Lower Bounds for Global Optimization

Problem

$$\min_{x \in B} f(x)$$

where $B = \{ x \in \mathbb{R}^n \mid \|x\| \le R \}$

- $\|\cdot\|$ can arbitrary
- $R > 0$

Assume that f is Lipschitz continuous:

$$|f(x) - f(y)| \le L\|x-y\|, \quad \forall x, y \in B$$

- $L > 0$

Goal: Find $\bar{x} \in B$: $f(\bar{x}) - f^* \le \varepsilon$

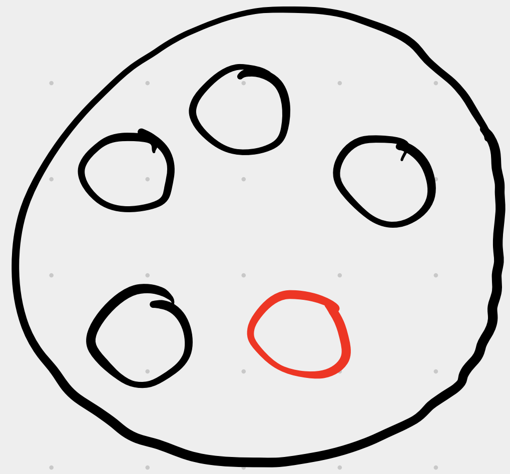
Zeroth-order algorithms: $O(x) = \{f(x)\}$

Question: What is a lower bound for complexity of zeroth-order methods?

Packing

B – a ball with radius $R>0$

$x_1, \dots, x_K \in B$ and $0<r<R$.



Denote "small balls":

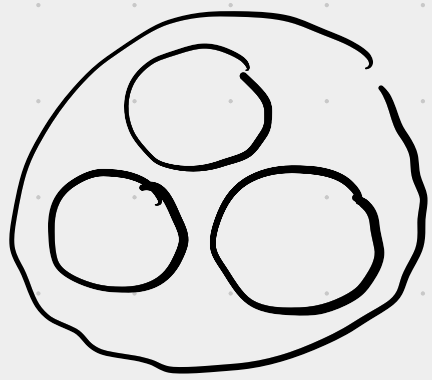
$$b_i = \{ y \in \mathbb{R}^n \mid \|y-x_i\| \le r \}$$

Def. We say $\{b_1, \dots, b_K\}$ is packing in B:

(1) Each $b_i \subset B$

(2) For any b_i, b_j: $\operatorname{int} b_i \cap \operatorname{int} b_j = \emptyset$.

Def. We say $\{b_1, \dots, b_K\}$ is a maximal packing in B if we cannot extend it.

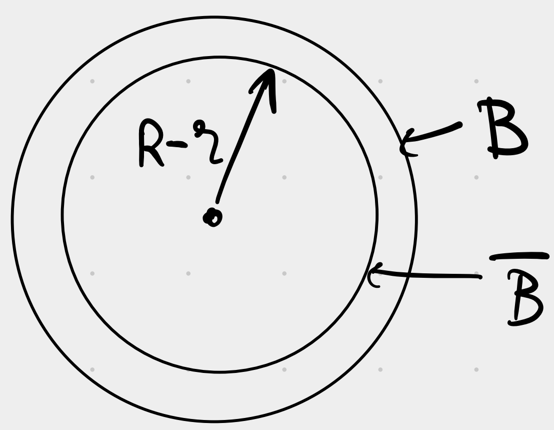


Theorem 1. Let $\{b_1, \dots, b_K\}$ – a maximal packing. Then

$$K \ge \left(\frac{R-r}{2r}\right)^n \approx \left(\frac{R}{r}\right)^n$$

Proof.

① Consider shrinked ball $\bar{B} = \{ y \in \mathbb{R}^n \mid \|y\| \le R-r \}$



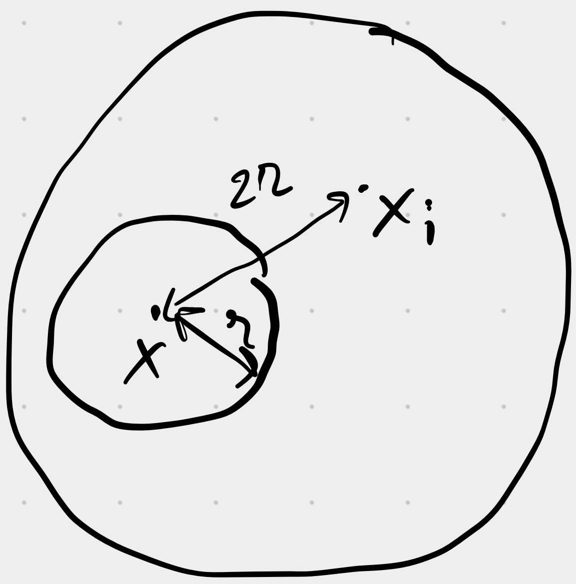
② For any $x \in \overline{B}$ $\exists$ x_i, $1 \leq i \leq K$ s.t.

