

Metoda Greedy

Prezentare [1]

Metoda Greedy (*greedy*=lacom) este aplicabilă problemelor de optim.

Un tipar posibil de probleme la care se aplică metoda Greedy este următorul:

*Se dă o mulțime finită A . Să se determine o submulțime finită $B \subseteq A$ care satisface o anumită proprietate propr (este **soluție posibilă**) și îndeplinește un criteriu de optim (este **soluție optimă**), adică minimizează/maximizează o **funcție obiectiv f***

Spre exemplu, să considerăm următoarea problemă: Dată o mulțime de intervale, să se determine o submulțime de cardinal maxim de intervale din mulțimea dată care nu se intersectează. Pentru această problemă A = mulțimea de intervale dată, o soluție $S \subseteq A$ este o submulțime de intervale care trebuie să verifice proprietatea că sunt disjuncte două câte două, iar funcția obiectiv care trebuie maximizată este $f(S) = |S|$.

Formalizând, un cadru posibil pentru aplicarea metodei Greedy este următorul: Considerăm mulțimea finită $A = \{a_1, \dots, a_n\}$ și o proprietate *propr* definită pe mulțimea submulțimilor lui A (notată $P(A)$):

$$\text{propr}: P(A) \rightarrow \{\text{True}, \text{False}\} \text{ cu } \begin{cases} \text{propr}(\emptyset) = \text{True} \\ \text{propr}(X) \Rightarrow \text{propr}(Y), \forall Y \subset X \end{cases}$$

O submulțime $S \subset A$ se numește **soluție** (*soluție posibilă*) dacă $\text{propr}(S) = 1$.

Dintre soluțiile posibile dorim să alegem una care optimizează (minimizează/maximizează) o funcție $f: P(A) \rightarrow \mathbb{R}$ dată.

Metoda Greedy urmărește evitarea parcurgerii tuturor submulțimilor (ceea ce ar necesita un timp de calcul exponențial), **mergându-se "direct" spre soluția optimă**. Astfel, soluția se construiește element cu element, elementul ales la un pas pentru a se adăuga în soluție fiind cel care pare cel mai "bun" la acel pas, conform criteriului de optim f . Nu este însă garantată obținerea unei soluții optime, de aceea aplicarea metodei Greedy trebuie însoțită neapărat de o demonstrație.

Distingem două variante generale de aplicare a metodei Greedy: cu prelucrare inițială a elementelor din A (de obicei ordonare) sau fără:

<pre> S ← ∅ for i=1,n x ← alege(A) A ← A \ {x} if propr(S ∪ {x}) = True S ← S ∪ {x} </pre>	<pre> prelucreaza(A) S ← ∅ for x in A if propr(S ∪ {x}) = True S ← S ∪ {x} return S </pre>
--	--

Observații:

- în algoritm nu apare funcția f
- dificultatea constă în a concepe procedurile *alege*, respectiv *prel*, în care este "ascunsă" funcția f .

Exemplul 1 Se consideră mulțimea de valori reale $A = \{a_1, \dots, a_n\}$. Să se determine o submulțime a lui A a cărei sumă a elementelor este maximă.

Soluție. La fiecare pas trebuie să adăugăm la submulțime un element care să crească suma elementelor din submulțime cu o valoare cât mai mare. Astfel, vom adăuga inițial la submulțime elementul cel mai mare din A . Apoi vom parcurge mulțimea și vom selecta numai elementele pozitive, dacă astfel de elemente există (dacă nu există elemente pozitive, soluția este dată de cel mai mare element din mulțime). Corectitudinea soluției este evidentă.

Exemplul 2 Se consideră mulțimea de valori reale $A = \{a_1, \dots, a_n\}$ și $k < n$. Să se determine o submulțime a lui A de cardinal k a cărei sumă a elementelor este maximă.

Soluție. Vom alege în S cele mai mari k elemente.

