Information complexity of mixed-integer convex optimization

Phillip Kerger

Work with Amitabh Basu, Hongyi Jiang, and Marco Molinaro

Johns Hopkins University

June 2023, IPCO

Outline

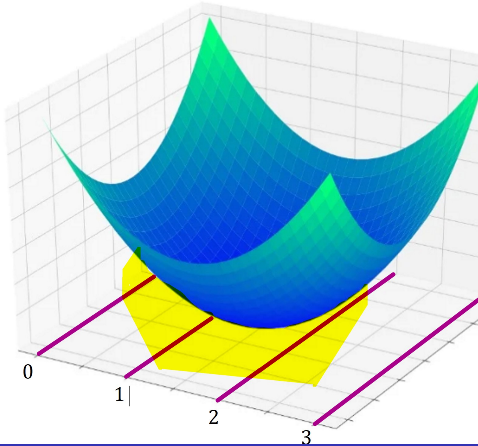
Outline

Setting

Mixed-integer convex optimization:

Min $f(x, y) : \mathbb{R}^{n+d} \rightarrow \mathbb{R}$ with

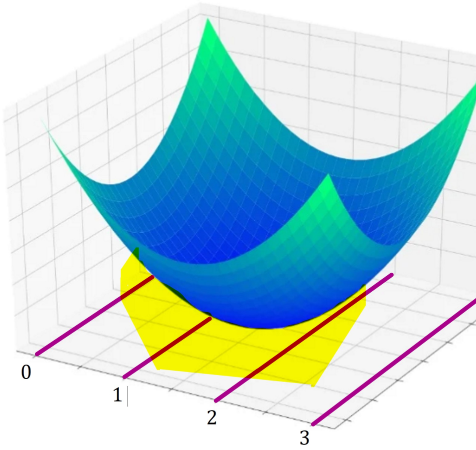
- f convex (possibly nonsmooth)
- $x \in \mathbb{Z}^n, y \in \mathbb{R}^d$
- $(x, y) \in C \subset \mathbb{R}^{n+d}$, C a convex set



Mixed-integer convex optimization:

Can't solve this exactly!

$\Rightarrow$ Only ask for ε-solution: Feasible point $(x, y) : f(x, y) - OPT_f \leq \varepsilon$.



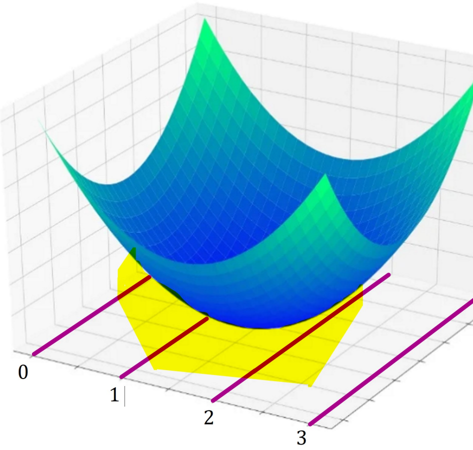
Setting

Mixed-integer convex optimization:

Can't solve this exactly!

$\Rightarrow$ Only ask for ε-solution:
Feasible point
$(x, y) : f(x, y) - OPT_f \leq \varepsilon$.

But if C is tiny...
there is also a problem



Setting

Parameterize the instances:

Definition

$\mathcal{I}_{n,d,R,\rho,M}$ is the set of all MICO instances such that

(i) C is a subset of the box $[-R, R]^{n+d}$.

(ii) Fiber containing optimum isn't degenerate, but a ball of size ρ in it is feasible.

(iii) f is M-Lipschitz

Setting

Information Complexity:

How much information is needed to solve a problem?

Setting

Information Complexity:
How much information is needed to solve a problem?

Classical First-Order:
How many (exact) evaluations of function value and gradient or separating hyperplanes?

Information Complexity:
How much information is needed to solve a problem?

Classical First-Order:
How many (exact) evaluations of function value and gradient or separating hyperplanes?

Notice: Gradient length grows with dimension...

Setting

Information Complexity:
How much information is needed to solve a problem?

