

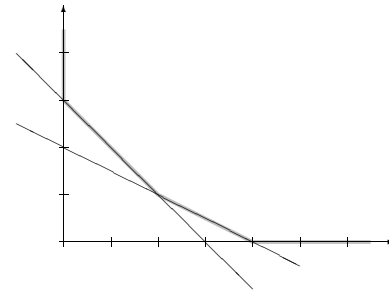
Interior point methods  
 — an introduction

- Linear Programming, duality
- Simplex and its complexity
- Interior point methods — history
- Newton's method
- Obtaining  $x \geq 0$  through barrier function
- Putting the pieces together
- Discussion

matrix notation, omit transposition  $x^T$

**Linear programming**

$$\begin{aligned} &\text{minimize} && 2x_1 + 3x_2 \\ &\text{subject to} && 1x_1 + 2x_2 \geq 4 \\ &&& 1x_1 + 1x_2 \geq 3 \\ &&& x_1, x_2 \geq 0 \end{aligned}$$



In matrix form

$$\begin{aligned} &\text{minimize} && cx \\ &\text{subject to} && Ax = b \\ &&& x \geq 0 \end{aligned}$$

**Linear programming**

- Primal

$$\begin{aligned} &\text{minimize} && cx \\ &\text{subject to} && Ax = b \\ &&& x \geq 0 \end{aligned}$$

- Dual

$$\begin{aligned} &\text{maximize} && by \\ &\text{subject to} && yA + s = c \\ &&& s \geq 0, y \in \mathbb{R} \end{aligned}$$

- *Weak duality*: Assume that  $x$  primal feasible and  $y, s$  dual feasible

$$by \leq cx$$

$$\text{Duality gap: } cx - by \geq 0$$

- *Strong duality*: If the problem has a feasible solution, then there exists a primal-dual feasible pair  $(x^*, y^*, s^*)$ , so that

$$cx^* = by^*$$

- *Complementary slackness* (alternative formulation of strong duality). If  $x \geq 0, s \geq 0, y \in \mathbb{R}$  satisfy

$$\begin{aligned} Ax &= b \\ yA + s &= c \\ sx &= 0 \end{aligned}$$

then  $(x, y, s)$  optimal.

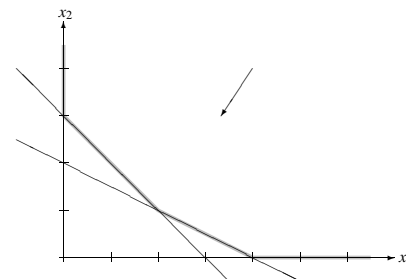
$$cx - by = (yA + s)x - y(Ax) = sx$$

**Solving Linear Programs**

Optimization problem with slack variables added

$$\begin{aligned} &\text{maximize} && 2x_1 + 3x_2 \\ &\text{subject to} && 1x_1 + 2x_2 - x_3 = 4 \\ &&& 1x_1 + 1x_2 - x_4 = 3 \\ &&& x_1, x_2, x_3, x_4 \geq 0 \end{aligned}$$

The set of constraints form a polyhedral.  
 Optimal solution is found at extreme points



Extreme points (basic solutions)

$$(2, 1, 0, 0) \quad (0, 3, 2, 0) \quad (4, 0, 0, 1)$$

Basic set can be chosen in  $\binom{n}{m}$  ways (i.e. exponential).

## Complexity of Simplex

Klee and Minty (1975) proved that the Simplex algorithm may use exponential time

$$\begin{aligned}
 & \text{maximize} \\
 & 2^{n-1}x_1 + 2^{n-2}x_2 + \dots + 2x_{n-1} + 1x_n \\
 & \text{subject to} \\
 & 1x_1 + \quad \quad \quad + \quad \quad \quad + \quad \quad \quad \leq 5 \\
 & 4x_1 + \quad 1x_2 + \quad \quad \quad + \quad \quad \quad \leq 5^2 \\
 & 8x_1 + \quad 4x_2 + 1x_3 + \quad \quad \quad \leq 5^3 \\
 & \quad \vdots + \quad \quad \quad + \quad \quad \quad + \quad \quad \quad \leq \quad \vdots \\
 & 2^n x_1 + 2^{n-1}x_2 + \dots + 4x_{n-1} + 1x_n \leq 5^n \\
 & x_i \geq 0, i = 1, \dots, n
 \end{aligned}$$

