Wykłady z programowania liniowego

A. Paweł Wojda

Wydział Matematyki Stosowanej AGH

Spis treści

Rozdział 1

Wstęp

Jeśli nie jesteś w stanie czegoś wytłumaczyć 5-letniemu dziecku,
to znaczy, że nie rozumiesz tego do końca
Albert Einstein

Niniejsza książka może być przydatna studentom różnych kierunków, szczególnie zaś kierunków ekonomicznych i technicznych. Pisałem ją jednak głównie z myślą o studentach Wydziału Matematyki Stosowanej AGH, dla których prowadzę od trzech lat semestralny wykład "Programowanie liniowe".

Mimo że istnieje wiele bardzo dobrych książek do nauki programowania liniowego, chcę przekazać studentom podręcznik, który zawiera dokładnie treść moich wykładów, w tym samym stopniu szczegółowości. Wykłady te obejmują również związane z programowaniem liniowym elementy optymalizacji kombinatorycznej: metody sieciowe i ich zastosowania w teorii grafów.

Mam nadzieję, że studenci Wydziału Matematyki Stosowanej zetkną się w tym podręczniku z matematyką, która rozwiązuje w niezbyt skomplikowany sposób trudne problemy praktyczne.

Programowanie liniowe jest szczególnym zagadnieniem programowania matematycznego. Ogólny **problem programowania matematycznego**[1] można sformułować następująco:

Niech X będzie dowolnym zbiorem i niech f będzie funkcją zdefiniowaną w zbiorze X o wartościach w zbiorze liczb rzeczywistch. Znajdź wartość maksymalną $f(x)$ dla $x \in X$.

Będziemy pisać krótko:

$f(x) \to \max$

[1] Warto zapamiętać tę definicję dlatego, że słowo "programowanie"występuje także w informatyce w znaczeniu bardzo różnym od rozważanego.

$x \in X$

W przypadku programowania liniowego o zbiorze X zakładamy, że jest podzbiorem $\mathbf{R}^\mathbf{n}$ zdefiniowanym przez układ nierówności liniowych, zaś funkcja f jest funkcją liniową.

Rozdziały 2–6 są poświęcone samemu programowaniu liniowemu. Kolejno więc przedstawiona jest metoda sympleks (rozdz. 3), zasada dualności (rozdz. 4), zredukowana metoda sympleks (rozdz. 5).

W rozdziale 6 omówiona jest szczególna sytuacja zadania ograniczonego.

Pozostałe rozdziały to interpretacje i zastosowania geometryczne programowania liniowego (rozdz. 7) i metody sieciowe z przykładami ich zastosowań w teorii grafów (rozdz. 8). Podrozdział 7.1, w którym jest mowa o interpretacji programowania liniowego w przypadkach dwu- i trzy-wymiarowym, może być czytany znacznie wcześniej niż występuje w tekście, jednak koniecznie po opisie metody sympleks – a więc po rozdziale 3.

Poświęcony przepływom w sieciach rozdział 8 zawarty jest w podręczniku z dwóch powodów. Powodem pierwszym jest fakt, że choć metoda sieci rozwiązuje tylko pewien szczególny problem programowania liniowego, niemniej dla tego szczególnego problemu algorytm sieciowy jest szybszy. Po drugie, mamy w ten sposób okazję do przedstawienia pięknego twierdzenia o maksymalnym przepływie i minimalnym przekroju. Wynikające z niego twierdzenia: Halla, Königa–Egerváryego i Mengera, dowodzą, jak głębokim jest ono wynikiem. Przekonamy się także, że twierdzenie to jest konsekwencją zasady dualności.

Pozostaje mi wyjaśnić znaczenie dwóch znaczków pojawiających się w różnych miejscach tekstu. $\blacksquare$ oznacza, że coś definitywnie się skończyło, najczęściej jest to znak końca dowodu twierdzenia (lub sygnał, że dowodu w ogóle nie będzie), czasami występuje na końcu przykładu. $\square$ występuje w tekście, gdy na chwilę przerywamy dany fragment (na ogół przykład) do którego jeszcze wrócimy.

Adam Paweł Wojda

Kraków, 2013 wojda@uci.agh.edu.pl

Rozdział 2

Problem programowania liniowego

2.1 PPL

Rozważania o programowaniu liniowym rozpocznijmy od przykładu.

Przykład 2.1.1 Właściciel ciężarówki przewozi cukier, mąkę i chipsy z miejscowości A do B. W ciężarówce mieści się towar o objętości co najwyżej 7000 litrów i o wadze co najwyżej 5 ton. 1 kilogram cukru ma objętość 1,5 litra, 1 kilogram mąki 2 litry, natomiast 1 kilogram chipsów ma objętość 4 litrów. Załóżmy także, że nasz przewoźnik zobowiązał się do dostarczenia co najmniej 1000 kg mąki i cukru. Zysk od przewozu poszczególnych towarów jest następujący:

– 8 zł za 100 kg cukru,

– 10 zł za 100 kg mąki,

– 25 zł za 100 kg chipsów.

Ile cukru, mąki i chipsów powinien załadować właściciel ciężarówki, aby zmaksymalizować swój zysk?

Matematyczny model tak postawionego zadania jest następujący:
Oznaczmy przez x_1 – wagę cukru, x_2 – wagę mąki, x_3 – wagę chipsów (za każdym razem w setkach kilogramów). Skoro ciężarówka może zabrać co najwyżej 5 ton towarów, musi zachodzić nierówność

$$100x_1 + 100x_2 + 100x_3 \leq 5000.$$

Z kolei ograniczenie objętości wyraża się wzorem

$$150x_1 + 200x_2 + 400x_3 \leq 7000.$$

Zobowiązanie dostarczenia 1000 kg mąki i cukru oznacza spełnienie wzoru

$$100x_1 + 100x_2 \geq 1000$$

Zysk właściciela wynosi

$$z = 8x_1 + 10x_2 + 25x_3,$$

x_1, x_2 i x_3 muszą być oczywiście nieujemne.

Po uproszczeniach otrzymamy więc problem programowania matematycznego:

$$f(x_1, x_2, x_3) = 8x_1 + 10x_2 + 25x_3 \to \max,$$

$$\text{w zbiorze} \left\{ \begin{array}{cccl} x_1 & +x_2 & +x_3 & \leq 50 \\ 1,5x_1 & +2x_2 & +4x_3 & \leq 70 \\ x_1 & +x_2 & & \geq 10 \\ & x_j \geq 0 & \text{dla} & j = 1, 2, 3. \end{array} \right.$$

Problemem programowania liniowego, lub krótko: **PPL**, będziemy nazywać problem programowania matematycznego sformułowany następująco:

Niech c_j oraz a_{ij} dla $i = 1, ..., m,\ j = 1, ..., n$ będą liczbami rzeczywistymi.

Znaleźć maksimum funkcji $f(x_1, ..., x_n) = \sum_{j=1}^{n} c_j x_j$ przy ograniczeniach:

$$\left\{ \begin{array}{rcll} \sum_{j=1}^{n} a_{ij} x_j & \leq & b_i & (i = 1, ..., m) \\ x_j & \geq & 0 & (j = 1, .., n) \end{array} \right. \tag{2.1}$$

Często wygodnie będzie nam pisać:

$$\left\{ \begin{array}{rcll} \sum_{j=1}^{n} a_{ij} x_j & \leq & b_i & (i = 1, ..., m) \\ x_j & \geq & 0 & (j = 1, .., n) \\ \hline f(x_1, ..., x_n) & = & \sum_{j=1}^{n} c_j x_j & \to \max \end{array} \right. \tag{2.2}$$

$f(x_1, ..., x_n) = \sum_{j=1}^{n} c_j x_j$ nazywamy **funkcją celu**, zaś nierówności (2.2) **ograniczeniami**. Zmienne $x_1, ..., x_n$ nazywamy **zmiennymi decyzyjnymi**. Zauważmy, że zarówno funkcja celu, jak i lewe strony nierówności we wzorze (2.2) są funkcjami liniowymi.

Problem programowania liniowego można także sformułować następująco:

$$\left\{ \begin{array}{ll} \text{zmaksymalizować} & \mathbf{cx} \\ \text{przy warunkach:} & \mathbf{Ax} \leq \mathbf{b} \\ & \mathbf{x} \geq \mathbf{\Theta}_n \end{array} \right. \tag{2.3}$$

gdzie:

$$\begin{array}{lcl} \mathbf{c} & = & (c_1, ..., c_n), \\ \mathbf{x} & = & (x_j)_{j=1,...,n} \in \mathbf{R}^n, \\ \mathbf{\Theta}_n & = & (0, ..., 0) \in \mathbf{R}^n, \\ \mathbf{A} & = & (a_{ij})_{i=1,...,m; j=1,...,n} \in \mathbf{R}^{m \times n}. \end{array}$$

Postać PPL przedstawioną wzorem (2.2) lub (2.3) nazywamy **postacią standardową PPL**.

Uwaga. Biznesmeni mają dwie strategie zachowań:

1) maksymalizują swój zysk (model optymistyczny),

2) minimalizują koszty (model pesymistyczny) – oczywiście przy założonych z góry minimalnych efektach.

Pesymistyczny model zachowań może mieć formę analityczną taką, jak w poniższym przykładzie.

Przykład 2.1.2 PPL w postaci niestandardowej
Niech będzie dany następujący problem:

$$\text{zminimalizować:} \quad x_1 - x_2 + x_3$$

$$\text{przy ograniczeniach:} \left\{ \begin{array}{rrrcl} x_1 & -2x_2 & +x_3 & \leq & 2 \\ x_1 & & -x_3 & \geq & 1 \\ x_1 & +x_2 & +2x_3 & = & 4 \\ & & x_i & \geq & 0 \quad (i = 1, 2, 3). \end{array} \right.$$

Zauważmy, że za pomocą banalnych operacji arytmetycznych można powyższy problem sprowadzić do PPL w postaci standardowej:

$$\text{zmaksymalizować:} \quad -x_1 + x_2 - x_3$$

$$\text{przy ograniczeniach:} \left\{ \begin{array}{rrrcr} x_1 & -2x_2 & +x_3 & \leq & 2 \\ -x_1 & & +x_3 & \leq & -1 \\ x_1 & +x_2 & +2x_3 & \leq & 4 \\ -x_1 & -x_2 & -2x_3 & \leq & -4 \\ & & x_i & \geq & 0 \quad (i = 1, 2, 3). \end{array} \right.$$

Zadanie polegające na maksymalizacji funkcji $f : X \to \mathbf{R}$ jest równoważne minimalizacji funkcji $g : X \to \mathbf{R}$ zdefiniowanej wzorem $g(x) = -f(x)$. Każdą nierówność postaci

$$\sum_{j=1}^{n} a_{ij} x_j \geq b_i$$

można zastąpić równoważną jej nierównością

$$\sum_{j=1}^{n} (-a_{ij}) x_j \leq -b_i,$$

zaś równość

$$\sum_{j=1}^{n} a_{ij} x_j = b_i$$

dwiema nierównościami:

$$\sum_{j=1}^{n} a_{ij}x_j \leq b_i$$

oraz

$$\sum_{j=1}^{n} a_{ij}x_j \geq b_i.$$

Tak więc mimo że wcześniej zdefiniowaliśmy PPL w postaci na pozór dość szczególnej (nazwaliśmy ją *standardową*), równie dobrze można powiedzieć, że PPL polega na maksymalizacji lub minimalizacji formy liniowej w zbiorze, który jest rozwiązaniem układu nierówności i równań liniowych w $\mathbf{R}^n$.

Problemem rozstrzygania niesprzeczności i rozwiązywania układów nierówności i równań liniowych będziemy się zajmować w rozdziale 7.

2.2 Definicje

Niech dany będzie problem programowania liniowego (w postaci standardowej)

$$\left\{\begin{array}{lcl} \mathbf{Ax} & \leq \mathbf{b} & \\ \mathbf{x} & \geq \mathbf{\Theta}_n & \\ \hline & & \\ \mathbf{cx} & \rightarrow & \max \end{array}\right. \tag{2.4}$$

$\mathbf{x} = (x_1, ..., x_n)$ (lub: x_j $(j = 1, ..., n)$) nazywamy **rozwiązaniem dopuszczalnym** PPL (2.4) jeśli

$\mathbf{x} \geq \mathbf{\Theta}_n$ (inaczej: $x_j \geq 0$, dla $j = 1, ..., n$),

$\mathbf{Ax} \leq \mathbf{b}$ (inaczej: $\sum_{j=1}^{n} a_{ij}x_j \leq b_i, i = 1, ..., m$).

Rozwiązaniem optymalnym PPL (2.4) (lub (2.3)) nazywamy takie rozwiązanie dopuszczalne $\mathbf{x}$, dla którego funkcja $f(\mathbf{x}) = \mathbf{cx}$ przyjmuje wartość maksymalną.

Oczywiście, może się zdarzyć, że PPL w ogóle nie ma rozwiązań dopuszczalnych, tzn. zbiór

$$\{\mathbf{x} \in \mathbf{R^n} : \mathbf{x} \geq \mathbf{\Theta}_n \text{ i } \mathbf{Ax} \leq \mathbf{b}\} \tag{2.5}$$

jest zbiorem pustym. Mówimy wówczas, że PPL (2.4) jest problemem **sprzecznym** (lub, że warunki: $\mathbf{x} \geq \mathbf{\Theta}_n$ i $\mathbf{Ax} \leq \mathbf{b}$ są sprzeczne).

Jeśli zbiór (2.5) jest niepusty i funkcja f w tym zbiorze nie ma maksimum, to mówimy że PPL jest **nieograniczony**. W rzeczy samej, zbiór (2.5) jest oczywiście domknięty w $\mathbf{R^n}$, a funkcja f - ciągła. Jeśli więc zbiór (2.5) jest niepusty i f nie przyjmuje w tym zbiorze maksimum, to f jest w zbiorze nieograniczona (sam zbiór także jest nieograniczony).

Przykład 2.2.1 Problem

$$\begin{array}{rcrll} x_1 & - & x_2 & \leq 2 & \\ x_2 & - & x_1 & \leq 3 & \\ & & x_i & \geq 0 & (i=1,2) \\ \hline x_1 & + & x_2 & \rightarrow & \max \end{array}$$

jest oczywiście sprzeczny.

Przykład 2.2.2 Problem

$$\begin{array}{rcrll} x_1 & - & x_2 & \leq 0 & \\ -3x_1 & + & x_2 & \leq 0 & \\ & & x_i & \geq 0 & (i=1,2) \\ \hline x_1 & + & x_2 & \rightarrow & \max \end{array}$$

jest nieograniczony.

Rzeczywiście, dla dowolnego $t \geq 0$, $x_1 = t$, $x_2 = 2t$ jest rozwiązaniem dopuszczalnym. Wartość funkcji $f(x_1, x_2) = x_1 + x_2 = 3t$ może być wówczas dowolnie duża.

2.3 ćwiczenia

Ćwiczenie 2.3.1 Sprowadź poniższe problemy programowania liniowego do postaci standardowej:

a) zmaksymalizuj: $x_1 + 2x_2 - 3x_3$

przy warunkach:
$$\begin{cases} x_1 \quad -x_2 \quad +x_3 \leq 3 \\ x_1 \quad -x_3 \geq 2 \\ x_1 \quad +x_2 \quad -2x_3 = 4 \\ x_i \geq 0 \quad (i=1,2,3), \end{cases}$$

b) zminimalizuj: $-x_1 + 2x_2 - 3x_3 - x_4$

przy warunkach:
$$\begin{cases} x_1 \quad -x_2 \quad +3x_3 \quad -2x_4 = 2 \\ -2x_1 \quad -x_2 \quad +2x_4 \geq 3 \\ x_i \geq 0 \quad (i=1,...,4), \end{cases} .$$

c)
$$\begin{array}{rrrrll} x_1 & +2x_2 & -x_3 & -3x_4 & \geq 2 & \\ x_1 & +2x_2 & +x_3 & +3x_4 & = 1 & \\ & & & x_i & \geq 0 & (i=1,...,4) \\ \hline \multicolumn{5}{c}{-x_1 - 2x_2 + 3x_3 + x_4 \rightarrow \min} & . \end{array}$$

Ćwiczenie 2.3.2 Problem diety Jest to znany, klasyczny wręcz problem. Przedstawmy go w wielkim uproszczeniu (pełne dane można znaleźć w [23]).

Ola postanowiła odżywiać się najtaniej, jak to możliwe, dostarczając jednak swojemu organizmowi odpowiednich ilości białka, witamin A i C, wapnia oraz energii (czyli składników żywności). Postanowiła odżywiać się mlekiem, serem, chlebem, cielęciną i marchewką. W poniższym zestawieniu podano zawartości odpowiednich składników w tych produktach (na 100 g produktu).

składniki żywności	mleko	ser	chleb	cielęcina	marchew
białko (g)	3	38	0	20	1
wit A (jedn.)	140	120	0	0	5760
wit C (mg)	1	0	0	0	3
wapń (mg)	120	1450	90	8	19
energia (kcal)	53	200	240	82	21

Zestawienia prezentujące zapotrzebowania dobowe Oli na poszczególne składniki oraz ceny produktów podajemy poniżej.

białko	70
wit A	5000
wit C	75
wapń	70
energia	2700

mleko	1,5 zł/ł
ser	5 zł//kg
chleb	1,5 zł//kg
cielęcina	12 zł//kg
marchew	0,8 zł//kg

Ułóż PPL, którego rozwiązanie będzie dla Oli wskazówką, jak za najniższą cenę zaspokoić potrzeby swojego organizmu. Waga dziennego pożywienia nie może przekroczyć 2 kg[1].

[1]Po to, by Ola nie musiała jeść zbyt dużo marchewki z chlebem.

Rozdział 3

Opis algorytmu sympleks

3.1 Wprowadzenie

W niniejszym rozdziale przedstawiony zostanie algorytm sympleks. Jak zobaczymy, algorytm ten rozwiązuje problemy programowania liniowego. Zaczniemy od rozwiazania prostych przykładów. Będziemy systematycznie zwiększać stopień trudności przykładów, by nieco później, w podrozdziale 3.3, podać szczegółowy opis algorytmu, łącznie ze sposobami wybrnięcia z problemów, które mogą sie pojawić (wyznaczenie pierwszego rozwiązania dopuszczalnego i wystąpienie zjawiska cykliczności).

Przykład 3.1.1
Rozważmy PPL

$$\begin{array}{rrrrrl} & & x_1 & +2x_2 & +3x_3 & \leq 5 \\ & & 2x_1 & +3x_2 & +5x_3 & \leq 8 \\ & & 3x_1 & +x_2 & +3x_3 & \leq 4 \\ & & & & & x_i \geq 0 \quad (i = 1, ..., 3) \\ \hline z & = & 2x_1 & +x_2 & +3x_3 & \rightarrow \qquad \max \end{array} \tag{3.1}$$

Pierwszym krokiem będzie wprowadzenie **zmiennych dodatkowych** x_4, x_5, x_6, które zdefiniujemy następująco

$$\begin{cases} x_4 & = 5 \quad -x_1 \quad -2x_2 \quad -3x_3 \\ x_5 & = 8 \quad -2x_1 \quad -3x_2 \quad -5x_3 \\ x_6 & = 4 \quad -3x_1 \quad -x_2 \quad -3x_3 \end{cases} \tag{3.2}$$

Rozważmy teraz problem następujący

$$\begin{array}{rlrrrll} x_4 & = 5 & -x_1 & -2x_2 & -3x_3 & & \\ x_5 & = 8 & -2x_1 & -3x_2 & -5x_3 & & \\ x_6 & = 4 & -3x_1 & -x_2 & -3x_3 & & \\ x_i & \geq 0 & & (i = & 1, ..., 6) & & \\ \hline z & = & 2x_1 & +x_2 & +3x_3 & \rightarrow & \max \end{array} \tag{3.3}$$

Jest zupełnie oczywiste, że problemy (3.1) i (3.3) są równoważne w tym sensie, że każde rozwiązanie optymalne problemu (3.1) daje nam pewne rozwiązanie optymalne problemu (3.3) (wystarczy wartości zmiennych x_4, x_5 i x_6 wyznaczyć z równości (3.2)). Podobnie, każde optymalne rozwiązanie problemu (3.3) dostarcza nam optymalnego rozwiązania (3.1).

Metoda będzie polegała na tym, że mając pewne rozwiązanie dopuszczalne problemu (3.3) $\mathbf{x}$, będziemy szukali następnego rozwiązania $\mathbf{x}'$ takiego, że $f(\mathbf{x}') > f(\mathbf{x})$, a więc lepszego. Takie postępowanie będziemy powtarzali wielokrotnie, otrzymując rozwiązania coraz bardziej zbliżone do optymalnego (o ile takie rozwiązanie istnieje).

O pierwsze rozwiązanie dopuszczalne w przykładzie nietrudno. Niech

$$x_1 = x_2 = x_3 = 0.$$

Wtedy oczywiście:

$$x_4 = 5, \quad x_5 = 8, \quad x_6 = 4, \quad z = 0.$$

Przyjrzyjmy się teraz wzorowi na z w (3.3). Ponieważ współczynniki przy x_i $(i = 1, 2, 3)$ są dodatnie, jeśli zwiększymy wartość którejkolwiek ze zmiennych x_i $(i = 1, 2, 3)$, wówczas zwiększy się także wartość z. Wartość ta będzie się tym szybciej powiększać, im większy jest współczynnik (dodatni) przy powiększanym x we wzorze na z.

Współczynnikiem przy x_1 jest 2, przy x_2 – 1, natomiast przy x_3 – 3. Będziemy więc zwiększać wartość x_3. O ile można powiększyć x_3 bez zmieniania wartości pozostałych zmiennych, by otrzymane w ten sposób nowe rozwiązanie było dopuszczalne? Wartość x_3 można zwiększyć o tyle, by zachowane zostały w dalszym ciągu nierówności:

$$x_4 \geq 0, \quad x_5 \geq 0, \quad x_6 \geq 0$$

(pamiętamy, że w naszym rozwiązaniu $x_1 = x_2 = x_3 = 0$).

łatwo zauważyć, że:

$$x_4 \geq 0 \Rightarrow x_3 \leq \frac{5}{3}, \quad x_5 \geq 0 \Rightarrow x_3 \leq \frac{8}{5}, \quad x_6 \geq 0 \Rightarrow x_3 \leq \frac{4}{3}.$$

Największą wartością x_3, jaką możemy wybrać, jest

$$x_3 = \frac{4}{3}.$$

Wtedy

$$z = 4.$$

Zauważmy jak oczywista jest tu metoda postępowania. Funkcja celu wyraża się za pomocą zmiennych, które w rozwiązaniu dopuszczalnym miały początkową wartość równą zero. Jasne więc było, że jeśli zwiększyć wartość tej zmiennej, która we wzorze na z ma współczynnik dodatni, zwiększy się także wartość

z. Z kolei ograniczenia $x_i \geq 0$ i wzory na x_4, x_5, x_6 pozwoliły z łatwością ustalić, o ile wolno nam zwiększyć wartości poszczególnych zmiennych tak, by otrzymać rozwiązanie dopuszczalne.

Teraz wartość zero przyjmują zmienne x_1, x_2 oraz x_6, natomiast $x_3 = \frac{4}{3}$. Przedstawmy x_3 za pomocą x_1, x_2 i x_6 korzystając z trzeciego z równań (3.2)

$$x_3 = \frac{4}{3} - x_1 - \frac{1}{3}x_2 - \frac{1}{3}x_6. \tag{3.4}$$

Tak otrzymane x_3 wstawmy do wzoru na z:

$$z = 4 - x_1 - \frac{1}{2}x_6.$$

Wobec warunków $x_1 \geq 0$ i $x_6 \geq 0$ jest oczywiste, że maksymalną wartością, jaką może przyjąć z, jest 4.

Rozwiązanie:

$$x_1 = 0, \quad x_2 = 0, \quad x_3 = \frac{4}{3}, \quad z = 4$$

jest rozwiązaniem optymalnym. ■

To co zrobiliśmy w przykładzie 3.1.1 spróbujmy teraz przenieść na przypadek ogólny. Niech będzie dany PPL

$$\begin{array}{rll} \sum_{j=1}^{n} a_{ij}x_j & \leq b_i & (i = 1, ..., m) \\ x_j & \geq 0 & (j = 1, .., n) \end{array} \tag{3.5}$$

$$\overline{}$$

$$z \;=\; \sum_{j=1}^{n} c_j x_j \quad \to \quad \max$$

Problem ten jest oczywiście równoważny problemowi znalezienia maksymalnej wartości z spełniającej układ równań:

$$\begin{cases} x_{n+i} &= b_i - \sum_{j=1}^{n} a_{ij}x_j \quad (i=1,...,m) \\ z &= \sum_{j=1}^{n} c_j x_j \end{cases} \tag{3.6}$$

przy założeniu, że $x_l \geq 0$, dla $l = 1, ..., n+m$.

Podobnie jak to miało miejsce w przykładzie (3.1.1), także w przypadku ogólnym będziemy rozważali układ równań

$$\begin{cases} \bar{x}_{n+i} &= \bar{b}_i - \sum_{j=1}^{n} \bar{a}_{ij}\bar{x}_j \quad (i=1,...,m) \\ \bar{z} &= z_0 + \sum_{j=1}^{n} \bar{c}_j \bar{x}_j \end{cases} \tag{3.7}$$

skojarzony z PPL (3.5) równoważny układowi (3.6) w tym sensie, że otrzymany z niego przez być może wielokrotne zamiany ról zmiennych znajdujących się po prawej i lewej stronie układu.

Definicja 3.1.1 *Układ równań (3.7) nazywamy* **słownikiem** *PPL (3.5). Słownik (3.7) nazywamy* **słownikiem dopuszczalnym** *PPL (3.5), jeżeli po podstawieniu $\bar{x}_j = 0$ dla $j = 1, ..., n$, otrzymamy rozwiązanie dopuszczalne, inaczej mówiąc, jeżeli*

$$\bar{x}_1 = ... = \bar{x}_n = 0; \quad \bar{x}_{n+1} = \bar{b}_1, ..., \bar{x}_{n+m} = \bar{b}_m \tag{3.8}$$

jest rozwiązaniem dopuszczalnym PPL (3.5).

Zmienne po lewej stronie znaków równości w słowniku nazywamy **zmiennymi bazowymi**, *zmienne po prawej stronie –* **niebazowymi**.

Zwróćmy uwagę, że dla słownika (3.7) wektor $\bar{\mathbf{x}} = (\bar{x}_1, ..., \bar{x}_{n+m})$ zdefiniowany wzorem (3.8) nie musi być rozwiązaniem dopuszczalnym PPL (3.5). żeby $\bar{\mathbf{x}}$ był rozwiązaniem dopuszczalnym wszystkie liczby $\bar{b}_1, ..., \bar{b}_m$ muszą być nieujemne, a to wcale nie musi zachodzić. Stąd następna definicja.

Definicja 3.1.2 *Wektor $(\bar{x}_k)_{k=1,...,n+m}$ dany równaniami (3.8) nazywamy* **rozwiązaniem bazowym**.

W przykładzie 3.1.1 mieliśmy do czynienia z dwoma słownikami:

$$\begin{cases} x_4 &= 5 \quad -x_1 \quad -2x_2 \quad -3x_3 \\ x_5 &= 8 \quad -2x_1 \quad -3x_2 \quad -5x_3 \\ x_6 &= 4 \quad -3x_1 \quad -x_2 \quad -3x_3 \\ z &= \quad 2x_1 \quad +x_2 \quad +3x_3 \end{cases} \tag{3.9}$$

oraz

$$\begin{cases} x_3 &= \frac{4}{3} & -x_1 & -\frac{1}{3}x_2 & -\frac{1}{3}x_6 \\ x_4 &= 1 & +2x_1 & -x_2 & +x_6 \\ x_5 &= \frac{4}{3} & +3x_1 & -\frac{4}{3}x_2 & +\frac{5}{3}x_6 \\ z &= 4 & -x_1 & & -x_6 \end{cases} \tag{3.10}$$

Słownika (3.10) nie mieliśmy potrzeby wypisywać wcześniej. żeby go otrzymać, należy x_3 ze wzoru (3.4) wstawić do wzorów na z, x_4 i x_5 w słowniku (3.9).

W słowniku (3.9) zmiennymi bazowymi są x_4, x_5 i x_6, zaś x_1, x_2, x_3 są zmiennymi niebazowymi. W (3.10) zmienne bazowe to x_3, x_4, x_5, zmienne niebazowe – x_1, x_2, x_6.

Oba słowniki są dopuszczalne bo, dla przykładu, rozwiązanie bazowe (3.10) $x_1 = x_2 = x_6 = 0,\ x_3 = \frac{4}{3},\ x_4 = 1,\ x_5 = \frac{4}{3}$, jest równocześnie rozwiązaniem dopuszczalnym ($x_i \geq 0$ dla $i = 1, ..., 6$)).

W przykładzie 3.1.1 wystarczyła jedna iteracja do otrzymania rozwiązania optymalnego. W następnym przykładzie zobaczymy, że tak być nie musi. Opis tego przykładu ułatwią nam nowe definicje.

Przykład 3.1.2
Rozwiążemy PPL:

$$\begin{array}{rrrrl} x_1 & +2x_2 & +x_3 & \leq 4 & \\ 2x_1 & +x_2 & +x_3 & \leq 3 & \\ x_1 & -x_2 & +4x_3 & \leq 2 & \\ & & & x_i \geq 0 & i = 1, 2, 3 \\ \hline \multicolumn{5}{c}{z = 3x_1 + x_2 + 2x_3 \to \max} \end{array} \tag{3.11}$$

Zdefiniujmy zmienne dodatkowe (są one równocześnie zmiennymi bazowymi pierwszego słownika).

$$\begin{cases} x_4 &= 4 & -x_1 & -2x_2 & -x_3 \\ x_5 &= 3 & -2x_1 & -x_2 & -x_3 \\ x_6 &= 2 & -x_1 & +x_2 & -4x_3 \\ & x_i \geq 0 & (i = 1, ..., 6) & & \end{cases} \tag{3.12}$$

Największym współczynnikiem dodatnim w funkcji celu jest 3 (współczynnik przy x_1), stąd nową zmienną bazową będzie x_1. Taką zmienną $\bar{x}_j$ będziemy nazywali **zmienną wchodzącą**.

Zmienną wychodzącą, czyli taką, która przestanie w następnym kroku być zmienną bazową, jest $\bar{x}_{n+i}$ $(i \in \{1, ..., m\})$, dla którego nierówność $\bar{x}_{n+i} \geq 0$ daje najostrzejsze ograniczenie od góry na wartość zmiennej wchodzącej. Inaczej mówiąc, $\bar{x}_{n+i_0}$ jest zmienną wychodzącą słownika (3.7), jeżeli

$$\frac{\bar{b}_{i_0}}{\bar{a}_{i_0 j}} \geq 0 \quad \text{i} \quad i_0 : \frac{\bar{b}_{i_0}}{\bar{a}_{i_0 j}} = \min\left\{ \frac{\bar{b}_i}{\bar{a}_{ij}} : \bar{a}_{ij} \neq 0 \wedge \frac{\bar{b}_i}{\bar{a}_{ij}} \geq 0,\ i = 1, ..., m \right\}.$$

W naszym przykładzie zmienną wychodzącą będzie x_5, łatwo bowiem sprawdzić, że:

$x_4 \geq 0 \Rightarrow x_1 \leq 4,$

$x_5 \geq 0 \Rightarrow x_1 \leq \frac{3}{2},$

$x_6 \geq 0 \Rightarrow x_1 \leq 2.$

Obliczamy więc x_1 z drugiego równania słownika (3.12)

$$x_1 = \frac{3}{2} - \frac{1}{2}x_2 - \frac{1}{2}x_3 - \frac{1}{2}x_5 \tag{3.13}$$

i tworzymy nowy słownik o zmiennych bazowych x_1, x_4, x_6, przepisując (3.13) jako pierwsze równanie i tworząc pozostałe dwa z pierwszego i trzeciego z równań słownika (3.12), zastępując w nich x_1 prawą stroną wzoru (3.13). Nową funkcję celu otrzymamy, wstawiając x_3 dane wzorem (3.13) do *starej* funkcji celu.

W rezultacie otrzymamy słownik dopuszczalny:

$$\left\{\begin{array}{llllll} x_1 & = \frac{3}{2} & -\frac{1}{2}x_2 & -\frac{1}{2}x_3 & -\frac{1}{2}x_5 \\ x_4 & = \frac{5}{2} & -\frac{3}{2}x_2 & -\frac{1}{2}x_3 & +\frac{1}{2}x_5 \\ x_6 & = \frac{1}{2} & +\frac{3}{2}x_2 & -\frac{7}{2}x_3 & +\frac{1}{2}x_5 \\ z & = \frac{9}{2} & -\frac{1}{2}x_2 & +\frac{1}{2}x_3 & -\frac{3}{2}x_5 \end{array}\right. \tag{3.14}$$

Teraz jako zmienną wchodzącą musimy wybrać x_3, ponieważ tylko ta zmienna ma współczynnik dodatni we wzorze na z. Nierówności $x_i \geq 0$ dla $i = 1, 4, 6$ dają, odpowiednio, nierówności dla wchodzącej zmiennej bazowej: $x_3 \leq 3,\ x_3 \leq 5,\ x_3 \leq \frac{1}{7}$.

