Advanced Algorithmics (6EAP)
Dynamic Programming

Jaak Vilo
2016 Fall

Example

• Wind has blown away the +, *, (, ) signs
• What’s the maximal value?
• Minimal?

2 1 7 1 4 3

• Q: How to maximize the value of any expression?

2 4 5 1 9 8 12 1 9 8 7 2 4 1 1 2 3 = ?

• 2 1 7 1 4 3
• (2+1)*7*(1+4)*3 = 21*15 = 315
• 2*1 + 7 + 1*4 + 3 = 16

• Dynamic programming, like the divide-and-conquer method, solves problems by combining the solutions to subproblems.
  – Dünaamiline planeerimine.
  – Divide-and-conquer algorithms partition the problem into independent subproblems, solve the subproblems recursively, and then combine their solutions to solve the original problem.
• In contrast, dynamic programming is applicable when the subproblems are not independent, that is, when subproblems share subsubproblems.
Dynamic programming

- Avoid calculating repeating subproblems
  - \( \text{fib}(1) = \text{fib}(0) = 1; \)
  - \( \text{fib}(n) = \text{fib}(n-1) + \text{fib}(n-2) \)
- Although natural to encode (and a useful task for novice programmers to learn about recursion) recursively, this is inefficient.

Structure within the problem

- The fact that it is not a tree indicates overlapping subproblems.

Topp-down (recursive, memoized)

- A dynamic-programming algorithm solves every subsubproblem just once and then saves its answer in a table, thereby avoiding the work of recomputing the answer every time the subsubproblem is encountered.

- **Top-down approach:** This is the direct fall-out of the recursive formulation of any problem. If the solution to any problem can be formulated recursively using the solution to its subproblems, and if its subproblems are overlapping, then one can easily memoize or store the solutions to the subproblems in a table. Whenever we attempt to solve a new subproblem, we first check the table to see if it is already solved. If a solution has been recorded, we can use it directly, otherwise we solve the subproblem and add its solution to the table.

Bottom-up

- **Bottom-up approach:** This is the more interesting case. Once we formulate the solution to a problem recursively as in terms of its subproblems, we can try reformulating the problem in a bottom-up fashion: try solving the subproblems first and use their solutions to build-on and arrive at solutions to bigger subproblems. This is also usually done in a tabular form by iteratively generating solutions to bigger and bigger subproblems by using the solutions to small subproblems. For example, if we already know the values of \( F_{41} \) and \( F_{40} \), we can directly calculate the value of \( F_{42} \).

- Dynamic programming is typically applied to optimization problems. In such problems there can be many possible solutions. Each solution has a value, and we wish to find a solution with the optimal (minimum or maximum) value.
- We call such a solution an optimal solution to the problem, as opposed to the optimal solution, since there may be several solutions that achieve the optimal value.
The development of a dynamic-programming algorithm can be broken into a sequence of four steps.

1. Characterize the structure of an optimal solution.
2. Recursively define the value of an optimal solution.
3. Compute the value of an optimal solution in a bottom-up fashion.
4. Construct an optimal solution from computed information.

Edit distance (Levenshtein distance)

- Smallest nr of edit operations to convert one string into the other

**Edit (Levenshtein) distance**

- **Definition** The edit distance $D(A,B)$ between strings $A$ and $B$ is the minimal number of edit operations to change $A$ into $B$. Allowed edit operations are deletion of a single letter, insertion of a letter, or replacing one letter with another.
- Let $A = a_1 a_2 ... a_n$ and $B = b_1 b_2 ... b_m$.
  - $E_1$: Deletion $a_i \rightarrow \varepsilon$.
  - $E_2$: Insertion $\varepsilon \rightarrow b_i$.
  - $E_3$: Substitution $a_i \rightarrow b_j$ (if $a_i \neq b_j$)
- Other possible variants:
  - $E_4$: Transposition $a_i a_{i+1} \rightarrow b_j b_{j+1}$ and $a_i = b_{j+1}, a_{i+1} = b_j$ (e.g. lecture $\rightarrow$ lecture)

**How can we calculate this?**

\[
D(a, b) = \begin{cases} 
D(a, b) & \text{if } a = b \\
D(a, b) + 1 & \text{if } a \neq b \\
D(a[i-1], b[j-1]) + (a[i] = b[j])? 0 : 1 \\
\min & \\
\end{cases}
\]

Other possible variants:

\[
d(i,j) = \begin{cases} 
\min & \\
\end{cases}
\]