The problem has

- $n$  variables
- $n$  constraints
- $2^n$  extreme points
- Simplex, starting at  $x = (0, \dots, 0)$ , visits all extreme points
- optimal solution  $(0, 0, \dots, 0, 5^n)$

5

## Complexity of Simplex

For  $n = 3$  simplex visits  $2^3 = 8$  extreme points  
Assume  $(s_1, s_2, s_3)$  slack variables:

basis	nonbasis			RHS
	$x_1$	$x_2$	$x_3$	
$s_1$	$1^*$			5
$s_2$	4	1		25
$s_3$	8	4	1	125
$-z$	4	2	1	0

basis	nonbasis			RHS
	$s_1$	$x_2$	$x_3$	
$x_1$	1			5
$s_2$	-4	$1^*$		5
$s_3$	-8	4	1	85
$-z$	-4	2	1	-20

basis	nonbasis			RHS
	$s_1$	$s_2$	$x_3$	
$x_1$	$1^*$			5
$x_2$	-4	1		5
$s_3$	8	-4	1	65
$-z$	4	-2	1	-30

basis	nonbasis			RHS
	$x_1$	$s_2$	$x_3$	
$s_1$	1			5
$x_2$	4	1		25
$s_3$	-8	-4	$1^*$	25
$-z$	-4	-2	1	-50

basis	nonbasis			RHS
	$s_1$	$s_2$	$s_3$	
$s_1$	$1^*$			5
$x_2$	4	1		25
$x_3$	-8	-4	1	25
$-z$	4	2	-1	-75

basis	nonbasis			RHS
	$s_1$	$s_2$	$s_3$	
$x_1$	1			5
$x_2$	-4	$1^*$		5
$x_3$	8	-4	1	65
$-z$	-4	2	-1	-95

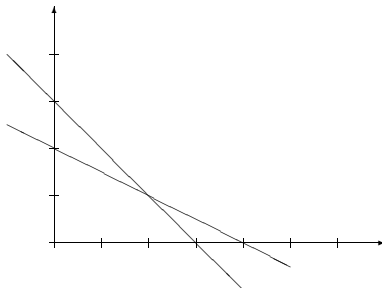
basis	nonbasis			RHS
	$s_1$	$x_2$	$s_3$	
$x_1$	$1^*$			5
$s_2$	-4	1		5
$s_3$	-8	4	1	85
$-z$	4	-2	-1	-105

basis	nonbasis			RHS
	$x_1$	$x_2$	$s_3$	
$s_1$	$1^*$			5
$s_2$	4	1		25
$x_3$	8	4	1	125
$-z$	-4	-2	-1	-125

6

## Interior-point methods — history

- 40ies: Simplex algorithm, Dantzig.
- Klee and Minty (1975) proved that a variant of Simplex may use exponential time. Stimulated research in alternatives.
- Khachiyan (1979) polynomial algorithm, *ellipsoid method*  
Bad performance in practice.
- Karmarkar (1984) polynomial algorithm *path-following*
- Path following method now described by Newton's method, barrier function

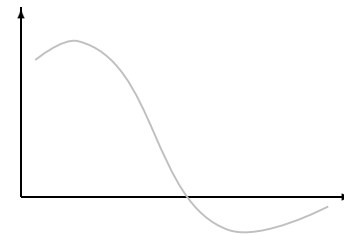


7

## Newton's method

$f(x) : \mathbb{R}^n \rightarrow \mathbb{R}^n$  smooth nonlinear function, solve:

$$f(x) = 0$$



Taylor's theorem (linearization)

$$f(x^0 + d_x) \approx f(x^0) + \nabla f(x^0) d_x$$

If  $x^0$  initial guess, compute  $d_x$  such that  $f(x^0 + d_x) = 0$ .

$$f(x^0) + \nabla f(x^0) d_x = 0 \quad d_x = -(\nabla f(x^0))^{-1} f(x^0)$$

$d_x$  defines search direction, new point  $x^+$

$$x^+ = x^0 + \alpha d_x$$

where  $0 < \alpha < 1$  is step size.

