# Dual-fitting analysis of Greedy for Set Cover

We showed earlier that the greedy algorithm for set cover gives a $H_n$ approximation

We will show that greedy produces a solution of cost at most $H_n \, OPT_{LP}$

Note that $OPT_{LP} \leq OPT$

For this we need the dual of the LP for set cover

# Primal and Dual for Set cover

Primal: $\min \sum_{i=1}^{m} c(i)\, x(i)$

$\sum_{i:\, e \in S_i} x(i) \geq 1$ for each $e \in \mathcal{U}$

$x(i) \geq 0 \quad 1 \leq i \leq m$

Dual: $\max \sum_{e \in \mathcal{U}} y(e)$

$\sum_{e \in S_i} y(e) \leq c(i)$ for $1 \leq i \leq m$

$y(e) \geq 0 \quad e \in \mathcal{U}$

# Dual-fitting

Recall the analysis of $H_n$ approximation for greedy. Essentially dual-fitting in disguise as we see now.

Let $h_j$ be the index of set picked by Greedy in iteration $j$

Let $S'_{h_j}$ be the *new* elements covered by $S_{h_j}$ : that is those elements covered in iteration $j$

for each $e$ in $S'_{h_j}$ we set $p(e) = c(h_j)/|S'_{h_j}|$

# Dual-fitting

Note that by construction $\sum_e p(e)$ = cost of solution output by Greedy

Define $y'(e) = p(e)/H_n$

Lemma: $y'$ is a *feasible* solution to the dual LP

Assuming lemma, $\sum_e y'(e) \leq OPT_{LP}$ (by weak duality).

Therefore cost of Greedy $\leq H_n\ OPT_{LP}$

# Dual-fitting

Define $y'(e) = p(e)/H_n$

Lemma: $y'$ is a *feasible* solution to the dual LP

Need to show that for any set $S_i$

$\sum_{e \text{ in } S_i} y'(e) \leq c(i)$ or in other words

$\sum_{e \text{ in } S_i} p(e) \leq H_n\, c(i)$

Renumber elements s.t $e_1, e_2, \ldots, e_n$ are covered in that order

# Dual-fitting

Renumber elements s.t $e_1, e_2, \ldots, e_n$ are covered in that order

For a set $S_i$ let its elements be $e_{i_1}, e_{i_2}, \ldots, e_{i_t}$ where $|S_i| = t$

for $l = 1$ to $t$ we claim that $p(e_{i_l}) \leq c(i)/(t-l+1)$

This is because when $e_{i_l}$ was covered, Greedy could have picked $S_i$ at density $c(i)/(t-l+1)$

# Dual-fitting

for $l = 1$ to $t$ we claim that $p(e_{i_l}) \leq c(i)/(t - l + 1)$

This is because when $e_{i_l}$ was covered, Greedy could have picked $S_i$ at density $c(i)/(t-l+1)$

Therefore $\sum_{e \in S_i} p(e) \leq \sum_{l=1}^{t} c(i)/(t-l+1) \leq c(i)\, H_t$

Remark: notice that the above analysis shows that Greedy's cost is can be upper bounded by $H_k\ OPT_{LP}$ where $k$ is the size of the largest set

# Primal Dual for Set Cover

Primal-dual method is a general paradigm - see Vazirani's book for extensive discussion

Here we apply it to obtain an f-approximation for Set Cover where f is the max # of sets that any element belongs to

Vertex Cover is a special case with $f = 2$ and hence this yields a 2-approx for weighted vertex cover

# Primal Dual for Set Cover

We obtain a feasible *integral* primal solution x and a feasible dual solution y

The pair (x,y) satisfy primal complementary slackness condition

$x(i) > 0$ implies $\sum_{e \in S_i} y(e) = c(i)$

since x is an integral solution

$x(i) = 1$ implies $\sum_{e \in S_i} y(e) = c(i)$

# Cost of solution

If we can find such a pair $(x, y)$ then cost of solution is
$\sum_i c(i)\, x(i) = \sum_i x(i) \sum_{e \in S_i} y(e)$

Changing the order of summation

$\leq \sum_e y(e) \sum_{i:\, e \in S_i} x(i)$

$\leq \sum_e y(e)\, f$

Since $y$ is feasible $\sum_e y(e) \leq OPT_{LP}$

Therefore cost of solution $\leq f\, OPT_{LP} \leq f\, OPT$

# Primal-Dual algorithm

Start with feasible dual $y = 0$ and infeasible primal $x = 0$

While some element is not covered

- increase $y(e)$ for all uncovered $e$ *uniformly* until some constraint becomes tight (that is for some $i$, $\sum_{e \in S_i} y(e) = c(i)$ )
- for all tight $i$, set $x(i) = 1$ and all elements in $S_i$ are covered

