NTNU | Norwegian University of Science and Technology

TTK4135 – Lecture 4
2nd order optimality conditions
Linear programming

Lecturer: Lars Imsland

Purpose of Lecture

- Finishing optimality conditions: 2nd order
- Brief recap: linear algebra
- Linear programming (LP): formulation and standard form
- KKT conditions for LP
- Dual LP, weak & strong duality

Reference: Chapter 13.1 (12.9) in N&W (Linear algebra: App A.1)

OPTIMALITY CONDITIONS

KKT Conditions (Thm 12.1)

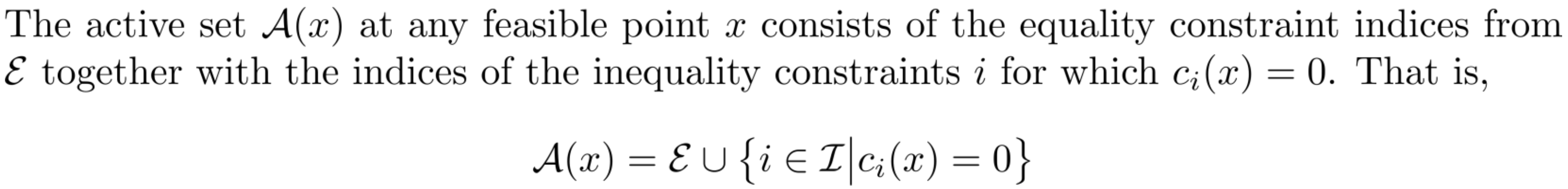
$$\min_{x \in \mathbb{R}^n} f(x) \qquad \text{subject to} \quad \begin{aligned} c_i(x) &= 0, \quad i \in \mathcal{E}, \\ c_i(x) &\geq 0, \quad i \in \mathcal{I}. \end{aligned}$$

Lagrangian: $$\mathcal{L}(x, \lambda) = f(x) - \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i c_i(x)$$

KKT-conditions (First-order necessary conditions): If x^* is a local solution and LICQ holds, then there exist λ^* such that

$$\begin{aligned}
\nabla_x \mathcal{L}(x^*, \lambda^*) &= 0, & & \qquad \text{(stationarity)} \\
\left.\begin{aligned} c_i(x^*) &= 0, \quad \forall i \in \mathcal{E}, \\ c_i(x^*) &\geq 0, \quad \forall i \in \mathcal{I}, \end{aligned}\right\} & & & \qquad \text{(primal feasibility)} \\
\lambda_i^* &\geq 0, & \forall i \in \mathcal{I}, & \qquad \text{(dual feasibility)} \\
\lambda_i^* c_i(x^*) &= 0, & \forall i \in \mathcal{E} \cup \mathcal{I}. & \qquad \text{(complementarity condition/ complementary slackness)}
\end{aligned}$$

Active Set



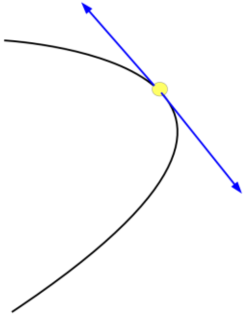
The active set $\mathcal{A}(x)$ at any feasible point x consists of the equality constraint indices from $\mathcal{E}$ together with the indices of the inequality constraints i for which $c_i(x) = 0$. That is,

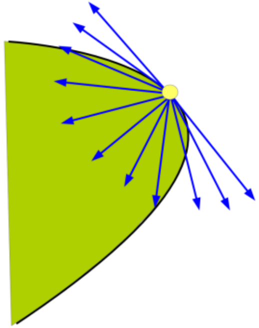
$$\mathcal{A}(x) = \mathcal{E} \cup \left\{ i \in \mathcal{I} \middle| c_i(x) = 0 \right\}$$

Set of (linearized) Feasible Directions

Given a feasible point x and the active constraint set $\mathcal{A}(x)$, the set of linearized feasible directions $\mathcal{F}(x)$ is

$$\mathcal{F}(x) = \left\{ d \;\middle|\; \begin{array}{ll} d^\top \nabla c_i(x) = 0, & \text{for all } i \in \mathcal{E}, \\ d^\top \nabla c_i(x) \geq 0, & \text{for all } i \in \mathcal{A}(x) \cap \mathcal{I} \end{array} \right\}$$





2^{nd} Order Conditions: Critical Cone

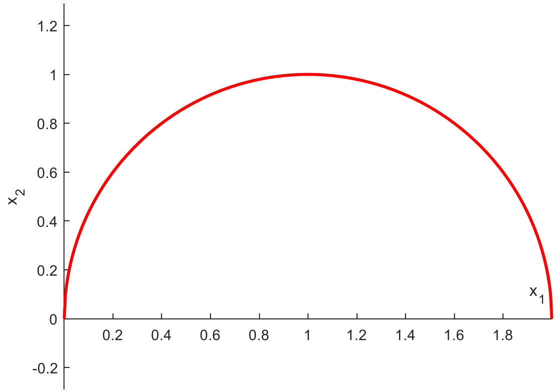
- We have found a point x^* that fulfills KKT conditions
- Say there are directions $w \in \mathcal{F}(x^*)$ that does not lead to an increase in the objective function, that is $w^\mathsf{T} \nabla f(x^*) = 0,\ w \neq 0$. How do we decide whether x^* is actually a minimum?
- Second-order conditions answer this by looking at the curvature (2^{nd} derivative) in these directions

