

Chapter 4

Greedy Algorithms



Greedy Algorithms



A greedy algorithm always makes the choice that looks best at the moment

- Our everyday examples:
 - Walking & Travelling
 - Queueing
- The hope: a locally optimal choice will lead to a globally optimal solution
- For some problems, it works
- But in most cases, it doesn't work



Dynamic programming can be overkill; greedy algorithms tend to be easier to code

Greedy Analysis Strategies

Greedy algorithm stays ahead. Show that after each step of the greedy algorithm, its solution is at least as good as any other algorithm's.

Exchange argument. Gradually transform any solution to the one found by the greedy algorithm without hurting its quality.

Structural. Discover a simple "structural" bound asserting that every possible solution must have a certain value. Then show that your algorithm always achieves this bound.





4.1 Interval Scheduling

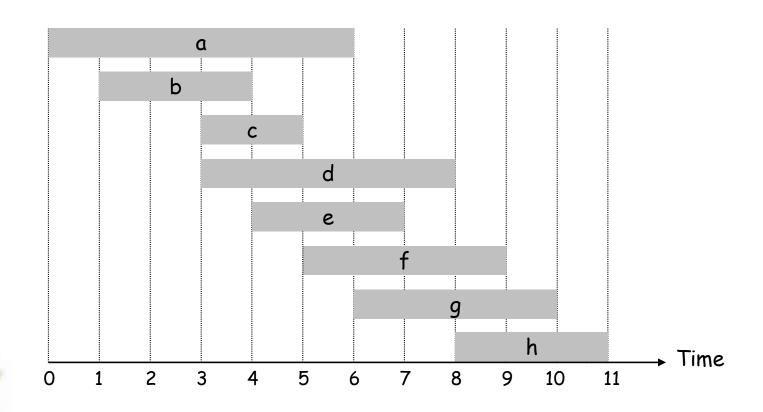


Interval Scheduling



Interval scheduling.

- Job j starts at s_j and finishes at f_j .
- Two jobs compatible if they don't overlap.
- Goal: find maximum subset of mutually compatible jobs.



Interval Scheduling: Greedy Algorithms

lad §

Greedy template. Consider jobs in some order. Take each job provided it's compatible with the ones already taken.

- [Earliest start time] Consider jobs in ascending order of start time s_{j} .
- [Earliest finish time] Consider jobs in ascending order of finish time f_j .
- [Shortest interval] Consider jobs in ascending order of interval length f_j s_j .
- [Fewest conflicts] For each job, count the number of conflicting jobs c_j . Schedule in ascending order of conflicts c_j .



Interval Scheduling: Greedy Algorithms

ad V

Greedy template. Consider jobs in some order. Take each job provided it's compatible with the ones already taken.



Interval Scheduling: Greedy Algorithm

Greedy algorithm. Consider jobs in increasing order of finish time. Take each job provided it's compatible with the ones already taken.

Implementation. O(n log n).

- Remember job j* that was added last to A.
 - Job j is compatible with A if $s_j \ge f_{j*}$.

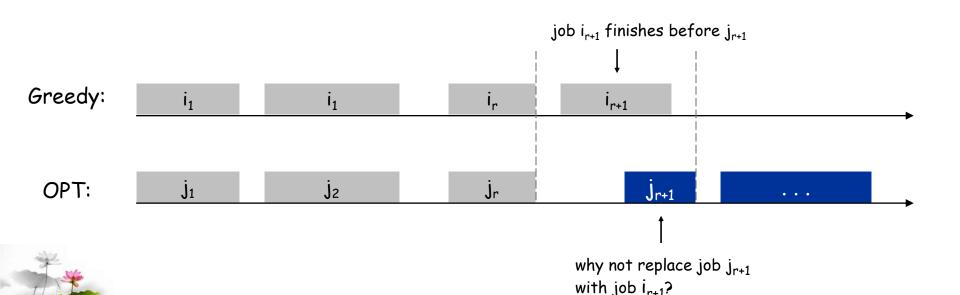
Interval Scheduling: Analysis



Theorem. Greedy algorithm is optimal.