Classical First-Order:
How many (exact) evaluations of function value and gradient or separating hyperplanes?

Notice: Gradient length grows with dimension...
Gives **more information with dimension**!
Can we refine this in a meaningful way?

Setting

Information Complexity:
How much information is needed to solve a problem?

Classical First-Order:
How many (exact) evaluations of function value and gradient or separating hyperplanes?

Notice: Gradient length grows with dimension...
Gives **more information with dimension**!
Can we refine this in a meaningful way?

Goals of this work:

- Study information complexity beyond exact first-order model
- Tighten and generalize existing bounds

Oracles Using First-Order Information

What kinds of oracles?

What kinds of oracles?

Definition

An *oracle using first-order information* consists of two parts:

Oracles Using First-Order Information

What kinds of oracles?

Definition

An *oracle using first-order information* consists of two parts:

1. For each point $\mathbf{z} \in \mathbb{R}^{n+d}$, a map $g_{\mathbf{z}}$ that maps instances to first-order information at $\mathbf{z}$:
   - Function value and gradient (objective)
   - Feasibility flag and separating hyperplane (feasibility)
   - A collection of these for all $\mathbf{z}$: "First order chart" $\mathcal{G}$

Oracles Using First-Order Information

What kinds of oracles?

Definition

An *oracle using first-order information* consists of two parts:

1. For each point $\mathbf{z} \in \mathbb{R}^{n+d}$, a map $g_{\mathbf{z}}$ that maps instances to first-order information at $\mathbf{z}$:
   - Function value and gradient (objective)
   - Feasibility flag and separating hyperplane (feasibility)
   - A collection of these for all $\mathbf{z}$: "First order chart" $\mathcal{G}$

What kinds of oracles?

Definition

An *oracle using first-order information* consists of two parts:

1. For each point $\mathbf{z} \in \mathbb{R}^{n+d}$, a map $g_{\mathbf{z}}$ that maps instances to first-order information at $\mathbf{z}$:
   - Function value and gradient (objective)
   - Feasibility flag and separating hyperplane (feasibility)
   - A collection of these for all $\mathbf{z}$: "First order chart" $\mathcal{G}$
2. A set of *permissible queries* $\mathcal{H}$
   $h \in \mathcal{H}$ functions taking first order information as input.

Oracles Using First-Order Information

What kinds of oracles?

Definition

An *oracle using first-order information* consists of two parts:

1. For each point $\mathbf{z} \in \mathbb{R}^{n+d}$, a map $g_{\mathbf{z}}$ that maps instances to first-order information at $\mathbf{z}$:
   - Function value and gradient (objective)
   - Feasibility flag and separating hyperplane (feasibility)
   - A collection of these for all $\mathbf{z}$: "First order chart" $\mathcal{G}$
2. A set of *permissible queries* $\mathcal{H}$
   $h \in \mathcal{H}$ functions taking first order information as input.

Oracles Using First-Order Information

What kinds of oracles?

Definition

An *oracle using first-order information* consists of two parts:

1. For each point $\mathbf{z} \in \mathbb{R}^{n+d}$, a map $g_{\mathbf{z}}$ that maps instances to first-order information at $\mathbf{z}$:
   - Function value and gradient (objective)
   - Feasibility flag and separating hyperplane (feasibility)
   - A collection of these for all $\mathbf{z}$: "First order chart" $\mathcal{G}$
2. A set of *permissible queries* $\mathcal{H}$
   $h \in \mathcal{H}$ functions taking first order information as input.

For query h at $\mathbf{z}$, for instance J, oracle answers $h(g_{\mathbf{z}}(J))$

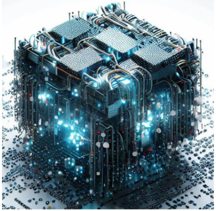
$\rightarrow$ Pair $(\mathcal{G}, \mathcal{H})$ defines an oracle!

