Algoritmer og Datastrukturer

Datalogisk Institut, Københavns Universitet,
Pawel Winter & Christian Wulff-Nilsen

24-timers hjemmeprøve, Digital Eksamen, 1. april kl. 9:00 til 2. april kl. 9:00, 2020

Alle hjælpemidler så som kursusbogen og noter må benyttes under eksamen. Alle elektroniske hjælpemidler kan benyttes. Besvarelsen kan være enten på dansk eller engelsk og skal helst skrives i LaTeX, OpenOffice, Word eller lignende. Dog tillades også håndskrevne besvarelser, såfremt de er letlæselige for censor og eksaminator. Figurer må gerne laves i hånden. Det er valgfrit, om man vil digitalisere eventuelle håndskrevne dele af besvarelsen via scan eller via fotografi med kamera, smartphone eller lignende. **Aflevering er i Digital Eksamen.** Angiv eksamensnummer men ikke navn på første side i afleveringen. **Krav til afleveringsformat:**

- **en enkelt pdf indeholdende din besvarelse,**
- **ingen side i pdf'en må indeholde dele af svar fra mere end en delopgave; start derfor med en ny side for hver delopgave (og ikke kun for hver opgave). Besvarelsen af en delopgave skal være på hinanden følgende sider i pdf'en (eks.: side 3–5 er fint for samme delopgave, men ikke side 3, 5 og 6).**

Bemærk: deadline for aflevering er 2. april kl. 9:00 om morgenen. Forsinkede afleveringer accepteres ikke. Sørg for at aflevere i god tid. Der er mulighed for at uploade afleveringer flere gange indtil deadline. Skulle der mod forventning være problemer med at aflevere i Digital Eksamen tæt på deadline, send da straks en e-mail til `uddannelse@diku.dk` og `koolooz@di.ku.dk` med en beskrivelse af, hvad der er gået galt, og sørg for som vedhæftninger til denne e-mail at have besvarelsen som pdf samt et screendump, der viser problemet med at uploade i Digital Eksamen. Din besvarelse vil blive vurderet, hvis du har en gyldig begrundelse for ikke at aflevere i Digital Eksamen, og hvis du afleverer inden deadline.

Det er ikke tilladt at kommunikere med andre omkring sættet under eksamen. Den eneste undtagelse er spørgsmål omkring uklarheder i formuleringen af eksamenssættet, som kan stilles på Absalon i tråden "Spørgsmål til eksamenssættet" eller via e-mail til Christian (`koolooz@di.ku.dk`). Eventuelle rettelser eller kommentarer til eksamenssættet under eksamen vil blive givet via Announcements på Absalon.

Fuld besvarelse er besvarelse af alle opgaver. De enkelte opgavers og delopgavers vægt ved bedømmelsen er angivet i procent. Der må gerne refereres til algoritmer, datastrukturer og resultater fra pensumdelen af kursusbogen, dog ikke resultater i Exercises eller Problem-opgaverne i bogen. Specielt må man gerne begrunde en påstand med at henvise til, at det umiddelbart følger fra et bestemt lemma, en sætning eller lignende i pensumdelen af kursusbogen. **Det er ikke tilladt at basere sine svar på andre kilder end pensumdelen af kursusbogen.**

ADVARSEL: eksamen er individuel. Ved mistanke om eksamenssnyd har vi pligt til at melde dette til studielederen. Eksamenssnyd kan føre til bortvisning fra universitetet. Udsæt ikke dig selv eller dine medstuderende for denne risiko!

Dette eksamenssæt består af 4 opgaver og er på 6 sider, inkl. denne forside.

Opgave 1 (15%)

Denne opgave omhandler datastrukturer.

