty(\boldsymbol{x},r_1)}[\Delta f(\boldsymbol{y})] \le (2r_1)^{-n} \cdot 2n(2r_1)^{n-1} \cdot L = \frac{nL}{r_1}$
- Define coordinate-wise deviations:

$$\omega_i(\boldsymbol{z}) \coloneqq \langle \nabla f(\boldsymbol{z}) - g(\boldsymbol{y}), \boldsymbol{e}_i\rangle \quad \forall i \in [n]$$

  We have $\mathbb{E}_{\boldsymbol{z}\sim B_\infty(\boldsymbol{y},r_2)}[\omega_i(\boldsymbol{z})] = 0$
- By the Poincaré inequality on the box,

$$\int_{B_\infty(\boldsymbol{y},r_2)} |\omega_i(\boldsymbol{z})|\mathrm{d}\boldsymbol{z} \le r_2 \int_{B_\infty(\boldsymbol{y},r_2)} \|\nabla \omega_i(\boldsymbol{z})\|_2 \mathrm{d}\boldsymbol{z} = r_2 \int_{B_\infty(\boldsymbol{y},r_2)} \|\nabla^2 f(\boldsymbol{z})\boldsymbol{e}_i\|_2 \mathrm{d}\boldsymbol{z}$$

Proof.

- Thus, $\mathbb{E}_{\boldsymbol{y}\sim B_\infty(\boldsymbol{x},r_1)}[\Delta f(\boldsymbol{y})] \leq (2r_1)^{-n} \cdot 2n(2r_1)^{n-1} \cdot L = \frac{nL}{r_1}$
- By the Poincaré inequality on the box,

$$\int_{B_\infty(\boldsymbol{y},r_2)} |\omega_i(\boldsymbol{z})| \mathrm{d}\boldsymbol{z} \leq r_2 \int_{B_\infty(\boldsymbol{y},r_2)} \|\nabla\omega_i(\boldsymbol{z})\|_2 \mathrm{d}\boldsymbol{z} = r_2 \int_{B_\infty(\boldsymbol{y},r_2)} \|\nabla^2 f(\boldsymbol{z})\boldsymbol{e}_i\|_2 \mathrm{d}\boldsymbol{z}$$

- Since f is convex, $\nabla^2 f \succcurlyeq 0$ and $\|\nabla^2 f(\boldsymbol{z})\|_F \leq \Delta f(\boldsymbol{z})$
- We get that

$$\int_{B_\infty(\boldsymbol{y},r_2)} \|\nabla f(\boldsymbol{z}) - g(\boldsymbol{y})\|_1 \mathrm{d}\boldsymbol{z} \leq \sqrt{n} r_2 \int_{B_\infty(\boldsymbol{y},r_2)} \Delta f(\boldsymbol{z}) \mathrm{d}\boldsymbol{z}$$

- Thus, we have

$$\mathbb{E}_{\boldsymbol{y}}\mathbb{E}_{\boldsymbol{z}}[\|\nabla f(\boldsymbol{z}) - g(\boldsymbol{y})\|_1] \leq n^{3/2} L \frac{r_2}{r_1}$$

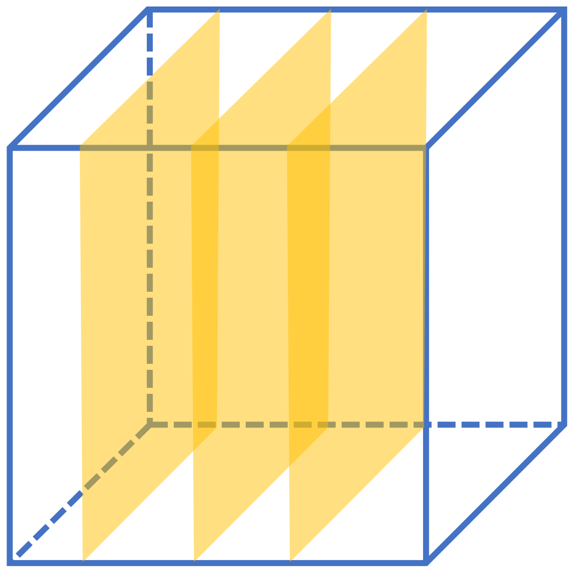
∎

Detour: L_1- Poincaré Inequality

Theorem. Let Ω be connected, bounded and open. Then the following inequality holds for any smooth function $f:\Omega \to \mathbb{R}$

$$\left\| f - \frac{1}{|\Omega|}\int_\Omega f(x)dx \right\|_{L^1(\Omega)} \leq \left(\sup_{S\subset\Omega} \frac{2|S||\Omega\backslash S|}{|\partial S||\Omega|} \right) \|\nabla f\|_{L^1(\Omega)}$$

- If $\Omega = B_\infty(0,R)$, then the L^1-Poincaré constant equal to R (dim-independent)



Gradient Estimation

Algorithm 10: SubgradConvexFunc(f, x, r_1, ε)

Require: $r_1 > 0$, $\|\partial f(z)\|_\infty \leq L$ 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.