Pf. (by contradiction)

- Assume greedy is not optimal, and let's see what happens.
- Let i_1 , i_2 , ... i_k denote set of jobs selected by greedy.
- Let j_1 , j_2 , ... j_m denote set of jobs in the optimal solution with $i_1 = j_1$, $i_2 = j_2$, ..., $i_r = j_r$ for the largest possible value of r.



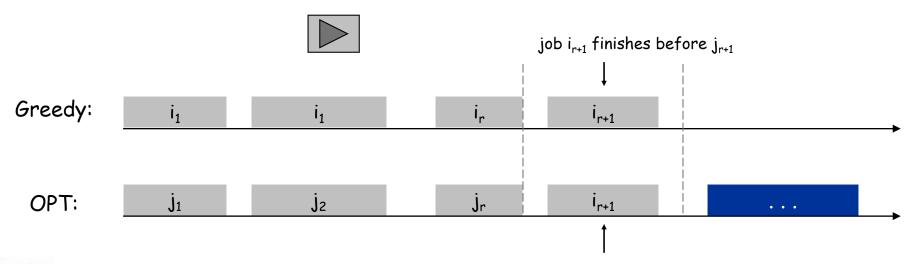
Interval Scheduling: Analysis



Theorem. Greedy algorithm is optimal.

Pf. (by contradiction)

- Assume greedy is not optimal, and let's see what happens.
- Let i_1 , i_2 , ... i_k denote set of jobs selected by greedy.
- Let j_1 , j_2 , ... j_m denote set of jobs in the optimal solution with $i_1 = j_1$, $i_2 = j_2$, ..., $i_r = j_r$ for the largest possible value of r.



solution still feasible and optimal, but contradicts maximality of r.



4.1 Interval Partitioning



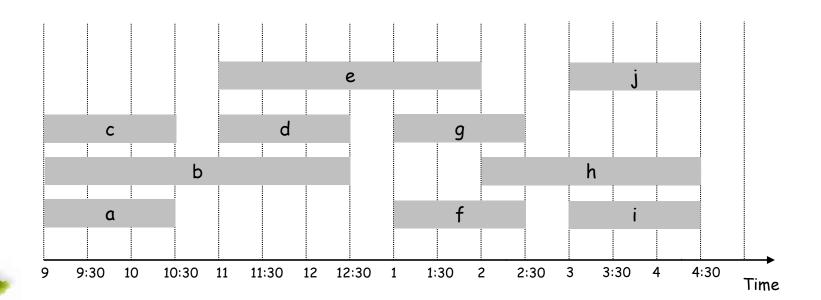
Interval Partitioning



Interval partitioning.

- Lecture j starts at s_j and finishes at f_j .
- Goal: find minimum number of classrooms to schedule all lectures so that no two occur at the same time in the same room.

Ex: This schedule uses 4 classrooms to schedule 10 lectures.



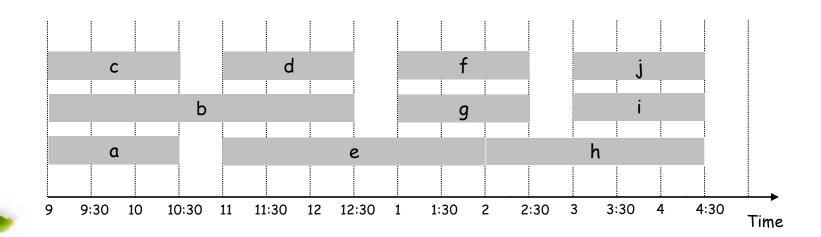
Interval Partitioning



Interval partitioning.

- Lecture j starts at s_j and finishes at f_j .
- Goal: find minimum number of classrooms to schedule all lectures so that no two occur at the same time in the same room.

Ex: This schedule uses only 3.