Oracles Using First-Order Information

Query point $z \in \mathbb{R}^{n+d}$ with query $h \in \mathcal{H}$

Query (z, h) →

← Answer $h(g_z(J))$



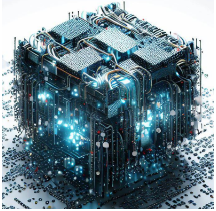
g_z determined by first-order chart $\mathcal{G}$

Oracles Using First-Order Information

Query point $z \in \mathbb{R}^{n+d}$ with query $h \in \mathcal{H}$

Query (z, h) →

← Answer $h(g_z(J))$



g_z determined by first-order chart $\mathcal{G}$

- Large class of oracles with answers using only first-order info

Examples: Oracles Using First-Order Information

What kinds of oracles can we get?

- **Standard first-order oracle**: $\mathcal{H}$ has only the identity

Examples: Oracles Using First-Order Information

What kinds of oracles can we get?

- **Standard first-order oracle**: $\mathcal{H}$ has only the identity
- **Bit Queries**: $\mathcal{H}$ output some designated bits of the function value/gradient/sep. hyperplane.
  $\rightarrow$ *How many* **bits** *of information does one need?*

Examples: Oracles Using First-Order Information

What kinds of oracles can we get?

- **Standard first-order oracle**: $\mathcal{H}$ has only the identity
- **Bit Queries**: $\mathcal{H}$ output some designated bits of the function value/gradient/sep. hyperplane.
  $\rightarrow$ *How many* **bits** *of information does one need?*
- **Directional Queries**: Elements of $\mathcal{H}$ give the sign of the inner product of the gradient/sep. hyperplane with a vector $h_v = sign\langle v, g_{\mathbf{z}}\rangle$.

What kinds of oracles can we get?

- **Standard first-order oracle**: $\mathcal{H}$ has only the identity
- **Bit Queries**: $\mathcal{H}$ output some designated bits of the function value/gradient/sep. hyperplane.
  $\rightarrow$ *How many* **bits** *of information does one need?*
- **Directional Queries**: Elements of $\mathcal{H}$ give the sign of the inner product of the gradient/sep. hyperplane with a vector $h_v = sign\langle v, g_{\mathbf{z}} \rangle$.
- **General Binary Queries**: Let $\mathcal{H}$ contain all binary queries

Outline

Lower Bounds

Theorem (Lower-Bound for Standard Oracle)

$$\text{icomp} = \Omega\left(2^n \cdot d \log\left(\frac{MR}{\rho\varepsilon}\right)\right).$$

Lower Bounds

Theorem (Lower-Bound for Standard Oracle)

$$\text{icomp} = \Omega\left(2^n \cdot d \log\left(\frac{MR}{\rho\varepsilon}\right)\right).$$

- Improvement on existing best bound (by incorporating M, ε).

Lower Bounds

Theorem (Lower-Bound for Standard Oracle)

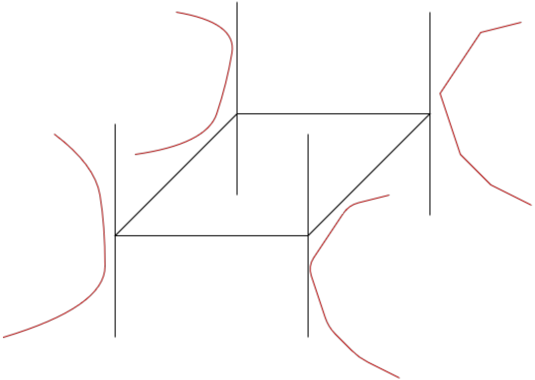
$$\text{icomp} = \Omega\left(2^n \cdot d \log\left(\frac{MR}{\rho\varepsilon}\right)\right).$$

- Improvement on existing best bound (by incorporating M, ε).
- Presents a 2^n "transfer" of continuous case LB to MICO

Lower Bounds

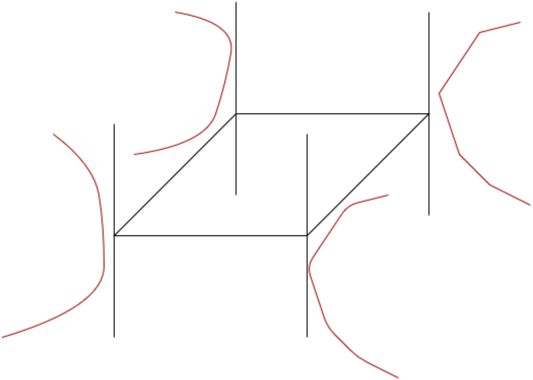
How to think about these lower-bounds...

