Duality in Mathematical Programming

Leo Liberti

LIX, École Polytechnique

liberti@lix.polytechnique.fr

Lecture material

- Website:
  `http://www.lix.polytechnique.fr/~liberti/teaching/mpri/07/`
- Lecture notes:
  `http://www.lix.polytechnique.fr/~liberti/teaching/mpri/06/linear_programming.pdf`
- S. Boyd and L. Vandenberghe, *Convex Optimization*, CUP, Cambridge, 2004
  `http://www.stanford.edu/~boyd/cvxbook/`
- J.-B. Hiriart-Urruty, *Optimisation et analyse convexe*, PUF, Paris 1998 (Ch. 5)
- C. Papadimitriou, K. Steiglitz, *Combinatorial Optimization: Algorithms and Complexity*, Dover, New York, 1998

Definitions

- Mathematical programming formulation:

$$\left.\begin{array}{rl} \min_x & f(x) \\ \text{s.t.} & g(x) \leq 0 \end{array}\right\} [P] \qquad (1)$$

- A point x^* is *feasible* in P if $g(x^*) \leq 0$;
  $F(P) =$ set of feasible points of P
- A feasible x^* is a *local minimum* if $\exists B(x^*, \varepsilon)$ s.t. $\forall x \in F(P) \cap B(x^*, \varepsilon)$ we have $f(x^*) \leq f(x)$
- A feasible x^* is a *global minimum* if $\forall x \in F(P)$ we have $f(x^*) \leq f(x)$
- Thm.: if f and $F(P)$ convex, any local min. is also global
- If $g_i(x^*) = 0$ for some i, g_i is *active at* x^*

LP Canonical form

- P is a *linear programming problem* (LP) if $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$ are **linear forms**
- LP in *canonical form*:

$$\left.\begin{array}{rl} \min_x & c^\mathsf{T} x \\ \text{s.t.} & Ax \le b \\ & x \ge 0 \end{array}\right\} [C] \qquad (2)$$

- Can reformulate inequalities to equations by adding a non-negative *slack variable* $x_{n+1} \ge 0$:

$$\sum_{j=1}^{n} a_j x_j \le b \;\Rightarrow\; \sum_{j=1}^{n} a_j x_j + x_{n+1} = b \;\wedge\; x_{n+1} \ge 0$$

- Can reformulate maximization to minimization by $\max f(x) = -\min -f(x)$

LP Standard form

- LP in *standard form*: all inequalities transformed to equations

$$\left.\begin{array}{rl} \min_x & (c')^{\mathsf{T}} x \\ \text{s.t.} & A'x = b \\ & x \geq 0 \end{array}\right\} [S] \qquad (3)$$

- where $x = (x_1, \ldots, x_n, x_{n+1}, \ldots, x_{n+m})$, $A' = (A, I_m)$, $c' = (c, \underbrace{0, \ldots, 0}_{m})$
- Standard form is useful because linear systems of equations are computationally easier to deal with than systems of inequalities
- Used in simplex algorithm

Diet problem I

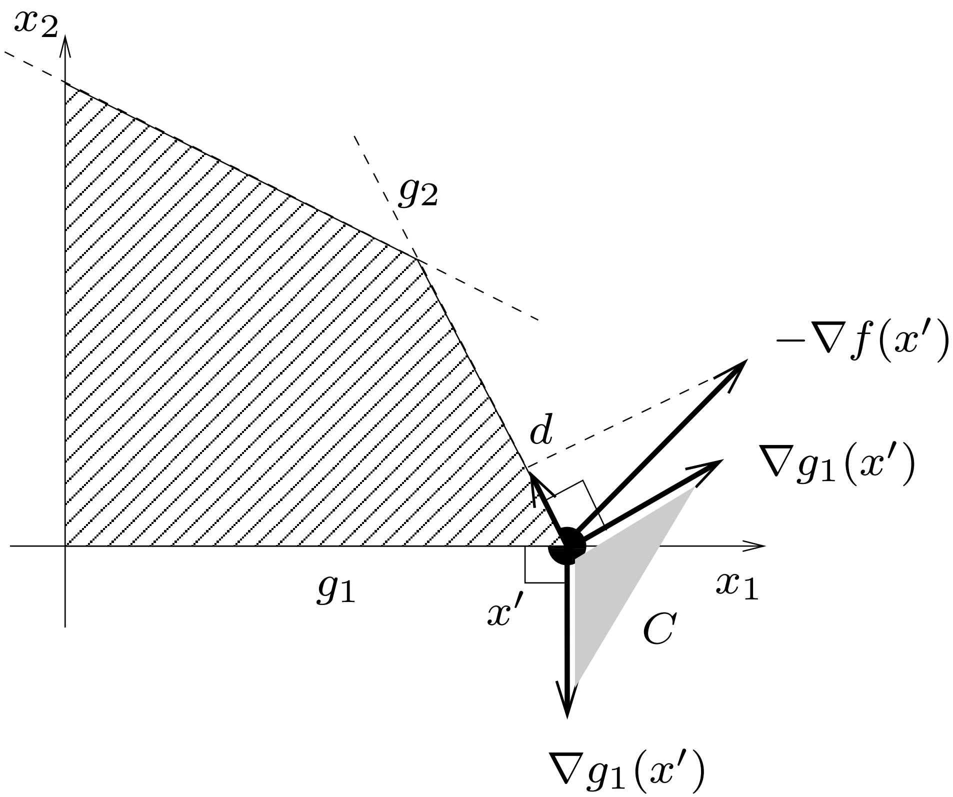
- Consider set M of m nutrients (e.g. sugars, fats, carbohydrates, proteins, vitamins, . . .)
- Consider set N of n types of food (e.g. pasta, steak, potatoes, salad, ham, fruit, . . .)
- A diet is healthy if it has at least b_i units of nutrient $i \in M$
- Food $j \in N$ contains a_{ij} units of nutrient $i \in M$
- A unit of food $j \in N$ costs c_j
- Find a healthy diet of minimum cost

