Weighted interval scheduling

Weighted Interval Scheduling (WIS) problem.

- Job \( j \) starts at \( s_j \), finishes at \( f_j \), and has weight or value \( v_j \).
- Two are jobs \emph{compatible} if they don’t overlap.
- Goal: find \emph{maximum-weight/ max-value} subset of mutually compatible jobs.
Recursive subproblems

two cases:

• $j_n$ is part of the optimal schedule $O$
  • recurse on the last job compatible to $j_n$
• $j_n$ is not part of $O$
  • recurse on job $j_{n-1}$

We will explore these two options to find the full solution

The recursive step corresponds to solving a subproblem:

• a problem considering fewer jobs
• note that the subset of jobs is sequential — it contains all jobs before a certain index.
**WIS — notation for compatibility**

**Notation.** Label jobs by finishing time: $f_1 \leq f_2 \leq \ldots \leq f_n$.

**Def.** $p(j) =$ largest index $i < j$ s.t. job $i$ is compatible with $j$. (if none, then $p(j) = 0$)

**Ex.** $p(8) = 5$, $p(7) = 3$, $p(2) = 0$.

$OPT(i) =$ maximum total value selection from jobs $1, 2, \ldots, i$

**Observation:**

Suppose job 8 is part of the optimal solution. Since $p(8) = 5$, the maximum value containing $j_8$ is $OPT(n) = OPT(5) + 8$
DP for WIS: recursive formula

Notation. \( OPT(j) \) = opt solution, i.e. max total value selection from jobs 1, 2, ..., \( j \).
\( OPT(n) \) = value of optimal solution to the original problem.

Case 1. \( OPT(j) \) selects job \( j \).
  • Collect profit \( v_j \).
  • Can’t use incompatible jobs \{ \( p(j) + 1 \), \( p(j) + 2 \), ..., \( j - 1 \) \}.
  • Must include optimal solution to problem consisting of remaining compatible jobs 1, 2, ..., \( p(j) \). This is \( OPT(p(j)) \).

Case 2. \( OPT(j) \) does not select job \( j \).
  • Must include optimal solution to problem consisting of remaining jobs 1, 2, ..., \( j - 1 \). This is \( OPT(j-1) \).

Recursive formula: Choose the better from Case 1 and 2

\[
OPT(j) = \begin{cases} 
0 & \text{if } j = 0 \\
\max \left\{ v_j + OPT(p(j)), \ OPT(j-1) \right\} & \text{otherwise}
\end{cases}
\]
Algorithm 1: NaiveRecursiveWIS(n jobs: $s_i, f_i, v_i$)

1. sorted ← sort jobs by increasing finish time $f_1 < \ldots < f_n$;
2. Compute $p(1), p(2), \ldots, p(n)$ /* can be done in $O(n)$ */
3. return RecOpt(n)

Algorithm 1: RecOpt(job index $j$)

1. if $j == 0$ then
2. | return 0
3. else
4. | $Opt(j) \leftarrow \max\{v_j + \text{RecOpt}(p(j)); \text{RecOpt}(j - 1)\}$;
5. | return $Opt(j)$

Running time: $\Omega(2^n)$
WIS — DP algorithm (recursive)

\[
OPT(j) = \begin{cases} 
0 & \text{if } j = 0 \\
\max \{ v_j + OPT(p(j)), \ OPT(j-1) \} & \text{otherwise}
\end{cases}
\]

Memoization table \(M\):

\(M[j] = OPT(j)\), array that contains the max value for jobs 0,1,…,j

---

**Algorithm 1:** WIS(n jobs: \(s_j, f_j, v_j\))

1. Sort jobs by finish time \(f_1 \leq \ldots \leq f_n\);
2. Compute \(p(1), p(2), \ldots p(n)\);
3. \(M \leftarrow \text{array}(n+1)\)// Empty array of size n+1, indexed 0…n
4. \(M[0] \leftarrow 0\)// no jobs selected
5. return WISCompute(n);

**Algorithm 2:** WISCompute(j)

1. if \(M[j]\) is empty then
2. \(M[j] \leftarrow \max\{v_j + WISCompute(p(j)) + WISCompute(j-1)\};
3. return \(M[j]\);
bottom-up algorithm to compute the optimal solution for WIS

