Videregående Algoritmik, DIKU 2005/06
Torsdag 15. december
David Pisinger

Løsning af $\mathcal{NP}$-hårde problemer

- Løs til optimalitet i eksponentiel tid
- Find tilnærmet løsning i polynomiel tid

Oversigt

- Grænseværdier
- Branch-and-bound algoritmens komponenter
- Eksempler på branch-and-bound algoritmer
- Eksempler på grænseværdier
- Dominans af grænseværdi
- Kritiske og semikritiske knuder

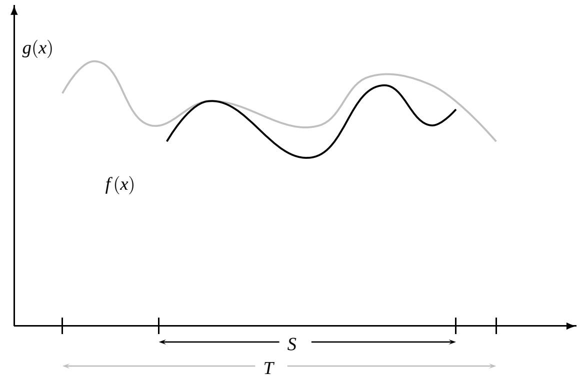
Selv om alle $\mathcal{NP}$-fuldstændige problemer kan reduceres til hinanden skal specifik struktur udnyttes i praksis

Grænseværdier

$$z = \max\{f(x) \mid x \in S\} \qquad \text{f.x. } S \text{ delproblem}$$

- *nedre grænseværdi* $\mathcal{L} \in \mathbb{R}$: $\mathcal{L} \leq z$
- *øvre grænseværdi* $\mathcal{U} \in \mathbb{R}$: $\mathcal{U} \geq z$

For enhver løsning $x' \in S$ er $f(x')$ nedre grænseværdi



Definition 1 $\quad P : z = \max\{f(x) \mid x \in S\}$

$\quad R : z_R = \max\{g(x) \mid x \in T\}$

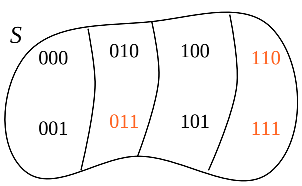
R er en relaksering af P hvis

(i) $S \subseteq T$

(ii) $g(x) \geq f(x)$ *for alle* $x \in S$

Grænseværdier

Hvis R er en relaksering af P så er z_R en øvre grænseværdi for S



Grænseværditest:

$$\mathcal{U}(S_i) \leq \mathcal{L} \qquad \Rightarrow \qquad x^* \notin S_i$$

hvor x^* optimal løsning med $f(x^*) > \mathcal{L}$

Knapsack problem, øvre grænseværdi

Fraktionelle knapsack problem i Cormen

$$\mathcal{U}^1_{\text{KP}} = \max\left\{\sum_{j=1}^{n} p_j x_j \,\middle|\, \sum_{j=1}^{n} w_j x_j \leq c,\ 0 \leq x_j \leq 1\right\}$$

Køretid

$O(n)$ mediansøgning

Relaksering

Originale knapsack problem

$$f(x) = \sum_{j=1}^{n} p_j x_j$$

$$S = \left\{(x_1,\ldots,x_n) \,\middle|\, \sum_{j=1}^{n} w_j x_j \leq c, x_j \in \{0,1\}\right\}$$

Fraktionelle knapsack problem

$$g(x) = \sum_{j=1}^{n} p_j x_j$$

$$T = \left\{(x_1,\ldots,x_n) \,\middle|\, \sum_{j=1}^{n} w_j x_j \leq c, 0 \leq x_j \leq 1\right\}$$