Diet problem II

- Parameters: $m \times n$ matrix $A = (a_{ij})$, $b = (b_1, \ldots, b_m)$, $c = (c_1, \ldots, c_n)$
- Decision variables: $x_j =$ quantity of food j in the diet
- Objective function: $\min\limits_{x} \sum\limits_{j=1}^{n} c_j x_j$
- Constraints: $\forall i \in M \ \sum\limits_{j=1}^{n} a_{ij} x_j \geq b_i$
- Limits on variables: $\forall j \in N \ x_j \geq 0$
- Canonical form: $\min\{c^{\mathsf{T}} x \mid -Ax \leq -b\}$
- Standard form: add slack variables $y_i =$ surplus quantity of i-th nutrient, get $\min\{c^{\mathsf{T}} x \mid -Ax + I_m y = -b\}$

Optimality Conditions I

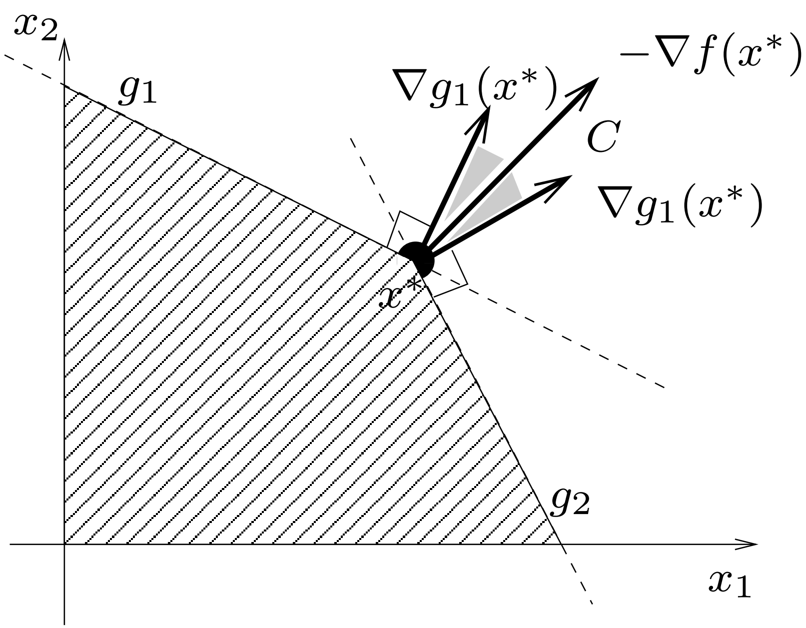
- If we can project improving direction $-\nabla f(x')$ on an active constraint g_2 and obtain a feasible direction d, point x' is not optimal



- Implies $-\nabla f(x') \notin C$ (*cone generated by active constraint gradients*)

Optimality Conditions II

- Geometric intuition: situation as above does not happen when $-\nabla f(x^*) \in C$, x^* optimum



- Projection of $-\nabla f(x^*)$ on active constraints is never a feasible direction

Optimality Conditions III

- If:
  1. x^* is a local minimum of problem $P \equiv \min\{f(x) \mid g(x) \le 0\}$,
  2. I is the index set of the active constraints at x^*, i.e. $\forall i \in I \ (g_i(x^*) = 0)$
  3. $\nabla g_I(x^*) = \{\nabla g_i(x^*) \mid i \in I\}$ is a linearly independent set of vectors
- then $-\nabla f(x^*)$ is a conic combination of $\nabla g_I(x^*)$, i.e. $\exists y \in \mathbb{R}^{|I|}$ such that

$$\begin{aligned} \nabla f(x^*) + \sum_{i \in I} y_i \nabla g_i(x^*) &= 0 \\ \forall i \in I \ y_i &\ge 0 \end{aligned}$$

Karush-Kuhn-Tucker Conditions

- Define

$$L(x,y) = f(x) + \sum_{i=1}^{m} y_i g_i(x)$$

as the *Lagrangian* of problem P

- KKT: If x^* is a local minimum of problem P and $\nabla g(x^*)$ is a linearly independent set of vectors, $\exists y \in \mathbb{R}^m$ s.t.

$$\begin{aligned} \nabla_{x^*} L(x,y) &= 0 \\ \forall i \le m \quad (y_i g_i(x^*) &= 0) \\ \forall i \le m \quad (y_i &\ge 0) \end{aligned}$$

Weak duality

Thm.

Let $\bar{L}(y) = \min_{x \in F(P)} L(x,y)$ and x^* be the global optimum of P. Then $\forall y \geq 0 \quad \bar{L}(y) \leq f(x^*)$.

Proof

Since $y \geq 0$, if $x \in F(P)$ then $y_i g_i(x) \leq 0$, hence $L(x,y) \leq f(x)$; result follows as we are taking the minimum over all $x \in F(P)$.

- Important point: $\bar{L}(y)$ is a lower bound for P for all $y \geq 0$
- The problem of finding the tightest Lagrangian lower bound

$$\max_{y \geq 0} \min_{x \in F(P)} L(x,y)$$

is the *Lagrangian dual* of problem P

Dual of an LP I

- Consider LP P in form: $\min\{c^\mathsf{T}x \mid Ax \geq b \wedge x \geq 0\}$
- $L(x,s,y) = c^\mathsf{T}x - s^\mathsf{T}x + y^\mathsf{T}(b - Ax)$ where $s \in \mathbb{R}^n$, $y \in \mathbb{R}^m$
- Lagrangian dual:

$$\max_{s,y \geq 0} \min_{x \in F(P)} (yb + (c^\mathsf{T} - s - yA)x)$$

- KKT: for a point x to be optimal,

$$\begin{aligned} c^\mathsf{T} - s - yA &= 0 \text{ (KKT1)} \\ \forall j \leq n \ (s_j x_j = 0),\ \forall i \leq m \ (y_i(b_i - A_i x) &= 0) \text{ (KKT2)} \\ s, y &\geq 0 \text{ (KKT3)} \end{aligned}$$

- Consider Lagrangian dual s.t. (KKT1), (KKT3):

Dual of an LP II

- Obtain:

$$\left.\begin{array}{rrcl} \max\limits_{s,y} & yb & & \\ \text{s.t.} & yA+s & = & c^{\mathsf{T}} \\ & s,y & \geq & 0 \end{array}\right\}$$