How to think about these lower-bounds...
Some continuous instance on each integer fiber:



Lower Bounds

How to think about these lower-bounds...
Some continuous instance on each integer fiber:



In unit cube, $2^n \cdot \ell$ means you need to solve all of them.

Lower Bounds

Theorem (Lower-Bound for Standard Oracle)

$$\text{icomp} = \Omega\left(2^n \cdot d \log\left(\frac{MR}{\rho\varepsilon}\right)\right).$$

- Improvement on existing best bound by incorporating M, ε.
- Presents a 2^n "transfer" of continuous case LB to MICO.

Lower Bounds

Theorem (Lower-Bound for Standard Oracle)

$$\text{icomp} = \Omega\left(2^n \cdot d \log\left(\frac{MR}{\rho\varepsilon}\right)\right).$$

- Improvement on existing best bound by incorporating M, ε.
- Presents a 2^n "transfer" of continuous case LB to MICO.

Conjecture (Transferring Lower-Bounds from Continuous to Mixed)

- *Oracle $(\mathcal{G}, \mathcal{H})$ using first-order info*
- *A family of instances with a lower bound $\ell(d, R, \varepsilon, M)$ for the continuous case with this oracle*

Then there exist a family of instances with lower-bound $2^n \cdot \ell(d, R, \varepsilon, M)$ for the general MICO case.

Conjecture proven for pure optimization case!

Proof idea for optimization case of conjecture:

- "Place" one continuous instance on each integer fiber $\mathbf{x} \times \mathbb{R}^d$, $\mathbf{x} \in \{0,1\}^n$.

Lower Bounds

Proof idea for optimization case of conjecture:

- "Place" one continuous instance on each integer fiber $\mathbf{x} \times \mathbb{R}^d$, $\mathbf{x} \in \{0,1\}^n$.
- Want: No query in the full space to give more information than a query on one of the fiber functions

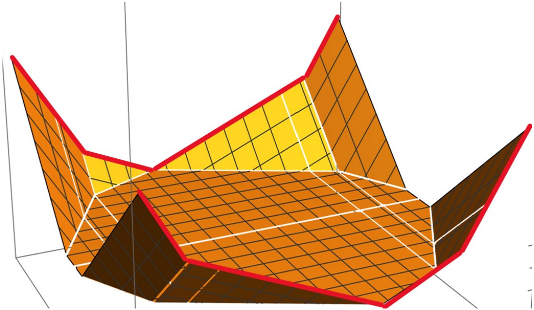
Lower Bounds

Proof idea for optimization case of conjecture:

- "Place" one continuous instance on each integer fiber $\mathbf{x} \times \mathbb{R}^d$, $\mathbf{x} \in \{0,1\}^n$.
- Want: No query in the full space to give more information than a query on one of the fiber functions
  $\rightarrow$ Extend the fiber functions to $\mathbb{R}^{n+d}$ such that any query not on a fiber can be simulated by a query on a fiber

Proof idea for optimization case of conjecture:

- "Place" one continuous instance on each integer fiber $\mathbf{x} \times \mathbb{R}^d$, $\mathbf{x} \in \{0,1\}^n$.
- Want: No query in the full space to give more information than a query on one of the fiber functions
  $\rightarrow$ Extend the fiber functions to $\mathbb{R}^{n+d}$ such that any query not on a fiber can be simulated by a query on a fiber



Construction:

Lower Bounds

Theorem

If H contains only binary functions, with access to oracle $(\mathcal{G}, \mathcal{H})$

$$\text{icomp} = \tilde{\Omega}\left(d^{\frac{8}{7}}\right).$$

Lower Bounds

Theorem

If H contains only binary functions, with access to oracle $(\mathcal{G}, \mathcal{H})$

$$\text{icomp} = \tilde{\Omega}\left(d^{\frac{8}{7}}\right).$$

Uses ideas from recent results on memory-constrained algorithms.

