Flere eksempler

Eksempel 1: Længste fælles delsekvens

```
\label{eq:Alfabet} \begin{split} & \mathsf{Alfabet} = \mathsf{mængde} \ \mathsf{af} \ \mathsf{tegn:} \\ & \{\mathsf{a},\mathsf{b},\mathsf{c},\dots,\mathsf{z}\}, \quad \  \{\mathsf{A},\mathsf{C},\mathsf{G},\mathsf{T}\}, \quad \  \{\mathsf{0},\mathsf{1}\} \end{split}
```

Eksempel 1: Længste fælles delsekvens

```
Alfabet = mængde af tegn:  \{a,b,c,\dots,z\}, \qquad \{A,C,G,T\}, \qquad \{0,1\}  Streng = sekvens x_1x_2x_3\dots x_n af tegn fra et alfabet: helloworld  \text{GATAAATCTGGTCTTATTTCC}  0010110010101010101111
```

Eksempel 1: Længste fælles delsekvens

Alfabet = mængde af tegn:

$$\big\{a,b,c,\ldots,z\big\},\qquad \big\{A,C,G,T\big\},\qquad \big\{0,1\big\}$$

Streng = sekvens $x_1x_2x_3...x_n$ af tegn fra et alfabet:

helloworld

GATAAATCTGGTCTTATTTCC

00101100101010001111

Delsekvens = delmængde af tegnene i streng, i uændret rækkefølge:



Længste fælles delsekvens

Fælles delsekvens for to strenge:



Længste fælles delsekvens

Fælles delsekvens for to strenge:



Længste fælles delsekvens (Longest Common Subsequence, LCS):

For to strenge

$$X = x_1 x_2 x_3 \dots x_m$$

$$Y = y_1 y_2 y_3 \dots y_n$$

af længde m og n, find en længste fælles delsekvens for dem.

Længden af denne kan ses som et mål for similaritet mellem strenge (f.eks. dna-strenge).

Rekursiv løsning?

Vi vil arbejde på at lave en rekursiv løsning. Vi definerer derfor en størrelse for del-problemer:

Rekursiv løsning?

Vi vil arbejde på at lave en rekursiv løsning. Vi definerer derfor en størrelse for del-problemer:

- $X_i = x_1 x_2 x_3 \dots x_i \text{ for } 1 \le i \le m.$
- ▶ $Y_j = y_1 y_2 y_3 \dots y_j$ for $1 \le j \le n$.
- \triangleright X_0 og Y_0 er den tomme streng.
- ightharpoonup lcs(i,j) er længden af længste fælles delsekvens af X_i og X_j .

Rekursiv løsning?

Vi vil arbejde på at lave en rekursiv løsning. Vi definerer derfor en størrelse for del-problemer:

- $X_i = x_1 x_2 x_3 \dots x_i \text{ for } 1 < i < m.$
- ► $Y_j = y_1 y_2 y_3 \dots y_j$ for $1 \le j \le n$.
- \triangleright X_0 og Y_0 er den tomme streng.
- ightharpoonup lcs(i,j) er længden af længste fælles delsekvens af X_i og X_j .

Vi vil gerne finde lcs(m, n).

Mere generelt: Vi søger en rekursiv formel for lcs(i, j).

Basistilfælde: Det er klart at lcs(0,j) = lcs(i,0) = 0.

Optimale delproblemer I

Formel for lcs(i, j):

Case I: $x_i = y_j$

Observation: en fælles delsekvens Z for X_i og Y_j af længde k består af

- ightharpoonup Et sidste tegn z_k .
- ▶ En streng $Z' = z_1 z_2 z_3 \dots z_{k-1}$ af længde k-1, som må være en fælles delsekvens af X_{i-1} og Y_{j-1} (tegnene i Z skal komme i samme rækkefølge som i X og Y, så kun sidste tegn i Z har mulighed for at være x_i og y_j).

Optimale delproblemer I

Formel for
$$lcs(i, j)$$
:

Case I:
$$x_i = y_j$$

Observation: en fælles delsekvens Z for X_i og Y_j af længde k består af

- ightharpoonup Et sidste tegn z_k .
- ▶ En streng $Z' = z_1 z_2 z_3 \dots z_{k-1}$ af længde k-1, som må være en fælles delsekvens af X_{i-1} og Y_{j-1} (tegnene i Z skal komme i samme rækkefølge som i X og Y, så kun sidste tegn i Z har mulighed for at være x_i og y_j).

