Polskie Towarzystwo Badań Operacyjnych i Systemowych

Instytut Badań Systemowych Polskiej Akademii Nauk

Wojskowa Akademia Techniczna

Redaktorzy:

Zbigniew Nahorski

Marian Chudy

Andrzej Straszak

# OPTYMALIZACJA


Warszawa 1991

POLSKIE TOWARZYSTWO
BADAŃ OPERACYJNYCH I SYSTEMOWYCH

INSTYTUT BADAŃ SYSTEMOWYCH
POLSKIEJ AKADEMII NAUK

WOJSKOWA AKADEMIA TECHNICZNA

# O P T Y M A L I Z A C J A

## Z A D A N I A, M E T O D Y, A L G O R Y T M Y

**Redaktorzy**

***Zbigniew Nahorski, Marian Chudy, Andrzej Straszak***

**WARSZAWA 1991**

OPTYMALIZACJA:
Zadania, metody, algorytmy
PTBOiS, Warszawa 1991.

# ALGORYTMY RELAKSACYJNE I SCHEMAT OPTYMALIZACYJNY DLA WIELOMASZYNOWEGO ZAGADNIENIA PRZYDZIAŁU PRAC

Henryk Potrzebowski
Instytut Badań Systemowych PAN
ul. Newelska 6, 01-447 Warszawa

*Streszczenie*: W pracy omówiono metody rozwiązywania ogólnego zagadnienia przydziału prac niepodzielnych rozpoczynając od wskazania na 1-drzewo reprezentację bazy i schemat minimalizacji czasu obsługi prac sygnalizowany w [3]. Pokazano algorytm eliminacji zmiennych i sposób generowania odcięć w bazie. Dokonano krytycznej oceny metody relaksacji Lagrangea oraz innych podejść.

## 1. Wprowadzenie

Rozważane są własności i metody obliczeniowe znanego z badań operacyjnych uogólnionego zagadnienia przydziału prac, minimalizującego czas lub koszt (-dochód) wnikający ze sposobu przydziału tych prac, być może przy pewnych ograniczeniach komplikujących. Z zagadnieniem takim zwykle spotykamy się w szeregowaniu równoległym (współbieżnym) zadań produkcyjnych, w rozstrzyganiu pewnych problemów pakowania, kredytowania itp.

Rozważania rozpoczynamy od sformułowania ogólnego zagadnienia przydziału prac podzielnych. Wykonanie j-tej pracy na maszynie i-tej pomniejsza jej zasób czasu pracy $b_i$ o wielkość $a_{ij}$ tworząc ewentualny dochód ($c_{ij} > 0$) lub koszt ($c_{ij} < 0$). Jeżeli z wykonaniem części pracy j na maszynie i zwiążemy zmienną $x_{ij} \geq 0$, to rozważane zagadnienie polegało będzie na znalezieniu ekstremalnej wartości liniowej funkcji dochodu, kosztu, lub też funkcji czasu na zbiorze przydziałów dopuszczalnych

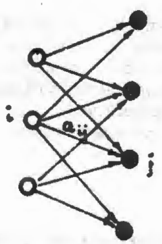
$$P = \{ x: \sum_j a_{ij}x_{ij} \leq b_i, \ i=1,\dots,m, \tag{1}$$

$$\sum_i x_{ij} = 1, \quad j=1,\dots,n, \tag{2}$$

$$x_{ij} \geq 0 \quad \text{dla każdego } i,j \ \} \tag{3}$$

gdzie m - liczba maszyn, n - liczba prac.

Zagadnienie jest równoważne wyznaczeniu ekstremalnego strumienia na dwudzielnym grafie $G=(I,J,E)$ ze zbiorami wierzchołków $I = \{1,\dots,m\}$, $J = \{1,\dots,n\}$ i zbiorem łuków $E =\{(i,j)$: dopuszcza się przydział $j$ na $i\}$. Przykład takiego grafu pokazano na rys. 1. Dla tego grafu suma strumieni związanych z wierzchołkiem $i$, pomnożonych przez mnożniki $a_{ij}$ nie może przekraczać $b_i$; suma strumieni związanych z wierzchołkiem $j$ równa jest 1. Istnieje wzajemnie jednoznaczna odpowiedniość pomiędzy kolumnami macierzy ograniczeń układu (1)-(3), a łukami grafu G. Wyznaczanie rozwiązania dopuszczalnego układu (1)-(3) odpowiada wyznaczaniu strumienia dopuszczalnego $x$ na grafie G przy uwzględnieniu współczynników $a_{ij}$.