- Define the *critical cone*:

$$w \in \mathcal{C}(x^*, \lambda^*) \Leftrightarrow \begin{cases} \nabla c_i(x^*)^\top w = 0, & \forall i \in \mathcal{E}, \\ \nabla c_i(x^*)^\top w = 0, & \forall i \in \mathcal{A}(x^*) \cap \mathcal{I} \text{ with } \lambda_i^* > 0, \\ \nabla c_i(x^*)^\top w \geq 0, & \forall i \in \mathcal{A}(x^*) \cap \mathcal{I} \end{cases}$$

- Note: $\mathcal{C}(x^*, \lambda^*) \subseteq \mathcal{F}(x^*)$. Difference: Inequalities with positive Lagrange multiplier treated as equalities
- $\mathcal{C}(x^*, \lambda^*)$ contains the “undecided” directions from $\mathcal{F}(x^*)$, the directions where decrease/increase cannot be decided from $\nabla f(x^*)$ alone

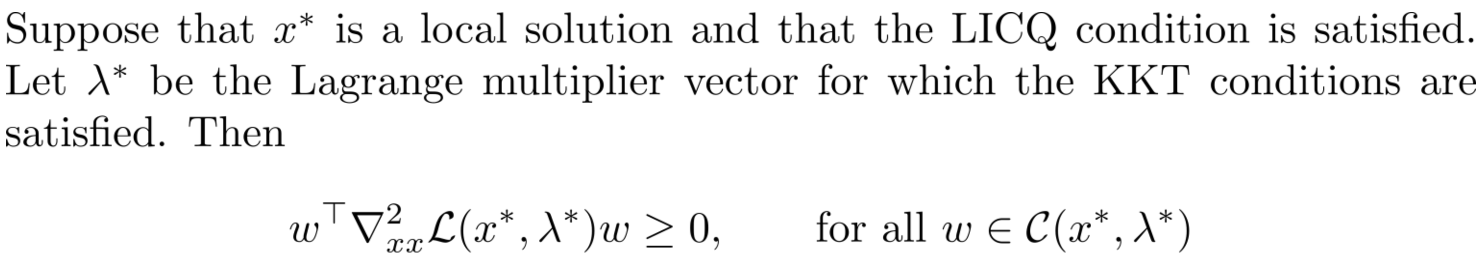
Critical cone Ex.



$$\begin{aligned} \min_{x\in\mathbb{R}^2} \quad & x_1 \\ \text{s.t.} \quad & c_1(x) = x_2 \geq 0 \\ & c_2(x) = -(x_1-1)^2 - x_2^2 + 1 \geq 0 \end{aligned}$$

2^{nd} Order Conditions: Necessary & Sufficient

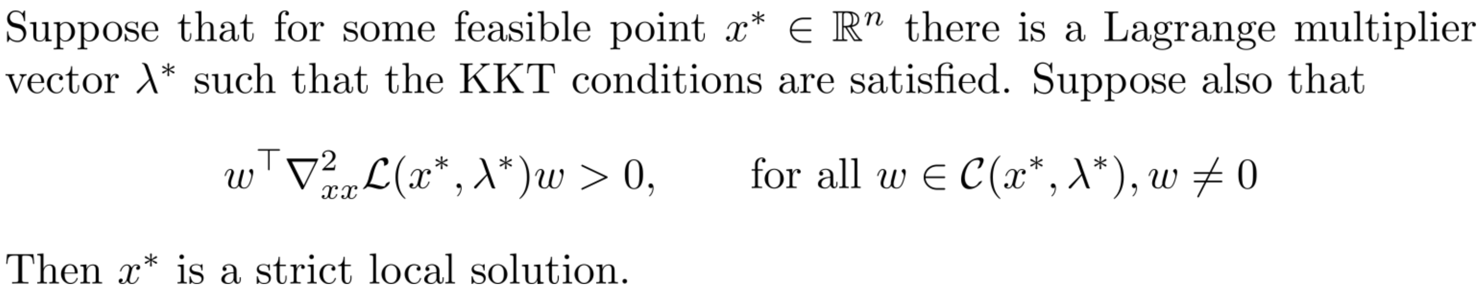
- Second-order necessary conditions (Theorem 12.5):