- Interpret s as slack variables, get *dual of LP*:

$$\left.\begin{array}{rrcl} \min\limits_{x} & c^{\mathsf{T}}x & & \\ \text{s.t.} & Ax & \geq & b \\ & x & \geq & 0 \end{array}\right\}[P] \longrightarrow \left.\begin{array}{rrcl} \max\limits_{y} & yb & & \\ \text{s.t.} & yA & \leq & c^{\mathsf{T}} \\ & y & \geq & 0 \end{array}\right\}[D]$$

Alternative derivation of LP dual

- Lagrangian dual $\Rightarrow$ find tightest lower bound for LP $\min c^{\mathsf{T}}x$ s.t. $Ax \geq b$ and $x \geq 0$
- Multiply constraints $Ax \geq b$ by coefficients $y_1, \ldots, y_m$ to obtain the inequalities $y_i Ax \geq y_i b$, valid provided $y \geq 0$
- Sum over i: $\sum_i y_i Ax \geq \sum_i y_i b = yAx \geq yb$
- Look for y such that obtained inequalities are as stringent as possible whilst still a lower bound for $c^{\mathsf{T}}x$
- $\Rightarrow yb \leq yAx$ and $yb \leq c^{\mathsf{T}}x$
- Suggests setting $yA = c^{\mathsf{T}}$ and maximizing yb
- Obtain LP dual: $\max yb$ s.t. $yA = c^{\mathsf{T}}$ and $y \geq 0$.

Strong Duality for LP

Thm.

If x is optimum of a linear problem and y is the optimum of its dual, primal and dual objective functions attain the same values at x and respectively y.

Proof

- Assume x optimum, KKT conditions hold
- Recall (KKT2) $\forall j \le n (s_i x_i = 0)$,
  $\forall i \le m \ (y_i(b_i - A_i x) = 0)$
- Get $y(b - Ax) = sx \Rightarrow yb = (yA + s)x$
- By (KKT1) $yA + s = c^\mathsf{T}$
- Obtain $yb = c^\mathsf{T} x$

Strong Duality for convex NLPs I

- Theory of KKT conditions derived for generic NLP $\min f(x)$ s.t. $g(x) \leq 0$, independent of linearity of f, g
- Derive strong duality results for convex NLPs
- Slater condition $\exists x' \in F(P)$ $(g(x') < 0)$ requires non-empty interior of $F(P)$
- Let $f^* = \min_{x:g(x)\leq 0} f(x)$ be the optimal objective function value of the primal problem P
- Let $p^* = \max_{y\geq 0} \min_{x\in F(P)} L(x,y)$ be the optimal objective function value of the Lagrangian dual

Thm.

If f, g are convex functions and Slater's condition holds, then $f^* = p^*$.

Strong Duality for convex NLPs II

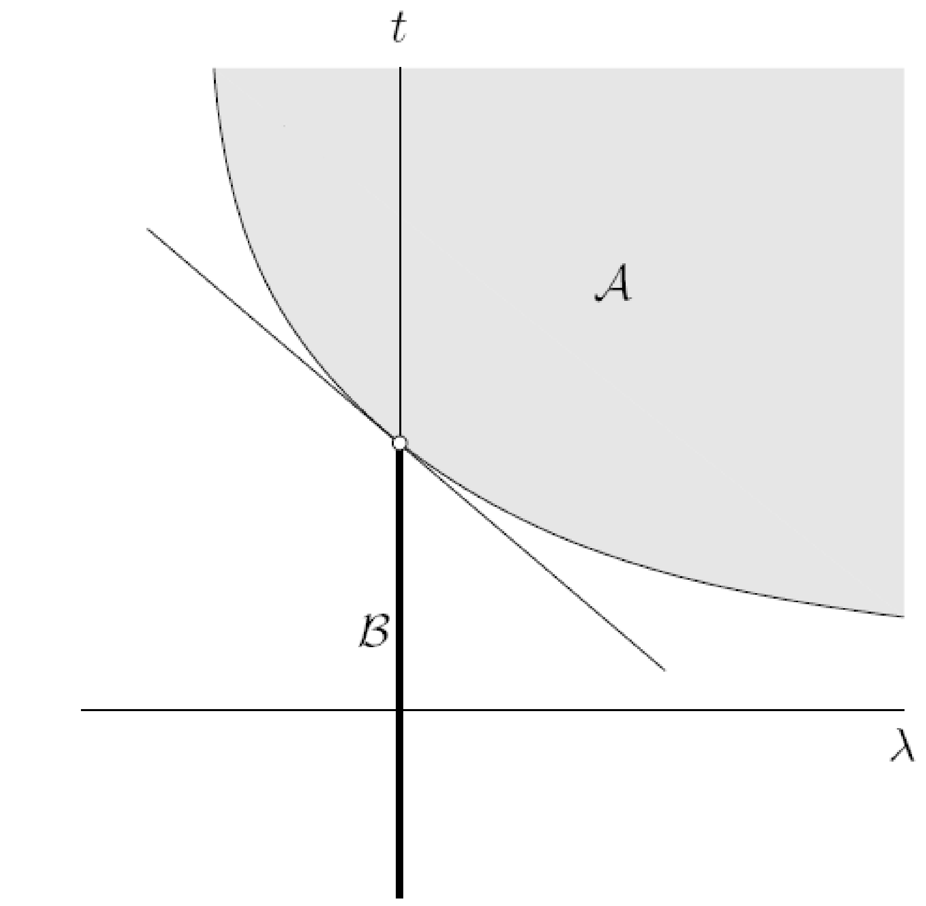
Proof

- Let $\mathcal{A} = \{(\lambda, t) \mid \exists x \ (\lambda \geq g(x) \wedge t \geq f(x))\}$, $\mathcal{B} = \{(0, t) \mid t < f^*\}$

  - $\mathcal{A} =$set of values taken by constraints and objectives
  - $\mathcal{A} \cap \mathcal{B} = \emptyset$ for otherwise f^* not optimal
  - P is convex $\Rightarrow \mathcal{A}, \mathcal{B}$ convex
  - $\Rightarrow \exists$ separating hyperplane $u\lambda + \mu t = \alpha$ s.t.

$$\forall (\lambda, t) \in \mathcal{A} \ (u\lambda + \mu t \geq \alpha) \quad (4)$$

$$\forall (\lambda, t) \in \mathcal{B} \ (u\lambda + \mu t \leq \alpha) \quad (5)$$