Interval Partitioning: Lower Bound on Optimal Solution

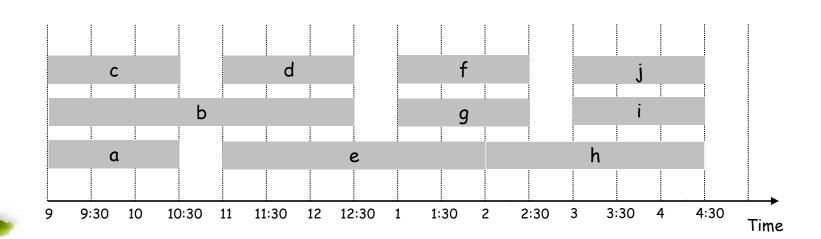
Def. The depth of a set of open intervals is the maximum number that contain any given time.

Key observation. Number of classrooms needed ≥ depth.

Ex: Depth of schedule below = $3 \Rightarrow$ schedule below is optimal.

a, b, c all contain 9:30

Q. Does there always exist a schedule equal to depth of intervals?



Interval Partitioning: Greedy Algorithm



Greedy algorithm. Consider lectures in increasing order of start time: assign lecture to any compatible classroom.

```
Sort intervals by starting time so that s_1 \leq s_2 \leq \ldots \leq s_n. d \leftarrow 0 \leftarrow \text{number of allocated classrooms}

for j = 1 to n \in \{1 \text{ if (lecture } j \text{ is compatible with some classroom } k) \}
\text{schedule lecture } j \text{ in classroom } k \in \{1 \text{ else} \}
\text{allocate a new classroom } k \in \{1 \text{ else} \}
\text{allocate } k \in \{1 \text{ else } k \in \{1 \text{ else} \}
\text{allocate } k \in \{1 \text{ else } k \in \{
```

Implementation. O(n log n).

- For each classroom k, maintain the finish time of the last job added.
 - Keep the classrooms in a priority queue.

Interval Partitioning: Greedy Analysis

A STATE OF THE STA

Observation. Greedy algorithm never schedules two incompatible lectures in the same classroom.

Theorem. Greedy algorithm is optimal. Pf.

- Let d = number of classrooms that the greedy algorithm allocates.
- Classroom d is opened because we needed to schedule a job, say j, that is incompatible with all d-1 other classrooms.
- Since we sorted by start time, all these incompatibilities are caused by lectures that start no later than s_j .
- Thus, we have d lectures overlapping at time $s_j + \epsilon$.
- Key observation \Rightarrow all schedules use \geq d classrooms. \blacksquare





4.2 Scheduling to Minimize Lateness



Scheduling to Minimizing Lateness

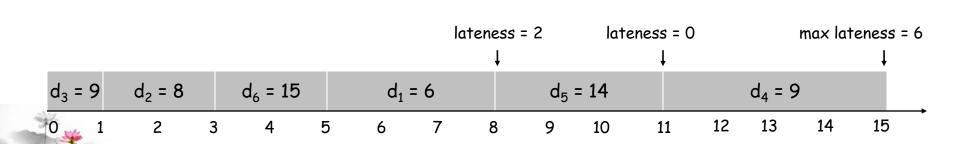


Minimizing lateness problem.

- Single resource processes one job at a time.
- Job j requires t_j units of processing time and is due at time d_j .
- If j starts at time s_j , it finishes at time $f_j = s_j + t_j$.
- Lateness: $l_j = \max \{ 0, f_j d_j \}$.
- Goal: schedule all jobs to minimize maximum lateness $L = \max_{j} l_{j}$.

Ex:

	1	2	3	4	5	6
† _j	3	2	1	4	3	2
dj	6	8	9	9	14	15





Greedy template. Consider jobs in some order.

[Shortest processing time first] Consider jobs in ascending order of processing time t_j .

[Earliest deadline first] Consider jobs in ascending order of deadline d_j .

[Smallest slack] Consider jobs in ascending order of slack $d_i - t_i$.





Greedy template. Consider jobs in some order.

[Shortest processing time first] Consider jobs in ascending order of processing time t_i .

	1	2
† _j	1	10
dj	100	10

counterexample

[Smallest slack] Consider jobs in ascending order of slack $d_j - t_j$.

	1	2
† _j	1	10
dj	2	10

Counterexample, wrong





Greedy template. Consider jobs in some order.