Exemplul 3 (Contraexemplu) Se consideră mulțimea $A = \{a_1, \dots, a_n\}$ cu elemente pozitive. Să se determine o submulțime a lui A de sumă maximă, dar cel mult egală cu o valoare M dată.

Dacă alegem la fiecare pas cel mai mare element care poate fi adăugat la soluție fără a depăși M , pentru $A = \{6, 3, 4, 2\}$ și $M = 7$ obținem $\{6\}$. Dar soluția optimă este $\{3, 4\}$ cu suma egală cu 7.

Continuăm cu prezentarea unor exemple clasice.

• Problema selectării activităților (problema spectacolelor) [2]

Să presupunem că avem o mulțime $A = \{1, 2, \dots, n\}$ de n activități (spectacole) care doresc să folosească aceeași resursă (sala de spectacole). Această resursă poate fi folosită de o singură activitate la un moment dat. Fiecare activitate i are un timp de start s_i și un timp de terminare t_i , deci se poate desfășura doar în intervalul $[s_i, t_i)$. Problema selectării activităților (spectacolelor) constă în selectarea unei mulțimi de activități compatibile între ele de cardinal maxim (compatibile = având intervalele de desfășurare disjuncte).

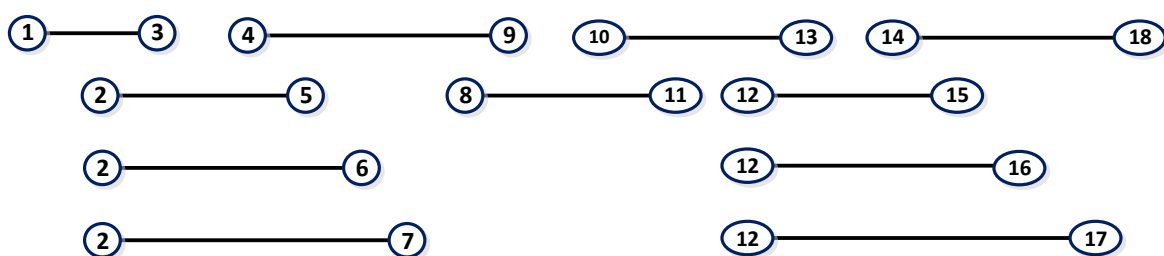
Variantă de enunț: Știind intervalele de desfășurare a unor conferințe la care dorim să asistăm în timpul unei zile (sau emisiuni pe care dorim să le vizionăm), să se determine numărul maxim de conferințe la care putem participa (= cu intervale de desfășurare disjuncte).

Variantă de enunț cu intervale: Se consideră o mulțime A de intervale închise. Să se determine o submulțime de cardinal maxim de intervale disjuncte din A .

Soluție. După ce criteriu alegem prima activitate, sau, mai general, activitatea planificată la pasul curent?

Posibile criterii „greedy” de alegere:

- activitatea cu durata cea mai mică, pentru a ocupa cât mai puțin resursa – nu este corect, deoarece acest criteriu nu ține cont de intervalul în care se desfășoară activitatea, care se poate intersecta cu intervalele altor activități. Spre exemplu, pentru activitățile cu intervalele $[1,10)$, $[9,12)$, $[13,18)$ soluția optimă este formată cu activitățile 1 și 3 (nu conține activitatea 2 cu lungimea cea mai mică)
- activitatea care începe cel mai devreme, pentru a ocupa cât mai repede resursa – nu este corect, această activitate poate avea durată mare: contraexemplu: $[1,10)$, $[2,4)$, $[5,8)$
- activitatea care este compatibilă cu cât mai multe activități (sau, echivalent, se intersectează cu cât mai puține intervalele de desfășurare ale celorlalte activități), pentru ca prin programarea ei să fie excluse din lista activităților care se mai pot programa cât mai puține activități – nici acest criteriu nu este corect, pentru că o mare parte dintre activitățile cu care este compatibilă pot să nu fie compatibile două câte două, ca în exemplul din desenul de mai jos (unde sunt indicate intervalele de desfășurare ale activităților). Intervalul $[8,11)$ se intersectează doar cu două intervale, dar soluția optimă este formată cu activitățile cu intervalele $[1,3)$, $[4,9)$, $[10,13)$, $[14,18)$



