DAT1

27/2/16

Agenda

- Review: optimisation
- Databases and MapReduce
- Dimensionality Reduction

Linear Programming

Find numbers x_1 and x_2 that maximize the sum $x_1 + x_2$ subject to the constraints $x_1 \geq 0,\ x_2 \geq 0$, and

$$\begin{aligned} x_1 + 2x_2 &\leq 4 \\ 4x_1 + 2x_2 &\leq 12 \\ -x_1 + x_2 &\leq 1 \end{aligned}$$

Try it!

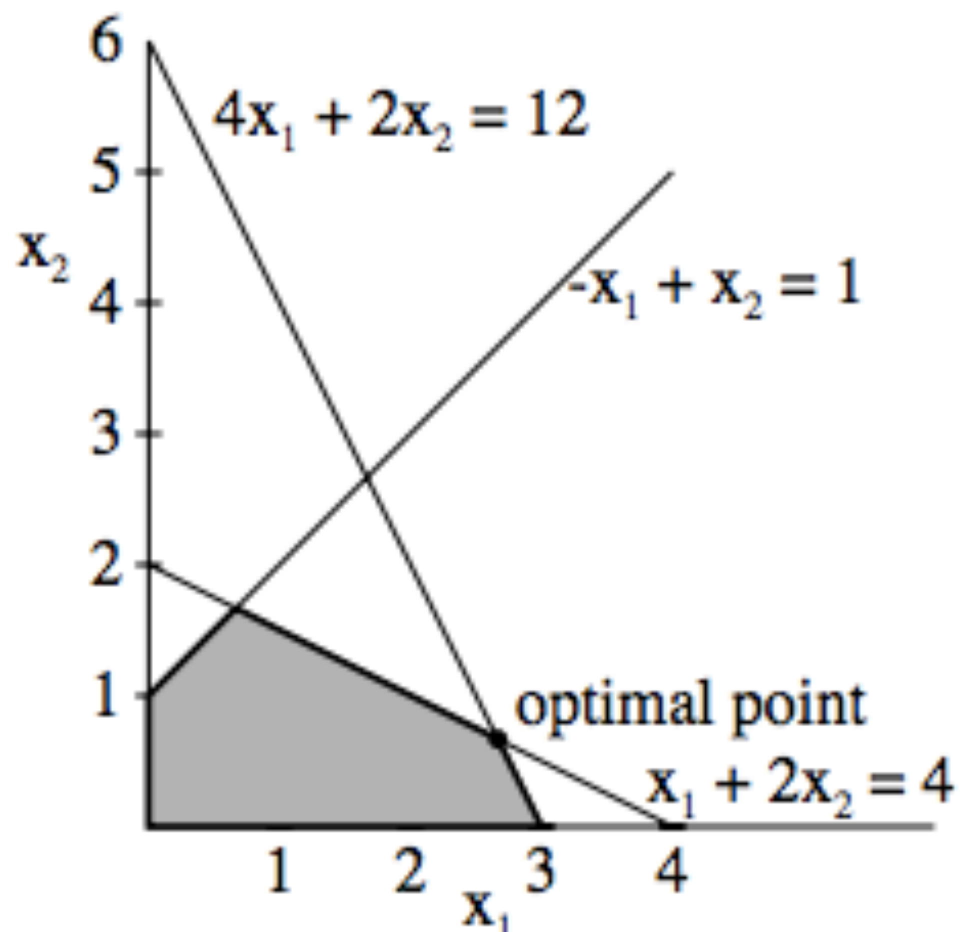


Figure 1.

Solvers: cvxopt in Matlab

$$\begin{array}{ll} \text{minimize} & 2x_1 + x_2 \\ \text{subject to} & -x_1 + x_2 \leq 1 \\ & x_1 + x_2 \geq 2 \\ & x_2 \geq 0 \\ & x_1 - 2x_2 \leq 4 \end{array}$$