Observation (optimale delproblemer) for Case I:

Hvis Z er en længste fælles delsekvens for for X_i og Y_j , må Z' være en længste fælles delsekvens af X_{i-1} og Y_{j-1} . For hvis der fandtes en længere fælles delsekvens for X_{i-1} og Y_{j-1} , kunne den tilføjes tegnet x_i (= y_j) og blive en længere fælles delsekvens for X_i og Y_j .

$$x_1 x_2 x_3 x_4 \dots x_{i-1}$$
 $y_1 y_2 y_3 y_4 \dots y_{j-1}$

Optimale delproblemer I

Af observationen haves i Case I $(x_i = y_i)$:

- ightharpoonup lcs(i, j) = lcs(i 1, j 1) + 1
- ▶ En længste fælles delsekvens for X_{i-1} og Y_{j-1} tilføjet tegnet x_i (= y_j) er en længste fælles delsekvens for X_i og Y_j .



Optimale delproblemer II

Formel for lcs(i, j):

Case II:
$$x_i \neq y_j$$

Observation: en fælles delsekvens $Z = z_1 z_2 z_3 \dots z_k$ for X_i og Y_j kan ikke have z_k værende en parring af x_i og y_j (da disse er forskellige).

Så Z må være en fælles delsekvens for enten X_{i-1} og Y_j eller for X_i og Y_{j-1} (eller evt. begge).

Observation (optimale delproblemer) for Case II:

Hvis Z er en længste fælles delsekvens for X_i og Y_j , må den være en længste fælles delsekvens for enten X_{i-1} og Y_j eller for X_i og Y_{j-1} (eller evt. begge). For hvis der fandtes en længere fælles delsekvens for enten X_{i-1} og Y_j eller for X_i og Y_{j-1} , ville denne også være en længere fælles delsekvens for X_i og Y_j .

$$x_1 x_2 x_3 x_4 \dots x_{i-1}$$
 A $y_1 y_2 y_3 y_4 \dots y_{j-1}$ C

$$x_1 x_2 x_3 x_4 \dots x_{i-1}$$
 A $y_1 y_2 y_3 y_4 \dots y_{j-1}$ C

Optimale delproblemer II

Lad T_1 være en længste fælles delsekvens for X_{i-1} og Y_j , og lad T_2 være en længste fælles delsekvens for X_i og Y_{j-1} .

Af observationen i Case II $(x_i \neq y_j)$ haves, at blandt T_1 og T_2 er der (mindst) én, som er en længste fælles delsekvens for X_i og Y_j .

Ingen af T_1 og T_2 kan være længere end den længste fælles delsekvens for X_i og Y_j (da de begge er delsekvenser af X_i og Y_j).

Så af observationen haves i Case II $(x_i \neq y_j)$:

- ▶ Hvis $lcs(i-1,j) \ge lcs(i,j-1)$, er en længste fælles delsekvens for X_{i-1} og Y_j også en længste fælles delsekvens for X_i og Y_j . Et symmetrisk udsagn gælder for "≤" og X_i og Y_{j-1} .

$$x_1 x_2 x_3 x_4 \dots x_{i-1}$$
 A $y_1 y_2 y_3 y_4 \dots y_{j-1}$ C

$$x_1 x_2 x_3 x_4 \dots x_{i-1}$$
 A $y_1 y_2 y_3 y_4 \dots y_{j-1}$ C

Rekursiv formel for lcs(i, j)

Alt i alt har vi fundet flg. rekursive formel for lcs(i, j):

$$\mathsf{lcs}(i,j) = \left\{ \begin{array}{ll} 0 & \text{if } i = 0 \text{ or } j = 0 \\ \mathsf{lcs}(i-1,j-1) + 1 & \text{if } i,j > 0 \text{ and } x_i = y_j \\ \mathsf{max}(\mathsf{lcs}(i-1,j),\mathsf{lcs}(i,j-1)) & \text{if } i,j > 0 \text{ and } x_i \neq y_j \end{array} \right.$$

Rekursiv formel for lcs(i, j)

Alt i alt har vi fundet flg. rekursive formel for lcs(i, j):

$$|cs(i,j)| = \begin{cases} 0 & \text{if } i = 0 \text{ or } j = 0 \\ |cs(i-1,j-1) + 1 & \text{if } i,j > 0 \text{ and } x_i = y_j \\ max(|cs(i-1,j),|cs(i,j-1)) & \text{if } i,j > 0 \text{ and } x_i \neq y_j \end{cases}$$

Den giver anledning til en naturlig, simpel rekursiv algoritme.

MEN: det er nemt at se at der er gentagelser blandt delproblemers delproblemer.