**Algorithm 1: WIS(n jobs: s_j, f_j, v_j)**

1. Sort jobs by finish time \( f_1 \leq \ldots \leq f_n \);
2. Compute \( p(1), p(2), \ldots, p(n) \);
3. \( M \leftarrow \text{array}(n + 1) // \text{empty array of size } n+1, \text{ indexed } 0 \ldots n \)
4. \( M[0] \leftarrow 0; \)
5. for \( j = 1 \) to \( n \) do
6. \hspace{1em} \( M[j] \leftarrow \max \{ v_j + M[p(j)]; M[j - 1] \} \);
7. \textbf{return} \( M[n] \);

Running time?
WIS — DP algorithm — how to write a complete solution

1.: clearly define the subproblems, with proper indexing

\[ \text{OPT}(j) = \text{maximum value selection from job requests } 1, \ldots, j \]

2.: write the recursive formula:

\[ p(j) = \max \{ i : i < j \text{ and job } i \text{ is compatible with } j \} = \text{highest index of a job that doesn’t overlap with } j. \]

\[ \text{OPT}(j) = \begin{cases} 
0 & \text{if } j = 0 \\
\max \left\{ v_j + \text{OPT}(p(j)), \text{OPT}(j-1) \right\} & \text{otherwise}
\end{cases} \]

3.: bottom-up algorithm to compute the optimal solution

Algorithm 1: WIS\((n \text{ jobs: } s_j, f_j, v_j)\)

1. Sort jobs by finish time \(f_1 \leq \ldots \leq f_n\);
2. Compute \(p(1), p(2) \ldots p(n)\);
3. \( M \leftarrow \text{array}(n+1) // \text{empty array fo size } n+1, \text{indexed } 0 \ldots n \)
4. \( M[0] \leftarrow 0; \)
5. for \( j = 1 \) to \( n \) do
6. \( \quad M[j] \leftarrow \max\{v_j + M[p(j)]; M[j-1]\}; \)
7. return \( M[n]; \)

4.: use backtracking to find set of jobs in optimal solution
WIS — DP algorithm (bottom-up/iterative)

Weighted Interval Scheduling: given $n$ jobs, each with start time $s_j$, finish time $f_j$ and value $v_j$ find the compatible schedule with maximum total value.

$\text{OPT}(j) = \text{optimal solution for jobs } (0), 1, 2, \ldots, n$

$$OPT(j)=\begin{cases} 0 & \text{if } j=0 \\ \max \{ v_j + OPT(p(j)),\ OPT(j-1) \} & \text{otherwise} \end{cases}$$

Memoization table $M$:

$M[j] = \text{OPT}(j)$, array that contains the max value for jobs $0, 1 \ldots, j$

---

**Algorithm 1: WIS($n$ jobs: $s_j, f_j, v_j$)**

1. Sort jobs by finish time $f_1 \leq \ldots \leq f_n$;
2. Compute $p(1), p(2) \ldots p(n)$;
3. $M \leftarrow \text{array}(n + 1)$ // empty array fo size $n+1$, indexed $0 \ldots n$
4. $M[0] \leftarrow 0$;
5. for $j = 1$ to $n$ do
6. \hspace{1em} $M[j] \leftarrow \max\{v_j + M[p(j)]; M[j - 1]\}$;
7. return $M[n]$;
WIS — DP algorithm (top-down/recursive)

\[ OPT(j) = \begin{cases} 
0 & \text{if } j = 0 \\
\max \{ v_j + OPT(p(j)), \ OPT(j-1) \} & \text{otherwise}
\end{cases} \]

Memoization table \( M \):

\( M[j] = OPT(j) \), array that contains the max value for jobs 0,1...,j

---

**Algorithm 1:** WIS\( (n \) jobs: \( s_j, f_j, v_j)\)

1. Sort jobs by finish time \( f_1 \leq \ldots \leq f_n \);
2. Compute \( p(1), p(2), \ldots p(n) \);
3. \( M \leftarrow array(n+1) \)/ Empty array of size \( n+1 \), indexed 0...n
4. \( M[0] \leftarrow 0 \)/ no jobs selected
5. return \( WISCompute(n) \);

---

**Algorithm 2:** WISCompute\( (j)\)

1. if \( M[j] \) is empty then
2. \( M[j] \leftarrow \max \{ v_j + WISCompute(p(j)) + WISCompute(j - 1) \} \);
3. return \( M[j] \);
Finding the set of optimal jobs — backtracking

A dynamic programming algorithm computes the optimal value.