- Since λ, t may increase indefinitely, (4) bounded below $\Rightarrow u \geq 0, \mu \geq 0$

Strong Duality for convex NLPs III

Proof

- Since $\lambda = 0$ in $\mathcal{B}$, (5) $\Rightarrow \forall t < f^* \ (\mu t \le \alpha)$
- Combining latter with (4) yields

$$\forall x \ (ug(x) + \mu f(x) \ge \mu f^*) \tag{6}$$

- Suppose $\mu = 0$: (6) becomes $ug(x) \ge 0$ for all feasible x; by Slater's condition $\exists x' \in F(P) \ (g(x') < 0)$, so $u \le 0$, which together with $u \ge 0$ implies $u = 0$; hence $(u, \mu) = 0$ contradicting separating hyperplane theorem, thus $\mu > 0$
- Setting $\mu y = u$ in (6) yields $\forall x \in F(P) \ (f(x) + yg(x) \ge f^*)$
- Thus, for all feasible x we have $L(x, y) \ge f^*$
- In particular, $p^* = \max_y \min_x L(x, y) \ge f^*$
- Weak duality implies $p^* \le f^*$
- Hence, $p^* = f^*$

The dual of the Diet Problem

- Recall diet problem: select minimum-cost diet of n foods providing m nutrients
- Suppose firm wishes to set the prices $y \geq 0$ for m nutrient pills
- To be competitive with normal foods, the equivalent in pills of a food $j \leq n$ must cost less than the cost of the food c_j
- Objective: $\max \sum_{i \leq m} b_i y_i$
- Constraints: $\forall j \leq n \quad \sum_{i \leq m} a_{ij} y_i \leq c_j$
- Economic interpretation:
  at optimum, cost of pills = cost of diet

Examples: LP dual formulations

- Primal problem P and canonical form:

$$\left.\begin{array}{rl} \max\limits_{x_1,x_2} & x_1 + x_2 \\ \text{s.t.} & x_1 + 2x_2 \le 2 \\ & 2x_1 + x_2 \le 2 \\ & x \ge 0 \end{array}\right\} \Rightarrow \left.\begin{array}{rl} -\min\limits_{x_1,x_2} & -x_1 - x_2 \\ \text{s.t.} & -x_1 - 2x_2 \ge -2 \\ & -2x_1 - x_2 \ge -2 \\ & x \ge 0 \end{array}\right\}$$

- Dual problem D and reformulation:

$$\left.\begin{array}{rl} -\max\limits_{y_1,y_2} & -2y_1 - 2y_2 \\ \text{s.t.} & -y_1 - 2y_2 \le -1 \\ & -2y_1 - y_2 \le -1 \\ & y \ge 0 \end{array}\right\} \Rightarrow \left.\begin{array}{rl} \min\limits_{y_1,y_2} & 2y_1 + 2y_2 \\ \text{s.t.} & y_1 + 2y_2 \ge 1 \\ & 2y_1 + y_2 \ge 1 \\ & y \ge 0 \end{array}\right\}$$

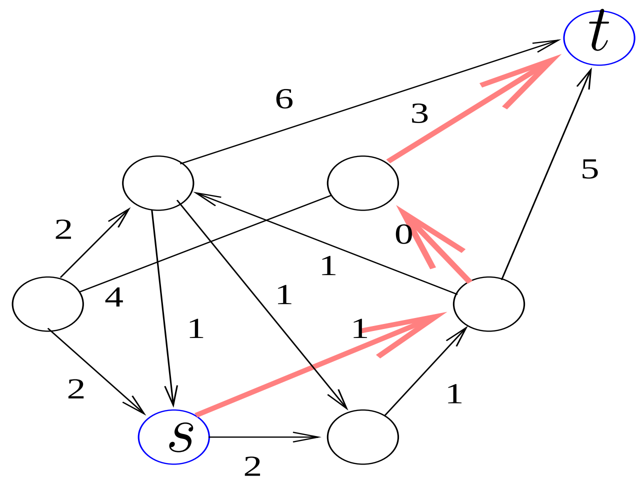
Rules for LP dual

Primal	Dual
min	max
variables x	**constraints**
constraints	**variables** y
objective coefficients c	constraint right hand sides c
constraint right hand sides b	objective coefficients b
$A_i x \geq b_i$	$y_i \geq 0$
$A_i x \leq b_i$	$y_i \leq 0$
$A_i x = b_i$	y_i unconstrained
$x_j \geq 0$	$yA^j \leq c_j$
$x_j \leq 0$	$yA^j \geq c_j$
x_j unconstrained	$yA^j = c_j$

A_i: **i-th row of** A A^j: **j-th column of** A

Example: Shortest Path Problem

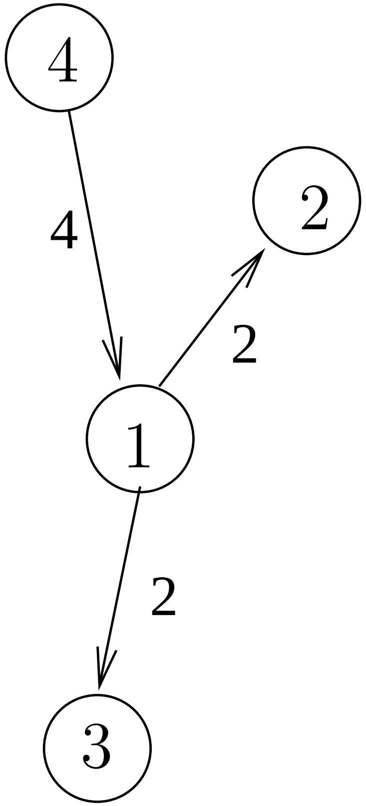
SHORTEST PATH PROBLEM.
Input: weighted digraph $G = (V, A, c)$, and $s, t \in V$.
Output: a minimum-weight path in G from s to t.



$$\left.\begin{array}{lll} \min\limits_{x \geq 0} & \sum\limits_{(u,v)\in A} c_{uv} x_{uv} & \\ \forall\, v \in V & \sum\limits_{(v,u)\in A} x_{vu} - \sum\limits_{(u,v)\in A} x_{uv} & = \left\{\begin{array}{ll} 1 & v = s \\ -1 & v = t \\ 0 & \text{othw.} \end{array}\right. \end{array}\right\} [P]$$