Del 1.1 (8%) Forklar overordnet, hvad en Fibonacci heap er, hvilke algoritmer inden for pensum denne datastruktur anvendes i. Kom herunder ind på køretider, både for udvalgte Fibonacci heap-operationer og for de algoritmer, der anvender Fibonacci heaps. Der forventes ikke nogen detaljeret gennemgang af algoritmerne eller operationerne. Brug gerne figurer og undlad beviser. Eventuelle håndkørsler skal være på små instanser. Teksten i din besvarelse skal fylde max en side; der er ingen grænse på, hvor meget figurer må fylde.

I de resterende delopgaver fokuserer vi på amortiseret analyse.

I kapitlet "Amortized Analysis" i kursusbogen gennemgås eksemplet med en binær tæller. I denne opgave betragter vi i stedet en tæller for decimaltal. Tælleren består af k cifre hvert mellem 0 og 9 og disse repræsenteres ved et 0-indekseret array A med k indgange, som til at starte med alle er lig 0. Der udføres nu n INCREMENT-operationer på A, og hver operation øger decimaltallet i tælleren med 1. Hvis alle cifre i tælleren er 9, vil en INCREMENT-operation ændre disse cifre til 0.

Del 1.2 (2%) Giv pseudo-kode for INCREMENT og forklar kort i ord, hvordan denne procedure fungerer. Undlad et korrekthedsbevis.

Del 1.3 (5%) Vis at worst-case-køretiden for INCREMENT er $O(k)$, og brug accounting-metoden til at vise, at den amortiserede køretid for INCREMENT er $O(1)$.

Opgave 2 (31%)

Denne opgave omhandler paradigmet Dynamisk Programmering.

Del 2.1 (8%) Giv en overordnet beskrivelse af dette paradigme og kom ind på det, du vurderer er det vigtigste inden for emnet Dynamisk Programmering. Brug gerne figurer og undlad beviser. Eventuelle håndkørsler skal være på små instanser. Teksten i din besvarelse skal fylde max en side; der er ingen grænse på, hvor meget figurer må fylde.

I resten af denne opgave kigger vi på *plate-cutting*-problemet. En input-instans for dette problem består af

1. en rektangulær plade, specificeret ved en heltalsbredde $b \geq 1$ og en heltalshøjde $h \geq 1$,
2. en pris p_{ij} for et delrektangel af bredde i og højde j for alle heltal i og j, hvor $1 \leq i \leq b$ og $1 \leq j \leq h$, og
3. et heltal $s \geq 0$.

Målet er at finde en udskæring af pladen i mindre rektangulære delplader, så den samlede fortjeneste ved at sælge disse plader maksimeres. Saven, der anvendes, kan dog kun foretage udskæringer, der kan opnås via vandrette eller lodrette snit, der hver især opsplitter en rektangulær delplade i to mindre rektangulære delplader. Alle delplader skal have heltalsbredde og heltalshøjde. Da saven efter noget tid bliver sløv, er der yderligere den begrænsning, at summen af længderne af alle snit i udskæringen højst må være s. Ligesom i rod-cutting-problemet er der ikke nogen omkostning ved at lave de enkelte snit.

Et delproblem er specificeret ved heltal i, j og s', hvor $1 \leq i \leq b$, $1 \leq j \leq h$ og $0 \leq s' \leq s$. Lad $r_{i,j,s'}$ være den maksimale fortjeneste, der kan opnås ved udskæring af en delplade med bredde i og højde j, og hvor summen af alle snitlængder i udskæringen af denne delplade højst er s'.

Betragt følgende rekursionsformel for tilfældet, hvor enten $b > 1$ eller $h > 1$:

$$\begin{aligned} r_{b,h,s} = \max\{&p_{b,h}, \max\{r_{i,h,s_1} + r_{b-i,h,s_2} | 1 \leq i < b, s_1 + s_2 \leq s - h\}, \\ &\max\{r_{b,j,s_1} + r_{b,h-j,s_2} | 1 \leq j < h, s_1 + s_2 \leq s - b\}\}. \end{aligned}$$

Her antages det implicit, at alle parametre i, j, s_1 og s_2 er heltal, at s_1 og s_2 er ikke-negative, samt at max over en tom mængde er $-\infty$.