**How can we calculate this efficiently?**

\[
D(S,T) = \min \begin{cases} 
1. D(S[1..n-1], T[1..m-1]) + (S[n]=T[m]?) 0 : 1 \\
2. D(S[1..n], T[1..m-1]) +1 \\
3. D(S[1..n-1], T[1..m]) +1 \\
\end{cases}
\]

Define:\n
\[
d(i,j) = D( S[1..i], T[1..j] )
\]

\[
d(i,j) = \min \begin{cases} 
1. d(i-1,j-1) + (S[i]=T[j]?) 0 : 1 \\
2. d(i-1,j) +1 \\
3. d(i,j-1) +1 \\
\end{cases}
\]

**Recursion?**

\[
d(i,j) = \min \begin{cases} 
1. d(i-1,j-1) + (S[i]=T[j]?) 0 : 1 \\
2. d(i-1,j) +1 \\
3. d(i,j-1) +1 \\
\end{cases}
\]
Recursion?

Algorithm Edit distance \(D(A,B)\) using Dynamic Programming (DP)

Input: \(A=a_1, a_2, ..., a_n\), \(B=b_1, b_2, ..., b_m\).
Output: Value \(d_{mn}\), in matrix \((d_{ij})\), \(0 \leq i \leq n, 0 \leq j \leq m\).

for \(i=0\) to \(m\) do \(d_{i0}=i\); 
for \(j=0\) to \(n\) do \(d_{0j}=j\); 
for \(j=1\) to \(n\) do 
  for \(i=1\) to \(m\) do 
    \(d_{ij} = \min(d_{i-1,j-1} + (\text{if } a_i = b_j \text{ then } 0 \text{ else } 1), d_{i-1,j} + 1, d_{i,j-1} + 1)\) 
return \(d_{mn}\)

Dynamic Programming

Matrix multiplication

- for \(i=1..n\) 
  for \(j=1..k\) 
    \[ c_{ij} = \sum_{x=1}^{m} a_{ix} b_{xj} \]

\[O(nmk)\]
MATRIX-MULTIPLY(A,B)
1 if columns [A] ≠ rows [B]
2 then error "incompatible dimensions"
3 else for i = 1 to rows [A]
4 do for j = 1 to columns [B]
5 do C[i,j] = 0
6 for k = 1 to columns [A]
8 return C

The **matrix-chain multiplication problem** can be stated as follows: given a chain <A₁, A₂, ..., Aᵣ> of n matrices

- matrix Aᵢ has dimension pᵢ₋₁ × pᵢ
- fully parenthesize the product A₁A₂...Aᵣ in a way that minimizes the number of scalar multiplications.

---

**Chain matrix multiplication**

![Diagram of matrix multiplication](image)

Figure 6.6: A = C = D = (A × (B × C)) × D.

<table>
<thead>
<tr>
<th>Parenthesization</th>
<th>Cost computation</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>A × ((B × C)) × D</td>
<td>20 · 1 · 10 + 20 · 10 + 100 · 100</td>
<td>120,200</td>
</tr>
<tr>
<td>(A × B) × (C × D)</td>
<td>20 · 1 · 10 + 50 · 10 + 100 · 100</td>
<td>7,000</td>
</tr>
</tbody>
</table>

As you can see, the order of multiplications makes a big difference in the final running time! Moreover, the naive greedy approach, to always perform the cheapest matrix multiplication available, leads to the second parenthesization shown here and is therefore a failure.
Let’s crack the problem

\[ A_{i,j} = A_i \cdot A_{i+1} \cdot \ldots \cdot A_j \]

- Optimal parenthesization of \( A_1 \cdot A_2 \cdot \ldots \cdot A_n \) splits at some \( k, k+1 \).
- Optimal = \( A_{1:k} \cdot A_{k+1:n} \)
- \( T(A_{1:n}) = T(A_{1:k}) + T(A_{k+1:n}) + T(A_{1:k} \cdot A_{k+1:n}) \)
- \( T(A_{1,k}) \) must be optimal for \( A_1 \cdot A_2 \cdot \ldots \cdot A_k \)

Let’s define the following:

\[ m(i,j) = \begin{cases} 0 & \text{if } i = j \\ \min_{k \in [i,j]} \left( m(i,k) + m(k+1,j) + \sum_{p=i}^{j} p \cdot p_k \cdot p_j \right) & \text{if } i < j \end{cases} \]  
(16.2)

- Since the optimal parenthesization must use one of these values for \( k \), we need only check them all to find the best. Thus, our recursive definition for the minimum cost of parenthesizing the product \( A_1 \cdot A_2 \cdot \ldots \cdot A_n \) becomes

\[ m(i,j) = \begin{cases} \min_{k \in [i,j]} \left( m(i,k) + m(k+1,j) + \sum_{p=i}^{j} p \cdot p_k \cdot p_j \right) & \text{if } i < j \end{cases} \]

- To help us keep track of how to construct an optimal solution, let us define \( s(i,j) \) to be a value of \( k \) at which we can split the product \( A_i \cdot A_{i+1} \cdot \ldots \cdot A_j \), to obtain an optimal parenthesization. That is, \( s(i,j) \) equals a value \( k \) such that \( m(i,j) = m(i,k) + m(k+1,j) + \sum_{p=i}^{j} p \cdot p_k \cdot p_j \).