8

## Nonlinear programming

Assume  $c : \mathbb{R}^n \rightarrow \mathbb{R}$  and  $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$  smooth functions.

$$\begin{aligned} & \text{minimize } c(x) \\ & \text{subject to } g(x) = 0 \end{aligned}$$

To use Newton's method formulate as

$$f(x) = 0$$

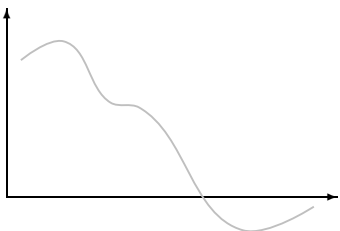
Lagrangian function,  $y$  Lagrangian multipliers:

$$L(x, y) = c(x) - y g(x)$$

First order optimality criterion

$$\begin{aligned} \nabla_x L(x, y) &= \nabla c(x) - \nabla g(x) y = 0 \\ \nabla_y L(x, y) &= -g(x) = 0 \end{aligned}$$

Necessary but not sufficient condition



If  $c(x)$  is convex, and  $g(x)$  is affine, then sufficient

9

## Newton's method for optimality criterion

Looking for solution to " $f(x, y) = 0$ " i.e.

$$\begin{aligned} \nabla_x L(x, y) &= \nabla c(x) - \nabla g(x) y = 0 \\ \nabla_y L(x, y) &= -g(x) = 0 \end{aligned}$$

Taylor expansion in  $(x, y) = (x^0, y^0)$ :

$$f(x^0, y^0) + \nabla f(x^0, y^0) (d_x, d_y) = 0$$

thus

$$\begin{pmatrix} \nabla c(x^0) - \nabla g(x^0)^T y^0 \\ -g(x^0) \end{pmatrix} + \begin{pmatrix} \nabla^2 c(x^0) - \sum_{i=1}^m y_i \nabla^2 g_i(x^0) & -\nabla g(x^0)^T \\ -\nabla g(x^0) & 0 \end{pmatrix} \begin{pmatrix} d_x \\ d_y \end{pmatrix} = 0$$

next point

$$\begin{pmatrix} x^+ \\ y^+ \end{pmatrix} = \begin{pmatrix} x^0 \\ y^0 \end{pmatrix} + \alpha \begin{pmatrix} d_x \\ d_y \end{pmatrix}$$

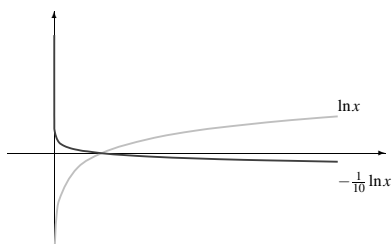
10

## Primal problem, barrier function

Linear programming model

$$\begin{aligned} & \text{minimize } cx \\ & \text{subject to } Ax = b \\ & \quad x \geq 0 \end{aligned}$$

Newton's method cannot handle inequalities.



New objective:

$$\begin{aligned} & \text{minimize } cx - \mu \sum_{j=1}^n \ln(x_j) \\ & \text{subject to } Ax = b \\ & \quad x \geq 0 \end{aligned}$$

where  $\mu > 0$  is a small constant.

$$\lim_{x_j \rightarrow 0} -\mu \ln(x_j) = \infty$$

If  $x > 0$  initially, then *barrier function* maintains  $x > 0$ .

Solution to barrier problem is approx. of original problem

11

## Primal problem, optimality condition

Define optimality conditions for barrier problem

$$L(x, y) = cx - \mu \sum_{j=1}^n \ln(x_j) - y(Ax - b)$$

Differentiation gives

$$\frac{\partial L}{\partial x_j} = c_j - \mu x_j^{-1} - A_{:j}^T y \quad \text{and} \quad \frac{\partial L}{\partial y_i} = b_i - A_{i \cdot} x$$

In vector notation

$$\begin{aligned} \nabla_x L(x, y) &= c - \mu X^{-1} e - A^T y = 0 \\ \nabla_y L(x, y) &= b - Ax = 0 \end{aligned}$$

where

$$X = \begin{pmatrix} x_1 & 0 & \cdots & 0 \\ 0 & x_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & x_n \end{pmatrix}$$

since  $x > 0$  the inverse  $X^{-1}$  exists

If we introduce  $s = \mu X^{-1} e$  then

$$\begin{aligned} c - s - A^T y &= 0 \\ b - Ax &= 0 \\ s &= \mu X^{-1} e \end{aligned}$$

or equivalent (Kuhn-Tucker optimality condition)

$$\begin{cases} A^T y + s = c \\ Ax = b \\ Xs = \mu e \end{cases} \quad \boxed{x > 0}$$