Rys.1. Graf G(I,J,E)

Znane z programowania liniowego wyniki teoretyczne ( Chaczijan, Karmarkar ) klasyfikują ogólne zagadnienie przydziału podzielnego prac do klasy P zagadnień o złożoności wielomianowej. Praktyka jednak uprzywiliowuje teoretycznie gorsze, bo w ogólności nie gwarantujące wielomianowej zbieżności, algorytmy, będące sieciową specjalizacją metody prymalnej symplex. Algorytmy takie omawiane są w pracach [1,4,5].

Wprowadzenie warunku niepodzielności prac, tj. warunku

$$x_{ij} \text{ - całkowitoliczbowe dla każdych } i,j \tag{4}$$

znacznie komplikuje rozważane zagadnienie. Już w przypadku $m = 2$ zagadnienie (1)-(4) równoważne jest NP-trudnemu zagadnieniu rozbicia Knapsack, co oznacza, że ogólne zagadnienie przydziału niepodzielnych prac jest zagadnieniem NP-trudnym. W szczególnym przypadku $m = n$, $b_i = 1\ \forall i$, $a_{ij} = 1\ \forall i,j$ dla kryteriów $\sum_{ij} c_{ij} x_{ij}$ i $\mathrm{Max}_{ij}\{ c_{ij} x_{ij}\}$ znane są algorytmy wielomianowe.

**2. Szczególna postać bazy**

Zauważmy, że zbiór P rozwiązań ułamkowych jest wielościanem wypukłym, zawartym w jednostkowej kostce. Przyjmijmy, że zagadnienie przydziału ułamkowego jest niesprzeczne, tj. $P \neq \emptyset$, wtedy

**Twierdzenie 1.** Wierzchołek (baza dopuszczalna) $x \in P$ posiada w grafie G reprezentację 1-drzewa.

**D.** Niech $E(x) = \{(i,j): x > 0\}$. Każdy niezdegenerowany wierzchołek $x \in P$ wyznacza jednoznacznie mn ograniczeń jakie muszą przejść w równości spośród ogólnej liczby mn+m+n ograniczeń (1)-(3). Oczywista, $E(x) \le E$ i $m \le |E(x)| \le m+n$ skoro (2)-ograniczenia równościowe. Ponieważ drzewo spinające wszystkie wierzchołki grafu G posiada m+n-1 krawędzi, w przypadku $|E(x)|=m+n$ zbiór E(x) wyznacza w grafie G 1-drzewo, tj. drzewo spinające z dodatkowym łukiem. W przypadku $|E(x)| < m+n$ w ogólności baza będzie reprezentowana przez las, którego składowymi będą 1-drzewo, drzewo i izolowane wierzchołki $i \in I$ ▪

Wymieniona własność, właściwa różnym problemom sieciowym, pozwoliła zbudować wysoce efektywne specjalizacje metody prymalnej symplex, znane z pionierskich prac [1,4], sygnalizowane w [5]. Dwa kolejne stwierdzenia wnoszą wiele do metod wyznaczania przydziałów całkowitoliczbowych.

