

Information complexity of convex optimization with integer variables

Amitabh Basu

Johns Hopkins University

Optimization and learning: theory and applications
Centre de Recherches Mathematiques (CRM)
Montreal, May 2025



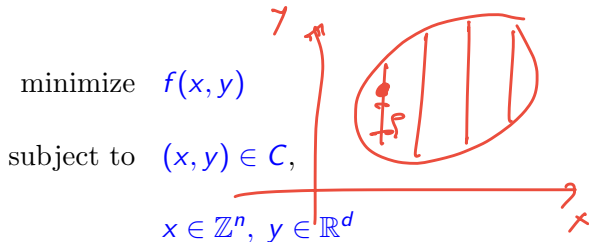
Convex optimization with integer variables

$$\begin{array}{ll}\text{minimize} & f(x, y) \\ \text{subject to} & (x, y) \in C, \\ & \underline{x \in \mathbb{Z}^n, y \in \mathbb{R}^d}\end{array}$$

$C \subseteq \mathbb{R}^n \times \mathbb{R}^d$ closed, convex

$f : \underline{\mathbb{R}^n \times \mathbb{R}^d} \rightarrow \mathbb{R}$ convex

Convex optimization with integer variables



$C \subseteq \mathbb{R}^n \times \mathbb{R}^d$ closed, convex.

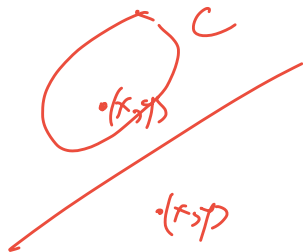
Contained in ℓ_∞ ball of radius $R > 0$.

Contains ℓ_∞ ball of radius $\rho > 0$ in the optimal fiber.

$f : \mathbb{R}^n \times \mathbb{R}^d \rightarrow \mathbb{R}$ convex, Lipschitz continuous with constant M over $[-R, R]^{n+d}$.

Convex optimization with integer variables

$$\begin{array}{ll}\text{minimize} & f(x, y) \\ \text{subject to} & (x, y) \in C, \\ & x \in \mathbb{Z}^n, y \in \mathbb{R}^d\end{array}$$



$C \subseteq \mathbb{R}^n \times \mathbb{R}^d$ closed, convex.

Contained in ℓ_∞ ball of radius $R > 0$.

Contains ℓ_∞ ball of radius $\rho > 0$ in the optimal fiber.

$f : \mathbb{R}^n \times \mathbb{R}^d \rightarrow \mathbb{R}$ convex, Lipschitz continuous with constant M over $[-R, R]^{n+d}$.

First order oracle: separation for C , value + subgradient for f

Information (a.k.a oracle) complexity

Information (a.k.a oracle) complexity

$$\epsilon > 0$$
$$(\hat{x}, \hat{y}) \in \mathcal{C} \wedge f(\hat{x}, \hat{y}) \leq f(x, y) + \epsilon$$

Notion of ϵ -approximate solutions

$$\forall (x, y) \in \mathcal{C}$$

Information complexity: Minimum number of queries needed to (provably) report ϵ -approximate solutions.

Information (a.k.a oracle) complexity

Notion of ϵ -approximate solutions

Information complexity: Minimum number of queries needed to ~~(provably) report~~ distinguish between instances with disjoint ϵ -approximate solutions.

Information (a.k.a oracle) complexity

Notion of ϵ -approximate solutions

Information complexity: Minimum number of queries needed to ~~(provably) report~~ distinguish between instances with disjoint ϵ -approximate solutions.

Algorithmic complexity: Total amount of computation needed to report ϵ -approximate solutions.

Information (a.k.a oracle) complexity

Notion of ϵ -approximate solutions

Information complexity: Minimum number of queries needed to ~~(provably) report~~ distinguish between instances with disjoint ϵ -approximate solutions.