as follows:

```
>>> from cvxopt import matrix, solvers
>>> A = matrix([ [-1.0, -1.0, 0.0, 1.0], [1.0, -1.0, -1.0, -2.0] ])
>>> b = matrix([ 1.0, -2.0, 0.0, 4.0 ])
>>> c = matrix([ 2.0, 1.0 ])
>>> sol=solvers.lp(c,A,b)
     pcost       dcost       gap    pres   dres   k/t
 0:  2.6471e+00 -7.0588e-01  2e+01  8e-01  2e+00  1e+00
 1:  3.0726e+00  2.8437e+00  1e+00  1e-01  2e-01  3e-01
 2:  2.4891e+00  2.4808e+00  1e-01  1e-02  2e-02  5e-02
 3:  2.4999e+00  2.4998e+00  1e-03  1e-04  2e-04  5e-04
 4:  2.5000e+00  2.5000e+00  1e-05  1e-06  2e-06  5e-06
 5:  2.5000e+00  2.5000e+00  1e-07  1e-08  2e-08  5e-08
>>> print(sol['x'])
[ 5.00e-01]
[ 1.50e+00]
```

General Linear Form

$$\begin{array}{ll} \text{Maximize} & \mathbf{c}^\top \mathbf{x} \\ \text{subject to} & \mathtt{A}\mathbf{x} \leq \mathbf{b}, \mathbf{x} \geq 0 \end{array}$$

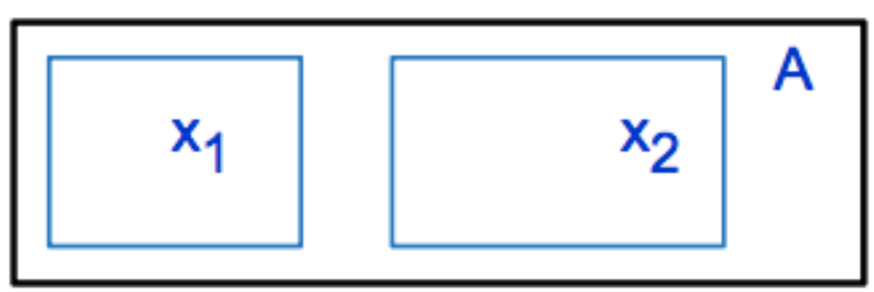
Example 1

Suppose that a farmer has a piece of farm land, say A square kilometers large, to be planted with either wheat or barley or some combination of the two.

The farmer has a limited permissible amount F of fertilizer and P of insecticide which can be used, each of which is required in different amounts per unit area for wheat (F_1, P_1) and barley (F_2, P_2).

Let S_1 be the selling price of wheat, and S_2 the price of barley, and denote the area planted with wheat and barley as x_1 and x_2 respectively.

The optimal number of square kilometers to plant with wheat vs. barley can be expressed as a linear programming problem.



Example 1 cont.

Maximize $S_1x_1 + S_2x_2$ (the revenue – this is the "objective function")

subject to $x_1 + x_2 \leq A$ (limit on total area)

$F_1x_1 + F_2x_2 \leq F$ (limit on fertilizer)

$P_1x_2 + P_2x_2 \leq P$ (limit on insecticide)

$x_1 >= 0,\ x_2 \geq 0$ (cannot plant a negative area)

which in matrix form becomes

maximize

$$\begin{bmatrix} S_1 & S_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

subject to

$$\begin{bmatrix} 1 & 1 \\ F_1 & F_2 \\ P_1 & P_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \leq \begin{bmatrix} A \\ F \\ P \end{bmatrix}, \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \geq \mathbf{0}$$