**Twierdzenie 2.** Dla bazy $x \in P$ istnieje 1-1 skojarzenie prac podzielonych z maszynami.

**D.** Niech j odpowiada pracy podzielonej. Jeżeli jest on wierzchołkiem 1-drzewa, który nie należy do jego cyklu, to można go skojarzyć z dowolnym i, jego bezpośrednim następnikiem. Takie i zawsze istnieje, ponieważ z j podzielonym są incydentne co najmniej dwa łuki . Jeżeli j należy do cyklu, kojarzymy go z i, który jest bezpośrednim następnikiem w cyklu (przy ustalonej, arbitralnej orientacji cyklu). Ostatecznie, wierzchołki j o stopniu 1, tj. prace niepodzielone wraz z incydentnymi łukami oraz łuki incydentne tylko z jednym wierzchołkiem pomijamy ▪

Bezpośrednio z twierdzenia wynika, że:

**Wniosek 1.** Dla ułamkowej bazy $x \in P$ istnieje "zaokrąglony" przydział całkowitoliczbowy wynikający z podniesienia zmiennych ułamkowych $x_{ij}$ do 1 dla wszystkich łuków $(i,j)$, które weszły w skojarzenie. Jeżeli $b$ jest wektorem zasobów czasu, to długość zaokrąglonego uszeregowania nie przekracza $\mathrm{Max}_i\{b_i\} + \tau$, gdzie $\tau$ jest masymalną wartością $a_{ij}$ dla łuków $(i,j)$ wchodzących w skojarzenie.

**Wniosek 2.** Liczba prac podzielonych nie przekracza liczby aktywnych w wierzchołku $x$ ograniczeń (1), tj. liczby tych $i$, dla których zasób $b_i$ jest wykorzystany w pełni.

Wniosek drugi potwierdza obserwowaną w tego typu zagadnieniach tezę o stosunkowo niewielkiej liczbie zmiennych ułamkowych w bazie i o możliwości oszczędnego pamięciowo opisu bazy. Wniosek pierwszy pozwala zbudować schemat optymalizacyjny dla zagadnienia minimalizacji czasu obsługi prac $C_{max}$. Dla danego $\varepsilon$, $\varepsilon > 0$, rozpatrzymy algorytm.

**Procedura S**

(i) użyj (dowolny) algorytm zachłanny dla wyznaczenia niepodzielnego przydziału startowego $x^{\circ}$. Niech $u^{\circ}$ będzie czasem obsługi zadań przy przydziale $x^{\circ}$. Rozpoczynamy od oszacowania z góry $u=u^{\circ}$ i oszacowania z dołu $l=\lfloor u/m \rfloor$.

(ii) utwórz zbiór $A_{\varepsilon}= \{j:p_{ij} > \varepsilon D\ \forall i\}$ prac "długich" dla $D = \lfloor (u+l)/2 \rfloor$. Dla (kolejnego) ustalonego przydziału prac "długich" wyznacz podzielny przydział $x \in P$ prac "krótkich" $j \notin A_{\varepsilon}$ przy $b_i= D\ \forall i$.

Jeżeli istnieje przydział prac "krótkich" to zgodnie z wnioskiem 1 zaokrąglamy go, otrzymując $C_{max} \le (1+\varepsilon)D$. Po przypisaniu $u:= C_{max}$ ($l$ pozostaje bez zmian) powtarzamy (ii). Jeżeli nie istnieje przydział prac "krótkich" i możliwości zmiany przydziału prac "długich" zostały wyczerpane, $l$ przypisujemy $D+1$ ($u$ pozostaje bez zmian) i powtarzamy (ii). Obliczenia kończymy dla $u = l$.

**Twierdzenie 3.** Dla całkowitoliczbowych $a_{ij}$, ustalonych $m$ i $\varepsilon$ procedura S jest $(1+\varepsilon)$-schematem optymalizacyjnym dla minimalno-czasowego zagadnienia przydziału prac przy $\varepsilon = 1$. D. Dla danego D liczba możliwych przydziałów prac z $A_\varepsilon$ nie przekracza $(n+1)^{m/\varepsilon}$, liczbę różnych D dla zadanych $a_{ij}$ ogranicza $\log u^{\circ}(1-1/m) <$ size A. Każdy podproblem zadań "krótkich" jest wielomianowy, ponieważ jest to problem liniowy. Wartość $\varepsilon = 1$ jest osiągalna: dla optymalnego $x^*$ istnieje takie i, że $a_{ij}x_{ij} \le C^*_{max}$ $\forall j$. Ponieważ oszacowanie liczby możliwych przydziałów prac z $A_\varepsilon$ jest wielomianem tylko od n a nie od $\varepsilon$, procedura S jest schematem a nie pełnym schematem ▪