Algorithmic complexity: Total amount of computation needed to report ϵ -approximate solutions.

$$\text{Information complexity} \leq \text{Algorithmic complexity}$$

Complexity of convex optimization with integer variables

Information complexity (standard first-order oracle):

Lower bounds

$$d \geq 1 : \Omega \left(d 2^n \log \left(\frac{MR}{\rho \epsilon} \right) \right)$$

$$d = 0 : \Omega (2^n \log(R))$$

Upper bounds

$$n, d \geq 1 : O \left(d(n + d) 2^n \log \left(\frac{MR}{\rho \epsilon} \right) \right)$$

$$d = 0 : O (n 2^n \log(R))$$

$$n = 0 : O \left(d \log \left(\frac{MR}{\rho \epsilon} \right) \right)$$

Complexity of convex optimization with integer variables

Information complexity (standard first-order oracle):

Lower bounds

$$d \geq 1 : \Omega \left(d 2^n \log \left(\frac{MR}{\rho \epsilon} \right) \right)$$

Nemirovskii-Yudin 1970s

B.-Oertel 2017

B. 2023

$$d = 0 : \Omega (2^n \log(R))$$

B.-Oertel 2017

B. 2023

Upper bounds

$$n, d \geq 1 : O \left(\overbrace{d(n+d)}^{2^{O(n)} \log R} 2^n \log \left(\frac{MR}{\rho \epsilon} \right) \right)$$

B.-Oertel 2017

$$d = 0 : O(n 2^n \log(R))$$

B.-Oertel 2017

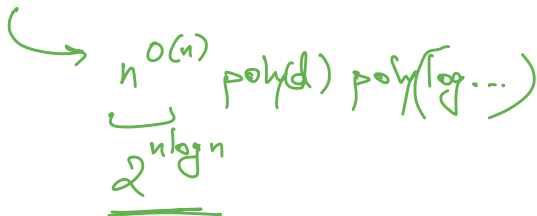
$$n = 0 : O \left(d \log \left(\frac{MR}{\rho \epsilon} \right) \right)$$

Nemirovskii-Yudin 1970s

Complexity of convex optimization with integer variables

Overall (deterministic) complexity:

$$(n + d)^{O(n)} \text{poly} \left(\log \left(\frac{MR}{\rho\epsilon} \right) \right)$$

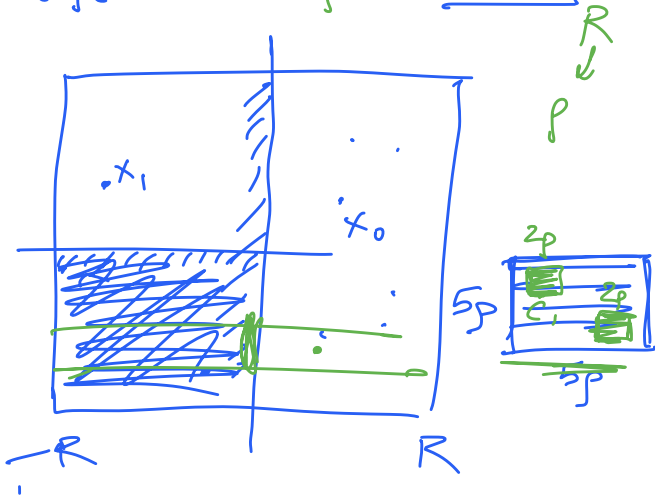

$$\begin{aligned} &\rightarrow n^{O(n)} \text{poly}(d) \text{poly}(\log \dots) \\ &\quad \underbrace{\hspace{1cm}}_{2^{n \log n}} \end{aligned}$$