How to find the solution itself?
We can reconstruct it from the table.

- backtrack based on the memoization table without explicitly storing values (by checking which case was chosen)

---

**Algorithm 1: FindSolution(\(j\))**

```
1 if \(j == 0\) then
2  return \(\emptyset\);
3 else if \(v_j + M[p(j)] > M[j - 1]\) then
4  return \(\{j\} \cup \text{FindSolution}(p(j))\);
5 else
6  return \text{FindSolution}(j - 1);
```
Finding the set of optimal jobs – backtracking

\[
OPT(j) = \begin{cases} 
0 & \text{if } j = 0 \\
\max \{ \nu_j + OPT(p(j)), OPT(j-1) \} & \text{otherwise}
\end{cases}
\]

<table>
<thead>
<tr>
<th>j</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>OPT(j)</td>
<td>$0$</td>
<td>$4$</td>
<td>$3$</td>
<td>$10$</td>
<td>$10$</td>
<td>$12$</td>
<td>$19$</td>
<td>$19$</td>
<td>$20$</td>
</tr>
<tr>
<td>predecessor</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>3 (or 6)</td>
<td>5</td>
</tr>
</tbody>
</table>

Finding the set of optimal jobs — backtracking
Problem: Given an integer $n$, find the minimum number of operations to get from 0 to $n$, if you are only allowed to perform two specific operations: (1.) *add* 1 (2.) *multiply* by 2.

example compute 12:

$0 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 = 12$  //12 operations

$(((0+1) \times 2 ) \times 2) + 1 +1 +1 +1 = 12$  // 7 operations

$(((0 +1) +1 + 1) \times 2 ) \times 2 = 12$  // 5 operations

Question: Suppose $\text{OPT}(j)$ is the minimum number of operations required to make the number $j$. What is the recursive formula for computing $\text{OPT}(j)$? we may assume the base case $\text{OPT}(0) = 0$.

A. $\text{OPT}(j) = \begin{cases} \text{OPT}(j - 1) & \text{if } j \text{ is odd} \\ \text{OPT}(j/2) & \text{if } j \text{ is even} \end{cases}$

B. $\text{OPT}(j) = \begin{cases} \text{OPT}(j - 1) + 1 & \text{if } j \text{ is odd} \\ \text{OPT}(j/2) + 1 & \text{if } j \text{ is even} \end{cases}$

C. $\text{OPT}(j) = \begin{cases} \text{OPT}(j - 1) + 1 & \text{if } j \text{ is odd} \\ 1 + \min\{\text{OPT}(j/2); \text{OPT}(j - 1)\} & \text{if } j \text{ is even} \end{cases}$

D. $\text{OPT}(j) = \begin{cases} \text{OPT}(j - 1) & \text{if } j \text{ is odd} \\ \min\{\text{OPT}(j/2); \text{OPT}(j - 1)\} & \text{if } j \text{ is even} \end{cases}$
Minimum Number of Operations

Problem: Given an integer \( n \), find the minimum number of operations to get from 0 to \( n \), if you are only allowed to perform two specific operations: (1.) add 1 (2.) multiply by 2.

Algorithm 1: MinOperations(\( n \))

1. \( M \leftarrow \) length-(\( n+1 \)) array;
2. \( M[0] = 0; \)
3. for \( j = 1 \) to \( n \) do
   4. if \( j \) is odd then
      5. \( M[j] = M[j - 1] + 1; \)
   6. else if \( M[j - 1] + 1 < M[j/2] + 1 \) then
      7. \( M[j] = M[j - 1] + 1; \)
   8. else
      9. \( M[j] = M[j/2] + 1; \)
4. return \( M \)
Finding the sequence of operations—backtracking

A dynamic programming algorithm computes the optimal value.

How to find the solution itself?
We can reconstruct it from the table.