Suppose that x^* is a local solution and that the LICQ condition is satisfied. Let λ^* be the Lagrange multiplier vector for which the KKT conditions are satisfied. Then

$$w^\top \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) w \geq 0, \qquad \text{for all } w \in \mathcal{C}(x^*, \lambda^*)$$

- Second-order sufficient conditions (Theorem 12.6):

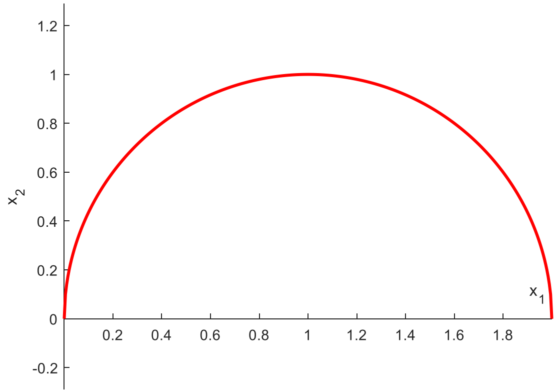


Suppose that for some feasible point $x^* \in \mathbb{R}^n$ there is a Lagrange multiplier vector λ^* such that the KKT conditions are satisfied. Suppose also that

$$w^\top \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) w > 0, \qquad \text{for all } w \in \mathcal{C}(x^*, \lambda^*), w \neq 0$$

Then x^* is a strict local solution.

2nd order cond., Ex.



$$\begin{aligned} \min_{x\in\mathbb{R}^2} \quad & x_1 \\ \text{s.t.} \quad & c_1(x) = x_2 \geq 0 \\ & c_2(x) = -(x_1-1)^2 - x_2^2 + 1 \geq 0 \end{aligned}$$

BRIEF RECAP: LINEAR ALGEBRA

Norms

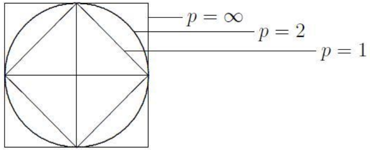
```
>> help norm
 norm   Matrix or vector norm.
     norm(X,2) returns the 2-norm of X.
     norm(X) is the same as norm(X,2).
     norm(X,1) returns the 1-norm of X.
     norm(X,Inf) returns the infinity norm of X.
     norm(X,'fro') returns the Frobenius norm of X.
```

Vector norm: A mapping $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}^+$ that satisfies

- $\|x\| = 0 \Rightarrow x = 0$
- $\|x + z\| \leq \|x\| + \|z\|$, for all $x, z \in \mathbb{R}^n$
- $\|\alpha x\| = |\alpha|\|x\|$, for all $\alpha \in \mathbb{R}$ and $x \in \mathbb{R}^n$

Common norms (p-norms):

- 1-norm: $\|x\|_1 = |x_1| + |x_2| + \ldots + |x_n|$ (sum norm)
- 2-norm: $\|x\|_2 = \sqrt{x_1^2 + x_2^2 + \ldots + x_n^2}$ (Euclidean norm)
- ∞-norm: $\|x\|_\infty = \max_{i=1,\ldots,n} |x_i|$ (max norm)



$\|x\|_p = 1$

Induced matrix norms, $A \in \mathbb{R}^{m\times n}$: $\|A\| := \sup_{x\neq 0} \frac{\|Ax\|}{\|x\|}$

- 1-norm: $\|A\|_1 = \max_{j=1,\ldots,n} \sum_{i=1}^m |A_{ij}|$
- 2-norm: $\|A\|_2 = \lambda_{\max}(\sqrt{A^\top A})$
- ∞-norm: $\|A\|_\infty = \max_{i=1,\ldots,m} \sum_{j=1}^n |A_{ij}|$

$$\begin{array}{c} \begin{array}{cc} \begin{pmatrix} 4 & -1 & 2 & 0 \\ 0 & 3 & 0 & 2 \\ -2 & 1 & 5 & 1 \end{pmatrix} & \begin{matrix} 7 \\ 5 \\ \boxed{9} \;\; \|A\|_\infty \end{matrix} \end{array} \\ \begin{matrix} 6 & \;\; 5 & \;\; \boxed{7} & \;\; 3 \end{matrix} \qquad \\ \|A\|_1 \end{array}$$

Other matrix norms, not induced:

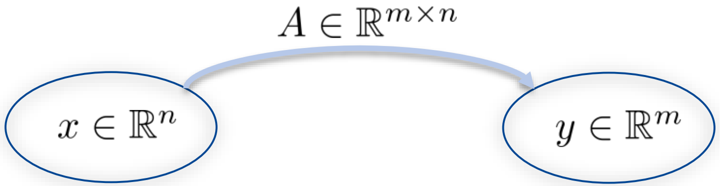
- Frobenius-norm $\|A\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n A_{ij}^2}$

Useful property, induced matrix norms:

- $\|Ax\| \leq \|A\|\|x\|$

Fundamental Theorem of Linear Algebra

A matrix $A \in \mathbb{R}^{m \times n}$ is a mapping:

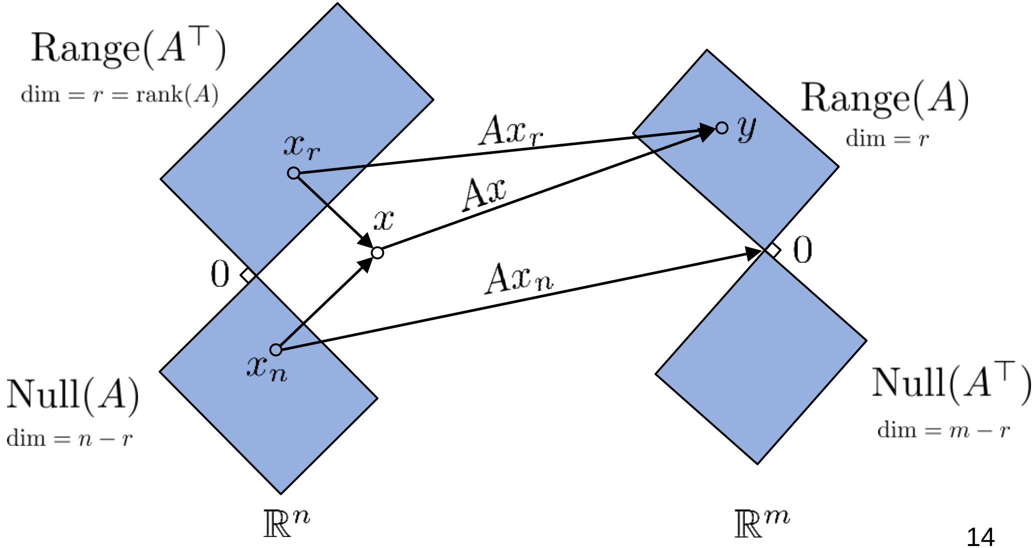


Nullspace of A: $\text{Null}(A) = \{v \in \mathbb{R}^n \mid Av = 0\}$

Rangespace (columnspace) of A: $\text{Range}(A) = \{w \in \mathbb{R}^m \mid w = Av, \text{ for some } v \in \mathbb{R}^n\}$

Fundamental theorem of linear algebra:

$$\boxed{\text{Null}(A) \oplus \text{Range}(A^\top) = \mathbb{R}^n}$$



Matrix Factorizations

```
>> help \
 \   Backslash or left matrix divide.
    A\B is the matrix division of A into B, which is roughly the
    same as INV(A)*B , except it is computed in a different way.
    If A is an N-by-N matrix and B is a column vector with N
    components, or a matrix with several such columns, then
    X = A\B is the solution to the equation A*X = B. A warning
    message is printed if A is badly scaled or nearly singular.
```

“All” algorithms in this course involve solving linear equation systems:

$$Ax = b \quad \Rightarrow x = A^{-1}b$$

In practice, **never** use the matrix inverse. It is inefficient and numerically unstable.

Instead, use matrix factorizations:

- General matrix *A*: Use LU-decomposition (*Gaussian elimination*)

$$A = LU: \quad Ax = L\underbrace{Ux}_{y} = b \quad \Rightarrow \quad Ly = b \quad \Rightarrow \quad Ux = y$$

  - Due to triangular structure of *L* and *U*, we easily solve the two linear systems by substitution
- Symmetric positive definite matrix *A*: Use Cholesky decomposition

$$A = LL^\top$$

  - Half the cost of LU. Solve system as for LU.
  - For symmetric, indefinite matrices: Use LDL-factorization instead (book: $A = LBL^\top$)
- Generally, algorithms use permutations:

$$PA = LU, \quad PAP^\top = LL^\top$$

- Other important factorizations
  - *A* = *QR*: Finds orthogonal (orthonormal) basis for rangespace of *A* and nullspace of A^T
  - Eigenvalue (spectral) decomposition, singular value decomposition

Condition Number of a Matrix (when solving *Ax* = *b*)

- Well-conditioned: Small perturbations give small changes in solution:

$$\begin{pmatrix}1 & 2\\ 1 & 1\end{pmatrix}\begin{pmatrix}x_1\\ x_2\end{pmatrix} = \begin{pmatrix}3\\ 2\end{pmatrix} \quad\Rightarrow\quad \begin{pmatrix}x_1\\ x_2\end{pmatrix} = \begin{pmatrix}1\\ 1\end{pmatrix}$$

$$\begin{pmatrix}1 & 2\\ 1 & 1\end{pmatrix}\begin{pmatrix}x_1\\ x_2\end{pmatrix} = \begin{pmatrix}3.00001\\ 2\end{pmatrix} \quad\Rightarrow\quad \begin{pmatrix}x_1\\ x_2\end{pmatrix} = \begin{pmatrix}0.99999\\ 1.00001\end{pmatrix}$$