Del 2.2 (10%) Argumentér for at rekursionsformlen er korrekt og kom herunder ind på optimal delstruktur. Vis desuden at $r_{b,h,s} = p_{b,h}$, når $b = h = 1$.

Del 2.3 (13%) Beskriv (i ord og/eller pseudo-kode) en dynamisk programmerings-algoritme til at finde $r_{b,h,s}$, argumentér for korrektheden af denne og analysér dens køretid og pladsforbrug. Både køretid og plads skal være polynomiel i b og h. Gør det klart, om din algoritme er en bottom-up-algoritme eller en top-down-algoritme med memoization.

Vink: for køretidsanalysen, vis først at den er polynomiel i b, h og s. Vis dernæst, at vi kan antage, at $s = O(bh)$.

Opgave 3 (30%)

Denne opgave omhandler paradigmet Grådige Algoritmer.

Del 3.1 (8%) Giv en overordnet beskrivelse af dette paradigme og kom ind på det, du vurderer er det vigtigste inden for emnet Grådige Algoritmer. Brug gerne figurer og undlad beviser. Eventuelle håndkørsler skal være på små instanser. Teksten i din besvarelse skal fylde en halv til en hel side; der er ingen grænse på, hvor meget figurer må fylde.

I resten af opgaven betragter vi en variant af strengkomprimerings-problemet fra afsnit 16.3 i kursusbogen. I stedet for at komprimere en tekststreng til en binær streng, kigger vi her i stedet på problemet med at komprimere en tekststreng til en streng bestående af symboler fra mængden $\Sigma = \{0, 1, 2\}$ vha. prefix codes. Som i kursusbogen er input en liste af n heltal, der specificerer frekvenserne for hvert af de n typer af bogstaver i strengen, der skal komprimeres [1]. Vi ønsker en komprimeret streng af minimal længde bestående af symboler fra Σ.

For at simplicifere opgaven antager vi, at n er ulige og mindst 3. Der gælder da, at i ethvert optimalt parse-træ for input har hver indre knude (knude der ikke er et blad) netop tre børn. Dette resultat må gerne benyttes i det følgende og skal ikke bevises.

I det følgende kigger vi på en modificeret version af Huffmans algoritme, der finder et optimalt parse-træ, når mængden af symboler er Σ. I stedet for at vælge to symboler i hvert skridt (som i kursusbogen) vælger den modificerede version tre symboler med de mindste frekvenser og tilføjer dem som børn af en fælles forælder, der har frekvens lig med summen af børnenes frekvenser. Ellers fungerer algoritmen på samme måde som beskrevet i kursusbogen.

Del 3.2 (6%) Håndkør den modificerede Huffman-algoritme på instansen bestående af de $n = 7$ bogstaver a til g, hvor $a.freq = 2$, $b.freq = 3$, $c.freq = 4$, $d.freq = 5$, $e.freq = 6$, $f.freq = 7$ og $g.freq = 8$. Hvert skridt i håndkørslen skal illustreres, og den resulterende prefix-kode for hvert bogstav skal angives.

Del 3.3 (6%) Analysér køretiden for den modificerede Huffman-algoritme.

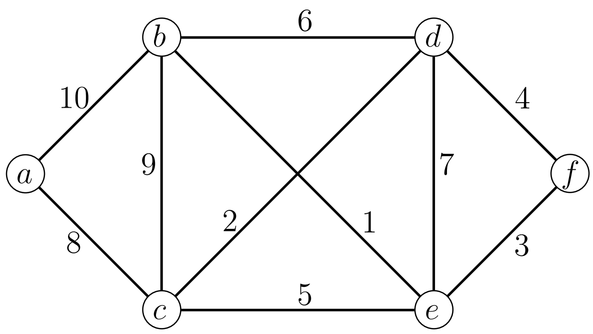
Del 3.4 (2%) I kursusbogen defineres omkostningen $B(T)$ af et parse-træ T, og denne omkostning er netop antallet af bits i den tilhørende komprimerede bitstreng. Kan vi definere $B(T)$ på samme måde, når vi betragter mængden af symboler Σ, så $B(T)$ ligeledes udtrykker længden af den komprimerede streng? Argumentér for dit svar.