Recursion

- Checks all possibilities...

- But – there is only a few subproblems – choose \( i, j \) s.t. \( 1 \leq i \leq j \leq n \) - \( O(n^2) \)

A recursive algorithm will encounter each subproblem many times in different branches of its recursion tree. This property of overlapping subproblems is the second hallmark of the applicability of dynamic programming.
A simple inspection of the nested loop structure of MATRIX-CHAIN-ORDER yields a running time of $O(n^3)$ for the algorithm. The loops are nested three deep, and each loop index ($l$, $i$, and $k$) takes on at most $n$ values.

- Time $\Omega(n^3) \Rightarrow O(n^3)$
- Space $O(n^2)$

Step 4 of the dynamic-programming paradigm is to construct an optimal solution from computed information.

Use the table $s[1 \ldots n, 1 \ldots n]$ to determine the best way to multiply the matrices.

Multiply using S table

```
MATRIX-CHAIN-MULTIPLY(A, s, i, j)
1 if j > i
2 then X = MATRIX-CHAIN-MULTIPLY(A, s, i, s[i, j])
3 Y = MATRIX-CHAIN-MULTIPLY(A, s, s[i, j]+1, j)
4 return MATRIX-MULTIPLY(X, Y)
5 else return A_i
((A_iA_j)(A_jA_k))
```

Elements of dynamic programming

- Optimal substructure within an optimal solution
- Overlapping subproblems
- Memoization
• A memoized recursive algorithm maintains an entry in a table for the solution to each subproblem. Each table entry initially contains a special value to indicate that the entry has yet to be filled in. When the subproblem is first encountered during the execution of the recursive algorithm, its solution is computed and then stored in the table. Each subsequent time that the subproblem is encountered, the value stored in the table is simply looked up and returned. (tabulated)

This approach presupposes that the set of all possible subproblem parameters is known and that the relation between table positions and subproblems is established. Another approach is to memoize by using hashing with the subproblem parameters as keys.

Longest Common Subsequence (LCS)

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Overlapping subproblems

Optimal triangulation

The problem is to find a triangulation that minimizes the sum of the weights of the triangles in the triangulation.

Two ways of triangulating a convex polygon. Every triangulation of this 7-sided polygon has 7 - 3 = 4 chords and divides the polygon into 7 - 2 = 5 triangles.

Optimal triangulation

\[
d(i,j) = \begin{cases} 
0 & \text{if } j = i \\
\min_{k \in S, k \neq i} \{d(i,k) + d(k+1,j) + w(v_i, v_k, v_j)\} & \text{if } i < j 
\end{cases} \tag{10.7}
\]

Parse tree

Parse trees. (a) The parse tree for the parenthesized product 
\((\text{A} \text{(A}\text{A})(\text{A}\text{A}\text{A}))\) and for the triangulation of the 7-sided polygon
(b) The triangulation of the polygon with the parse tree overlaid. Each matrix \(A_i\) corresponds to the side \(v_i \text{v}_{i+1}\) for \(i = 1, 2, \ldots, 6\).
Dynamic programming

- Avoid re-calculating same subproblems by
  - Characterising optimal solution
  - Clever ordering of calculations

**Edit distance is a metric**

- It can be shown, that $D(A,B)$ is a metric
  - $D(A,B) \geq 0$, $D(A,B)=0$ iff $A=B$
  - $D(A,B) = D(B,A)$
  - $D(A,C) \leq D(A,B) + D(B,C)$

**Path of edit operations**

- Optimal solution can be calculated afterwards
  - Quite typical in dynamic programming
  
  $d[i-1,j-1] \quad d[i-1,j] \quad d[i,j-1] \quad d[i,j]$

- Memorize sets $\text{pred}[i,j]$ depending from where the $d_j$ was reached.
Three possible minimizing paths