- Ill-conditioned: Small perturbations give large changes in solution:

$$\begin{pmatrix}1.00001 & 1\\ 1 & 1\end{pmatrix}\begin{pmatrix}x_1\\ x_2\end{pmatrix} = \begin{pmatrix}2.00001\\ 2\end{pmatrix} \quad\Rightarrow\quad \begin{pmatrix}x_1\\ x_2\end{pmatrix} = \begin{pmatrix}1\\ 1\end{pmatrix}$$

$$\begin{pmatrix}1.00001 & 1\\ 1 & 1\end{pmatrix}\begin{pmatrix}x_1\\ x_2\end{pmatrix} = \begin{pmatrix}2\\ 2\end{pmatrix} \quad\Rightarrow\quad \begin{pmatrix}x_1\\ x_2\end{pmatrix} = \begin{pmatrix}0\\ 2\end{pmatrix}$$

- Condition number:

$$\boxed{\kappa(A) = \|A\|\|A^{-1}\|}$$

  - A small condition number (say, 1-100) implies the matrix is well-conditioned, a large condition number (say, >10 000) implies the matrix is ill-conditioned.

    The condition number (2-norm) of the above matrices are 6.9 and 400 000, respectively.

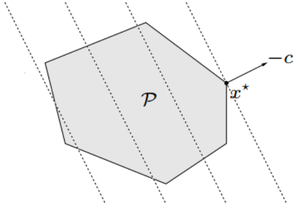
```
>> help cond
 cond   Condition number with respect to inversion.
    cond(X) returns the 2-norm condition number (the
    ratio of the largest singular value of X to the smallest).
    Large condition numbers indicate a nearly singular
    matrix.

    cond(X,P) returns the condition number of X in P-
norm:

       NORM(X,P) * NORM(INV(X),P).

    where P = 1, 2, inf, or 'fro'.
```

LINEAR PROGRAMMING

Types of Constrained Optimization Problems

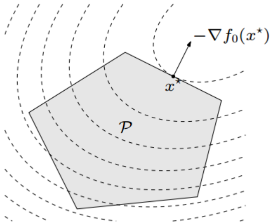
- Linear programming
  - Convex problem
  - Feasible set polyhedron

$$\begin{array}{ll} \text{min} & c^\mathsf{T} x \\ \text{subject to} & Ax \le b \\ & Cx = d \end{array}$$



- Quadratic programming
  - Convex problem if $P \ge 0$
  - Feasible set polyhedron

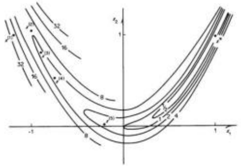
$$\begin{array}{ll} \text{min} & \frac{1}{2} x^\mathsf{T} P x + q^\mathsf{T} x \\ \text{subject to} & Ax \le b \\ & Cx = d \end{array}$$



- Nonlinear programming
  - In general non-convex!

$$\begin{array}{ll} \text{min} & f(x) \\ \text{subject to} & g(x) = 0 \\ & h(x) \ge 0 \end{array}$$

