Linear Algebra

ETH Zürich, HS 2024, 401-0131-00L

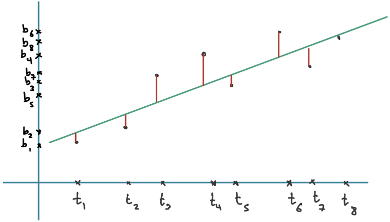
The Computer Science Lens

Linear Programming

Bernd Gärtner

November 14, 2024

Linear Regression, reloaded



With least squares and linear algebra, we can find a line that minimizes the sum of *quadratic* errors,

$$\min_{\alpha_0,\alpha_1} \sum_{k=1}^{m} (b_k - (\alpha_0 + \alpha_1 t_k))^2$$

Outliers (points far up or down) have too much influence: large errors get "inflated" by squaring them. Why not minimize the sum of *absolute* errors?

$$\min_{\alpha_0,\alpha_1} \sum_{k=1}^{m} |b_k - (\alpha_0 + \alpha_1 t_k)|$$

Linear Regression with absolute errors

A line that minimizes the sum of *absolute* errors

$$\min_{\alpha_0,\alpha_1} \sum_{k=1}^{m} |b_k - (\alpha_0 + \alpha_1 t_k)|$$

can be found efficiently...

- ...but this is much more complicated (and much slower) than least squares
- ...and uses a technique that is known to be efficient only since 1979!
- In contrast, least squares is well-understood since the 18th century.

The 20th century technique is called **Linear Programming**.

Observation: the sum of *absolute* errors can be minimized by linear programming [MG07, Section 2.4 + Section 6.1].

Linear programming: from linear equations... to linear inequalities

Problem: Solve $A\mathbf{x} = \mathbf{b}$! **Algorithm:** Gauss-Jordan elimination

Problem: Solve $A\mathbf{x} \leq \mathbf{b}$!

Example:

$$\begin{aligned} x_1 &\geq 0 \\ x_2 &\geq 0 \\ x_2 - x_1 &\leq 1 \\ x_1 + 6x_2 &\leq 15 \\ 4x_1 - x_2 &\leq 10 \end{aligned}$$

$$A = \begin{bmatrix} -1 & 0 \\ 0 & -1 \\ -1 & 1 \\ 1 & 6 \\ 4 & -1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 15 \\ 10 \end{bmatrix}$$

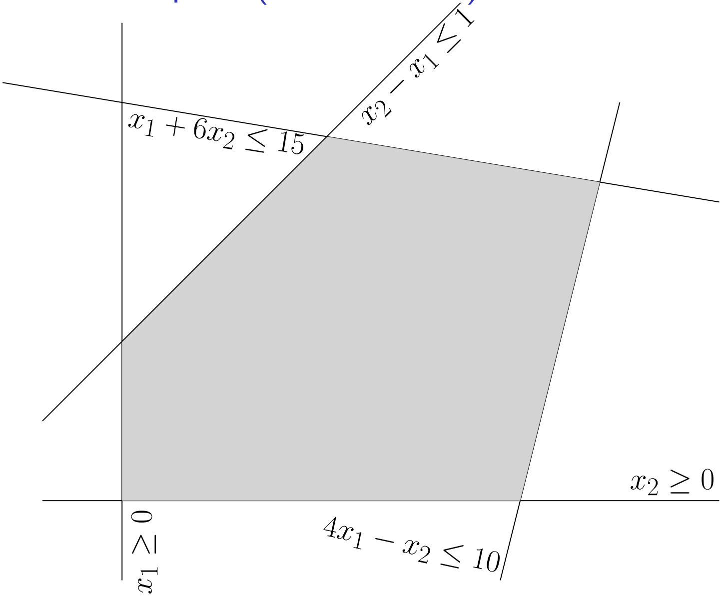
$$\begin{array}{rlrlcr} -1 & x_1 & +0 & x_2 & \leq & 0 \\ 0 & x_1 & -1 & x_2 & \leq & 0 \\ -1 & x_1 & +1 & x_2 & \leq & 1 \\ 1 & x_1 & +6 & x_2 & \leq & 15 \\ 4 & x_1 & -1 & x_2 & \leq & 10 \end{array}$$

This is called (the decision version) of **Linear Programming** and has many important applications [MG07].