Z tych nierówności najostrzejszym ograniczeniem na wartość x_2 od góry jest $x_3 \leq \frac{1}{7}$, otrzymana z $x_6 \geq 0$, tak więc wychodzącą zmienną bazową jest x_6. Po analogicznych do poprzednich obliczeniach otrzymamy nowy słownik (już trzeci dla naszego problemu):

$$\left\{\begin{array}{llllll} x_1 & = \frac{10}{7} & -\frac{5}{7}x_2 & -\frac{4}{7}x_5 & +\frac{1}{7}x_6 \\ x_3 & = \frac{1}{7} & +\frac{3}{7}x_2 & +\frac{1}{7}x_5 & -\frac{2}{7}x_6 \\ x_4 & = \frac{17}{7} & -\frac{12}{7}x_2 & +\frac{3}{7}x_5 & +\frac{1}{7}x_6 \\ z & = \frac{32}{7} & -\frac{2}{7}x_2 & -\frac{10}{7}x_5 & -\frac{1}{7}x_6 \end{array}\right. \tag{3.15}$$

Ten słownik jest ostatni. Jest dopuszczalny, a jego rozwiązanie bazowe

$$x_1 = \frac{10}{7},\ x_2 = 0,\ x_3 = \frac{1}{7},\ x_4 = \frac{17}{7},\ x_5 = x_6 = 0,\ z = \frac{32}{7}$$

jest rozwiazaniem optymalnym. $\square$

3.2 Tabele sympleksowe

Przyjrzyjmy się jeszcze raz przykładowi 3.1.1 zapisując PPL rozważany w nim nieco inaczej. Przeanalizujmy mianowicie układ równań:

$$\begin{array}{rrrrrrrr}
 & x_1 & +2x_2 & +3x_3 & +x_4 & & & =5 \\
 & 2x_1 & 3x_2 & +5x_3 & & +x_5 & & =8 \\
 & 3x_1 & +x_2 & +3x_3 & & & +x_6 & =4 \\
\hline
-z & +2x_1 & +x_2 & +3x_3 & & & & =0
\end{array}$$

Powyżej linii poziomej zapisaliśmy ograniczenia naszego PPL (pominęliśmy $x_j \geq 0,\ j = 1, ..., 6$, ale w programowaniu liniowym wystarczy zapamiętać, że ograniczenia na nieujemność wszystkich zmiennych występują często, dopiero w dalszej części książki omówimy sytuację w której ograniczenia na znak występują tylko dla niektórych zmiennych), poniżej zaś funkcję celu. Skojarzone z tym PPL rozwiązanie bazowe jest w naszym przykładzie postaci: $x_1 = x_2 = x_3 = 0,\ x_4 = 5,\ x_5 = 8,\ x_6 = 4,\ z = 0$. Schematycznie można to zapisać w tabeli:

$$\begin{array}{cccccc|c}
1 & 2 & 3 & 1 & 0 & 0 & 5 \\
2 & 3 & 5 & 0 & 1 & 0 & 8 \\
3 & 1 & 3 & 0 & 0 & 1 & 4 \\
\hline
2 & 1 & 3 & 0 & 0 & 0 & 0
\end{array}$$

Naszą metodę przedstawioną na przykładach 3.1.1 i 3.1.2 można teraz przedstawić następująco:

Krok 1. Wyznaczanie zmiennej wchodzącej. Sprawdzamy ostatni wiersz tabeli, pomijając ostatni wyraz tego wiersza (odpowiadający wartości z). Jeśli w tym wierszu wszystkie współczynniki są mniejsze od zera **STOP**, tabela opisuje rozwiązanie optymalne.

Jeśli niektóre ze współczynników w tym wierszu są dodatnie, wybieramy największy z nich. Numer j kolumny, w której się znajduje, jest numerem *zmiennej wchodzącej* x_j (inaczej: x_j jest nową zmienną bazową). W naszym przypadku zmienną bazową wchodzącą jest x_3.

$$\begin{array}{cc|c|ccc|c}
 & & a_{ij} & & & & b_i \\
1 & 2 & 3 & 1 & 0 & 0 & 5 \\
2 & 3 & 5 & 0 & 1 & 0 & 8 \\
3 & 1 & 3 & 0 & 0 & 1 & 4 \\
\hline
2 & 1 & \mathbf{3} & 0 & 0 & 0 & 0
\end{array}$$

Krok 2. Wybieramy zmienną wychodzącą. Wybieramy wiersz i_0, w którym $a_{ij} > 0$ i iloraz $\frac{b_i}{a_{ij}}$ jest najmniejszy. W naszym przykładzie najmniejszą z liczb $\frac{5}{3},\ \frac{8}{5},\ \frac{4}{3}$ jest $\frac{4}{3}$. Wybieramy więc trzeci wiersz ($i_0 = 3$).

$$\begin{array}{cc|c|ccc|c}
1 & 2 & 3 & 1 & 0 & 0 & 5 \\
2 & 3 & 5 & 0 & 1 & 0 & 8 \\
\hline
3 & 1 & 3 & 0 & 0 & 1 & 4 \\
\hline\hline
2 & 1 & \mathbf{3} & 0 & 0 & 0 & 0
\end{array}$$

Krok 3. Tworzymy nowy słownik. Mnożymy wiersz i_0 przez odpowiednie współczynniki i odejmujemy od pozostałych tak, by w j-tej kolumnie jedynym niezerowym wyrazem był wyraz w wierszu i_0. Dzielimy wiersz i_0 przez $a_{i_0 j}$.

Tak więc w naszym przykładzie:

1. odejmujemy trzeci wiersz od pierwszego,
2. mnożymy trzeci wiersz przez $\frac{5}{3}$ i odejmujemy od drugiego,
3. odejmujemy trzeci wiersz od ostatniego (tego *pod kreską*, odpowiadającego zmiennej z,
4. dzielimy trzeci wiersz przez 3.

Otrzymamy następną tabelę:

$$\begin{array}{rrrrrr|r} -2 & 1 & 0 & 1 & 0 & -1 & 1 \\ -3 & -\frac{4}{3} & 0 & 0 & 1 & -\frac{5}{3} & \frac{4}{3} \\ -1 & \frac{1}{3} & 1 & 0 & 0 & \frac{1}{3} & \frac{4}{3} \\ \hline -1 & 0 & 0 & 0 & 0 & -1 & -4 \end{array} \tag{3.16}$$

Zauważmy, że otrzymana tabela dokładnie odpowiada słownikowi. Przekonujemy się o tym stwierdzając, że:

- pierwszy wiersz tabeli (3.16) odpowiada drugiemu równaniu słownika (3.10),
- drugi wiersz tabeli (3.16) trzeciemu równaniu (3.10),
- trzeci wiersz pierwszemu równaniu (3.10),
- należy uwzględnić zmiany znaków (np. w ostatnim wierszu i kolumnie tabeli (3.16) występuje minus wartość zmiennej z, podczas gdy w słowniku (3.10) występuje wartość z itd.

Jest oczywiste, że przedstawienie algorytmu przy pomocy ciągu tabel sympleksowych niczym istotnym nie różni się od przedstawienia go w postaci ciągu zmieniających się słowników. Wielu autorów (np. [17], [14], [21]) tak właśnie robi. W dalszym ciągu prezentacji metody simpleks nie będziemy się tymi tabelami posługiwać.

Słownik jest układem równań liniowych – tabela macierzą jednoznacznie mu przyporządkowaną. Jeden słownik powstaje z drugiego przez operacje algebraiczne prowadzące do układów równoważnych. Stąd oczywista jest własność wszystkich słowników każdego PPL, którą wygodnie będzie nam zapisać w postaci twierdzenia.

Twierdzenie 3.2.1 *Dla ustalonego PPL dowolne dwa słowniki są równoważne. W szczególności każde rozwiązanie dowolnego słownika jest rozwiązaniem każdego innego słownika.*

3.3 Szczegóły metody

Zapoznając się z metodą simpleks skrupulatnie omijaliśmy wszelkie trudności, na które można natrafić. Przykłady były dobrane tak, by:

1) wiadomo było od czego trzeba zacząć bowiem $x_j = 0$ dla $j = 1, ..., n$ było rozwiązaniem dopuszczalnym;

2) rozpoczęty algorytm działał niezawodnie w tym sensie, że na żadnym etapie jego funkcjonowania nie mieliśmy wątpliwości co należy zrobić w następnym kroku a następny krok był zawsze wykonywalny;

3) po trzech (co najwyżej!) zmianach zmiennych bazowych otrzymywaliśmy optymalne rozwiązanie.

Czy zawsze funkcjonowanie algorytmu jest takie proste? Okazuje się, że nie. Rozważmy następujący przykład.

Przykład 3.3.1

$$\begin{array}{rcrcrcl} x_1 & - & 2x_2 & + & 3x_3 & \le & 5 \\ -x_1 & + & 4x_2 & - & x_3 & \le & -6 \\ 3x_1 & - & 2x_2 & + & x_3 & \le & -3 \\ \hline z = x_1 & + & x_2 & + & 2x_3 & \to & max \\ & & & & x_i & \ge & 0 \quad i = 1, 2, 3 \end{array} \tag{3.17}$$

Tym razem $x_1 = x_2 = x_3 = 0$ nie jest rozwiązaniem dopuszczalnym problemu, nie spełnia drugiej ani trzeciej nierówności w PPL (3.17). Co więcej, wcale nie jest oczywiste, że takie rozwiązanie w ogóle istnieje.

3.3.1 Od czego zacząć?

W ogólnym przypadku problemu PL w postaci standardowej:

$$\begin{array}{rcll} \sum_{j=1}^{n} a_{ij}x_j & \le & b_i & i = 1, ..., m \\ x_j & \ge & 0 & j = 1, ..., n \\ \hline \sum_{j=1}^{n} c_j x_j & \to & & \qquad\qquad \max \end{array} \tag{3.18}$$

rozwiązanie bazowe $x_1 = x_2 = ... = x_n = 0$ jest dopuszczalne, wtedy i tylko wtedy, gdy $b_i \ge 0$ dla każdego $i = 1, ..., m$.
Jak postępować gdy choć jedna z liczb b_i jest ujemna?

Rozwiążemy ten problem za pomocą metody sympleks. Rozważmy następujący PPL:

$$\begin{array}{rcll} \sum_{j=1}^{n} a_{ij}x_j - x_0 & \le & b_i & i = 1, ..., m \\ x_j & \ge & 0 & j = 0, 1, ..., n \\ \hline x_0 & \to & \min & \end{array} \tag{3.19}$$

Dla dowolnie ustalonych $x_1, ..., x_n$ istnieje takie x_0, że nierówności

$$\sum_{j=1}^{n} a_{ij}x_j - x_0 \ \leq \ b_i \ \text{ dla } \ i = 1, ..., m$$

są spełnione, bowiem można przyjąć dowolnie duże x_0. Tak więc problem (3.19) jest niesprzeczny. Jest także oczywiste, że problem wyjściowy (3.18) jest niesprzeczny wtedy i tylko wtedy jeśli rozwiązaniem PPL (3.19) jest $x_0 = 0$. Pierwszą zmienną wchodzącą będzie x_0, natomiast zmienną wychodzącą x_{i_0} taka, że b_{i_0} jest najmniejszą ze stałych b_i, $i = 1, ..., m$.

Sprawdźmy na naszym przykładzie jak ta metoda funkcjonuje.

Przykład 3.3.1 (cd.)
Zgodnie przyjętym przez nas zwyczajem, będziemy maksymalizować, a nie minimalizować, pewną funkcję celu.

Problem

$$\begin{array}{rrrrl} x_1 & -2x_2 & +3x_3 & -x_0 & \leq 5 \\ -x_1 & +4x_2 & -x_3 & -x_0 & \leq -6 \\ 3x_1 & -2x_2 & +x_3 & -x_0 & \leq -3 \\ & x_i & \geq & 0 & i = 0,1,2,3 \\ \hline & w & = -x_0 & \rightarrow & \max \end{array} \tag{3.20}$$

po wprowadzeniu zmiennych bazowych x_4, x_5, x_6 przyjmie postać

$$\begin{array}{rlrrrr} x_4 & = 5 & +x_0 & -x_1 & +2x_2 & -3x_3 \\ x_5 & = -6 & +x_0 & +x_1 & -4x_2 & +x_3 \\ x_6 & = -3 & +x_0 & -3x_1 & +2x_2 & -x_3 \\ & x_i & \geq 0 & i = 0, ..., 6 & & \\ \hline & w & = -x_0 & \rightarrow & \max & \end{array} \tag{3.21}$$

(3.21) nie jest słownikiem dopuszczalnym, bowiem rozwiązaniem bazowym jest $x_0 = x_1 = x_2 = x_3 = 0, x_4 = 5, x_5 = -6, x_6 = -3$, więc x_5 i x_6 przyjmują wartości ujemne. Ten defekt zniknie po wprowadzeniu x_0 jako zmiennej bazowej w miejsce x_5, inaczej mówiąc przyjmujemy x_0 jako zmienną wchodzącą zaś x_5 jako zmienną wychodzącą.

Nowym słownikiem będzie

$$\begin{array}{rlrrrr} x_0 & = 6 & -x_1 & +4x_2 & -x_3 & +x_5 \\ x_4 & = 11 & -2x_1 & +6x_2 & -4x_3 & +x_5 \\ x_6 & = 3 & -4x_1 & +6x_2 & -2x_3 & +x_5 \\ \hline w & = -6 & +x_1 & -4x_2 & +x_3 & -x_5 \end{array} \tag{3.22}$$

Teraz już (3.22) jest słownikiem dopuszczalnym (rozwiązanie bazowe jest rozwiązaniem dopuszczalnym: $x_0 = 6, \ x_1 = x_2 = x_3 = x_5 = 0, \ x_4 = 11, x_6 = 3, w = -6$).

Zauważmy, że także w ogólnym przypadku, zawsze będzie możliwy taki wybór zmiennej bazowej wychodzącej, aby pierwsza iteracja dała słownik dopuszczalny.

Wystarczy w tym celu tak wybrać x_{n+i_0} żeby b_{i_0} było najmniejsze spośród b_i, $i = 1, ..., m$.

Nową zmienną bazową wchodzącą może być x_1 lub x_3 (we wzorze na w w słowniku (3.22) współczynniki przy tych zmiennych są dodatnie i równe 1, przy pozostałych zmiennych współczynniki są ujemne).

Mamy teraz:

$$\begin{array}{lllllll} x_0 & \geq & 0 & \Rightarrow & x_1 & \leq & 6, \\ x_4 & \geq & 0 & \Rightarrow & x_1 & \leq & \frac{11}{2}, \\ x_6 & \geq & 0 & \Rightarrow & x_1 & \leq & \frac{3}{4}. \end{array}$$

Wobec tego, zmienną wychodzącą jest x_6. Wyliczamy więc x_1 z trzeciego z równań (3.22) i obliczamy nowe wzory na x_0, x_4 oraz w. Po elementarnych rachunkach otrzymamy trzeci słownik

$$\begin{array}{llllll} x_0 & = 5\frac{1}{4} & +2\frac{1}{2}x_2 & -\frac{1}{2}x_3 & +\frac{3}{4}x_5 & +\frac{1}{4}x_6 \\ x_1 & = \frac{3}{4} & +\frac{3}{2}x_2 & -\frac{1}{2}x_3 & +\frac{1}{4}x_5 & -\frac{1}{4}x_6 \\ x_4 & = 9\frac{1}{2} & +3x_2 & -3x_3 & +\frac{1}{2}x_5 & +\frac{1}{2}x_6 \\ \hline w & = -\frac{21}{4} & -\frac{5}{2}x_2 & +\frac{1}{2}x_3 & -\frac{3}{4}x_5 & -\frac{1}{4}x_6 \end{array} \tag{3.23}$$

Zmienną wchodzącą jest teraz x_3 a wychodzącą x_1. Wyliczamy więc x_3 z drugiego ze wzorów słownika (3.22), a następnie wstawiamy do pozostałych równań eliminując w ten sposób x_3.

Nowym słownikiem jest następujący układ równań:

$$\begin{array}{llllll} x_0 & = 4\frac{1}{2} & +x_1 & +x_2 & +\frac{1}{2}x_5 & +\frac{1}{2}x_6 \\ x_3 & = \frac{3}{2} & -2x_1 & +3x_2 & +\frac{1}{2}x_5 & -\frac{1}{2}x_6 \\ x_4 & = 5 & +6x_1 & -6x_2 & -x_5 & +2x_6 \\ \hline w & = -\frac{9}{2} & -x_1 & -x_2 & -\frac{1}{2}x_5 & -\frac{1}{2}x_6 \end{array} \tag{3.24}$$

Wartością maksymalną jaką może przyjąć w w problemie (3.20) jest $w = -\frac{9}{2}$, a więc liczba ujemna, nie zero. Oznacza to, że problem (3.17) jest sprzeczny. □

3.3.2 Czy sympleks może się zaciąć?

Pytanie postawione w tytule należy rozumieć tak: czy na każdym etapie algorytmu wiadomo co należy czynić dalej? Czy zawsze kolejny krok jest wykonalny?

Działanie algorytmu polega na wykonywaniu następujących kroków:

1. **Wybór zmiennej bazowej wchodzącej**
   Przypomnijmy, że nową zmienną wchodzącą jest zmienna $\bar{x}_{j_0}$ taka, że $\bar{c}_{j_0} > 0$. Jeśli istnieje więcej niż jedna taka zmienna, można, jak to robiliśmy w przykładach, wybrać tę dla której dodatnie $\bar{c}_{j_0}$ jest największe (z nadzieją – ale bez gwarancji – że taki wybór doprowadzi nas najszybciej do celu).
   Jeśi taki wybór jest niemożliwy to wszystkie współczynniki $\bar{c}_j$ dla $j = 1, ..., n$ we wzorze
   $$z = a + \bar{c}_1\bar{x}_1 + ... + \bar{c}_n\bar{x}_n$$
   są ujemne. *To zaś oznacza, że rozwiązanie dopuszczalne stowarzyszone ze słownikiem dopuszczalnym jest optymalne (w szczególności maksymalna wartość z jest równa a).*

2. **Wybór zmiennej wychodzącej**
   Jako zmienną wychodzącą przyjmujemy tę która daje najostrzejsze ograniczenie od góry na zmienną wchodzącą x_{j_0}.
   Na przykład, dla słownika dopuszczalnego
   $$\begin{array}{ccrrrr} x_3 & = & 5 & -x_1 & +x_2 & -x_4 \\ x_5 & = & 3 & -x_1 & -x_2 & +2x_4 \\ \hline z & = & 15 & +2x_1 & -x_2 & -x_4 \end{array}$$
   zmienną wchodzącą jest x_1. Nierówności $x_3 \geq 0$ i $x_5 \geq 0$ dają dla odpowiedniego rozwiązania dopuszczczalnego (tj. takiego w którym $x_2 = x_4 = 0$) następujące ograniczenia:
   $$x_3 \geq 0 \Rightarrow x_1 \leq 5, \qquad x_5 \geq 0 \Rightarrow x_1 \leq 3.$$
   Wybieramy zmienną wychodzącą x_5.
   Jeśli na wartości zmiennej wchodzącej nie otrzymujemy ograniczenia od góry, jak w przykładzie słownika
   $$\begin{array}{ccrrrr} x_3 & = & 5 & +x_1 & -x_2 & +x_4 \\ x_5 & = & 3 & +x_1 & -x_2 & +2x_4 \\ \hline z & = & 15 & +2x_1 & -x_2 & -x_4 \end{array}$$
   czyli jeśli współczynniki przy x_{j_0} są nieujemne, to wartość z jest oczywiście nieograniczona od góry. Nie ma rozwiązania maksymalnego, z może być dowolnie duże.

3. **Tworzenie nowego słownika**
   W tym kroku algorytmu tworzymy nowy słownik przy pomocy prostych operacji arytmetycznych.

Można więc z całą pewnością stwierdzić, że w każdym momencie swojego działanie algorytm albo poda rozwiązanie problemu (tzn. rozwiązanie optymalne lub informację, że problem jest sprzeczny lub nieograniczony), albo będzie wykonywał następne kroki. Czy oznacza to, że algorytm sympleks zawsze rozwiązuje każdy PPL? Przekonamy się wkrótce, że nie rozwiązaliśmy jeszcze wszstkich trudności.

3.3.3 Cykliczność

Zacznijmy od przykładu.

Przykład 3.3.2
Rozważmy słownik

$$\begin{array}{rcrrrr} x_4 & = & & -x_1 & -x_2 & +3x_3 \\ x_5 & = & 2 & +x_1 & -3x_2 & +x_3 \\ x_6 & = & & -2x_1 & +x_2 & -x_3 \\ \hline z & = & 3 & +2x_1 & -x_2 & -x_3 \end{array} \tag{3.25}$$

Rozwiązaniem bazowym tego słownika jest $x_1 = x_2 = x_3 = x_4 = x_6 = 0, x_5 = 2, z = 3$. Zmienną wchodzącą będzie x_1, zaś zmienną wychodzącą może być x_4 lub x_6. Wybierzmy na przykład x_4.

Nowym słownikiem będzie:

$$\begin{array}{rcrrrr} x_1 & = & & x_2 & -3x_3 & -x_4 \\ x_5 & = & 2 & +4x_2 & -2x_3 & -x_4 \\ x_6 & = & & -x_2 & +5x_3 & +2x_4 \\ \hline z & = & 3 & +x_2 & -7x_3 & -2x_4 \end{array} \tag{3.26}$$

Zauważmy więc, że na naszej operacji zamiany słownika nic nie zyskaliśmy w tym sensie, że wartość z nie wzrosła.

Sytuacja w której zastąpienie jednej ze zmiennych bazowych zmienną niebazową według podanych przez nas reguł nie spowoduje zwiększenia wartości z w rozwiązaniu bazowym może zajść tylko wtedy, gdy wartość jednej ze zmiennych w rozwiązaniu bazowym wynosi zero. W naszym przykładzie takimi zmiennymi były x_4 i x_6 dla słownika (3.25) i x_1 i x_6 dla słownika (3.26).

Definicja 3.3.1. *Mówimy, że* **rozwiązanie bazowe słownika jest zdegenerowane** *jeżeli co najmniej jedna zmienna bazowa w tym rozwiązaniu przyjmuje wartość zero. Odpowiedni słownik także nazywamy* **zdegenerowanym**.

Przykład 3.3.2 (cd)
Czytelnik zechce sprawdzić, że następnym (po (3.26)) słownikiem będzie

$$\begin{array}{rcrrrr} x_1 & = & & 2x_3 & +x_4 & -x_6 \\ x_2 & = & & 5x_3 & +2x_4 & -x_6 \\ x_5 & = & 2 & +18x_3 & +7x_4 & -4x_6 \\ \hline z & = & 3 & -2x_3 & & -x_6 \end{array}$$

a więc sympleks doprowadził do rozwiązania optymalnego PPL (3.25), rozwiązaniem optymalnym jest $x_1 = x_2 = x_3 = x_4 = x_6 = 0$, $x_5 = 2$, $z = 3$.

Ponieważ przy przejściu ze słownika (3.25) do (3.26) rozwiązanie nie polepszyło się (wartość z w tych słownikach jest taka sama) powstaje pytanie, *czy może zaistnieć sytuacja, w której po kilku zmianach słownika algorytm wróci do* *słownika który był już rozważany?*

Odpowiedź na to pytanie brzmi *tak!* Jest jasne, że jeśli konsekwentnie wybieramy zmienne wchodzące i wychodzące z jakąś regułą i na ten sam słownik wpadamy dwa razy, to te same słowniki będą sie pojawiały cyklicznie (nazwiemy to zjawisko **zapętlaniem się algorytmu**).

Zauważmy, że taka sytuacja może się pojawić tylko wtedy gdy pojawi się słownik zdegenerowany – jeśli nie, to w każdej iteracji rozwiązanie bazowe jest inne, różni się od wszystkich poprzednich chociażby wartością z.

W przykładzie 3.3.2 oba słowniki (3.25) i (3.26) są zdegenerowane. Czytelnik zechce sprawdzić, że przykład Chvátala [4] zaproponowany poniżej jako ćwiczenie 3.7.6 jest problemem PL, w którym po kilku odpowiednio przeprowadzonych iteracjach wracamy do wyjściowego słownika [1].

Z twierdzenia 3.2.1 wynika, że każde rozwiązanie bazowe jest rozwiązaniem każdego innego słownika (nie tylko tego, którego jest rozwiązaniem bazowym). Stąd wniosek, że wszystkich możliwych rozwiązań bazowych dla PPL jest co najwyżej tyle, na ile sposobów można wybrać n elementów (zmiennych bazowych) spośród $n+m$ elementów (tyle jest wszystkich zmiennych), a więc $\begin{pmatrix} m+n \\ n \end{pmatrix}$. Tak więc dla danego PPL istnieje skończona liczba słowników i jeśli w każdej iteracji otrzymujemy inny słownik, wówczas algorytm sympleks musi nas doprowadzić do rozwiązania problemu w skończonej liczbie iteracji (co najwyżej $\begin{pmatrix} m+n \\ n \end{pmatrix}$). Wykazaliśmy więc bardzo ważne twierdzenie:

Twierdzenie 3.3.1 *Jeśli algorytm simpleks nie zapętla się*[2]*, to rozwiązuje PPL w skończonej liczbie iteracji. Co więcej, jeśli problem jest niesprzeczny i ograniczony, to istnieje rozwiązanie bazowe, które jest równocześnie rozwiązaniem optymalnym.* ■

Wśród metod unikania cykliczności elegancją [3] wyróżnia się **reguła R.G. Blanda** [2] zwana także **regułą najmniejszego indeksu**.

Reguła Blanda.

1. Wybierz jako zmienną wchodzącą tę zmienną która:

   (i) ma dodatni współczynnik we wzorze na z,

[1] Można wykazać, że PPL w którym występuje zjawisko cykliczności musi mieć co najmniej 7 zmiennych decyzyjnych. Przykład Chvátala jest więc, w pewnym sensie, najmniejszym PPL który się zapętla.

[2] Inaczej: jeśli nie występuje zjawisko cykliczności.

[3] elegancja (*mat.*) = prostota (*wulg.*)

(ii) spośród wszystkich zmiennych o dodatnim współczynniku we wzorze na z ma najmniejszy indeks.

2. Wybierz jako zmienną wychodzącą tę która:

(iii) powoduje najsilniejsze ograniczenie od góry dla zmiennej wchodzącej,

(iv) spośród wszystkich kandydatów na zmienną wychodzącą (tj. zmiennych bazowych spełniających (iii)) ma najmniejszy indeks.

Regułę Blanda można podać nieco bardziej formalnie w sposób następujący. Niech będzie dany słownik

$$\begin{array}{ccccl} x_i & = & b_i & - & \sum_{j \notin B} a_{ij} x_j \quad (i \in B) \\ \hline z & = & z_0 & + & \sum_{j \notin B} c_j x_j \end{array} \tag{3.27}$$

(w słowniku (3.27) B jest zbiorem indeksów zmiennych bazowych).

- Wybierz jako zmienną wchodzącą x_{j_0} dla której

  (a) $c_{j_0} > 0$,

  (b) $j_0 = \min\{j : j \notin B \wedge c_j > 0\}$.

- Wybierz jako zmienną wychodzącą x_{i_0} dla której

  (a) $a_{i_0 j_0} > 0$,

  (b) $\frac{b_{i_0}}{a_{i_0 j_0}} = \min\{\frac{b_i}{a_{ij_0}} : a_{ij_0} > 0, i = 1, ..., m\} = a$,

  (c) $i_0 = \min\{i : i \in B \wedge \frac{b_i}{a_{ij_0}} = a\}$.

W porównaniu z metodą stosowaną do tej pory różnica pojawia się w punkcie (ii) – stosujemy tę regułę zamiast wybierać zmienną wchodzącą o największym współczynniku (dodatnim) w funkcji celu.

Przykład 3.3.3
Dla słownika

$$\begin{array}{llrrr} x_4 & = 4 & +x_1 & +2x_2 & -x_3 \\ x_5 & = 2 & -2x_1 & -2x_2 & +x_3 \\ x_6 & = 3 & -x_1 & -3x_2 & +2x_3 \\ \hline z & = 1 & -x_1 & +2x_2 & +3x_3 \end{array}$$

postępując według reguły Blanda wybierzemy x_2 jako zmienną wchodzącą i x_5 jako zmienną wychodzącą.

Twierdzenie 3.3.2 (Bland [2]). *Jeśli wybory zmiennych wchodzących i wychodzących w algorytmie simpleks dokonywane są według reguły Blanda, to algorytm kończy się po skończonej liczbie iteracji.*

Idea dowodu twierdzenia Blanda zaczerpnięta jest z [4].

Dowód. Ze względu na twierdzenie 3.3.1 wystarczy wykazać, że jeżeli zmienne wchodząca i wychodząca będą wybierane według reguły Blanda to nie może wystąpić cykliczność.

Dla dowodu nie wprost przypuśćmy, że zmienne są wybierane regułą Blanda, a jednak algorytm zapętla się. Oznacza to, że poczynając od pewnego słownika S_1 w kolejnych iteracjach otrzymujemy k–elementowy ciąg słowników $S_1, S_2, ..., S_{k-1}, S_k$ taki, że $k > 1$ i $S_k = S_1$. Oczywiście wszystkie słowniki $S_1, ..., S_{k-1}$ muszą być zdegenerowane.

Zmienną będziemy nazywali *błędną*[4] jeżeli dla pewnych słowników spośród $S_1, ..., S_{k-1}$ jest zmienną bazową a dla innych niebazową.

Przypuśćmy, że x_t jest zmienną błędną o największym indeksie t. Niech S będzie jednym ze słowników $S_1, S_2, ..., S_{k-1}, S_k$, takim, że x_t jest zmienną bazową wychodzącą słownika S i niech x_s będzie zmienną wchodzącą S.

Inaczej mówiąc: x_t i x_s są takie, że

– x_t jest zmienną bazową S i nie jest zmienną bazową słownika następnego,

– x_s jest zmienną niebazową słownika S i zmienną bazową słownika następnego.

Oznaczmy przez B zbiór indeksów zmiennych bazowych słownika S i powiedzmy, że S jest postaci

$$\begin{array}{ccccl} x_i & = & b_i & - & \sum_{i \notin B} a_{ij} x_j, \quad i \in B \\ \hline z & = & w & + & \sum_{j \notin B} c_j x_j \end{array} \tag{3.28}$$

Skoro $S_1, ..., S_{k-1}, S_k$ jest cyklem, tzn. $S_k = S_1$, musi istnieć inny słownik, nazwijmy go S^* taki, że x_t jest zmienną wchodzącą S^*. Z tego samego powodu wartość zmiennej z (funkcji celu) w rozwiązaniu bazowym każdego słownika $S_1, ..., S_{k-1}, S_k$ jest taka sama. Tak więc wzór na z w słowniku S^* jest postaci

$$z = w + \sum_{j=1}^{m+n} c_j^* x_j \tag{3.29}$$

przy czym $c_j^* = 0$ dla $j \in B^*$, gdzie B^* jest zbiorem indeksów zmiennych bazowych słownika S^*. Na mocy twierdzenia 3.2.1, każde rozwiązanie słownika S jest także rozwiązaniem słownika S^*. Jednym z rozwiązań słownika S jest, jak łatwo zauważyć:

$x_j = 0$ dla wszystkich zmiennych niebazowych S , poza x_s (czyli dla $j \notin B, j \neq s$),

$$\left.\begin{array}{l} x_s = y \\ x_i = b_i - a_{is} y \quad \text{dla} \quad i \in B \\ z = w + c_s y \end{array}\right\} \text{dla dowolnego } y.$$

[4] Tu określenie *błędna* nie pochodzi od słowa *błąd* a raczej od błądzić, w sensie pojawiania się i znikania jak *błędny rycerz* lub *błędny ognik*.

Wstawiając to rozwiązanie do wzoru (3.29) otrzymamy

$$w + c_s y = w + c_s^* y + \sum_{i \in B} c_i^* (b_i - a_{is} y)$$

Stąd

$$(c_s - c_s^* + \sum_{i \in B} c_i^* a_{is}) y = \sum_{i \in B} c_i^* b_i$$

dla każdego $y \in \mathbf{R}$. Wobec dowolności y mamy

$$c_s - c_s^* + \sum_{i \in B} c_i^* a_{is} = 0 \tag{3.30}$$

Zauważmy następujące fakty:

– Ponieważ x_s jest zmienną wchodzącą słownika S zachodzi $c_s > 0$.

– Przypomnijmy, że x_t jest błędną zmienną o największym indeksie. Stąd oczywiście wynika, że $s < t$ (x_s jest także zmienną błędną).