12

## Convexity of barrier function

If objective function is convex, then optimality condition is sufficient

$$cx - \mu \sum_{j=1}^n \ln(x_j)$$

- The function  $\ln(x)$  is concave, i.e.  $-\ln(x)$  is convex
- Barrier function is sum of convex functions
- Barrier function is convex

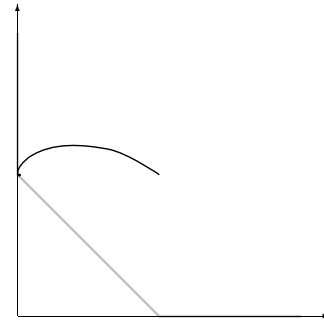
13

## Path following methods

primal *central path* is  $\{x(\mu) : \mu > 0\}$

dual *central path* is  $\{y(\mu), s(\mu) : \mu > 0\}$

### Example



14

## How large is the error?

$$\begin{aligned} A^T y + s &= c \\ Ax &= b \\ Xs &= \mu e \end{aligned}$$

A feasible solution  $(x, y, s)$  to the system is

- Primal feasible:  $Ax = b$  and  $x > 0$ .
- Dual feasible:  $A^T y + s = c$  and  $s = \mu X^{-1}e > 0$
- Duality gap:

$$\begin{aligned} cx - yb &= (yA + s)x - y(Ax) \\ &= xs \\ &= x(\mu X^{-1}e) \\ &= \mu ee = \mu n \end{aligned}$$

Since an optimal solution  $x^*$  must satisfy  $by \leq cx^* \leq cx$

$$cx - cx^* \leq n\mu$$

15

## Overall principle

Problem: find primal-dual pair  $(x, y, s)$  with  $x, s \geq 0$  such that

$$F(x, y, s) = \begin{pmatrix} Ax - b \\ yA + s - c \\ sx \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Algorithm:

- Select a tolerance:

$$\begin{aligned} \|Ax - b\| &\leq \epsilon_P \\ \|yA + s - c\| &\leq \epsilon_D \\ sx &\leq \epsilon_G \end{aligned}$$

- Choose an initial solution  $x > 0, y > 0, s > 0$
- Based on the tolerance choose a sufficiently small  $\mu$
- Use Newton's method on the barrier function
- Terminate when the above tolerance is satisfied

Note

$$\nabla F(x, y, s) = \begin{pmatrix} A & 0 & 0 \\ 0 & A & I \\ S & 0 & X \end{pmatrix}$$

16

## Step size, and convergence

$$\alpha^{\max} = \arg \max_{0 \leq \alpha} \left\{ \begin{pmatrix} x^k \\ s^k \end{pmatrix} + \alpha \begin{pmatrix} d_x \\ d_s \end{pmatrix} \geq 0 \right\}$$

and for some  $\theta \in ]0, 1[$

$$\alpha := \min(\theta \alpha^{\max}, 1)$$

ensures convergence. In practice  $\theta = 0.9$  is good (but no polynomial guarantee!)

## Number of steps

“interior point methods solve an LP in less than 100 iterations even if the problem contains millions of variables”

## Complexity of each step

Derive  $d_x, d_y, d_s$  from

$$\begin{pmatrix} A & 0 & 0 \\ 0 & A & I \\ S & 0 & X \end{pmatrix} \begin{pmatrix} d_x \\ d_y \\ d_s \end{pmatrix} = \begin{pmatrix} Ax - b \\ yA + s - c \\ sx \end{pmatrix}$$

next point

$$\begin{pmatrix} x^+ \\ y^+ \\ s^+ \end{pmatrix} = \begin{pmatrix} x^0 \\ y^0 \\ s^0 \end{pmatrix} + \alpha \begin{pmatrix} d_x \\ d_y \\ d_s \end{pmatrix}$$