Lenstra i inni w [3] zauważają np., że jeżeli $a_{ij} \in \{1,3\}$, $b_i = 3$ $\forall i$ lub $a_{ij} \in \{1,2,3\}$, $b_i = 2$ $\forall i$, minimalno-casowy problem przydziału niepodzielnego jest wielomianowo równoważny 3-wymiarowemu zagadnieniu skojarzenia. Ponieważ jest to problem NP-zupełny stąd wniosek, że odpowiednio dla $\varepsilon < 1/3$ przy $C_{max} \ge 3$ i dla $\varepsilon < 1/2$ przy $C_{max} \ge 2$ nie istnieje schemat optymalizacyjny, chyba że P=NP.

## 3. Algorytmy odcięć

Omówimy metody wyznaczania x całkowitoliczbowych polegające na dodaniu do ograniczeń (1)-(3) odcięć, tj. ograniczeń liniowych, w przypadku których pewne $x \in P$ są niedopuszczalne z wyjątkiem x całkowitoliczbowych. Dla ustalenia uwagi przyjmijmy, że P zawiera co najmniej jedno x całkowitoliczbowe. Oczywista metoda polegała by na wyznaczeniu wszystkich ścian powłoki wypukłej przydziałow całkowitoliczbowych i na rozwiązaniu problemu liniowego. Chociaż realizację metody dla zadań załadunku w znacznej mierze umożliwiają techniki generowania minimalnych odcięć, pominiemy ją, ponieważ (niezależnie od problemów numerycznych) liczba ścian takiej powłoki jest w ogólności funkcją wykładniczą liczby zmiennych. Za [8] zwrócimy uwagę na informacje, jakie wnosi tutaj baza rozwiązania liniowego.

**Definicja 1.** Dla bazy x wielościanu P zmienna $x_{ij}$ jest nadmiarowa, jeżeli przy ustalonych przydziałach całkowitoliczbowych jej podniesienie do 1 narusza $b_i$, nawet po przypisaniu pozostałym zmiennym ułamkowym wartości 0.

**Twierdzenie 4.** Jeżeli x jest bazą ułamkową P, to co najmniej jedna zmienna ułamkowa $x_{ij}$ jest nadmiarowa.

**D.** Niech j odpowiada pracy podzielonej. Przyjmijmy najpierw, że j jest wierzchołkiem 1-drzewa najbardziej oddalonym i nie należącym do cyklu. Dla tego j istnieje w bazie bezpośredni następnik i taki, że $x_{ij} < 1$. Inaczej j nie mogłoby być podzielone. Łatwo sprawdzić, że to $x_{ij}$ jest nadmiarowe. Dla bazy niezdegenerowanej, natomiast, jeżeli j jest elementem cyklu to istnieje możliwość wyboru kolejnych wierzchołków cyklu k,i,j takich, że przy $x_{ij}+ x_{ik} \leq 1$ $a_{ij} > a_{ik}$. W tym przypadku również zmienna $x_{ij}$ jest nadmiarowa ∎

Jeżeli pominąć w grafie G łuki (i,j) dla których $a_{ij} > b_i$, to algorytm wyznaczania przybliżonych rozwiązań całkowitoliczbowych $x \in P$ może przyjąć następującą postać.

**Procedura E** (eliminacji zmiennych)

(i) wyznacz bazę optymalną x zbioru P (faza liniowa)

(ii) dla danego x ustal wszystkie niepodzielnie przydzielone j, oblicz nowe $b_i$. W zredukowanym grafie G usuń łuki odpowiadające zmiennym nadmiarowym. Jeżeli pozostał w grafie G łuk, idź do (i).

Ponieważ po pierwszym (najbardziej pracochłonnym) użyciu procedury liniowej może pozostać podproblem m podzielonych prac, w każdej kolejnej iteracji można ustalić jedną pracę (inaczej nie wystąpią zmienne nadmiarowe ), najdalej w iteracji m+1 procedura zakończy pracę, eliminując co najwyżej m-1 prac (uwagi szczegółowe patrz [8]).