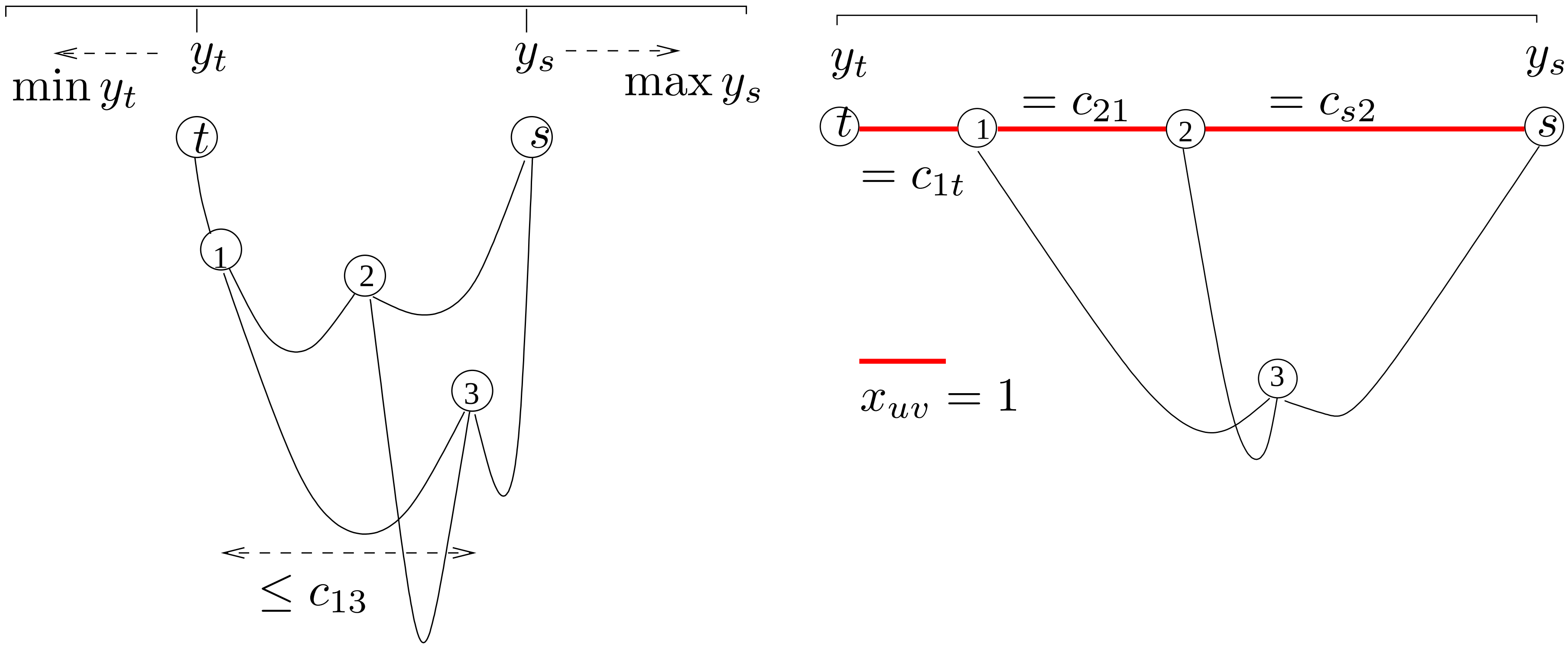
$$\left.\begin{array}{lll} \max\limits_{y} & y_s - y_t & \\ \forall\, (u,v) \in A & y_v - y_u & \leq \quad c_{uv} \end{array}\right\} [D]$$

Shortest Path Dual



cols	(1,2)	(1,3)	...	(4,1)	...		
rows\\c	2	2	...	4	...	b	
1	1	1	...	-1	...	0	y_1
2	-1	0	...	0	...	0	y_2
3	0	-1	...	0	...	0	y_3
4	0	0	...	1	...	0	y_4
⋮	⋮	⋮		⋮		⋮	⋮
s	0	0	...	0	...	1	y_s
⋮	⋮	⋮		⋮		⋮	⋮
t	0	0	...	0	...	-1	y_t
⋮	⋮	⋮		⋮		⋮	⋮
	x_{12}	x_{13}	...	x_{41}	...		

SP Mechanical Algorithm



KKT2 on [D] $\Rightarrow \quad \forall (u,v) \in A \; (x_{uv}(y_v - y_u - c_{uv}) = 0) \Rightarrow$

$$\forall (u,v) \in A \; (x_{uv} = 1 \rightarrow y_v - y_u = c_{uv})$$

Sensitivity analysis I

- Suppose we solved an LP to optimality, get x^*
- Ask the question: if b is varied by a certain "noise" vector ε, how does the objective function change?
- In practice, this addresses the problem of stability:
  - we found an optimal solution with lowest associated cost f^*
  - all coefficients deriving from real world carry some measurement uncertainties (suppose b are uncertain)
  - so x^* may not be optimal for the practical application
  - however, there may be a "close" feasible solution
  - we hope the "real" optimal cost doesn't change too much from f^*
  - can we say by how much?

Sensitivity analysis II

- Consider an LP with primal optimal solution x^* and dual optimal solution y^*
- Perturb b coefficients to $b + \varepsilon$
- The objective function value becomes $y(b + \varepsilon) = yb + y\varepsilon$
- Suppose $||\varepsilon||$ is small enough so that the optimal solution does not change
- $c^\mathsf{T} x^* = y^* b$ (strong LP duality) implies the optimal objective function value for the perturbed problem is $c^\mathsf{T} x^* + y^* \varepsilon$
- In other words: y^* **is the variation of the objective function with respect to a unit variation in** b