Kontrollerer betingelser

(i) $S \subseteq T$

(ii) $g(x) \geq f(x)$ for alle $x \in S$

Branch-and-bound

$$z = \max\{f(x) \mid x \in S\}$$

$\mathcal{L} := -\infty$; $L := \{S\}$
while $L \neq \emptyset$
  vælg et delproblem S_i fra L
  $L := L \setminus \{S_i\}$
  if $S_i \neq \emptyset$ **then**
    find en øvre grænseværdi $\mathcal{U}(S_i)$
    find en lovlig løsning x i S_i
    if $\mathcal{U}(S_i) > \mathcal{L}$ **then** *(grænseværditest)*
      if $f(x) > \mathcal{L}$ **then** $\mathcal{L} := f(x)$; $x^* := x$
      opdel S_i i delproblemer $S_i^1, \ldots, S_i^k$
      tilføj delproblemerne til L, dvs. sæt
        $L := L \cup \{S_i^1, \ldots, S_i^k\}$
    endif
  endif
endwhile

- L liste af delproblemer S_i ofte organiseret prioritetskø
- Delproblemerne i L er *åbne delproblemer*
- Behandlede deproblemer er *lukkede delproblemer*
- global nedre grænseværdi $\mathcal{L}$, tilhørende løsning x^*

Branch-and-bound, fire komponenter

1. *Øvre grænseværdifunktion*
   - lineær relaksering
   - lagrange relaksering
   - surrogat relaxering
   - semidefinite relaxering
2. *Nedre grænseværdi* (hidtil bedst kendte løsning)
   - i hver branch-and-bound knude, check om $|S_i| = 1$
   - omform øvre grænseværdi til lovlig løsning
   - heuristik for hvert delproblem
   - god initiel løsning inden branch-and-bound
3. *Søgestrategi* (rækkefølge af delproblemer)
   - dybde-først søgning
   - bedste-først søgning
   - bredde-først søgning
   - heuristisk styret søgning
4. *Forgreningsregel* (opdeling af løsningsrum)
   - tilføj ekstra begrænsning (problem skal stadig kunne løses effektivt)
   - opdeling i 2 eller flere delproblemer (ikke for mange)
   - søgetræet bør vokse symmetrisk

Knapsack problem, nedre grænseværdi

Grådig algoritme

j	1	2	3	4	5	6	7
p_j	6	5	8	9	6	7	3
w_j	2	3	6	7	5	9	4

$c = 9$, $n = 7$

Øvre grænseværdi: genstande $1, 2$ brøk af $b = 3$

$$\mathcal{U}^1_{\text{KP}} = 11 + 8\frac{2}{3} = 16\frac{1}{3},$$

kan rundes ned til $\mathcal{U}^1_{\text{KP}} = 16$

Nedre grænseværdi: genstande $1, 2, 7$

$$\mathcal{L} = 14$$

Knapsack problem, branch-and-bound

j	1	2	3	4	5	6	7
p_j	6	5	8	9	6	7	3
w_j	2	3	6	7	5	9	4

$c = 9$, $n = 7$

Dybde-først søgning (stak lagrer åbne delproblemer L)
Heuristisk styret søgestrategi

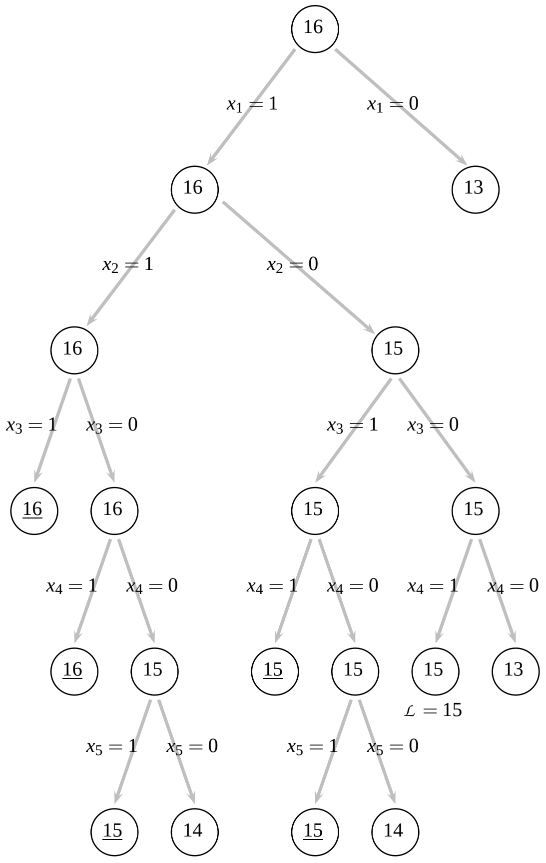
BranchBoundKnapsack$(\overline{p}, \overline{w}, i)$
if $\overline{w} > c$ **then return**
if $\overline{p} > \mathcal{L}$ **then** $\mathcal{L} = \overline{p}$; $x^* := x$
if $i > n$ **then return**
Udregn $\mathcal{U}^1_{\text{KP}}$ defi neret på genstande $\{i, \ldots, n\}$ og med kapacitet $c - \overline{w}$.
Øvre grænseværdi for det betragtede delproblem er $\mathcal{U} = \overline{p} + \lfloor \mathcal{U}^1_{\text{KP}} \rfloor$.
if $\mathcal{U} > \mathcal{L}$ **then**
  $x_i := 1$; BranchBoundKnapsack$(\overline{p} + p_i, \overline{w} + w_i, i + 1)$
  $x_i := 0$; BranchBoundKnapsack$(\overline{p}, \overline{w}, i + 1)$
endif