– Skoro x_s nie jest zmienną wchodzącą w słowniku S^*, $s < t$, i wybór zmiennej wchodzącej następuje według reguły Blanda, zachodzi $c_s^* \leq 0$.

Ze wzoru (3.30) wynika więc, że

$$c_{i_0}^* a_{i_0 s} < 0 \quad \text{dla pewnego } \; i_0 \in B \tag{3.31}$$

Oczywiście x_{i_0} jest zmienną bazową w słowniku S (bo $i_0 \in B$) i zmienną niebazową w słowniku S^* ($c_{i_0}^* \neq 0$, w przeciwnym przypadku nie moglibyśmy mieć $c_{i_0}^* a_{i_0 s} < 0$).
x_{i_0} jest więc zmienną błędną i wobec tego $i_0 \leq t$. Co więcej: $i_0 \neq t$, bowiem $a_{ts} > 0$ skoro x_t jest zmienną wychodzącą a x_s zmienną wchodzącą słownika S. Pamiętamy, że x_t jest zmienną wchodzącą słownika S^*. Wobec tego $c_t^* > 0$ i w konsekwencji $c_t^* a_{ts} > 0$. Ostatecznie $i_0 < t$.

x_{i_0} nie jest jednak zmienną wchodzącą słownika S^* (jest nią przecież x_t, a właśnie wykazaliśmy, że $i_0 \neq t$), a ponieważ zmienną wchodzącą wybieramy według reguły Blanda, musi być $c_{i_0}^* \leq 0$. Ze wzoru (3.31) mamy więc $a_{i_0 s} > 0$.

Z faktu, że x_{i_0} jest błędną zmienną oraz, że wartość z w rozwiązaniu bazowym jest stale równa w dla wszystkich słowników $S_1, ..., S_k$ wynika, że $b_{i_0} = 0$. Tak więc x_{i_0} była zmienną, która powinna według reguły Blanda, być zmienną wychodzącą ze słownika S ($i_0 < t$). Ta sprzeczność kończy dowód twierdzenia Blanda.

■

3.4 Ile jest rozwiązań optymalnych?

Problemowi ilości rozwiązań optymalnych nie będziemy poświęcali wiele miejsca i czasu. Niemniej często można łatwo stwierdzić, że rozwiązanie jest jedyne. Najlepiej wyjaśni to poniższy przykład.

Przykład 3.4.1
Niech będzie dany problem PL

$$\begin{array}{rrrrcr} & x_1 & -x_2 & +2x_3 & \leq & 1 \\ & x_1 & +2x_2 & +x_3 & \leq & 4 \\ & x_1 & -x_2 & -x_3 & \leq & 2 \\ & & & x_1,x_2,x_3 & \geq & 0 \\ \hline z & =x_1 & -x_2 & +x_3 & \to & \max \end{array} \tag{3.32}$$

Kolejnymi słownikami dla tego problemu będą:

$$\begin{array}{rcrrrrcl} x_4 & = & 1 & -x_1 & +x_2 & -2x_3 & & \\ x_5 & = & 4 & -x_1 & -2x_2 & -x_3 & & \\ x_6 & = & 2 & -x_1 & +x_2 & +x_3 & & \\ \hline z & = & & x_1 & -x_2 & +x_3 & \to & \max \end{array} \tag{3.33}$$

$$\begin{array}{rcrrrrcl} x_1 & = & 1 & +x_2 & -2x_3 & -x_4 & & \\ x_5 & = & 3 & -3x_2 & +x_3 & +x_4 & & \\ x_6 & = & 1 & & +3x_3 & +x_4 & & \\ \hline z & = & 1 & & -x_3 & -x_4 & \to & \max \end{array} \tag{3.34}$$

Zauważmy, że wartość maksymalna $z = 1$ jest przyjęta dla więcej niż jednego wektora $\mathbf{x}$. Rzeczywiście, we wzorze na z słownika (3.34), a więc ostatniego słownika w metodzie simpleks dla naszego problemu, nie występuje zmienna bazowa x_2. **Jeżeli we wzorze na z w ostatnim słowniku występują wszystkie zmienne niebazowe, to rozwiązanie jest jedyne: zmienne niebazowe muszą w rozwiązaniu maksymalizującym z być równe zeru, zaś zmienne bazowe wyznaczone są jednoznacznie przez odpowiednie równania słownika**. W naszym przykładzie, aby z było równe 1 musi zachodzić $x_3 = 0$ oraz $x_4 = 0$, natomiast x_2 niekoniecznie jest równe zero. Na przykład dla $x_3 = x_4 = 0$, $x_2 = t \in< 0, 1 >$, $x_1 = 1 + t$, $x_5 = 3 - 3t$, $x_6 = 1$ wartość z jest także równa wartości maksymalnej 1. Dla problemu (3.40) oznacza to, jak łatwo stwierdzić, że zbiorem wszystkich rozwiązań jest zbiór punktów odcinka $\{(x_1, x_2, x_3) : x_1 = 1 + t, x_2 = t, x_3 = 0, t \in< 0, 1 >\}$.

3.5 Skuteczność algorytmu sympleks

Panuje powszechna, jak się wydaje, zgoda co do tego, że sympleks jest w praktyce bardzo skutecznym i szybkim algorytmem, pozwalającym rozwiązywać problemy w których występują setki czy nawet tysiące zmiennych. Dowodzą tego różnego rodzaju testy (por. [4, 10, 14, 17, 20, 22]). A jednak, okazuje się, że jeśli dobrać wystarczająco złośliwie przykład, to liczba iteracji może wynosić nawet $2^n - 1$, gdzie n oznacza ilość zmiennych. Oznacza to konieczność utworzenia 2^n słowników. To zaś z kolei znaczy, że sympleks jest *teoretycznie* beznadziejny: licząc 1 mikrosekundę na utworzenie słownika (przy dużej liczbie zmiennych to kompletnie nierealne, nawet przy bardzo szybkim komputerze), oznaczałoby to konieczność oczekiwania trudno wyobrażalnej liczby lat na obliczenie zadania o 1000 zmiennych[5].

Takim *złośliwym* przykładem jest zadanie Klee–Minty'ego (por. [4, 20]) które przedstawione poniżej (lemat 3.5.2) posłuży do wykazania następującego twierdzenia.

Twierdzenie 3.5.1 *Algorytm sympleks nie jest algorytmem wielomianowym.*

Dowód twierdzenia 3.5.1 wynika z lematu:

Lemat 3.5.2 *Algorytm sympleks rozwiązuje problem PL_n:*

$$\begin{array}{rrcccccr}
 & x_1 & & & & & & \le 5 \\
 & 4x_1+ & x_2 & & & & & \le 25 \\
 & 8x_1+ & 4x_2 & + & x_3 & & & \le 125 \\
 & & & & \dots & & & \\
 & 2^n x_1+ & 2^{n-1}x_2 & + & \dots & + & 4x_{n-1} \quad + & x_n \le 5^n \\
\hline
z = & 2^{n-1}x_1+ & 2^{n-2}x_2 & + & \dots & + & 2x_{n-1} \quad + & x_n \to max
\end{array} \tag{3.35}$$

w $2^n - 1$ iteracjach (tj. 2^n słownikach).

Dowód lematu 3.5.2. Udowodnijmy najpierw dwa proste fakty.

Fakt 1. *$z = 5^n$ jest wartością optymalną, zaś*

$$x_1 = \dots = x_{n-1} = 0, \ \ x_n = 5^n \tag{3.36}$$

rozwiązaniem optymalnym problemu (3.35).

Rzeczywiście, pierwszy słownik problemu (3.35) można zapisać następująco [6]:

[5]Policz sam jeśli nie wierzysz!

[6]Wyjątkowo w tym przypadku, wygodniej nam będzie oznaczać zmienne sztuczne przez y_i zamiast, jak to było zawsze do tej pory, przez x_{n+i}.

$$\begin{array}{llll}
y_1 = 5 - x_1 & & & \\
y_2 = 25 - 4x_1 - x_2 & & & \\
y_3 = 125 - 8x_1 - 4x_2 - x_3 & & & \\
... & & ... & \\
y_{n-1} = 5^{n-1} - 2^{n-1}x_1 - 2^{n-2}x_2 - & ... & - & x_{n-1} \\
y_n = 5^n - 2^n x_1 - 2^{n-1}x_2 - & ... & - & 4x_{n-1} - x_n \\
\hline
z = 2^{n-1}x_1 + 2^{n-2}x_2 + & ... & + & 2x_{n-1} + x_n
\end{array} \tag{3.37}$$

Słownikiem, w którym zmiennymi niebazowymi są $x_1, ..., x_{n-1}, y_n$, zaś zmiennymi bazowymi $y_1, ..., y_{n-1}, x_n$ jest

$$\begin{array}{llll}
y_1 = 5 - x_1 & & & \\
y_2 = 25 - 4x_1 - x_2 & & & \\
y_3 = 125 - 8x_1 - 4x_2 - x_3 & & & \\
... & & ... & \\
y_{n-1} = 5^{n-1} - 2^{n-1}x_2^{n-2} - & ... & - & x_{n-1} \\
x_n = 5^n - 2^n x_1 - 2^{n-1}x_2 - & ... & - & 4x_{n-1} - y_n \\
\hline
z = 5^n + \bar{c}_1 x_1 + \bar{c}_2 x_2 + & ... & + & \bar{c}_{n-1}x_{n-1} - y_n
\end{array} \tag{3.38}$$

gdzie $\bar{c}_i = 2^{n-i} - 2^{n-i+1}$ dla każdego $i = 1, ..., n-1$. Jest oczywiste, że słownik (3.38) jest ostatnim słownikiem algorytmu sympleks dla (3.35), zaś (3.36) optymalnym rozwiązaniem (współczynniki $\bar{c}_i$ są oczywiście ujemne).□

Fakt 2. *W każdym słowniku problemu (3.35), dla każdego $i \in \{1, ..., n\}$ dokładnie jedna ze zmiennych x_i, y_i jest zmienną bazową.*
Gdyby tak nie było, to w w pewnym słowniku dla pewnego i_0 zarówno x_{i_0} jak i y_{i_0} byłyby zmiennymi niebazowymi.
Dla odpowiadającego temu słownikowi bazowego rozwiązania dopuszczalnego $(\bar{x}_1, ..., \bar{x}_n, \bar{y}_1, ..., \bar{y}_n)$ mamy oczywiście $\bar{x}_{i_0} = 0$ i $\bar{y}_{i_0} = 0$. Napiszmy równania $i_0 - 1$ i i_0 słownika (3.37).

$$2^{i_0-1}x_1 + 2^{i_0-2}x_2 + ... + 4x_{i_0-2} + x_{i_0-1} + y_{i_0-1} = 5^{i_0-1} \tag{3.39}$$

$$2^{i_0}x_1 + 2^{i_0-1}x_2 + ... + 8x_{i_0-2} + 4x_{i_0-1} + x_{i_0} + y_{i_0} = 5^{i_0} \tag{3.40}$$

Dla naszego rozwiązania bazowego w którym $x_{i_0} = 0$ i $y_{i_0} = 0$ mamy więc z (3.40)

$$2^{i_0}\bar{x}_1 + 2^{i_0-1}\bar{x}_2 + ... + 4\bar{x}_{i_0-1} = 5^{i_0}$$

Z drugiej zaś strony, korzystając z (3.40), (3.39), nierówności $2\bar{x}_{i_0-1} \leq 2 \cdot 5^{i_0-1}$ wynikającej w sposób banalny z (3.39) i pamiętając, że wszystkie współrzędne rozwiązania są nieujemne, otrzymamy

$$5^{i_0} = 2^{i_0}\bar{x}_1 + 2^{i_0-1}\bar{x}_2 + ... + 8\bar{x}_{i_0-2} + 4\bar{x}_{i_0-1} =$$

$$= 2(2^{i_0-1}\bar{x}_1 + 2^{i_0-2}\bar{x}_2 + ... + 4\bar{x}_{i_0-2} + \bar{x}_{i_0-1}) + \\ + 2\bar{x}_{i_0-1} \leq 2 \cdot 5^{i_0-1} + 2 \cdot 5^{i_0-1} = 4 \cdot 5^{i_0-1}$$

– sprzeczność która kończy dowód faktu 2.□

Reszta dowodu lematu 3.5.2 przez indukcję ze względu na n.
Dla $n = 1$ problem jest banalny: PL_n ma postać

$$\begin{array}{c} x_1 \leq 5 \\ \hline z = x_1 \end{array}$$

i rozwiązuje się w 2 słownikach: $\begin{array}{rr} y_1 = & 5 - x_1 \\ \hline z = & x_1 \end{array}$, $\begin{array}{rr} x_1 = & 5 - y_1 \\ \hline z = & 5 - y_1 \end{array}$.
Pierwszy słownik dla PL_n $(n \geq 2)$ jest postaci

$$\begin{array}{rrrlll} y_1 = & 5- & x_1 & & & \\ y_2 = & 25- & 4x_1 & -x_2 & & \\ y_3 = & 125- & 8x_1 & -4x_2 - x_3 & & \\ & & & & ... & \\ y_n = & 5^n- & 2^n x_1 & -2^{n-1}x_2- & ... - 4x_{n-1} & -x_n \\ \hline z = & & 2^{n-1}x_1 & +2^{n-2}x_2+ & ... + 2x_{n-1} & +x_n \end{array} \tag{3.41}$$

a ostatni (o rozwiązaniu bazowym $\bar{x}_1 = ... = \bar{x}_{n-1} = 0, \bar{x}_n = 5^n, \bar{z} = 5^n$)

$$\begin{array}{rrrrrrlll} y_1 = & 5 & -x_1 & & & & & & \\ y_2 = & 25 & -4x_1 & -x_2 & & & & & \\ y_3 = & 125 & -8x_1 & -4x_2 & -x_3 & & & & \\ & & & ... & & & & & \\ y_{n-1} = & 5^{n-1} & -2^{n-1}x_1 & -2^{n-2}x_2 & & - & ... - 4x_{n-2} & -x_{n-1} & \\ x_n = & 5^n & -2^n x_1 & -2^{n-1}x_2 & & - & ... - 8x_{n-2} & -4x_{n-1} & -y_n \\ \hline z = & 5^n & -2^{n-1}x_1 & -2^{n-2}x_2 & & - & ... - 4x_{n-2} & -2x_{n-1} & -y_n \end{array}$$

Napiszmy teraz pierwszy słownik dla PL_{n+1}

$$\begin{array}{l} y_1 = 5 - x_1 \\ y_2 = 25 - 4x_1 - x_2 \\ ... \\ y_{n-1} = 5^{n-1} - 2^{n-1}x_1 - 2^{n-2}x_2 - ... - 4x_{n-2} - x_{n-1} \\ y_n = 5^n - 2^n x_1 - 2^{n-1}x_2 - ... - 8x_{n-2} - 4x_{n-1} - x_n \\ y_{n+1} = 5^{n+1} - 2^{n+1}x_1 - 2^n x_2 - ... - 2^n x_{n-2} - 8x_{n-1} - 4x_n - x_{n+1} \\ \hline z = 2^n x_1 + 2^{n-1}x_2 + ... + 2^3 x_{n-2} + 4x_{n-1} + 2x_n + x_{n+1} \end{array}$$

Słownik ten różni się od (3.41) tym, że dodaliśmy do niego jeden wiersz ograniczeń, a funkcję celu najpierw pomnożyliśmy przez 2, a potem dodaliśmy do niej x_{n+1}. Wszystkie współczynniki (a także wyrazy wolne) we wszystkich słownikach są całkowite. Rzeczywiście: skoro, jak stwierdziliśmy wyżej, dla każdego i dokładnie jedna ze zmiennych x_i, y_i jest bazowa, za każdym razem zmienna wchodząca (x_i lub y_i) jest obliczana z $i-$tego równania w którym występuje zawsze ze współczynnikiem -1. Tak więc, dopóki we wzorze na funkcję z któregokolwiek słownika problemu PL_{n+1} występuje któraś ze zmiennych x_i, $i < n+1$, ze współczynnikiem dodatnim, zmienną wchodzącą jest jedna z nich (ta której współczynnik w funkcji z jest największy). Tak więc algorytm sympleks będzie funkcjonował na pierwszych n wierszach ograniczeń dopóty, dopóki w funkcji celu wszystkie współczynniki przy zmiennych $x_1, ..., x_n, y_1, ..., y_n$ będą niedodatnie.

W ostatnim z tych słowników podstawiamy

$$x_n = 5^n - 2^n x_1 - 2^{n-1} x_2 - ... - 2^3 x_{n-2} - 2^2 x_{n-1} - y_n$$

do

$$y_{n+1} = 5^{n+1} - 2^{n+1} x_1 - 2^n x_2 - ... - 2^4 x_{n-2} - 2^3 x_{n-1} - 4x_n - x_{n+1}$$

i do

$$z = 2^n x_1 + 2^{n-1} x_2 + ... + 4x_{n-1} + 2x_n + x_{n+1}$$

otrzymując

$$\begin{array}{ll}
y_1 & = 5 - x_1 \\
y_2 & = 5^2 - 4x_1 - x_2 \\
 & ... \\
y_{n-1} & = 5^{n-1} - 2^{n-1} x_1 - 2^{n-2} x_2 - ... - 4x_{n-2} - x_{n-1} \\
x_n & = 5^n - 2^n x_1 - 2^{n-1} x_2 - ... - 8x_{n-2} - 4x_{n-1} - y_n \\
y_{n+1} & = 5^n + 2^{n+1} x_1 + 2^n x_2 + ... + 2^4 x_{n-2} + 2^3 x_{n-1} + 4y_n - x_{n+1} \\
\hline
z & = 2 \cdot 5^n - 2^n x_1 - 2^{n-1} x_2 - 2^{n-2} x_3 - ... - 4x_{n-1} - 2y_n + x_{n+1}
\end{array}$$

Zmienną wchodzącą teraz jest $x_{n+1} = 5^n + 2^{n+1} x_1 + 2^n x_2 + ... + 2^4 x_{n-2} + 2^3 x_{n-1} + 4y_n - y_{n+1}$. Tak więc następnym, $2^n + 1-$szym słownikiem PL_{n+1} jest słownik następujący

$$\begin{array}{ll}
y_1 & = 5 - x_1 \\
y_2 & = 5^2 - 4x_1 - x_2 \\
 & ... \\
y_{n-1} & = 5^{n-1} - 2^{n-1} x_1 - 2^{n-2} x_2 - ... - 4x_{n-2} - x_{n-1} \\
x_n & = 5^n - 2^n x_1 - 2^{n-1} x_2 - ... - 8x_{n-2} - 4x_{n-1} - y_n \\
x_{n+1} & = 5^{n+1} + 2^{n+1} x_1 + 2^n x_2 + ... + 2^4 x_{n-2} + 2^3 x_{n-1} + 4y_n - y_{n+1} \\
\hline
z & = 3 \cdot 5^n + 2^n x_1 + 2^{n-1} x_2 + ... + 2^3 x_{n-2} + 2^2 x_{n-1} + 2y_n - y_{n+1}
\end{array}$$

Otrzymaliśmy problem w którym następne iteracje będą polegały na minimalizacji funkcji

$$z = 3 \cdot 5^n + 2^n x_1 + 2^{n-1} x_2 + ... + 2^3 x_{n-2} + 2^2 x_{n-1} + 2y_n. \tag{3.42}$$

Operacje na wierszach ograniczeń dotyczyć będą wyłącznie pierwszych n wierszy. Funkcja z dana wzorem (3.42) równa jest stałej $3 \cdot 5^n$ plus 2 razy funkcja celu PL_n, potrzebować będziemy więc jeszcze 2^n słowników, czyli w sumie $2 \cdot 2^n = 2^{n+1}$ słowników, co należało udowodnić. ■

3.6 Wyjaśnienie nazwy *sympleks*

Pozostaje wyjaśnić skąd wzięła się nazwa *algorytm sympleks*? Sympleksem w matematyce nazywa się obwiednie wypukłą zbioru $n+1$ elementowego w przestrzeni rzeczywistej n wymiarowej (odcinek w $\mathbf{R}$, trójkąt w $\mathbf{R^2}$, czworościan w $\mathbf{R^3}$). Metoda (algorytm) sympleks polega zaś na sprawdzaniu kolejnych wierzchołków wielościanów (w $\mathbf{R^n}$ – por. rozdział 7). A więc *poruszamy się* po krawędziach od wierzchołka do sąsiedniego wierzchołka n–wymiarowego wielościanu, zawsze w pewnym sympleksie zwiększając, o ile to możliwe, wartość funkcji celu.

3.7 ćwiczenia

Ćwiczenie 3.7.1
Rozwiąż następujące problemy:

$$\begin{array}{rrl} x_1 & +x_2 & \le 3 \\ 2x_1 & -x_2 & \le 4 \\ & x_1, x_2 & \ge 0 \\ \hline 2x_1 & +x_2 & \to \ \max \end{array} \qquad \begin{array}{rrrll} 2x_1 & +4x_2 & +x_3 & \le 6 & \\ x_1 & -4x_2 & +x_3 & \le 4 & \\ 2x_1 & +2x_2 & +x_3 & \le 2 & \\ & & x_j & \ge 0 & (j=1,2,3) \\ \hline x_1 & +2x_2 & +3x_3 & \to & \max \end{array}$$

Ćwiczenie 3.7.2
Posługując sie tabelami simpleksowymi rozwiąż przykład 3.1.2

Ćwiczenie 3.7.3
Posługując sie tabelami simpleksowymi rozwiąż PPL:

$$\begin{array}{rcrcrcl} x_1 & - & 2x_2 & + & 3x_3 & \le & 10 \\ 2x_1 & + & x_2 & + & x_3 & \le & 12 \\ x_1 & + & 3x_2 & + & x_3 & \le & 9 \\ \hline z = x_1 & + & 3x_2 & + & 4x_3 & \to & \max \\ & & & & x_i & \ge & 0 \quad i=1,2,3 \end{array}$$

Ćwiczenie 3.7.4
Rozwiąż następujące PPL:

$$\begin{array}{rrl} -3x_1 & +x_2 & \le -3 \\ x_1 & +x_2 & \ge 1 \\ & x_1, x_2 & \ge 0 \\ \hline x_1 & +3x_2 & \to \ \max \end{array}$$

$$\begin{array}{rrrl} -x_1 & -x_2 & +x_3 & \leq -1 \\ 2x_1 & +x_2 & +x_3 & \leq 8 \\ x_1 & +x_2 & & \leq 10 \\ \hline x_1 & +2x_2 & +x_3 & \to \max \end{array}$$

$x_i \geq 0,\ i = 1, 2, 3.$

Ćwiczenie 3.7.5
Sprawdź, że PPL

$$\begin{array}{rrrl} 2x_1 & -x_2 & +x_3 & \leq 12 \\ x_1 & +2x_2 & +x_3 & \leq 10 \\ x_1 & +x_2 & -x_3 & \geq 4 \\ \hline 2x_1 & +x_2 & -3x_3 & \to \max \end{array}$$

$x_i \geq 0,\ i = 1, 2, 3.$
jest niesprzeczny. Wskaż dla tego problemu bazowe rozwiązanie dopuszczalne.

Ćwiczenie 3.7.6 (przykład Chvatála)
Zastosuj algorytm simpleks do PPL

$$\begin{array}{rcrrrrr} x_5 & = & & -0,5x_1 & +5,5x_2 & +2,5x_3 & -9x_4 \\ x_6 & = & & -0,5x_1 & +1,5x_2 & +0,5x_3 & -x_4 \\ x_7 & = & 1 & -x_1 & & & \\ \hline z & = & & 10x_1 & -57x_2 & -9x_3 & -24x_4 \end{array} \tag{3.43}$$

wybierając za każdym razem

- zmienną wchodzącą o największym współczynniku we wzorze na z,
- zmienną wychodzącą o najmniejszym indeksie spośród tych które dają najostrzejsze ograniczenie od góry na zmienną wchodzącą (np pierwszą zmienną wychodzącą będzie x_5).

Ćwiczenie 3.7.7
Wykaż, że każda wypukła kombinacja rozwiązań optymalnych PPL jest rozwiązaniem tego problemu (kombinacja wypukła zdefiniowana jest w podrozdziale 7.2).

Ćwiczenie 3.7.8
Algorytmem sympleks rozwiąż zadanie Klee–Minty'ego dla $n = 3$ (a jeśli masz cierpliwość i potrzebę lepszego zrozumienia dowodu twierdzenia 3.5.1 to dla $n = 4$). Ile otrzymasz iteracji a ile słowników w każdym z tych przypadków (por. przykład 7.1.2)?

Rozdział 4

Dualizm

W tym rozdziale zmienimy metodę prezentacji. Dotychczas każde nowe zagdnienie rozpoczynaliśmy od podania przykładów których celem było uzasadnienie tego co robiliśmy później. Tym razem najpierw podamy ogólną zasadę dualności w programowaniu liniowym. Dopiero później zostanie podane kryterium optymalności które można otrzymać dzieki tej zasadzie a następnie interpretację ekonomiczną.

4.1 Problem dualny programowania liniowego

Definicja 4.1.1 *Niech będzie dany PPL*

$$\begin{array}{rcl} \sum_{j=1}^{n} a_{ij}x_j & \leq & b_i \quad (i=1,...,m) \\ x_j & \geq & 0 \quad (j=1,...,n) \\ \hline z & = & \sum_{j=1}^{n} c_j x_j \to max \end{array} \tag{4.1}$$

PPL

$$\begin{array}{rcl} \sum_{i=1}^{m} a_{ij}y_j & \geq & c_j \quad (j=1,...,n) \\ y_i & \geq & 0 \quad (i=1,...,m) \\ \hline w & = & \sum_{i=1}^{m} b_i y_i \to min \end{array} \tag{4.2}$$

nazywamy **problemem dualnym** *problemu (4.1). Problem (4.1) nazywamy* **problemem prymalnym**.

Z łatwością można się przekonać, że po sprowadzeniu problemu (4.2) do postaci standardowej otrzymamy PPL którego problemem dualnym jest wyjściowy PPL (4.1). Tak więc prawdziwy jest natychmiast następujący wniosek.

Wniosek 4.1.1 *Problemem dualnym do problemu dualnego do PPL jest wyjściowy PPL.*

Przykład 4.1.1 Problemem dualnym do

$$
\begin{array}{rrrcl}
2x_1 & +x_2 & +5x_3 & \leq & 2 \\
x_1 & -2x_2 & +x_3 & \leq & 1 \\
x_1 & +4x_2 & -x_3 & \leq & 3 \\
 & & x_i & \geq & 0 \quad (i = 1, 2, 3) \\
\hline
2x_1 & -x_2 & +3x_3 & \rightarrow & \max
\end{array}
\tag{4.3}
$$

jest

$$
\begin{array}{rrrcl}
2y_1 & +y_2 & +y_3 & \geq & 2 \\
y_1 & -2y_2 & +4y_3 & \geq & -1 \\
5y_1 & +y_2 & -y_3 & \geq & 3 \\
 & & y_i & \geq & 0 \quad (i = 1, 2, 3) \\
\hline
2y_1 & +y_2 & +3y_3 & \rightarrow & \min
\end{array}
\tag{4.4}
$$

Następne twierdzenie bywa nazywane **słabą zasadą dualności**.

Twierdzenie 4.1.2 *Niech (4.1) będzie problemem prymalnym a (4.2) problemem do niego dualnym. Jeżeli* $\bar{x}_1, ..., \bar{x}_n$ *oraz* $\bar{y}_1, ..., \bar{y}_m$ *są rozwiązaniami dopuszczalnymi, odpowiednio, problemów (4.1) i (4.2) to*

$$
\sum_{j=1}^{n} c_j \bar{x}_j \leq \sum_{i=1}^{m} b_i \bar{y}_i
$$

Dowód twierdzenia 4.1.2 jest wyjątkowo prosty i mieści się w jednej linijce:

$$
\sum_{j=1}^{n} c_j \bar{x}_j \leq \sum_{j=1}^{n} \left(\sum_{i=1}^{m} a_{ij} \bar{y}_i \right) \bar{x}_j = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} a_{ij} \bar{x}_j \right) \bar{y}_i \leq \sum_{i=1}^{m} b_i \bar{y}_i \blacksquare
$$

Przykład 4.1.1 (c.d.) $y_1 = 2, y_2 = 1, y_3 = 0$ *jest rozwiązaniem dopuszczalnym (4.4) o wartości funkcji celu* $w = 5$*. Stąd i z twierdzenia 4.1.2 wynikają dwa wnioski:*

1. *problem (4.3) jest ograniczony,*
2. *wartość maksymalna funkcji celu jest co najwyżej równa* 5*.*

Oczywisty jest następujący wniosek z twierdzenia 4.1.2.

Wniosek 4.1.3 *Jeżeli problem prymalny jest nieograniczony, to problem do niego dualny jest sprzeczny.*

Przykład 4.1.2 Może się zdarzyć, że oba problemy: prymalny i dualny są sprzeczne. Tak jest na przykład dla problemu prymalnego:

$$
\begin{array}{rrcr}
2x_1 & -x_2 & \leq & 2 \\
-2x_1 & +x_2 & \leq & -3 \\
\hline
3x_1 & -x_2 & \rightarrow & \max
\end{array}
$$

Z twierdzenia 4.1.2 wynika, że gdyby udało nam się znaleźć takie rozwiązanie dopuszczalne $x_1, ..., x_n, z$ problemu prymalnego (4.1) i $y_1, ..., y_m, w$ problemu do niego dualnego (4.2), że $z = w$, to każde z tych rozwiązań byłoby rozwiązaniem optymalnym swojego problemu. Powstaje pytanie: kiedy takie rozwiązania istnieją? Zaraz zobaczymy, że zawsze wtedy gdy problemy (4.1) i (4.2) są niesprzeczne. Mówi o tym twierdzenie sformułowane w ostatecznej formie przez D. Gale'a, H.W. Kuhna i A.W. Tuckera w 1951 roku.

Twierdzenie 4.1.4 (Zasada dualności.) *Jeżeli zagadnienie prymalne (4.1) ma rozwiązanie optymalne* $(\bar{x}_1, ..., \bar{x}_n)$ *to problem dualny (4.2) ma także rozwiązanie optymalne* $(\bar{y}_1, ..., \bar{y}_m)$ *i zachodzi równość*

$$\sum_{j=1}^{n} c_j \bar{x}_j = \sum_{i=1}^{m} b_i \bar{y}_i \tag{4.5}$$

Dowód. Przypuśćmy, że $(\bar{x}_1, ..., \bar{x}_n)$ jest optymalnym rozwiązaniem problemu prymalnego (4.1). Zgodnie z twierdzeniem 4.1.2 wystarczy wskazać takie rozwiązanie $(\bar{y}_1, ..., \bar{y}_m)$ problemu dualnego (4.2), żeby zachodziła równość (4.5). Przypuśćmy, że ostatnim słownikiem otrzymanym dla problemu (4.1) jest słownik w którym funkcja celu wyraża się wzorem

$$z = \bar{z} + \sum_{k=1}^{m+n} \bar{c}_k x_k \tag{4.6}$$

Pamiętamy, że we wzorze (4.6)

$\bar{c}_k = 0$ gdy x_k jest zmienną niebazową ostatniego słownika

$\bar{c}_k \leq 0$ dla każdego $k = 1, ..., m+n$, bowiem zakładamy, że (4.1) ma rozwiązanie optymalne.

$\bar{z}$ jest oczywiście wartością maksymalną funkcji celu problemu (4.1) i, skoro $(\bar{x}_1, ..., \bar{x}_n)$ jest rozwiązaniem optymalnym, zachodzi

$$\bar{z} = \sum_{j=1}^{n} c_j \bar{x}_j$$

Wykażemy, że szukanym rozwiązaniem optymalnym problemu dualnego jest

$$\bar{y}_1 = -\bar{c}_{n+1}, \bar{y}_2 = -\bar{c}_{n+2}, ..., \bar{y}_m = -\bar{c}_{n+m}$$

W tym celu przepiszemy wzór (4.6) po dokonaniu na nim następujących manipulacji:

– zastępujemy z przez $\sum_{j=1}^{n} c_j x_j$,

– korzystając z faktu, że $\sum_{k=1}^{n+m} \bar{c}_k x_k = \sum_{j=1}^{n} \bar{c}_j x_j + \sum_{i=1}^{m} \bar{c}_{n+i} x_{n+i}$ (x_{n+i} są zmiennymi sztucznymi), zastępujemy

$\bar{c}_{n+i}$ przez $-\bar{y}_i$ (inaczej mówiąc, definiujemy $\bar{y}_i := -\bar{c}_{n+i}$)
x_{n+i} przez $b_i - \sum_{j=1}^{n} a_{ij}x_j$.