Strategia Greedy (corectă) pentru rezolvarea acestei probleme este următoarea: la fiecare pas este selectată o activitate nouă care se termină cel mai **devreme** (cu timpul de terminare cel mai mic) compatibilă cu activitățile deja selectate, pentru a lăsa după terminarea ei

resursa liberă cât mai mult timp. Pentru aceasta este util să procesăm inițial datele, mai exact să ordonăm activitățile crescător după timpul de terminare.

Pentru a simplifica scrierea pseudocodului și justificarea corectitudinii, să presupunem că activitățile sunt ordonate crescător după timpul de terminare: $t_1 \leq t_2 \leq \dots \leq t_n$

Algoritmul Greedy este descris de următoarea funcție, scrisă în pseudocod.

```
function selectie_activitati(s, t)
    S ← {1}
    us ← 1 //indicele ultimei activitati selectate
    pentru ac = 2,n executa //activitatea curenta
        daca  $s_{ac} \geq t_{us}$  atunci //  $t_{us} = \max\{t_k, k \in S\}$ 
            S ← S ∪ {ac}
            us ← ac
    return S
```

În mulțimea S se introduc activitățile care au fost selectate. Variabila us indică ultima activitate introdusă în S.

Deoarece activitățile sunt considerate în ordinea crescătoare a timpilor lor de terminare, t_{us} va reprezenta timpul maxim de terminare a oricărei activități din S:

$$t_{us} = \max\{t_k, k \in S\}$$

De aceea, pentru a vedea dacă activitatea curentă a_c este compatibilă cu toate celelalte activități existente la momentul curent în S este suficient ca $s_{ac} \geq t_{us}$.

Corectitudine - Suplimentar. Presupunem (ca și în algoritm) că activitățile sunt ordonate crescător după timpul de terminare: $t_1 \leq t_2 \leq \dots \leq t_n$. Fie $S = \{g_1=1, \dots, g_m\}$ (cu $g_1 < \dots < g_m$) soluția Greedy.

Varianta 1 [2]:

Vom folosi o strategie întâlnită în multe demonstrații de corectitudine pentru algoritmi de tip Greedy, similară celei din problema anterioară. Pentru a demonstra că soluția Greedy S este optimă considerăm o soluție optimă O care are un număr maxim de elemente inițiale în comun cu Greedy (o soluție optimă care “seamănă” cel mai mult cu soluția Greedy). Evidențiem prima diferență dintre O și S și arătăm că eliminând această diferență (înlocuind elementul din O care nu este în soluția Greedy cu unul din S – exchange argument) obținem o soluție cel puțin la fel de bună ca O, care are însă mai multe elemente în comun cu soluția greedy S (deci **O se poate transforma element cu element printr-o strategie de tip “exchange argument” în S**)

Considerăm o soluție optimă $O = \{o_1, \dots, o_p\}$ (cu activitățile notate astfel încât $o_1 < \dots < o_p$, deci $t_{o_1} \leq \dots \leq t_{o_p}$) care are **un număr maxim de elemente (inițiale) în comun cu soluția Greedy S**. Avem $m \leq p$. Presupunem prin absurd $|O| > |S|$.