Del 3.5 (8%) Bevis at den modificerede algoritme har Greedy Choice Property ved at tilpasse beviset for Greedy Choice Property for Huffmans algoritme i kursusbogen. Husk at nævne, hvad det grådige valg for den modificerede algoritme er. Optimal Substructure skal ikke bevises.

[1] Som i kursusbogen tillader vi n at være vilkårlig stor, så strengen, der skal komprimeres, kan bestå af andet end blot bogstaverne a til å.

Opgave 4 (24%)

Denne opgave omhandler minimum udspændende træ-problemet (Minimum Spanning Tree, forkortet MST).



Figur 1: Grafen H med vægte på kanterne.

Lad H være grafen i Figur 1. For hver kant er der angivet en vægt.

Del 4.1 (6%) Vis forløbet af Kruskals algoritme, når vi ønsker at finde et minimum udspændende træ i H. For hvert skridt skal der vises den del af en minimum udspændede skov der er fundet indtil nu. Desuden skal der vises, hvorledes det besluttes, om den aktuelt betragtede kant skal tages med i det endelige træ. Det er ikke nødvendigt at tegne alle kanterne med vægte for hvert skridt. Man skal blot vise de kanter der indtil nu er inkluderet i den aktuelle skov.

I de to næste delopgaver betegner T det MST i H, der består af kanterne (a, c), (c, d), (b, e), (d, f) og (e, f).

Del 4.2 (2%) Argumentér for, at hvis vi ændrer vægten af kanten (d, f) til 6, så vil T ikke længere være et MST i H.

Del 4.3 (2%) Argumentér for, at hvis vi ændrer vægten af kanten (b, c) til 3, så vil T ikke længere være et MST i H (vægten af (d, f) er her 4 som i figuren).

(*Eksamenssættet fortsætter på næste side*)

Nedenstående to spørgsmål relaterer sig til et fast MST M i en graf $G = (V, E, w)$ med kantvægte givet ved vægt-funktion w. Lad x være en fast kant i E og lad $k \in \mathbb{N}$ være et fast positivt heltal.

Lad $G' = (V, E, w')$, hvor w' er identisk med w for alle kanter i $E \setminus \{x\}$, og hvor $w'(x) = k$.

Del 4.4 (7%) Antag at x tilhører M. Beskriv en egenskab ved G (herunder en egenskab ved vægtfunktionen w), så

1. hvis M er et MST i G', da må denne egenskab gælde, og
2. hvis egenskaben gælder, da er M et MST i G'.

Argumentér for dine svar.

Vink: Se på M minus kanten x og en passende valgt delmængde af E. Hvad må der gælde i G' om vægten $w'(x) = k$ ift. vægtene af kanterne i denne delmængde? For at vise del 2, argumentér for, at egenskaben sikrer, at den generiske MST-algoritme i kursusbogen kan finde M i G'.

Del 4.5 (7%) Antag nu at x ikke tilhører M. Beskriv en egenskab ved G (herunder en egenskab ved vægtfunktionen w), så

1. hvis M er et MST i G', da må denne egenskab gælde, og
2. hvis egenskaben gælder, da er M et MST i G'.

Argumentér for dine svar.

Vink: se på vejen mellem de to endepunkter af x i M. Hvad må der gælde i G' om vægten $w'(x) = k$ ift. vægten af kanterne på denne vej? For at vise del 2, argumentér (ligesom i den tidligere delopgave) for, at egenskaben sikrer, at den generiske MST-algoritme i kursusbogen kan finde M i G'.

EKSAMENSSÆTTET SLUTTER HER