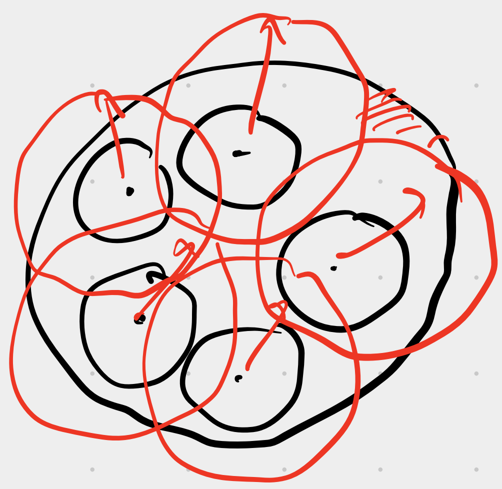
$$\|x - x_i\| < 2r.$$

Otherwise, it means we can add to our packing

$$\{y \in \mathbb{R}^n \mid \|y - x\| \leq r\}$$

(?!) the packing is maximal.





③ We got a covering

$$c_i = \{y \in \mathbb{R}^n : \|y - x_i\| \leq 2r\} \Rightarrow$$

$$\overline{B} \subset \bigcup_{i=1}^{K} c_i$$

④ Take the volume:

$$\mathrm{Vol}(\overline{B}) \leq \mathrm{vol}\left(\bigcup_{i=1}^{K} c_i\right) \leq \sum_{i=1}^{K} \mathrm{vol}(c_i)$$

$$= K \cdot \mathrm{vol}(c_1)$$

Denote $\alpha = \mathrm{Vol}(\{y \in \mathbb{R}^n : \|y\| \leq 1\})$.

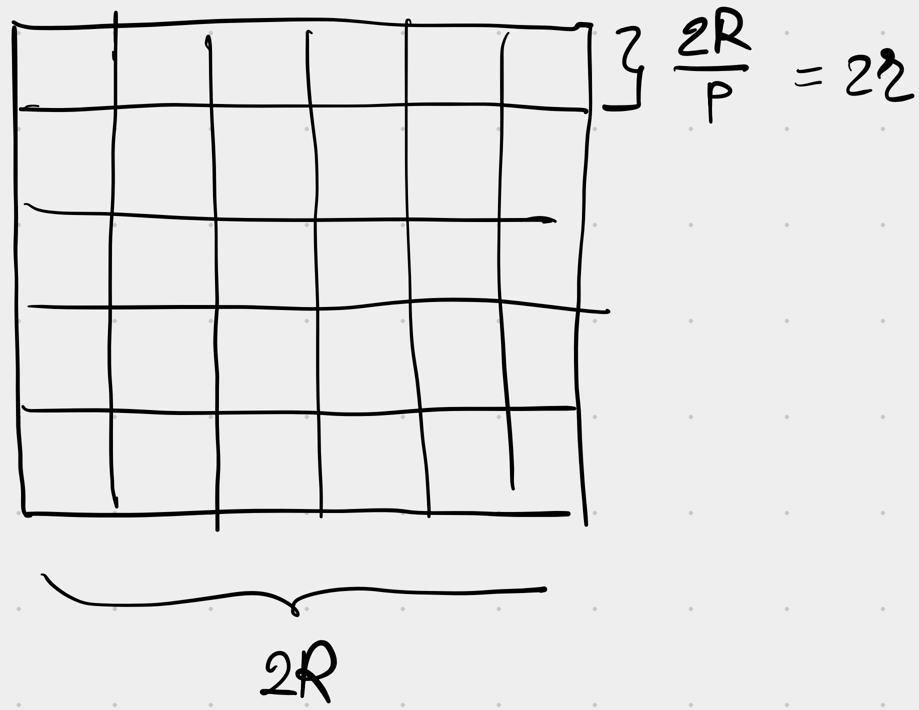
Then: $\mathrm{vol}(c_1) = \alpha \cdot (2r)^n$

$\mathrm{vol}(\overline{B}) = \alpha \cdot (R - r)^n$. $\Rightarrow K \geq \dfrac{(R-r)^n}{(2r)^n}$. $\square$

Remark. $\|\cdot\| = \|\cdot\|_\infty$

$p \geq 1$

$$K = p^n = \left(\frac{R}{r}\right)^n.$$



Resisting Oracle

An optimization algorithm:

$$x_{k+1} = \mathcal{A}\left(O(x_0), \dots, O(x_k)\right).$$

Resisting oracle $O(\cdot)$

1. Should return information that is the worst for the algorithm.

2. In the end: $\exists$ at least one funct. $f \in \mathcal{F}$ that is consistent with $O(x_0), \dots, O(x_k)$.

In our case: $\boxed{O(x) = \{0\}.}$

The result x_k of an algorithm will be $f(x_k) = 0$.