- backtrack based on the memoization table without explicitly storing values (by checking which case was chosen)

---

**Algorithm 1:** Findsolution\(j, M\)

1. if \(j == 0\) then
2. \hspace{1em} return \(\emptyset\)
3. if \(j\) is odd then
4. \hspace{1em} return Findsolution\((j - 1, M)\).append(’+1’)
5. else if \(M[j - 1] + 1 < M[j/2] + 1\) then
6. \hspace{1em} return Findsolution\((j - 1, M)\).append(’+1’)
7. else
8. \hspace{1em} return Findsolution\((j/2, M)\).append(’x2’)
Subset sum — dynamic programming

Subset Sum problem: given a set of $n$ positive integer weights $w_1, w_2, \ldots, w_n$ and a weight limit $W$. Find the subset of weights $S$ with maximum total weight that doesn’t exceed $W$. That is, find

$$S : \max_{S \subseteq \{w_1 \ldots w_n\}} \sum_{w_i \in S} w_i \leq W$$

DP: $\text{OPT}(j) =$ the max weight solution among weights $w_1, w_2, \ldots, w_j$

$\text{OPT}(j) = \max\{ \quad ? \quad \}$

Complications:

- there are no compatibility issues as with overlapping jobs (good)
- once a weight is chosen the available weight limit is decreased. Can we express this with just a single variable in $\text{OPT}$?
Subset sum — dynamic programming — 2-dimensional DP

Subset Sum problem: given a set of $n$ positive integer weights $w_1, w_2, \ldots, w_n$ and a weight limit $W$. Find the subset of weights $S$ with maximum total weight that doesn’t exceed $W$. That is, find

$$S:\ \max_{S\subseteq\{w_1\ldots w_n\}} \sum_{w_i\in S} w_i \leq W$$

$OPT(j, w) =$ the max weight solution among weights $w_1, w_2, \ldots, w_j$ with available weight limit $w$.

$$OPT(j, w) = \begin{cases} 0 & \text{if } j = 0 \\ OPT(j - 1, w) & \text{if } w_j > w \\ \max\{w_j + OPT(j - 1, w - w_j); OPT(j - 1, w)\} & \text{otherwise} \end{cases}$$
Subset sum — 2D memoization table — TopHat

\[
OPT(j, w) = \begin{cases} 
0 & \text{if } j = 0 \\
OPT(j - 1, w) & \text{if } w_j > w \\
\max\{w_j + OPT(j - 1, w - w_j); OPT(j - 1, w)\} & \text{otherwise}
\end{cases}
\]

**Input:** \(w_1, w_2, \ldots, w_n\) and \(W\) (assume weights are ints)

**Output:** \(OPT(n, W)\)

**Implementation:**

**Question:** What is the size of the memoization table and what is the running time of the resulting DP algorithm?

A. \(n^2 \& O(n^3)\) 
B. \(W^2 \& O(W^2)\) 
C. \(nW \& O(n^2W)\) 
D. \((n+1)(W+1) \& O(nW)\)
Subset sum — 2D DP

\[
OPT(j, w) = \begin{cases} 
0 & \text{if } j = 0 \\
OPT(j - 1, w) & \text{if } w_j > w \\
\max\{w_j + OPT(j - 1, w - w_j); OPT(j - 1, w)\} & \text{otherwise}
\end{cases}
\]

Input: \(w_1, w_2, \ldots, w_n\) and \(W\) (assume weights are ints)
Output: \(OPT(n, W)\)

**Algorithm 1: SubsetSum\((w_1, w_2, \ldots, w_n, W)\)**

1. \(M \leftarrow (n + 1) \times (W + 1)\) table/* 2D array/ matrix */
   /* set border cases */
2. \(M[0][\ast] = 0\)/* set row 0 to zeros */
3. \(M[\ast][0] = 0\)/* set column 0 to zeros */
4. for \(j = 1 \ldots n\) do
5.   for \(w = 1 \ldots W\) do
6.     /* apply recursive formula */
7.       \(M[j][w] = \max\{w_j + M[j - 1][w - w_j]; M[j - 1][w]\}\);
7. return \(M[n][W]\)
Subset sum — 2D DP — backtracking the solution

\[
OPT(j, w) = \begin{cases} 
0 & \text{if } j = 0 \\
OPT(j - 1, w) & \text{if } w_j > w \\
\max\{w_j + OPT(j - 1, w - w_j); OPT(j - 1, w)\} & \text{otherwise}
\end{cases}
\]