Otrzymamy wtedy wzór

$$\sum_{j=1}^{n} c_j x_j = \bar{z} + \sum_{j=1}^{n} \bar{c}_j x_j - \sum_{i=1}^{m} \bar{y}_i (b_i - \sum_{j=1}^{n} a_{ij} x_j)$$

a stąd po łatwych przekształceniach

$$\sum_{j=1}^{n} c_j x_j = (\bar{z} - \sum_{i=1}^{m} \bar{y}_i b_i) + \sum_{j=1}^{n} (\bar{c}_j + \sum_{i=1}^{m} \bar{y}_i a_{ij}) x_j \tag{4.7}$$

Równość (4.7) jest równością wielomianów zmiennych $x_1, ..., x_n$, stąd

$$c_j = \bar{c}_j + \sum_{i=1}^{m} a_{ij} \bar{y}_i \quad \text{dla} \quad j = 1, ..., n \tag{4.8}$$

oraz

$$\bar{z} = \sum_{i=1}^{m} \bar{y}_i b_i \tag{4.9}$$

Równość (4.9) oznacza, że zachodzi równość (4.5). Ponieważ $\bar{c}_j \leq 0$ dla wszystkich j, równość (4.8) pociąga

$$\sum_{j=1}^{n} a_{ij} \bar{y}_i \geq c_j \quad \text{dla} \quad j = 1, ..., n$$

a więc $\bar{y}_1, ..., \bar{y}_m$ jest rozwiązaniem dopuszczalnym problemu (4.2) i twierdzenie zostało wykazane. ■

Wzajemne zależności pomiędzy problemem prymalnym a dualnym wygodnie będzie mieć zapisane w poniższej tabeli.

Jeśli problem prymalny	to problem dualny
ma rozwiązanie optymalne skończone	ma rozwiązanie optymalne o tej samej wartości (twierdzenie 4.1.4)
jest nieograniczony	jest sprzeczny (wniosek 4.1.3)
jest sprzeczny	jest sprzeczny lub nieograniczony

4.2 Korzyści

Z pewnością łatwiej jest rozwiązywać PPL w którym jest mniej równań – liczba zmiennych odgrywa tu mniejszą rolę. Jeśli więc mamy PPL o bardzo wielu (np. $m = 100$) równaniach i znacznie mniejszej liczbie (np $n = 5$) niewiadomych, to łatwiej będzie nam rozwiązywać, zamiast problemu oryginalnego, problem dualny, z 5 równaniami i 100 niewiadomymi [1].

Powiedzmy, że rozwiązaliśmy prymalny PPL

$$\begin{array}{rcl} \sum_{j=1}^{n} a_{ij}x_j & \leq & b_i \quad (i=1,...,m) \\ x_j & \geq & 0 \quad (j=1,...,n) \\ \hline \sum_{j=1}^{n} c_j x_j & \to & \max \end{array} \tag{4.10}$$

Rozwiązaniem optymalnym (4.10) niech będzie $x_1^*, ..., x_n^*$. Powiedzmy dalej, że $y_1^*, ..., y_m^*$ jest rozwiązaniem optymalnym problemu dualnego

$$\begin{array}{rcl} \sum_{i=1}^{m} a_{ij}y_i & \geq & c_j (j=1,...,n) \\ y_i & \geq & 0 \quad (i=1,...,n) \\ \hline \sum_{i=1}^{m} b_i y_i & \to & \min \end{array} \tag{4.11}$$

Przekonanie kogokolwiek wystarczającego światłego, na przykład naszego zleceniodawcy, że rozwiązania te są rzeczywiście rozwiązaniami optymalnymi swoich problemów (odpowiednio (4.10) i (4.11)) jest bardzo proste: wystarczy sprawdzić, że $x_1^*, ..., x_n^*$ oraz $y_1^*, ..., y_m^*$ są rozwiązaniami dopuszczalnymi swoich problemów, czyli

$$\sum_{j=1}^{n} a_{ij}x_j \leq b_i \quad (i=1,...,m)$$

oraz

$$\sum_{i=1}^{m} a_{ij}y_i \geq c_i \quad (j=1,...,n)$$

oraz zachodzi wzór

$$\sum_{j=1}^{n} c_j x_j^* = \sum_{i=1}^{m} b_i y_i^* \tag{4.12}$$

Twierdzenia 4.2.1 i 4.2.2 pozwolą nam nie tylko sprawdzić optymalność rozwiązań problemów prymalnego i dualnego kiedy już je mamy, ale także, w pewnych sytuacjach, łatwo znaleźć optymalne rozwiązanie problemu dualnego, gdy będzie dane rozwiązanie problemu prymalnego.

Twierdzenie 4.2.1 *Niech $x_1^*, ..., x_n^*$ będzie rozwiązaniem dopuszczalnym problemu (4.10), zaś $y_1^*, ..., y_m^*$ rozwiązaniem dopuszczalnym problemu dualnego (4.11). $x_1^*, ..., x_n^*$ oraz $y_1^*, ..., y_m^*$ są – równocześnie – rozwiązaniami optymalnymi swoich problemów wtedy i tylko wtedy, gdy dla każdego $j = 1, ..., n$*

$$\sum_{i=1}^{m} a_{ij}y_i^* = c_j \quad \textit{lub} \quad x_j^* = 0$$

[1] Pamiętamy z dowodu twierdzenia 3.3.1, że w najgorszym przypadku iteracji może być $\binom{m+n}{n}$ w pierwszym i $\binom{n+m}{m}$ drugim przypadku, a więc tyle samo.

oraz dla każdego $i = 1, ..., m$

$$\sum_{j=1}^{n} a_{ij} x_j^* = b_i \quad lub \quad y_i^* = 0$$

Dowód. Niech $x_1^*, ..., x_n^*$ będzie rozwiązaniem dopuszczalnym problemu prymalnego (4.10), zaś $y_1^*, ..., y_m^*$ rozwiązaniem dopuszczalnym problemu dualnego (4.11). Podobnie jak w dowodzie twierdzenia 4.1.2

$$\sum_{j=1}^{n} c_j x_j^* \leq \sum_{j=1}^{n} \left(\sum_{i=1}^{m} a_{ij} y_i^* \right) x_j^* = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} a_{ij} x_j^* \right) y_i^* \leq \sum_{i=1}^{m} b_i y_i^*$$

Zachodzenie równości

$$\sum_{j=1}^{n} c_j x_j^* = \sum_{i=1}^{m} b_i y_i^*$$

która, na mocy twierdzenia 4.1.4 jest warunkiem koniecznym i wystarczającym optymalności rozwiązań $x_1^*, ..., x_n^*$ i $y_1^*, ..., y_n^*$ jest równoważne równościom

$$\sum_{j=1}^{n} c_j x_j^* = \sum_{j=1}^{n} \left(\sum_{i=1}^{m} a_{ij} y_i^* \right) x_j^* \quad \text{i} \quad \sum_{i=1}^{m} \left(\sum_{j=1}^{n} a_{ij} x_j^* \right) y_i^* = \sum_{i=1}^{m} b_i y_i^*$$

czyli, wobec $c_j \leq \sum_{i=1}^{m} a_{ij} y_i^*$ oraz $x_j^* \geq 0, (j = 1, ..., n)$

$$x_j^* = 0 \;\; \text{lub} \;\; c_j = \sum_{i=1}^{m} a_{ij} y_i^* \quad (j = 1, ..., n)$$

oraz (wobec $\sum_{j=1}^{n} a_{ij} x_j^* \leq b_i$ $y_i^* \geq 0$ dla $i = 1, ..., m)$

$$y_i^* = 0 \;\; \text{lub} \;\; \sum_{j=1}^{n} a_{ij} x_j^* = b_i \; (i = 1, ..., m). \blacksquare$$

Z twierdzeń 4.2.1 i 4.1.4 wynika następne twierdzenie które nazwiemy **twierdzeniem o odstępach** .

Twierdzenie 4.2.2 *Rozwiązanie dopuszczalne* $x_1^*, ..., x_n^*$ *problemu (4.10) jest optymalne wtedy i tylko wtedy, jeżeli istnieje ciąg* m *elementowy* $y_1^*, ..., y_m^*$ *taki, że*

1. *jeśli* $x_j^* > 0$ *to* $\sum_{i=1}^{m} a_{ij} y_i^* = c_j$ *(dla* $j = 1, ..., n$*),*
2. *jeśli* $\sum_{j=1}^{n} a_{ij} x_j^* < b_i$ *to* $y_i^* = 0$ *(dla* $i = 1, ..., m$*),*

przy czym:

$$\begin{array}{ccll} \sum_{i=1}^{m} a_{ij} y_i^* & \geq & c_j & (j = 1, ..., n) \\ y_i^* & \geq & 0 & (i = 1, ..., m). \end{array} \tag{4.13}$$

Dowód.

- Jeśli PPL (4.10) ma rozwiązanie optymalne $x_1^*, ..., x_n^*$, to na mocy twierdzenia 4.1.4 istnieje rozwiązanie optymalne $y_1^*, ..., y_m^*$ problemu dualnego. Tak więc $y_1^*, ..., y_m^*$ spełniają (4.13) zaś $x_1^*, ..., x_n^*$ spełniają warunki 1 i 2 na mocy twierdzenia 4.2.1.

- Z drugiej strony, jeśli $y_1^*, ..., y_m^*$ jest ciągiem spełniającym (4.13), to stanowi on rozwiązanie dopuszczalne problemu dualnego 4.11. Z twierdzenia 4.2.1 wynika, że jest to rozwiązanie optymalne. ■

4.2.1 Przykład

Niech będzie dany problem prymalny

$$\left\{\begin{array}{rcrcll} -x_1 & + & x_2 & \leq & 1 & \\ x_1 & - & 2x_2 & \leq & 0 & \\ x_1 & + & x_2 & \leq & 3 & \\ & & x_j & \geq & 0 & (j=1,2,3) \\ \hline & z & = & x_1+2x_2 & \to \max & \end{array}\right. \tag{4.14}$$

Wykażemy, że $x_1^* = 1, x_2^* = 2$ jest rozwiązaniem optymalnym problemu (4.14). Problemem dualnym do (4.14) jest

$$\left\{\begin{array}{rcrcrcll} -y_1 & + & y_2 & + & y_3 & \geq & 1 & \\ y_1 & - & 2y_2 & + & y_3 & \geq & 2 & \\ & & & & y_i & \geq & 0 & (i=1,2,3) \\ \hline & & & & w & = & y_1+3y_3 & \to \min \end{array}\right. \tag{4.15}$$

Oznaczmy przez y_1^*, y_2^*, y_3^* optymalne rozwiązania problemu (4.15). Zauważmy, że

- $x_1^* = 1, x_2^* = 2$ jest rozwiązaniem dopuszczalnym problemu (4.14).

- Spośród trzech pierwszych nierówności problemu (4.14) x_1^*, x_2^* spełnia drugą jako nierówność silną ($x_1^* - 2x_2^* = 1 - 4 = -3 < 0$), a więc, na mocy twierdzenia 4.2.2, $y_2^* = 0$.

- Ponieważ $x_1^*, x_2^* > 0$, na mocy twierdzenia 4.2.2 obie nierówności w problemie (4.15) są równościami, otrzymujemy więc

$$\left\{\begin{array}{rcrcl} -y_1^* & + & y_3^* & = & 1 \\ y_1^* & + & y_3^* & = & 2 \end{array}\right.$$

  i stąd $y_1^* = \frac{3}{2}, y_2^* = 0, y_3^* = \frac{1}{2}$. Ponieważ $z^* = x_1^* + 2x_2^* = y_1^* + 3y_3^* = w^* = 5$, x_1^*, x_2^* jest rozwiązaniem optymalnym (4.14) zaś y_1^*, y_2^*, y_3^* rozwiązaniem optymalnym (4.15).

4.3 Interpretacja ekonomiczna zmiennych dualnych

Problemem praktycznym który modeluje PPL

$$\begin{array}{llll} \text{zmaksymalizować} & \sum_{j=1}^{n} c_j x_j & & (j=1,...,n) \\ \text{przy warunkach} & \sum_{j=1}^{n} a_{ij} x_j & \le b_i & (i=1,...,m) \\ & x_j & \ge 0 & (j=1,...,n) \end{array}$$

jest następujące zagadnienie ekonomiczne.
Firma F produkuje towary $X_1, ..., X_n$. Do wyprodukowania tych towarów potrzebne są materiały (surowce) $Y_1, ..., Y_m$ w ilości a_{ij} jednostek materiału Y_i na jednostkę towaru X_j dla $i = 1, ..., m$, $j = 1, ..., n$. Firma dysponuje b_i jednostkami materiału Y_i dla $i = 1, ..., m$. Zysk z produkcji jednej jednostki towaru X_j wynosi c_j (ten zysk jednostkowy może być ujemny co dla firmy oznacza oczywiście stratę).
Ta ekonomiczna interpretacja problemu prymalnego jest dość oczywista. Interpretacja problemu dualnego

$$\begin{array}{lllll} \text{zminimalizować} & \sum_{i=1}^{m} b_i y_i & & & \\ \text{przy ograniczeniach} & \sum_{i=1}^{m} a_{ij} y_i \ge c_j & (j=1,...,n) & & \\ & y_i & & \ge 0 & (i=1,...,m) \end{array}$$

jest mniej jasna. Inspirująca jest tutaj kwestia zgodności jednostek fizycznych[2] (p. [13]). W ograniczeniach

$$\sum_{i=1}^{m} a_{ij} y_i \ge c_j$$

po prawej stronie nierówności jednostką fizyczną jest wartość jednostki produktu, jednostkowy zysk. Jednostkami a_{ij} są zaś ilości jednostek materiału Y_i potrzebnego do wytworzenia jednostki produktu X_j, czyli

$$\sum_{i=1}^{m} \left(\frac{\text{ilość jednostek materiału } Y_i}{\text{ilość jednostek produktu } X_j} \right) y_i \ge \left(\frac{\text{wartość}}{\text{ilość jednostek produktu } X_j} \right).$$

Stąd jednostką fizyczną zmiennej dualnej y_i musi być

$$\left(\frac{\text{wartość}}{\text{ilość jednostek materiału } Y_i} \right).$$

W problemie dualnym minimalizujemy dualną funkcję celu

$$\sum_{i=1}^{m} b_i y_i.$$

Zmienne dualne $y_1, ..., y_m$ interpretujemy jako koszty jednostki materiałow $Y_1, ..., Y_m$. W ten sposób dualna funkcja celu $\sum_{i=1}^{m} b_i y_i$ jest sumą składników postaci $b_i y_i$

[2] W tym fragmencie tekstu słowo *jednostka* występuje w dwóch zupełnie różnych znaczeniach.

gdzie b_i – jest ilością jednostek materiału (surowca) Y_i
zaś y_i – kosztem jednostkowym materiału (surowca)Y_i.

Fakt, że w problemie dualnym poszukujemy minimum $\sum_{i=1}^{m} b_i y_j$ oznacza, że szukamy wartości y_i przy których nakład (koszt materiałów (surowców)) jest minimalny.
W dualnych ograniczeniach lewa strona wzoru

$$\sum_{i=1}^{m} a_{ij} y_i \geq c_j$$

oznacza wtedy koszty produkcji towaru X_j. Mamy tu bowiem sumę iloczynów

(ilość jednostek materiału Y_i potrzebnych do wytworzenia jednostki X_j)·
(koszt jednostkowy Y_j)

która, jak zakładamy, jest co najmniej równa c_j – jednostkowej wartości produktu X_j.

Zmienne dualne y_i, $(i = 1, ..., m)$ nazywane są często *cenami dualnymi.* Zauważmy jednak, że *cena dualna* ma niewiele wspólnego z "prawdziwąćeną surowca. Z twierdzenia o odstępach komplementarnych wynika, że gdy dla pewnego i w rozwiązaniu optymalnym $(x_1^*, ..., x_n^*)$ zachodzi $\sum_{j=1}^{n} a_{ij} x_j^* < c_i$, czyli surowca Y_i jest zbyt dużo, to cena dualna tego surowca jest równa zero.

Dodatkowym uzasadnieniem interpretacji zmiennych dualnych jest następujące twierdzenie (które podajemy tu bez dowodu).

Twierdzenie 4.3.1 *Jeżeli problem prymalny*

$$\begin{array}{lcll} \sum_{j=1}^{n} a_{ij} x_j & \leq & b_i & (i = 1, ..., m) \\ \hline \sum_{j=1}^{n} c_j x_j & \rightarrow & max & \end{array} \tag{4.16}$$

ma niezdegenerowane bazowe rozwiązanie optymalne, to istnieje $\varepsilon > 0$ *takie, że jeśli* $\mid t_i \mid \leq \varepsilon$ *dla* $i = 1, ..., m$, *to problem*

$$\begin{array}{lcll} \sum_{j=1}^{n} a_{ij} & \leq & b_i + t_i & (i = 1, ..., m) \\ \hline \sum_{j=1}^{n} c_j x_j & \rightarrow & max & \end{array}$$

ma rozwiązanie optymalne o wartości $z^* + \sum_{i=1}^{m} t_i y_i^*$, *gdzie* z^* *jest wartością optymalną problemu (4.16), zaś* $y_1^*, ..., y_m^*$ *optymalnym rozwiązaniem dla problemu dualnego.*

4.4 Dualność ogólniej

Jak zobaczymy w ustępie 7.3, czasami wygodnie będzie nam dysponować problemem dualnym do PPL w następującej postaci:

$$\begin{cases} \text{zmaksymalizuj:} & \sum_{j=1}^{n} c_j x_j \\ \text{przy ograniczeniach:} & \sum_{j=1}^{n} a_{ij} x_j \leq b_i \quad (i \in N) \\ & \sum_{j=1}^{n} a_{ij} x_j = b_i \quad (i \in R) \\ & x_j \geq 0 \quad (j \in P) \end{cases} \tag{4.17}$$

W PPL postaci (4.17) tylko o części zmiennych (x_j) zakładamy, że są nieujemne (dla $j \in P$), pozostałe zmienne nazywamy *wolnymi*, a zbiór indeksów zmiennych wolnych oznaczamy przez W[3]

Definicja 4.4.1 Problemem dualnym *do (4.17) nazywamy problem programowania liniowego następującej postaci:*

$$\begin{cases} \text{zminimalizuj:} & \sum_{i=1}^{m} b_i y_i \\ \text{przy ograniczeniach:} & \sum_{i=1}^{m} a_{ij} y_i \geq c_j \quad \text{dla } j \in P \\ & \sum_{i=1}^{m} a_{ij} y_i = c_j \quad \text{dla } j \in W \\ & y_i \geq 0 \quad \text{dla } i \in N \end{cases} \tag{4.18}$$

Problem (4.17) nazywamy wtedy **problemem prymalnym** *(względem (4.18)).*

Przykład 4.4.1 Problemem dualnym do PPL

$$\begin{array}{lrcrcrcl} \text{zmaksymalizuj:} & 3x_1 & + & x_2 & + & 2x_3 & & \\ \text{przy warunkach:} & x_1 & + & 2x_2 & - & x_3 & \leq & 3 \\ & 4x_1 & - & 2x_2 & + & x_3 & \leq & 2 \\ & -x_1 & + & 5x_2 & + & x_3 & = & 1 \\ & & & & & x_2 & \geq & 0 \end{array}$$

jest PPL

$$\begin{array}{lrcrcrcl} \text{zminimalizuj:} & 3y_1 & + & 2y_2 & + & y_3 & & \\ \text{przy warunkach:} & 2y_1 & - & 2y_2 & + & 5y_3 & \geq & 1 \\ & y_1 & + & 4y_2 & - & y_3 & = & 3 \\ & -y_1 & + & y_2 & + & y_3 & = & 2 \\ & & & & & y_1, y_2 & \geq & 0 \end{array}$$

Rzeczywiście, w naszym przykładzie mamy $N = \{1, 2\}$, $R = \{3\}$, $P = \{2\}$, $W = \{1, 3\}$.$\Box$

Wyniki jakie otrzymamy dla problemów (4.17) i (4.18) są uogólnieniami twierdzeń ustępu 4.1 w tym sensie, że jeśli w problemach (4.17) i (4.18) położymy $R = \emptyset$ i $W = \emptyset$ (a więc $P = 1, ..., n$), to poniższe twierdzenia 4.4.1 i 4.4.2 sprowadzają się do twierdzeń 4.1.2 i 4.1.4.

---

[3]Pozostałe zbiory indeksów też zostały nazwane tak, by było je można łatwiej zapamiętać: N od **n**ierówność, R od **r**ówność, P od **p**ozytywne (tu w znaczeniu: nieujemne).

Twierdzenie 4.4.1 *Jeżeli $\bar{x}_1, ..., \bar{x}_n$ i $\bar{y}_1, ..., \bar{y}_m$ są rozwiązaniami dopuszczalnymi problemów, odpowiednio (4.17) i (4.18), to zachodzi nierówność*

$$\sum_{j=1}^{n} c_j \bar{x}_j \leq \sum_{i=1}^{m} b_i \bar{y}_i$$

Dowód twierdzenia 4.4.1 niewiele różni się od dowodu twierdzenia 4.1.2.
Dla $j \in P$ mamy

$$c_j \leq \sum_{i=1}^{m} a_{ij} \bar{y}_i$$

co po wymnożeniu przez $\bar{x}_j$ (pamiętamy, że wtedy $\bar{x}_j \geq 0$) daje

$$c_j \bar{x}_j \leq (\sum_{i=1}^{m} a_{ij} y_i) \bar{x}_j$$

Z kolei dla $j \in W$

$$c_j = \sum_{i=1}^{m} a_{ij} \bar{y}_i$$

i wobec tego

$$c_j \bar{x}_j = (\sum_{i=1}^{m} a_{ij} \bar{y}_i) \bar{x}_j$$

Sumując te wzory po wszystkich $j = 1, ..., n$ otrzymamy

$$\sum_{j=1}^{n} c_j \bar{x}_j \leq \sum_{j=1}^{n} (\sum_{i=1}^{m} a_{ij} \bar{y}_i) \bar{x}_j$$

co po łatwych przekształceniach daje

$$\sum_{j=1}^{n} c_j \bar{x}_j \leq \sum_{j=1}^{n} (\sum_{i=1}^{m} a_{ij} \bar{y}_i) \bar{x}_j$$

Dla $i \in N$ mamy $\sum_{j=1}^{n} a_{ij} \bar{x}_j \leq b_i$ oraz $\bar{y}_i \geq 0$, więc

$$(\sum_{j=1}^{n} a_{ij} \bar{x}_j) \bar{y}_i \leq b_i \bar{y}_i$$

Podobnie, dla $i \in R$

$$(\sum_{j=1}^{n} a_{ij} \bar{x}_j) \bar{y}_i = b_i \bar{y}_i$$

Znowu sumując po $i \in N \cup R = \{1, ..., m\}$ otrzymamy ostatecznie

$$\sum_{i=1}^{m} (\sum_{j=1}^{n} a_{ij} \bar{x}_j) \bar{y}_i \leq \sum_{i=1}^{m} b_i \bar{y}_i$$

i dowód twierdzenia 4.4.1 jest zakończony. ■

Czytelnika który nie jest jeszcze znużony kolejnym wykorzystaniem metody sympleksowej którą poznał w rozdziale 3, zachęcamy do samodzielnego udowodnienia przedstawionej poniżej uogólnionej zasady dualności (lub do odszukania dowodu w literaturze (np. w [4]).

Twierdzenie 4.4.2 (Ogólna zasada dualności) *Jeżeli problem prymalny (4.17) ma rozwiązanie optymalne, to także problem dualny (4.18) ma rozwiązanie optymalne i wartości funkcji celu obu tych rozwiązań są równe.* ■

Uwaga. łatwo zauważyć, że jeśli wymnożymy w obu problemach (4.17) i (4.18) funkcje celu oraz nierówności w pierwszych linijkach ograniczeń przez -1, to otrzymamy sytuację w której problem (4.18) jest problemem w którym maksymalizujemy funkcję celu i problemem do niego dualnym jest (4.17). Stąd prawdziwy jest wniosek:

Wniosek 4.4.3 *Jeśli (4.18) ma rozwiązanie optymalne, to także (4.17) ma rozwiązanie optymalne i wartości tych rozwiązań są identyczne.*

4.5 ćwiczenia

Ćwiczenie 4.5.1 Wskaż problemy dualne do poniższych problemów.

1.

$$\begin{cases} \text{zmaksymalizuj} & x_1 + 3x_2 - 2x_3 \\ \text{przy ograniczeniach:} & 2x_1 - 3x_2 - x_3 \geq 8 \\ & x_1 - x_2 + x_3 \leq 6 \\ & -x_1 + x_2 + 2x_3 = 3 \\ & x_2 \geq 0 \end{cases}$$

2.

$$\begin{cases} \text{zminimalizuj} & -2x_1 + x_2 + 2x_3 \\ \text{przy ograniczeniach:} & x_1 - 3x_2 - 4x_3 \leq 8 \\ & x_1 + 2x_2 - 4x_3 = 6 \\ & -x_1 - x_2 + 3x_3 = 5 \\ & x_1 \geq 0 \end{cases}$$

Ćwiczenie 4.5.2 Dla PPL

$$\begin{array}{rcll} 2x_1 + 2x_2 - x_3 & \geq & 1 & \\ x_1 + 3x_2 + 2x_3 & = & 2 & \\ x_i & \geq & 0 & i = 1, 2, 3 \\ \hline \end{array}$$

$$z = \quad 8x_1 + 4x_2 - 2x_3 - 3 \quad \to \quad \min$$

znajdź i rozwiąż zadanie dualne. Odczytaj rozwiązanie zadania prymalnego.

Ćwiczenie 4.5.3 Sprawdź, że $x_1 = 2,5,\ x_2 = 1,5,\ x_3 = 0$ jest optymalnym rozwiązaniem PPL

$$\begin{array}{lcl} x_1 + x_2 + 2x_3 & \leq & 4 \\ 2x_1 + 3x_3 & \leq & 5 \\ 2x_1 + x_2 + 3x_3 & \leq & 7 \\ x_1, x_2, x_3 & \geq & 0 \\ \hline 3x_1 + 2x_2 + 4x_3 & \to & \max \end{array}$$

Napisz problem dualny i rozwiązanie problemu dualnego.

Ćwiczenie 4.5.4 Wykaż, że $x_1^* = \frac{10}{7}, x_2^* = 0, x_3^* = \frac{1}{7}$ jest rozwiązaniem optymalnym problemu

$$(*) \left\{ \begin{array}{rrrrcrl} & x_1+ & 2x_2+ & x_3 & \leq & 4 & \\ & 2x_1+ & x_2+ & x_3 & \leq & 3 & \\ & x_1- & x_2+ & 4x_3 & \leq & 2 & \\ & & & x_i & \geq & 0 & i = 1,2,3 \\ \hline z = & 3x_1+ & x_2+ & 2x_3 & \to & \max & \end{array} \right.$$

Napisz program dualny do $(*)$ i wskaż jego rozwiązanie optymalne.

Rozdział 5

Zredukowana metoda sympleksowa

W niniejszym rozdziale okaże się, że nie wszystkie obliczenia które wykonywaliśmy w każdej iteracji są niezbędne. Więcej, większość z nich można pominąć. żeby to dostrzec, przyjrzymy się macierzowemu opisowi słownika.

5.1 Macierzowy opis słownika

Dla potrzeb niniejszego rozdziału słownik który do tej pory zapisywaliśmy jako

$$\begin{array}{ccccl} x_i & = & b_i^* & - & \sum_{j\notin B} \bar{a}_{ij} x_j \quad (i \in B) \\ \hline z & = & z^* & + & \sum_{j\notin B} \bar{c}_j x_j \end{array} \tag{5.1}$$

gdzie B jest zbiorem indeksów bazowych, będzie nam wygodniej zapisywać w postaci

$$\begin{array}{ccccccl} \sum_{j\notin B} \bar{a}_{ij} x_j & + & x_i & = & b_i^* & & (i \in B) \\ \hline & & z & = & z^* + & \sum_{j\notin B} \bar{c}_j x_j & \end{array} \tag{5.2}$$

W szczególności pierwszy słownik jest postaci

$$\begin{array}{cccccl} \sum_{j\notin B} a_{ij} x_j & +x_i & = & b_i & (i \in B) & \\ \hline & z & = & z & \sum_{j\notin B} c_j x_j & \end{array} \tag{5.3}$$

ponieważ zaś w pierwszym słowniku zbiór indeksów bazowych $B = \{n+1, ..., n+m\}$, słownik ten wygląda następująco

$$\begin{array}{ccccl} \sum_{j=1}^{n} a_{ij} x_j & +x_i & = b_i & & (i = n+1, ..., n+m) \\ \hline & z & = & z^* & + \sum_{j=1}^{n} c_j x_j \end{array} \tag{5.4}$$

W zapisie macierzowym to oczywiście

$$\begin{array}{rclcl} \mathbf{A}^*\mathbf{x} & = & \mathbf{b} & & \\ \hline z & = & z^* & + & \mathbf{c}\mathbf{x}_N \end{array} \tag{5.5}$$

gdzie

$$\mathbf{A}^* = \begin{bmatrix} a_{11} & ... & a_{1n} & 1 & 0 & ... & 0 \\ a_{21} & ... & a_{2n} & 0 & 1 & ... & 0 \\ & ... & & & & ... & \\ a_{m1} & ... & a_{mn} & 0 & 0 & ... & 1 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} b_1 \\ . \\ . \\ . \\ b_m \end{bmatrix},$$

$$\mathbf{c} = [c_1, ..., c_n], \mathbf{x} = \begin{bmatrix} x_1 \\ . \\ . \\ . \\ x_n \\ x_{n+1} \\ . \\ . \\ . \\ x_{n+m} \end{bmatrix} \quad \mathbf{x}_N = \begin{bmatrix} x_1 \\ . \\ . \\ . \\ x_n \end{bmatrix}$$

Ten sam słownik możemy zapisać także tak:

$$\begin{array}{ccccc} \mathbf{A}\mathbf{x}_N & + & \mathbf{I}_m\mathbf{x}_B & = \mathbf{b} & \\ \hline z & = & z^* & + & \mathbf{c}\mathbf{x}_N \end{array} \tag{5.6}$$

gdzie $\mathbf{I}_m$ jest macierzą jednostkową o m wierszach i m kolumnach, zaś

$$\mathbf{x}_B = \begin{bmatrix} x_{n+1} \\ . \\ . \\ . \\ x_{n+m} \end{bmatrix}$$

Przykład 5.1.1 Czytelnik który rozwiązał ćwiczenie 3.7.5 stwierdził z pewnością, że z pierwszego słownika

$$\begin{array}{rrrrrrrcr} & 2x_1- & x_2+ & x_3+ & x_4 & & & = & 12 \\ & x_1+ & 2x_2+ & x_3 & + & x_5 & & = & 10 \\ & -x_1- & x_2+ & x_3 & & + & x_6 & = & -4 \\ \hline z = & 2x_1+ & x_2+ & x_3 & & & & \to & \max \end{array} \tag{5.7}$$

w którym rozwiązanie bazowe $x_1 = x_2 = x_3 = 0$, $x_4 = 12$, $x_5 = 10$, $x_6 = -4$ nie jest rozwiązaniem dopuszczalnym, otrzymał słownik w którym rozwiązanie bazowe jest dopuszczalne, mianowicie

$$\begin{array}{rrrrr} x_1 = & 4- & x_2+ & x_3+ & x_6 \\ x_4 = & 4+ & 3x_2- & 3x_3- & 2x_6 \\ x_5 = & 6- & x_2- & 2x_3- & x_6 \\ \hline z = & 8- & x_2- & x_3+ & 2x_6 \end{array} \tag{5.8}$$

Zmiennymi bazowymi słownika (5.8) są x_1, x_4 i x_5. Macierzowo słownik (5.8) można zapisać następująco

$$\begin{array}{ccccc} \mathbf{x}_B & = & \mathbf{b}^* & - & \mathbf{P}\mathbf{x}_N \\ \hline z & = & z^* & + & \mathbf{c}^*\mathbf{x}_N \end{array} \tag{5.9}$$

gdzie $\mathbf{P} = \begin{bmatrix} 1 & -1 & -1 \\ -3 & 3 & 2 \\ 1 & 2 & 1 \end{bmatrix}$, $\mathbf{b}^* = \begin{bmatrix} 4 \\ 4 \\ 6 \end{bmatrix}$, $z^* = 8$, $\mathbf{c}^* = [-1, -1, 2]$, zaś $\mathbf{x}_N = \begin{bmatrix} x_2 \\ x_3 \\ x_6 \end{bmatrix}$

□.

Jak można wyrazić słownik (5.9) przy pomocy słownika (5.6) lub (5.5)? W tym celu wystarczy zauważyć następujące fakty:

1. W macierzy $\mathbf{A}^*$ zmiennym bazowym x_1, x_4, x_5 odpowiadają pierwsza, czwarta i piąta kolumna.

2. Pierwszy wiersz słownika (5.5) (to znaczy wszystko co ponad kreską w tym wzorze), można zapisać w postaci

$$\mathbf{A}_B\mathbf{x}_B + \mathbf{A}_N\mathbf{x}_N = \mathbf{b} \tag{5.10}$$

   gdzie macierz $\mathbf{A}_B$ jest utworzona przez kolumny macierzy $\mathbf{A}^*$ odpowiadające zmiennym bazowym, zaś $\mathbf{A}_B$ powstaje z $\mathbf{A}^*$ przez wybranie kolumn odpowiadających zmiennym niebazowym. Wtedy $\mathbf{x}_B$ i $\mathbf{x}_N$ są utworzone, odpowiednio, ze współrzędnych będących bazowymi i niebazowymi współrzędnymi wektora $\mathbf{x}$.