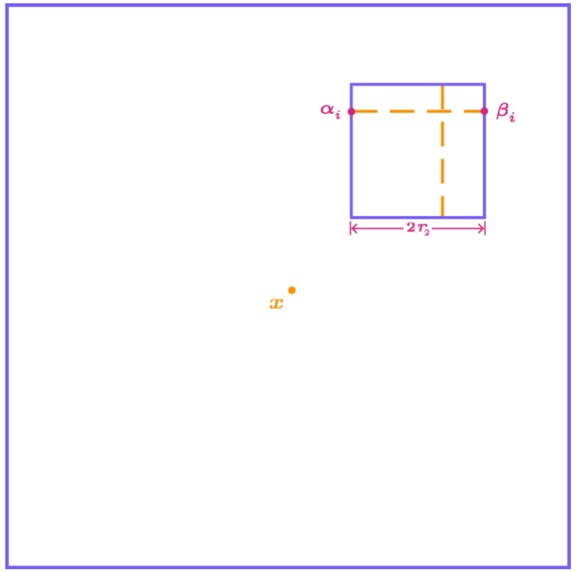
for $i = 1, 2, \cdots, n$ **do**

  Let α_i and β_i denote the end points of the interval $B_\infty(y, r_2) \cap \{z + s e_i : s \in \mathbb{R}\}$.

  Set $\tilde{g}_i = \frac{f(\beta_i) - f(\alpha_i)}{2r_2}$ where we compute f to within ε additive error.

end

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



Lemma (Approximate subgradient). Let $r_1 > 0$ and f be a convex function. Suppose that $\|\partial f(\boldsymbol{z})\|_\infty \leq L$ for any $\boldsymbol{z} \in B_\infty(\boldsymbol{x}, 2r_1)$ and suppose that we can evaluate f to within ϵ additive error for $\epsilon \leq r_1 \sqrt{n} L$. Let $\tilde{\boldsymbol{g}} =$ SubgradConvexFunc$(f, \boldsymbol{x}, r_1, \epsilon)$. Then,

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \langle \tilde{\boldsymbol{g}}, \boldsymbol{y} - \boldsymbol{x} \rangle - \zeta \|\boldsymbol{y} - \boldsymbol{x}\|_\infty - 4 n r_1 L$$

where $\zeta \geq 0$ is a random variable with $\mathbb{E}[\zeta] \leq 3 \sqrt{\frac{L\epsilon}{r_1}} n^{5/4}$

Gradient Estimation

Lemma (Approximate subgradient). Let $r_1 > 0$ and f be a convex function. Suppose that $\|\partial f(\boldsymbol{z})\|_\infty \le L$ for any $\boldsymbol{z} \in B_\infty(\boldsymbol{x}, 2r_1)$ and suppose that we can evaluate f to within ϵ additive error for $\epsilon \le r_1\sqrt{n}L$. Let $\widetilde{\boldsymbol{g}}$ $= \mathrm{SubgradConvexFunc}(f, \boldsymbol{x}, r_1, \epsilon)$. Then,

$$f(\boldsymbol{y}) \ge f(\boldsymbol{x}) + \langle \widetilde{\boldsymbol{g}}, \boldsymbol{y} - \boldsymbol{x} \rangle - \zeta\|\boldsymbol{y} - \boldsymbol{x}\|_\infty - 4nr_1L$$

where $\zeta \ge 0$ is a random variable with $\mathbb{E}[\zeta] \le 3\sqrt{\frac{L\epsilon}{r_1}}n^{5/4}$

Proof.

- Assume f is twice-differentiable and $\epsilon = 0$. Let $g(\boldsymbol{y}) := \mathbb{E}_{\boldsymbol{z}\sim B_\infty(\boldsymbol{y}, r_2)}[\nabla f(\boldsymbol{z})]$

$$\mathbb{E}_{\boldsymbol{z}\sim B_\infty(\boldsymbol{y}, r_2)}[|\widetilde{\boldsymbol{g}}_i - \boldsymbol{g}(\boldsymbol{y})_i|] = \mathbb{E}_{\boldsymbol{z}}\left[\left|\frac{f(\boldsymbol{\beta}_i) - f(\boldsymbol{\alpha}_i)}{2r_2} - g(\boldsymbol{y})_i\right|\right] \le \mathbb{E}_{\boldsymbol{z}}\left[\frac{1}{2r_2}\int\left|\frac{\partial f}{\partial x_i}(\boldsymbol{z} + s\boldsymbol{e}_i) - g(\boldsymbol{y})_i\right| \mathrm{d}s\right]$$

$$= \mathbb{E}_{\boldsymbol{z}}\left[\left|\frac{\partial f}{\partial x_i}(\boldsymbol{z}) - g(\boldsymbol{y})_i\right|\right] \qquad (\boldsymbol{z} + s\boldsymbol{e}_i \sim B_\infty(\boldsymbol{y}, r_2))$$

Gradient Estimation

Proof.