$f(x) = 100\,(x_2 - x_1^2)^2 + (1 - x_1)^2$

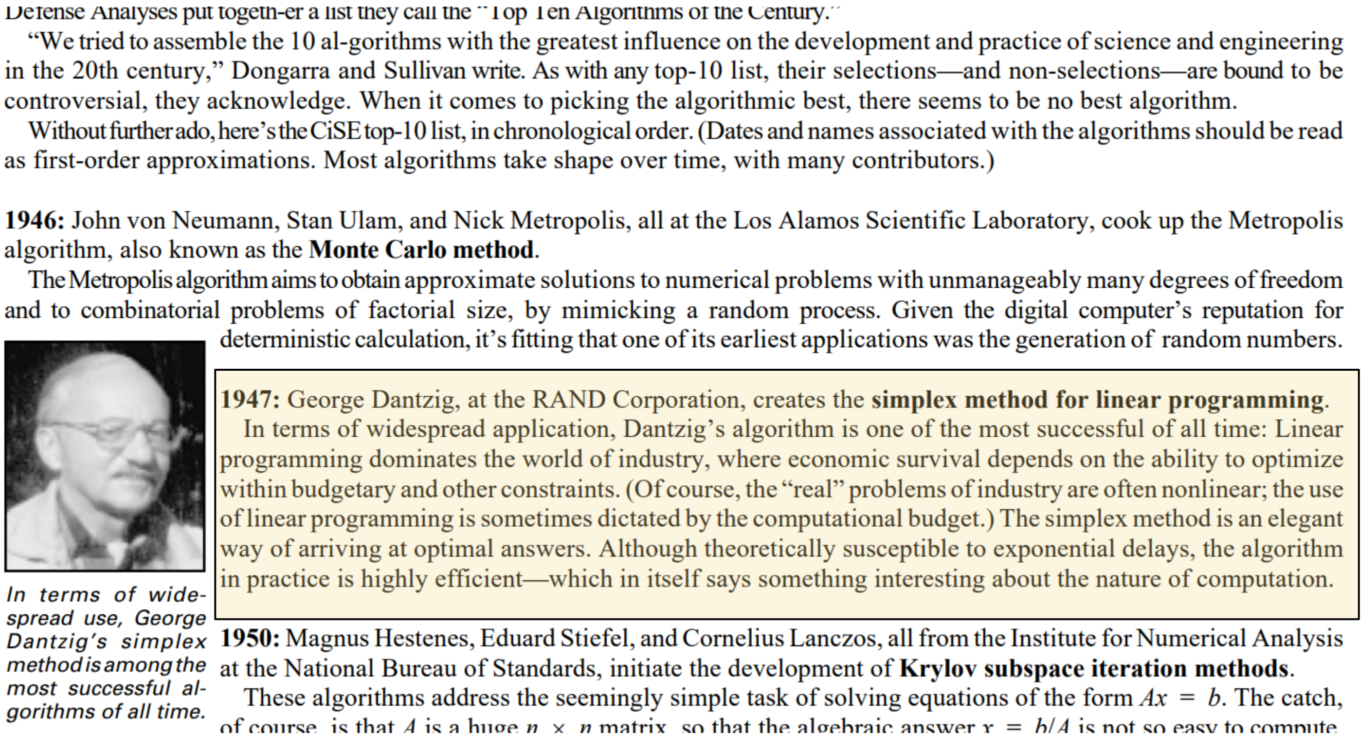


$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{subject to} \quad \begin{array}{ll} c_i(x) = 0, & i \in \mathcal{E}, \\ c_i(x) \ge 0, & i \in \mathcal{I}. \end{array}$$

from *SIAM News*, Volume 33, Number 4

The Best of the 20th Century: Editors Name Top 10 Algorithms

By Barry A. Cipra



Defense Analyses put togeth-er a list they call the "Top Ten Algorithms of the Century."

"We tried to assemble the 10 al-gorithms with the greatest influence on the development and practice of science and engineering in the 20th century," Dongarra and Sullivan write. As with any top-10 list, their selections—and non-selections—are bound to be controversial, they acknowledge. When it comes to picking the algorithmic best, there seems to be no best algorithm.

Without further ado, here's the CiSE top-10 list, in chronological order. (Dates and names associated with the algorithms should be read as first-order approximations. Most algorithms take shape over time, with many contributors.)

1946: John von Neumann, Stan Ulam, and Nick Metropolis, all at the Los Alamos Scientific Laboratory, cook up the Metropolis algorithm, also known as the **Monte Carlo method**.

The Metropolis algorithm aims to obtain approximate solutions to numerical problems with unmanageably many degrees of freedom and to combinatorial problems of factorial size, by mimicking a random process. Given the digital computer's reputation for deterministic calculation, it's fitting that one of its earliest applications was the generation of random numbers.

1947: George Dantzig, at the RAND Corporation, creates the **simplex method for linear programming**.

In terms of widespread application, Dantzig's algorithm is one of the most successful of all time: Linear programming dominates the world of industry, where economic survival depends on the ability to optimize within budgetary and other constraints. (Of course, the "real" problems of industry are often nonlinear; the use of linear programming is sometimes dictated by the computational budget.) The simplex method is an elegant way of arriving at optimal answers. Although theoretically susceptible to exponential delays, the algorithm in practice is highly efficient—which in itself says something interesting about the nature of computation.

In terms of widespread use, George Dantzig's simplex method is among the most successful algorithms of all time.

1950: Magnus Hestenes, Eduard Stiefel, and Cornelius Lanczos, all from the Institute for Numerical Analysis at the National Bureau of Standards, initiate the development of **Krylov subspace iteration methods**.

These algorithms address the seemingly simple task of solving equations of the form $Ax = b$. The catch, of course, is that A is a huge $n \times n$ matrix, so that the algebraic answer $x = b/A$ is not so easy to compute.

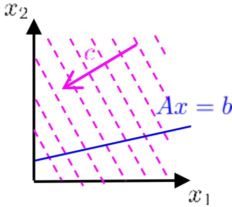
Standard form LP: $\min_{x \in \mathbb{R}^n} c^\top x \quad \text{subject to} \quad \begin{cases} Ax = b \\ x \geq 0 \end{cases}$

LPs does not always have a solution

$$\min_{x\in\mathbb{R}^n} c^\top x \quad \text{subject to} \quad \begin{cases} Ax = b \\ x \geq 0 \end{cases}$$

There are two sources for no solution:

1. Infeasibility: $Ax = b$ has no solution (the feasible set Ω is empty)
2. Unboundedness: There exists a sequence $x^k \in \Omega$ such that $c^\top x^k \to -\infty$



We can assume without loss of generality that $A \in \mathbb{R}^{m\times n}$, $m < n$