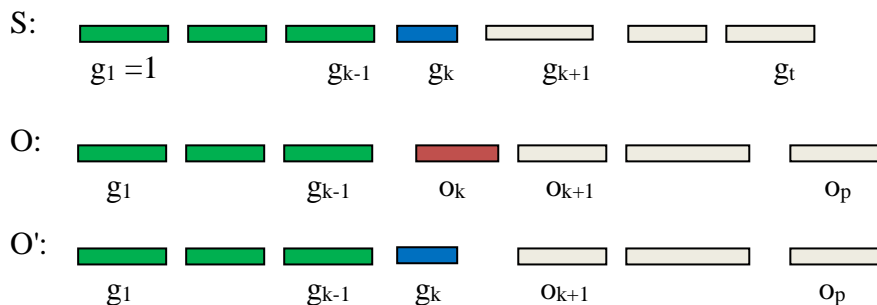
Atunci există un indice $k \leq m$ astfel încât $g_k \neq o_k$ (nu putem avea $S \subset O$ și $m < p$ deoarece activitatea o_{m+1} este compatibilă cu toate activitățile din $\{o_1, \dots, o_m\} = \{g_1, \dots, g_m\}$ și algoritmul Greedy `selectie_activitati` ar mai fi avut activități compatibile cu $\{g_1, \dots, g_m\}$ din care să selecteze, deci ar fi furnizat o soluție cu mai multe elemente).

Fie $k \leq m$ cel mai mic indice astfel încât $g_k \neq o_k$ (prima poziție pe care soluția greedy S și soluția optimă O diferă). La pasul la care `selectie_activitati` a ales activitatea g_k , activitatea o_k era de asemenea neselectată și compatibilă cu activitățile deja selectate $g_1 = o_1, \dots, g_{k-1} = o_{k-1}$. Deoarece activitatea g_k a fost cea selectată, rezultă că $t_{g_k} \leq t_{o_k}$ (g_k a fost selectată deoarece avea

timpul de terminare mai mic), deci g_k este compatibilă cu o_{k+1}, \dots, o_m (deoarece acestea încep după ce se termină o_k) – v. desenul de mai jos:

$$t_{g_k} \leq t_{o_k} \leq s_{o_{k+1}} \leq t_{o_{k+1}} \leq \dots \leq s_{o_p} \leq t_{o_p}$$

Atunci putem înlocui în soluția optimă O activitatea o_k cu g_k și obținem tot o soluție posibilă $O' = O - \{o_k\} \cup \{g_k\}$ (formată din activități compatibile din A). Avem $|O'| = |O|$, deci O' este tot o soluție optimă, dar care are mai multe elemente în comun cu soluția greedy S , contradicție.



Varianta 2 (similar) – Demonstrăm prin inducție după n că algoritmul greedy `selectie_activitati` construiește o soluție optimă.

Pentru $n = 0, 1$ afirmația este evidentă.

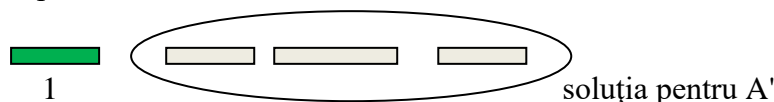
Fie $n \geq 2$. Presupunem că algoritmul `selectie_activitati` construiește o soluție optimă pentru orice mulțime de activități A' cu $|A'| < n$

Fie A o mulțime de n activități.

- a) **Există o soluție optimă care conține prima activitate selectată de algoritmul greedy, anume $g_1 = 1$**

Într-adevăr, fie $O = \{o_1, \dots, o_p\}$ o soluție optimă (o submulțime de cardinal maxim de activități compatibile) cu activitățile notate astfel încât $o_1 < \dots < o_p$, deci $t_{o_1} \leq \dots \leq t_{o_p}$. Dacă activitatea 1 aparține lui O , atunci afirmația a) este adevărată. Altfel considerăm $O' = O - \{o_1\} \cup \{1\}$. Deoarece $t_1 \leq t_{o_1}$, activitatea 1 este compatibilă cu toate celelalte activități din $O - \{o_1\}$, deci O' este o submulțime de activități compatibile cu $|O'| = |O|$, adică o soluție optimă care conține activitatea 1.