- Assume f is twice-differentiable and $\epsilon = 0$. Let $g(\boldsymbol{y}) \coloneqq \mathbb{E}_{\boldsymbol{z}\sim B_\infty(\boldsymbol{y},r_2)}[\nabla f(\boldsymbol{z})]$

$$\mathbb{E}_{\boldsymbol{z}\sim B_\infty(\boldsymbol{y},r_2)}[|\widetilde{\boldsymbol{g}}_i - \boldsymbol{g}(\boldsymbol{y})_i|] = \mathbb{E}_{\boldsymbol{z}}\left[\left|\frac{f(\boldsymbol{\beta}_i) - f(\boldsymbol{\alpha}_i)}{2r_2} - g(\boldsymbol{y})_i\right|\right] \le \mathbb{E}_{\boldsymbol{z}}\left[\frac{1}{2r_2}\int\left|\frac{\partial f}{\partial x_i}(\boldsymbol{z} + s\boldsymbol{e}_i) - g(\boldsymbol{y})_i\right| \mathrm{d}s\right]$$

$$= \mathbb{E}_{\boldsymbol{z}}\left[\left|\frac{\partial f}{\partial x_i}(\boldsymbol{z}) - g(\boldsymbol{y})_i\right|\right] \qquad (\boldsymbol{z} + s\boldsymbol{e}_i \sim B_\infty(\boldsymbol{y}, r_2))$$

- Thus, $\mathbb{E}_{\boldsymbol{z}}[\|\widetilde{\boldsymbol{g}} - \nabla f(\boldsymbol{z})\|_1] \le \mathbb{E}_{\boldsymbol{z}}[\|\widetilde{\boldsymbol{g}} - g(\boldsymbol{y})\|_1] + \mathbb{E}_{\boldsymbol{z}}[\|g(\boldsymbol{y}) - \nabla f(\boldsymbol{z})\|_1] \le 2\mathbb{E}_{\boldsymbol{z}}[\|\nabla f(\boldsymbol{z}) - g(\boldsymbol{y})\|_1]$
- The convexity of f gives that

$$\begin{aligned} f(\boldsymbol{y}) &\ge f(\boldsymbol{z}) + \langle \nabla f(\boldsymbol{z}), \boldsymbol{y} - \boldsymbol{z}\rangle = f(\boldsymbol{z}) + \langle \widetilde{\boldsymbol{g}}, \boldsymbol{y} - \boldsymbol{x}\rangle + \langle \nabla f(\boldsymbol{z}) - \widetilde{\boldsymbol{g}}, \boldsymbol{y} - \boldsymbol{x}\rangle + \langle \nabla f(\boldsymbol{z}), \boldsymbol{x} - \boldsymbol{z}\rangle \\ &\ge f(\boldsymbol{z}) + \langle \widetilde{\boldsymbol{g}}, \boldsymbol{y} - \boldsymbol{x}\rangle - \|\nabla f(\boldsymbol{z}) - \widetilde{\boldsymbol{g}}\|_1 \|\boldsymbol{y} - \boldsymbol{x}\|_\infty - \|\nabla f(\boldsymbol{z})\|_\infty \|\boldsymbol{x} - \boldsymbol{z}\|_1 \\ &\ge f(\boldsymbol{x}) + \langle \widetilde{\boldsymbol{g}}, \boldsymbol{y} - \boldsymbol{x}\rangle - \|\nabla f(\boldsymbol{z}) - \widetilde{\boldsymbol{g}}\|_1 \|\boldsymbol{y} - \boldsymbol{x}\|_\infty - 2L\|\boldsymbol{x} - \boldsymbol{z}\|_1 \end{aligned}$$

Gradient Estimation

Proof.