[Shortest processing time first] Consider jobs in ascending order of processing time t_j .

	1	2
† _j	1	10
dj	100	10

[Smallest slack] Consider jobs in ascending order of slack d_j - t_j .

	1	2
† _j	2	5
dj	4	6



Greedy template. Consider jobs in some order.

[Shortest processing time first] Consider jobs in ascending order of processing time $t_{\rm j}$.

	1	2
tj	1	10
dj	100	10

[Smallest slack] Consider jobs in ascending order of slack d_j - t_j .

	1	2
† _j	2	7
dj	6	9

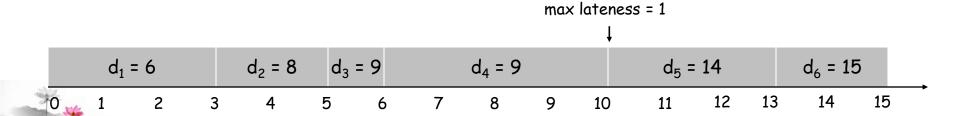
Counterexample..





Greedy algorithm. Earliest deadline first.

```
Sort n jobs by deadline so that d_1 \leq d_2 \leq ... \leq d_n t \leftarrow 0 for j = 1 to n  \text{Assign job j to interval } [t, \ t + t_j]  s_j \leftarrow t, \ f_j \leftarrow t + t_j  t \leftarrow t + t_j output intervals [s_j, \ f_j]
```



Minimizing Lateness: No Idle Time



Observation. There exists an optimal schedule with no idle time.

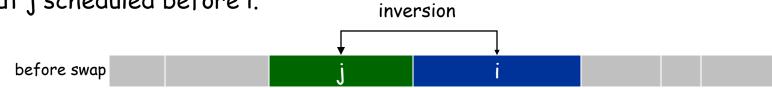
	d = 4			d :	= 6				d =	12	
0	1	2	3	4	5	6	7	8	9	10	11
	d = 4		d :	= 6		d =	: 12				
0	1	2	3	4	5	6	7	8	9	10	11

Observation. The greedy schedule has no idle time.



Minimizing Lateness: Inversions

Def. An inversion in schedule 5 is a pair of jobs i and j such that: i < j but j scheduled before i.



Observation. Greedy schedule has no inversions.

Observation. If a schedule (with no idle time) has an inversion, it has one with a pair of inverted jobs scheduled consecutively.

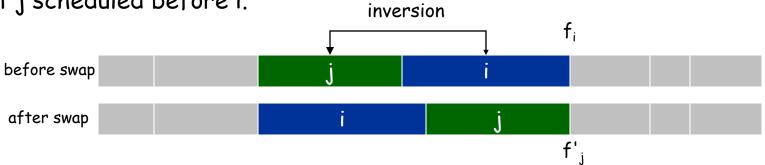


Minimizing Lateness: Inversions



Def. An inversion in schedule S is a pair of jobs i and j such that:

i < j but j scheduled before i.



Claim. Swapping two adjacent, inverted jobs reduces the number of inversions by one and does not increase the max lateness.

Pf. Let L be the lateness before the swap, and let L' be it afterwards.

$$l'_k = l_k$$
 for all $k \neq i, j$

$$_{n}$$
 $|'_{i} \leq |_{i}$

If job j is late:

$$\ell'_{j} = f'_{j} - d_{j}$$
 (definition)
 $= f_{i} - d_{j}$ (j finishes at time f_{i})
 $\leq f_{i} - d_{i}$ (i < j)
 $\leq \ell_{i}$ (definition)

Minimizing Lateness: Analysis of Greedy Algorithm



Theorem. Greedy schedule S is optimal.

- Pf. Define 5* to be an optimal schedule that has the fewest number of inversions, and let's see what happens.
 - Can assume 5* has no idle time.
 - If S^* has no inversions, then $S = S^*$.
 - If S* has an inversion, let i-j be an adjacent inversion.
 - swapping i and j does not increase the maximum lateness and strictly decreases the number of inversions
 - this contradicts definition of S* •



Greedy Analysis Strategies

Greedy algorithm stays ahead. Show that after each step of the greedy algorithm, its solution is at least as good as any other algorithm's.