Przykład 5.1.1 (c.d.) Dla zmiennych bazowych x_1, x_4, x_5 mamy:

$$\mathbf{A}_B = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 0 & 1 \\ -1 & 0 & 0 \end{bmatrix}, \ \mathbf{A}_N = \begin{bmatrix} -1 & 1 & 0 \\ 2 & 1 & 0 \\ -1 & 1 & 1 \end{bmatrix}, \ \mathbf{x}_B = \begin{bmatrix} x_1 \\ x_4 \\ x_5 \end{bmatrix}, \ \mathbf{x}_N = \begin{bmatrix} x_2 \\ x_3 \\ x_6 \end{bmatrix}$$

Ponieważ w naszym przykładzie macierz $\mathbf{A}_B$ jest nieosobliwa, możemy z równania (5.10) obliczyć $\mathbf{x}_B$

$$\mathbf{x}_B = \mathbf{A}_B^{-1}(\mathbf{b} - \mathbf{A}_N\mathbf{x}_\mathbf{N}) \tag{5.11}$$

■

Nieosobliwość macierzy $\mathbf{A}_B$ w przykładzie 5.1.1 nie jest przypadkowa. Udowodnimy następujące twierdzenie.

Twierdzenie 5.1.1 *Jeśli dany jest słownik postaci*

$$\begin{array}{rcrcl} \mathbf{A}_B\mathbf{x}_B & = & \mathbf{b} & - & \mathbf{A}_N\mathbf{x}_N \\ \hline z & = & \mathbf{c}_B\mathbf{x}_B & + & \mathbf{c}_N\mathbf{x}_N \end{array} \tag{5.12}$$

gdzie $\mathbf{x}_B = [x_k]_{k\in B}$*,* $\mathbf{x}_N = [x_k]_{k\notin B}$*,* B *– zbiór indeksów zmiennych bazowych, to macierz* $\mathbf{A}_B$ *jest nieosobliwa.*

Dowód. Dla wykazania naszego twierdzenia wystarczy udowodnić, że równanie $\mathbf{A}_B\mathbf{x}_B = \mathbf{b}$ ma dokładnie jedno rozwiązanie.
Niesprzeczność równania $\mathbf{A}_B\mathbf{x}_B = \mathbf{b}$ jest oczywista: niech $\mathbf{x}^*$ będzie dopuszczalnym rozwiązaniem bazowym – czyli takim, że $\mathbf{x}^*{}_N = \mathbf{0}$. Wstawiając $\mathbf{x}^*$ do (5.12) otrzymamy $\mathbf{A}_B\mathbf{x}^*{}_B = \mathbf{b}$. Przypuśćmy, że $\bar{\mathbf{x}}_B \in \mathbf{R^n}$ także spełnia równanie $\mathbf{A}_B\bar{\mathbf{x}}_B = \mathbf{b}$. Kładąc $\bar{\mathbf{x}}_N = \mathbf{0}$ otrzymamy $\mathbf{A}\bar{\mathbf{x}} = \mathbf{A}_B\bar{\mathbf{x}}_B + \mathbf{A}_N\bar{\mathbf{x}}_N = \mathbf{b}$.
Tak więc $\bar{\mathbf{x}}$ jest rozwiązaniem dopuszczalnym, co więcej, skoro $\bar{\mathbf{x}}_N = \mathbf{0}$, jest to dopuszczalne rozwiązanie bazowe (dla zmiennych bazowych $(x_k)_{k\in B}$), czyli $\bar{\mathbf{x}}_B = \mathbf{x}^*_B$. ■

Definicja 5.1.1 *Macierz* $\mathbf{B} = \mathbf{A}_B$ *powstałą z* $\mathbf{A}$ *przez usunięcie kolumn odpowiadających zmiennym niebazowym nazywamy* **macierzą bazową** [1]*.*

Oznaczmy przez $\tilde{\mathbf{c}}$ wektor $\tilde{\mathbf{c}} = (c_1, ..., c_n, 0, ..., 0) = (\tilde{c}_1, ..., \tilde{c}_{n+m})$. Niech $\mathbf{c}_N$ oznacza wektor $\mathbf{R^n}$ którego współrzędne są współrzędnymi wektora $\tilde{\mathbf{c}}$ odpowiadającymi zmiennym niebazowym, $\mathbf{c}_N = (\tilde{c}_i)_{i\notin B}$, zaś $\mathbf{c}_B$ niech będzie wektorem $\mathbf{R^n}$ którego współrzędne są współczynnikami przy zmiennych bazowych w funkcji celu (w wyjściowej postaci). Inaczej:

$$\mathbf{c}_B = (\tilde{c}_j)_{j\in B}$$

We wzorze (5.6) możemy teraz zastąpić $\mathbf{c}\mathbf{x}_N$ przez

$$\mathbf{c}_B\mathbf{x}_B + \mathbf{c}_N\mathbf{x}_N$$

i otrzymamy

$$z = \mathbf{c}_B\mathbf{x}_B + \mathbf{c}_N\mathbf{x}_N$$

Po zastąpieniu $\mathbf{x}_B$ przez $\mathbf{B}^{-1}(\mathbf{b} - \mathbf{A}_N\mathbf{x}_N)$ otrzymamy

$$z = \mathbf{c}_B\mathbf{B}^{-1}(\mathbf{b} - \mathbf{A}_N\mathbf{x}_N) + \mathbf{c}_N\mathbf{x}_N$$

$$z = \mathbf{c}_B\mathbf{B}^{-1}\mathbf{b} + (\mathbf{c}_N - \mathbf{c}_B\mathbf{B}^{-1}\mathbf{A}_N)\mathbf{x}_N$$

(gdzie $\mathbf{B}$ jest macierzą bazową).
Ostatecznie, postacią słownika przy bazie $\mathbf{B}$ jest

$$\begin{array}{lcl} \mathbf{x}_B & = & \mathbf{B}^{-1}(\mathbf{b} - \mathbf{A}_N\mathbf{x}_N) \\ \hline z & = & \mathbf{c}_B\mathbf{B}^{-1}\mathbf{b} + (\mathbf{c}_N - \mathbf{c}_B\mathbf{B}^{-1}\mathbf{A}_N)\mathbf{x}_N \end{array} \tag{5.13}$$

[1] Macierz bazową czasami wygodnie jest oznaczać przez $\mathbf{A}_B$ (by podkreślić, że powstała z macierzy $\mathbf{A}$), a czasami przez $\mathbf{B}$ (bo krócej).

Przykład 5.1.1 (dokończenie) W ostatnim słowniku naszego przykładu jedynym współczynnikiem dodatnim funkcji celu jest współczynnik przy x_6. Zmienną wchodzącą będzie więc x_6. Równie łatwo zauważyć, że zmienną wychodzącą jest x_4. W nowym słowniku zmiennymi bazowymi będą więc x_1, x_5 i x_6, zmiennymi niebazowymi x_2, x_3, x_4. Wobec tego macierzą bazową $\mathbf{B}$ jest macierz złożona z pierwszej, piątej i szóstej kolumny macierzy

$$\mathbf{A} = \begin{bmatrix} 2 & -1 & 1 & 1 & 0 & 0 \\ 1 & 2 & 1 & 0 & 1 & 0 \\ -1 & -1 & 1 & 0 & 0 & 1 \end{bmatrix}$$

czyli

$$\mathbf{B} = \begin{bmatrix} 2 & 0 & 0 \\ 1 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix}, \quad \mathbf{A}_N = \begin{bmatrix} -1 & 1 & 1 \\ 2 & 1 & 0 \\ -1 & 1 & 0 \end{bmatrix}$$

$\mathbf{c}_B = [2, 0, 0]$, $\mathbf{c}_N = [1, 1, 0]$.
Obliczmy najpierw wektor

$$\mathbf{y} = \mathbf{c}_B \mathbf{B}^{-1}$$

z równania $\mathbf{yB} = \mathbf{c}_B$

$$[y_1, y_2, y_3] \begin{bmatrix} 2 & 0 & 0 \\ 1 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix} = [2, 0, 0]$$

$$\begin{cases} 2y_1 & +y_2 & -y_3 & = 2 \\ & y_2 & & = 0 \\ & & y_3 & = 0 \end{cases}$$

$y_1 = 1, y_2 = y_3 = 0$

$$\mathbf{c}_B \mathbf{B}^{-1} \mathbf{A}_N = [1, 0, 0] \begin{bmatrix} -1 & 1 & 1 \\ 2 & 1 & 0 \\ -1 & 1 & 0 \end{bmatrix} = [-1, 1, 1]$$

$$\mathbf{c}_N - \mathbf{c}_B \mathbf{B}^{-1} \mathbf{A}_N = [1, 1, 0] - [-1, 1, 1] = [2, 0, -1]$$

Stąd oczywiście zmienną wchodzącą jest koniecznie x_2 (jedyną dodatnią współrzędną wektora $[2, 0, -1]$ jest 2, współrzędna odpowiadająca zmiennej x_2).
żeby wyznaczyć zmienną wychodzącą, musimy w pierwszym równaniu (5.13) wyznaczyć wektor $\mathbf{B}^{-1}\mathbf{b}$ i **pierwszą kolumnę** (tę odpowiadającą zmiennej x_2) macierzy $\mathbf{B}^{-1}\mathbf{A}_N$.

$$\begin{cases} 2a & & & = 12 \\ a & +b & & = 10 \\ -a & & +c & = -4 \end{cases}$$

$a = 6,\ b = 4,\ c = 2$, a więc

$$\mathbf{B}^{-1}\mathbf{b} = \begin{bmatrix} 6 \\ 4 \\ 2 \end{bmatrix}$$

Z kolei pierwszą kolumną $\begin{bmatrix} k_1 \\ k_2 \\ k_3 \end{bmatrix}$ macierzy $\mathbf{D} = \mathbf{B}^{-1}\mathbf{A}_N$ wyznaczymy z równania $\mathbf{BD} = \mathbf{A}_N$, a więc

$$\begin{bmatrix} 2 & 0 & 0 \\ 1 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \\ k_3 \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \\ -1 \end{bmatrix}$$

(po lewej stronie tej równości jest iloczyn macierzy $\mathbf{B}$ przez pierwszą kolumnę macierzy $\mathbf{D} = \mathbf{B}^{-1}\mathbf{A}_N$, z prawej pierwsza kolumna $\mathbf{A}_N$).

$$\begin{cases} 2k_1 & & & = & -1 \\ k_1 & +k_2 & & = & 2 \\ -k_1 & & +k_3 & = & -1 \end{cases}$$

$k_1 = -\frac{1}{2}$, $k_2 = \frac{5}{2}$, $k_3 = -\frac{3}{2}$.
Pomijając nieistotne dla nas zmienne niebazowe x_3 i x_4 równanie $\mathbf{x}_B = \mathbf{B}^{-1}\mathbf{b} - \mathbf{B}^{-1}\mathbf{A}_N\mathbf{x}_N$ słownika (5.13) oznacza w w naszym przypadku

$$\begin{array}{rcl} x_1 & = & 6 + \frac{1}{2}x_2 \quad ... \\ x_5 & = & 4 - \frac{5}{2}x_2 \quad ... \\ x_6 & = & 2 + \frac{3}{2}x_2 \quad ... \end{array}$$

Jedynym współczynnikiem ujemnym przy x_2 jest $-\frac{5}{2}$ i zmienną wychodząca jest x_5.
Tak więc nowymi zmiennmi bazowymi są

x_1, x_2, x_6, a zmiennymi niebazowymi

x_3, x_4, x_5. Nowe macierze i wektory $\mathbf{B}, \mathbf{A}_N, \mathbf{c}_B, \mathbf{c}_N$ to teraz

$$\mathbf{B} = \begin{bmatrix} 2 & -1 & 0 \\ 1 & 2 & 0 \\ -1 & -1 & 1 \end{bmatrix}, \mathbf{A}_N = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}, \mathbf{c}_B = [2, 1, 0], \mathbf{c}_N = [1, 0, 0].$$

Znowu liczymy $\mathbf{y} = [y_1, y_2, y_3]$ z równania $\mathbf{y} = \mathbf{c}_B\mathbf{B}^{-1}$, czyli $\mathbf{yB} = \mathbf{c}_B$.

$$[y_1, y_2, y_3] \begin{bmatrix} 2 & -1 & 0 \\ 1 & 2 & 0 \\ -1 & -1 & 1 \end{bmatrix} = [2, 1, 0]$$

$$\begin{cases} 2y_1 & +2y_2 & -y_3 & = 2 \\ -y_1 & +2y_2 & -y_3 & = 1 \\ & & y_3 & = 0 \end{cases}$$

$$\begin{cases} -y_1 & +2y_2 & = 1 \\ & 5y_2 & = 4 \end{cases}$$

$$y_2 = \frac{4}{5}, \quad y_1 = \frac{8}{5} - 1 = \frac{3}{5}$$

Teraz liczymy $\mathbf{c}_B\mathbf{B}^{-1}\mathbf{A}_N = \mathbf{y}\mathbf{A}_N = [\frac{3}{5}, \frac{4}{5}, 0] \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} = [\frac{7}{5}, \frac{3}{5}, \frac{4}{5}]$ i ostatecznie otrzymamy

$$\mathbf{c}_N - \mathbf{c}_B\mathbf{B}^{-1}\mathbf{A}_N = [1, 0, 0] - [\frac{7}{5}, \frac{3}{5}, \frac{4}{5}] = [-\frac{2}{5}, -\frac{3}{5} - \frac{4}{5}]$$

Stąd oczywiście wniosek, że nasze rozwiązanie bazowe jest optymalne. W rozwiązaniu tym oczywiście $x_3 = x_4 = x_5 = 0$. Czytelnik zechce sprawdzić, że wartościami zmiennych bazowych x_1, x_2 i x_6 z równania

$$\begin{bmatrix} x_1 \\ x_2 \\ x_6 \end{bmatrix} = \mathbf{B}^{-1}\mathbf{b}$$

są $x_1 = \frac{34}{5}, x_2 = \frac{8}{5}, x_6 = 0$ i wartością optymalną funkcji celu jest $z = z^* + \mathbf{c}_B\mathbf{B}^{-1}\mathbf{b} = 0 + [\frac{3}{5}, \frac{4}{5}, 0] \begin{bmatrix} 12 \\ 10 \\ -4 \end{bmatrix} = \frac{36}{5} + \frac{40}{5} = \frac{76}{5}$.

5.2 Podsumowanie

W zrewidowanej metodzie sympleksowej oszczędzamy – w porównaniu z metodąoryginalną przedstawioną w rozdziale trzecim – czas i pamięć (komputera).

Czas – bowiem nie liczymy iloczynu $\mathbf{A}_B^{-1}\mathbf{A}_N$ a tylko jego jedną kolumnę. To w praktycznych zadaniach PL bardzo istotna oszczędność, bowiem na ogół zmiennych, a więc kolumn macierzy jest bardzo dużo.

Pamięć – ponieważ do kolejnych obliczeń w ogóle nie są nam potrzebne żadne słowniki poprzednie, poza słownikiem pierwszym.

Każdą iterację w metodzie zredukowanej można opisać następująco:

Krok 1. Obliczamy $\mathbf{y} = \mathbf{c}_B\mathbf{B}^{-1}$ (z równania $\mathbf{y}\mathbf{B} = \mathbf{c}_B$)[2]

Krok 2. Jeśli $\mathbf{c}_N \leq \mathbf{y}\mathbf{A}_N$ to aktualne rozwiązanie bazowe jest optymalne. Jeśli nie, to znaczy, że dla pewnej kolumny $\mathbf{k}$ macierzy $\mathbf{A}_N$ zachodzi $c_j - \mathbf{y}\mathbf{k} > 0$, dla odpowiedniej współrzędnej c_j wektora $\mathbf{c}_N$. Kolumna $\mathbf{k}$ wyznacza zmienną wychodzącą.

[2]Nigdy nie obliczamy $\mathbf{y}$ przez wyliczenie macierzy $\mathbf{B}^{-1}$ co jest wysiłkiem dużym i zupełnie zbędnym. Szczególnie wtedy, gdybyśmy chcieli w tym celu wykorzystać zupełnie nieefektywną metodę wyznacznikową.

Krok 3. Obliczamy kolumnę $\mathbf{d}$ macierzy $\mathbf{D} = \mathbf{B}^{-1}\mathbf{A}_N$, odpowiadającą zmiennej wchodzącej. Jeśli wszystkie jej współrzędne są ujemne, to problem jest nieograniczony. W przeciwnym przypadku jako zmienną wychodzącą wybieramy tę dla której iloraz

$$\frac{\text{wartość w rozwiązaniu bazowym}}{\text{wartość współrzędnej w kolumnie } \mathbf{d}}$$

jest minimalny (i nieujemny).

Krok 4. Ustalamy nowe zmienne bazowe i niebazowe i dla nich wartości macierzy $\mathbf{B}, \mathbf{A}_N$ oraz wektorów $\mathbf{c}_B$ i $\mathbf{c}_N$.

5.3 Programowanie liniowe w liczbach całkowitych

Bardzo często w zadaniach praktycznych jest tak, że jedynymi interesującymi rozwiązaniami problemów programowania matematycznego w ogólności, a liniowego w szczególności, są rozwiązania całkowite.
Dla przykładu, jeśli x_i są zmiennymi decyzyjnymi, a więc mogącymi przyjmować wartości 0 lub 1, to najczęściej rozwiązania niecałkowite nie mają żadnej sensownej interpretacji.

Przykład 5.3.1 W pewnym przedsiębiorstwie n osób ma wykonywać n czynności, przy czym:

- każda osoba ma wykonywać jedną (i tylko jedną) czynność,
- koszt przystosowania $i-$tej osoby do wykonywania $j-$ej czynności wynosi d_{ij}.

Zadanie polega na przyporządkowaniu każdej z osób zadania w taki sposób, by suma kosztów była możliwie najmniejsza.
Połóżmy

$$x_{ij} = \begin{cases} 1 & \text{jeśli } i-\text{tej osobie przydzielono } j-\text{te zadanie} \\ 0 & \text{w przeciwnym przypadku.} \end{cases}$$

Problem programowania liniowego jaki otrzymamy dla tego zadania jest następujący.
Problem przydziału:

$$\begin{array}{rcll} \sum_{i=1}^{n} x_{ij} & = & 1 & (j = 1, ..., n) \\ \sum_{j=1}^{n} x_{ij} & = & 1 & (i = 1, ..., n) \\ x_{ij} & \in & \{0,1\} & \\ \hline \sum_{i=1}^{n}\sum_{j=1}^{n} d_{ij}x_{ij} & \to & \min & \end{array}$$

Okazuje się, że zadanie programowania liniowego w liczbach całkowitych jest problemem trudnym, któremu jest poświęcona bardzo obszerna literatura. Zainteresowanego czytelnika odsyłam do rozdziału osiemnastego [20], w którym poza wielu innymi rezultatami znajdzie twierdzenie mówiące, że jeśli współczynniki macierzy $\mathbf{A}$ i wektora $\mathbf{b}$ są wymierne, to problem

$$\begin{array}{rcl} \mathbf{Ax} & \leq & \mathbf{b} \\ \mathbf{x} & - & \text{wektor o współczynnikach całkowitych} \\ \hline \mathbf{cx} & \to & \max \end{array}$$

jest $\mathcal{NP}-$zupełny.

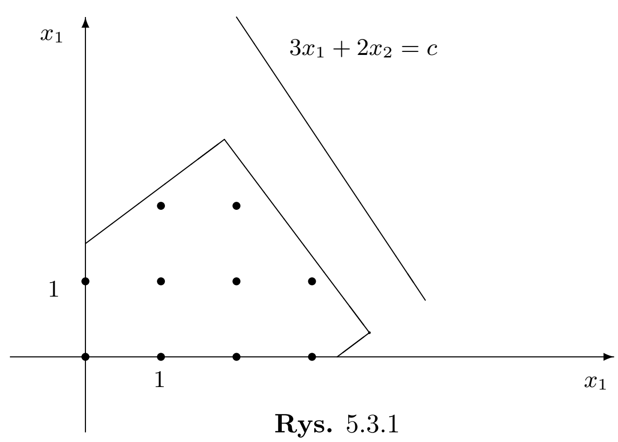
Ogólna postać programowania liniowego w liczbach całkowitych jest następująca:

$$\left\{ \begin{array}{rcl} \mathbf{Ax} & \leq & \mathbf{b} \\ \mathbf{x} & \in & \mathbf{Z}_+^n \\ \hline \mathbf{cx} & \to & \max \end{array} \right. \tag{5.14}$$

Przykład 5.3.2 Rozważmy PPL

$$\left\{ \begin{array}{rcr} -3x_1 + 4x_2 & \leq & 6 \\ 4x_1 + 3x_2 & \leq & 16 \\ 3x_1 - 4x_2 & \leq & 10 \\ x_1 + x_2 & \in & \mathbf{Z}_+ \\ \hline 3x_1 + 2x_2 & \to & \max \end{array} \right. \tag{5.15}$$

Na rysunku 5.3.1 przedstawiono interpretację geometryczną zadania (5.15).



Rys. 5.3.1

łatwo zauważyć, że rozwiązaniem optymalnym problemu 5.15) jest $x_1 = 3, x_2 = 1$, a więc punkt znajdujący sie wewnątrz zbioru rozwiązań dopuszczalnych problemu powstałego z (5.15) przez usunięcie założeń o całkowitości zmiennych. Taki punkt jest oczywiście nie do wykrycia przez algorytm sympleks.□

W dalszym ciągu zajmiemy się sytuacją w której problem programowania liniowego w liczbach całkowitych ma rozwiązanie które można otrzymać stosując metodę sympleks. W rozdziale 8 okaże się, że z takimi właśnie sytuacjami mamy do czynienia częściej niż mogłoby się, na pierwszy rzut oka, wydawać.

Definicja 5.3.1 *Mówimy, że macierz A jest* **totalnie unimodularna**, *jeżeli dla każdej podmacierzy kwadratowej A' macierzy A zachodzi* $\det(A') \in \{0, 1, -1\}$.

Przykład 5.3.3 Z łatwością można sprawdzić, że macierz

$$A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}$$

jest totalnie unimodularna. ■

Z definicji macierzy unimodularnych wynika natychmiast, że jej współrzędne są równe 0, 1 lub −1.
Udowodnimy następujące twierdzenie.

Twierdzenie 5.3.1 *Niech będzie dany PPL*

$$\begin{cases} \mathbf{Ax} = \mathbf{b} \\ \hline \mathbf{cx} \to \max \end{cases} \tag{5.16}$$

gdzie $A \in Z^{m\times n}$ jest macierzą rzędu m i $\mathbf{b} \in \mathbf{Z}^m$. Jeśli dla każdej macierzy bazowej $\mathbf{A}_B$ zachodzi $\det(\mathbf{A})_B \in \{1, -1\}$, *to*

1. $\mathbf{A}_B^{-1} \in \mathbf{Z}^{m\times m}$,

2. *każde rozwiązanie bazowe $\bar{\mathbf{x}}_B$ jest całkowite.*

Dowód. W podrozdziale 5.1 widzieliśmy, że ograniczenia w problemie (5.16) można zapisać w postaci

$$\mathbf{x}_B = \mathbf{A}_B^{-1}(\mathbf{b} - \mathbf{A}_N\mathbf{x}_N), \tag{5.17}$$

gdzie $\mathbf{A}_B$ i $\mathbf{A}_N$ są pewnymi podmacierzami macierzy $\mathbf{A}$, zaś $\mathbf{A}_B$ jest nieosobliwą macierzą kwadratową rzędu m. Jeśli oznaczymy wyrazy macierzy $\mathbf{A}_B^{-1}$ przez $d_{ij},\ i, j = 1, ..., m$, to

$$d_{ij} = \frac{(-1)^{i+j}\det((\mathbf{A}_B)_{ij})}{\det(\mathbf{A}_B)}$$

gdzie $(\mathbf{A}_B)_{ij}$ jest macierzą powstałą z **A** przez skreślenie $j-$tego wiersza i $i-$tej kolumny. Jest więc zupełnie oczywiste, że $d_{ij} \in \mathbf{Z}$ dla wszystkich $i, j = 1, ..., m$. W rozwiązaniu bazowym $\bar{\mathbf{x}}$ mamy $\bar{\mathbf{x}}_N = \mathbf{0}$ i $\bar{\mathbf{x}}_B = \mathbf{A}_B^{-1}\mathbf{b}$ (por. (5.17)). Ponieważ **b** jest z założenia wektorem o współczynnikach całkowitych, także wektor $\bar{\mathbf{x}}_B$ ma współczynniki całkowite. ■

Twierdzenie 5.3.1 pociąga natychmiast podany niżej ważny wniosek który ostatecznie wyjaśnia rolę macierzy totalnie unimodularnych w całkowitoliczbowym programowaniu liniowym.

Wniosek 5.3.2 *Niech będzie dany PPL*

$$\left\{ \begin{array}{ccc} \mathbf{Ax} & = & \mathbf{b} \\ \hline \mathbf{cx} & \rightarrow & \max \end{array} \right. \tag{5.18}$$

gdzie A jest pewną macierzą totalnie unimodularną, zaś **b** *wektorem o współczynnikach całkowitych.*

1. *Jeżeli (5.18) jest niesprzeczny, to jego rozwiązania bazowe mają wszystkie współrzędne całkowite.*
2. *Jeżeli istnieje rozwiązania optymalne problemu (5.18), to istnieją takie rozwiązania optymalne których wszystkie współrzędne są całkowite.*
3. *Jeśli problem (5.18) jest niesprzeczny i ograniczony to algorytm sympleks znajduje takie rozwiązanie (5.18) którego wszystkie współrzędne są liczbami całkowitymi.* ■

Wobec wniosku 5.3.2 ważnym jest problem żozpoznania"macierzy totalnie unimodularnych. W rozdziale 8 okaże się jak ważnym warunkiem wystarczającym dla tych macierzy jest poniższe twierdzenie.

Twierdzenie 5.3.3 *Niech* $\mathbf{A} \in \mathbf{R}^{m \times n}$ *będzie macierzą spełniającą następujące warunki:*

1. *każdy wyraz macierzy* **A** *jest równy* $1, -1$ *lub* 0,
2. *każda kolumna macierzy* **A** *zawiera co najwyżej dwa wyrazy niezerowe,*
3. *istnieje taki podział zbioru wierszy na dwa podzbiory* W_1 *i* W_2 *takie, że:*
   a) *jeśli istnieją dwa wyrazy tego samego znaku występujące w jednej kolumnie, to jeden z nich jest w wierszu zbioru* W_1 *a drugi w wierszu zbioru* W_2,
   b) *jeśli w pewnej kolumnie występują dwa wyrazy niezerowe przeciwnych znaków, to oba wiersze w których występują należą albo do zbioru* W_1 *albo do zbioru* W_2.

Wtedy **A** *jest macierzą totalnie unimodularną.*

Przykład 5.3.4 Macierz

$$\mathbf{B} = \begin{array}{c} \left[\begin{array}{cccccc} 1 & 0 & -1 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 1 & -1 \end{array} \right] \end{array} \begin{array}{l} W_1 \\ W_1 \\ W_1 \end{array},$$

(podobnie jak macierz z przykładu 5.3.3) spełnia założenia twierdzenia 5.3.3, w więc jest macierzą totalnie unimodularną.

Dowód twierdzenia 5.3.3. Wystarczy wykazać, że dla każdej macierzy $\mathbf{A}$ o n wierszach i n kolumnach spełniającej założenia twierdzenia 5.3.3 $\det(\mathbf{A}) \in \{0, 1, -1\}$. Dowód przeprowadzimy przez indukcję ze względu na n.
Dla $n = 1$ przwdziwość twierdzenia jest oczywista. Przypuśćmy, więc, że $n > 1$ i twierdzenie jest prawdziwe dla macierzy o $n-1$ wierszach i $n-1$ kolumnach. Rozpatrzymy trzy przypadki:

Przypadek 1. Jedna z kolumn macierzy $\mathbf{A}$ ma wszystkie wyrazy równe 0. Wtedy oczywiście $\det(\mathbf{A}) = 0$.

Przypadek 2. W jednej z kolumn jest jeden wyraz różny od zera, powiedzmy że tym wyrazem jest a_{i_0,j_0}. Wtedy $a_{i_0,j_0} \in \{-1, 1\}$. Rozwijamy wyznacznik macierzy$\mathbf{A}$ względem kolumny j_0 i otrzymujemy

$$\det(\mathbf{A}) = (-1)^{i+j} a_{i_0,j_0} \det(\mathbf{A}_{i_0 j_0}), \tag{5.19}$$

gdzie $\mathbf{A}_{i_0 j_0}$ jest macierzą powstałą przez skreślenie w $\mathbf{A}$ wiersza i_0 i kolumny j_0. Zauważmy, że macierz $\mathbf{A}_{i_0 j_0}$ jest macierzą kwadratową o $n-1$ wierszach i kolumnach spełniającą warunki twierdzenia 5.3.3 i wobec tego $\det(\mathbf{A}_{i_0 j_0}) \in \{0, 1, -1\}$. Stąd i (5.19) otrzymujemy $\det(\mathbf{A}) \in \{0, 1, -1\}$.

Przypadek 3. Nie zachodzą przypadki 1 ani 2. Wtedy

$$\sum_{i \in W_1} a_{ij} = \sum_{i \in W_2} a_{ij} \tag{5.20}$$

dla każdego $j = 1, ..., m$. Udowodnienie, że w tym przypadku $\det(\mathbf{A}) = 0$ jest bardzo prostym ćwiczeniem, które pozostawiam czytelnikowi (p. ćwiczenie 5.4.2). ■

5.4 ćwiczenia

Ćwiczenie 5.4.1 Niech będzie dany PPL:

$$\left\{ \begin{array}{rcrrrrl} x_4 & = & 5 & -x_1 & -2x_2 & +x_3 & \\ x_5 & = & 6 & -x_1 & +2x_2 & -x_3 & \\ x_6 & = & 2 & -2x_1 & -x_2 & +2x_3 & \\ x_i & \geq & 0 & & & & (i = 1, ..., 6) \\ \hline z & = & & x_1 & +x_2 & +x_3 & \to \max \end{array} \right.$$

(a) Wykaż, że x_2, x_4, x_5 są zmiennymi bazowymi pewnego słownika.

(b) Znajdź wartość funkcji celu dla tego słownika.

(c) Jakie są wtedy możliwe zmienna wchodzące?

Ćwiczenie 5.4.2 Wykaż, że jeśli zbiór wierszy macierzy $A \in \mathbf{R}^{n\times n}$ można podzielić na dwa rozłączne podzbiory W_1 i W_2 takie, że dla każdego $j = 1, ..., n$ spełniona jest równość (5.20), to $\det A = 0$ (prawdziwe także gdy jeden ze zbiorów W_1, W_2 jest pusty[3]).

Ćwiczenie 5.4.3 Niech będą dane macierz totalnie unimodularna $\mathbf{A}$, układ nierówności

$$\mathbf{Ax} \leq \mathbf{b} \tag{5.21}$$

oraz równoważny mu układ

$$\bar{\mathbf{A}}\mathbf{x} = \mathbf{b} \tag{5.22}$$

powstały z (5.21) przez wprowadzenie zmiennych dodatkowych. Wykaż, że macierz $\bar{\mathbf{A}}$ jest totalnie unimodularna.

Ćwiczenie 5.4.4 Z twierdzenia 5.3.1 nie wynika, że jeśli są spełnione założenia tego twierdzenia, to *wszystkie* rozwiązania problemu (5.16) są całkowite. Skonstruuj przykład PPL w którym macierz ograniczeń $\mathbf{A}$ jest totalnie unimodularna i wektor $\mathbf{b}$ całkowity, dla którego istnieją optymalne rozwiązania niecałkowite.

[3]Na mocy definicji przyjmujemy $\sum_{i\in\emptyset} a_{ij} = 0$.