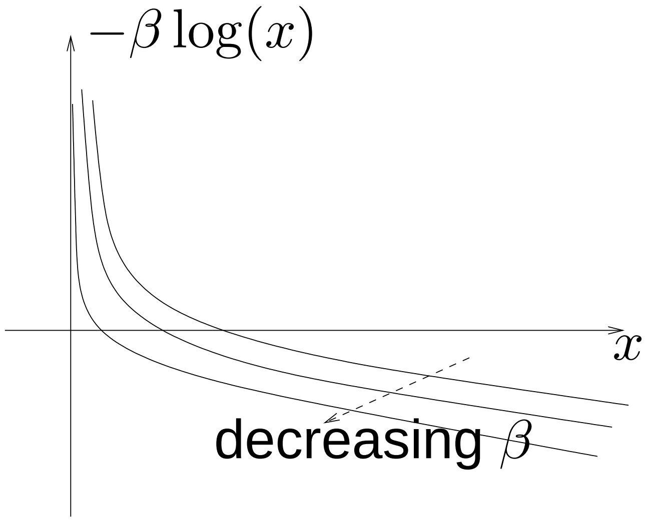
Interior point methods

- Simplex algorithm is practically efficient but nobody ever found a pivot choice rule that proves that it has polynomial worst-case running time
- Nobody ever managed to prove that such a rule does not exist
- With current pivoting rules, simplex worst-case running time is exponential
- Kachiyan managed to prove in 1979 that LP $\in$ **P** using a polynomial algorithm called *ellipsoid method* (`http://www.stanford.edu/class/msande310/ellip.pdf`)
- Ellipsoid method has polynomial worst-case running time but performs badly in practice
- Barrier interior point methods for LP have both polynomial running time and good practical performances

IPM I: Preliminaries

- Consider LP P in standard form: $\min\{c^\mathsf{T} x \mid Ax = b \wedge x \geq 0\}$.
- Reformulate by introducing "logarithmic barriers":

$$P(\beta) : \min\{c^\mathsf{T} x - \beta \sum_{j=1}^{n} \log x_j \mid Ax = b\}$$



- The term $-\beta \log(x_j)$ is a "penalty" that ensures that $x_j > 0$; the "limit" of this reformulation for $\beta \to 0$ should ensure that $x_j \geq 0$ as desired
- Notice $P(\beta)$ is convex $\forall \beta > 0$

IPM II: Central path

- Let $x^*(\beta)$ the optimal solution of $P(\beta)$ and x^* the optimal solution of P
- The set $\{x^*(\beta) \mid \beta > 0\}$ is called the *central path*
- Idea: determine the central path by solving a sequence of convex problems $P(\beta)$ for some decreasing sequence of values of β and show that $x^*(\beta) \to x^*$ as $\beta \to 0$
- Since for $\beta > 0$, $-\beta \log(x_j) \to +\infty$ for $x_j \to 0$, $x^*(\beta)$ will never be on the boundary of the feasible polyhedron $\{x \geq 0 \mid Ax = b\}$; thus the name “interior point method”

IPM III: Dual feasibility

Thm.

For all $\beta > 0$, $x^*(\beta)$ determines a dual feasible point y for P.

Proof

The Langrangian of P is

$$L_1(x, y, \nu) = c^{\mathsf{T}} x - \sum_{j \le n} y_j x_j + \nu(Ax - b), \tag{7}$$

where $y \in \mathbb{R}^n_+$ (corresponds to constraints $-x \le 0$) and $\nu \in \mathbb{R}^m$ (A is $m \times n$). The Lagrangian of $P(\beta)$ is

$$L_2(x, \nu) = c^{\mathsf{T}} x - \sum_{j \le n} \beta \log(x_j) + \nu(Ax - b). \tag{8}$$

Derive KKT1 ($\nabla L = 0$) for L_1, L_2:

$$\forall j \le n \; (c_j - y_j + \nu A^j = 0) \wedge (c_j - \frac{\beta}{x^*_j} + \nu A^j = 0)$$

Letting $y_j = \frac{\beta}{x^*_j}$ shows that $x^*(\beta)$ yields a point (y, ν) feasible in the dual

IPM III: Convergence

Thm.

$x^*(\beta) \to x^*$ as $\beta \to 0$.

Proof

Notice first that $x^*(\beta)$ determines a converging sequence for each sequence of values of β that converges to 0, because $P(\beta)$ is not unbounded for any $\beta > 0$; let the limit be x'. By previous thm., for each $x^*(\beta)$ there is a dual feasible point $(y(\beta), \nu(\beta))$ s.t. $\forall j \leq n$ $(y_j(\beta) = \frac{\beta}{x_j^*(\beta)})$. This also shows that any sequence $y(\beta)$ is convergent for $\beta \to 0$; let the limit be y^*. Since $\forall j \leq n$ $(x_j^*(\beta) y_j(\beta) = \beta)$, as $\beta \to 0$ we have $x_j^*(\beta) y_j(\beta) \to 0$. But since $x_j^*(\beta) y(\beta) \to x_j' y_j^*$, then $x_j' y_j^* = 0$. By the KKT complementarity conditions, x', y^* are a pair of primal/dual optimal solutions, so $x' = x^*$.

IPM IV: Optimal partitions

- An LP may have more than one optimal solution (try solving $\max x_1 + x_2$ s.t. $x_1 + x_2 \leq 1$ and $x \geq 0$)
- If this happens, all the solutions are on the same face ϕ of the feasible polyhedron
- The simplex method fails to detect this situation
- In this case, the barrier IPM gives a *strictly complementary* solution (i.e. $(x^*)^{\mathsf{T}} y^* = 0$ and $x^* + y^* > 0$) in the interior of the face ϕ
- This solution can be used to determine the *optimal partition* (B, N) such that $B = \{j \leq n \mid x_i^* > 0\}$ and $N = \{j \leq n \mid y_i^* > 0\}$.
- The optimal partition is unique and does not depend on the optimal solution used to define it — thus it provides a well-defined characterization of optimal faces

IPM V: Strict complementarity

Thm.

(x^*, y^*) is a strictly complementary primal-dual optimal solution of P.