**Input:** filled memoization table \( M \)

**Output:** set of weights in the optimal solution \( S \)

**runtime?**

---

**Algorithm 1:** SubsetSumSolution\((M, w = [w_1, \ldots w_n], W)\)

```
1  S ← [ ] /* set of opt weights */
2  i ← n, j ← W;
3  while i > 0 AND j > 0 do
4    if \( M[i][j] > M[i - 1][j] \) then
5      /* the case where \( w_i \) is chosen */
6        S.append(w[i]);
7        i ← i - 1;
8        j ← j - w_i;
9    else
10      /* \( w_i \) is not chosen */
11        i ← i - 1;
12  return S
```
Dynamic programming: adding a new variable

Def. $OPT(i, w) = \text{max-profit on items 1, \ldots, } i \text{ with weight limit } w$.
Goal. $OPT(n, W)$.

Case 1. $OPT(i, w)$ does not select item $i$.
- $OPT(i, w)$ selects best of \{ 1, 2, \ldots, \} \ i \ - \ 1 \} \text{ using weight limit } w$.

Case 2. $OPT(i, w)$ selects item $i$.
- Collect value $v_i$.
- New weight limit = $w - w_i$.
- $OPT(i, w - w_i)$ selects best of \{ 1, 2, \ldots, \} \ i \ - \ 1 \} \text{ using this new weight limit}$.

$$OPT(i, w) = \begin{cases} 0 & \text{if } i = 0 \\ OPT(i - 1, w) & \text{if } w_i > w \\ \max \{ OPT(i - 1, w), v_i + OPT(i - 1, w - w_i) \} & \text{otherwise} \end{cases}$$
### Knapsack problem example

<table>
<thead>
<tr>
<th>$i$</th>
<th>$v_i$</th>
<th>$w_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>18</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>22</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>28</td>
<td>7</td>
</tr>
</tbody>
</table>

OPT($i$, $w$) = \[
\begin{cases}
0 & \text{if } i = 0 \\
OPT(i - 1, w) & \text{if } w_i > w \\
\max\{OPT(i - 1, w), v_i + OPT(i - 1, w - w_i)\} & \text{otherwise}
\end{cases}
\]

**OPT(i, w)** = max-profit subset of items 1, ..., $i$ with weight limit $w$.  

![Knapsack problem example](image-url)
Knapsack problem: running time

**Theorem.** There exists an algorithm to solve the knapsack problem with $n$ items and maximum weight $W$ in $\Theta(nW)$ time and $\Theta(nW)$ space.

**Pf.**

- Takes $O(1)$ time per table entry.
- There are $\Theta(nW)$ table entries.
- After computing optimal values, can trace back to find solution:
  take item $i$ in $OPT(i, w)$ iff $M[i, w] > M[i-1, w]$. □
Knapsack problem is NP-complete

Knapsack is in fact not polynomial in the *input size!*

**Input:**
2n integers: v\textsubscript{i} and w\textsubscript{i}
one additional integer W

**How many bits** to describe the input?
- W requires $\log W$ bits, w\textsubscript{i} requires $O(\log W)$ bits
- overall $O(n \log W)$

The algorithm would be polynomial in the input size, if the running time was a polynomial of $n$ and $\log W$

But the running time is $O(nW) = O(n 2^{\log W})$

- Decision version of knapsack problem is **NP-complete.** [Chapter 8]
- There exists a poly-time algorithm that produces a feasible solution that has value within 1% of optimum. [Section 11.8]
Here is how you would properly write out the solution to a DP problem:

1. precisely define the subproblem with proper indexing
   - OPT(i) = …….    or OPT(i,j) = …. is also possible! (or even more variables)
2. give the recursive formula to compute OPT( ) and argue about its correctness
   - make sure to define everything that needs to be, e.g. p(j) = …
   - don’t forget about border cases (sometimes you may want to add a dummy index, e.g. j=0, OPT(0) = 0)
3. write the DP algorithm.
   - bottom-up and recursive are equally good. The asymptotic running time is the same.
   - be clear about what values your memoization table holds,
     - e.g. M[i,j] = OPT(i,j), size of M is n x W
     - don’t forget initialization steps for border cases
4. write an algorithm that prints the elements (e.g. jobs) in the optimal solution
   - sometimes called “back-tracking” the solution
Bounded Knapsack problem