Upper Bounds ($n = 0$ case)

Theorem

- *Instances with function values in* $[-U, U]$
- *Permissible queries* $\mathcal{H}$ *are the bit or directional sign queries*

$$\text{icomp} = O\left(d^2 \log^2\left(\frac{dMR}{\varepsilon}\right)\right) \cdot \log \frac{U}{\varepsilon}$$

Upper Bounds ($n = 0$ case)

Theorem

- *Instances with function values in* $[-U, U]$
- *Permissible queries* $\mathcal{H}$ *are the bit or directional sign queries*

$$\text{icomp} = O\left(d^2 \log^2\left(\frac{dMR}{\varepsilon}\right)\right) \cdot \log\frac{U}{\varepsilon}$$

– Roughly LB under *exact* oracle times $d\log(\frac{dMR}{\varepsilon})$
– ε-approximate gradient contains roughly $d\log(\frac{1}{\varepsilon})$ bits

Conjecture ("The above upper-bound is good")

When permissible queries $\mathcal{H}$ *are binary functions,*

$$\text{icomp} = \Omega\left(d^2 \log\left(\frac{MR}{\varepsilon}\right)^2\right)$$

.

Theorem (Similar result in the general MICO case)

For the general MICO case with binary permissible queries $\mathcal{H}$,

$$\text{icomp} = O\left(2^n d\,(n+d)^2 \log^2\left(\frac{dMR}{\min\{\rho, 1\}\varepsilon}\right)\right)$$

Upper Bounds

Theorem

- **Finitely many** *instances* $\mathcal{I} \subset \mathcal{I}_{n,d,R,\rho,M}$
- $\mathcal{H}$ *for the oracle consisting of all binary functions*

$$icomp = O\left(\log|\mathcal{I}| + d\log\left(\frac{MR}{\rho\varepsilon}\right)\right)$$

Compare to:

$\tilde{\Omega}\left(d^{\frac{8}{7}}\right)$ lower-bound

$\Omega\left(d^2\log\left(\frac{MR}{\varepsilon}\right)^2\right)$ conjectured lower-bound

Upper Bounds

Proof idea for feasibility case:
Maintain a family $\mathcal{U} \subseteq \mathcal{I}$ of the instances that are possible, and a polyhedron P containing C.

Start with $\mathcal{U} = \mathcal{I}$ and $P = [-R, R]^d$.
We will be able to either **reduce** $|\mathcal{U}|$ **or** $\text{vol}(P)$ **by a constant fraction** with each query.

Upper Bounds

Proof idea for feasibility case:
Maintain a family $\mathcal{U} \subseteq \mathcal{I}$ of the instances that are possible, and a polyhedron P containing C.

Start with $\mathcal{U} = \mathcal{I}$ and $P = [-R, R]^d$.
We will be able to either **reduce** $|\mathcal{U}|$ **or** vol(P) **by a constant fraction** with each query.

While $|\mathcal{U}| > 1$ do the following:

- Set $\mathbf{p}$ equal to be the centroid of P. If the separation oracle at $\mathbf{p}$ reports that $\mathbf{p} \in C$, then we return $\mathbf{p}$.

  Otherwise...

Upper Bounds

- **Case 1:** For all $\mathbf{v}$, no more than half the instances $C' \in \mathcal{U}$ give the answer $g_p^{sep}(C') = \mathbf{v}$.

Upper Bounds

- **Case 1:** For all **v**, no more than half the instances $C' \in \mathcal{U}$ give the answer $g_p^{sep}(C') = \mathbf{v}$.

  $\Rightarrow$ there is a set of answers $V \subseteq \mathbb{R}^d$ such that between $\frac{1}{4}|\mathcal{U}|$ and $\frac{3}{4}|\mathcal{U}|$ of the sets give an answer in V.

Upper Bounds

- **Case 1:** For all $\mathbf{v}$, no more than half the instances $C' \in \mathcal{U}$ give the answer $g_p^{sep}(C') = \mathbf{v}$.