Hvor

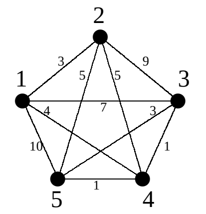
- i: næste genstand vi skal forgrene på
- $\overline{p}$: profitsummen af valgte genstande
- $\overline{w}$: vægtsummen af valgte genstande

init: $x = (0, 0, \ldots, 0)$, $\mathcal{L} = 0$ (eller bedre)
kald: BranchBoundKnapsack$(0, 0, 1)$

Knapsack problem, branch-and-bound



Dense subgraph problem



- komplet (orienteret) graf $G = (V, E, c)$, heltal k
- udvælg $U \subseteq V$ af størrelse $|U| = k$
- sum af kantvægte mellem knuder i U maksimeres

$\mathcal{NP}$-hårdt ved "reduktion" fra klike

$$z = \max \left\{ \sum_{i \in U} \sum_{j \in U} c_{ij} \,\middle|\, U \subseteq V, |U| = k \right\}$$

Eksempel $G = (V, E, c), k = 3$

i \ j	1	2	3	4	5	6	7
1	0	3	7	4	10	5	7
2	3	0	9	5	5	10	6
3	7	9	0	1	3	2	4
4	4	5	1	0	1	9	1
5	10	5	3	1	0	3	2
6	5	10	2	9	3	0	3
7	7	6	4	1	2	3	0

Optimal løsning $U = \{2, 4, 6\}$, løsningsværdi 48

Dense subgraph problem, øvre grænseværdi

Sæt en parantes i objektfunktionen

$$z = \max \left\{ \sum_{i \in U} \left(\sum_{j \in U} c_{ij} \right) \,\middle|\, U \subseteq V, |U| = k \right\}$$

Lad $\overline{c}_i$ øvre grænse på kantvægt-sum fra knude i

$$\overline{c}_i = \max \left\{ \sum_{j \in U} c_{ij} \,\middle|\, U \subseteq V, |U| = k \right\}$$

Grænseværdi

$$\mathcal{U}^1_{\text{DSP}} = \max \left\{ \sum_{i \in U} \overline{c}_i \,\middle|\, U \subseteq V, |U| = k \right\}$$

Køretid

$V \cdot O(\lvert V \rvert)$	vælg k største kantvægte fra knude i (Problem 9-1 side 194 i Cormen)
$O(\lvert V \rvert)$	vælg k største af $\overline{c}_1, \overline{c}_2, \ldots, \overline{c}_n$

Samlet: $O(|V|^2)$ tid

Dense subgraph problem, øvre grænseværdi

Relaksering

Originale problem

$$f(x) = \textstyle\sum_{i \in U} \left(\sum_{j \in U} c_{ij} \right) \qquad S = \{U \subseteq V, |U| = k\}$$

Nye problem

$$g(x) = \textstyle\sum_{i \in U} \overline{c}_i \qquad T = \{U \subseteq V, |U| = k\}$$