Example 2: Max flow

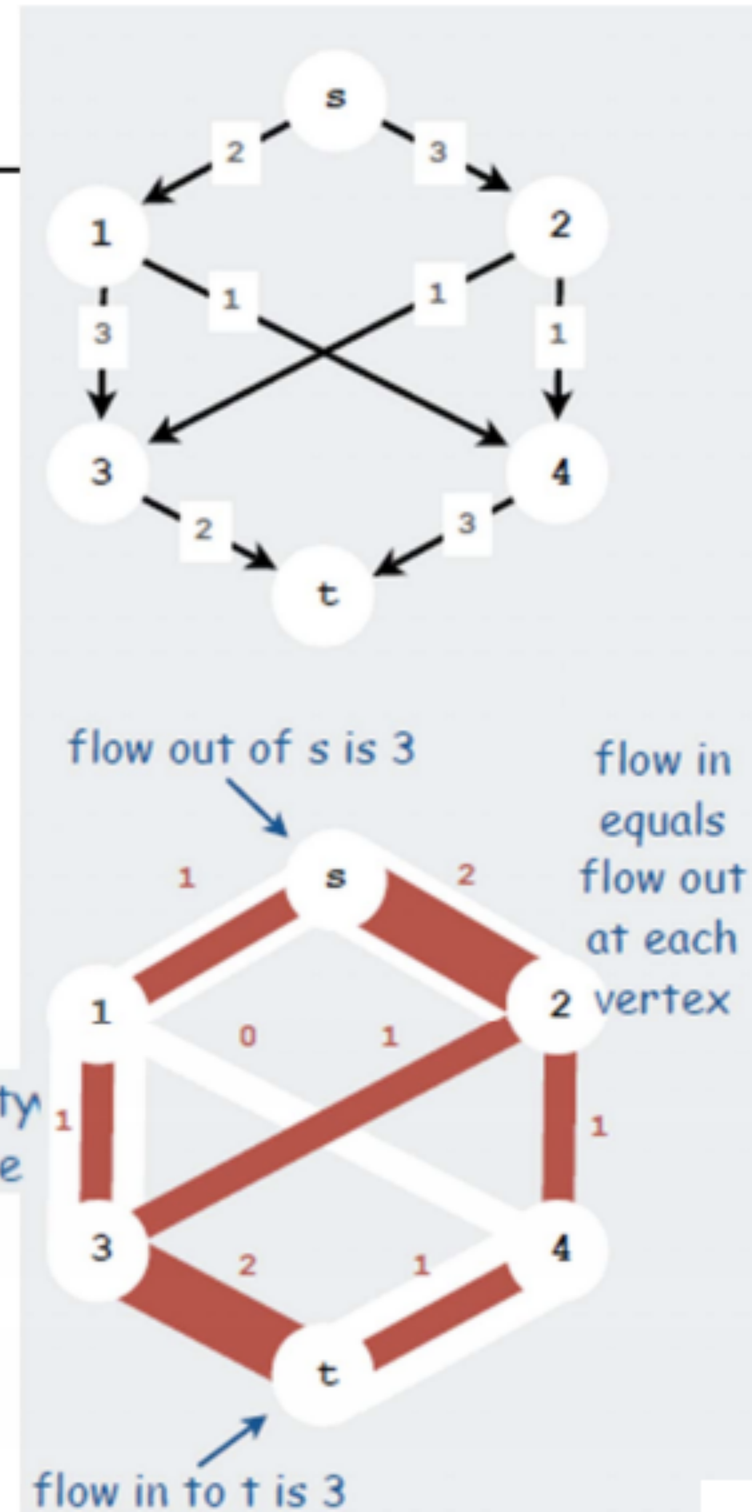
Given: a weighted directed graph, source s, destination t

Interpret edge weights as capacities

Goal: Find maximum flow from s to t

- Flow does not exceed capacity in any edge
- Flow at every vertex satisfies equilibrium

[flow in equals flow out]



e.g. oil flowing through pipes, internet routing

LP formulation of maxflow problem

One variable per edge.

One inequality per edge, one equality per vertex.

maximize x_{ts}

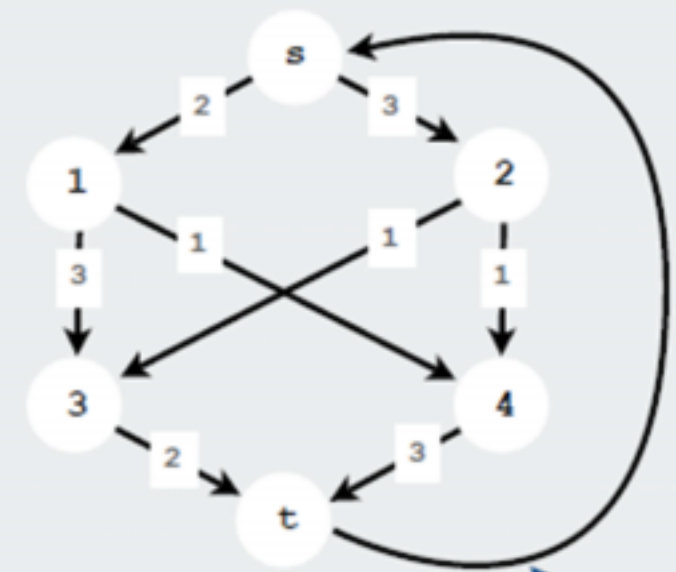
subject to the constraints

$$
\left.\begin{aligned}
x_{s1} &\le 2 \\
x_{s2} &\le 3 \\
x_{13} &\le 3 \\
x_{14} &\le 1 \\
x_{23} &\le 1 \\
x_{24} &\le 1 \\
x_{3t} &\le 2 \\
x_{4t} &\le 3
\end{aligned}\right\} \text{capacity constraints}
$$