Optimization version (used for regression with absolute errors): find solution of $A\mathbf{x} \leq \mathbf{b}$ that maximizes a linear function $\mathbf{c}^\top\mathbf{x}$! Reduces to decision version [MG07, Section 6.1].

Algorithm: ?

The solution space (subset of $\mathbb{R}^2$)



Intersection of five **halfplanes**

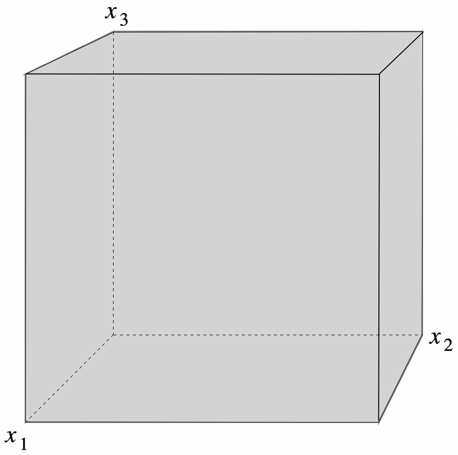
.

Halfplane: everything on one side of a line, including the line

The solution space (subset of $\mathbb{R}^3$)

$$
\begin{aligned}
0 &\leq x_1 \leq 1 \\
0 &\leq x_2 \leq 1 \\
0 &\leq x_3 \leq 1
\end{aligned}
$$

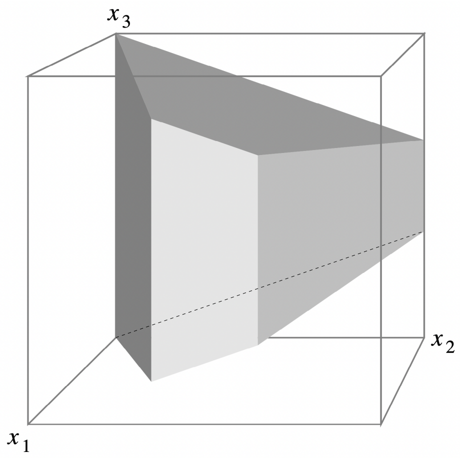
Intersection of six **halfspaces**



Unit cube

The solution space (subset of $\mathbb{R}^3$)

$$
\begin{aligned}
0 &\leq x_1 \leq 1 \\
\tfrac{1}{3}x_1 &\leq x_2 \leq 1 - \tfrac{1}{3}x_1 \\
\tfrac{1}{3}x_2 &\leq x_3 \leq 1 - \tfrac{1}{3}x_2
\end{aligned}
$$



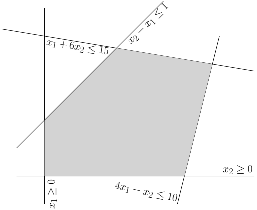
Klee-Minty Cube

The solution space (subset of $\mathbb{R}^n$)

A an $m \times n$ matrix, $b \in \mathbb{R}^m$

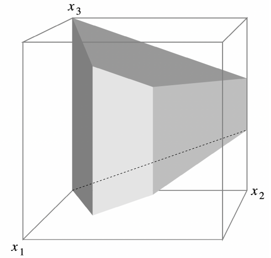
Solutions of $A\mathbf{x} \leq \mathbf{b}$: intersection of m halfspaces in $\mathbb{R}^n$, a **convex polyhedron**

$\mathbb{R}^2$



Convex polygon

$\mathbb{R}^3$



Convex polyhedron

$\mathbb{R}^n$

Beast

Taming the beast

Title Page, New York Times, November 7, 1979 (bottom of the page)

The New York Times / Teresa Zabala

President Carter after announcing that Robert S. Strauss, right, would head his campaign. At left are Secretary of State Cyrus R. Vance and Sol M. Linowitz, who is succeeding Mr. Strauss as Mideast negotiator.

A Soviet Discovery Rocks World of Mathematics

By MALCOLM W. BROWNE

A surprise discovery by an obscure Soviet mathematician has rocked the world of mathematics and computer analysis, and experts have begun exploring its practical applications.