Kontrollerer betingelser

(i) $S \subseteq T$

(ii) $g(x) \geq f(x)$ for alle $x \in S$

Eksempel $n = 7, k = 3$

i \ j	1	2	3	4	5	6	7	$\overline{c}_i$
1	0	3	7	4	10	5	7	24
2	3	0	9	5	5	10	6	25
3	7	9	0	1	3	2	4	20
4	4	5	1	0	1	9	1	18
5	10	5	3	1	0	3	2	18
6	5	10	2	9	3	0	3	24
7	7	6	4	1	2	3	0	17

Øvre grænseværdi $\mathcal{U}^1_{\text{DSP}} = 73$

Dense subgraph problem, branch-and-bound

Heuristisk styret søgestrategi
$\overline{c}_i$ øvre grænseværdi for kantvægt-summen fra knude i

$_i\ ^j$	1	2	3	4	5	6	7
1	0	3	7	4	10	5	7
2	3	0	9	5	5	10	6
3	7	9	0	1	3	2	4
4	4	5	1	0	1	9	1
5	10	5	3	1	0	3	2
6	5	10	2	9	3	0	3
7	7	6	4	1	2	3	0

$V_3 \notin U$

$_i\ ^j$	1	2	3	4	5	6	7
1	0	3	7	4	10	5	7
2	3	0	9	5	5	10	6
3	7	9	0	1	3	2	4
4	4	5	1	0	1	9	1
5	10	5	3	1	0	3	2
6	5	10	2	9	3	0	3
7	7	6	4	1	2	3	0

$V_3 \in U$

$_i\ ^j$	1	2	3	4	5	6	7
1	**14**	3	7	4	10	5	7
2	3	**18**	9	5	5	10	6
3	7	9	0	1	3	2	4
4	4	5	1	**2**	1	9	1
5	10	5	3	1	**6**	3	2
6	5	10	2	9	3	**4**	3
7	7	6	4	1	2	3	**8**

- Knude i forbydes: slet tilhørende række og søjle
- Knude i vælges: slet den tilhørende række og søjle modificer (c_{ij})

$$c_{jj} := c_{jj} + c_{ij} + c_{ji}$$

for hvert $j = 1, \ldots, n$ hvor $j \neq i$

Traveling salesman problem

Symmetrisk traveling salesman problem

- vægtet graf (V, E, d), hvor d_{ij} afstand
- symmetriske afstande $d_{ij} = d_{ji}$
- find minimal længde Hamilton-kreds H

$$z = \min \left\{ \sum_{(i,j) \in H} d_{ij} \,\middle|\, H \subseteq E, H \text{ er en Hamilton-kreds} \right\}$$

Otte byer på Bornholm

Traveling salesman problem, eksempel

$_i\ ^j$	1	2	3	4	5	6	7	8
1	0	11	24	25	30	29	15	15
2	11	0	13	20	32	37	17	17
3	24	13	0	16	30	39	29	22
4	25	20	16	0	15	23	18	12
5	30	32	30	15	0	9	23	15
6	29	37	39	23	9	0	14	21
7	15	17	29	18	23	14	0	7
8	15	17	22	12	15	21	7	0

Optimal løsning:

- $1 \to 2 \to 3 \to 4 \to 5 \to 6 \to 7 \to 8 \to 1$
- længde af Hamilton-kreds $z = 100$

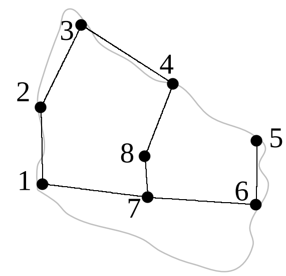
Traveling salesman problem, nedre grænseværdi

Vægtet graf (V, E, d)

- **1-træ**: udspændende træ på knuderne $2, 3, \ldots, n$
  knude 1 forbindes med to vilkårlige kanter
- **minimalt 1-træ**: mindste udspændende træ på knuderne $2, 3, \ldots, n$
  knude 1 forbindes med billigste kanter

Bevis: uafhængige delproblemer optimeres hver for sig

Eksempel 1-træ relaksering



$\mathcal{L}^{1}_{\text{TSP}} = 97$

Køretid

$O(\lvert E\rvert \log \lvert V\rvert)$	MST Kruskal
$O(\lvert E\rvert + \lvert V\rvert \log \lvert V\rvert)$	MST Prim, Fibonacci hobe
$O(\lvert V\rvert)$	to billigste kanter fra 1