# Primal-Dual

By induction on while loop iterations it can easily be seen that

y remains a feasible solution through out the algorithm

thus at end of algorithm all elements covered are covered (by x) and y is feasible

by construction x(i) = 1 iff i is tight hence primal complementary slackness is satisfied for pair (x, y)

# Integral Polyhedra

A rational polyhedron given by a system of inequalities $Ax \leq b$ is *integral* iff all its vertices have integer coords

Theorem: The polyhedron $Ax \leq b$ is integral iff for each integral $w$, the optimum value of $\max wx$ s.t $Ax \leq b$ is an integer if it is finite.

# Totally Unimodular Matrices

A 0,1 matrix A is called totally unimodular (TUM) iff for each square submatrix A' of A,
$\det(A') \in \{0,1,-1\}$

A simple consequence of the definition is the following

Theorem: For *any* integer vector b and TUM matrix A, the polyhedron $A\,x \leq b$ is an integral polyhedron

# Totally Unimodular Matrices

Theorem: For *any* integer vector $b$ and TUM matrix $A$, the polyhedron $A x \leq b$ is an integral polyhedron

Proof: Any vertex of the polyhedron is given by the solution to the system $A' x' = b'$ for some square non-singular sub-matrix $A'$ of $A$.

Therefore $x' = (A')^{-1} b' = ((A')^t/\det(A'))\, b'$

Since $A$ is TUM, $\det(A') \in \{+1, -1\}$. Therefore $x'$ is integral since both $A'$ and $b'$ are integral

# TUM matrices

Several operations preserve total unimodularity
For example we can add box constraints to $Ax \leq b$

The system $A x \leq b, \ u \leq x \leq l$ for integer $b, u, l$ is an integral polyhedron if $A$ is TUM

The dual of $\max c x, A \leq b$ is $\min y b, y A = c$

If $c, b$ are integer then both primal and dual are integer polyhedra

# Examples of TUM matrices

Given a directed graph $G = (V, E)$ consider the adjacency matrix $A = V \times E$ with

$A(u, e) = +1$ if $e = (u, v)$ for some $v \in V$

$A(u, e) = -1$ if $e = (v, u)$ for some $v \in V$

$A(u, e) = 0$ if $u$ is not the head or tail of $e$

$A$ is TUM

From above one can deduce integrality of single-commodity flows with integer capacities and also the maxflow-mincut theorem

# Bipartite graphs

Bipartite graph G = (U, V, E) with U, V as the vertex sets of the two partitions

Adjacency matrix $A = U \times V$ with
$A(u, v) = 1$ if $uv \in E$

$= 0$ otherwise

A is TUM

From above one can deduce Konig's theorem and Hall's theorems etc regarding matchings and vertex covers in bipartite graphs

# Consecutive 1's matrices

Let A be a 0,1 matrix such that the 1's in each row occur consecutively. Then A is TUM.

From above one can solve various optimization problems on interval graphs

# Network matrices

The previous three cases are examples of *network matrices.* A network matrix is obtained from a directed graph $G=(V, E)$ and a directed tree $T = (V, E')$ on the same vertex set. Note that the arcs of T can be oriented in any way. Given G, T we obtain a matrix A = E x E' as follows

For $e = (u,v)$ in E let $P_{uv}$ be the *unique* path from u to v in T (the path is simply a list of edges in E')

$A(e, e') = +1$ if e' occurs in the direction of traversal of $P_{uv}$ from u to v

$A(e, e') = -1$ if e' occurs in the opposite direction of traversal from u to v

$A(e, e') = 0$ if e' is not in $P_{uv}$

# Network matrices

Theorem: Network matrices are TUM

*Not all TUM matrices are network matrices*

Exercise: show how each of the three classes seen before are examples of network matrices

(difficult theorem of Seymour)

Theorem: Given a matrix A, poly-time algorithm to check if A is TUM

# Integer decomposability of TUM matrices

Useful to know following.

Theorem: Let $x$ be a *feasible* point in $A x \leq b$ for some TUM $A$ and integral $b$. Let $k x$ be integral for some integer $k$. Then $k x = x_1 + x_2 + \ldots + x_k$ for integral vectors $x_1, \ldots, x_k$ each of which is a feasible point in $A x \leq b$.

Above also holds for system $A x \leq b, l \leq x \leq u$ for integral bounds $l, u$