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \langle \widetilde{\boldsymbol{g}}, \boldsymbol{y} - \boldsymbol{x} \rangle - \|\nabla f(\boldsymbol{z}) - \widetilde{\boldsymbol{g}}\|_1 \|\boldsymbol{y} - \boldsymbol{x}\|_\infty - 2L\|\boldsymbol{x} - \boldsymbol{z}\|_1$$

- By definition, $\|\boldsymbol{x} - \boldsymbol{z}\|_1 \leq n\|\boldsymbol{x} - \boldsymbol{z}\|_\infty \leq n(r_1 + r_2) \leq 2nr_1$
- By the previous lemma,

$$\mathbb{E}_{\boldsymbol{y}}\mathbb{E}_{\boldsymbol{z}}[\|\nabla f(\boldsymbol{z}) - \widetilde{\boldsymbol{g}}\|_1] \leq 2\mathbb{E}_{\boldsymbol{y}}\mathbb{E}_{\boldsymbol{z}}[\|\nabla f(\boldsymbol{z}) - g(\boldsymbol{y})\|_1] \leq 2n^{3/2}L\frac{r_2}{r_1}$$

- Thus, $f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \langle \widetilde{\boldsymbol{g}}, \boldsymbol{y} - \boldsymbol{x} \rangle - \zeta\|\boldsymbol{y} - \boldsymbol{x}\|_\infty - 4nr_1L$
- If f can only be evaluated with ϵ additive error, then it introduces $\frac{\epsilon}{r_2}$ error to $\widetilde{\boldsymbol{g}}_i$, which gives

$$\mathbb{E}[\zeta] \leq 2n^{3/2}L\frac{r_2}{r_1} + \frac{\epsilon n}{r_2} \leq 3\sqrt{L\epsilon/r_1}\, n^{5/4}$$

■

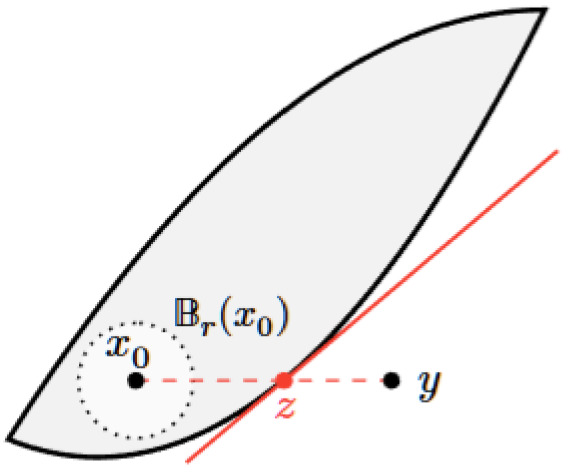
Today's Lecture

- Oracles for Convex Sets
- Oracles for Convex Functions
- Separation for Convex Set

Separation for Convex Set

Theorem (MEM to SEP, Lee-Sidford-Vempala '17). Let K be a convex body s.t. $B_2^n \subseteq K \subseteq RB_2^n$. Then, for any $0 < \eta \leq 1/2$, we can compute an η-approximate separation oracle for K using $\mathcal{O}(n \log(nR/\eta))$ queries to a membership oracle for K

- If $\|\boldsymbol{y}\|_2 > R$, then we just return the separating plane for RB_2^n:
$$\{\boldsymbol{x} : 0 \geq \langle \boldsymbol{x} - \boldsymbol{y}, \boldsymbol{y} \rangle\}$$
- Consider the segment from $\mathbf{0}$ to $\boldsymbol{y}$, and let $\boldsymbol{z}$ be the point on the boundary of K
- We will compute a supporting hyperplane of K at $\boldsymbol{z}$



Separation for Convex Set

We introduce the depth function:

$$d_{\boldsymbol{y}}(\boldsymbol{x}) \coloneqq \sup\left\{\alpha \geq 0 : \boldsymbol{x} + \alpha \frac{\boldsymbol{y}}{\|\boldsymbol{y}\|_2} \in K\right\}$$

- The maximum is guaranteed to exist since the set on the right is nonempty, closed, and bounded
- We will show that the subgradient of $d_{\boldsymbol{y}}$ can serve as an approximate separating hyperplane

Lemma.

- For any $\boldsymbol{x} \in K$, $d_{\boldsymbol{y}}(\boldsymbol{x}) \in [0,2R]$. We can compute $d_{\boldsymbol{y}}(\boldsymbol{x})$ up to δ additive error using $\mathcal{O}(\log(R/\delta))$ queries to a membership oracle for K (using binary search)
- $d_{\boldsymbol{y}}$ is concave
- $d_{\boldsymbol{y}}$ is $\left(\frac{R+\delta}{r-\delta}\right)$-Lipshitz on $B(0,\delta)$ for any $\delta < r$

Separation for Convex Set

- Recall the δ-neighborhood $K_\delta := \{\boldsymbol{x} \in \mathbb{R}^n : d(\boldsymbol{x}, K) \le \delta\}$
- We define $K_{-\delta} := \{\boldsymbol{x} \in K : \boldsymbol{x} + \delta B_2^n \subseteq K\}$