Information complexity lower bounds

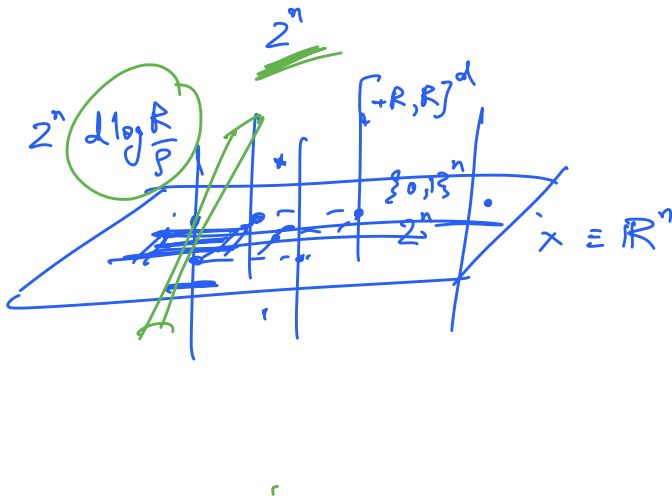
$$d \log \frac{MR}{p\epsilon}$$

$$2^n d \log \frac{R}{p}$$

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

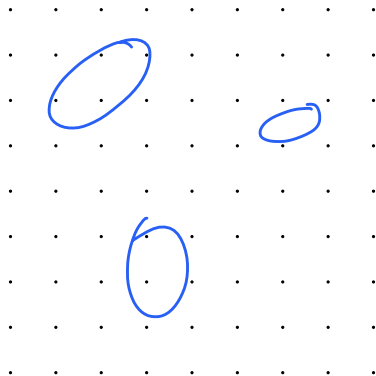


Information complexity lower bounds



Information complexity upper bounds

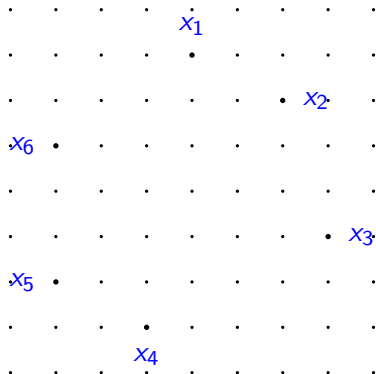
Feasibility problem with separation oracle for $\mathcal{C}$



Information complexity upper bounds

Feasibility problem with separation oracle for C

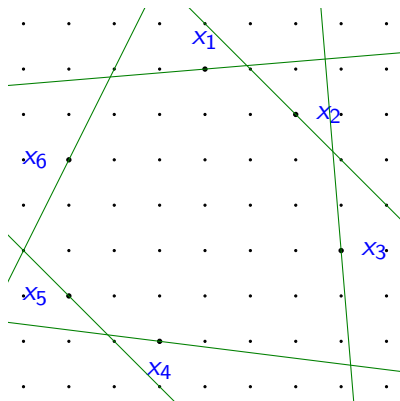
Assume we have queried at $x_1, x_2, \dots, x_N$.



Information complexity upper bounds

Feasibility problem with separation oracle for C

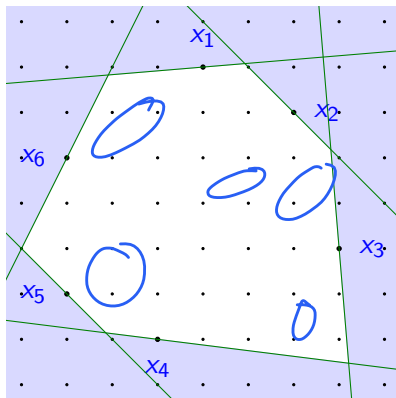
Assume we have queried at $x_1, x_2, \dots, x_N$.



Information complexity upper bounds

Feasibility problem with separation oracle for C

Assume we have queried at $x_1, x_2, \dots, x_N$.

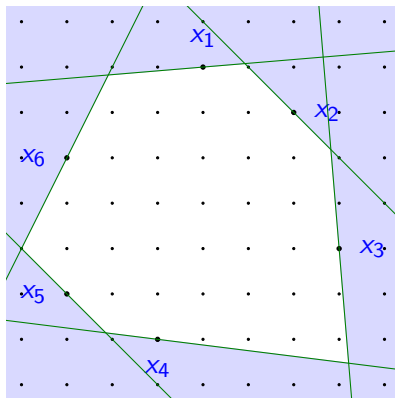


Information complexity upper bounds

Feasibility problem with separation oracle for C

Assume we have queried at $x_1, x_2, \dots, x_N$.

How should we choose x_{N+1} ?

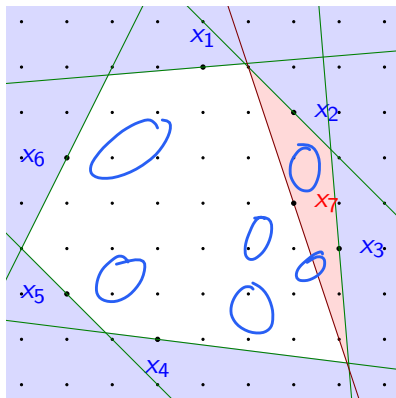


Information complexity upper bounds

Feasibility problem with separation oracle for C

Assume we have queried at $x_1, x_2, \dots, x_N$.

How should we choose x_{N+1} ?