Proof

Let $x' = x^*(\beta)$, $y' = y(\beta)$, $\nu' = \nu(\beta)$ for some $\beta > 0$ and ν^* be the limit of the sequence $\nu(\beta)$ as $\beta \to 0$. $x^*(\beta), x^*$ are both primal feasible (hence $A(x^* - x') = 0$), and $(y^*, \nu^*), (y', \nu')$ are both dual feasible (hence $(\nu^* - \nu')A = y^* - y'$). In other words, $x^* - x'$ is in the null space of A and $y^* - y'$ in the range of A^T. Thus, the two vectors are orthogonal: hence $0 = (x^* - x')^\mathsf{T}(y^* - y') = (x^*)^\mathsf{T} y^* + (x')^\mathsf{T} y' - (x^*)^\mathsf{T} y - (x')^\mathsf{T} y^*$. Since $(x^*)^\mathsf{T} y^* = 0$ and $(x')^\mathsf{T} y' = \sum_{j \le n} \beta = n\beta$, we obtain $(x^*)^\mathsf{T} y' + (x')^\mathsf{T} y^* = n\beta$. We now divide throughout by $\beta = x'_j y'_j$, obtaining $\sum_{j \le n} (\frac{x^*_j}{x'_j} + \frac{y^*_j}{y'_j}) = n$.

Notice that $\lim_{\beta \to 0} \frac{x^*_j}{x^*_j(\beta)} = \begin{cases} 1 & \text{if } x^*_j > 0 \\ 0 & \text{otherwise} \end{cases}$ and similarly for y. So for each $j \le n$ exactly one of x^*_j, y^*_j is zero and the other is positive.

IPM VI: Prototype algorithm

1. Consider an initial point $x(\beta_0)$ feasible in $P(\beta)$, a parameter $\alpha < 1$ and a tolerance $\varepsilon > 0$. Let $k = 0$.
2. Solve $P(\beta)$ with initial point $x(\beta_k)$ to get a solution x^*.
3. If $n\beta_k < \varepsilon$, stop with solution x^*.
4. Update $\beta_{k+1} = \alpha\beta_k$, $x(\beta_{k+1}) = x^*$ and $k \leftarrow k + 1$.
5. Go to step 2.

Since $L_1(x_k, y, \nu) = c^{\mathsf{T}} x_k - n\beta_k$, the *duality gap* is $n\beta_k$ (i.e. x_k is never more than $n\beta_k$-suboptimal). Each problem $P(\beta)$ can be solved by Newton's method.

IPM VII: Newton's method

- The Newton descent direction d for an unconstrained problem $\min f(x)$ at a point $\bar{x}$ is given by

$$d = -(\nabla^2 f(\bar{x}))^{-1} \nabla f(\bar{x}) \tag{9}$$

- If $\nabla^2 f(\bar{x})$ is positive definite, we obtain

$$(\nabla f(\bar{x}))^\mathsf{T} d = -(\nabla f(\bar{x}))^\mathsf{T} (\nabla^2 f(\bar{x}))^{-1} \nabla f(\bar{x}) < 0,$$

  so d is a descent direction

- Direction d needs to be feasible (i.e. $Ad = 0$), thus solve for (d, ν)

$$\begin{pmatrix} \nabla^2 f(\bar{x}) & A^\mathsf{T} \\ A & 0 \end{pmatrix} \begin{pmatrix} d \\ \nu^\mathsf{T} \end{pmatrix} = \begin{pmatrix} -\nabla f(\bar{x}) \\ 0 \end{pmatrix}$$

- Step 4 in the alg. becomes $x(\beta_{k+1}) = x(\beta_k) + \gamma d$, where γ is the result of a line search

$$\gamma = \mathrm{argmin}_{s \geq 0} f(\bar{x} + sd) \tag{10}$$

IPM VIII: Example

- Consider the LP in canonical form (P is the corresponding standard form problem)

$$\left.\begin{array}{lrcl} \min & x_1 - x_2 & & \\ & -x_1 + x_2 & \leq & 1 \\ & x_1 + x_2 & \leq & 3 \\ & x_1, x_2 & \geq & 0 \end{array}\right\}$$

- with associated $P(\beta)$:

$$\left.\begin{array}{lrcl} \min\limits_{x_1,x_2,x_3,x_4} & x_1 - x_2 & - & \beta \sum\limits_{j=1}^{4} \log x_j \\ & -x_1 + x_2 + x_3 & = & 1 \\ & x_1 + x_2 + x_4 & = & 3 \end{array}\right\}$$

IPM IX: Example

- The constraint matrix A is

$$\begin{pmatrix} -1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{pmatrix}$$

- The objective function gradient is

$$(1 - \frac{\beta}{x_1}, -1 - \frac{\beta}{x_2}, -\frac{\beta}{x_3}, -\frac{\beta}{x_4})^{\mathsf{T}},$$

- The Hessian is

$$\mathrm{diag}(\frac{\beta}{x_1^2}, \frac{\beta}{x_2^2}, \frac{\beta}{x_3^2}, \frac{\beta}{x_4^2}).$$

(hence it is positive semidefinite)

IPM X: Example

- The Newton system to be solved is:

$$\begin{pmatrix} \frac{\beta}{x_1^2} & & & & -1 & 1 \\ & \frac{\beta}{x_2^2} & & & 1 & 1 \\ & & \frac{\beta}{x_3^2} & & 1 & 0 \\ & & & \frac{\beta}{x_4^2} & 0 & 1 \\ -1 & 1 & 1 & 0 & & \\ 1 & 1 & 0 & 1 & & \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \\ d_3 \\ d_4 \\ \nu_1 \\ \nu_2 \end{pmatrix} = \begin{pmatrix} -1 + \frac{\beta}{x_1} \\ 1 + \frac{\beta}{x_2} \\ \frac{\beta}{x_3} \\ \frac{\beta}{x_4} \\ 0 \\ 0 \end{pmatrix}$$

- We can now write a Matlab (or GNU Octave) code to implement the IPM algorithm

IPM XI: Code

```
function [xstar, ystar, k, B] = ipm(c, A, b, beta, xfeas, alpha, epsilon)
  %% initialization
  OPTIONS = [ ];
  [m, n] = size(A);
  Ineq = A(:, 1 : n-m);
  nx = size(xfeas);
  if nx < n
    s = b - Ineq*xfeas;
    x = [ xfeas ; s ];
  else
    x = xfeas;
  end
  J = zeros(n, 1);
  H = zeros(n, n);
  d = zeros(n, 1);
  nu = zeros(m, 1);
  termination = 0;
  counter = 1;
...
```