Samlet: $O(|E| + |V| \log |V|)$

Traveling salesman problem, nedre grænseværdi

$$\mathcal{L}^1_{\text{TSP}} = \min\left\{ \sum_{(i,j)\in H} d_{ij} \,\middle|\, H \subseteq E, H \text{ er et 1-træ} \right\}$$

Relaksering

Originale problem

$$f(x) = \sum_{(i,j)\in H} d_{ij} \quad S = \left\{ H \subseteq E \,\middle|\, H \text{ er en Hamilton-kreds} \right\}$$

Nye problem

$$g(x) = \sum_{(i,j)\in H} d_{ij} \quad T = \left\{ H \subseteq E \,\middle|\, H \text{ er et 1-træ} \right\}$$

Kontrollerer betingelser (minimieringsproblem)

(i) $S \subseteq T$

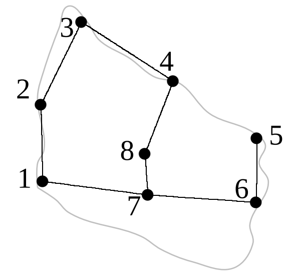
(ii) $g(x) \leq f(x)$ for alle $x \in S$

Traveling salesman problem, forgreningsstrategi

Metode 1

- Vælg en kant fra knude 1
- Vælg næste kant
- ...

Ikke god

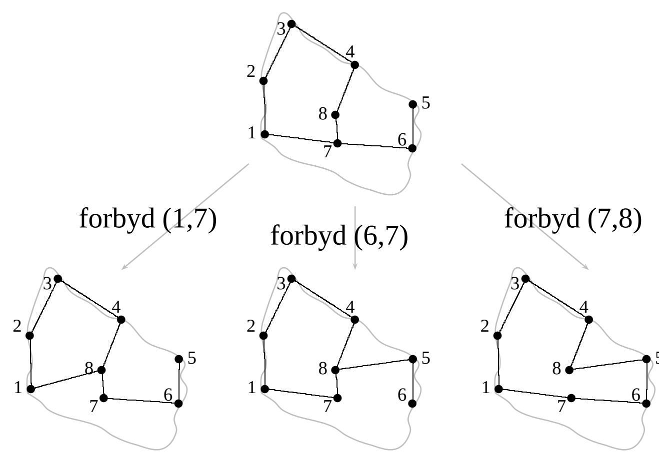


$\mathcal{L}^1_{\text{TSP}} = 97$

Metode 2:

- Udgangspunkt i *1-træ*
- Hvis alle knuder har valens 2, har vi Hamilton-kreds (og kan bortskære delproblem)
- Ellers vælg en knude med valens større end 2
- Forbyd på skift de kanter, som udgår fra knuden

Traveling salesman problem, forgreningsstrategi



Knude 7 har valens 3

- Forbyd $(1,7)$ prisen af 1-træet $\mathcal{L} = 97$
- Forbyd $(6,7)$ prisen af 1-træet $\mathcal{L} = 98$
- Forbyd $(7,8)$ 1-træet giver $\mathcal{L} = \mathcal{U} = 105$

Kvalitet af grænseværdifunktionen

maksimeringsproblem

$$z = \max\{f(x) \mid x \in S\}$$

to øvre grænseværdifunktioner $\mathcal{U}_1$ og $\mathcal{U}_2$

Definition 2 *grænseværdifunktion* $\mathcal{U}_1$ dominerer *grænseværdifunktion* $\mathcal{U}_2$ *hvis*

- $\mathcal{U}_1 \leq \mathcal{U}_2$ *for alle instanser*
- $\mathcal{U}_1 < \mathcal{U}_2$ *for mindst en instans*

NB: kræver ikke at $\mathcal{U}_1$ *altid* er bedre end $\mathcal{U}_2$

Dominans af grænseværdier, knapsack problem

$$\mathcal{U}^0_{\text{KP}} = \max \left\{ \sum_{j=1}^{n} p_j x_j \,\middle|\, \sum_{j=1}^{n} w_j x_j \leq c,\ x_j \geq 0 \right\}$$