- b) Fie $A' = \{k \in S, s_k \geq t_1\}$ mulțimea activităților care încep după ce se termină activitatea 1 (compatibile cu activitatea 1). Conform ipotezei de inducție soluția S' construită de algoritmul `selectie_activitati` pentru A' este soluție optimă. Rezultă că $S = S' \cup \{1\}$ este soluție optimă pentru A (altfel, conform punctului a) ar exista o soluție optimă O pentru A care conține activitatea 1 cu cardinal mai mare decât S : $|O| > |S|$; dar atunci $O - \{1\}$ este soluție posibilă pentru A' cu $|O - \{1\}| > |S - \{1\}| = |S'|$, ceea ce contrazice optimalitatea lui S')

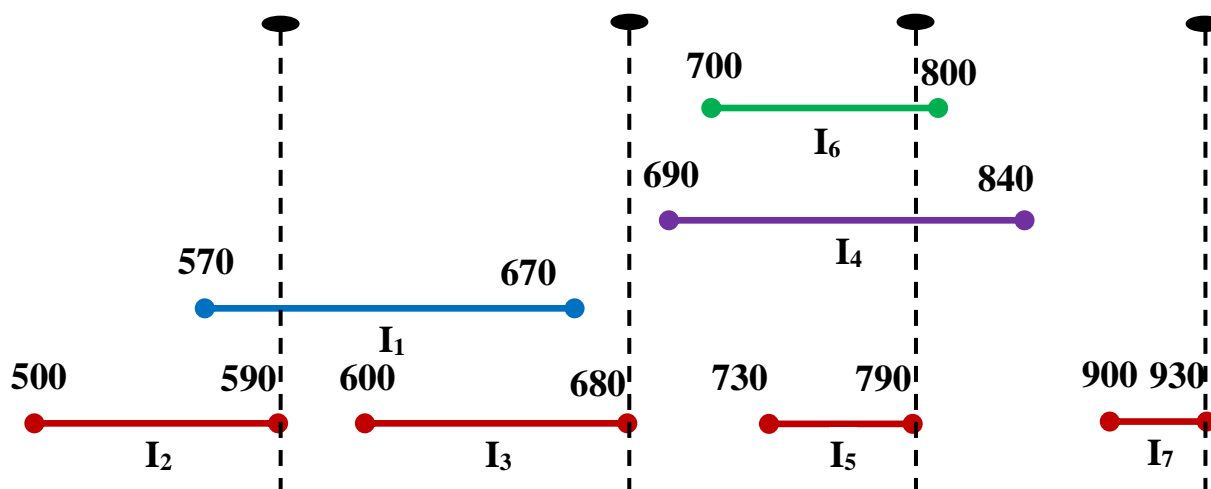


• Problema mulțimii de acoperire

Fie n intervale închise $I_1 = [a_1, b_1], \dots, I_n = [a_n, b_n]$. Să se determine o mulțime M cu număr minim de elemente astfel încât $\forall k \in \overline{1, n}, \exists x \in M$ astfel încât $x \in I_k = [a_k, b_k]$. Mulțimea M se numește *mulțime de acoperire* a șirului de intervale respectiv.

Problema mai este cunoscută și sub numele de *problema cuielor*: "Se consideră n scânduri, fiecare fiind dată printr-un interval închis de pe axa reală. Să se determine numărul minim de cui pe care trebuie să le batem astfel încât în fiecare scândură să fie bătut cel puțin un cui. Se consideră faptul că orice cui are o lungime suficient de mare pentru a trece prin oricâte scânduri este nevoie.".

intervale.txt	acoperire.txt
570 670	590
500 590	680
600 680	790
690 840	930
730 790	
700 800	
900 930	



Parcurgând scândurile de la stânga spre dreapta, în ordinea crescătoare a extremităților din dreapta, se observă faptul că primul cui trebuie să fie bătut în extremitatea dreaptă a primei scânduri (i.e., scândura cu extremitatea dreaptă minimă). Pentru fiecare dintre scândurile rămase verificăm dacă există deja un cui bătut în ea (i.e., extremitatea sa stângă este strict mai mică decât poziția ultimului cui bătut), iar în cazul în care nu există, batem cuiul în extremitatea sa dreaptă.