Denote by $K(R, r, n)$ – a lower bound on size of a maximal packing of ball of radis $R > 0$ by small balls of radius $r > 0$ in $\mathbb{R}^n$.

By Th 1. $\Rightarrow$ $K(R, r, n) = \left\lceil \left(\frac{R-r}{2r}\right)^n \right\rceil$.

Theorem 2. Let $0<r<R$. Consider any zeroth-order algorithm running for $k < K(R,r,n) - 1$ iterations. Then,

$\exists$ a Lipschitz function (with const. $L>0$)

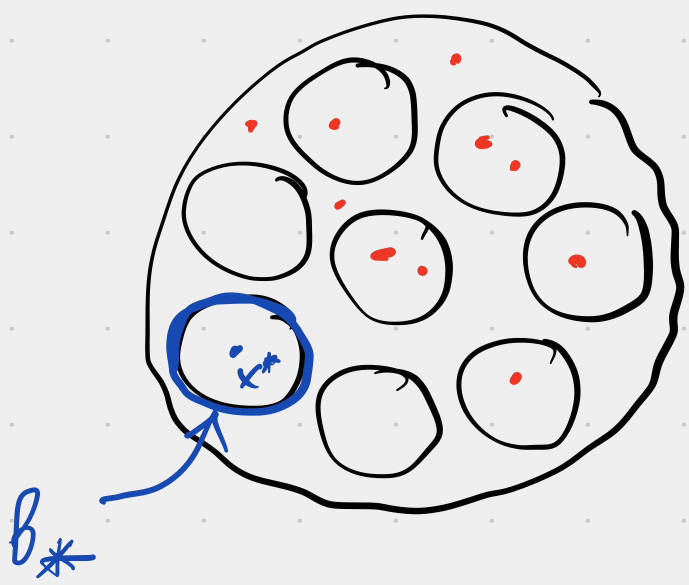
s.t. $$f(x_k) - f^* = Lr.$$

Proof Let $\{x_0, \dots, x_k\} \in B$ be a sequence of points generated by the algorithm, when applied to the resist. oracle. $\Rightarrow$ $f(x_k) = 0$.

Now, consider any maximal packing of size $N \geq K(R,r,n) > k+1$.
$\Rightarrow$ $\exists$ a small ball B_* in our maximal packing s.t.

$$x_i \notin B_* \quad \forall\, 0 \le i \le k.$$

Denote by x^* the center of B_*.



Introduce: $f(x) = L \cdot \min\{0, \|x - x^*\| - r\}$.

(1) $f(x^*) = -rL$

(2) $f(x) = 0 \quad \forall x: \|x - x^*\| \geq r$.

(3) f is Lipschitz.

Let $x, y \in B_*$: $\|x - x^*\| \leq r$, $\|y - x^*\| \leq r$. Then

$$f(x) - f(y) = L \cdot \left(\|x - x^*\| - r - \|y - x^*\| + r \right)$$
$$= L \left(\|x - x^*\| - \|y - x^*\| \right) \leq L \|x - y\|.$$

Let $x \notin B_*$, $y \in B_*$.

$$f(x) - f(y) \leq L \left(\|x - x^*\| - r \right) - L \left(\|y - x^*\| - r \right)$$
$$\leq L \|x - y\|.$$

Let $x \notin B_*$, $y \notin B_*$:

$$f(x) - f(y) = 0 \leq L \|x - y\|.$$

Therefore

$$f(x_n) - f^* = rL. \quad \square$$

Lower Bound

Theorem 3. Let $L>0$, $R>0$, $\varepsilon>0$. Assume: $\varepsilon < \frac{LR}{2}$.
Then, for <u>any</u> zeroth-order algorithm, the complexity is bounded:

$$K \geq \left\lceil \left(\frac{LR}{4\varepsilon} - \frac{1}{2}\right)^n \right\rceil - 1.$$

Proof. $\quad r = \frac{2\varepsilon}{L} < R.$

Notice,

$$K(R,r,n) = \left\lceil \left(\frac{R-r}{2r}\right)^n \right\rceil = \left\lceil \left(\frac{R}{2r} - \frac{1}{2}\right)^n \right\rceil$$

$$= \left\lceil \left(\frac{LR}{4\varepsilon} - \frac{1}{2}\right)^n \right\rceil.$$

Th2: Assume $k < K(R,r,n) - 1$. Then,

$$f(x_k) - f^* = Lr = 2\varepsilon. \quad (?!)$$

$\Rightarrow$ To solve the problem we need

$$k \geq K(R,r,n) - 1. \quad \square$$

$\|\cdot\| := \|\cdot\|_\infty \Rightarrow$ Grid Search Algorithm

has the complexity:

$$\Theta\left(\left\lceil \frac{LR}{\varepsilon} \right\rceil^n \right).$$

- Grid Search is optimal for $\|\cdot\|_\infty$.
- Other norms — open question?