Relaksering

Originale knapsack problem

$f(x) = \sum_{j=1}^{n} p_j x_j$

$S = \left\{ (x_1, \dots, x_n) \,\middle|\, \sum_{j=1}^{n} w_j x_j \leq c, x_j \in \{0,1\} \right\}$

nye knapsack problem

$g(x) = \sum_{j=1}^{n} p_j x_j$

$T = \left\{ (x_1, \dots, x_n) \,\middle|\, \sum_{j=1}^{n} w_j x_j \leq c, 0 \leq x_j \right\}$

Kontrollerer betingelser

(i) $S \subseteq T$

(ii) $g(x) \geq f(x)$ for alle $x \in S$

Køretid

Hvis sorteret profit-vægt forhold, udregnes $O(1)$ tid

$$\mathcal{U}^0_{\text{KP}} = \frac{c \cdot p_1}{w_1}$$

Dominans

- $\boxed{\mathcal{U}^1_{\text{KP}} \leq \mathcal{U}^0_{\text{KP}} \text{ for alle instanser}}$

  To relakseringer har samme objektfuktion

  $$\mathcal{U}^1_{\text{KP}} : T' = \{(x_1, \dots, x_n) \mid \sum_{j=1}^{n} w_j x_j \leq c, 0 \leq x_j \leq 1\}$$

  $$\mathcal{U}^0_{\text{KP}} : T = \{(x_1, \dots, x_n) \mid \sum_{j=1}^{n} w_j x_j \leq c, 0 \leq x_j \qquad \}$$

  Da $T' \subseteq T$ er

  $$\max_{x \in T'} f(x) \leq \max_{x \in T} f(x)$$

- $\boxed{\mathcal{U}^1_{\text{KP}} < \mathcal{U}^0_{\text{KP}} \text{ for en instanser}}$

  eksemplet: $\mathcal{U}^0_{\text{KP}} = 27$ mens $\mathcal{U}^1_{\text{KP}} = 16\frac{1}{3}$

Dominans af grænseværdier, traveling salesman

Strammere grænseværdi ved at omformulere afstandsmatricen (d_{ij})

- længden af en Hamilton-kreds er uændret
- 1-træ relakseringen returnerer strammere grænseværdi

Vi "straffer" 1-træer som ikke er en Hamilton-kreds

Omformuler afstandsmatrix

- "straf" knuder med valens større end 2
- "beløn" knuder med valens mindre end 2

  $$d'_{ij} = d_{ij} + (v_i - 2) + (v_j - 2)$$

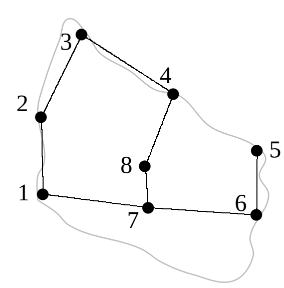
- ny grænseværdi ved mindste 1-træ for (d'_{ij})

$$\mathcal{L}^2_{\text{TSP}} = \min \left\{ \sum_{(i,j) \in H} d'_{ij} \,\middle|\, H \subseteq E, H \text{ er et 1-træ} \right\}$$

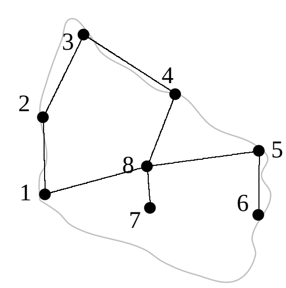
Dominans af grænseværdier, traveling salesman

$i \backslash j$	1	2	3	4	5	6	7	8
1	0	11	24	25	30	29	15	15
2	11	0	13	20	32	37	17	17
3	24	13	0	16	30	39	29	22
4	25	20	16	0	15	23	18	12
5	30	32	30	15	0	9	23	15
6	29	37	39	23	9	0	14	21
7	15	17	29	18	23	14	0	7
8	15	17	22	12	15	21	7	0