Rozdział 6

Zadanie ograniczone

W wielu praktycznych zagadnieniach obok ograniczeń

$$\sum_{j=1}^{n} a_{ij}x_j \leq b_i \quad (i = 1, ..., m)$$

występują indywidualne ograniczenia od góry na zmienne (wszystkie lub niektóre postaci

$$x_j \leq g_j \quad (j = 1, ..., n)$$

Zadanie PL z tak określonymi ograniczeniami jest wtedy oczywiście zadaniem w postaci standardowej

$$\sum_{j=1}^{n} a_{ij}x_j \leq b_i \quad (i = 1, ..., m)$$

$$x_j \leq g_j \quad (j = 1, ..., n)$$

$$x_j \geq 0 \quad (j = 1, ..., n)$$

Jednak po dołączeniu zmiennych sztucznych otrzymujemy wtedy macierz ograniczeń o rozmiarze $(m + 2n) \times (m + 2n)$. Jeśli liczba zmiennych (a więc liczba n) jest duża, wielkość problemu który należało będzie rozwiązać może być kłopotliwa. Stąd istotna może się okazać metoda zasugerowana po raz pierwszy przez Dantziga w [5].
Problem postawimy jeszcze ogólniej niż wspomniano wyżej, mianowicie będziemy zakładać ograniczenia nie tylko górne ale i dolne i to **na wszystkie** zmienne, czyli

$$d_j \leq x_j \leq g_j \quad (j = 1, ..., n)$$

przy czym dopuszczać będziemy $-\infty$ dla dolnego i $+\infty$ dla górnego ograniczenia (co oczywiście oznacza, że odpowiednich ograniczeń nie ma) [1] Będziemy też

[1] Jeżeli dolne ograniczenia są skończone, to można zmienne x_j zastąpić nowymi zmiennymi $y_j = x_j - d_j$ i przyjąć standardowe ograniczenia $y_j \geq 0$ (por. Gass [10]). Ograniczenia od góry

zakładali równości $\sum_{j=1}^{n} a_{ij}x_j = b_i,\ (i = 1, ..., m)$, czyli sytuację którą mamy po wprowadzeniu zmiennych sztucznych. Problem nasz będzie więc postaci

$$\begin{array}{lcll} \text{zmaksymalizować} & z = & \sum_{j=1}^{n} c_j x_j & \\ \text{przy ograniczeniach} & & \sum_{j=1}^{n} a_{ij}x_j = b_i & (i = 1, ..., m) \\ & & d_j \le x_j \le g_j & (j = 1, ..., n) \end{array}$$

lub w formie macierzowej

$$\left\{ \begin{array}{lccc} \text{zmaksymalizować} & z & = & \mathbf{cx} \\ \text{przy ograniczeniach} & \mathbf{Ax} & = & \mathbf{b} \\ & \mathbf{d} & \le \mathbf{x} \le & \mathbf{g} \end{array} \right. \tag{6.1}$$

gdzie $\mathbf{A} \in \mathbf{R}^{n\times n},\ \mathbf{b}, \mathbf{d}, \mathbf{g}, \mathbf{x} \in \mathbf{R}^{n\times 1}, \mathbf{c} \in \mathbf{R}^{1\times n}$.

6.1 Sympleks dla zadania ograniczonego

Różnica pomiędzy algorytmem dla zadań ograniczonych a poznanym wcześniej polega, jak zobaczymy, na żądaniu, by w rozwiązaniach bazowych wartości zmiennych niebazowych były równe dolnym lub górnym ograniczeniom [2].

Definicja 6.1.1 *Rozwiązanie* $\bar{\mathbf{x}}$ *układu równań* $\mathbf{Ax} = \mathbf{b}$ *nazywamy* **rozwiązaniem bazowym** *jeżeli* n *współczynników wektora* $\mathbf{x}$ *można podzielić na*

(a) m *zmiennych* **bazowych** *oraz*

(b) $n - m$ *zmiennych* **niebazowych**

w ten sposób, że

1. *kolumny macierzy* $\mathbf{A}$ *których elementy są współczynnikami przy zmiennych bazowych tworzą macierz (kwadratową o* m *wierszach i* m *kolumnach) nieosobliwą,*

2. *wartości zmiennych niebazowych w wektorze* $\bar{\mathbf{x}}$ *równe są ich ograniczeniom górnym lub dolnym.*

Rozwiązanie bazowe $\bar{\mathbf{x}}$ *jest dopuszczalne jeżeli spełnia warunek*

$$\mathbf{d} \le \bar{\mathbf{x}} \le \mathbf{g}$$

---

na poszczególne zmienne w praktyce występują prawie zawsze. Na przykład, jeśli zmienne x_j oznaczają wielkość produktów X_j, $j = 1, ..., n$, to górne ograniczenia mogą oznaczać popyt na odpowiednie produkty.

[2] W metodzie sympleks z którą obcujemy od trzeciego rozdziału przyjmujemy wartość 0 dla zmiennych niebazowych, a więc także wartość *indywidualnego* ograniczenia (dolnego, innego nie bylo) na zmienne.

Przykład 6.1.1 Dla problemu

$$
\begin{array}{lrrrrrrcl}
\text{zmaksymalizuj} & x_1+ & x_2+ & x_3+ & x_4+ & x_5+ & x_6 & & \\
\text{przy warunkach} & x_1+ & 2x_2+ & x_3+ & 3x_4+ & 5x_5+ & 7x_6 & = & 39 \\
 & 2x_1+ & 3x_2+ & x_3+ & 4x_4+ & + & x_6 & = & 19 \\
 & 2 \le & x_1 & \le & 3 & & & & \\
 & 1 \le & x_2 & \le & 4 & & & & \\
 & 5 \le & x_3 & \le & 8 & & & & \\
 & 0 \le & x_4 & \le & 6 & & & & \\
 & 0 \le & x_5 & \le & 2 & & & & \\
 & 1 \le & x_6 & \le & 4 & & & &
\end{array}
$$

zmienne x_2, x_6 są[3] bazowymi, zaś x_1, x_3, x_4, x_5 niebazowymi, a bazowym rozwiązaniem dopuszczalnym jest $\bar{\mathbf{x}} = [2, 2, 5, 0, 0, 4]^T$. ■

Oczywiście dla danego wyboru zmiennych bazowych może być więcej niż jedno rozwiązanie bazowe (por. ćwiczenie 6.2.1).

Dalszy ciąg postępowania będzie w dużej mierze przypominał metodę sympleks: będziemy utrzymywać ograniczenie

$$\mathbf{Ax} = \mathbf{b}$$

starając sie wybierać nową bazę i nowe rozwiązania bazowe tak, by powiększać wartość funkcji celu. Zacznijmy od przykładu.

Przykład 6.1.2

$$
\begin{array}{rrrrrcl}
\text{zmaksymalizuj} & x_1+ & 2x_2+ & 3x_3+ & 2x_4 & & \\
\text{przy ograniczeniach:} & x_1+ & x_2+ & 2x_3+ & x_4 & = & 110 \\
 & x_1+ & 3x_2+ & x_3+ & 2x_4 & = & 120 \\
 & 0 \le & x_1 \le & 40 & & & \\
 & 10 \le & x_2 \le & 30 & & & \\
 & 20 \le & x_3 \le & 30 & & & \\
 & 0 \le & x_4 \le & 20 & & &
\end{array}
$$

Przyjmijmy jako zmienne bazowe x_1, x_2 i weźmy rozwiązanie bazowe

$$x_1 = 30,\ x_2 = 20,\ x_3 = 30,\ x_4 = 0$$

Wartość funkcji celu dla naszego rozwiązania bazowego wynosi $z = 160$.
Będziemy, tak jak w metodzie sympleks, zmieniać wartości jednej ze zmiennych niebazowych tak, aby powiększała się wartość rozwiązania, przy zachowaniu warunków ograniczających, a w szczególności pierwszych dwóch równań warunków ograniczających.
Wybierzmy zmienną x_4 jako tę zmienną niebazową niebazową której powiększenie powoduje zwiększanie funkcji celu.

[3] właściwie *mogą być*

Zauważmy, że jeżeli podstawimy

$$\begin{array}{lcrcl} x_1 & = & 30 & - & \frac{1}{2}t \\ x_2 & = & 20 & - & \frac{1}{2}t \\ x_3 & = & 30 & & \\ x_4 & = & t & & \qquad (t \geq 0) \end{array}$$

to także takie zmienne spełniają oba równania w ograniczeniach naszego zdania PL.
t możemy zwiększyć do 20. Wtedy otrzymamy

$$\begin{array}{lcl} \bar{x}_1 & = & 20 \\ \bar{x}_2 & = & 10 \\ \bar{x}_3 & = & 30 \\ \bar{x}_4 & = & 20 \end{array}$$

Przyjąć więc można: x_1 oraz x_4 jako zmienne bazowe oraz x_2 i x_3 jako zmienne niebazowe. Nowe rozwiązanie ma wartość $\bar{z} = 170$. Teraz zmienną wchodzącą będzie x_2. Podobnie jak poprzednio zauważmy, że podstawiając

$$\begin{array}{lcrcr} x_1 & = & 20 & + & t \\ x_2 & = & 10 & + & t \\ x_3 & = & 30 & & \\ x_4 & = & 20 & - & 2t \end{array}$$

t możemy zwiększyć do 10 (rozpoczynając od $t = 0$), wtedy jednak funkcja celu maleje – a więc naszego rozwiązania nie da się już poprawić. ■

W ogólnym przypadku możemy napisać rozważany PPL postaci macierzowej, tak jak to zrobiliśmy w prezentacji zrewidowanej metodzie sympleks. Przyjmijmy więc następujące oznaczenia:

$\mathbf{A}_N$ – macierz otrzymana przez wybór z $\mathbf{A}$ kolumn o indeksach niebazowych,

$\mathbf{B}$ – macierz powstała z $\mathbf{A}$ przez wybór kolumn o indeksach bazowych,

$\mathbf{x}_N, \mathbf{x}_B, \mathbf{c}_N, \mathbf{c}_B$ – odpowiednie wektory otrzymane z $\mathbf{x}$ i $\mathbf{c}$.

Wtedy ograniczenie $\mathbf{Ax} = \mathbf{b}$ można zapisać w postaci

$$\mathbf{A}_N\mathbf{x}_N + \mathbf{B}\mathbf{x}_B = \mathbf{b}$$

i

$$\mathbf{x}_B = \mathbf{B}^{-1}\mathbf{b} - \mathbf{B}^{-1}\mathbf{A}_N\mathbf{x}_N$$

Funkcja celu będzie postaci

$$\mathbf{cx} = \mathbf{c}_B(\mathbf{B}^{-1}\mathbf{b} - \mathbf{B}^{-1}\mathbf{A}_N\mathbf{x}_N) + \ c_N\mathbf{x}_N$$

$$\mathbf{cx} = \mathbf{c}_B\mathbf{B}^{-1}\mathbf{b} + (\mathbf{c}_N - \mathbf{c}_B\mathbf{B}^{-1}\mathbf{A}_N)\mathbf{x}_N$$

lub krócej

$$\mathbf{cx} = \mathbf{yb} + (\mathbf{c}_N - \mathbf{yA}_N)\mathbf{x}_N$$

gdzie

$$\mathbf{y} = \mathbf{c}_B\mathbf{B}^{-1}$$

Niech teraz $\bar{\mathbf{x}}$ będzie rozwiązaniem dopuszczalnym i bazowym problemu i niech

x_j będzie pewną współrzędną niebazową wektora $\mathbf{x}$, zaś

$\mathbf{a} = (\mathbf{A_N})_j$ – kolumną macierzy $\mathbf{A}_N$ odpowiadającą współrzędnej x_j wektora $\mathbf{x}$ (a więc $j-$tą kolumną macierzy A).

Wektor $\bar{\mathbf{x}}$ spełnia równanie

$$\bar{\mathbf{x}}_B = \mathbf{B}^{-1}\mathbf{b} - \mathbf{B}^{-1}\mathbf{A}_N\bar{\mathbf{x}}_N \tag{6.2}$$

Utwórzmy teraz nowy wektor $\bar{\bar{\mathbf{x}}}$ spełniający ograniczenia, czyli taki, że

$$\bar{\bar{\mathbf{x}}} = \mathbf{B^{-1}b} - \mathbf{B^{-1}A_N}\bar{\bar{\mathbf{x}}}_\mathbf{N}$$

gdzie $\bar{\bar{x}}_j + t$, zaś pozostałe współrzędne niebazowe wektora $\bar{\bar{\mathbf{x}}}$ nie ulegają zmianie (to znaczy $\bar{\bar{x}}_k = \bar{x}_j$ dla niebazowych indeksów $k \neq j$). Będziemy wtedy mieli

$$\bar{\bar{\mathbf{x}}}_\mathbf{B} = \mathbf{B^{-1}b} - \mathbf{B^{-1}A_N}\bar{\mathbf{x}}_\mathbf{N} - \mathbf{tB^{-1}a} \tag{6.3}$$

Ze wzorów (6.2) i (6.3)

$$\bar{\bar{\mathbf{x}}}_\mathbf{B} = \bar{\mathbf{x}}_\mathbf{B} - \mathbf{tB^{-1}a}$$

a oznaczając $\mathbf{d} = \mathbf{B}^{-1}\mathbf{a}$

$$\bar{\bar{\mathbf{x}}}_\mathbf{B} = \bar{\mathbf{x}}_\mathbf{B} - \mathbf{td}$$

($\mathbf{d}$ to wektor który w przykładzie 6.1.2 był równy $\mathbf{d} = [-\frac{1}{2}, -\frac{1}{2}]^T$).
Wartość funkcji celu $\mathbf{cx}$ dla $\mathbf{x} = \bar{\bar{\mathbf{x}}}$ jest równa

$$\mathbf{c}\bar{\bar{\mathbf{x}}} = \mathbf{yb} + (\mathbf{c_N} - \mathbf{yA_N})\bar{\mathbf{x}}_\mathbf{N} + (\mathbf{c_j} - \mathbf{ya})\mathbf{t}$$

(niebazową część wektora $\bar{\bar{\mathbf{x}}}$ otrzymaliśmy z $\bar{\mathbf{x}}_N$ przez zastąpienie x_j przez zastąpienie $j-$tej współrzędnej $\bar{x}_j$ przez $\bar{x}_j + t$, pozostawiając inne współrzędne niebazowe bez zmiany).
Stąd wynika ważna obserwacja.

Obserwacja 6.1.1 *Jeżeli w rozwiązaniu bazowym $\bar{\mathbf{x}}$ zastąpimy $\bar{x}_j$ przez $\bar{\bar{x}}_j = \bar{x}_j + t$, a pozostałe współrzędne niebazowe wektora $\bar{\mathbf{x}}$ nie ulegną zmianie, to*

$$\bar{\bar{\mathbf{x}}}_\mathbf{B} = \bar{\mathbf{x}} - \mathbf{td}$$

gdzie $\mathbf{d} = \mathbf{B}^{-1}\mathbf{a}$, $\mathbf{a}$ *jest $j-$tą kolumną* $\mathbf{A}_N$ *oraz*

$$\mathbf{c}\bar{\bar{\mathbf{x}}} = \mathbf{c}\bar{\mathbf{x}} + (\mathbf{c_j} - \mathbf{ya})\mathbf{t}$$

Naszym celem jest zwiększenie wartości funkcji celu. Jest to możliwe w każdej z następujących sytuacji:

1. $c_j - \mathbf{ya} > 0,\ \bar{x}_j < g_j$
   Wartości funkcji celu można zwiększyć przyjmując pewne $t > 0$.

2. $\bar{x}_j > d_j,\ c_j - \mathbf{ya} < 0$
   W tym przypadku wartość z można zwiększyć przyjmując $\bar{\bar{x}}_j = \bar{x}_j + t$ dla pewnego t ujemnego.

Po tych objaśnieniach możemy przedstawić nasz algorytm następująco:

Algorytm sympleks dla PPL (6.1) (ze zmiennymi ograniczonymi)
Niech $\bar{\mathbf{x}}$ będzie rozwiązaniem bazowym dopuszczalnym.

Krok 1. Rozwiąż układ $\mathbf{yB} = \mathbf{c}_B$.

Krok 2. Wybór zmiennej wchodzącej:
Zmienną wchodzącą może byc zmienna niebazowa x_j taka, że

Przypadek (a): $\bar{x}_j < g_j$ i $c_j > \mathbf{ya}$ lub

Przypadek (b): $\bar{x}_j > d_j$ i $c_j < \mathbf{ya}$

gdzie $\mathbf{a}$ jest kolumną macierzy $\mathbf{A}$ odpowiadającą zmiennej x_j.
Jeśli ani przypadek (a) ani przypadek (b) nie zachodzi dla żadnej zmiennej niebazowej **STOP**: $\mathbf{x}$ jest rozwiązaniem optymalnym.

Krok 3. Rozwiąż układ $\mathbf{Bd} = \mathbf{a}$.
Zdefiniujmy: $x_j(t) = \bar{x}_j + t,\ \mathbf{x}_B(t) = \bar{\mathbf{x}}_B - \mathbf{td}$.

Krok 4. Przypadek (a). Jeśli dla dowolnego t dodatniego:

$$d_j \le x_j(t) \le g_j \quad \text{i} \quad \mathbf{d}_B \le \mathbf{x}_B(t) \le \mathbf{g}_B$$

(gdzie $\mathbf{d}_B$ i $\mathbf{g}_B$ są wektorami powstałymi z wektorów $\mathbf{d}$ i $\mathbf{g}$ przez opuszczenie współrzędnych niebazowych) STOP: problem jest nieograniczony.

Przypadek (b). Jeśli dla dowolnego $t < 0$

$$d_j \le x_j(t) \le g_j \quad \text{i} \quad \mathbf{d}_B \le \mathbf{x}_B(t) \le \mathbf{g}_B$$

STOP: problem jest nieograniczony.

Krok 5. Przypadek (a). Weźmy t_0 maksymalne spełniające

$$d_j \le x_j(t_0) \le g_j \quad \text{i} \quad \mathbf{d}_B \le \mathbf{x}_B(t_0) \le \mathbf{g}_B \tag{6.4}$$

Przypadek (b). Niech t_0 będzie minimalne spełniające

$$d_j \le x_j(t_0) \le g_j \quad \text{i} \quad \mathbf{d}_B \le \mathbf{x}_B(t_0) \le \mathbf{g}_B \tag{6.5}$$

Jeśli t_0 jest takie, że

$$d_j = x_j(t_0) \quad \text{lub} \quad g_j = x_j(t_0)$$

to idź do **Kroku 2** (x_j pozostaje zmienną niebazową). W przeciwnym przypadku wybierz indeks bazowy i_0 dla którego zachodzi

$$x_{i_0}(t_0) = d_{i_0} \quad \text{lub} \quad x_{i_0}(t_0) = g_{i_0}$$

x_{i_0} **jest zmienną wychodzącą.** ∎

Czytelnikowi pozostawiam wykazanie poniższych twierdzeń: pierwszy dowód jest bardzo łatwy, drugi wymaga zaadaptowania dowodu twierdzenia 3.3.2.

Twierdzenie 6.1.2 *Jeśli metoda sympleks dla problemu ograniczonego (6.1) nie doprowadza do rozwiązania w skończonej liczbie kroków, to algorytm wpada w pętlę.* ∎

Twierdzenie 6.1.3 *Jeśli do wyboru zmiennych wchodzących i wychodzących w algorytmie sympleks dla problemu ograniczonego (6.1) stosujemy regułę Blanda, to algorytm kończy się po skończonej liczbie iteracji.* ∎

6.2 Inicjalizacja

Podobnie jak w przypadku standardowego PPL także w przypadku zadania ograniczonego istnieje problem *inicjalizacji*, to znaczy znalezienia jakiegokolwiek bazowego rozwiązania dopuszczalnego czyli pierwszego dopuszczalnego słownika. Także w tym przypadku w poszukiwanie takiego słownika wprzęgniemy metodę sympleks.
Z zadaniem

$$\left\{ \begin{array}{lll} \text{zmaksymalizuj} & \sum_{j=1}^n c_j x_j & \\ \text{przy ograniczeniach} & \sum_{j=1}^n a_{ij} x_j = b_i & (i = 1, ..., m) \\ & d_j \le x_j \le g_j & (j = 1, ..., n) \end{array} \right. \tag{6.6}$$

skojarzymy zadanie PL w którym ograniczenia będą postaci:

$$\left\{ \begin{array}{lllllll} \sum_{j=1}^n a_{ij} x_j & + & x_{n+i} & = & b_i & & (i = 1, ..., m) \\ & d_k & \le & x_k & \le & g_k & (k = 1, ..., n+m) \end{array} \right. \tag{6.7}$$

przy czym dla $k > n$ dolne i górne ograniczenia we wzorze (6.7) są wyznaczane w sposób opisany poniżej.
Niech $\tilde{x}_1, ..., \tilde{x}_{n+m}$ będą dane następującymu wzorami:

$$\tilde{x}_j = d_j \quad \text{lub} \quad \tilde{x}_j = g_j \quad \text{dla} \quad j = 1, ..., n$$

$$\tilde{x}_{n+i} = b_i - \sum_{j=1}^{n} a_{ij}\tilde{x}_j \quad (i = 1, ..., m)$$

Jeśli $\tilde{x}_{n+i} \geq 0$ to w (6.7) przyjmujemy $d_{n+i} = 0,\ g_{n+i} = +\infty$, jeśli zaś $\tilde{x}_{n+i} < 0$ to $d_{n+i} = -\infty,\ g_{n+i} = 0$.
Jest oczywiste, że (6.6) ma rozwiązanie dopuszczalne wtedy i tylko wtedy, gdy (6.7) ma rozwiązanie w którym $x_{n+i} = 0$ $(i = 1, ..., m)$. Problemem ograniczonym PL który wystarczy rozwiązać jest

$$\left\{\begin{array}{lcccll} \sum_{i=1}^{m} w_{n+i}x_{n+i} & \rightarrow & \max & & & \\ \hline \sum_{j=1}^{n} a_{ij}x_j & + & x_{n+i} & = & b_i & (i = 1, ..., m) \\ & d_k & \leq & x_k & \leq\ g_k & (k = 1, ..., n+m) \end{array}\right. \tag{6.8}$$

gdzie $w_{n+i} = \left\{\begin{array}{rlcl} 1 & \text{gdy} & d_{n+i} & = 0 \\ -1 & \text{gdy} & g_{n+i} & = 0 \end{array}\right.$. Problem (6.8) jest niesprzeczny, $\tilde{x}_1, ..., \tilde{x}_{n+m}$ jest jego rozwiązaniem dopuszczalnym. Jeśli (6.8) ma rozwiązanie optymalne zerowe ($x_{n+i} = 0,\ i = 1, ..., m$), to pierwszych n współrzędnych tego rozwiązania $x_1, ..., x_n$ jest rozwiązaniem dopuszczalnym (6.6).

6.2.1 ćwiczenia

Ćwiczenie 6.2.1 Znajdź wszystkie rozwiązania bazowe odpowiadające w przykładzie 6.1.1 zmiennym bazowym x_2, x_6. Znajdź inne zbiory zmiennych bazowych dla tego przykładu.

Ćwiczenie 6.2.2 Rozwiąż PPL

$$\left\{\begin{array}{rclcl} x_1 - x_2 + 2x_3 & \leq & 7 & & \\ x_1 + 2x_2 - x_3 & \leq & 6 & & \\ -1 & \leq & x_1 & \leq & 2 \\ 1 & \leq & x_2 & \leq & 3 \\ -\infty & \leq & x_3 & \leq & 4 \\ \hline 2x_1 + 3x_2 + x_3 & \rightarrow & \max & & \end{array}\right.$$

Ćwiczenie 6.2.3 Dany jest PPL:

$$\begin{array}{rcr} z = 5x_1 + 2x_2 - 3x_3 + 3x_4 + 6x_5 + x_6 & \rightarrow & \max \\ \hline 3x_1 + x_2 - 4x_3 + 2x_4 + 5x_5 + x_6 & \leq & 3 \\ -5x_1 + 4x_2 + 2x_3 - 3x_4 + 2x_5 + 3x_6 & \leq & 25 \\ x_1 + x_2 + 2x_3 + x_4 + x_5 + 2x_6 & = & 4 \\ x_1 \geq 0,\quad 2 \leq x_2 \leq 10,\quad x_3 \leq 0,\quad -3 \leq x_4 \leq 3 & & \end{array}$$

i jego pierwszy słownik

$$\begin{array}{rcr} z = 5x_1 + 2x_2 - 3x_3 + 3x_4 + 6x_5 + x_6 & \rightarrow & \max \\ \hline 3x_1 + x_2 - 4x_3 + 2x_4 + 5x_5 + x_6 + x_7 & = & 3 \\ -5x_1 + 4x_2 + 2x_3 - 3x_4 + 2x_5 + 3x_6 + x_8 & = & 25 \\ x_1 + x_2 + 2x_3 + x_4 + x_5 + 2x_6 & = & 4 \\ x_1 \geq 0,\quad 2 \leq x_2 \leq 10,\quad x_3 \leq 0,\quad -3 \leq x_4 \leq 3 & & \end{array}$$

Przyjmując jako zmienne bazowe x_1, x_6, x_8 i rozwiązanie bazowe

$$(x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8) = (9, 2, 0, -3, -4, 0, 0, 61)$$

wykonaj następną iterację biorąc jako zmienną wchodzącą x_5.

Rozdział 7

Interpretacje i zastosowania

7.1 Interpretacja geometryczna

Naszej wyobraźni dostępnw jest interpretacja geometryczna przestrzeni $\mathbf{R}^\mathbf{n}$ w przypadkach $n = 1, 2$ i 3 (p. podrozdział 7.1.3). Przypadek $n = 1$ jest banalny, zajmiemy się więc tylko przypadkami kiedy $n = 2$ i $n = 3$.

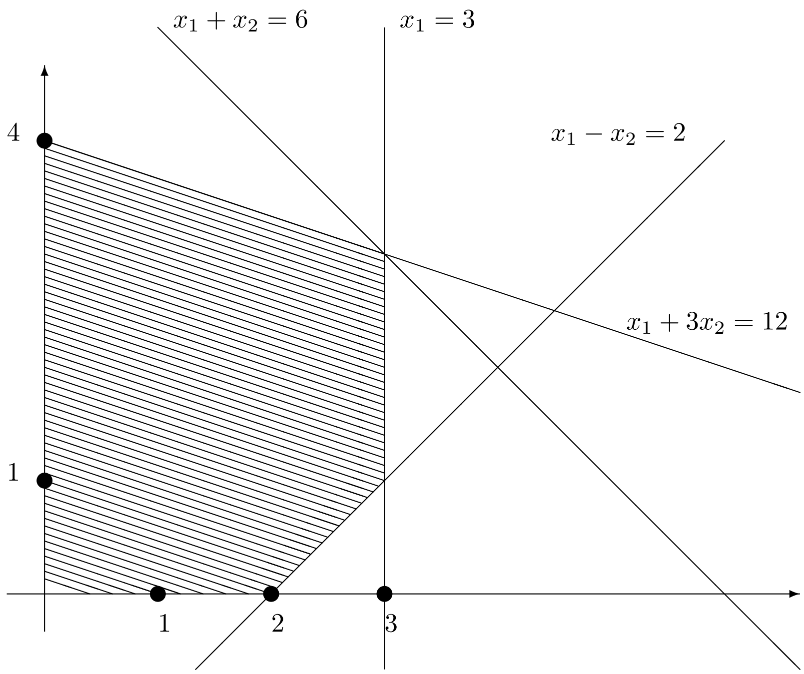
7.1.1 $n = 2$

Przyjrzyjmy się poniższemu przykładowi.

Przykład 7.1.1

$$\begin{array}{lcccrcr}
\text{zmaksymalizuj} & x_1 & + & & x_2 & & \\
\text{przy ograniczeniach} & x_1 & - & & x_2 & \leq & 2 \\
 & x_1 & + & & 3x_2 & \leq & 12 \\
 & x_1 & & & & \leq & 3 \\
 & & & & x_1, x_2 & \geq & 0
\end{array}$$

Zbiór punktów (x_1, x_2) płaszczyzny spełniających warunek $x_1 - x_2 \leq 2$ to wszystkie punkty płaszczyzny leżące nad prostą $x_1 - x_2 = 2$, punkty (x_1, x_2) spełniające $x_1 + 3x_2 \leq 12$ to wszystkie punkty leżące poniżej prostej $x_1 + 3x_2 = 12$, zbiór punktów (x_1, x_2) spełniających $x_1 \leq 3$ to punkty na lewo od prostej $x_1 = 3$ (oczywiście w każdym przypadku włacznie z punktami leżącymi na odpowiedniej prostej. Ograniczenia $x_1, x_2 \geq 0$ oznaczają, że chodzi nam o punkty pierwszej ćwiartki.
Tak więc szukać będziemy maksimum funkcji $z = x_1 + x_2$ dla (x_1, x_2) należących do wielokąta ograniczonego prostymi: $x_1 - x_2 = 2$, $x_1 + 3x_2 = 12$, $x_1 = 0$, $x_1 = 3$, $x_2 = 0$ (rys. 7.1.1).



Rys.6.1

Maksymalizacja funkcji $z = x_1 + x_2$ oznacza, że powinniśmy znaleźć taką maksymalną stała C, żeby prosta $x_1 + x_2 = c$ miała co najmniej jeden punkt wspólny z naszym wielobokiem. Inaczej mówiąc, musimy prostą $x_1 + x_2 = 0$ przesunąć równolegle najwyżej jak to możliwe, tak jednak aby miała z wielobokiem przecięcie niepuste. Jest jasne, że optymalne położenie prostej będzie takie, w którym jedynym punktem wspólnym prostej i wieloboku będzie $P = (3, 3)$. Prosta będzie miała równanie $(x_1 - 3) + (x_2 - 3) = 0$, a więc $x_1 + x_2 = 6$ i wartość maksymalna $z = 6$.$\square$

Zauważmy przy okazji, że z sytuacją więcej niż jednego rozwiązania będziemy mieli do czynienia w pewien sposób rzadko: odpowiedni bok wieloboku będzie musiał być równolegly do prostej reprezentującej funkcję celu. Zbiór rozwiązań optymalnych będzie wtedy nie punktem jak w przykładzie 7.1.1, a zbiorem punktów pewnego odcinka (boku wieloboku).

Dla ograniczeń zadanych w przykładzie 7.1.1 będzie tak na przykład wtedy, gdy funkcja celu będzie zadana wzorem $z = x_1 + 3x_2$. Wtedy zbiorem rozwiązań optymalnych będzie odcinek łączący punkty $(0, 4)$ i $(3, 3)$, a wartością optymalnę $z = 12$.

7.1.2 n=3

Studiowanie tego przypadku połączymy z badaniem przykładu Klee–Minty'ego dla $n = 3$.

Przykład 7.1.2 Dla $n = 3$ przykład Klee–Minty'ego wygląda następująco:

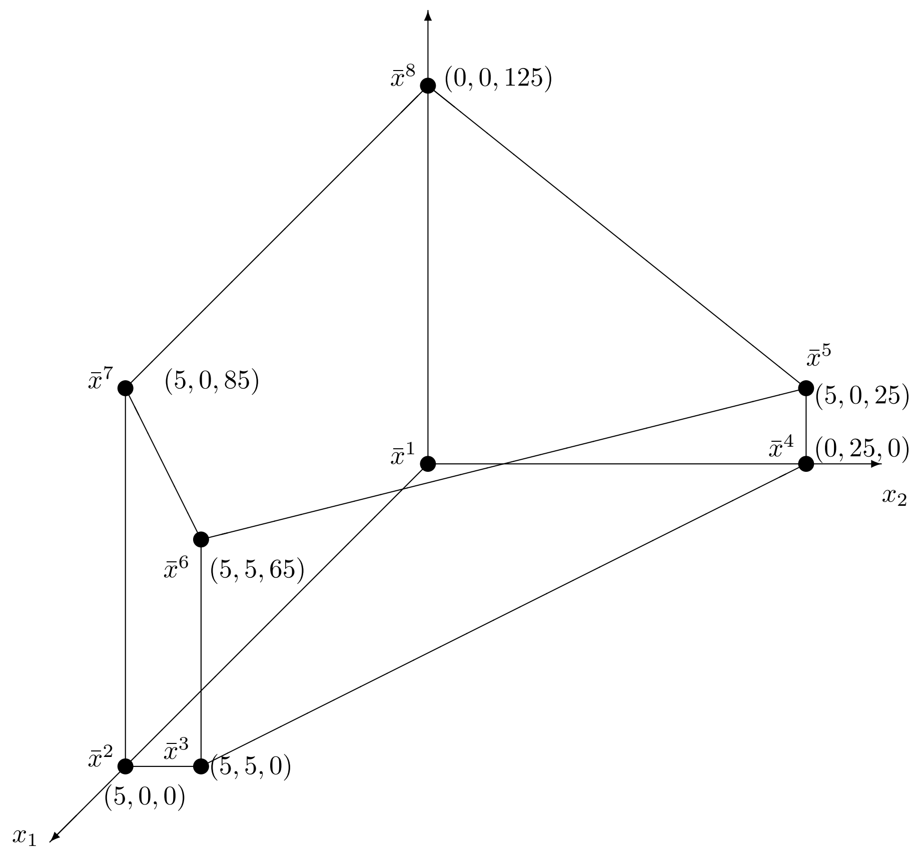
$$
\begin{array}{llll}
\text{zmaksymalizować} & z = & 4x_1 + 2x_2 + x_3 & \\
\text{przy ograniczeniach} & & x_1 & \leq 5 \\
 & & 4x_1 + x_2 & \leq 25 \\
 & & 8x_1 + 4x_2 + x_3 & \leq 125 \\
 & & x_1, x_2, x_3 & \geq 0
\end{array}
$$

Czytelnik który wykonał ćwiczenie 3.7.8 otrzymał osiem słowników i tyleż rozwiązań bazowych:

$$
\mathbf{x^1} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 5 \\ 25 \\ 125 \end{bmatrix}
\mathbf{x^2} = \begin{bmatrix} 5 \\ 0 \\ 0 \\ 0 \\ 5 \\ 85 \end{bmatrix}
\mathbf{x^3} = \begin{bmatrix} 5 \\ 5 \\ 0 \\ 0 \\ 0 \\ 65 \end{bmatrix}
\mathbf{x^4} = \begin{bmatrix} 0 \\ 25 \\ 0 \\ 0 \\ 0 \\ 125 \end{bmatrix}
$$

$$
\mathbf{x^5} = \begin{bmatrix} 0 \\ 25 \\ 25 \\ 5 \\ 0 \\ 0 \end{bmatrix}
\mathbf{x^6} = \begin{bmatrix} 5 \\ 5 \\ 65 \\ 0 \\ 0 \\ 0 \end{bmatrix}
\mathbf{x^7} = \begin{bmatrix} 5 \\ 0 \\ 85 \\ 0 \\ 5 \\ 0 \end{bmatrix}
\mathbf{x^8} = \begin{bmatrix} 0 \\ 0 \\ 125 \\ 5 \\ 25 \\ 0 \end{bmatrix}
$$

(trzy współrzędne tych rozwiązań odpowiadają zmiennym decyzyjnym x_1, x_2, x_3 podczas gdy trzy następne zmiennym sztucznym y_1, y_2, y_3).