- Add into pred[i,j]
  - $(i-1,j-1)$ if $d_{ij} = d_{i-1,j-1} + (if \ a_i = b_j \ then \ 0 \ else \ 1)$
  - $(i-1,j)$ if $d_{ij} = d_{i-1,j} + 1$
  - $(i,j-1)$ if $d_{ij} = d_{ij-1} + 1$

Multiple paths possible

- All paths are correct
- There can be many (how many?) paths

Calculation of $D(A,B)$ in space $\Theta(m)$

Input: $A=a_1,a_2,...,a_n$, $B=b_1,b_2,...,b_n$ (choose $m \leq n$)
Output: $d_{mn}=D(A,B)$
for $i=0$ to $m$ do $C[i]=i$
for $j=1$ to $n$ do $C[0]=j$
for $i=1$ to $m$ do
  $d = \min (C + (if \ a_i = b_j \ then \ 0 \ else \ 1), C[i-1] + 1, C[j] + 1)$
  $C = C[i]$ // memorize new "diagonal" value
  $C[i] = d$
write $C[m]$

Time complexity is $\Theta(mn)$ since $C[0..m]$ is filled $n$ times
Shortest path in the graph


All nodes at distance 1 from source

Observations?

• Shortest path is close to the diagonal
  – If a short distance path exists

• Values along any diagonal can only increase
  (by at most 1)

Diagonal

Property of any diagonal: The values of matrix \( (d_{ij}) \) can on any specific diagonal either increase by 1 or stay the same

Diagonal lemma

Lemma: For each \( d_{ij} \), \( i, j \in \{0, \ldots, m\} \) holds:
\[
d_{ij} = d_{i-1, j-1} + 1.
\]
(notice that \( d_{ij} \) and \( d_{i-1,j-1} \) are on the same diagonal)

Proof: Since \( d_{ij} \) is an integer, show:

1. \( d_{ij} \leq d_{i-1, j-1} + 1 \)
2. \( d_{ij} \geq d_{i-1, j-1} + 1 \)

From the definition of edit distance 1. holds since \( d_{ij} \leq d_{i-1, j-1} + 1 \)

Induction on \( i+j \):

– Basis is trivial when \( i=0 \) or \( j=0 \)

– Induction step: there are 3 possibilities:
  • On minimization the \( d_{ij} \) is calculated from entry \( d_{i-1,j-1} \), hence \( d_{ij} = d_{i-1,j-1} + 1 \)
  • On minimization the \( d_{ij} \) is calculated from entry \( d_{i-1,j} \), hence \( d_{ij} = d_{i-1,j} + 1 \)
  • On minimization the \( d_{ij} \) is calculated from entry \( d_{i,j-1} \), Analagous to 2.

  Hence, \( d_{ij} \leq d_{ij} \)
Transform the matrix into $f_{kp}$

- For each diagonal only show the position (row index) where the value is increased by 1.
- Also, one can restrict the matrix ($d_{ij}$) to only this part where $d_i \leq d_{mn}$ since only those $d_i$ can be on the shortest path.
- We’ll use the matrix ($f_{kp}$) that represents the diagonals of $d_{ij}$
  - $f_{kp}$ is a row index $i$ from $d_{ij}$, such that on diagonal $k$ the value $p$ reaches row $i$ ($d_{ip}$ and $j=i+k$).
  - Initialization: $f_{k-1} = 1$ and $f_{k,n} = \infty$ when $p \leq |k|-1$;
  - $d_{mm} = p$, such that $f_{m,m} = m$

Calculating matrix ($f_{kp}$) by columns

- Assume the column $p-1$ has been calculated in ($f_{kp}$), and we want to calculate $f_{kp}$. (the region of $d_{ij}=p$)
- On diagonal $k$ values $p$ reach at least the row $t = \max(f_{k,p-1}+1, f_{k-1,p-1}, f_{k+1,p-1}+1)$ if the diagonal $k$ reaches so far.
- If on row $t+1$ additionally $a_i = b_j$ on the same diagonal, then $d_{ij}$ cannot increase, and value $p$ reaches row $t+1$.
- Repeat previous step until $a_i \neq b_j$ on diagonal $k$.