Consider the approximate MEM and SEP:

- Separation oracle ($\mathrm{SEP}_\delta(K)$):
  - **Input:** $\boldsymbol{y} \in \mathbb{R}^n$, $\delta > 0$
  - **Output:** "$\boldsymbol{y} \in K_\delta$" or unit $\boldsymbol{c} \in \mathbb{R}^n$ s.t. $\langle \boldsymbol{c}, \boldsymbol{x}\rangle \le \langle \boldsymbol{c}, \boldsymbol{y}\rangle + \delta \;\; \forall \boldsymbol{x} \in K_{-\delta}$
- Membership oracle ($\mathrm{MEM}_\delta(K)$):
  - **Input:** $\boldsymbol{y} \in \mathbb{R}^n$, $\delta > 0$
  - **Output:** "$\boldsymbol{y} \in K_\delta$" or "$\boldsymbol{y} \notin K_{-\delta}$"

Algorithm: $\mathrm{Separate}_{\epsilon,\rho}(K, \boldsymbol{y})$

- If $\boldsymbol{y} \in K_\epsilon$ or $\boldsymbol{y} \notin RB_2^n$

  $\dots$

- $\kappa := R/r$ and $r_1 := n^{1/6}\epsilon^{1/3}R^{2/3}\kappa^{-1}$
- Compute $\tilde{\boldsymbol{g}} = \mathrm{SubgradConvexFunc}(-d_{\boldsymbol{y}}, \boldsymbol{0}, r_1, 4\epsilon)$
- Output:

$$\left\{\boldsymbol{x} : \frac{50}{\rho} n^{7/6} R^{2/3} \kappa \epsilon^{1/3} \ge \langle \tilde{\boldsymbol{g}}, \boldsymbol{x} - \boldsymbol{y}\rangle\right\}$$

Separation for Convex Set

Lemma. Let K be a convex set with $rB_2^n \subset K \subset RB_2^n$. For any $\rho \in (0,1)$ and sufficiently small ϵ, with probability $1-\rho$, $\text{Separate}_{\epsilon,\rho}(K, \boldsymbol{y})$ outputs a halfspace that contains K

Proof.

- We know that $-d_{\boldsymbol{y}}$ is convex and 3κ-Lipschitz on $\left(\frac{r}{2}\right) B_2^n$
- For $\epsilon \le \mathcal{O}(R/n^2)$, $B_\infty(\mathbf{0}, 2r_1) \subset \left(\frac{r}{2}\right) B_2^n$
- Thus, SubgradConvexFunc ensures that

$$-d_{\boldsymbol{y}}(\boldsymbol{x}) \ge -d_{\boldsymbol{y}}(\mathbf{0}) + \langle \tilde{\boldsymbol{g}}, \boldsymbol{x} \rangle - \zeta \|\boldsymbol{x}\|_\infty - 12nr_1\kappa \qquad \forall \boldsymbol{x} \in K$$

- Note that $-\kappa^{-1}\boldsymbol{y} = -r\frac{\boldsymbol{y}}{R} \in rB_2^n \subset K$ and $d_{\boldsymbol{y}}(-\kappa^{-1}\boldsymbol{y}) = d_{\boldsymbol{y}}(\mathbf{0}) + \kappa^{-1}\|\boldsymbol{y}\|_2$

$$d_{\boldsymbol{y}}(\mathbf{0}) + \frac{1}{\kappa}\|\boldsymbol{y}\|_2 = d_{\boldsymbol{y}}\left(-\frac{\boldsymbol{y}}{\kappa}\right) \le d_{\boldsymbol{y}}(\mathbf{0}) - \left\langle \tilde{\boldsymbol{g}}, -\frac{\boldsymbol{y}}{\kappa} \right\rangle + \frac{\zeta}{\kappa}\|\boldsymbol{y}\|_\infty + 12nr_1\kappa$$

Separation for Convex Set

Lemma. Let K be a convex set with $rB_2^n \subset K \subset RB_2^n$. For any $\rho \in (0,1)$ and sufficiently small ϵ, with probability $1-\rho$, $\text{Separate}_{\epsilon,\rho}(K, \boldsymbol{y})$ outputs a halfspace that contains K

Proof.