Przykład Klee–Minty'ego ($n = 3$)

Rys. 6.2

Przyjrzyjmy się tej sytuacji na rysunku (rys. 6.2) na którym oczywiście uwzględniamy jedynie współrzędne x-owe rozwiązań zaś y_1, y_2, y_3 ignorujemy (z oczywistych powodów rysunek jest *przeskalowany* – skala na każdej z osi jest inna).

Zbiorem rozwiązań dopuszczalnych jest wielościan wypukły o wierzchołkach $\bar{x}^1, ..., \bar{x}^8$. ściany wielościanu odpowiadają równościom w ograniczeniach. Od dołu nasz wielobok jest ograniczony przez płaszczyznę $x_3 = 0$, od góry przez płaszczyznę $8x_1 + 4x_2 + x_3 = 125$. Krawędzie wyznaczone są przez przecięcia płaszczyzn. Na przykład krawędź łącząca wierzchołki $(5, 5, 65)$ i $(0, 25, 25)$ zawarta jest w prostej

$$\begin{cases} 8x_1 + 4x_2 + x_3 & = & 125 \\ 4x_1 + x + 2 & = & 25 \end{cases}$$

Wierzchołki to oczywiście przecięcia trzech płaszczyzn – a więc nasze rozwiązania bazowe! Teraz jasnym sie staje, jak bardzo jak bardzo *pechowy* dla algorytmu sympleks jest przykład Klee–Minty'ego. Sympleks wybiera wierzchołki wielościanu, przechodząc kolejno od wierzchołka do wierzchołka sąsiedniego po kzawędzi, za każdym razem zwiększając (a raczej nie zmniejszając) wartość funkcji celu. W przykładzie Klee–Minty'ego zaczyna od najgorszego rozwiązania a następnie najgorszy możliwie wierzchołek sąsiedni. W ten sposób sympleks musi sprawdzić wszystkich osiem dla $n = 3$, a 2^n w ogólnym przypadku, wierzchołków (inaczej: rozwiązań bazowych) – a więc wszystkie możliwe.

7.1.3 Komentarz

W większości książek poświęconych programowaniu liniowemu interpretacja geometryczna prezentowana jest na samym początku, zaraz po przedstawieniu problemu programowania liniowego lub algorytmu sympleks (por. [10], [17], [22]). Taka metoda dydaktyczna ma dużo zalet. Interpretacja geometryczna jest bowiem bardzo prosta i daje dobre intuicje. Niemniej, w tym bardzo pożytecznym geometrycznym ujęciu kryje się pewna pułapka. Sugeruje ono, że wszystko jest tu bardzo proste i łatwe do zobaczenia i narysowania. Tymczasem wcale tak nie jest. Narysować i wyobrazić sobie można jedynie przestrzenie $1-$, $2-$ i $3-$wymiarowe. Próby – niewiedzieć czemu bardzo modne – wyobrażenia sobie przestrzeni więcej niż $3-$wymiarowych [1] są z góry skazane na niepowodzenie z prostego powodu: *przestrzeń $4-$ (i więcej) wymiarowa nie ma swojej interpretacji geometrycznej w przestrzeniach mniej wymiarowych* [2]. $4-$wymiarowa przestrzeń w której czwartym wymiarem jest czasnie stanowi tu zadnej pomocy. Przestrzen $n-$wymiarowa jest dla $n \geq 4$ pojęciem czysto algebraicznym. Fakt, że język tu używany jest zapożyczony z geometrii daje nam tyle, że w naszej wyobraźni jesteśmy w stanie kreować obrazy – interpretacje w przestrzeniach tej wyobraźni dostępnych, czyli $2-$ i $3-$wymiarowych.

7.2 Powłoki wypukłe zbiorów

Definicja 7.2.1 *Mówimy, że podzbiór $W \subset \mathbf{R}^n$ jest* **wypukły** [3] *jeżeli dla każdych dwóch elementów $x \in W$ i $y \in W$ odcinek je łączący:*

$$[x, y] = \{\alpha x + \beta y : \alpha, \beta \geq 0, \alpha + \beta = 1\}$$

jest zawarty w W.

Przykład 7.2.1 Jest oczywistym, że jedynymi podzbiorami wypukłymi $\mathbf{R}$ sa przedziały.

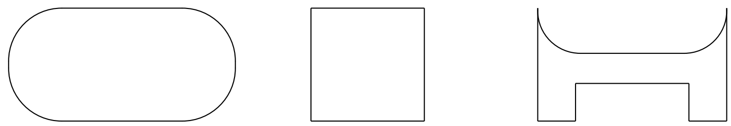
---

[1] Chvatál w swojej książce [4] tak pisze na ten temat: *... do not try to visualize $n-$dimentional objects for $n \geq 4$. Such an effort is not only doomed to failure – it may be dangerous to your mantional health. (If you do succeed, then you are in trouble).*

[2] Bardziej dobitnie i prosto: przestrzeń $4-$wymiarowa nie jest $3-$wymiarowa (bo $3 < 4$).

[3] Zbiory wypukłe odgrywają ważną rolę w wielu działach matematyki, szczególnie w analizie funkcjonalnej (por. [1]) i optymalizacji (por. [11]).

W $\mathbf{R}^2$ zbiory S_1 i S_2 przedstawione na rysunku 7.3.1 są wypukłe, zaś S_3 nie jest zbiorem wypukłym.



S_1 S_2 S_3

Rys. 7.3.1

Banalnie prosty do udowodnienia jest następujący wniosek którego dowód wobec tego pozostawiamy czytelnikowi.

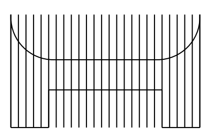
Wniosek 7.2.1 *Niech $\{W_i : i \in I\}$ będzie rodziną zbiorów wypukłych. Wtedy przecięcie $\bigcap_{i \in I} W_i$ jest także zbiorem wypukłym.*$\square$

Wniosek 7.2.1, jakkolwiek by nie był prosty do udowodnienia jest *odpowiedzialny* za bardzo ważne pojęcie *powłoki wypukłej zbioru.*

Definicja 7.2.2 *Niech X będzie dowolnym podzbiorem $\mathbf{R}^n$.* **Powłoką wypukłą** *X nazywamy najmniejszy (w sensie zawierania zbiorów) zbiór wypukły $W \subset \mathbf{R}^n$ zawierający X. Piszemy wtedy*

$$W = conv(X)$$

Przykład 7.2.2 Dla zbioru S_3 z rysunku 7.3.1 powłoka wypukła przedstawiona jest na rysunku 7.3.2.



$conv(S_3)$

Rys. 7.3.2

Wniosek 7.2.2 *Niech X będzie dowolnym zbiorem zawartym w $\mathbf{R}^n$ i niech $\mathcal{W}$ będzie rodziną wszystkich zbiorów wypukłych zawierających X. Wtedy*

$$conv(X) = \bigcap_{W \in \mathcal{W}} W$$

Dowód. Ponieważ $conv(X) \in \mathcal{W}$ $\bigcup_{W \in \mathcal{W}} W \subset conv(X)$. Z drugiej strony, skoro $conv(X)$ jest najmniejszym, w sensie zawierania zbiorów, zbiorem wypukłym zawierającym X, zachodzi $conv(X) \subset W$ dla każdego $W \in \mathcal{W}$. Stąd zaś $conv(X) \subset \bigcap_{W \in \mathcal{W}} W$. $\blacksquare$

Wniosek 7.2.3 *Dla dowolnego zbioru* $X \subset \mathbf{R}^n$ *zachodzi związek* $conv(X) = \{\alpha_1\mathbf{v}_1 + ... + \alpha_k\mathbf{v}_k : k \in \mathbf{N}; \mathbf{v}_1, ..., \mathbf{v}_k \in X; \alpha_1, ..., \alpha_k \geq 0; \alpha_1 + ... + \alpha_k = 1\}$ [4].

Dowód. Jest oczywiste, że zbiór $A = \{\alpha_1\mathbf{v}_1 + ... + \alpha_k\mathbf{v}_k : k \in \mathbf{N}; \mathbf{v}_1, ..., \mathbf{v}_k \in X; \alpha_1, ..., \alpha_k \geq 0; \alpha_1 + ... + \alpha_k = 1\}$ zawiera X (aby otrzymać wszystkie elementy zbioru X wystarczy położyć we wzorze na A $k = 1$ i $\alpha_1 = 1$).
łatwo także wykazać, że A jest zbiorem wypukłym. Rzeczywiście, weźmy $\mathbf{u}, \mathbf{v} \in A$. Wykażemy, że dla wszystkich $\alpha, \beta \in \mathbf{R}, \alpha, \beta \geq 0$ i takich, że $\alpha + \beta = 1$, $\alpha\mathbf{u} + \beta\mathbf{v} \in A$.
Skoro $\mathbf{u}, \mathbf{v} \in A$, to istnieją $k, l \in \mathbf{N}$ oraz $\alpha_1, ..., \alpha_k, \beta_1, ..., \beta_l \geq 0$ oraz $\mathbf{u}_1, ..., \mathbf{u}_k, \mathbf{v}_1, ..., \mathbf{v}_l \in X$ takie, że $\alpha_1 + ... + \alpha_k = \beta_1 + ... + \beta_l = 1$ i $u = \alpha_1\mathbf{u}_1 + ... + \alpha_k\mathbf{u}_k, v = \beta_1\mathbf{v}_1 + ... + \beta_l\mathbf{v}_l$.
Wtedy

$$\begin{aligned} \alpha\mathbf{u} + \beta\mathbf{v} &= \alpha(\alpha_1\mathbf{u}_1 + ... + \alpha_k\mathbf{u}_k) + \beta(\beta_1\mathbf{v}_1 + ... + \beta_l\mathbf{v}_l) \\ &= \alpha\alpha_1\mathbf{u}_1 + ... + \alpha\alpha_k\mathbf{u}_k + \beta\beta_1\mathbf{v}_1 + ... + \beta\beta_l\mathbf{v}_l \end{aligned}$$

Tak więc $\alpha\mathbf{u}+\beta\mathbf{v}$ jest kombinacją wektorów $\mathbf{u}_1, ..., \mathbf{u}_k, \mathbf{v}_1, ..., \mathbf{v}_l \in X$. Co więcej, współczynniki tej kombinacji są oczywiście nieujemne i $\alpha\alpha_1 + ... + \alpha\alpha_k + \beta\beta_1 + ... + \beta\beta_l = \alpha(\alpha_1 + ... + \alpha_k) + \beta(\beta_1 + ...\beta_l) = \alpha + \beta = 1$. Czyli $\alpha\mathbf{u} + \beta\mathbf{v} \in A$.

Dla udowodnienia prawdziwości wniosku 7.2.3 pozostaje jeszcze wykazać, że A jest najmniejszym zbiorem wypukłym zawierającym X czyli, że dla każdego zbioru wypukłego W zawierającego X zachodzi $A \subset W$.
Niech W będzie zbiorem wypukłym, $W \supset X$ i niech $\mathbf{u} \in A$. Istnieją wobec tego $k \geq 1 (k \in \mathbf{N}), \mathbf{u}_1, ..., \mathbf{u}_k \in X, \alpha_1, ..., \alpha_k \geq 0$, takie, że $\alpha_1 + ... + \alpha_k = 1$ i $\mathbf{u} = \alpha_1\mathbf{u}_1 + ... + \alpha_k\mathbf{u}_k$. Korzystając z indukcji matematycznej ze względu na k wykażemy, że $\mathbf{u} \in W$.
Dla $k = 1$ zachodzi uczywiście $u \in X$ i wobec $W \supset X$ zachodzi także $u \in W$.
Przypuśćmy, że $k > 1$ i nasza teza jest prawdą dla $k-1$, to znaczy że każda kombinacja $\beta_1\mathbf{v}_1 + ... + \beta_{k-1}\mathbf{v}_{k-1}$ w której $\mathbf{v}_1, ..., \mathbf{v}_{k-1} \in X$, $\beta_1, ..., \beta_{k-1} \geq 0, \beta_1 + ... + \beta_{k-1} = 1$, należy do W.
Rozważmy teraz kombinację wypukłą k elementów: $\mathbf{u} = \alpha_1\mathbf{v_1} + ... + \alpha_k\mathbf{v_k}$, $\alpha_1, ..., \alpha_k \geq 0, \mathbf{v_1}, ..., \mathbf{v_k} \in X, \alpha_1 + ... + \alpha_k = 1$. Wykażemy, że $\mathbf{u} \in W$.
Możemy założyć, że $\alpha_k \neq 1$ (w przeciwnym przypadku $\mathbf{u} = 1\mathbf{u}_k = \mathbf{u}_k \in X$). Mamy

$$u = \alpha_1\mathbf{u}_1 + ... + \alpha_{k-1}\mathbf{u}_{k-1} + \alpha_k\mathbf{u}_k =$$

$$(1 - \alpha_k)(\frac{\alpha_1}{1 - \alpha_k}\mathbf{u}_1 + ... + \frac{\alpha_{k-1}}{1 - \alpha_k}\mathbf{u}_{k-1}) + \alpha_k\mathbf{u}_k$$

Oczywiście $\frac{\alpha_i}{1-\alpha_k} \geq 0$ dla $i = 1, ..., k - 1$. Co więcej,

$$\frac{\alpha_1}{1 - \alpha_k} + ... + \frac{\alpha_{k-1}}{1 - \alpha_k} = \frac{\alpha_1 + ... + \alpha_{k-1}}{1 - \alpha_k} = \frac{1 - \alpha_k}{1 - \alpha_k} = 1$$

[4]Kombinację $\alpha_1 v_1 + ... + \alpha_k v_k$, gdzie $\alpha_1, ..., \alpha_k \geq 0, \alpha_1 + ... + \alpha_k = 1$ nazywamy **kombinacją wypukłą** $v_1, ..., v_k$.

Stąd i z założenia indukcyjnego $\mathbf{v} = \frac{\alpha_1}{1-\alpha_k}\mathbf{u}_1 + ... + \frac{\alpha_{k-1}}{1-\alpha_k}\mathbf{u}_{k-1} \in W$.
Ponieważ W jest z założenia zbiorem wypukłym

$$\mathbf{u} = (1-\alpha_k)\mathbf{v} + \alpha_k\mathbf{u}_k \in W$$

co kończy dowód wniosku 7.2.3. ■

Wniosek 7.2.3 posłuży nam do wykazania poniżej twierdzenia Carathéodory'ego które mówi, że w przestrzeni n wymiarowej k we wniosku 7.2.3 można zastąpić przez $n+1$.

Twierdzenie 7.2.4 (Carathéodory) *Dla dowolnego zbioru $X \subset \mathbf{R}^n$ zachodzi związek*

$$conv(X) = \{\alpha_1\mathbf{v}_1 + ... + \alpha_k\mathbf{v}_k : k \leq n+1, \mathbf{v}_1, ..., \mathbf{v}_k \in X,$$

$$\alpha_1, ..., \alpha_k > 0, \alpha_1 + ... + \alpha_k = 1\}$$

Dowód. Z wniosku 7.2.3 wynika, że jeśli istnieją k, $\alpha_1, ..., \alpha_k$ oraz $\mathbf{v}_1, ..., \mathbf{v}_k \in X$ spełniające $\alpha_i \geq 0$ $(i = 1..., k)$, $\sum_{i=1}^k \alpha_i = 1$, takie, że $\mathbf{u} = \alpha_1\mathbf{v}_1 + ...\alpha_k\mathbf{v}_k$ to $\mathbf{u} \in conv(X)$ (w szczególności wtedy gdy $k \leq n+1$).
Wystarczy więc wykazać, że jeśli $\mathbf{u} \in conv(X)$, to istnieje nie więcej niż $n+1$ wektorów $\mathbf{u}_1, ..., \mathbf{u}_k$ i skalarów $\alpha_1, ..., \alpha_k$, $\alpha_i > 0$ $(i = 1, ..., k)$, $\sum_{i=1}^k \alpha_i = 1$, takich, że $\mathbf{u} = \alpha_1\mathbf{u}_1 + ...\alpha_k\mathbf{u}_k$. Dzięki wnioskowi 7.2.3 wiemy, że istnieje l, wektory $\mathbf{v}_1, ..., \mathbf{v}_l \in X$ oraz współczynniki $\beta_j > 0$, $j = 1, ..., l, \beta_1 + ... + \beta_l = 1$ takie, że $\mathbf{u} = \beta_1\mathbf{v}_1 + ... + \beta_l\mathbf{v}_l$.
Zauważmy teraz, że równania

$$\begin{cases} \sum_{i=1}^l \beta_i\mathbf{v}_i & = & \mathbf{u} \\ \sum_{i=1}^l \beta_i & = & 1 \end{cases} \tag{7.1}$$

dają nam układ $k+1$ równań rzeczywistych liniowych, których zmiennymi są $\beta_1, ..., \beta_l$ a współczynnikami współrzędne wektorów $\mathbf{v}_1, ..., \mathbf{v}_l$ (i jedynki w ostatnim równaniu), zaś wyrazami wolnymi są współrzędne wektora $\mathbf{u}$ (i jedynka w ostatnim równaniu).
Nasz nieoceniony wniosek 7.2.3 zapewnia nam, że układ (7.1) jest niesprzeczny. Z twierdzenia 3.3.1 wynika więc, że ma nieujemne (!) rozwiązanie w którym zmiennych bazowych jest $n+1$ (tyle ile równań). Pamiętamy także, że zmienne niebazowe w rozwiązaniu bazowym przyjmują wartość zero. To kończy dowód twierdzenia 7.2.4. ■

7.3 Układy nierówności i równań liniowych

Niniejszy ustęp poświęcimy twierdzeniu Kuhna dotyczącemu układów nierówności i równań liniowych które udowodnimy metodami programowania liniowego. Już w następnym ustępie będziemy z tego twierdzenia korzystać w dowodzie

twierdzenia o rozdzielaniu (twierdzenie 7.4.1).
Układem równań i nierówności liniowych będziemy nazywać układ postaci

$$\begin{cases} \sum_{j=1}^{n} a_{ij}x_j & \leq & b_i & (i \in N) \\ \sum_{j=1}^{n} a_{ij}x_j & = & b_i & (i \in R) \end{cases} \tag{7.2}$$

gdzie $N \cup R = \{1, ..., m\}$.
Z tej postaci nierównościami zetknęliśmy się już w ustępie 4.4 z tą niewielką różnicą, że tym razem nie wyróżniamy specjalnie nierówności postaci $x_j \geq 0$, które mogą występować jako specjalne przypadki nierówności postaci $\sum_{j=1}^{n} a_{ij}x_j \leq b_i$.
Tak jak już to robiliśmy wielokrotnie, możemy i tym razem wymnożyć każdą z nierówności $\sum_{j=1}^{n} a_{ij}x_j \leq b_i$ przez *nieujemne* y_i otrzymując

$$(\textstyle\sum_{j=1}^{n} a_{ij}x_j)y_i \quad \leq \quad y_ib_i \quad (i \in N)$$

oraz każdą z równości $\sum_{j=1}^{n} a_{ij}x_j = b_i$ przez, tym razem *dowolne* y_i otrzymując

$$(\textstyle\sum_{j=1}^{n} a_{ij}x_j)y_i \quad = \quad y_ib_i \quad (i \in R)$$

by po zsumowaniu po $i = 1, ..., m$ a następnie zmienieniu kolejności sumowania po prawej stronie nierówności otrzymać

$$\textstyle\sum_{j=1}^{n}(\sum_{i=1}^{m} a_{ij}y_i)x_j \quad \leq \quad \sum_{i=1}^{m} y_ib_i \tag{7.3}$$

Zwróćmy uwagę, że w ten prosty sposób otrzymaliśmy pewne kryterium niesprzeczności układu (7.2). Nim kryterium to sformułujemy w postaci ogólnej, przyjrzymy się jak funkcjonuje ono na przykładzie.

Przykład 7.3.1 Niech będzie dany następujący układ nierówności i równań liniowych:

$$\begin{cases} 2x_1 & - & 7x_2 & + & x_3 & & & \leq & 1 \\ -x_1 & + & 3x_2 & + & 2x_3 & + & x_4 & \leq & 2 \\ x_1 & + & x_2 & - & x_3 & & & \leq & 0 \\ 3x_1 & + & 2x_2 & + & 2x_3 & + & 2x_4 & = & 6 \end{cases} \tag{7.4}$$

Jeśli w nierówności (7.3) podstawimy $y_1 = 1,\ y_2 = 2,\ y_3 = 3$ i $y_4 = -1$, wtedy otrzymamy $0x_1 + 0x_2 + 0x_3 - 0x_4 \leq -1$ (a więc sprzeczność), przy czym tylko zmienna odpowiadająca równości (y_4) przyjmuje wartość ujemną ($y_4 = -1$). Stąd wniosek, że nasz układ (7.4) jest sprzeczny. ∎

Definicja 7.3.1 *Mówimy, że układ 7.2 jest* **niezgodny** *jeśli istnieją liczby* $y_1, ..., y_m$ *spełniające następujące warunki*

$$\begin{cases} y_i & \geq & 0 & dla & i \in N \\ \sum_{i=1}^{m} a_{ij}y_i & = & 0 & dla & j = 1, ..., n \\ \sum_{i=1}^{m} b_iy_i & < & 0 \end{cases} \tag{7.5}$$

Wobec tego co powiedziano wyżej jest zupełnie oczywiste, że jeśli układ jest niezgodny, to jest sprzeczny. Okazuje się, że zależność pomiędzy niezgodnością a sprzecznością układu jest znacznie głębsza i niebanalna. Prawdziwe jest bowiem następujące twierdzenie, udowodnione przez H.W. Kuhna w 1956 roku.

Twierdzenie 7.3.1 (Kuhn) *Układ nierówności i równań liniowych (7.2) jest sprzeczny wtedy i tylko wtedy, jeżeli jest niezgodny.*

Dowód. Po tym co już zostało powiedziane, wystarczy udowodnić, że jeżeli układ 7.2 jest sprzeczny, to jest także niezgodny.
Rozważmy następujący problem programowania liniowego

$$\left\{\begin{array}{rcll} \text{zminimalizuj} \ \sum_{i=1}^m x_{n+i} & & & \\ \text{przy ograniczeniach:} & & & \\ \sum_{j=1}^n a_{ij}x_j + w_i x_{n+i} & \leq & b_i & i \in N \\ \sum_{j=1}^n a_{ij}x_j + w_i x_{n+i} & = & b_i & i \in R \\ x_{n+i} & \geq & 0 & (i = 1, ..., m) \end{array}\right. \tag{7.6}$$

gdzie

$$w_{n+i} = \left\{\begin{array}{rcl} 1 & \text{gdy} & b_i \geq 0 \\ -1 & \text{gdy} & b_i < 0 \end{array}\right.$$

Jest oczywiste, że układ (7.2) jest niesprzeczny wtedy i tylko wtedy, gdy (7.6) posiada rozwiązanie optymalne które ma wartość zero (co może mieć miejsce jedynie jeśli $x_{n+1} = ... = x_{n+m} = 0$). Nasz problem (7.6) jest oczywiście niesprzeczny ($x_i = 0,\ x_{n+i} =\mid b_i \mid$ dla $i = 1, ..., m$ są rozwiązaniem) i ograniczony (nie ma rozwiązania o wartości mniejszej niż zero). Rozwiązanie optymalne istnieje więc na mocy twierdzenia 3.3.1. Zauważmy, że minimalizacja funkcji $\sum_{i=1}^m x_{n+i}$ jest równoważna maksymalizacji $\sum_{i=1}^m(-x_{n+i})$. Tak więc problemem dualnym do (7.6) jest

$$\left\{\begin{array}{rcll} \text{zminimalizuj} \ \sum_{i=1}^m b_i y_i & & & \\ \text{przy ograniczeniach} \ \sum_{i=1}^m a_{ij} y_i & = & 0 & j = 1, ..., n \\ w_i y_i & \geq & -1 & i = 1, ..., m \\ y_i & \geq & 0 & i \in N \end{array}\right. \tag{7.7}$$

Jeżeli układ nierówności i rownań liniowych (7.2) jest sprzeczny, to optymalna wartość funkcji celu $\sum_{i=1}^m(-x_{n+i})$ jest ściśle ujemna, a stąd na mocy ogólnej zasady dualności (twierdzenie 4.4.2)

$$\sum_{i=1}^m b_i y_i \ < \ 0$$

Ta ostatnia nierówność łącznie z ograniczeniami problemu (7.7) oznacza, że system (7.2) jest niezgodny. ■

7.4 Wielościany i półprzestrzenie

Definicja 7.4.1 Niech $a_1, ..., a_n$ będą liczbami rzeczywistymi nie równymi równocześnie zero (inaczej: $a_1^2 + ... + a_n^2 > 0$). Zbiór wektorów $x \in \mathbf{R}^n$ o współrzędnych $x_1, ..., x_n$ spełniających $a_1x_1 + ... + a_nx_n \geq c$ (gdzie c jest pewną liczbą rzeczywistą nazywamy **półprzestrzenią**. **Wielościanem** w $\mathbf{R}^n$ nazywamy przecięcie skończonej liczby półprzestrzeni.

Czytelnik zechce samodzielnie sprawdzić, że każdy wielościan (w sensie powyższej definicji) jest zbiorem wypukłym (por ćwiczenie 7.6.8).

Celem niniejszego podrozdziału jest pokazanie jak z twierdzenia 7.3.1 wynika podane niżej twierdzenie o rozdzielaniu wielościanów.

Twierdzenie 7.4.1 (O rozdzielaniu wielościanów) *Dla dowolnych wielościanów W_1 i W_2 rozłącznych i niepustych, istnieją rozłączne półprzestrzenie P_1 i P_2 takie, że $W_1 \subset P_1$ i $W_2 \subset P_2$.*

Dowód. Przypuśćmy, że wielościany W_1 i W_2 dane są w następujący sposób:

$$W_1 = \{\mathbf{x} \in \mathbf{R}^n : \mathbf{A}_1\mathbf{x} \leq \mathbf{b}_1\}$$

$$W_2 = \{\mathbf{x} \in \mathbf{R}^n : \mathbf{A}_2\mathbf{x} \leq \mathbf{b}_2\}$$

Skoro $W_1 \cap W_2 = \emptyset$, to układ nierówności:

$$\begin{cases} \mathbf{A}_1\mathbf{x} \leq \mathbf{b}_1 \\ \mathbf{A}_2\mathbf{x} \leq \mathbf{b}_2 \end{cases}$$

jest sprzeczny, a więc, na mocy twierdzenia 7.3.1 istnieją wektory $\mathbf{y}_1 \in \mathbf{R}^{m_1}, \mathbf{y}_2 \in \mathbf{R}^{m_2}$ o współrzędnych nieujemnych (gdzie m_i jest liczbą wierszy macierzy $\mathbf{A}_i$, $i = 1, 2$) takie, że

$$\mathbf{y}_1\mathbf{A}_1 + \mathbf{y}_2\mathbf{A}_2 = 0 \tag{7.8}$$

$$\mathbf{y}_1\mathbf{b}_1 + \mathbf{y}_2\mathbf{b}_2 < 0 \tag{7.9}$$

Z drugiego z powyższych związków wynika, że $\mathbf{y}_1\mathbf{b}_1 < 0$ lub $\mathbf{y}_2\mathbf{b}_2 < 0$. Bez straty ogólności możemy założyć, że $\mathbf{y}_1\mathbf{b}_1 < 0$. Stąd zaś wynika, że $\mathbf{y}_1\mathbf{A}_1 \neq \mathbf{0}$, w przeciwnym bowiem przypadku układ $\mathbf{A}_1\mathbf{x} \leq \mathbf{b}_1$ byłby niezgodny, czyli wielościan P_1 pusty, a to jest sprzeczne z założeniami. Zbiory $\{\mathbf{x} : \mathbf{y}_1\mathbf{A}_1\mathbf{x} \leq c\}$ oraz $\{\mathbf{x} : \mathbf{y}_1\mathbf{A}_1\mathbf{x} \geq c\}$ są więc niepuste i różne od całej przestrzeni dla dowolnego $c \in \mathbf{R}$, czyli są półprzestrzeniami. W szczególności półprzestrzeniami są

$$H_1 = \{\mathbf{x} : \mathbf{y}_1\mathbf{A}_1\mathbf{x} \leq \mathbf{y}_1\mathbf{b}_1\}$$

i

$$H_2 = \{\mathbf{x} : \mathbf{y}_1\mathbf{A}_1\mathbf{x} \geq -\mathbf{y}_2\mathbf{b}_2\}$$

Z (7.9) wynika, że $H_1 \cap H_2 = \emptyset$. Ponieważ współrzędne wektora y_1 są nieujemne, dla każdego wektora $x \in W_1$, czyli takiego, że $\mathbf{A}_1\mathbf{x} \leq \mathbf{b}_1$, spełniona jest nierówność $\mathbf{y}_1\mathbf{A}_1\mathbf{x} \leq \mathbf{y}_1\mathbf{b}_1$, a więc $\mathbf{x} \in H_1$, Stąd oczywiście $W_1 \subset H_1$. Wzór (7.8) implikuje z kolei równość

$$\mathbf{y}_1\mathbf{A}_1 = -\mathbf{y}_2\mathbf{A}_2$$

Każdy $\mathbf{x}$ należący do W_2, czyli taki, że $\mathbf{A}_2\mathbf{x} \leq \mathbf{b}_2$ spełnia więc $\mathbf{y}_1\mathbf{A}_1\mathbf{x} = -\mathbf{y}_2\mathbf{A}_2\mathbf{x}$. Z drugiej strony $\mathbf{A}_2 x \leq \mathbf{b}_2$, a ponieważ współrzędne wektora $\mathbf{y}_2$ są nieujemne, zachodzi $\mathbf{y}_2\mathbf{A}_2 x \leq \mathbf{y}_2\mathbf{b}$ i w konsekwencji $-\mathbf{y}_2\mathbf{b} \leq -\mathbf{y}_2\mathbf{A}_2\mathbf{x} = \mathbf{y}_1\mathbf{A}_1\mathbf{x}$ Wykazaliśmy więc, że $W_2 \subset H_2$, co kończy dowód twierdzenia. ■

7.5 Metoda Fouriera–Motzkina redukcji układów nierówności

Niech będzie dany układ nierówności

$$\sum_{j=1}^{n} a_{ij}x_j \leq b_i \quad i = 1, ..., m \tag{7.10}$$

Metoda Fouriera–Motzkina polega na stopniowej redukcji zmiennych tak, by w końcu otrzymać jedną zmienną niezależną, a więc układ trywialny. Powiedzmy od razu, że to zmniejszanie liczby zmiennych będzie sie odbywało kosztem bardzo szybkiego wzrostu liczby nierówności.
bez straty ogólności można przyjąć, że w układzie (7.10)

- pierwszych m_1 zmiennych ma przy x_1 współczynnik dodatni,
- następne m_2 nierówności mają przy x_1 współczynnik ujemny, zaś
- w pozostałych $m - (m_1 + m_2)$ nierównościach współczynnik przy x_1 jest równy zero.

Dzieląc każdą z $m_1{+}m_2$ nierówności przez $\mid a_{i1} \mid$ $(i = 1, ..., m_1{+}m_2)$ otrzymamy z (7.10) równoważny mu układ postaci

$$\left\{ \begin{array}{rlcll} x_1+ & \sum_{j=2}^{n} a'_{ij}x_j & \leq & b'_i & i = 1, ..., m_1 \\ -x_1+ & \sum_{j=2}^{n} a'_{ij}x_j & \leq & b'_i & i = m_1 + 1, ..., m_1 + m_2 \\ & \sum_{j=2}^{n} a'_{ij}x_j & \leq & b'_i & i = m_1 + m_2 + 1, ..., m \end{array} \right. \tag{7.11}$$

czyli

$$\left\{ \begin{array}{rlcll} x_1+ & \sum_{j=2}^{n} a''_{ij}x_j & \leq & b''_i & i = 1, ..., m_1 \\ x_1+ & \sum_{j=2}^{n} a''_{ij}x_j & \geq & b''_i & i = m_1 + 1, ..., m_1 + m_2 \\ & \sum_{j=2}^{n} a''_{ij}x_j & \leq & b''_i & i = m_1 + m_2 + 1, ..., m \end{array} \right. \tag{7.12}$$

gdzie $a''_{ij} = a'_{ij}$ i $b''_i = b'_i$ dla $i = 1, ..., m_1$ i $i = m_1{+}m_2{+}1, ..., m$ oraz $a''_{ij} = -a'_{ij}$ i $b''_i = -b'_i$ dla $i = m_1 + 1, ..., m_1 + m_2$.