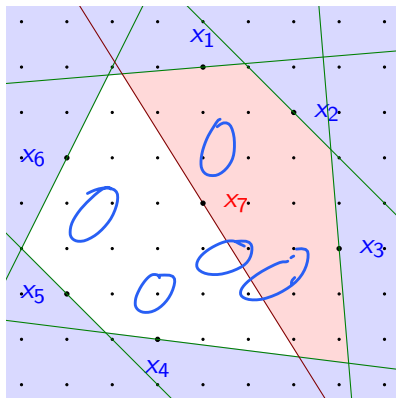


Information complexity upper bounds

Feasibility problem with separation oracle for C

Assume we have queried at $x_1, x_2, \dots, x_N$.

How should we choose x_{N+1} ?



Helly numbers and centerpoints

THEOREM Helly 1923, Doignon-Bell-Scarf 1970s,
Averkov-Weismantel 2012

$C_1, \dots, C_k \subseteq \mathbb{R}^n \times \mathbb{R}^d$ convex. If

$$C_1 \cap C_2 \cap \dots \cap C_k \cap (\mathbb{Z}^n \times \mathbb{R}^d) = \emptyset,$$

then $\exists i_1, \dots, i_{2^n(d+1)}$ such that

$$C_{i_1} \cap C_{i_2} \cap \dots \cap C_{i_k} \cap (\mathbb{Z}^n \times \mathbb{R}^d) = \emptyset.$$

Helly numbers and centerpoints

THEOREM Helly 1923, Doignon-Bell-Scarf 1970s,
Averkov-Weismantel 2012

$C_1, \dots, C_k \subseteq \mathbb{R}^n \times \mathbb{R}^d$ convex. If

$$C_1 \cap C_2 \cap \dots \cap C_k \cap (\mathbb{Z}^n \times \mathbb{R}^d) = \emptyset,$$

then $\exists i_1, \dots, i_{2^n(d+1)}$ such that

$$C_{i_1} \cap C_{i_2} \cap \dots \cap C_{i_k} \cap (\mathbb{Z}^n \times \mathbb{R}^d) = \emptyset.$$

$$1 - \frac{1}{2^{n(d+1)}}$$

THEOREM Grünbaum 1960, Oertel 2013, B.-Oertel 2017

Probability measure μ supported on $\mathbb{Z}^n \times \mathbb{R}^d$. There exists $z^\mu \in \mathbb{Z}^n \times \mathbb{R}^d$ such that for all halfspaces H containing z^μ ,

centerpoint

$$\mu(H) \geq \frac{1}{2^{n(d+1)}}.$$

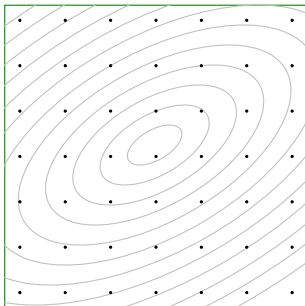


Centerpoint based algorithm

- ▶ Let $P_0 := [-R, R]^{n+d}$.
- ▶ For $i = 1$ to N
 - ▶ Compute centerpoint z^i of P_{i-1} .
 - ▶ If $z^i \in C$, then return z^i and STOP.
 - ▶ Else, let H be a separating halfspace
 - ▶ Define $P_i = P_{i-1} \cap H$.
- ▶ Return “INFEASIBLE”.

Centerpoint based algorithm

- ▶ Let $P_0 := [-R, R]^{n+d}$.
- ▶ For $i = 1$ to N
 - ▶ Compute centerpoint z^i of P_{i-1} .
 - ▶ If $z^i \in C$, then return z^i and STOP.
 - ▶ Else, let H be a separating halfspace
 - ▶ Define $P_i = P_{i-1} \cap H$.
- ▶ Return “INFEASIBLE”.

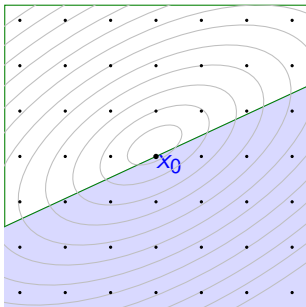


Centerpoint based algorithm

- ▶ Let $P_0 := [-R, R]^{n+d}$.
- ▶ For $i = 1$ to N
 - ▶ Compute centerpoint z^i of P_{i-1} .
 - ▶ If $z^i \in C$, then return z^i and STOP.
 - ▶ Else, let H be a separating halfspace
 - ▶ Define $P_i = P_{i-1} \cap H$.
- ▶ Return "INFEASIBLE".

