Gijs de Leve prize talk

Geometric Aspects of Linear Programming

Sophie Huiberts

CNRS & LIMOS
Linear Programming

maximize $c^T x$
subject to $Ax \leq b$

we get $A \in \mathbb{R}^{n \times d}$
$b \in \mathbb{R}^n$
$c \in \mathbb{R}^d$

we compute $x \in \mathbb{R}^d$
Linear Programming

maximize $c^T x$
subject to $Ax \leq b$

we get $A \in \mathbb{R}^{n \times d}$
$b \in \mathbb{R}^n$
$c \in \mathbb{R}^d$

we compute $x \in \mathbb{R}^d$
In Practice

The simplex method visits \( 2(n+d) \) vertices before reaching an optimal one.

Only a few documented cases where \( > 10(n+d) \) iterations were performed.

See, e.g., Shamir '87.
Worst-case complexity

Theorem The simplex method has exponential worst-case complexity.

Klee Minty '72
Many, many others '72-'23

*terms and conditions apply
Simplex method is slow in theory but fast in practice.
Average case analysis

**Theorem** There is a simplex method which, if the rows of $A$ are iid uniform from the sphere and $b=1$, visits $d^2 \ln^3$ vertices in expectation.
Average case analysis

Theorem There is a simplex method which, if the rows of $A$ are iid uniform from the sphere and $b=1$, visits $d^2/n$ vertices in expectation.

Theorem If $n \gg 2^d$ then this concentrates around the mean.
Extension to diameter

Theorem

If \( n \gg 2^{d^2} \) then the diameter is

\[ d \cdot n^{\frac{1}{d-1}} \]

and

\[ d^2 \cdot n^{\frac{1}{d-1}} \]

with high probability.
How Realistic is Average Case?

real world photo

random bitmap
Smoothed analysis

Let $\bar{A} \in \mathbb{R}^{n \times d}$ have rows of norm $\leq 1$.

$\bar{b} \in [-1, 1]^n$, $c \in \mathbb{R}^d$

Let $\hat{A}, \hat{b}$ have i.i.d. $N(0, \sigma^2)$ entries.
Smoootted analysis

Let $\bar{A} \in \mathbb{R}^{n \times d}$ have rows of norm $\leq 1$.
$\bar{b} \in [-1, 1]^n$, $c \in \mathbb{R}^d$

Let $\hat{A}, \hat{b}$ have i.i.d. $N(0, \sigma^2)$ entries.

Then

$$\max_{\bar{A}, \bar{b}, c} \mathbb{E}_{\hat{A}, \hat{b}} \left[ \begin{array}{c}
\text{time to solve} \\
\text{maximize } c^T x \\
\text{s.t. } (\bar{A} + \hat{A}) x \leq \bar{b} + \hat{b}
\end{array} \right] \leq \text{poly}(n, d, \sigma^{-1})$$
Why smoothed analysis?

Independent measurement/numerical errors do not conspire against your algorithm.

Interpolate between worst case and average case analysis.

Shows algorithm is fast on average in every large enough neighborhood.
Smoothed analysis results

<table>
<thead>
<tr>
<th></th>
<th>Expected Number of Pivots</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spielman, Teng ’01</td>
<td>$O(n^{86}d^{55}\sigma^{-30})$</td>
</tr>
<tr>
<td>Vershynin ’09</td>
<td>$O(d^3 \log^3 n\sigma^{-4})$</td>
</tr>
<tr>
<td>Dadush, Huiberts ’18</td>
<td>$O(d^2 \sqrt{\log n}\sigma^{-2})$</td>
</tr>
<tr>
<td>Huiberts, Lee, Zhang ’23</td>
<td>$O(d^{13/4} \log^{7/4} n\sigma^{-3/2})$</td>
</tr>
<tr>
<td>Borgwardt ’87</td>
<td>$\Omega(d^{3/2} \sqrt{\log n})$</td>
</tr>
<tr>
<td>Huiberts, Lee, Zhang ’23</td>
<td>$\Omega(\min(2^d, \frac{1}{\sqrt{\sigma d \sqrt{\log n}}}))$</td>
</tr>
</tbody>
</table>
Linear Programming

maximize $c^T x$
subject to $Ax \leq b$

we get $A \in \mathbb{R}^{n \times d}$
$b \in \mathbb{R}^n$
$c \in \mathbb{R}^d$

we compute $x \in \mathbb{R}^d$
Basic IPM analysis

Theorem: An IPM can solve an LP using
\[ \sqrt{n} \cdot L_{A,b,c} \]
linear system solves

\[ L_{A,b,c} : \# \text{ of bits to write down } A,b,c \]
Basic IPMs are scale-invariant

For $D$ positive diagonal, consider

maximize $c^T x$

subject to $DAx \leq Db$

$\rightarrow$ Same feasible region
$\rightarrow$ Predictable change to central path
$\rightarrow$ Same change to IPM path
Sophisticated IPM analysis

Theorem  Specific IPMs can solve an LP using

\[ \text{poly}(n) \cdot L_A \]

linear system solves

\[ L_A : \text{# of bits to write down } A \]
Question: Can we have both properties for a single IPM?
Question: Can we have both properties for a single IPM?

Answer: Yes
Separation oracle

\[ k \in \mathbb{R}^d \text{ unknown convex.} \]

\[ z + r B_2^d \leq k \leq r B_2^d \] with \( z, r \) unknown.

Can ask: is \( x \in k \)?

- Yes
- No because:
  \[ a^\top x > b \]
  \[ a^\top y \leq b \text{ for all } y \in k \]
Gradient oracle

\( f : \mathbb{R}^d \rightarrow \mathbb{R} \) unknown convex

\( \| \nabla f(y) \| \leq L \) for all \( y \in \mathcal{K} \)

Can query \( x \in \mathbb{R}^d \):

- value \( f(x) \)
- gradient \( \nabla f(x) \)
We want to minimize $f(x)$ over $x \in k$

with a small # of queries.
Ellipsoid method

1. Have an ellipsoid containing all optimal points
2. Query center point
3. if not in $k$:
   find smaller ellipsoid using cut
4. if in $k$:
   find smaller ellipsoid using gradient
5. go to 1.
Ellipsoid method

**Pros**
- Convergence guarantee

**Cons**
- Slow in practice
LP based loop

1. have a set $S$ of valid constraints.

2. solve minimize $c^T x$
   subject to $a^T x \leq b$
   for all $(a,b) \in S$

3. query LP optimal solution

4. if not in $K$
   add $(a,b)$ to $S$
LP based loop

Pros
- mostly fast in practice
- LP solutions are cheap

Cons
- no convergence guarantee
- only lower bounds, no feasible points
Our new algorithm

1. have a set $S$ of valid constraints.
2. solve a quadratic program
3. query its solution $x^*$
4. if not in $K$
   add $(a,b)$ to $S$
5. if in $K$
   add $(\nabla f(x^*), \langle \nabla f(x^*), x^* \rangle)$ to $S$
Our new algorithm

**Pros**
- Convergence guarantee
- Good in experiments

**Cons**
- Need to solve a quadratic program every round
Experimental results

![Graph showing experimental results with different methods and their performance over iterations.](image-url)
Linear programming
- diameter of random polyhedra
- smoothed analysis of simplex method
- scale-invariant IPM

Oracle model
- new cutting plane method