Mathematicians describe the discovery by L.G. Khachian as a method by which computers can find guaranteed solutions to a class of very difficult problems that have hitherto been tackled on a kind of hit-or-miss basis.

Apart from its profound theoretical interest, the discovery may be applicable in weather prediction, complicated industrial processes, petroleum refining, the scheduling of workers at large factories, secret codes and many other things.

"I have been deluged with calls from virtually every department of government for an interpretation of the significance of this," a leading expert on computer methods, Dr. George B. Dantzig of Stanford University, said in an interview.

The solution of mathematical problems by computer must be broken down into a series of steps. One class of problem sometimes involves so many steps that it could take billions of years to compute.

The Russian discovery offers a way by which the number of steps in a solution can be dramatically reduced. It also offers the mathematician a way of learning quickly whether a problem has a solution or not, without having to complete the entire immense computation that may be required.

According to the American journal Sci-

Continued on Page A20, Column 3

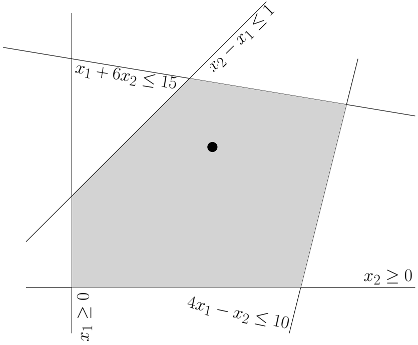
Taming the beast, in theory and in practice

- Leonid Khachiyan, 1979: Linear Programming can in theory be solved efficiently.
- However, Khachiyan's algorithm (the ellipsoid method) is useless in practice.
- George Dantzig, 1940's: Linear Programming can in practice be solved efficiently.
- However, Dantzig's algorithm (the simplex method) is useless in theory.
- What this means: It is not known whether there is a variant of the simplex method that is *provably* efficient.
- This is ongoing research...

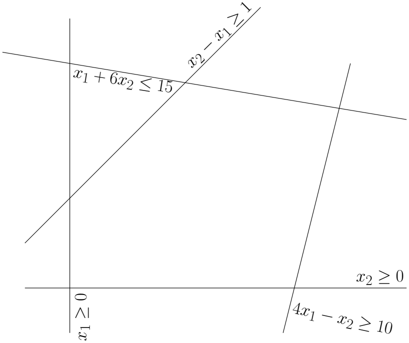
Linear inequalities, geometrically

Problem, algebraically: Solve $A\mathbf{x} \leq \mathbf{b}$!

Problem, geometrically: Find a point in a convex polyhedron, or conclude that it is empty!



point in convex polyhedron = solution

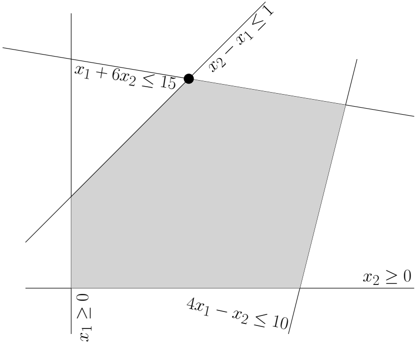


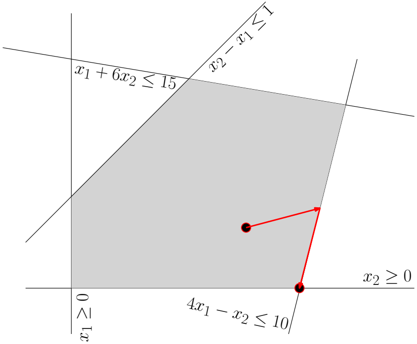
convex polyhedron empty, no solution

Finding a point in a convex polyhedron...

...is "the same" as finding a corner.[1]

If we have a corner, we have a point, and if we have a point, we can easily find a corner (walk until we hit a wall, walk inside the wall until we hit another wall,...)

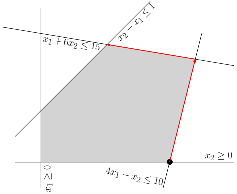




[1]For this, we need to assume that the convex polyhedron is bounded, but this is no problem.