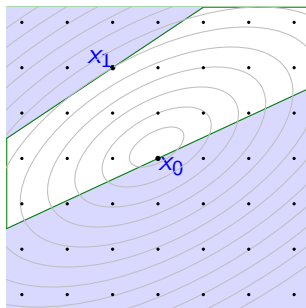
$$\frac{1}{2^n(d+1)} = \frac{1}{2^n}$$

$$\left(1 - \frac{1}{2^n}\right)$$



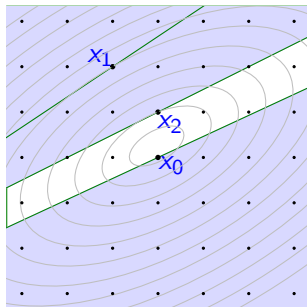
Centerpoint based algorithm

- ▶ Let $P_0 := [-R, R]^{n+d}$.
- ▶ For $i = 1$ to N
 - ▶ Compute centerpoint z^i of P_{i-1} .
 - ▶ If $z^i \in C$, then return z^i and STOP.
 - ▶ Else, let H be a separating halfspace
 - ▶ Define $P_i = P_{i-1} \cap H$.
- ▶ Return “INFEASIBLE”.



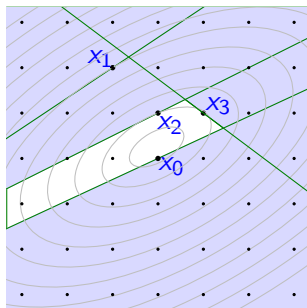
Centerpoint based algorithm

- ▶ Let $P_0 := [-R, R]^{n+d}$.
- ▶ For $i = 1$ to N
 - ▶ Compute centerpoint z^i of P_{i-1} .
 - ▶ If $z^i \in C$, then return z^i and STOP.
 - ▶ Else, let H be a separating halfspace
 - ▶ Define $P_i = P_{i-1} \cap H$.
- ▶ Return “INFEASIBLE”.



Centerpoint based algorithm

- ▶ Let $P_0 := [-R, R]^{n+d}$.
- ▶ For $i = 1$ to N
 - ▶ Compute centerpoint z^i of P_{i-1} .
 - ▶ If $z^i \in C$, then return z^i and STOP.
 - ▶ Else, let H be a separating halfspace
 - ▶ Define $P_i = P_{i-1} \cap H$.
- ▶ Return "INFEASIBLE".



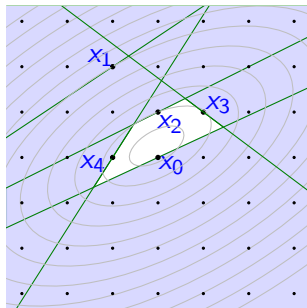
Centerpoint based algorithm

- ▶ Let $P_0 := [-R, R]^{n+d}$.
- ▶ For $i = 1$ to N
 - ▶ Compute centerpoint z^i of P_{i-1} .
 - ▶ If $z^i \in C$, then return z^i and STOP.
 - ▶ Else, let H be a separating halfspace
 - ▶ Define $P_i = P_{i-1} \cap H$.
- ▶ Return "INFEASIBLE".

$$< 1$$

$$R^n$$

$$\left(1 - \frac{1}{2^n}\right)$$



Centerpoint based algorithm

- ▶ Let $P_0 := [-R, R]^{n+d}$.
- ▶ For $i = 1$ to N
 - ▶ Compute centerpoint z^i of P_{i-1} .
 - ▶ If $z^i \in C$, then return z^i and STOP.
 - ▶ Else, let H be a separating halfspace
 - ▶ Define $P_i = P_{i-1} \cap H$.
- ▶ Return “INFEASIBLE”.