$$
\text{equilibrium constraints} \left\{\begin{aligned}
x_{ts} &= x_{s1} + x_{s2} \\
x_{s1} &= x_{13} + x_{14} \\
x_{s2} &= x_{23} + x_{24} \\
x_{13} + x_{23} &= x_{3t} \\
x_{14} + x_{24} &= x_{4t} \\
x_{3t} + x_{4t} &= x_{ts}
\end{aligned}\right.
$$

all $x_{ij} \ge 0$

interpretation:
x_{ij} = flow in edge i-j



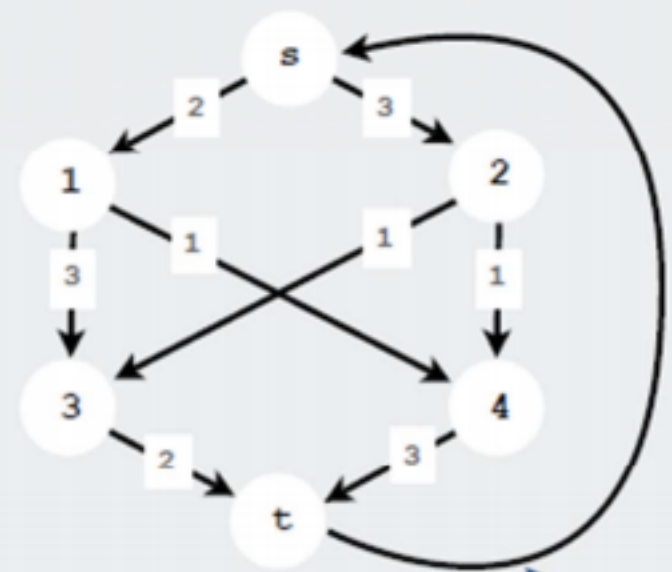
add dummy edge from t to s

Slide: Robert Sedgewick and Kevin Wayne

LP formulation of maxflow problem

One variable per edge.

One inequality per edge, one equality per vertex.



add dummy edge from t to s

maximize x_{ts}

subject to the constraints

$$
\left.\begin{aligned}
x_{s1} &\le 2\\
x_{s2} &\le 3\\
x_{13} &\le 3\\
x_{14} &\le 1\\
x_{23} &\le 1\\
x_{24} &\le 1\\
x_{3t} &\le 2\\
x_{4t} &\le 3
\end{aligned}\right\}\ \text{capacity constraints}
$$

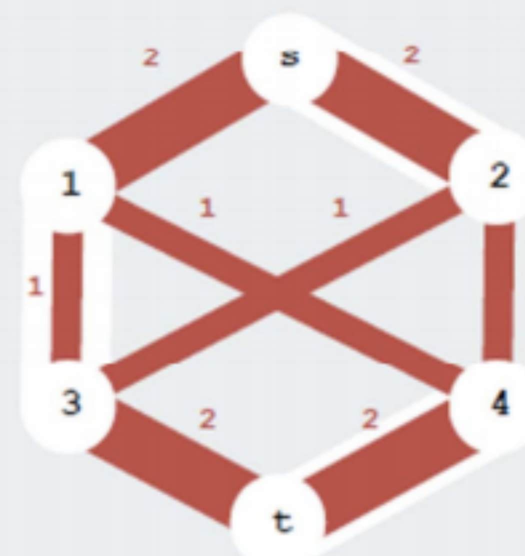
interpretation:
x_{ij} = flow in edge i-j

$$
\text{equilibrium constraints}\ \left\{\begin{aligned}
x_{ts} &= x_{s1} + x_{s2}\\
x_{s1} &= x_{13} + x_{14}\\
x_{s2} &= x_{23} + x_{24}\\
x_{13} + x_{23} &= x_{3t}\\
x_{14} + x_{24} &= x_{4t}\\
x_{3t} + x_{4t} &= x_{ts}
\end{aligned}\right.
$$

all $x_{ij} \ge 0$

solution

$x_{s1} = 2$
$x_{s2} = 2$
$x_{13} = 1$
$x_{14} = 1$
$x_{23} = 1$
$x_{24} = 1$
$x_{3t} = 2$
$x_{4t} = 2$
$x_{ts} = 4$ maxflow value