Problema se rezolvă într-un mod foarte asemănător cu problema spectacolelor, deoarece cardinalul minim al unei mulțimi de acoperire este mai mare sau egal cu cardinalul maxim al unei mulțimi de intervale disjuncte, iar prin strategia de mai sus obținem egalitate (adăugăm în acoperire extremitățile din dreapta ale intervalelor din soluția de la problema spectacolelor)

• Problema continuă a rucsacului [1]

Se consideră un rucsac de capacitate (greutate) maximă G și n obiecte caracterizate prin:

- greutatea lor g_1, \dots, g_n ;
- câștigurile c_1, \dots, c_n obținute la încărcarea lor în totalitate în rucsac.

Din fiecare obiect poate fi încărcată orice fracțiune a sa (și valoarea asociată va fi aceea fracțiune din câștigul total). Să se determine o modalitate de încărcare de (fracțiuni de) obiecte în rucsac, astfel încât câștigul total să fie maxim.

De exemplu, pentru $G = 8$, $n = 4$, obiectul 1 cu greutatea 4 și câștigul 8, obiectul 2 cu greutatea 8 și câștigul 12, obiectul 3 cu greutatea 3 și câștigul 6 și obiectul 4 cu greutatea 10 și câștigul 10 o soluție optimă este dată de vectorul de fracțiuni $(1, 1/8, 1, 0)$, adică se vor încărca obiectele 1 și 3 întregi și $1/8$ din obiectul 2, iar câștigul total va fi $8 + 1/8 \cdot 12 + 6 = 15,5$

Variantă de enunț: T = timp total de funcționare a unei resurse, n activități cu durată d_i și profitul p_i . O activitate începută poate fi întreruptă obținându-se un profit parțial.

Soluție [1]. Prin *soluție* înțelegem un vector $x = (x_1, \dots, x_n)$ cu

$$\begin{cases} x_i \in [0, 1], \forall i \\ \sum_{i=1}^n g_i x_i \leq G \end{cases}$$

O *soluție optimă* este soluție care maximizează funcția $f(x) = \sum_{i=1}^n c_i x_i$.

Dacă suma greutateilor obiectelor este mai mică decât G , atunci încărcăm toate obiectele: $x = (1, \dots, 1)$. De aceea **presupunem în continuare că $g_1 + \dots + g_n > G$.**

Exemplu

Cum alegem obiectul care trebuie încărcat la pasul curent? Alegem dintre obiectele rămase pe cel mai valoros **pe unitatea de greutate** (cu valoarea cea mai mare a raportului c/g) și încărcăm cât putem de mult din el.

Astfel, pentru a implementa strategia Greedy, ar fi util să ordonăm inițial obiectele descrescător după câștigul la unitatea de greutate. Algoritmul constă în încărcarea în această ordine a obiectelor, atâta timp cât nu se depășește greutatea G , ultimul obiect adăugat putând fi eventual încărcat parțial. Mai exact, dacă presupunem obiectele ordonate astfel încât

$$\frac{c_1}{g_1} \geq \frac{c_2}{g_2} \geq \dots \geq \frac{c_n}{g_n}$$

algoritmul este următorul:

```
G1 ← G { G1 reprezintă greutatea disponibilă }
for i=1, n
  if  $g_i \leq G1$ 
     $x_i \leftarrow 1$ ;  $G1 \leftarrow G1 - g_i$ 
  else
     $x_i \leftarrow G1/g_i$ ;
    for  $j=i+1, n$ 
       $x_j \leftarrow 0$ 
    stop
write(x)
```

Am obținut deci soluția greedy $x = (1, \dots, 1, x_j, 0, \dots, 0)$ cu $x_j \in [0, 1)$.

Corectitudine: Intuitiv, să presupunem că există o soluție optimă cu câștigul mai mare decât soluția Greedy. Atunci în această soluție există două obiecte i și j cu $c_i/g_i > c_j/g_j$ astfel încât soluția optimă a ales o fracțiune mai mică decât soluția greedy din obiectul i și mai mare din obiectul j . Crescând fracțiunea aleasă din obiectul i de soluția optimă și scăzând fracțiunea din obiectul j astfel încât greutatea totală din rucsac să rămână G , obținem o soluție și mai bună decât soluția optimă, contradicție.

