Chapter 4
Greedy Algorithms

Section 4.1: 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.

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

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

- Breaks earliest start time
- Breaks shortest interval
- Breaks fewest conflicts

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

Sort jobs by finish times so that $f_1 \leq f_2 \leq \ldots \leq f_n$.

jobs selected

A $\leftarrow \phi$

for $j = 1$ to $n$

if (job $j$ compatible with $A$)

A $\leftarrow A \cup \{j\}$

return A

Implementation. $O(n \log n)$.
- Remember job $j^*$ that was added last to $A$.
- Job $j$ is compatible with $A$ if $s_j \geq f_{j^*}$.

Interval Scheduling: Analysis

**Theorem.** Greedy algorithm is optimal.

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

Greedy: $i_1, i_2, \ldots, i_r, i_{r+1}$

job $i_{r+1}$ finishes before $j_{r+1}$

why not replace job $j_{r+1}$ with job $i_{r+1}$?

OPT: $j_1, j_2, \ldots, j_r, j_{r+1}$

OPT: $j_1, j_2, \ldots, j_r, j_{r+1}$

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

**Theorem.** Greedy algorithm is optimal.

**Pf.** (by contradiction)
- Assume greedy is not optimal, and let’s see what happens.
- Let $i_1, i_2, \ldots, i_k$ denote set of jobs selected by greedy.
- Let $j_1, j_2, \ldots, j_m$ denote set of jobs in the optimal solution with $i_1 = j_1, i_2 = j_2, \ldots, i_r = j_r$, for the largest possible value of $r$.
4.1 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: 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 $\geq$ depth.

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

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 \( s_1 \leq s_2 \leq \ldots \leq s_n \).

\[
d \leftarrow 0 \quad \text{--- number of allocated classrooms}
\]

For each classroom \( k \), maintain the finish time of the last job added.

Keep the classrooms in a priority queue.

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

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

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.

4.2 Scheduling to Minimize 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: \( \lambda_j = \max \{ 0, f_j - d_j \} \).
- Goal: schedule all jobs to minimize maximum lateness \( L = \max \lambda_j \).

Ex:

<table>
<thead>
<tr>
<th>( d_j )</th>
<th>9</th>
<th>8</th>
<th>15</th>
<th>6</th>
<th>14</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>( t_j )</td>
<td>2</td>
<td>4</td>
<td>5</td>
<td>14</td>
<td>2</td>
<td>6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>lateness = 2</th>
<th>lateness = 0</th>
<th>max lateness = 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>( d_j = 9 )</td>
<td>( d_j = 8 )</td>
<td>( d_j = 15 )</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>
Minimizing Lateness: Greedy Algorithms

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 $d_j - t_j$.

Minimizing Lateness: Greedy Algorithm

Greedy algorithm. Earliest deadline first.

Sort $n$ jobs by deadline so that $d_1 \leq d_2 \leq \ldots \leq d_n$.

$t \leftarrow 0$

for $j = 1$ to $n$

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 $[s_j, f_j]$

max lateness = 1

Minimizing Lateness: Greedy Algorithms

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 $d_j - t_j$.

Minimizing Lateness: No Idle Time

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

Observation. The greedy schedule has no idle time.
Minimizing Lateness: Inversions

**Def.** An inversion in schedule \( S \) is a pair of jobs \( i \) and \( j \) such that:
\[
d_i < d_j \text{ but } j \text{ scheduled before } i.
\]

**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.

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

**Pf.** Let \( \lambda \) be the lateness before the swap, and let \( \lambda' \) be it afterwards.
- \( \lambda' = \lambda \) for all \( k \neq i, j \)
- \( \lambda' \) is strictly less than \( \lambda \)
- If job \( j \) is late:
\[
\lambda_i = f'_i - d_i = f_i - d_i = f_i - d_j = f'_j - d_j = \lambda
\]

Minimizing Lateness: Analysis of Greedy Algorithm

**Theorem.** Greedy schedule \( S \) is optimal.

**Pf.** Define \( S^* \) to be an optimal schedule that has the fewest number of inversions, and let’s see what happens.
- Can assume \( S^* \) 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

- Cache with capacity to store $k$ items.
- Sequence of $m$ item requests $d_1, d_2, \ldots, d_m$.
- Cache hit: item already in cache when requested.
- Cache miss: item not already in cache when requested; must bring 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: $a, b, c, b, a, b$.