Procedura E jest przykładem iteracyjnej deterministycznej metody ustalania zmiennych poprzez przypisanie im ich dotychczasowej wartości dla zmiennych związanych z j niepodzielonym i wartości 0 dla zmiennych nadmiarowych. Jest także przykładem metody przybliżonej: ogólny problem całkowitoliczbowego przydziału prac, jak zauważyliśmy jest NP-trudny, zatem problem czy P zawiera całkowitoliczbowe x może być rozstrzygnięty tylko za pomocą procedury niedeterministycznej (niezależnie od siły ewentualnych oszacowań). Poniżej, wykorzystując znane w programowaniu całkowitoliczbowym pojęcie pokrycia minimalnego pokażemy jak będąc w bazie x zbioru P dołączyć niektóre ściany otoczki wypukłej punktów całkowitych by taki proces przyśpieszyć.

**Twierdzenie 5.** Niech x będzie bazą P, a $E_i$- zbiorem łuków (i,j) nadmiarowych i łuków, dla których $x_{ij} = 1$. Odcięciem prawidłowym jest ograniczenie liniowe

$$\sum_{(i,j)\in E_i} x_{ij} \le |J_i| \quad \text{gdzie } J_i = \{j:(i,j)\in E_i,\ x_{ij} = 1\} \qquad (5)$$

D. Ograniczenie (5) wiąże $|E_i| \ge |J_i| + 1$ prac. Jeżeli $x_{ij} = 1$ dla każdego $j \in J_i$, to nie jest możliwe by $x_{ij} > 0$ dla jakiegokolwiek $j \notin J_i$. Oznacza to, że (5) nie dopuszcza ułamkowego $x_{ij}$, chyba, że $x_{ik} = 0$ dla pewnego $k \in J_i$. Każde całkowitoliczbowe $x \in P$ nie narusza (5) ∎

**Twierdzenie 6.** Dla bazy x zbioru P niech $(E_i,J_i),(E_k,J_k)$ będą określonymi jak w tezie twierdzenia 5 zbiorami łuków i prac określonymi odpowiednio dla wierzchołków i,k dla których istnieje j nie będące elementem ani zbioru $J_i$ ani zbioru $J_k$ takie, że $(i,j) \in E_i$ ,$(k,j) \in E_k$. Niech r oznacza liczbą takich j, wtedy odcięciem będzie ograniczenie

$$\sum_{(i,j)\in E_i \cup E_k} x_{ij} \le |J_i| + |J_k| - r \qquad (6)$$

Ze względu na podobieństwo do twierdzenia 5 pominiemy jego dowód. Zauważmy tylko, że 1-drzewo reprezentacja bazy pozwala w ten sposób wyznaczyć bardziej złożone odcięcia.

Innego rodzaju odcięcie wynika z oszacowania dualnego dochodu cx w bazie $x \in P$. Przy danym oszacowaniu z góry u i wektorze cen dualnych v zmienna $x_{ij}$ jest nadmiarowa,jeżeli

$$cx + c_{ij} - v_i a_{ij} - v_j > u - 1. \qquad (7)$$

## 4. Algorytmy kombinowane

Atrakcyjne teoretycznie odcięcia omówione w p.3 posiadają tę wadę, że po dołączeniu do ograniczeń (1)-(3) zakłócają czytelną 1-drzewa reprezentację bazy wikłając w konsekwencji w poszukiwaniu rozwiązań kombinowanych. Z podobną sytuacją spotykamy się w przypadku, gdy w zagadnieniu wyjściowym zachodzi potrzeba uwzględnienia dodatkowych ograniczeń zasobowych lub dodatkowych zmiennych.

Pierwsze podejście w rozwiązywaniu uogólnionych zagadnień przydziału pracy polegało będzie na wykorzystaniu w naszych algorytmach specjalizowanych metod sympleks, o kombinowanej sieciowo - analitycznej reprezentacji bazy. Taki kierunek badań reprezentuje np. McBride w [4].