Så samme delproblemer bliver gentagne gange beregnet forskellige steder i rekursionstræet, og køretiden bliver meget dårlig.

Kan evt. løses med memoization: hav en tabel med plads til svaret på alle de mulige delproblemer lcs(i,j), og gem svaret, når det er beregnet første gang. Siden, slå det bare op.

Dynamisk programmering: udfyld i stedet direkte denne tabel bottom-up på struktureret måde.

i^{j}	0	1	2		n
0					
1					
2					
•					
•					
m					

$$lcs(i,j) = \begin{cases} 0 & \text{if } i = 0 \text{ or } j = 0\\ lcs(i-1,j-1) + 1 & \text{if } i,j > 0 \text{ and } x_i = y_j\\ max(lcs(i-1,j), lcs(i,j-1)) & \text{if } i,j > 0 \text{ and } x_i \neq y_j \end{cases}$$

i^{j}	0	1	2				$\mid n \mid$
0	0	0	0	0	0	0	0
1	0						
2	0						
•	0						
•	0						
m	0						

$$|cs(i,j)| = \begin{cases} 0 & \text{if } i = 0 \text{ or } j = 0 \\ |cs(i-1,j-1) + 1 & \text{if } i,j > 0 \text{ and } x_i = y_j \\ max(|cs(i-1,j),|cs(i,j-1)) & \text{if } i,j > 0 \text{ and } x_i \neq y_j \end{cases}$$

i^{j}	0	1	2				$\mid n \mid$
0	0	0	0	0	0	0	0
1	0						
2	0			K	*		
•	0			\	7		
•	0						
m	0						

$$lcs(i,j) = \begin{cases} 0 & \text{if } i = 0 \text{ or } j = 0\\ lcs(i-1,j-1) + 1 & \text{if } i,j > 0 \text{ and } x_i = y_j\\ max(lcs(i-1,j), lcs(i,j-1)) & \text{if } i,j > 0 \text{ and } x_i \neq y_j \end{cases}$$

\sum_{i}^{j}	0	1	2				n
0	8	8	%	Ø	8	8	8
1	8						
2	8						
•	8				7		
	8						
m	8						

$$lcs(i,j) = \begin{cases} 0 & \text{if } i = 0 \text{ or } j = 0\\ lcs(i-1,j-1) + 1 & \text{if } i,j > 0 \text{ and } x_i = y_j\\ max(lcs(i-1,j), lcs(i,j-1)) & \text{if } i,j > 0 \text{ and } x_i \neq y_j \end{cases}$$

\sum_{i}^{j}	0	1	2				n
0	(%)	8	%	Ø	8	8	8
1	%						
2	8						
•						7	
•	(%)						
m	% /						

$$lcs(i,j) = \begin{cases} 0 & \text{if } i = 0 \text{ or } j = 0\\ lcs(i-1,j-1) + 1 & \text{if } i,j > 0 \text{ and } x_i = y_j\\ max(lcs(i-1,j), lcs(i,j-1)) & \text{if } i,j > 0 \text{ and } x_i \neq y_j \end{cases}$$

\sum_{i}^{j}	0	1	2				$\mid n \mid$
0	(%)	8	%	Ø	8	8	0
1	%						
2	8						
•						7	-
•	/	1	\setminus			^	
\overline{m}	% /						

$$lcs(i,j) = \begin{cases} 0 & \text{if } i = 0 \text{ or } j = 0\\ lcs(i-1,j-1) + 1 & \text{if } i,j > 0 \text{ and } x_i = y_j\\ max(lcs(i-1,j), lcs(i,j-1)) & \text{if } i,j > 0 \text{ and } x_i \neq y_j \end{cases}$$

\sum_{i}^{j}	0	1	2				n
0	8	8	%	Ø	%	8	0
1	8				1		
2					/		
	8			Z /			
•	% /				V		
m	% /			/			

$$lcs(i,j) = \begin{cases} 0 & \text{if } i = 0 \text{ or } j = 0\\ lcs(i-1,j-1) + 1 & \text{if } i,j > 0 \text{ and } x_i = y_j\\ max(lcs(i-1,j), lcs(i,j-1)) & \text{if } i,j > 0 \text{ and } x_i \neq y_j \end{cases}$$

i^j	0	1	2				n
0	(%)	/	%	Ø	8	8	0
1	%						1
2							
•	8			$\Delta '$	//		
	(%)						
\overline{m}	% /						

$$lcs(i,j) = \begin{cases} 0 & \text{if } i = 0 \text{ or } j = 0\\ lcs(i-1,j-1) + 1 & \text{if } i,j > 0 \text{ and } x_i = y_j\\ max(lcs(i-1,j), lcs(i,j-1)) & \text{if } i,j > 0 \text{ and } x_i \neq y_j \end{cases}$$

Køretid

Dynamisk programmering: udfyld tabel over lcs(i,j) bottom-up på struktureret måde.