(If $m \geq n$, the problem is either infeasible or can be transformed to an equivalent problem with $m < n$)

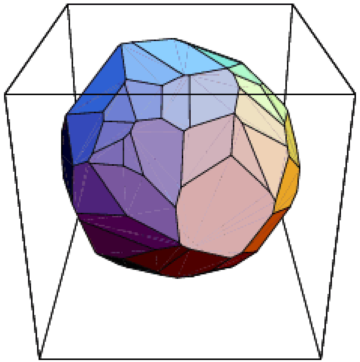
Linear Programming Solutions

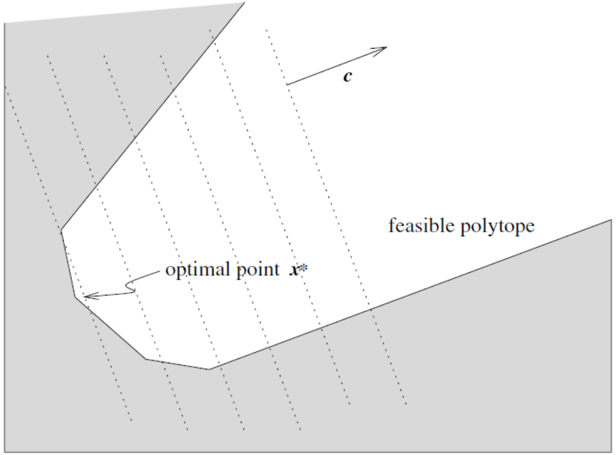
The feasible set is a polytope (a convex set with flat faces)

The objective function contours are planar

Three possible cases for solutions:

- No solutions: Feasible set is empty or problem is unbounded
- One solution: A vertex
- Infinite number of solutions: An "edge" is a solution





KKT Conditions for LPs

$$\min_{x \in \mathbb{R}^n} c^\top x \quad \text{subject to} \quad \begin{cases} Ax = b \\ x \geq 0 \end{cases}$$

Lagrangian:

$$\mathcal{L}(x, \lambda) = f(x) - \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i c_i(x)$$

KKT-conditions:

$$\begin{aligned} \nabla_x \mathcal{L}(x^*, \lambda^*) &= 0, \\ c_i(x^*) &= 0, \quad \forall i \in \mathcal{E}, \\ c_i(x^*) &\geq 0, \quad \forall i \in \mathcal{I}, \\ \lambda_i^* &\geq 0, \quad \forall i \in \mathcal{I}, \\ \lambda_i^* c_i(x^*) &= 0, \quad \forall i \in \mathcal{E} \cup \mathcal{I}. \end{aligned}$$

KKT for LPs is necessary and sufficient

The dual LP problem

$$\max_{\lambda \in \mathbb{R}^m} b^\top \lambda \quad \text{subject to} \quad A^\top \lambda \leq c$$

Lagrangian:

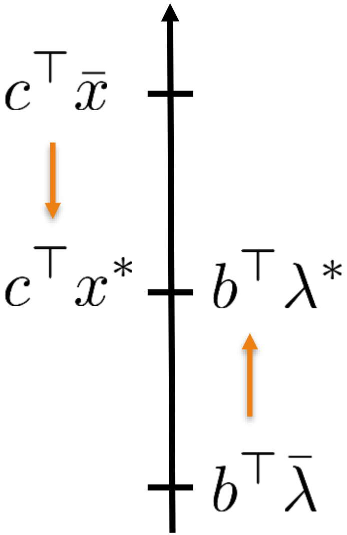
$$\mathcal{L}(x, \lambda) = f(x) - \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i c_i(x)$$

KKT-conditions:

$$\begin{aligned}
\nabla_x \mathcal{L}(x^*, \lambda^*) &= 0, \\
c_i(x^*) &= 0, \quad \forall i \in \mathcal{E}, \\
c_i(x^*) &\geq 0, \quad \forall i \in \mathcal{I}, \\
\lambda_i^* &\geq 0, \quad \forall i \in \mathcal{I}, \\
\lambda_i^* c_i(x^*) &= 0, \quad \forall i \in \mathcal{E} \cup \mathcal{I}.
\end{aligned}$$

Weak duality $c^\top \bar{x} \geq c^\top x^* = b^\top \lambda^* \geq b^\top \bar{\lambda}$

$$\min_{x \in \mathbb{R}^n} c^\top x \quad \text{subject to} \quad \begin{cases} Ax = b \\ x \geq 0 \end{cases}$$



$$\max_{\lambda \in \mathbb{R}^m} b^\top \lambda \quad \text{subject to} \quad A^\top \lambda \leq c$$

Strong duality

Theorem 13.1 Strong duality for LP

i) If primal or dual has a finite solution, so does the other, and $c^\top x^* = b^\top \lambda^*$

ii) If primal or dual is unbounded, the other is infeasible

Sensitivity

- Given LP in standard form:

$$\min_{x \in \mathbb{R}^n} c^\top x \quad \text{subject to} \quad \begin{cases} Ax = b \\ x \geq 0 \end{cases}$$

- Assume optimal solution x^*, corresponding Lagrangian multiplier λ^*
- Do a small perturbation in b_i: $\tilde{b}_i = b_i + \epsilon$
- New solution fulfills

$$c^\top x^*_{\text{new}} = c^\top x^* \pm \epsilon \lambda^*_i$$

Derivatives

Notation: $f(x) \in C^1$ means $f(x)$ once continuously differentiable, C^2 twice, etc.