Solving $A\mathbf{x} \leq \mathbf{b}$

Consider the inequalities

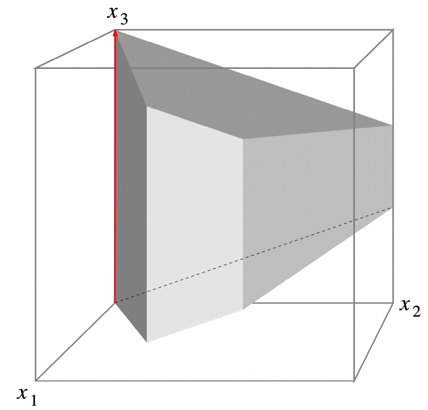


$$A\mathbf{x} + x_{n+1} \cdot \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \leq \mathbf{b}, \quad x_{n+1} \leq 0.$$

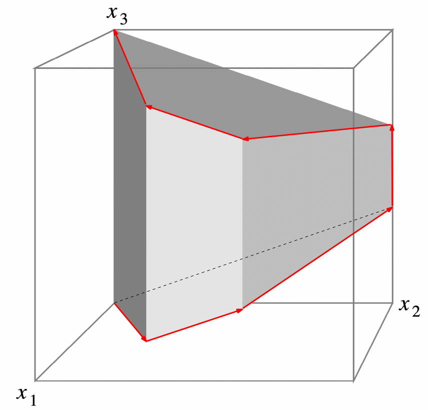
- The corresponding convex polyhedron in $\mathbb{R}^{n+1}$ is nonempty, and a point in it can be found easily (set $\mathbf{x} = \mathbf{0}$ and make x_{n+1} small enough).
- From this point, find a corner, as described before.
- **From this corner, "climb up" along edges to the highest corner (= highest point, the one with largest x_{n+1}-value).**
- If the highest corner has $x_{n+1} = 0$, we have solved $A\mathbf{x} \leq \mathbf{b}$, otherwise, $A\mathbf{x} \leq \mathbf{b}$ has no solution.

This is George Dantzig's **simplex method** from the 1940's [Dan63].

Depending on the climbing rule, the simplex method...



...can be very fast...



...or very slow.

n-dimensional Klee-Minty cube: a natural climbing rule visits all 2^n corners [KM72].

Runtime of the simplex method

- Fast in practice
- But: For every climbing rule that people have developed, there are (artificially constructed) beasts on which climbing takes very long when this rule is used.

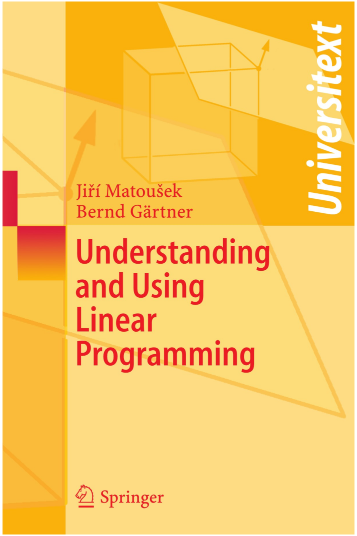
Open problem: Is there a climbing rule which climbs *every* beast quickly?

A positive answer would solve Smale's 9th problem for the 21st century:

`https://en.wikipedia.org/wiki/Smale%27s_problems`

- There are randomized climbing rules (using coin flips) which are (in expectation) faster than the known deterministic ones (no coin flips).
- One concrete result here: There is a rule that climbs every n-dimensional cube in at most $e^{2\sqrt{n}}$ steps in expectation (much better than the worst case 2^n) [Gär02].

More on linear programming



- VL *Algorithms, Probability, and Computing* (Kernfach, Vertiefung Theoretische Informatik)

References

G. B. Dantzig.
Linear Programming and Extensions.
Princeton University Press, Princeton, NJ, 1963.

B. Gärtner.
The Random-Facet simplex algorithm on combinatorial cubes.
Random Structures & Algorithms, 20(3), 2002.
https://doi.org/10.1002/rsa.10034.

V. Klee and G. J. Minty.
How good is the simplex algorithm?
In O. Shisha, editor, *Inequalities III*, pages 159–175. Academic Press, 1972.

J. Matoušek and B. Gärtner.
Understanding and Using Linear Programming.
Springer, 2007.
https://doi.org/10.1007/978-3-540-30717-4.