IPM XII: Code

```
...
  %% iterative method
  while termination == 0
    for i = 1 : n
      J(i) = c(i) - beta / x(i);
      H(i,i) = beta/x(i)ˆ2;
    end
    N = [ H, A’; A, zeros(m, m) ];
    bN = [ -J; zeros(m, 1) ];
    direction = N \ bN;
    d = direction(1 : n, 1);
    nu = direction(n + 1 : n + m);
    lambda = fminbnd(’linesearch’, 0, 1, OPTIONS, c, x, d, beta);
    xstar = x + lambda * d;
...
```

IPM XIII: Code

```
...
    if n * beta < epsilon
      termination = 1;
      k = counter;
      ystar = beta ./ xstar;
      B = zeros(1, n);
      for i = 1 : n
        if xstar(i) > ystar(i)
          B(i) = 1;
        end
      end
    end
    beta = alpha * beta;
    x = xstar;
    counter = counter + 1;
  end
%end function

function y = linesearch(lambda, c, x, d, beta)
  y = c*(x + lambda*d) - beta*sum(log(x + lambda*d),1);
```

IPM XIII: Solution

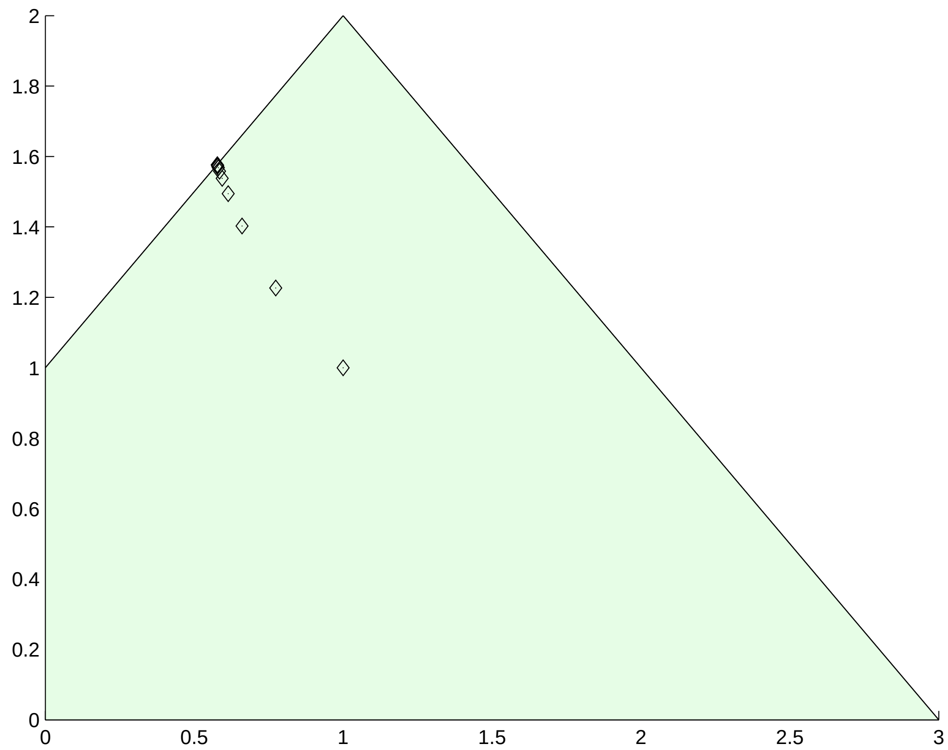
- Code parameters: $\beta_1 = 1$, $\alpha = 0.5$, $\varepsilon = 0.01$
- Problem parameters: $c^{\mathsf{T}} = (1, -1, 0, 0)$, $b = (1, 3)$ and $x^*(\beta_1)^{\mathsf{T}} = (1, 1)$
- Running the example:

```
ipm([1 -1 0 0 ], [-1 1 1 0 ; 1 1 0 1],
    [1; 3], 1, [1; 1], 0.5, 0.01)
```

- Solution (approximated to 10^{-2}): $x^* = (0.58, 1.58, 0.00, 0.85)$

IPM XIV: Solution

The central path:



determines a solution $x^* = (0.58.1.58, 0, 0.85)$ whose optimal partition is $B = \{1, 2, 4\}$ and $N = \{3\}$.

IPM XV: Comparison with Simplex

The solution found by the simplex method is $x^* = (0, 1, 0, 2)$ and $y^* = (0, 0, 1, 0)$, which is not strictly complementary, as $x_1^* + y_1^* = 0 + 0 = 0$.

History of LP I

- **1788**: Optimality conditions for equality-constrained programs (Lagrange)
- **1826**: Solution of a system of linear equations (Gauss)
- **1873**: Solution of a system of linear equations with nonnegative variables (Gordan)
- **1896**: Representation of convex polyhedra (Minkowski)
- **1936**: Solution of a system of linear inequalities (Motzkin)
- **1939**: Optimality conditions (Karush, Kuhn & Tucker)
- **1939-45**: Blackett's Circus, UK Naval Op. Res. , US Navy Antisubmarine Warfare Op. Res. Group, USAF Op. Res., Project RAND
- **1945**: The diet problem (Stigler)

History of LP II

- **1947**: The simplex method (Dantzig)
- **1953**: The revised simplex method (Dantzig)
- **1954**: Cutting planes applied to TSP (Dantzig, Fulkerson, Johnson)
- **1954**: Max flow / min cut theorem (Ford & Fulkerson), declassified 1999
- **1954**: Dual simplex method (Lemke)
- **1954**: Branch and Bound applied to TSP (Eastman)
- **1955**: Stochastic programming (Dantzig & Beale)
- **1956**: Dijkstra's algorithm (Dijkstra)
- **1958**: Cutting planes for integer programming (Gomory)
- **1958**: Dantzig-Wolfe decomposition (Dantzig & Wolfe)

History of LP III

- **1962**: Benders' decomposition (Benders)
- **1963**: *Linear programming and extensions* (Dantzig)
- **1970**: Lagrangian relaxation for integer programming (Held & Karp)
- **1971**: NP-completeness (Cook, Karp)
- **1972**: Simplex method is not polynomial (Klee & Minty)
- **1977**: Bland's rule for simplex method (Bland)
- **1979**: Kachiyan proves LP$\in$**P** using ellipsoid method
- **1982**: Average running time of simplex method (Borgwardt)
- **1984**: Interior point method for LP (Karmarkar)
- **1985**: Branch-and-cut on TSP (Padberg& Grötschel)