Exchange argument. Gradually transform any solution to the one found by the greedy algorithm without hurting its quality.

Structural. Discover a simple "structural" bound asserting that every possible solution must have a certain value. Then show that your algorithm always achieves this bound.





4.3 Optimal Caching



Optimal Offline Caching



Caching.

- . Cache with capacity to store k items.
- Sequence of m item requests d_1 , d_2 , ..., d_m .
- . Cache hit: item already in cache when requested.
- requested item into cache, and evict some existing item, if full.

Goal. Eviction schedule that minimizes number of cache misses.

Ex: k = 2, initial cache = ab,requests: a, b, c, b, c, a, a, b.Optimal eviction schedule: 2 cache misses.

α	а	b
b	а	b
С	С	b
b	С	b
С	С	b
α	а	b
α	а	b
b	а	b
requests	cad	che

Optimal Offline Caching: Farthest-In-Future



Farthest-in-future. Evict item in the cache that is not requested until farthest in the future.

current cache:

a b c d e f

future queries:

g a b c e d a b b a c d e a f a d e f g h ...

t cache miss

eject this one

Theorem. [Bellady, 1960s] FF is optimal eviction schedule. Pf. Algorithm and theorem are intuitive; proof is subtle.



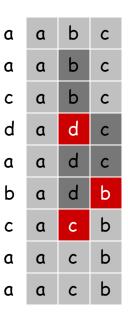
Reduced Eviction Schedules

Def. A reduced schedule is a schedule that only inserts an item into the cache in a step in which that item is requested.

Intuition. Can transform an unreduced schedule into a reduced one with no more cache misses.

α	а	Ь	С
а	а	×	С
С	а	d	С
d	α	d	Ь
α	α	С	Ь
b	а	х	Ь
С	а	С	Ь
α	а	b	С
а	а	Ь	С

an unreduced schedule



a reduced schedule

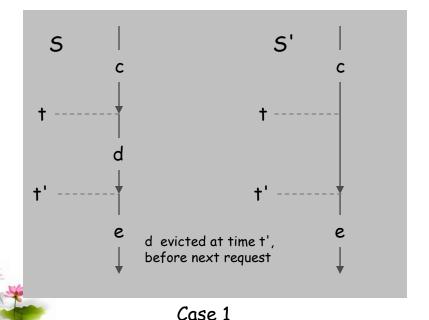


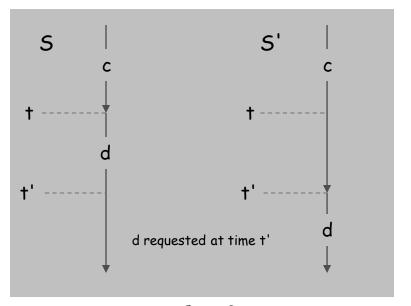
Reduced Eviction Schedules

Claim. Given any unreduced schedule 5, can transform it into a reduced schedule 5' with no more cache misses.

Pf. (by induction on number of unreduced items) time

- Suppose S brings d into the cache at time t, without a request.
- Let c be the item S evicts when it brings d into the cache.
- Case 1: d evicted at time t', before next request for d.
- . Case 2: d requested at time t' before d is evicted.





Case 2

Farthest-In-Future: Analysis



Theorem. FF is optimal eviction algorithm.

Pf. (by induction on number or requests j)

Invariant: There exists an optimal reduced schedule S that makes the same eviction schedule as S_{FF} through the first j+1 requests.

Let S be reduced schedule that satisfies invariant through j requests. We produce S' that satisfies invariant after j+1 requests.

- Consider $(j+1)^{s+1}$ request $d = d_{j+1}$.
- Since S and S_{FF} have agreed up until now, they have the same cache contents before request j+1.
- Case 1: (d is already in the cache). S' = S satisfies invariant.
- Case 2: (d is not in the cache and S and S_{FF} evict the same element). S' = S satisfies invariant.