- Thus, we get that
$$\langle \tilde{\boldsymbol{g}}, \boldsymbol{y} \rangle \geq \|\boldsymbol{y}\|_2 - \zeta \|\boldsymbol{y}\|_\infty - 12nr_1\kappa^2$$
- We have proved that $0 \leq d_{\boldsymbol{y}}(\boldsymbol{x}) \leq d_{\boldsymbol{y}}(\boldsymbol{0}) - \langle \tilde{\boldsymbol{g}}, \boldsymbol{x} \rangle + \zeta \|\boldsymbol{x}\|_\infty + 12nr_1\kappa^2$ for all $\boldsymbol{x} \in K$
- We claim that $\left(1 - \frac{\epsilon}{r}\right) K \subset K_{-\epsilon}$
  - For any $\boldsymbol{y} \in K$, let $\boldsymbol{z} \coloneqq \left(1 - \frac{\epsilon}{r}\right) \boldsymbol{y}$
  - For any $\boldsymbol{u} \in \epsilon B_2^n$, let $\boldsymbol{w} \coloneqq \frac{r}{\epsilon} \boldsymbol{u} \in rB_2^n \subset K$
  - Then, $\boldsymbol{z} + \boldsymbol{u} = \left(1 - \frac{\epsilon}{r}\right) \boldsymbol{y} + \frac{\epsilon}{r} \boldsymbol{w} \in K$ since K is convex

Separation for Convex Set

Lemma. Let K be a convex set with $rB_2^n \subset K \subset RB_2^n$. For any $\rho \in (0,1)$ and sufficiently small ϵ, with probability $1-\rho$, $\mathrm{Separate}_{\epsilon,\rho}(K, \boldsymbol{y})$ outputs a halfspace that contains K

Proof.

- Thus, we get that
$$\langle \widetilde{\boldsymbol{g}}, \boldsymbol{y} \rangle \geq \|\boldsymbol{y}\|_2 - \zeta \|\boldsymbol{y}\|_\infty - 12nr_1\kappa^2$$
- We claim that $\left(1-\frac{\epsilon}{r}\right)K \subset K_{-\epsilon}$
- Since $\boldsymbol{y} \notin K_{-\epsilon}$ by assumption, we have $\boldsymbol{y} \notin \left(1-\frac{\epsilon}{r}\right)K$
- By definition, $\left\{t\boldsymbol{y} : 0 \leq t \leq \frac{d_{\boldsymbol{y}}(\mathbf{0})}{\|\boldsymbol{y}\|_2}\right\} \subset K$ and in particular, $\left(1-\frac{\epsilon}{r}\right)\frac{d_{\boldsymbol{y}}(\mathbf{0})}{\|\boldsymbol{y}\|_2}\boldsymbol{y} \in \left(1-\frac{\epsilon}{r}\right)K$
- Thus, $\left(1-\frac{\epsilon}{r}\right)\frac{d_{\boldsymbol{y}}(\mathbf{0})}{\|\boldsymbol{y}\|_2} < 1 \implies \|\boldsymbol{y}\|_2 > d_{\boldsymbol{y}}(\mathbf{0}) - \frac{\epsilon}{r}d_{\boldsymbol{y}}(\mathbf{0}) \geq d_{\boldsymbol{y}}(\mathbf{0}) - \epsilon\kappa$ (since $K \subset RB_2^n$ and $\kappa = \frac{R}{r}$)

Separation for Convex Set

Lemma. Let K be a convex set with $rB_2^n \subset K \subset RB_2^n$. For any $\rho \in (0,1)$ and sufficiently small ϵ, with probability $1-\rho$, $\text{Separate}_{\epsilon,\rho}(K, \boldsymbol{y})$ outputs a halfspace that contains K

Proof.

- $0 \leq d_{\boldsymbol{y}}(\boldsymbol{0}) - \langle \widetilde{\boldsymbol{g}}, \boldsymbol{x} \rangle + \zeta \|\boldsymbol{x}\|_\infty + 12nr_1\kappa^2$
- $\langle \widetilde{\boldsymbol{g}}, \boldsymbol{y} \rangle \geq \|\boldsymbol{y}\|_2 - \zeta \|\boldsymbol{y}\|_\infty - 12nr_1\kappa^2$
- $d_{\boldsymbol{y}}(\boldsymbol{0}) \leq \|\boldsymbol{y}\|_2 + \epsilon\kappa$

$$\begin{aligned} 0 &\leq \|\boldsymbol{y}\|_2 + \epsilon\kappa - \langle \widetilde{\boldsymbol{g}}, \boldsymbol{x} \rangle + \zeta \|\boldsymbol{x}\|_\infty + 12nr_1\kappa^2 \\ &\leq \langle \widetilde{\boldsymbol{g}}, \boldsymbol{y} - \boldsymbol{x} \rangle + \zeta(\|\boldsymbol{x}\|_\infty + \|\boldsymbol{y}\|_\infty) + 24nr_1\kappa^2 + \epsilon\kappa \\ &\leq \langle \widetilde{\boldsymbol{g}}, \boldsymbol{y} - \boldsymbol{x} \rangle + 2\zeta R + 25nr_1\kappa^2 \end{aligned}$$