Drugie podejście polega na wykorzystaniu metody relaksacji Lagrangea i innych metod dualnych. Z przeprowadzonych w [9] rozważań nad całkowitoliczbowym zagadnieniem maksymalizacji dochodu cx, $x \in P$ wynikają co najmniej dwa istotne fakty:

1. Agregacja ograniczeń (1) w jedno ograniczenie zastępcze $uAx \leq ub$, $u \in R_+^m$ przy zachowaniu ograniczeń (2)-(4) prowadzi do rozwiązywania "$\min_u \max_x$" zagadnienia optymalizacji niegładkiej. Dla ustalonego u zagadnienie to jest równoważne jednowymiarowemu zagadnieniu załadunku z warunkami wyboru (2), dla którego, sądząc z [2], istnieją wysoce efektywne algorytmy. Otrzymane w ten sposób dla optymalnego u rozwiązanie trzeba będzie sprawdzić i ewentualnie skorygować ze względu na ograniczenia zasobowe (1).
2. Przeniesienie ograniczeń (2) w człon korygujący $v(\sum_j x_j - 1)$, $v \in R^n$ funkcji cx prowadzi do konkurencyjnego "$\min_v \max_x$" zagadnienia optymalizacji niegładkiej. Dla ustalonego v zagadnienie to jest równoważne wielowymiarowemu zagadnieniu załadunku. Zagadnienie takie było badane w [7], gdzie pokazano efektywne algorytmy przybliżone. Dla optymalnego v tak otrzymane rozwiązanie x trzeba będzie sprawdzić i ewentualnie skorygować ze względu na ograniczenia (2).

Wadą obydwu podejść jest konieczność rozwiązywania zadań optymalizacji niegładkiej dla stosunkowo dużej liczby zmiennych dualnych, przy jednoczesnej rezygnacji z wysoce efektywnych sieciowych metod symplex. O ile może to nie mieć większego znaczenia w przypadku zadań o mniejszych rozmiarach, w ogólności jest uzasadnione przenosić do funkcji Lagrangea w postaci członów korygujących tylko odcięcia i dodatkowe ograniczenia zasobowe.

Trzecie i ostatnie z sygnalizowanych tu podejść, wszechobecne zresztą w różnorodnych zagadnieniach kombinatorycznych, to algorytmy zachłanne i kombinatoryczne algorytmy poprawy rozwiązania. Algorytmy zachłanne pozwalają budować rozwiązania startowe wykorzystując w tym celu lokalne oceny oszczędności lub strat. Na celowość uwzględniania takich metod pokazują eksperymenty obliczeniowe znane np. z [8].

## Literatura

[1] Brown G.G., R.D.McBride: Solving Generalized Networks. ManagementSc. 20 (12) 1984, pp 1497-1523

[2] Dudziński K., S.Walukiewicz: A Fast Algorithm for the Linear Multiple-choice Knapsack Problem. O.R.Letters 4 (2) 1984 pp 205-209

[3] Lenstra J.K., D.B.Shmoys, E.Tardos: Approximation Algorithms for Scheduling Unrelated Parallel Machines. Math. Programming 46 1990 pp 259-271

[4] McBride R.D.: Solving Embidded Generalized Network Problems. E.J.of OR 21 1985 pp 82-92

[5] Nulty W.G., M.A.Trick.: GNO/PC Generalized Network Optimization System. OR Letters 7 (2) 1988

[6] Potrzebowski H.: Dokładne i przybliżone schematy obliczeniowe dla zagadnień szeregowania równoległego. Raport ZPM - 37/C-1/1990

[7] Sikorski J.: Algorytmy wyznaczania rozwiązań przybliżonych dla wielowymiarowych zadań załadunku. Raport ZPM-10/I1573/A1528/1990

[8] Trick M.A.: A Linear Relaxation Heuristc for the Generalized Assignment Problem. Report No. 89568-OR/1989 GS of IA, C.M.University, Pittsburg, PA.

[9] Walukiewicz S.: Metoda podziału i oszacowań dla uogólnionego zagadnienia przydziału pracy. Raport ZPM-35/A1528/CPBP 02.15/2.2.7/1990

ISBN 83-900412-1-9