Algorithm $A()$: calculate $f_{kp}$

$A(k,p)$
1. $t = \max(f_{k,p-1}+1, f_{k-1,p-1}, f_{k+1,p-1}+1)$
2. while $a_{t+1} = b_{t+k}$ do $t = t+1$
3. $f_{kp} = \text{if } t>m \text{ or } t+k > n \text{ then undefined else } t$
Algorithm: Diagonal method by columns

\[ p = -1 \]
\[ \text{while } f_{n,m,p} \neq m \]
\[ p = p + 1 \]
\[ \text{for } k = -p \text{ to } p \text{ do } \quad // \quad f_{kp} = A(k,p) \]
\[ t = \max(f_{k,p-1} + 1, f_{k-1,p}, f_{k+1,p}) \]
\[ \text{while } a_{i+k} = b_{i+1} \text{ do } \quad t = t + 1 \]
\[ f_{kp} = \begin{cases} \text{undefined} & \text{if } t > m \text{ or } t + k > n \text{ then } \text{undefined} \\ t \cdot p \text{ can only occur on diagonals } -p \leq k \leq p . \end{cases} \]
\[ \text{Method can be improved since } k \text{ is often such that } f_{kp} \text{ is undefined.} \]
\[ \text{We can decrease values of } k: \]
\[ -m \leq k \leq n \text{ (diagonal numbers)} \]
\[ \text{Let } m \leq n \text{ and } d_i \text{ on diagonal } k. \]
\[ \text{if } -m \leq k \leq 0 \text{ then } |k| d_i \leq m \]
\[ \text{if } 1 \leq k \leq n \text{ then } k d_i \leq k + m \]
\[ \text{Hence, } -m \leq k \leq m \text{ if } p \leq m \text{ and } p - m \leq p \text{ if } p \geq m \]

Extensions to basic edit distance

- New operations
- Variable costs
- Time Warping

Transposition (ab \rightarrow ba)

- \textbf{E4: Transposition} \quad a_i a_{i+1} \rightarrow b_j b_{j+1}, \text{ s.t. } a_i = b_{j+1} \text{ and } a_{i+1} = b_j
- (e.g.: lecture \rightarrow letcure)

\[ d(i,j) = \min \begin{cases} 1. & d(i-1,j-1) + (S[i]=T[j] \text{? } 0 \text{ : } 1) \\ 2. & d(i,j-1) + 1 \\ 3. & d(i-1,j) + 1 \\ 4. & d(i-2,j-2) + (\text{if } S[i-1]=T[j-1] \text{ then } 1 \text{ else } \infty) \end{cases} \]

Generalized edit distance

- Use more operations \text{E1...En}, and to provide different costs to each.
- \textbf{Definition.} Let \( x, y \in \Sigma^* \). Then every \( x \rightarrow y \) is an edit operation. Edit operation replaces \( x \) by \( y \).
- If \( A = uv \) then after the operation, \( A = uyv \)
- We note by \( w(x \rightarrow y) \) the cost or weight of the operation.
- Cost may depend on \( x \) and/or \( y \). But we assume \( w(x \rightarrow y) \geq 0 \).
Generalized edit distance

- If operations can only be applied in parallel, i.e. the part already changed cannot be modified again, then we can use the dynamic programming.
- Otherwise it is an algorithmically unsolvable problem, since question - can A be transformed into B using operations of G, is unsolvable.
- The diagonal method in general may not be applicable.
- But, since each diversion from diagonal, the cost slightly increases, one can stay within the narrow region around the diagonal.

Applications of generalized edit distance

- Historic documents, names
- Human language and dialects
- Transliteration rules from one alphabet to another
e.g. Tõugu => Tyugu (via Russian)
- ...

Examples

näituseks – näiteks
Ahwrika - Aafrika
weikese - väikese
materjaali - materjali

“kavalam” otsimine
Dush, dušš, dushsh?
Gorbatšov, Gorbatschov, Горбачов,
Gorbachev
režim, rezhim, riim
Links

- Est-Eng, Old Estonian, Est-Rus transliteration

- Pronunciation
  - [https://bit-devcs.ut.ee/~orasmaa/ing_ligikaudne/](https://bit-devcs.ut.ee/~orasmaa/ing_ligikaudne/)

- Github (Reina Uiba; Siim Orasmaa)
  - [https://github.com/soras/genEditDist](https://github.com/soras/genEditDist)


How?

- Apply Aho-Corasick to match for all possible edit operations
- Use minimum over all possible such operations and costs
- Implementation: Reina Käärik, Siim Orasmaa

Possible problems/tasks

- Manually create sensible lists of operations
  - For English, Russian, etc...
  - Old language,

- Improve the speed of the algorithm (testing)

- Train for automatic extraction of edit operations and respective costs from examples of matching words...

Advanced Dynamic Programming

- Robert Giegerich:
  - [http://www.techfak.uni-bielefeld.de/ags/pi/lehre/ADP/](http://www.techfak.uni-bielefeld.de/ags/pi/lehre/ADP/)

- Algebraic dynamic programming
  - Functional style
  - Haskell compiles into C