- We also have $\mathbb{E}[\zeta] \leq 3\sqrt{\frac{4L\epsilon}{r_1}} n^{5/4} = 6\sqrt{\frac{3\kappa\epsilon}{r_1}} n^{5/4}$

Separation for Convex Set

Lemma. Let K be a convex set with $rB_2^n \subset K \subset RB_2^n$. For any $\rho \in (0,1)$ and sufficiently small ϵ, with probability $1-\rho$, $\mathrm{Separate}_{\epsilon,\rho}(K, \boldsymbol{y})$ outputs a halfspace that contains K

Proof.

$$0 \le \langle \tilde{\boldsymbol{g}}, \boldsymbol{y} - \boldsymbol{x} \rangle + 2\zeta R + 25 n r_1 \kappa^2$$

- We also have $\mathbb{E}[\zeta] \le 3\sqrt{\frac{4L\epsilon}{r_1}}\, n^{5/4} = 6\sqrt{\frac{3\kappa\epsilon}{r_1}}\, n^{5/4}$
- Thus, for any $\boldsymbol{x} \in K$, $\langle \tilde{\boldsymbol{g}}, \boldsymbol{x} - \boldsymbol{y} \rangle \le \tilde{\zeta}$, where
$$\mathbb{E}[\tilde{\zeta}] \le 12\sqrt{3\kappa\epsilon / r_1}\, n^{5/4} R + 25 n r_1 \kappa^2 \le 50 n^{7/6} R^{2/3} \epsilon^{1/3} \kappa$$
if we take $r_1 = n^{1/6} R^{2/3} \epsilon^{1/3} / \kappa$
- By Markov inequality, with probability $1-\rho$, $\tilde{\zeta} \le 50 n^{7/6} R^{2/3} \epsilon^{1/3} \kappa / \rho$

■

Separation for Convex Set

Theorem. Let K be a convex set with $rB_2^n \subset K \subset RB_2^n$. For any $0 \leq \eta < 1/2$, we have that

$$\mathrm{SEP}_\eta(K) \leq \mathcal{O}(n \log(n\kappa R/\eta))\mathrm{MEM}_{(\eta/n\kappa R)^{\mathcal{O}(1)}}(K)$$

Proof.

- $h_y(x)$ can be computed within δ error using $\mathcal{O}(\log(R/\delta))$ queries to the membership oracle
- We have shown that $\widetilde{\boldsymbol{g}} = \mathrm{Separate}_{\delta,\rho}(K, \boldsymbol{y})$ is a separating hyperplane, except $\|\widetilde{\boldsymbol{g}}\|_2 \neq 1$
- We know that $\langle \widetilde{\boldsymbol{g}}, \boldsymbol{y} \rangle \geq \|\boldsymbol{y}\|_2 - \zeta\|\boldsymbol{y}\|_\infty - 12nr_1\kappa^2$. We claim that $\|\widetilde{\boldsymbol{g}}\|_2 \geq 1/(4\kappa)$
  - $\boldsymbol{y} \notin K_{-\delta} \implies \|\boldsymbol{y}\|_2 \geq r - \delta \geq r/2$
  - $12nr_1\kappa^2 = 12n^{7/6}R^{2/3}\delta^{1/3}\kappa$
  - $\zeta\|\boldsymbol{y}\|_\infty \leq \frac{6\sqrt{3}}{\rho}n^{7/6}R^{2/3}\delta^{1/3}\kappa$
  - If we take $\delta \leq \left(\frac{\rho R^{1/3}}{100n^{7/6}\kappa^2}\right)^3$, then $\langle \widetilde{\boldsymbol{g}}, \boldsymbol{y} \rangle \geq r/2 - r/4 = r/4$, and thus $\|\widetilde{\boldsymbol{g}}\|_2 \geq \frac{r}{4\|\boldsymbol{y}\|_2} \geq \frac{1}{4\kappa}$

Separation for Convex Set

Theorem. Let K be a convex set with $rB_2^n \subset K \subset RB_2^n$. For any $0 \le \eta < 1/2$, we have that

$$\mathrm{SEP}_\eta(K) \le \mathcal{O}(n \log(n\kappa R/\eta))\mathrm{MEM}_{(\eta/n\kappa R)^{\mathcal{O}(1)}}(K)$$

Proof.