**Optimal eviction schedule:** 2 cache misses.

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.

Optimal Offline Caching: Farthest-In-Future

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

**Theorem.** [Bellady, 1960s] FF is optimal eviction schedule.

**Pf.** Algorithm and theorem are intuitive; proof is subtle.
Reduced Eviction Schedules

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

Pf. (by induction on number of unreduced items)

- 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.

Farthest-In-Future: Analysis

Theorem. FF is optimal eviction algorithm.

Pf. (by induction on number of 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.

Farthest-In-Future: Analysis (continued)

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

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 $j'$.

Pf. (continued)

- 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 $e' 
eq e$, $S'$ 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 \( j' \). Let \( e \) or \( f \) (or both) must involve \( e \) or \( f \) (or both).

\[
\begin{array}{c|c|c}
  j' & S & S' \\
  \text{same} & e & \text{same} \\
\end{array}
\]

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.

\[
\begin{array}{c|c|c}
  j' & S & S' \\
  \text{same} & g & \text{same} \\
\end{array}
\]

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.

Theorem. FF is optimal offline eviction algorithm.
- Provides basis for understanding and analyzing online algorithms.
- LRU is \( k \)-competitive. \( \text{[Section 13.8]} \)
- LIFO is arbitrarily bad.

4.4 Shortest Paths in a Graph

Shortest Path Problem

Shortest path network:
- Directed graph \( G = (V, E) \).
- Source \( s \), destination \( t \).
- Length \( \ell_e \) = length of edge \( e \).

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

Cost of path = sum of edge costs in path

Shortest path from Princeton CS department to Einstein’s house

- Cost of path \( s-2-3-5-t \)
  \( = 9 + 23 + 2 + 16 = 48 \)
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
add $v$ to $S$, and set $d(v) = \pi(v)$.

\[ d(u) + \ell(u,v) \]

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 $|S| = k \geq 1$.

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 a $s$-$v$ 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 $S$.

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

nonnegative weights
inductive hypothesis

Dijkstra chose $v$ instead of $y$

Dijkstra’s Algorithm: Implementation

For each unexplored node, explicitly maintain $d(v) = \min_{e \in (v,u) \cup S} \ell(u,v)$.

Next node to explore = node with minimum $\pi(v)$.

When exploring $v$, for each incident edge $e = (v, w)$, update

$\pi(v) = \min \{ \pi(w), \pi(v) + \ell(e) \}$

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

<table>
<thead>
<tr>
<th>PQ Operation</th>
<th>Dijkstra</th>
<th>Array</th>
<th>Binary heap</th>
<th>d-way heap</th>
<th>Fibonacci heap</th>
</tr>
</thead>
<tbody>
<tr>
<td>Insert</td>
<td>$n$</td>
<td>$n$</td>
<td>$\log n$</td>
<td>$d \log n$</td>
<td>$n \log n$</td>
</tr>
<tr>
<td>ExtractMin</td>
<td>$n$</td>
<td>$n$</td>
<td>$\log n$</td>
<td>$d \log n$</td>
<td>$n \log n$</td>
</tr>
<tr>
<td>ChangeKey</td>
<td>$m$</td>
<td>$1$</td>
<td>$\log n$</td>
<td>$d \log n$</td>
<td>$n \log n$</td>
</tr>
<tr>
<td>InsertEmpty</td>
<td>$n$</td>
<td>$1$</td>
<td>$1$</td>
<td>$1$</td>
<td>$1$</td>
</tr>
<tr>
<td>Total</td>
<td>$m^2$</td>
<td>$m \log n$</td>
<td>$m \log n$</td>
<td>$m \log n$</td>
<td>$m \log n$</td>
</tr>
</tbody>
</table>

† Individual ops are amortized bounds