\sum_{i}^{j}	0	1	2				$\mid n \mid$
0	0	0	0	0	0	0	0
1	0						
2	0			K	^		
•	0			+	Z		
•	0						
m	0						

Tabelstørrelse: mn

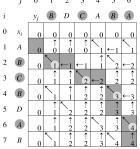
Udfyld tabelindgang: $O(\max \text{ størrelse af røde graf}) = O(1)$.

Tid i alt: O(produktet af de to) = O(mn).

Find en konkret løsning

lcs(m, n) er længden af en længste fælles delsekvens for $X = X_m$ og $Y = Y_n$.

Hvis vi gerne vil finde en konkret fælles delsekvens af denne længde: Gem for hvert felt i tabellen hvilken af de tre røde pile som gav lcs(i,j)-værdien i dette felt.



Følg gemte pile baglæns fra lcs(m, n). Når en skrå pil følges er det en Case I, og x_i (= y_i) udskrives. Ellers er den en Case II, og intet udskrives.

I alt udskrives en længste fælles delsekvens for X og Y i baglæns orden i tid O(m+n).

Pladsforbrug for LCS

Hvis vi kun skal bruge længden af længste fælles delsekvens, kan vi nøjes med $min\{m, n\}$ plads:

$i^{\mathcal{I}}$	0	1	2				n
0	0	0	0	0	0	0	0
1	0						
2	0						
•	9					7	_
-	0	1				*	
\overline{m}	0						

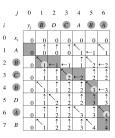
Pladsforbrug for LCS

Hvis vi kun skal bruge længden af længste fælles delsekvens, kan vi nøjes med $min\{m, n\}$ plads:

0 0 1 2 · · · n
0 0 0 0 0 0 0 0 0
1 0 0 0 0 0 0 0 0
2 0 0 0 0 0 0 0 0

m 0 0 0 0 0 0 0 0 0

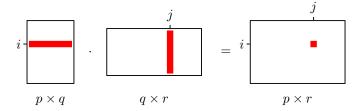
Hvis vi skal bruge en længste fælles delsekvens, må vi gemme hele tabellen, dvs. bruge $\Theta(mn)$ plads (da vi ikke kender stien tilbage, må vi gemme hele tabellen):



[Hirschberg gav i 1975 en metode til også at opnå dette med $\min\{m, n\}$ plads, men det er ikke pensum i DM507.]

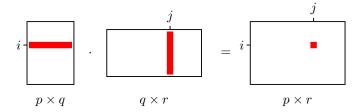
Eksempel 2: Multi-Matrix-multiplikation

En $p \times q$ matrix A_1 og en $q \times r$ matrix A_2 kan multipliceres i tid O(pqr). Resultatet er en $p \times r$ matrix.



Eksempel 2: Multi-Matrix-multiplikation

En $p \times q$ matrix A_1 og en $q \times r$ matrix A_2 kan multipliceres i tid O(pqr). Resultatet er en $p \times r$ matrix.



Matrix-multiplikation er associativ:

$$A_1\cdot (A_2\cdot A_3)=(A_1\cdot A_2)\cdot A_3$$

Matrix-multiplikation er associativ:

$$A_1 \cdot (A_2 \cdot A_3) = (A_1 \cdot A_2) \cdot A_3$$

Men køretiden er IKKE ens.

Matrix-multiplikation er associativ:

$$A_1 \cdot (A_2 \cdot A_3) = (A_1 \cdot A_2) \cdot A_3$$

Men køretiden er IKKE ens. Eksempel:

$$\begin{array}{cccc} A_1 & A_2 & A_3 \\ 10 \times 100 & 100 \times 5 & 5 \times 50 \end{array}$$

$$(A_2 \cdot A_3)$$
: 100×50 $(A_1 \cdot A_2)$: 10×5

Matrix-multiplikation er associativ:

$$A_1 \cdot (A_2 \cdot A_3) = (A_1 \cdot A_2) \cdot A_3$$

Men køretiden er IKKE ens. Eksempel:

$$A_1$$
 A_2 A_3 10×100 100×5 5×50 $(A_2 \cdot A_3)$: 100×50 $(A_1 \cdot A_2)$: 10×5

Tid for
$$A_1 \cdot (A_2 \cdot A_3)$$
 er $10 \cdot 100 \cdot 50 + 100 \cdot 5 \cdot 50 = 75.000$