17

## Dual problem, barrier function

Linear programming model

$$\begin{aligned} & \text{maximize } by \\ & \text{subject to } yA + s = c \\ & \quad \quad \quad s \geq 0, y \in \mathbb{R} \end{aligned}$$

Introduce barrier function

$$\begin{aligned} & \text{maximize } by + \mu \sum_{j=1}^n \ln(s_j) \\ & \text{subject to } yA + s = c \\ & \quad \quad \quad s > 0 \end{aligned}$$

Lagrangian function

$$L(x, y, s) = by + \mu \sum_{j=1}^n \ln(s_j) - x(yA + s - c)$$

Optimality conditions

$$\begin{aligned} \nabla_x L(x, y, s) &= c - s - yA = 0 \\ \nabla_y L(x, y, s) &= b - Ax = 0 \\ \nabla_s L(x, y, s) &= \mu S^{-1} e - x = 0 \end{aligned}$$

which can be reduced to

$$\begin{cases} yA + s = c, & \boxed{s > 0} \\ Ax = b \\ Xs = \mu e \end{cases}$$

Essentially the same as for primal problem

18

## Primal-dual approach

Two sets of optimality conditions

$$\begin{cases} Ax = b, & \boxed{x > 0} \\ yA + s = c, & \boxed{s > 0} \\ Xs = \mu e \end{cases}$$

perturbed KKT conditions. Let

$$F_\gamma(x, y, s) = \begin{pmatrix} Ax - b \\ yA + s - c \\ Xs - \gamma \mu e \end{pmatrix}$$

where  $\mu = xs/n$  and  $\gamma \geq 0$ . Assume  $(\bar{x}, \bar{y}, \bar{s})$  given. One iteration of Newton method to  $F_\gamma(\bar{x}, \bar{y}, \bar{s}) = 0$  i.e.

$$\nabla F_\gamma(\bar{x}, \bar{y}, \bar{s}) \begin{pmatrix} d_x \\ d_y \\ d_s \end{pmatrix} = -F_\gamma(\bar{x}, \bar{y}, \bar{s})$$

Since

$$\nabla F_\gamma(\bar{x}, \bar{y}, \bar{s}) = \begin{pmatrix} A & 0 & 0 \\ 0 & A & I \\ \bar{S} & 0 & \bar{X} \end{pmatrix}$$

we obtain

$$\begin{pmatrix} A & 0 & 0 \\ 0 & A & I \\ \bar{S} & 0 & \bar{X} \end{pmatrix} \begin{pmatrix} d_x \\ d_y \\ d_s \end{pmatrix} = \begin{pmatrix} \bar{r}_P \\ \bar{r}_D \\ -\bar{X}\bar{s} + \gamma \mu e \end{pmatrix}$$

primal residual  $\bar{r}_P = b - A\bar{x}$

dual residual  $\bar{r}_D = c - \bar{y}A - \bar{s}$

19

## The algorithm

1 choose  $(x^0, y^0, s^0)$  with  $x^0, s^0 > 0$ , and  $\epsilon_P, \epsilon_D, \epsilon_G > 0$

2 for  $k := 0$  to  $\infty$

3 calculate residuals

$$\begin{aligned} \bar{r}_P^k &= b - A\bar{x}^k \\ \bar{r}_D^k &= c - \bar{y}^k A - \bar{s}^k \\ \mu^k &= \bar{x}^k \bar{s}^k / n \end{aligned}$$

4 if  $\|\bar{r}_P^k\| \leq \epsilon_P, \|\bar{r}_D^k\| \leq \epsilon_D, \|\mu^k\| \leq \epsilon_G$  then stop

5 choose  $\gamma < 1$  and solve

$$\begin{pmatrix} A & 0 & 0 \\ 0 & A & I \\ \bar{S} & 0 & \bar{X} \end{pmatrix} \begin{pmatrix} d_x \\ d_y \\ d_s \end{pmatrix} = \begin{pmatrix} \bar{r}_P \\ \bar{r}_D \\ -\bar{X}\bar{s} + \gamma \mu e \end{pmatrix}$$

5 compute

$$\alpha^{\max} = \arg \max_{0 \leq \alpha} \left\{ \begin{pmatrix} x^k \\ s^k \end{pmatrix} + \alpha \begin{pmatrix} d_x \\ d_s \end{pmatrix} \geq 0 \right\}$$