As input we are given the weights $w_i$ and values $v_i$ of each of $n$ items, further we are given a maximum capacity of $W$. Suppose there are two identical copies of each item available. Select a maximum value subset of the items within the capacity limit $W$, such that we can take at most two of each item.
Bounded Knapsack problem - backtracking

Backtracking the maximum choice over multiple items is tedious. Instead: keep track of our decisions on the fly:

\[ C = \text{length (n+1)x(W+1) array} \]
\[ C[i][w] = \text{how many copies of i we select for OPT(i,w)} \]

---

**Algorithm 1: BoundedKnapsack(i = 1...n : (w_i, v_i), W)**

```plaintext
/* (w_i, v_i) weight and value of item i, W capacity */
1 M ← (n + 1) \times (W + 1) array/* DP table */
2 C ← (n + 1) \times (W + 1) array/* number of copies */
3 M[0][*] ← 0 and M[*][0] ← 0;
4 for i = 1 to n do
5     for w = 1 to W do
6         c_0 ← M[i - 1][w]/* 0 of item i */
7         c_1 = v_i + M[i - 1][w - w_i] if w_i < w else c_1 ← -1/* 1 of i */
8         c_2 = 2v_i + M[i - 1][w - 2w_i] if 2w_i < w else c_2 ← -1/* 2 of i */
9         M[i][w] ← max\{c_0, c_1, c_2\};
10        C[i][w] ← argmax\{c_0, c_1, c_2\}/* index of max case */
11    return M, C
```
Bounded Knapsack problem - backtracking

Backtracking the maximum choice over multiple items is tedious.
Instead: keep track of our decisions on the fly:

\[ C = \text{length } (n+1) \times (W+1) \text{ array} \]
\[ C[i][w] = \text{how many copies of } i \text{ we select for } \text{OPT}(i,w) \]

\begin{algorithm}
\caption{BKBacktrack($C, W$)}
\begin{algorithmic}[1]
\State $sol \leftarrow \text{empty list}$;
\State $i \leftarrow n$ and $w \leftarrow W$;
\While{$i > 0$ and $w > 0$} do
\State \text{sol.add}($C[i][w] \times \text{item } i$) /* add 0, 1 or 2 of item i */
\State $i \leftarrow i - 1$;
\State $w \leftarrow w - C[i][w] \cdot w_i$;
\EndWhile
\State \text{return } sol
\end{algorithmic}
\end{algorithm}
Bounded Knapsack problem - TopHat

As input we are given the weights $w_i$ and values $v_i$ of each of $n$ items, further we are given a maximum capacity of $W$. Suppose there are two $m$ identical copies of each item available. Select a maximum value subset of the items within the capacity limit $W$, such that we can take at most two $m$ of each item.

OPT($i,w$) = maximum value within capacity $w$ if we can consider items 1,…,i

What is the recursive formula for OPT($i,w$) - excluding boundary cases?

A. $OPT(i, w) = \max_{j=0\ldots m} \{ j \cdot v_i + OPT(i-1, w-j \cdot w_i) \}$

B. $OPT(i, w) = \max \{ OPT(i-1, w); m \cdot v_i + OPT(i-1, w-m \cdot w_i) \}$

C. $OPT(i, w) = \max_{j=0\ldots m} \{ j \cdot v_i + OPT(i-1, w-w_i) \}$

D. $OPT(i, w) = \max_{j=0\ldots m} \{ j \cdot v_i + OPT(i-j, w-w_i) \}$
Coin change problem

In a far away country there are four different valued coins, the dream dollar amounts are $1, $4, $7, $13. In this country people always try to pay with the fewest number of coins possible. Design a DP algorithm to pay $n with the fewest number of coins.

bonus: the greedy algorithm - pay with the largest denomination while possible - doesn’t work. However, you’ll probably need to implement it to find the smallest $n counter example
**String matching example — plagiarism**

Plagiarism is often not a verbatim copy.

source: https://www.turnitin.com/static/plagiarism-spectrum/  “Find-Replace” method