Complexity of convex optimization with integer variables

Information complexity (standard first-order oracle):

Lower bounds

$$d \geq 1: \Omega\left(d 2^n \log\left(\frac{MR}{\rho\epsilon}\right)\right)$$

$$d = 0: \Omega(2^n \log(R))$$



Upper bounds

$$n, d \geq 1: O\left(d(n+d)2^n \log\left(\frac{MR}{\rho\epsilon}\right)\right)$$

$$d = 0: O(n 2^n \log(R))$$

$$n = 0: O\left(d \log\left(\frac{MR}{\rho\epsilon}\right)\right)$$

Complexity of convex optimization with integer variables

THEOREM Grunbaum 1960, Oertel 2013, B.-Oertel 2017

Probability measure μ supported on $\mathbb{Z}^n \times \mathbb{R}^d$. There exists $z^\mu \in \mathbb{Z}^n \times \mathbb{R}^d$ such that for all halfspaces H containing z^μ ,

$$\mu(H) \geq \frac{1}{2^n(d+1)}.$$

Complexity of convex optimization with integer variables

Information complexity (standard first-order oracle):

Lower bounds

$$\begin{aligned} d \geq 1: & \quad \Omega \left(d 2^n \log \left(\frac{MR}{\rho \epsilon} \right) \right) \\ d = 0: & \quad \Omega \left(2^n \log(R) \right) \end{aligned} \left(1 - \frac{1}{2^n (d+1)} \right)$$

Upper bounds

$$\frac{1}{2^n e}$$

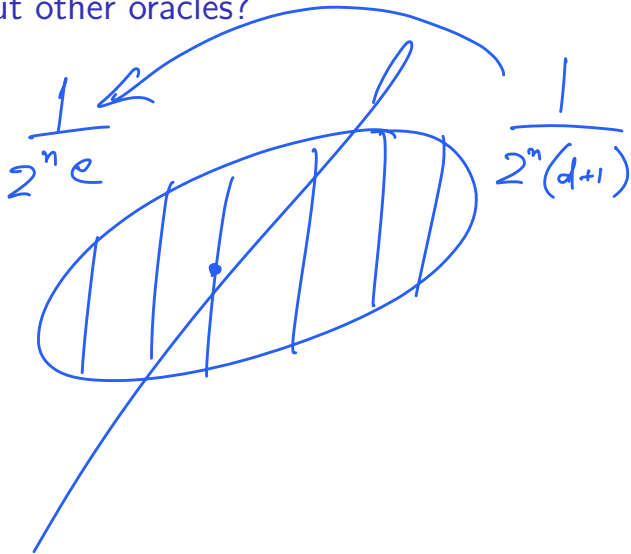
$$n, d \geq 1: \quad O \left(d(n+d) 2^n \log \left(\frac{MR}{\rho \epsilon} \right) \right) \frac{1}{e} \frac{1}{d+1}$$

$$d = 0: \quad O(n 2^n \log(R))$$

$$n = 0: \quad O \left(d \log \left(\frac{MR}{\rho \epsilon} \right) \right)$$



What about other oracles?



What about other oracles?

Family of all possible **binary** queries: $\Theta \left((n + d) \log \left(\frac{MR}{\rho\epsilon} \right) \right)$
information complexity.

What about other oracles?

Definition: Oracle based on **first-order information**:

1. **First-order chart:** Maps from instances to separating hyperplanes, function values and (sub)gradients
2. **Permissible queries:** Maps from vectors/real numbers to a response set.

A **query** is now a pair (z, h) , where $z \in \mathbb{R}^n \times \mathbb{R}^d$ and h is a permissible query.

What about other oracles?

Examples:

1. **Full-information first-order oracle:** Permissible queries are identity functions.
2. **Bit oracles:** Permissible queries ask for a **bit** of the function value/some specific coordinate of (sub)gradient or separating hyperplane.
3. **Directional oracles:** Permissible queries ask for a **sign of inner product** between a specified direction and (sub)gradient/separating hyperplane/function value.
4. **General binary oracles**

Information complexity with other oracles

A “transfer theorem” (B.-Kerger-Jiang-Molinaro 2024):

For any oracle using first-order information:

$\ell(d, R, \rho, M, \epsilon)$ lower bound for continuous convex optimization



Mixed-integer problem has information complexity at least

$$\Omega\left(2^n \ell(d, R, \rho, M, \epsilon)\right)$$

Information complexity with other oracles

Memory constrained convex optimization

(Marsden, Sharan, Sidford, Valiant 2022,
Blanchard, Zhang, Jaillet 2023):

$$\ell(d, R, \rho, M, \epsilon) = \max \left\{ d^{5/4}, d \log \left(\frac{MR}{\rho\epsilon} \right) \right\} \text{ for general binary oracle}$$

Information complexity with other oracles

Memory constrained convex optimization

(Marsden, Sharan, Sidford, Valiant 2022,
Blanchard, Zhang, Jaillet 2023):

$$\ell(d, R, \rho, M, \epsilon) = \max \left\{ d^{5/4}, d \log \left(\frac{MR}{\rho\epsilon} \right) \right\} \text{ for general binary oracle}$$

THEOREM (B.-Kerger-Jiang-Molinaro 2024)

$$\Omega \left(2^n \max \left\{ d^{5/4}, d \log \left(\frac{MR}{\rho\epsilon} \right) \right\} \right)$$

mixed-integer lower bound for general binary oracle.