- $h_{\boldsymbol{y}}(x)$ can be computed within δ error using $\mathcal{O}(\log(R/\delta))$ queries to $\mathrm{MEM}_\delta(K)$
- We have shown that $\widetilde{\boldsymbol{g}} = \mathrm{Separate}_{\delta,\rho}(K, \boldsymbol{y})$ is a separating hyperplane, except $\|\widetilde{\boldsymbol{g}}\|_2 \neq 1$
- We know that $\langle \widetilde{\boldsymbol{g}}, \boldsymbol{y} \rangle \ge \|\boldsymbol{y}\|_2 - \zeta \|\boldsymbol{y}\|_\infty - 12nr_1\kappa^2$. We claim that $\|\widetilde{\boldsymbol{g}}\|_2 \ge 1/(4\kappa)$
- After re-scaling $\tilde{g}$, we obtain a separation oracle with error $200n^{7/6}R^{2/3}\delta^{1/3}\kappa^2/\rho$ and failure probability $\mathcal{O}(n \log(R/\delta)\,\delta + \rho)$
- Thus, for $\mathrm{SEP}_\eta(K)$, we need to choose δ and ρ such that

$$\eta \le \max\{200n^{7/6}R^{2/3}\delta^{1/3}\kappa^2/\rho, \mathcal{O}(n \log(R/\delta)\,\delta + \rho)\}, \qquad \delta \le \left(\frac{\rho R^{1/3}}{100n^{7/6}\kappa^2}\right)^3 \le \mathcal{O}\left(\frac{R}{n^2}\right)$$

Separation for Convex Set

Theorem. Let K be a convex set with $rB_2^n \subset K \subset RB_2^n$. For any $0 \leq \eta < 1/2$, we have that

$$\mathrm{SEP}_\eta(K) \leq \mathcal{O}(n \log(n\kappa R/\eta))\mathrm{MEM}_{(\eta/n\kappa R)^{\mathcal{O}(1)}}(K)$$

Proof.

- Thus, for $\mathrm{SEP}_\eta(K)$, we need to choose δ and ρ such that

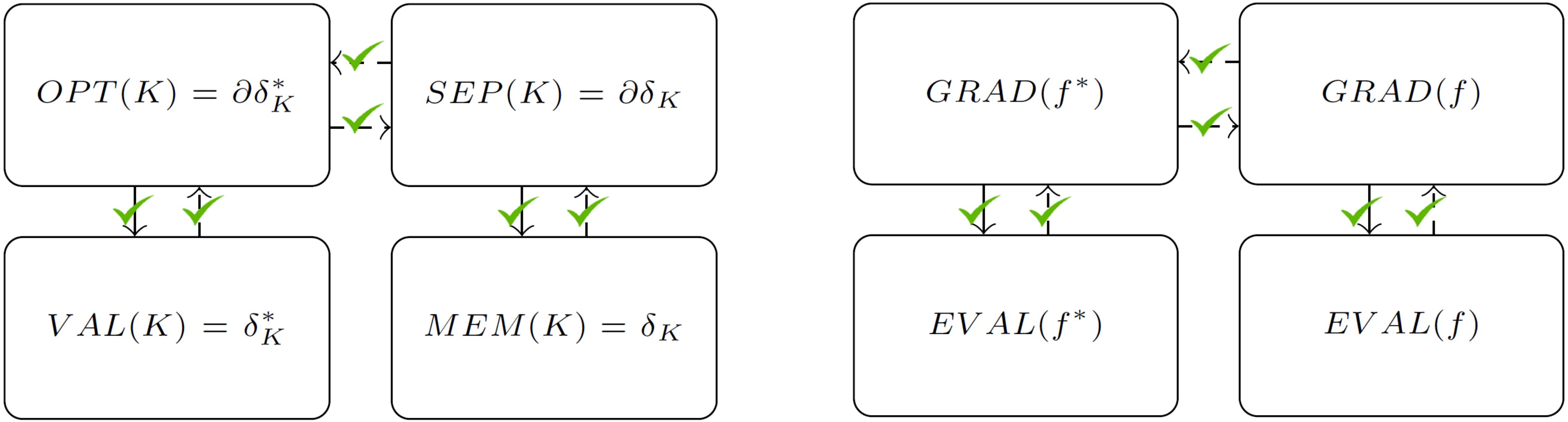
$$\eta \leq \max\{200n^{7/6}R^{2/3}\delta^{1/3}\kappa^2/\rho\,, \mathcal{O}(n\log(R/\delta)\,\delta + \rho)\}, \qquad \delta \leq \left(\frac{\rho R^{1/3}}{100n^{7/6}\kappa^2}\right)^3$$

- It suffices to choose

$$\rho = \sqrt{n^{7/6}R^{2/3}\kappa^2\delta^{1/3}}, \qquad \delta = \Theta\left(\frac{\eta^6}{n^{7/2}\kappa^6 R^2}\right)$$

■

Oracles for Convex Sets and Convex Functions



$\longrightarrow$ is $\tilde{O}(1)$

$-\,-\!\!\rightarrow$ is $\tilde{O}(n)$