- **Gradient and Hessian**: For a function $f : \mathbb{R}^n \to \mathbb{R}$, $f(x) \in C^2$, the gradient and Hessian are

$$\nabla f(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}, \quad \nabla^2 f(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 x_n} \\ \frac{\partial^2 f}{\partial x_2 x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n x_1} & \frac{\partial^2 f}{\partial x_n x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}$$

- **Directional derivative**: The directional derivative of $f : \mathbb{R}^n \to \mathbb{R}$ is

$$D(f(x);p) := \lim_{\epsilon \to 0} \frac{f(x + \epsilon p) - f(x)}{\epsilon}$$

Also valid when $f(x)$ is not continuously differentiable. When $f(x) \in C^1$,

$$D(f(x);p) = \nabla f(x)^\top p$$

- **Lipschitz continuity**: A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is Lipschitz continuous in a neighborhood $\mathcal{N}$, if

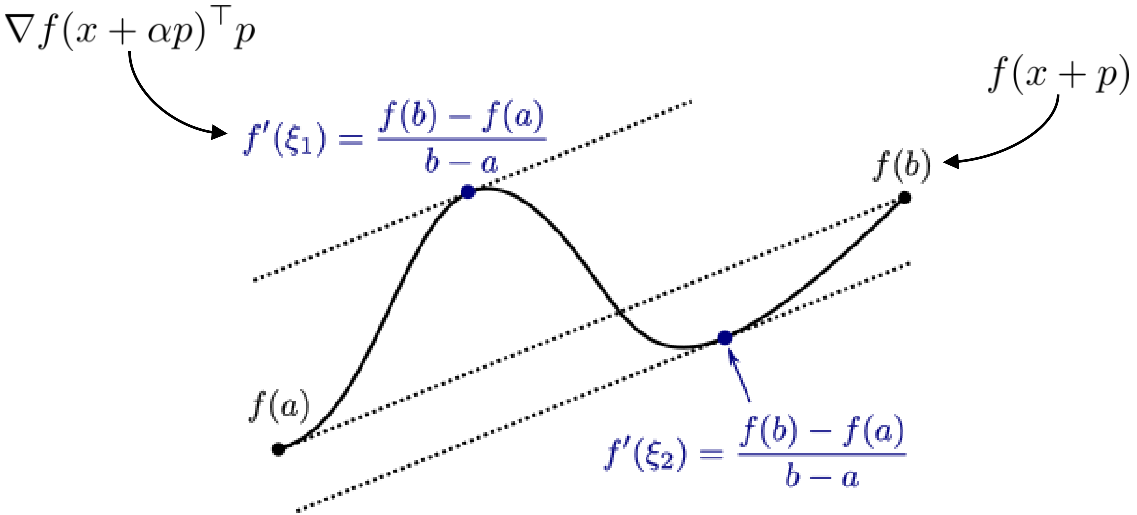
$$\|f(x) - f(y)\| \leq L\|x - y\|, \quad \text{for all } x, y \in \mathcal{N}$$

Taken from N&W A.2

Mean Value Theorem

- For $f : \mathbb{R}^n \to \mathbb{R}$, $f(x) \in C^1$, we have

$$f(x+p) = f(x) + \nabla f(x+\alpha p)^\top p \quad \text{for some } \alpha \in < 0, 1 >$$


$\nabla f(x+\alpha p)^\top p$
$f(x+p)$
$f'(\xi_1) = \frac{f(b)-f(a)}{b-a}$
$f(b)$
$f(a)$
$f'(\xi_2) = \frac{f(b)-f(a)}{b-a}$

wikipedia.org

LP, Standard Form, and KKT

LP, standard form: $\min_{x \in \mathbb{R}^n} c^T x \quad \text{subject to} \quad \begin{cases} Ax = b \\ x \geq 0 \end{cases}$

Lagrangian: $\mathcal{L}(x, \lambda, s) = c^T x - \lambda^T (Ax - b) - s^T x$

KKT-conditions (necessary *and* sufficient for LP):

$$\begin{aligned} A^T \lambda^* + s^* &= c, \\ Ax^* &= b, \\ x^* &\geq 0, \\ s^* &\geq 0, \\ x_i^* s_i^* &= 0, \quad i = 1, 2, \ldots, n \end{aligned}$$