Information complexity with other oracles

Examples:

1. **Full-information first-order oracle:** Permissible queries are identity functions.
2. **Bit oracles:** Permissible queries ask for a **bit** of the function value/some specific coordinate of (sub)gradient or separating hyperplane.
3. **Directional oracles:** Permissible queries ask for a **sign of inner product** between a specified direction and (sub)gradient/separating hyperplane/function value.
4. **General binary oracles**

Some open questions

Some open questions

Conjecture on **mixed-integer centerpoints**: improve bound from $\frac{1}{2^n(d+1)}$ to $\frac{1}{2^ne}$.

Unify upper bounds, remove gap between lower and upper bounds

Some open questions

Conjectured lower bound for **general binary queries based on first order information** (continuous convex optimization):

$$\Omega \left(d^2 \log \left(\frac{MR}{\rho \epsilon} \right) \right)$$

Sample complexity of stochastic convex opt

$$F : (\mathbb{R}^n \times \mathbb{R}^d) \times \mathcal{X} \rightarrow \mathbb{R}$$

$$C \subseteq \mathbb{R}^n \times \mathbb{R}^d$$

ξ is a random variable taking values in $\mathcal{X}$

$$\min_{z \in C \cap (\mathbb{Z}^n \times \mathbb{R}^d)} \mathbb{E}_{\xi} [F(z, \xi)]$$

Sample complexity of stochastic convex opt

$$F : (\mathbb{R}^n \times \mathbb{R}^d) \times \mathcal{X} \rightarrow \mathbb{R}$$

$$C \subseteq \mathbb{R}^n \times \mathbb{R}^d$$

ξ is a random variable taking values in $\mathcal{X}$

$$\min_{z \in C \cap (\mathbb{Z}^n \times \mathbb{R}^d)} \mathbb{E}_{\xi} [F(z, \xi)]$$

Solve the problem only given access to n i.i.d. samples of ξ .

Sample complexity of stochastic convex opt

$$\min_{z \in C \cap (\mathbb{Z}^n \times \mathbb{R}^d)} \mathbb{E}_{\xi} [F(z, \xi)]$$

Solve the problem only given access to n i.i.d. samples of ξ .

Natural idea: Given samples $\xi^1, \dots, \xi^n \in \mathcal{X}$, solve the deterministic problem

$$\min_{z \in C \cap (\mathbb{Z}^n \times \mathbb{R}^d)} \frac{1}{n} \sum_{i=1}^n F(z, \xi^i)$$

Sample complexity of stochastic convex opt

$$\min_{z \in C \cap (\mathbb{Z}^n \times \mathbb{R}^d)} \mathbb{E}_{\xi} [F(z, \xi)]$$

Solve the problem only given access to n i.i.d. samples of ξ .

Natural idea: Given samples $\xi^1, \dots, \xi^n \in \mathcal{X}$, solve the deterministic problem

$$\min_{z \in C \cap (\mathbb{Z}^n \times \mathbb{R}^d)} \frac{1}{n} \sum_{i=1}^n F(z, \xi^i)$$

Stochastic optimizers call this **sample average approximation (SAA)**; machine learners call this **empirical risk minimization (ERM)**.

Sample complexity of stochastic convex opt

$$\min_{z \in C \cap (\mathbb{Z}^n \times \mathbb{R}^d)} \mathbb{E}_{\xi} [F(z, \xi)]$$

Given access to n i.i.d. samples $\xi^1, \dots, \xi^n \in \mathcal{X}$.

Sample complexity of stochastic convex opt

$$\min_{z \in C \cap (\mathbb{Z}^n \times \mathbb{R}^d)} \mathbb{E}_{\xi} [F(z, \xi)]$$

Given access to n i.i.d. samples $\xi^1, \dots, \xi^n \in \mathcal{X}$.