$i \backslash j$	1	2	3	4	5	6	7	8
1	0	11	24	25	29	29	16	15
2	11	0	13	20	31	37	18	17
3	24	13	0	16	29	39	30	22
4	25	20	16	0	14	23	19	12
5	29	31	29	14	0	8	23	14
6	29	37	39	23	8	0	15	21
7	16	18	30	19	23	15	0	8
8	15	17	22	12	14	21	8	0



$\mathcal{L}^1_{\text{TSP}} = 97$



$\mathcal{L}^2_{\text{TSP}} = 97$

Man kan gentage iterationsprocessen et antal gange
Efterfølgende iterationer $\mathcal{L}^2_{\text{TSP}} = 98, 99, 99, 100$

Relaksering

Originale problem

$$f(x) = \sum_{(i,j)\in H} d_{ij}$$
$$S = \left\{ H \subseteq E \,\middle|\, H \text{ er en Hamilton-kreds} \right\}$$

Nye problem

$$g(x) = \sum_{(i,j)\in H} d'_{ij}$$
$$T = \left\{ H \subseteq E \,\middle|\, H \text{ er et 1-træ} \right\}$$

Kontrollerer betingelser (minimeringsproblem)

(i) $S \subseteq T$

(ii) $g(x) \leq f(x)$ for alle $x \in S$

For enhver Hamilton-kreds H gælder

$$\begin{aligned} g(x) &= \sum_{(i,j)\in H} (d_{ij} + (v_i - 2) + (v_j - 2)) \\ &= \sum_{(i,j)\in H} d_{ij} + \sum_{(i,j)\in H} (v_i - 2) + \sum_{(i,j)\in H} (v_j - 2) \\ &= \sum_{(i,j)\in H} d_{ij} + 2\left(\sum_{i\in V} v_i - 2|V|\right) = \sum_{(i,j)\in H} d_{ij} = f(x) \end{aligned}$$

Dominans

$$\mathcal{L}^3_{\text{TSP}} = \max_{i=1,\ldots,k} \mathcal{L}^{2\,(i)}_{\text{TSP}}$$

bedste grænseværdi med k iterationer

$\mathcal{L}^3_{\text{TSP}}$ dominerer $\mathcal{L}^1_{\text{TSP}}$ da

- $\boxed{\mathcal{L}^3_{\text{TSP}} \geq \mathcal{L}^2_{\text{TSP}} \text{ for alle instanser}}$

  første iteration foregår med originale afstandsmatrix

- $\boxed{\mathcal{L}^3_{\text{TSP}} > \mathcal{L}^2_{\text{TSP}} \text{ for en instanser}}$

  eksemplet

Kritiske og Semikritiske delproblemer

optimeringsproblem

$$z = \max\{ f(x) \mid x \in S\}$$

forgreningsregel som opdeler S i $S_1, \ldots, S_m$
øvre grænseværdi-funktion $\mathcal{U}(S_i)$

Definition 3 *Et delproblem S_i er* kritisk $\Leftrightarrow$

$$\mathcal{U}(S_i) > z$$

alle kritiske delproblemer skal behandles

Definition 4 *Et delproblem S_i er* semikritisk $\Leftrightarrow$

$$\mathcal{U}(S_i) \geq z$$

Sætning 1 *Hvis z er kendt fra starten, vil enhver branch-and-bound algoritme kun gennemsøge de kritiske delproblemer (svarende til den valgte forgreningsregel og grænseværdi-funktion).*

Kritiske og Semikritiske delproblemer

Sætning 2 *Uanset den initielle nedre grænseværdi vil bedste-først søgestrategien kun behandle de semikritiske delproblemer*

Indirekte bevis

- Antag bedste-først behandler S_i hvor $\mathcal{U}(S_i) < z$
- $\mathcal{L} < z$ på givne tidspunkt
- Bedste-først søgning: så for alle $S_j \in L$
$$\mathcal{U}(S_i) \geq \mathcal{U}(S_j)$$
- for alle $S_j \in L$
$$\mathcal{U}(S_j) \leq \mathcal{U}(S_i) < z$$
- z kan ikke findes i nogen af delløsningsrummene

Dvs: Bedste-først søgning sikrer at z bliver fundet inden man betragter delproblemer S_i med $\mathcal{U}(S_i) < z$.