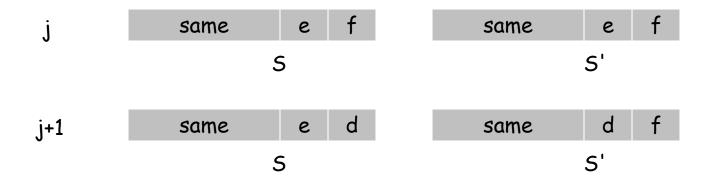


Farthest-In-Future: Analysis



Pf. (continued)

- Case 3: (d is not in the cache; S_{FF} evicts e; S evicts $f \neq e$).
 - begin construction of S' from S by evicting e instead of f



- now S' agrees with S_{FF} on first j+1 requests; we show that having element f in cache is no worse than having element e



Farthest-In-Future: Analysis

Let j' be the first time after j+1 that S and S' take a different action, and let g be item requested at time j'. \uparrow must involve e or f (or both)

j'	same	е	same	f
	5		5'	

- Case 3a: g = e. Can't happen with Farthest-In-Future since there must be a request for f before e.
- Case 3b: g = f. Element f can't be in cache of S, so let e' be the element that S evicts.
 - if e' = e, S' accesses f from cache; now S and S' have same cache
 - if e' ≠ e, S' evicts e' and brings e into the cache; now S and S' have the same cache

Note: S' is no longer reduced, but can be transformed into a reduced schedule that agrees with S_{FF} through step j+1

Farthest-In-Future: Analysis

Let j' be the first time after j+1 that S and S' take a different action, and let g be item requested at time j'. \uparrow must involve e or f (or both)



otherwise S' would take the same action

Case 3c: $g \neq e$, f. S must evict e. Make S' evict f; now S and S' have the same cache.

j' same g same g
S



Caching Perspective



Online vs. offline algorithms.

- Offline: full sequence of requests is known a priori.
- Online (reality): requests are not known in advance.
- Caching is among most fundamental online problems in CS.

LIFO. Evict page brought in most recently.

LRU. Evict page whose most recent access was earliest.

FF with direction of time reversed!

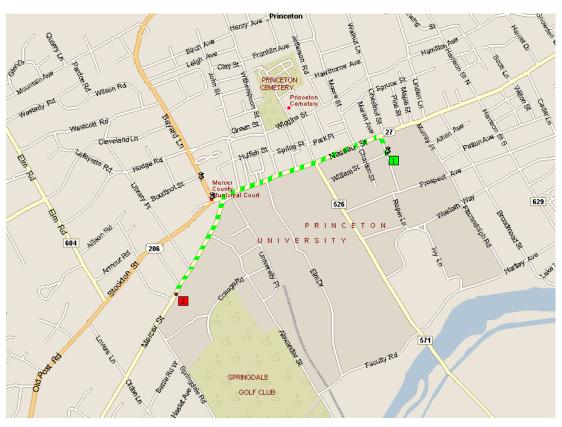
Theorem. FF is optimal offline eviction algorithm.

- Provides basis for understanding and analyzing online algorithms.
- LRU is k-competitive. [Section 13.8]
- LIFO is arbitrarily bad.





4.4 Shortest Paths in a Graph



shortest path from Princeton CS department to Einstein's house

Shortest Path Problem

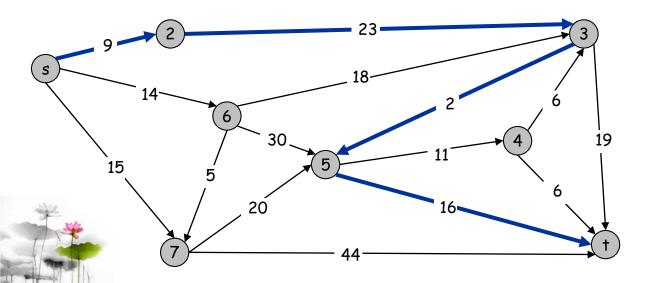


Shortest path network.

- Directed graph G = (V, E).
- Source s, destination t.
- Length l_e = length of edge e.