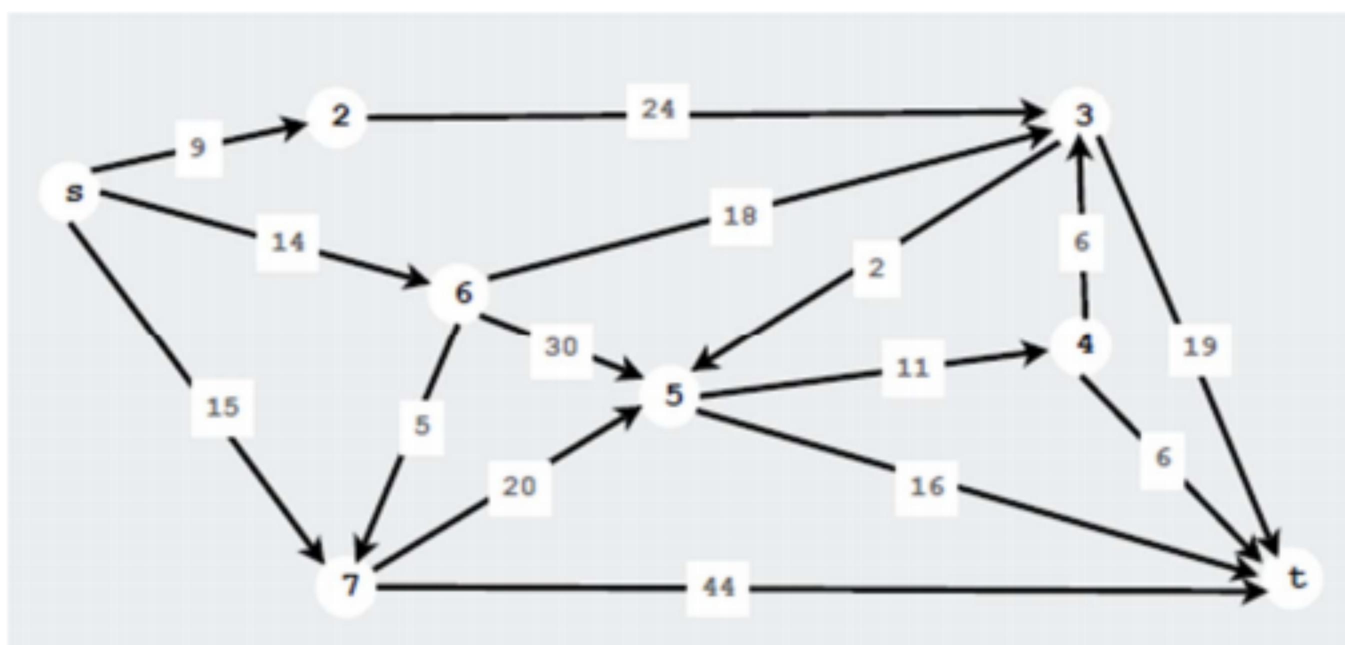


Example 3: shortest path

Given: a weighted directed graph, with a single source s

Distance from s to v: length of the shortest part from s to v

Goal: Find distance (and shortest path) to every vertex



e.g. plotting routes on Google maps

Application

Minimize number of stops (lengths = 1)

Minimize amount of time (positive lengths)

General Convex Form

Optimization problem in standard form

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p \end{array}$$

- $x \in \mathbf{R}^n$ is the optimization variable
- $f_0 : \mathbf{R}^n \to \mathbf{R}$ is the objective or cost function
- $f_i : \mathbf{R}^n \to \mathbf{R}$, $i = 1, \ldots, m$, are the inequality constraint functions
- $h_i : \mathbf{R}^n \to \mathbf{R}$ are the equality constraint functions

optimal value:

$$p^\star = \inf\{f_0(x) \mid f_i(x) \leq 0,\ i = 1, \ldots, m,\ h_i(x) = 0,\ i = 1, \ldots, p\}$$

- $p^\star = \infty$ if problem is infeasible (no x satisfies the constraints)