<table>
<thead>
<tr>
<th>SOURCE TEXT</th>
<th>STUDENT WORK</th>
</tr>
</thead>
<tbody>
<tr>
<td>A Natural Setting: A History of Exploration and Settlement in Yosemite Valley</td>
<td>A Beautiful Setting in Yosemite</td>
</tr>
<tr>
<td>Since its first discovery by non-indigenous people in the mid-nineteenth century, Yosemite Valley has held a special, even religious, hold on the American conscience because its beauty makes it an incomparable valley and one of the grandest of all special temples of Nature. While Yosemite holds a special grip on the western mind, perceptions about the Valley have evolved over time due to changing politics, migration patterns and environmental concerns as man has become more attuned to his relationship and impact on nature.</td>
<td>Since its first discovery by non-native people in the mid-19th century Yosemite Valley has held a special, even sacred, hold on the American psyche because its beauty makes it an incomparable valley and one of the grandest of all special temples of Nature. While Yosemite holds a special grip on the western mindset, perceptions about the Valley have evolved over time due to changing political movements, migration patterns and environmental issues as man has become more attuned to their relationship and impact on nature.</td>
</tr>
</tbody>
</table>

Some words, e.g. “the”, “and”, or given the topic of this text “Yosemite”, naturally appear in both and are not plagiarized. We want to assign some kind of similarity score between the two texts.
Type “Define occurrence” in Google. Google will correct your spelling. How does it know which word you wanted to find? (source: google.com search result)
Q. How similar are two strings? occurrence and occurrence.

Which alignment is best depends on relative cost of gap and mismatch penalties
• The cost $\alpha$ of a mismatch and $\delta$ of a gap is part of the input

\[
\text{cost} = \alpha \cdot \text{mismatch} + \delta \cdot \text{gap}
\]
String similarity — Levenshtein Distance

Edit distance:

- first introduced by Levenshtein (1966)
- number of single character edits (insertion, deletion or substitution) required to change one string into another.
  - sometimes called Levenshtein distance

Longest Common Subsequence: special case, only allows insertions and deletions, not substitutions
String similarity — Longest Common Subsequence (Substring)

Longest Common Subsequence: given two sequences \( x = [x_1, x_2, \ldots, x_n] \) and \( y = [y_1, y_2, \ldots, y_m] \) find a longest (not consecutive) subsequence common to them both.

- special case of the similarity problem with mismatch penalty = infinite, gap penalty = 1

application: Genome similarity

- used in computational biology
- algorithm is named after Needleman and Wunsch (1970s)

\[
\begin{align*}
\text{cgtacgtacgtacgtacgtacgtatcgtacgt} & \\
\text{acgtacgtacgtacgtacgtacgtacgtacgt} & \\
\text{cgtacgtacgtacgtacgtacgtacgtacgtacgt} & \\
\text{cgtacgtacgtacgtacgtacgta} & \text{t} \quad \text{cgtacgt} \\
\text{a} & \quad \text{cgtacgtacgtacgtacgtacgtacgta} \quad \text{cgtacgt}
\end{align*}
\]
Sequence alignment

Problem. Given two strings \( X = [x_1 \ x_2 \ldots x_m] \) and \( Y = [y_1 \ y_2 \ldots y_n] \) and costs \( \alpha, \delta \) find the minimum-cost alignment \( \text{Align}(X, Y) \).

Alignment. Given two strings \( X \) and \( Y \), their alignment \( \text{Align}(X, Y) \) is a set of ordered pairs (a matching) \( (x_i, y_j) \), such that

- each character is matched at most once
- there are no two pairs \( (x_i, y_k) \) and \( (x_j, y_l) \), such that \( x_i \) comes before \( x_j \) but \( y_k \) after \( y_l \) or vice verse, i.e. there are no crossing pairs

invalid alignment:
Sequence alignment

Problem. Given two strings $X = [x_1, x_2, \ldots, x_m]$ and $Y = [y_1, y_2, \ldots, y_n]$ and costs $\alpha, \delta$ find the minimum-cost alignment $\text{Align}(X,Y)$.