Shortest path problem: find shortest directed path from s to t.

cost of path = sum of edge costs in path



Cost of path s-2-3-5-t = 9 + 23 + 2 + 16 = 48.

Dijkstra's Algorithm



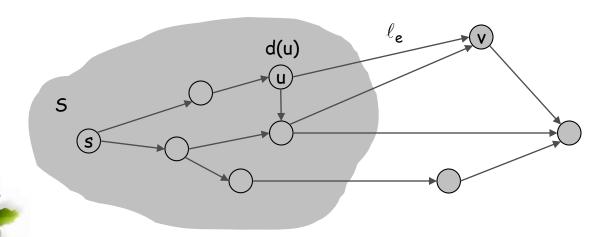
Dijkstra's algorithm.

- Maintain a set of explored nodes S for which we have determined the shortest path distance d(u) from s to u.
- Initialize $S = \{s\}, d(s) = 0.$
- Repeatedly choose unexplored node v which minimizes

$$\pi(v) = \min_{e = (u,v): u \in S} d(u) + \ell_e,$$

add v to S, and set $d(v) = \pi(v)$.

shortest path to some u in explored part, followed by a single edge (u, v)



Dijkstra's Algorithm



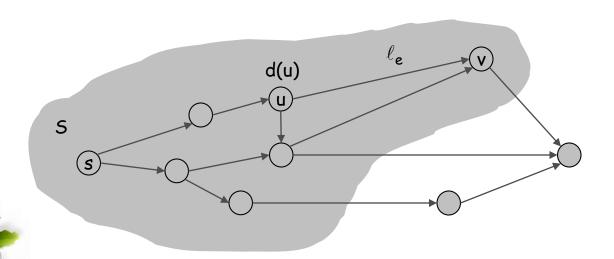
Dijkstra's algorithm.

- Maintain a set of explored nodes S for which we have determined the shortest path distance d(u) from s to u.
- Initialize $S = \{s\}, d(s) = 0.$
- Repeatedly choose unexplored node v which minimizes

$$\pi(v) = \min_{e = (u,v): u \in S} d(u) + \ell_e,$$
set $d(u) = \pi(u)$

add v to S, and set $d(v) = \pi(v)$.

shortest path to some u in explored part, followed by a single edge (u, v)



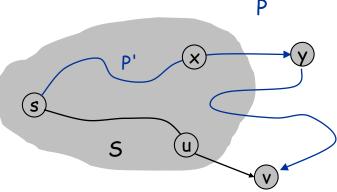
Dijkstra's Algorithm: Proof of Correctness

Invariant. For each node $u \in S$, d(u) is the length of the shortest s-u path. Pf. (by induction on |S|)

Base case: |S| = 1 is trivial.

Inductive hypothesis: Assume true for $|S| = k \ge 1$.

- Let v be next node added to S, and let u-v be the chosen edge.
- The shortest s-u path plus (u, v) is an s-v path of length $\pi(v)$.
- Consider any s-v path P. We'll see that it's no shorter than $\pi(v)$.
- Let x-y be the first edge in P that leaves S, and let P' be the subpath to x.
- P is already too long as soon as it leaves S.



$$\ell \ (P) \ge \ell \ (P') + \ell \ (x,y) \ge \ d(x) + \ell \ (x,y) \ge \ \pi(y) \ge \ \pi(v)$$

$$\uparrow \qquad \qquad \uparrow \qquad \qquad \uparrow$$

$$nonnegative \qquad inductive \qquad defn of \pi(y) \qquad Dijkstra \ chose \ v$$

$$weights \qquad hypothesis \qquad instead \ of \ y$$

Dijkstra's Algorithm: Implementation



For each unexplored node, explicitly maintain $\pi(v) = \min_{e = (u,v) : u \in S} d(u) + \ell_e$.

- Next node to explore = node with minimum $\pi(v)$.
- When exploring v, for each incident edge e = (v, w), update

$$\pi(w) = \min \{ \pi(w), \pi(v) + \ell_e \}.$$

Efficient implementation. Maintain a priority queue of unexplored nodes, prioritized by $\pi(v)$.