  $\Rightarrow$ there is a set of answers $V \subseteq \mathbb{R}^d$ such that between $\frac{1}{4}|\mathcal{U}|$ and $\frac{3}{4}|\mathcal{U}|$ of the sets give an answer in V.
  Query whether the true instance has $g_{\mathbf{p}}^{\text{sep}}(C) \in V$.
  $\Rightarrow$ Size of $\mathcal{U}$ decreases by at least $1/4$.

Upper Bounds

- **Case 1:** For all $\mathbf{v}$, no more than half the instances $C' \in \mathcal{U}$ give the answer $g_p^{sep}(C') = \mathbf{v}$.

  $\Rightarrow$ there is a set of answers $V \subseteq \mathbb{R}^d$ such that between $\frac{1}{4}|\mathcal{U}|$ and $\frac{3}{4}|\mathcal{U}|$ of the sets give an answer in V.
  Query whether the true instance has $g_{\mathbf{p}}^{\text{sep}}(C) \in V$.
  $\Rightarrow$ Size of $\mathcal{U}$ decreases by at least $1/4$.

- **Case 2:** There exists $\bar{\mathbf{v}} \in \mathbb{R}^d$ such that more than half of the instances have $g_{\mathbf{p}}^{\text{sep}}(C') = \bar{\mathbf{v}}$.

Upper Bounds

- **Case 1:** For all $\mathbf{v}$, no more than half the instances $C' \in \mathcal{U}$ give the answer $g_p^{sep}(C') = \mathbf{v}$.

  $\Rightarrow$ there is a set of answers $V \subseteq \mathbb{R}^d$ such that between $\frac{1}{4}|\mathcal{U}|$ and $\frac{3}{4}|\mathcal{U}|$ of the sets give an answer in V.
  Query whether the true instance has $g_{\mathbf{p}}^{\text{sep}}(C) \in V$.
  $\Rightarrow$ Size of $\mathcal{U}$ decreases by at least $1/4$.

- **Case 2:** There exists $\bar{\mathbf{v}} \in \mathbb{R}^d$ such that more than half of the instances have $g_{\mathbf{p}}^{\text{sep}}(C') = \bar{\mathbf{v}}$.
  Query whether the true instance has $g_{\mathbf{p}}^{\text{sep}}(C) = \bar{\mathbf{v}}$.
  $\Rightarrow$ Either size of $\mathcal{U}$ decreases by at least half, or we get an exact separating hyperplane for the true instance to reduce the volume of P by at least $1/e$ (Grünbaum's theorem).

Upper Bounds

- **Case 1:** For all $\mathbf{v}$, no more than half the instances $C' \in \mathcal{U}$ give the answer $g_p^{sep}(C') = \mathbf{v}$.

  $\Rightarrow$ there is a set of answers $V \subseteq \mathbb{R}^d$ such that between $\frac{1}{4}|\mathcal{U}|$ and $\frac{3}{4}|\mathcal{U}|$ of the sets give an answer in V.
  Query whether the true instance has $g_{\mathbf{p}}^{\text{sep}}(C) \in V$.
  $\Rightarrow$ Size of $\mathcal{U}$ decreases by at least $1/4$.

- **Case 2:** There exists $\bar{\mathbf{v}} \in \mathbb{R}^d$ such that more than half of the instances have $g_{\mathbf{p}}^{\text{sep}}(C') = \bar{\mathbf{v}}$.
  Query whether the true instance has $g_{\mathbf{p}}^{\text{sep}}(C) = \bar{\mathbf{v}}$.
  $\Rightarrow$ Either size of $\mathcal{U}$ decreases by at least half, or we get an exact separating hyperplane for the true instance to reduce the volume of P by at least $1/e$ (Grünbaum's theorem).

Reducing $\mathcal{U}$ can only happen $\log(|I|)$ times, and reducing the volume of P can only happen $d\log(\frac{MR}{\min\{\rho,1\}\varepsilon})$ times.

Thank you for your attention!

Questions?