# Dynamic Programming

## Sequence Alignment: Optimal Substructure

Algorithms: Design and Analysis, Part II

# Problem Definition

**Recall:** sequence alignment. [Needleman–Wunsch score = similarity measure between strings]

**Example:**

AGGGCT
AGG–CA

total penalty = $\alpha_{gap} + \alpha_{AT}$

**Input:** strings $X = x_1 \cdots x_m$, $Y = y_1 \cdots y_n$ over some alphabet $\Sigma$ (like $\{A, C, G, T\}$)
– penalty $\alpha_{gap} \geq 0$ for inserting a gap,
$\alpha_{ab}$ for matching a & b [presumably $\alpha_{ab} = 0$ if $a = b$]

**Feasible Solutions:** alignments – i.e., insert gaps to equalize lengths of the strings

**Goal:** alignment with the minimum-possible total penalty.

# A Dynamic Programming Approach

Key step: identify subproblems. As usual, will look at structure of an optimal solution for clues.

[i.e., develop a recurrence + then reverse engineer the subproblems]

Structure of optimal solution: Consider an optimal alignment of X, Y and its final position:

X + gaps

Y + gaps

final position

Question: How many relevant possibilities are there for the contents of the final position?

(A) 2

(B) 3

(C) 4

(D) m·n

Case 1: $x_m, y_n$ matched

Case 2: $x_m$ matched with a gap

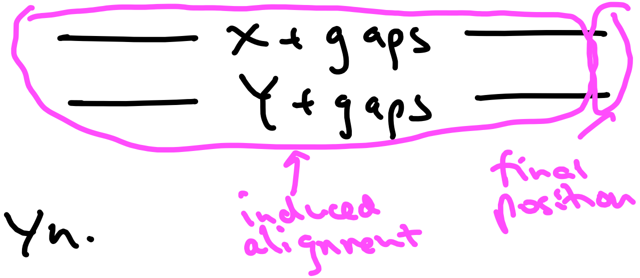
Case 3: $y_n$ matched with a gap

[pointless to have 2 gaps]

# Optimal Substructure

① $x_m$ & $y_n$
② $x_m$ & gap
③ $y_n$ & gap

**Point:** narrow optimal solution down to 3 candidates.




**Optimal substructure:** Let $X' = X - x_m$, $Y' = Y - y_n$.

If case ① holds, then induced alignment of $X'$ & $Y'$ is optimal.

If case ② holds, then induced alignment of $X'$ & $Y$ is optimal.

If case ③ holds, then induced alignment of $X$ & $Y'$ is optimal.

# Optimal Substructure (Proof)

Proof: [of Case 1, other cases are similar]

By contradiction. Suppose induced alignment of $X', Y'$ has penalty $P$ while some other one has penalty $P^* < P$.

$\Rightarrow$ appending $\begin{matrix} x_m \\ | \\ y_n \end{matrix}$ to the latter, get an alignment of $X$ and $Y$ with penalty $P^* + \alpha_{x_m y_n} < \underbrace{P + \alpha_{x_m y_n}}$

contents of final position

$\hookrightarrow$ penalty of original alignment

$\Rightarrow$ contradicts optimality of original alignment of $X$ & $Y$

QED!