Uniform convergence:

$$\left| \frac{1}{n} \sum_{i=1}^n F(z, \xi^i) - \mathbb{E}_{\xi} [F(z, \xi)] \right| \leq \epsilon$$

for all $z \in [-R, R]^{n+d} \cap (\mathbb{Z}^n \times \mathbb{R}^d)$ with prob. $1 - \delta$

Sample complexity of stochastic convex opt

$$\min_{z \in C \cap (\mathbb{Z}^n \times \mathbb{R}^d)} \mathbb{E}_{\xi} [F(z, \xi)]$$

Given access to n i.i.d. samples $\xi^1, \dots, \xi^n \in \mathcal{X}$.

Uniform convergence:

$$\left| \frac{1}{n} \sum_{i=1}^n F(z, \xi^i) - \mathbb{E}_{\xi} [F(z, \xi)] \right| \leq \epsilon$$

for all $z \in [-R, R]^{n+d} \cap (\mathbb{Z}^n \times \mathbb{R}^d)$ with prob. $1 - \delta$

SAA/ERM:

$$\mathbb{E}_{\xi} [F(z_{\text{SAA}}, \xi)] - \mathbb{E}_{\xi} [F(z^*, \xi)] \leq \epsilon$$

with prob. $1 - \delta$

Sample complexity of stochastic convex opt

$$\min_{z \in C \cap (\mathbb{Z}^n \times \mathbb{R}^d)} \mathbb{E}_{\xi} [F(z, \xi)]$$

Given access to n i.i.d. samples $\xi^1, \dots, \xi^n \in \mathcal{X}$.

Uniform convergence:

$$\left| \frac{1}{n} \sum_{i=1}^n F(z, \xi^i) - \mathbb{E}_{\xi} [F(z, \xi)] \right| \leq \epsilon$$

for all $z \in [-R, R]^{n+d} \cap (\mathbb{Z}^n \times \mathbb{R}^d)$ with prob. $1 - \delta$

SAA/ERM:

$$\mathbb{E}_{\xi} [F(z_{\text{SAA}}, \xi)] - \mathbb{E}_{\xi} [F(z^*, \xi)] \leq \epsilon$$

with prob. $1 - \delta$

General learning: Report $\hat{z}(\xi^1, \dots, \xi^n)$ such that

$$\mathbb{E}_{\xi} [F(\hat{z}, \xi)] - \mathbb{E}_{\xi} [F(z^*, \xi)] \leq \epsilon$$

with prob. $1 - \delta$

Continuous case ($n = 0$)

	Lower bound	Upper bound
UC	$\Omega \left(d \frac{M^2 R^2}{\epsilon^2} + \frac{M^2 R^2}{\epsilon^2} \log \frac{1}{\delta} \right)$	$O \left(d \frac{M^2 R^2}{\epsilon^2} \log \left(\frac{MR}{\epsilon} \right) + \frac{M^2 R^2}{\epsilon^2} \log \frac{1}{\delta} \right)$
ERM	$\Omega \left(d \frac{MR}{\epsilon} + \frac{M^2 R^2}{\epsilon^2} \log \frac{1}{\delta} \right)$	$O \left(d \frac{MR}{\epsilon} + \frac{M^2 R^2}{\epsilon^2} \log \frac{1}{\delta} \right)$
GL	$\Omega \left(\frac{M^2 R^2}{\epsilon^2} \log \frac{1}{\delta} \right)$	$O \left(\frac{M^2 R^2}{\epsilon^2} \log \frac{1}{\delta} \right)$

With integer variables

	Lower bound	Upper bound
UC	??	$O\left(\frac{M^2 R^2}{\epsilon^2} \left(n \log(R) + d \log \frac{MR}{\epsilon} + \log \frac{1}{\delta} \right)\right)$
ERM	??	??
GL	$\Omega\left(\frac{M^2 R^2}{\epsilon^2} \left(n + d + \log \frac{1}{\delta} \right)\right)$	??

Accompanying papers

<https://arxiv.org/abs/2110.06172>

(Math. Prog., Series B 2023)

<https://arxiv.org/abs/2308.11153>

(B.-Kerger-Jiang-Molinaro, Math. Prog., Series B 2024, IPCO
2023)

THANK YOU !

Questions/Comments ?