Example: The Paragraphing Problem

Problem: Given words $w_1 \ldots w_m$ and a line length $L$, break the sequence into lines subject to a penalty function $\alpha(i,j) \geq 0$ that measures how "bad" words $w_i \ldots w_j$ look on a line that is both right and left justified (or in the case of the last line, when $j=m$, only left justified). Minimize the sum of the penalty function over all lines of the paragraph.

Idea:

- Compute from right to left $cost[i]$, the cost of paragraphing $w_i \ldots w_m$.
- Only the values of $cost[j]$, $j > i$, are needed to compute $cost[i]$.
- To compute $cost[i]$, we try all ways of breaking off the first line of $w_i \ldots w_m$ and adding the cost of that line to the cost of paragraphing the remaining words (already computed).
- An array $break$ can store the best break points found.
(paragraphing problem continued)

**Computing the cost and break arrays:**

Input and store the words $w_1 \ldots w_m$ in a list of strings

$\text{cost}[m+1] := 0$

for $i := m$ downto 1 do begin

$\text{cost}[i] := \infty$

$\text{break}[i] := m$

$j := i$

while $j \leq m$ and $w_i \ldots w_j$ can fit on a line do begin

if $(\alpha(i,j) + \text{cost}[j+1]) < \text{cost}[i]$ then begin

$\text{cost}[i] := \alpha(i,j) + \text{cost}[j+1]$

$\text{break}[i] := j$

end

$j := j + 1$

end

end
Formatting the paragraph:

\[ i := 1 \]

\textbf{while} \( i \leq m \) \textbf{do begin}

Output a line consisting of \( w_i \) through \( w_{break[i]} \)

\( i := break[i] + 1 \)

\textbf{end}
(paragraphing problem continued)

**Complexity:**

- Let $n$ be the sum of the lengths of all input words.
- Reading the input is $O(n)$.
- Under the assumption that $\alpha(i,j)$ can be computed in $O(1)$ time, computing the cost and break arrays is at worst $O(mL) \leq O(nL)$, which is $O(n)$ assuming that $L$ is constant with respect to $n$.

  ** In fact, it is not necessary for this analysis that $\alpha(i,j)$ can be computed in $O(1)$ time, only that it can be computed in $O(1)$ time from values already computed.

- Formatting the paragraph takes an additional $O(n)$ time.
- In addition to the space to store the input, $O(m)$ space is used to store the cost and break arrays.