Detaliat- SUPLIMENTAR [1]. –Arătăm că soluția astfel obținută este optimă.

Fie y soluția optimă care are o subsecvență inițială maximă în comun cu x :

$$y = (\dots, y_k, \dots) \text{ cu } \begin{cases} \sum_{i=1}^n g_i y_i = G \\ \sum_{i=1}^n c_i y_i \text{ maxim} \end{cases}$$

Dacă $y \neq x$, fie k **prima poziție** pe care $y_k \neq x_k$. Atunci avem

- $k \leq j$: pentru $k > j$ se depășește G .
- $y_k < x_k$

Considerăm o soluție y' cu **mai multe elemente inițiale în comun** cu soluția Greedy x , obținută din y înlocuind y_k cu x_k și păstrând o fracțiune mai mică din obiectele $k+1, \dots, n$, astfel încât să obținem tot greutatea G : $y' = (y_1, \dots, y_{k-1}, x_k, \alpha y_{k+1}, \dots, \alpha y_n)$ cu $\alpha < 1$, astfel încât:

$$g_k(x_k - y_k) = (1 - \alpha)(g_{k+1}y_{k+1} + \dots + g_n y_n) \quad (*)$$

Comparăm performanța lui y' cu cea a lui y pentru a arăta că soluția y' este cel puțin la fel de bună ca y , dar cu mai multe elemente în comun cu soluția greedy x , ceea ce contrazice alegerea lui y . Avem

$$\begin{aligned} f(y') - f(y) &= c_k x_k + \alpha c_{k+1} y_{k+1} + \dots + \alpha c_n y_n - (c_k y_k + c_{k+1} y_{k+1} + \dots + c_n y_n) = \\ &= c_k(x_k - y_k) - (1 - \alpha)(c_{k+1} y_{k+1} + \dots + c_n y_n) = \\ &= c_k/g_k \cdot g_k(x_k - y_k) + (\alpha - 1)(c_{k+1}/g_{k+1} \cdot g_{k+1} y_{k+1} + \dots + c_n/g_n \cdot g_n y_n) = \end{aligned}$$

Dar $1 - \alpha > 0$ și $c_k/g_k \geq c_s/g_s, \forall s > k$. Atunci

$$f(y') - f(y) \geq c_k/g_k [g_k(x_k - y_k) + (\alpha - 1)(c_{k+1} y_{k+1} + \dots + c_n y_n)] = 0 \quad (\text{din } (*)),$$

deci $f(y') \geq f(y)$. Rezultă că și y' este soluție optimă ($f(y)$ este maxim), contradicție cu alegerea lui y , deoarece y' are mai multe elemente inițiale în comun cu x decât y .

Variante. Generalizări

Problema discretă a rucsacului diferă de cea continuă prin faptul că fiecare obiect poate fi încărcat numai în întregime în rucsac.

Să observăm că aplicarea metodei Greedy eșuează în acest caz. Într-adevăr, aplicarea ei pentru: $G=5, n=3$ și $g=(4, 3, 2), c=(6, 4, 2.5)$ are ca rezultat încărcarea primului obiect; câștigul obținut este 6. Dar încărcarea ultimelor două obiecte conduce la un câștig mai mare 6.5.

Bibliografie

1. Horia Georgescu. **Tehnici de programare**. Editura Universității din București 2005
2. **Jon Kleinberg, Éva Tardos** - Algorithm Design, 2005 Addison-Wesley Professional

Slideu-rile oficiale pentru această carte se pot găsi la

<http://www.cs.princeton.edu/~wayne/kleinberg-tardos/pearson/>

o versiune îmbogățită a acestora este accesibilă la adresa

<http://www.cs.princeton.edu/~wayne/kleinberg-tardos>