Matrix-multiplikation er associativ:

$$A_1 \cdot (A_2 \cdot A_3) = (A_1 \cdot A_2) \cdot A_3$$

Men køretiden er IKKE ens. Eksempel:

$$A_1$$
 A_2 A_3 10×100 100×5 5×50

$$(A_2 \cdot A_3)$$
: 100×50 $(A_1 \cdot A_2)$: 10×5

Tid for
$$A_1 \cdot (A_2 \cdot A_3)$$
 er $10 \cdot 100 \cdot 50 + 100 \cdot 5 \cdot 50 = 75.000$
Tid for $(A_1 \cdot A_2) \cdot A_3$ er $10 \cdot 100 \cdot 5 + 10 \cdot 5 \cdot 50 = 7.500$

Spørgsmålet:

For et produkt af n matricer

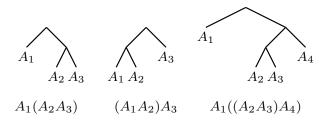
$$A_1 \cdot A_2 \cdot A_3 \cdot \cdots \cdot A_{n-1} \cdot A_n$$

med kombatible dimensioner

hvad er den billigste rækkefølge at gange dem sammen i?

Beregningstræer

Rækkefølge = parentes-sætning = binært beregningstræ:



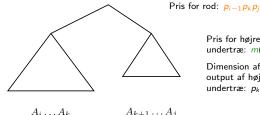
Optimale delproblemer og rekursiv ligning

Lad m(i, j) være prisen for bedste måde at gange A_i, \ldots, A_i sammen på.

Observation (optimale delproblemer):

Undertræerne for roden af et optimalt træ må selv være optimale beregningstræer.

Pris for venstre undertræ: m(i, k)Dimension af output af venstre undertræ: $p_{i-1} \times p_{\nu}$



Pris for høire undertræ: m(k+1, i)Dimension af

output af høire undertræ: $p_k \times p_i$

Prøv alle placeringer af rod, dvs. alle split A_i, \ldots, A_k og A_{k+1}, \ldots, A_i :

$$m(i,j) = \begin{cases} 0 & \text{if } i = j \\ \min_{i \le k < j} \{ m(i,k) + m(k+1,j) + \frac{p_{i-1}p_kp_j}{p_i} \} & \text{if } i < j \end{cases}$$

Gentagelser blandt delproblemers delproblemer. Lav tabel og udfyld systematisk. Målet er at kende m(1, n).

Gentagelser blandt delproblemers delproblemer. Lav tabel og udfyld systematisk. Målet er at kende m(1, n).

$$m(i,j) = \begin{cases} 0 & \text{if } i = j \\ \min_{i \le k < j} \{ m(i,k) + m(k+1,j) + \frac{p_{i-1}p_kp_j}{p_i} \} & \text{if } i < j \end{cases}$$

i^{j}	1	2	3				n
1	% /					/	
2						7	1
3			8				1
-				8			
					8		
\overline{n}							% /

Gentagelser blandt delproblemers delproblemer. Lav tabel og udfyld systematisk. Målet er at kende m(1, n).

$$m(i,j) = \begin{cases} 0 & \text{if } i = j \\ \min_{i \le k < j} \{m(i,k) + m(k+1,j) + \frac{p_{i-1}p_kp_j}{p_i}\} & \text{if } i < j \end{cases}$$

\sum_{i}^{j}	1	2	3				n
1	8					1	
2						7	1
3			0				1
-				8			
					8		
n							%

Tabelstørrelse: $O(n^2)$.

Udfyld tabelindgang: $O(\max \text{ størrelse af røde graf}) = O(n)$.

Tid i alt: $O(\text{produktet af de to}) = O(n^3)$.

Gentagelser blandt delproblemers delproblemer. Lav tabel og udfyld systematisk. Målet er at kende m(1, n).

$$m(i,j) = \begin{cases} 0 & \text{if } i = j \\ \min_{i \le k < j} \{m(i,k) + m(k+1,j) + \frac{p_{i-1}p_kp_j}{p_i}\} & \text{if } i < j \end{cases}$$

\underline{i}^{j}	1	2	3				n
1	8					/	
2						7	1
3			0				1
				8			
					8/		
						W	
n							%

Tabelstørrelse: $O(n^2)$.

Udfyld tabelindgang: $O(\max \text{ størrelse af røde graf}) = O(n)$.

Tid i alt: $O(\text{produktet af de to}) = O(n^3)$.

Find konkret løsning: følg de optimale valg baglæns.