PQ Operation	Dijkstra	Array	Binary heap	d-way Heap	Fib heap [†]
Insert	n	n	log n	d log _d n	1
ExtractMin	n	n	log n	d log _d n	log n
ChangeKey	m	1	log n	log _d n	1
IsEmpty	n	1	1	1	1
Total		n ²	m log n	m log _{m/n} n	m + n log n

[†] Individual ops are amortized bounds



Homework

Homework: 3,4,6,7





Selecting Breakpoints



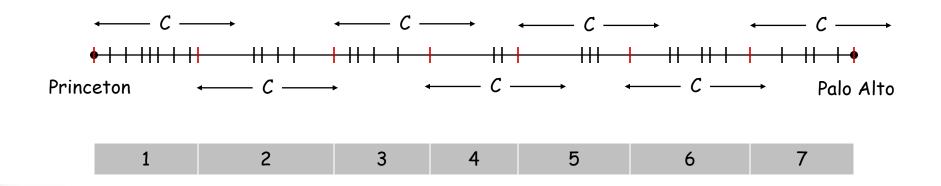
Selecting Breakpoints



Selecting breakpoints.

- Road trip from MIT to CalTech along fixed route.
- Refueling stations at certain points along the way.
- Fuel capacity = C.
- Goal: makes as few refueling stops as possible.

Greedy algorithm. Go as far as you can before refueling.



Selecting Breakpoints: Greedy Algorithm



Truck driver's algorithm.

```
Sort breakpoints so that: 0 = b_0 < b_1 < b_2 < \ldots < b_n = L
S \leftarrow \{0\} \leftarrow \text{breakpoints selected}
x \leftarrow 0 \leftarrow \text{current location}
\text{while } (x \neq b_n)
\text{let p be largest integer such that } b_p \leq x + C
\text{if } (b_p = x)
\text{return "no solution"}
x \leftarrow b_p
S \leftarrow S \cup \{p\}
\text{return S}
```

Implementation. O(n log n)

Use binary search to select each breakpoint p.

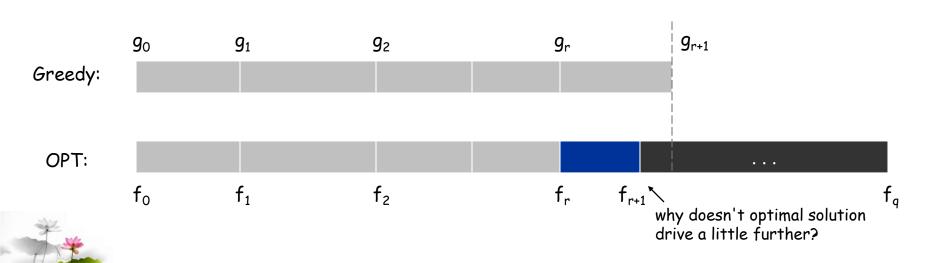
Selecting Breakpoints: Correctness



Theorem. Greedy algorithm is optimal.

Pf. (by contradiction)

- Assume greedy is not optimal, and let's see what happens.
- Let $0 = g_0 < g_1 < \ldots < g_p = L$ denote set of breakpoints chosen by greedy.
- Let $0 = f_0 < f_1 < ... < f_q = L$ denote set of breakpoints in an optimal solution with $f_0 = g_0, f_1 = g_1, ..., f_r = g_r$ for largest possible value of r.
- Note: $g_{r+1} > f_{r+1}$ by greedy choice of algorithm.



Selecting Breakpoints: Correctness



Theorem. Greedy algorithm is optimal.

Pf. (by contradiction)

- Assume greedy is not optimal, and let's see what happens.
- Let $0 = g_0 < g_1 < \ldots < g_p = L$ denote set of breakpoints chosen by greedy.
- Let $0 = f_0 < f_1 < ... < f_q = L$ denote set of breakpoints in an optimal solution with $f_0 = g_0, f_1 = g_1, ..., f_r = g_r$ for largest possible value of r.
- Note: $g_{r+1} > f_{r+1}$ by greedy choice of algorithm.