6 for some  $\theta \in ]0, 1[$

$$\alpha := \min(\theta \alpha^{\max}, 1)$$

7 update

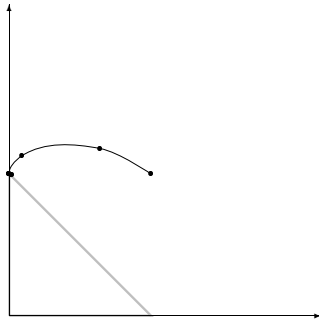
$$\begin{aligned} x^{k+1} &:= x^k + \alpha d_x, \\ y^{k+1} &:= y^k + \alpha d_y, \\ s^{k+1} &:= s^k + \alpha d_s \end{aligned}$$

20

## Numerical Example

Problem:

$$\begin{aligned} & \text{minimize} && -1x_1 - 2x_2 \\ & \text{subject to} && x_1 + x_2 \leq 1 \\ & && x_1, x_2 \geq 0 \end{aligned}$$



Iterations

k	$x_1^k$	$x_2^k$	$x^k s^k$
0	1.00e0	1.00e0	3.0e0
1	6.40e-1	1.18e0	1.1e0
2	0.91e-2	1.13e0	2.5e-1
3	0.91e-3	9.93e-1	4.1e-2
4	1.91e-3	9.98e-1	6.1e-3
5	2.42e-4	1.00e0	7.2e-4
6	2.45e-5	1.00e0	7.3e-5
7	2.45e-6	1.00e0	7.4e-6
8	2.56e-7	1.00e0	7.3e-7

21

## Worst-case time complexity

- *Simplex*: exponential, but heuristics often get round this problem
- *Interior*: approximate solution, polynomial time in  $n$  and  $\epsilon$

## Characterization of the iterative sequence

- *Simplex*: generate a sequence of feasible basic solutions
- *Interior*: generate a sequence of feasible primal and dual solutions
- *Simplex*: primal solution decreases monotonically
- *Interior*: Duality gap decreases monotonically
- *Simplex*: many iterations
- *Interior*: few iterations (20-30)
- *Simplex*: one simplex iteration  $O(n^2)$
- *Interior*: one iteration  $O(n^3)$

22

## Generated solutions

- *Simplex*: Returns an optimal basic solution
- *Interior*: Returns an  $\epsilon$ -optimal solution  $cx - by \leq \epsilon$
- *Interior*: Solution is at the analytic center of the optimal face
- *Interior*: Basis solution can be constructed in strongly polynomial time

## Initialization

- *Simplex*: Needs a feasible solution to start
- *Simplex*: First phase of algorithm finds feasible solution
- *Interior*: Self-dual version can be initialized by any positive vector

23

## Degeneracy

Basis solution  $x_B$  is degenerate if  $\exists i \in I_B : x_i = 0$

- *Simplex*: objective function remains same in simplex iteration, cycling
- *Interior*: theoretical and practical complexity is not affected by degeneracy

## Practical performance

- *Simplex*: often good performance despite worst-case exponential time
- *Interior*: best suited for huge problems, highly degenerate problems

## Warm-start

- *Simplex*: easy to warm-start from previous basis solution
- *Interior*: not as efficient as simplex for warm-start

24

## Integer-linear problems

- *Simplex*: clear winner, generate cuts from basic solutions, warm-start in branch-and-bound
- *Interior*: basic solution *can* be constructed, no warm-start

## Generalizations

- *Simplex*: generalized to nonlinear, and semi-infinite optimization problems
- *Interior*: same generalizations as simplex, moreover *conic* optimization

## Perspectivation

- Interior point methods is one of the best examples of theoretical goals leading to practical algorithms
- Interior point methods has started competition leading to innovation
- Interior point methods can be extended to a number of cones (*self-dual homogeneous cones*)
  - $\mathbb{R}^n$  (linear programming)
  - vectorized symmetric matrices over real numbers (semidefinite programming)
  - vectorized Hermitian matrices over complex numbers
  - vectorized Hermitian matrices over quaternions
  - vectorized Hermitian  $3 \times 3$  matrices over octonions
- With conic optimization we can solve “more” problems than we asked for
- Challenge to model builders to use new relaxations