Alignment. Given two strings $X$ and $Y$, their alignment $\text{Align}(X,Y)$ is a set of ordered pairs (a matching) $(x_i, y_j)$, such that

- each character is matched at most once
- there are no two pairs $(x_i, y_k)$ and $(x_j, y_l)$, such that $x_i$ comes before $x_j$ but $y_k$ after $y_l$ or vice verse, i.e. there are no crossing pairs

valid alignment:

\[
\begin{align*}
\text{x}_1 & \quad \text{x}_2 & \quad \text{x}_3 & \quad \text{x}_4 & \quad \text{x}_5 & \quad \text{x}_6 & \quad \text{x}_7 & \quad \text{x}_8 & \quad \ldots \ldots \ldots \ldots \quad \text{x}_{m-2} & \quad \text{x}_{m-1} & \quad \text{x}_m \\
\text{y}_1 & \quad \text{y}_2 & \quad \text{y}_3 & \quad \text{y}_4 & \quad \text{y}_5 & \quad \text{y}_6 & \quad \text{y}_7 & \quad \text{y}_8 & \quad \ldots \ldots \ldots \ldots & \quad \text{y}_{n-2} & \quad \text{y}_{n-1} & \quad \text{y}_n
\end{align*}
\]
Sequence alignment

Problem. Given two strings $X = [x_1 \ x_2 \ldots \ x_m]$ and $Y = [y_1 \ y_2 \ldots \ y_n]$ and costs $\alpha, \delta$ find the minimum-cost alignment $\text{Align}(X, Y)$.

Alignment. Given two strings $X$ and $Y$, their alignment $\text{Align}(X, Y)$ is a set of ordered pairs (a matching) $(x_i, y_j)$, such that
- each character is matched at most once
- there are no two pairs $(x_i, y_k)$ and $(x_j, y_l)$, such that $x_i$ comes before $x_j$ but $y_k$ after $y_l$ or vice versa, i.e. there are no crossing pairs

Cost of an alignment:

$$\text{cost}(\text{Align}(X, Y)) = \alpha \cdot \#(\text{mismatch}) + \delta \cdot \#(\text{gap})$$

$\alpha, \delta$ are given as input.

# of unmatched characters in $X$ + # of unmatched chars in $Y$
Question: cost of alignment?

mismatch: \( \alpha = 3 \)

gap: \( \delta = 2 \)
Brute-force approach — TopHat

**Algorithm:** Try *all* possible valid alignments and return the min-cost valid alignment:

- each character $x_i$ is matched to at most one character $y_j$
- there are no crossing pairs

**Question.** Guess how many valid alignments there are if $X$ is a string of $m$, and $Y$ a string of $n$ characters. (we may assume $m \leq n$)

A. $\sum_{k=0}^{m} \binom{m}{k} \binom{n}{k} = \binom{m+n}{m}$

B. $m! n! = O(m^n n^m)$

C. $m \cdot n$

D. $m^n$
Brute-force approach — number of valid alignments

Some facts:
- each character $x_i$ in $X$ is aligned to either a character $y_j$ or a gap.
- if $k$ characters in $X$ are matched to characters in $Y$, then the number of matched characters in $Y$ is also $k$.

compute number of valid assignments:

count how many ways there are to pick $k$ among $X$ and $k$ among $Y$. Those are the characters matched to each other.
- note that the order in which these $k$ are matched is fixed, and hence unambiguous

\[ \sum_{k=0}^{m} \binom{m}{k} \binom{n}{k} = \binom{m+n}{m} \]

if $m < n$ we get:

\[ \geq \binom{2m}{m} = \Theta \left( \frac{4^m}{\sqrt{m}} \right) \]

choose the $k$ characters in $X$ that are assigned to characters in $Y$ (as opposed to gaps)
choose the $k$ characters in $Y$ that are assigned to characters in $X$
tricks with binomial coefficients

Good to know:

\[(a + b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k\]

\[2^n = (1 + 1)^n = \sum_{k=0}^{n} \binom{n}{k} 1^{n-k} 1^k = \sum_{k=0}^{n} \binom{n}{k}\]

\[\binom{n}{k} = \binom{n}{n-k}\]

\[\sum_{k=0}^{n} \binom{n}{k} \binom{m}{m-k} = \binom{n+m}{m}\]

Computation on previous slide:

\[\sum_{k=0}^{n} \binom{n}{k} \binom{m}{k} = \sum_{k=0}^{n} \binom{n}{k} \binom{m}{m-k} = \binom{n+m}{m} = \binom{n+m}{n}\]