Układ (7.12) jest z kolei równoważny

$$\left\{\begin{array}{c} b''_{i_1} - \sum_{j=2}^n a''_{i_1 j} x_j \leq x_1 \leq b''_{i_2} - \sum_{j=2}^n a''_{i_2 j} x_j, \\ (\text{ dla } i_1 = m_1+1,...,m_2;\ i_2 = 1,...,m_1) \\ \sum_{j=2}^n a''_{ij} x_j \leq b''_j, \quad i = m_1+m_2+1,...,m \end{array}\right. \tag{7.13}$$

Zredukowane zadanie będzie następującym układem nierówności:

$$\left\{\begin{array}{ccl} \sum_{j=2}^n (a''_{i_2 j} - a''_{i_1 j}) x_j & \leq b''_{i_2} - b''_{i_1} & i_1 = m_1+1,...,m_2, \\ & & i_2 = 1,...,m_1 \\ \sum_{j=2}^n a''_{ij} x_j & \leq b''_i & i = m_1+m_2+1,...,m \end{array}\right. \tag{7.14}$$

Sprzeczność układu (7.14) oznacza sprzeczność układu (7.10). Natomiast jeśli układ (7.14) jest niesprzeczny to każde rozwiązanie $x_2,...,x_n$ można rozszerzyć do rozwiązania $x_1, x_2, ..., x_n$ dołączając dowolne x_1 spełniające (7.13).
Zobaczymy jak opisana metoda funkcjonuje na przykładzie.

Przykład 7.5.1 Niech będzie danu układ równań

$$\left\{\begin{array}{rrrcr} 2x_1+ & 4x_2- & 6x_3 & \leq & 8 \\ x_1- & 2x_2+ & x_3 & \leq & 5 \\ -x_1+ & x_2+ & x_3 & \leq & 3 \\ & 3x_2- & 4x_3 & \leq & 12 \end{array}\right.$$

Zastosujmy do tego układu kolejne etapy metody Fouriera–Motzkina. Po pierwsze, doprowadzamy do sytuacji w której współczynniki przy x_1 są równe 1, -1 lub 0

$$\left\{\begin{array}{rrrcr} x_1+ & 2x_2- & 3x_3 & \leq & 4 \\ x_1- & 2x_2+ & x_3 & \leq & 5 \\ -x_1+ & x_2+ & x_3 & \leq & 3 \\ & 3x_2- & 4x_3 & \leq & 12 \end{array}\right.$$

Stąd

$$\left\{\begin{array}{c} -3 + x_2 + x_3 \leq x_1 \leq 4 - 2x_2 + x_3 \\ -3 + x_2 + x_3 \leq x_1 \leq 5 + 2x_2 - x_3 \\ 3x_2 - 4x_3 \leq 12 \end{array}\right. \tag{7.15}$$

i zredukowany układ nierówności

$$\left\{\begin{array}{c} -3 + x_2 + x_3 \leq 4 - 2x_2 + 3x_3 \\ -3 + x_2 + x_3 \leq 5 + 2x_2 - x_3 \\ 3x_2 - 4x_3 \leq 12 \end{array}\right.$$

który po zredukowaniu przyjmie postać

$$\left\{\begin{array}{c} 3x_2 - 2x_3 \leq 7 \\ -x_2 + 2x_3 \leq 8 \\ 3x_2 - 4x_3 \leq 12 \end{array}\right. \quad \text{a więc} \quad \left\{\begin{array}{c} x_2 - \frac{2}{3}x_3 \leq \frac{7}{3} \\ -x_2 + 2x_3 \leq 8 \\ x_2 - \frac{4}{3}x_3 \leq 4 \end{array}\right.$$

Teraz redukujemy zmienną x_2:

$$\begin{cases} -8 + 2x_3 \le x_2 \le \frac{7}{3} + \frac{2}{3}x_3 \\ -8 + 2x_3 \le x_2 \le 4 + \frac{4}{3}x_3 \end{cases} \tag{7.16}$$

$$\begin{cases} \frac{4}{3}x_3 \le \frac{31}{3} \\ \frac{2}{3}x_3 \le 12 \end{cases}$$

i ostatecznie $x_3 \le \frac{31}{4}$. Ten wynik zaś oznacza, że dla każdego $x_3 \le \frac{31}{4}$ możemy dzieki (7.16) i (7.15) znaleźć odpowiednie wartości x_2 i x_3.

Zauważmy, że przy układzie co najwyżej trzech nierówności liczba nierówności przy kolejnych redukcjach nie zwiększa się (tak właśnie było w naszym przykładzie). Rozwiązując samodzielnie następne ćwiczenia czytelnik zauważy, że liczba nierówności na ogół rośnie bardzo szybko w miarę kolejnych redukcji. Więcej o redukcji układów nierówności w [20].

Metoda eliminacji Fouriera–Motzkina daje nam inny algorytm rozwiązywania PPL. Rzeczywiście, niech będzie dany PPL

$$\begin{cases} \mathbf{Ax} \le \mathbf{b} \\ \hline \\ \mathbf{cx} \to \max \end{cases} \tag{7.17}$$

W celu znalezienia rozwiązania optymalnego dla (7.17) zastosujmy metodę redukcji do układu

$$\begin{cases} \mathbf{Ax} \le \mathbf{b} \\ \lambda - \mathbf{cx} \le 0 \end{cases} \tag{7.18}$$

gdzie λ jest nową zmienną. Redukować będziemy kolejne współrzędne wektora $\mathbf{x}$ tak, że w końcu pozostanie nam tylko zmienna λ. Teraz wystarczy przyjąć λ największą możliwą, by otrzymać rozwiązanie optymalne problemu (7.17).
Zauważmy na koniec, że metodę Fouriera–Motzkina można stosować także do redukowania układów w których występują także silne nierówności.

7.6 ćwiczenia

Ćwiczenie 7.6.1 Niech będzie dany PPL

$$\left\{ \begin{array}{rcr} -x_1 + x_2 & \le & 3 \\ -x_1 + 2x_2 & \le & 7 \\ x_1 + 2x_2 & \le & 13 \\ 2x_1 + x_2 & \le & 14 \\ -4x_1 + x_2 & \ge & -19 \\ -x_1 + 3x_2 & \ge & -2 \\ x_1, x_2 & \ge & 0 \\ \hline z = 4x_1 + 5x_2 & \to & \max \end{array} \right.$$

(a) Napisz postać standardową i kanoniczną tego problemu.

(b) Podaj jego interpretację geometryczną (wykonaj rysunek).

Nie stosując algorytmu sympleks:

(c) Znajdź rozwiązanie optymalne. Czy istnieje tylko jedno optymalne rozwiązanie tego problemu?

(d) Zakładając, że wybór zmiennej wchodzący dokonowany jest tak jak zazwyczaj (tj. zmienną wchodzącą jest ta której dodatni współczynnik w funkcji celu jest największy). podaj rozwiązania bazowe każdej iteracji.

(e) To samo pytanie przy wyborze zmiennej wchodzącej regułą Blanda.

W (d) i (e) zakładamy, że pierwszym rozwiązaniem bazowym jest $x_1 = x_2 = z = 0$ (w rozwiązaniach bazowych należy podać wartości x_1, x_2 i z).

Ćwiczenie 7.6.2 Znajdź powłokę wypukłą zbiorów:

(a) $X = \{(x, y) \in \mathbf{R}^2 : x^2 + y^2 \leq 1, 0 \leq y, \mid x \mid\}$

(b) $X = \{(1, 1), (1, 0), (0, 1)\}$

(c) Podaj przykład zbioru $X \subset \mathbf{R}^n$ takiego, że jego powłoka wypukła jest wypukłą kombinacją mniej niż $n+1$ elementów zbioru X, oraz taki przykład by we wzorze

$$conv(X) = \{\alpha \mathbf{u}_1 + ... + \alpha_{n+1} : \alpha_i \geq 0 \text{ dla } i = 1, ..., n; \sum_{i=1}^{n} \alpha_i = 1\}$$

$n+1$ nie dała się zastąpić przez nic mniejszego.

Ćwiczenie 7.6.3 Zbadaj niesprzeczność poniższych układów.

1.
$$\begin{cases} x_1 - 3x_2 + x_3 - x_4 \leq 1 \\ -2x_1 + x_2 - x_3 + x_4 \leq 2 \\ x_1 - x_2 - 2x_3 + x_4 \leq -3 \\ 2x_1 + 3x_2 + x_3 - x_4 = 1 \end{cases}$$

2.
$$\begin{cases} x_1 + x_2 - x_3 = 3 \\ -x_1 + x_2 - 2x_3 \leq -1 \\ 3x_1 + x_2 \leq 2 \end{cases}$$

3.
$$\begin{cases} 3x_1 - 2x_2 + 4x_3 \leq 1 \\ 2x_1 + x_2 - 3x_3 \leq 3 \\ x_1 - 2x_2 + x_3 \geq 3 \\ x_1 + 7x_2 - 3x_3 \geq 2 \end{cases}$$

Ćwiczenie 7.6.4 Metodą Fouriera–Motzkina zredukuj układy nierówności z ćwiczeń 7.6.1 i 7.6.3 niniejszego rozdziału.

Ćwiczenie 7.6.5 Wykorzystując metodę Fouriera-Motzkina znajdź maksymalną wartość x_2^* zmiennej x_2 dla której spełniony jest układ nierówności

$$\begin{cases} x_1 - x_2 + 3x_3 & \leq & 4 \\ -x_1 - x_2 + 2x_3 & \leq & 6 \\ x_1 + x_2 + 5x_3 & \geq & 1 \\ x_1 + 3x_2 - x_3 & \leq & 3 \end{cases}$$

Ćwiczenie 7.6.6 Korzystając z metody Fouriera–Motzkina znajdź optymalne rozwiązania problemów:

$$\left\{ \begin{array}{ccc} 2x_1 + 3x_2 + x_3 & \leq & 3 \\ x_1 + x_2 - x_3 & \leq & 2 \\ x_1 + 2x_2 + x_3 & \leq & 3 \\ x_1 + 2x_2 - x_3 & \geq & -4 \\ x_1, x_2, x_3 & \geq & 0 \\ \hline \\ 2x_1 + x_2 + 4x_3 & \to & \max \end{array} \right.$$

$$\left\{ \begin{array}{ccc} x_1 - x_2 & \leq & 3 \\ x_1 + x_2 & \leq & 6 \\ 2x_1 - x_2 & \leq & -4 \\ x_1, x_2 & \geq & 0 \\ \hline \\ 2x_1 + x_2 & \to & \max \end{array} \right.$$

Ćwiczenie 7.6.7 Wykaż, że dla każdego zbioru skończonego $X \subset \mathbf{R}^n$ powłoką wypukłą jest przedział domknięty.

Ćwiczenie 7.6.8 Udowodnij, że każdy wielościan jest zbiorem wypukłym.

Ćwiczenie 7.6.9 Wskaż przykłady dowodzące, że w twierdzeniu 7.4.1 wielościanów nie da się zastąpić zbiorami wypukłymi.

Ćwiczenie 7.6.10 Niech będą dane dwa wielościany w $\mathbf{R}^4$:

$$W_1 : \begin{cases} x_1 & +2x_2 & -x_3 & +x_4 & \leq & 2 \\ -x_1 & -x_2 & +2x_3 & -x_4 & \leq & 3 \\ 3x_1 & +x_2 & -2x_3 & +x_4 & \geq & 4 \end{cases}$$

$$W_2 : \begin{cases} 2x_1 & -x_2 & +3x_3 & -x_4 & \leq & 2 \\ 3x_1 & -2x_2 & +x_3 & +2x_4 & \leq & 1 \\ 5x_1 & -8x_2 & +15x_3 & +2x_4 & \geq & 4 \end{cases}$$

(a) Wykaż, że $W_1 \neq \emptyset, W_2 \neq \emptyset$ i $W_1 \cap W_2 = \emptyset$.

(b) Posługując się metodą dowodu twierdzenia 7.4.1 wskaż rozłączne półprzestrzenie P_1 i P_2 w $\mathbf{R}^4$ takie, że $W_1 \subset P_1$ i $W_2 \subset P_2$.

Rozdział 8

Metody sieciowe

Problemy które będziemy rozwiązywali przy pomocy metod sieciowych są w rzeczywistości specjalnymi problemami programowania liniowego. Powstać może więc pytanie: czy przypadkiem nie wyważamy otwartych drzwi? Po co szukać specjalnych metod do rozwiązywania specjalnych przypadków, skoro umiemy rozwiązywać przypadek ogólny?
Jest kilka ważnych powodów zajmowania się metodami sieciowymi. Wspomnijmy kilka z nich.

1. Sieci i grafy są wyjątkowo *wdzięcznym* i działającym na wyobraźnię narzędziem pozwalającym łatwo modelować wiele zagadnień. Jeśli mamy obliczyć sprawność – czyli przepustowość – sieci komunikacyjnej, wodociągowej czy kanalizacyjnej (por. [16]), to wygodniej nam i łatwiej w języku i modelu sieci pozostać niż przechodzić na język równań i nierówności.

2. Algorytmy sieciowe (grafowe) dotyczące omawianych zagadnień są szybsze (o co nietrudno, skoro sympleks – przynajmniej teoretycznie jest, jak widzieliśmy w ustępie 3.5, beznadziejny[1]).

3. Wiele bardzo ważnych twierdzeń teorii grafów można łatwo udowodnić korzystając ze słynnego twierdzenia o maksymalnym przepływie i minimalnym przekroju (twierdzenie 8.3.4). Niektóre z tych twierdzeń poznamy w podrozdziale 8.7.

Pierwszy z powyższych argumentów ma charakter estetyczny i poznawczy. Zobaczymy, że wspomniane już twierdzenie o maksymalnym przepływie i minimalnym przekroju jest, w pewien sposób, konsekwencją zasady dualności poznanej w rozdziale 4 (twierdzenie 4.1.4).
Drugi argument ma charakter praktyczny, osłabiony przez wspomniane dobre spisywanie się sympleksu w praktyce.

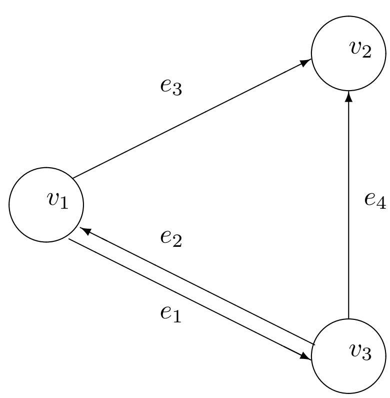
[1] Jeszcze raz podkreślmy, co znaczy to *teoretycznie*: sympleks jest algorytmem *niewielomianowym* jak wykazuje przykład Klee–Minty'ego, w praktycznych zaś przykładach spisuje się bardzo dobrze.

8.1 Grafy i sieci

Grafem zorientowanym nazywać będziemy parę $G = (V, A)$, gdzie V jest dowolnym skończonym zbiorem niepustym, nazywanym zbiorem *wierzchołków*, zaś A pewnym podzbiorem zbioru dwuelementowych par $(x, y) \in V \times V$ takich, że $x \neq y$. x jest wtedy *początkiem* zaś y końcem łuku $e = (x, y)$. x i y nazywamy [2] *końcami* łuku (x, y). Jeśli $(x, y) \in A$ to mówimy, że wierzchołki x i y są *połączone* w G. łuk (x, y) jest wtedy *incydentny* z wierzchołkami x i y. *Rzędem*
arzďż˝d grafu grafu G nazywamy liczbę $|V|$, zaś *rozmiarem* G liczbę $|A|$. Elementy zbioru A nazywamy *łukami.*
Dla wierzchołka $x \in V$ przez $N_G^+(x)$ nazywamy zbiór tych wierzchołków y dla których $(x, y) \in A$, zaś $N_G^-(x) = \{y : (y, x) \in A\}$.

Graf zorientowany (będziemy, dla wygody, mówili czasami po prostu *graf*) ma bardzo prostą interpretację graficzną: elementy zbioru V (a więc wierzchołki grafu) będziemy przedstawiali jako punkty płaszczyzny, zaś łuki jako zakończone strzałkami linie łaczące odpowiednie wierzchołki.

Przykład 8.1.1 Graf $G = (\{v_1, v_2, v_3\}, \{e_1 = (v_1, v_3), e_2 = (v_3, v_1), e_3 = (v_1, v_2), e_4 = (v_3, v_2)\})$ interpretujemy tak, jak to zostało przedstawione na rysunku 8.1.



Rys. 8.1

Z każdym grafem zorientowanym można skojarzyć wzajemnie jednoznacznie dwie macierze: macierz sąsiedztw i macierz incydencji.

[2] Niekonsekwentnie...

8.1.1 Macierz sąsiedztw grafu zorientowanego

Niech $G = (V, A)$ będzie pewnym grafem zorientowanym o n wierzchołkach $\{v_1, ..., v_n\}$. Wtedy macierz $T \in \mathbf{R}^{n\times n} = (a_{ij})_{i,j=1,...,n}$ zdefiniowaną wzorami

$$s_{ij} = \begin{cases} 1 & \text{jeśli } (v_i, v_j) \in A \\ 0 & \text{jeśli } (v_i, v_j) \notin A \end{cases}$$

Przykład 8.1.1 (c.d.) Dla grafu z rysunku 8.1 macierz sąsiedztw jest postaci

$$T = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix}$$

Zauważmy, że skoro założyliśmy, że dla dowolnego łuku (x, y) zachodzi $x \neq y$, macierz sąsiedztw ma same zera na przekątnej głównej.

8.1.2 Macierz incydencji

Macierzą incydencji grafu zorientowanego $G = (V, A)$ o zbiorze wierzchołków $V = \{v_1, ..., v_n\}$ i zbiorze łuków $A = \{e_1, ..., e_m\}$, nazywamy macierz $N = (b_{ij})_{\substack{i=1,...,n \\ j=1,...,m}}$ taką, że

$$b_{ij} = \begin{cases} 1 & \text{jeśli } v_i \text{ jest początkiem łuku } e_j \\ -1 & \text{jeśli } v_i \text{ jest końcem łuku } e_j \\ 0 & \text{jeśli } v_i \text{ i } e_j \text{ nie są incydentne} \end{cases}$$

Przykład 8.1.1 Dla grafu z rysunku 8.1 macierzą incydencji jest

$$N = \begin{bmatrix} 1 & -1 & 1 & 0 \\ 0 & 0 & -1 & -1 \\ -1 & 1 & 0 & 1 \end{bmatrix}$$

Wobec tego co wiemy z podrozdziału 5.3 o macierzach totalnie unimodularnych (wniosek 5.3.2) ważne będą konsekwencje poniższego twierdzenia.

Twierdzenie 8.1.1 *Macierz incydencji grafu zorientowanego jest totalnie unimodularna.*

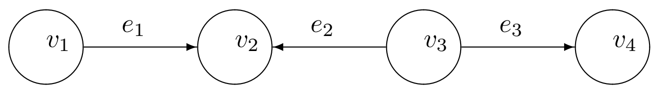
Dowód. Wykorzystamy twierdzenie 5.3.3. Zauważmy, że w macierzy incydencji dowolnego grafu zorientowanego każdej kolumnie odpowiada pewien łuk grafu i są w niej dwa wyrazy różne od zera: -1 i 1 w wierszach odpowiadających początkowi i końcowi łuku. Wystarczy więc przyjąć jako zbiory W_1 i W_2 występujące w założeniach twierdzenia 5.3.3 zbiór wszystkich wierszy macierzy incydencji grafu i zbiór pusty. ■

8.1.3 ścieżki i cykle

ścieżką o długości k w grafie $G = (V, A)$ nazywamy ciąg wierzchołków i łuków

$$P = (v_1, e_1, v_2, e_2, ..., v_i, e_i, v_{i+1}, ..., e_k, v_{k+1}) \tag{8.1}$$

takich, że $v_i \in V$ dla $i = 1, ..., k+1$, e_i jest łukiem o końcach v_i, v_{i+1} dla $i = 1, ..., k$, oraz $v_i \neq v_j$ dla $i \neq j$.



Rys. 8.2

O łuku $e_i = (v_i, v_{i+1})$ mówimy, że jest *zgodny* ze ścieżką P, zaś jeśli $e_i = (v_{i+1}, v_i)$ mówimy, że e_i jest *niezgodny* z P.
Na rysunku 8.2 łuki e_1 i e_3 są zgodne, zaś łuk e_2 niezgodny ze ścieżką $P = (v_1, e_1, v_2, e_2, v_3, e_3, v_4)$, o łukach $e_1 = (v_1, v_2), e_2 = (v_3, v_2), e_3 = (v_3, v_4)$.

Jeśli, dodatkowo, założymy, że $v_1 = v_n$ (przy zachowaniu $v_i \neq v_j$ dla wszystkich pozostałych przypadków), to ścieżkę (8.1) nazywamy *cyklem*.

8.2 Sieci

Graf $G = (V, A)$ nazywali będziemy *siecią* S, jeżeli zdefiniowana jest pewna funkcja

$$c : A \to \mathbf{R}$$

Pisali będziemy wtedy $S = (V, A, c)$.

W zależności od sytuacji, funkcję c będziemy nazywali przepustowością łuków[3] lub długością łuków[4], czy wreszcie kosztem łuków[5]. Z interpretacji sieci wynika, że na ogół nie będzie kłopotliwym założenie nieujemności funkcji c.

8.3 Przepływy w sieciach

W sieci S może być wyróżniony pewien wierzchołek s, nazywany *źródłem* i inny wierzchołek, powiedzmy t, który nazywać będziemy *odpływem*.

[3] Na przykład gdy będziemy myśleli o sieciach wodociągowych, kanalizacyjnych – ale także, czasami, drogowych.
[4] Na przykład dla sieci komunikacyjnej.
[5] W sieci komunikacyjnej: gdy myślimy raczej o cenie biletu niż odległości.

Niech $S = (V, A, c)$ będzie siecią o źródla s i odpływie t oraz nieujemnej funkcji przepustowości c. O nieujemnej funkcji $f : A \to \mathbf{R}$ mówimy, że jest *przepływem* z s do t w S jeżeli spełnia następujące warunki:

$$f(x,y) \leq c(x,y) \quad \text{dla każdego } (x,y) \in A \tag{8.2}$$

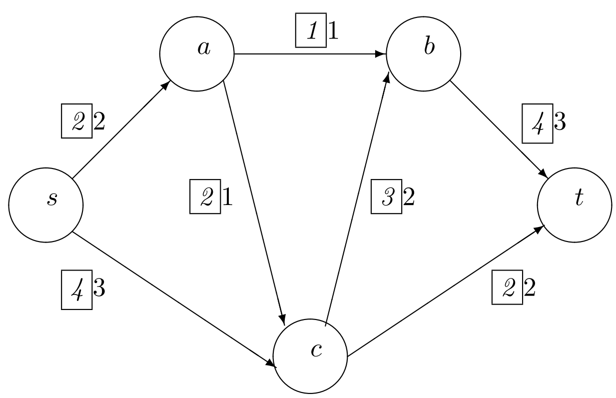
$$\sum_{y \in N^+(x)} f(x,y) - \sum_{z \in N^-(x)} f(z,x) = 0 \quad \text{dla każdego } x \neq s,t \tag{8.3}$$

Liczbę

$$w(f) = \sum_{y \in N^+(s)} f(s,y) - \sum_{z \in N^-(s)} f(z,s)$$

nazywamy *wartością przepływu* f.

Przykład 8.3.1 Niech S będzie siecią jak na rysunku 8.3. Liczby wypisane obok łuków w kwadracikach oznaczają przepustowości łuków, a te bez kwadracików przepływy. Jest bardzo łatwym sprawdzenie, że na rysunku poprawnie zdefiniowany jest przepływ (spełnia warunki (8.2) i (8.3)), a wartość tego przepływu wynosi 5. □



Rys. 8.3

Problem optymalizacyjny narzuca się teraz sam.

Problem maksymalnego przepływu.
Dla danej sieci $S = (V, A; c)$ o źródle s i odpływie t znaleźć przepływ o maksymalnej wartości.

W dalszym ciągu będzie nam wygodnie będzie nam przyjąć zamiast $(x,y) \notin A$, $c(x,y) = 0$ i istnienie wszystkich możliwych łuków w sieci. Interpretacja takiej umowy jest oczywista: nie ma przepływu (a raczej jest przepływ zerowy) na nie istniejących łukach[6]. Wtedy warunki (8.2) i (8.3) można zapisać, odpowiednio,

$$f(x,y) \le c(x,y) \quad \text{dla wszystkich } x,y \in V \tag{8.4}$$

$$\sum_{y\in V} f(x,y) - \sum_{z\in V} f(z,x) = 0 \quad \text{dla każdego } x \neq s,t \tag{8.5}$$

a wzór na wartość przepływu

$$w(f) = \sum_{y\in V} f(s,y) - \sum_{z\in V} f(z,s) \tag{8.6}$$

Zauważmy, że przepływ w przykładzie 8.3.1 nie jest maksymalny. Zwiększając przepływ na łukach: $(s,a),(a,c),(c,b)$ i (c,t) o 1 otrzymamy przepływ (warunki (8.2) i (8.3) są dalej spełnione) i wartość tego przepływu wynosi 6. Okaże się, że ten prosty pomysł znajdowania ścieżki prowadzącej z s do t *wzdłóż* której można zwiększyć przepływ na poszczególnych łukach o stałą wartość, jest kluczową ideą prowadzącą do rozwiązania problemu maksymalnego przepływu w sieciach.
Jest najzupełniej oczywiste, że problem maksymalnego przepływu jest problemem programowania liniowego. Wierzchołki sieci nazwiemy $v_0, v_1, ..., v_n$, z $v_0 = s, v_n = t$, przepływ na łuku (v_i, v_j) oznaczymy przez x_{ij} (zamiast $f(v_i, v_j)$ a przepustowość łuku (v_i, v_j) przez c_{ij}. Nasz problem maksymalnego przepływu jest PPL:

$$\left\{\begin{array}{l} \text{zmaksymalizuj} \quad z = \sum_{i:(v_0,v_i)\in A} x_{0i} - \sum_{j:(v_j,v_s)\in A} x_{j0} \\ \text{przy warunkach:} \\ \sum_{(v_{i_0},v_j)\in A} x_{i_0 j} - \sum_{(v_j,v_{i_0})\in A} x_{j i_0} = 0 \quad (i_0 = 1,...,n-1) \\ 0 \le x_{ij} \le c_{ij}, \quad i,j : (v_i,v_j) \in A \end{array}\right. \tag{8.7}$$

Zauważmy, że jeśli przepustowości łuków są liczbami skończonymi, to problem (8.7) jest ograniczony, posiada więc rozwiązanie optymalne. Z powyższego oraz twierdzenia 8.1.1 wynika następujący wniosek.

Wniosek 8.3.1 *Jeśli funkcja przepustowości łuków przyjmuje wartości całkowite, to problem maksymalnego przepływu w tej sieci ma całkowite rozwiązanie optymalne.* ■

Ważnym pojęciem dla problemu maksymalnego przepływu są przekrój i jego wartość.

Niech będzie dana sieć $S = (V, A, c)$ ze źródłem s i odpływem t i niech W będzie takim podzbiorem V, że $s \in W$ i $t \notin W$. Zbiór $V - W$ oznaczmy przeż $\bar{W}$. Zbiór łuków

$$(W, \bar{W}) = \{(x,y) : x \in W, y \in \bar{W}\}$$

[6]W praktyce oznacza to, że woda nie płynie jak nie ma rury.

nazywamy *przekrojem rozdzielającym s od t* lub krótko *przekrojem.* Liczbę

$$c(W,\bar{W}) = \sum_{(x,y)\in(W,\bar{W})} c(x,y)$$

nazywamy *przepustowością przekroju* $(W,\bar{W})$.

Twierdzenie 8.3.2 *Niech $S = (V, A, c)$ będzie siecią o źródle s i odpływie t. Dla dowolnego przepływu f i dla dowolnego przekroju $(W,\bar{W})$ zachodzi nierówność*

$$w(f) \leq c(W,\bar{W})$$

Dowód. Sumując odpowiednio wzory (8.6) i (8.5) po wszystkich $x \in W$ i pamiętając, że $s \in W, W \cap \bar{W} = \emptyset, W \cup \bar{W} = V$, otrzymamy

$$w(f) = \sum_{u\in V} f(s,u) - \sum_{v\in V} f(v,s) + \sum_{x\in W, x\neq s} \left(\sum_{y\in V} f(x,y) - \sum_{z\in V} f(z,x) \right) =$$

$$= \sum_{x\in W, y\in V} f(x,y) - \sum_{x\in W, z\in V} f(z,x) =$$

$$= \left(\sum_{x\in W, y\in W} f(x,y) + \sum_{x\in W, y\in\bar{W}} f(x,y) \right)$$

$$- \left(\sum_{x\in W, z\in W} f(z,x) + \sum_{x\in W, z\in\bar{W}} f(z,x) \right)$$

Zauważmy, że $\sum_{x\in W, y\in W} f(x,y) = \sum_{x\in W, z\in W} f(z,x)$, a wobec tego

$$w(f) = \sum_{x\in W, y\in\bar{W}} f(x,y) - \sum_{x\in W, z\in\bar{W}} f(z,x) \tag{8.8}$$

Wartości $f(z,x)$ są nieujemne, uwzględniając więc nierówności (8.4) otrzymujemy

$$w(f) \leq \sum_{x\in W, y\in\bar{W}} f(x,y) \leq \sum_{x\in W, y\in\bar{W}} c(x,y) = c(W,\bar{W}) \tag{8.9}$$

co kończy dowód. ■

Uwaga 8.1 *Jeżeli przepływ f i przekrój $(W,\bar{W})$ spełniają warunek*

$$w(f) = c(W,\bar{W})$$

to na mocy twierdzenia 8.3.2 f jest przepływem maksymalnym (a przekrój $(W,\bar{W})$ ma minimalną przepustowość).

W trakcie dowodu twierdzenia 8.3.2 otrzymaliśmy następującą ważną równość (8.8).

Uwaga 8.2 *Dla każdego przepływu f i przekroju $(W, \bar{W})$ zachodzi wzór*

$$w(f) = \sum_{x \in W, v \in \bar{W}} f(x, v) - \sum_{y \in W, w \in \bar{W}} f(w, y) \tag{8.10}$$

Wniosek 8.3.3 *Dla każdego przepływu f zachodzi wzór*

$$\sum_{x \in V} f(x, t) - \sum_{x \in V} f(t, x) = w(f)$$

Wniosek 8.3.3 oznacza, że *w przyrodzie nic nie ginie*, to co wypłynęło ze źródła wpływa do odpływu!

Dowód wniosku 8.3.3. Wystarczy zastosować wzór (8.10) do $\bar{W} = \{t\}$. ■

Twierdzenie 8.3.2 da nam możliwość udowodnienia *twierdzenia o maksymalnym przepływie i minimalnym przekroju*, odkrytego niezależnie przez Forda i Fulkersona [8] oraz Eliasa, Feinsteina i Shannona [7].

Twierdzenie 8.3.4 (Max flow min cut theorem) *Maksymalna wartość przepływu ze źródła s do odpływu t w sieci jest równa wartości minimalnego przekroju rozdzielającego s od t:*

$$max\{w(f) | f \text{ – przepływ z } s \text{ do } t\} = min\{c(W, \bar{W}) | s \in W, t \in \bar{W}\}$$

Dowód twierdzenia 8.3.4 polega *de facto* na wskazaniu algorytmu pozwalającego znaleźć maksymalny przepływ w sieci, tzw. *algorytm ścieżki powiększającej* Forda–Fulkersona (p. [9]). Formalnie algorytm ten zostanie przedstawiony później.

Niech f będzie przepływem maksymalnym w sieci $G = (V, A, c)$ ze źródła s do odpływu t. Utwórzmy rekurencjnie zbiór W w następujący sposób:

1. $s \in W$
2. jeśli $x \in W, (x, y) \in A, f(x, y) < c(x, y)$ to $y \in W$
3. jeśli $x \in W, (y, x) \in A$ i $f(y, x) > 0$ to $y \in W$.

Kluczową ideą dowodu jest obserwacja, że dla każdego wierzchołka $x \in W$ istnieje ścieżka

$$P = (x_0, e_1, x_1, ..., x_{k-1}, e_k, x_k)$$

taka, że $x_0 = s, x_k = x$ i dla każdego e_i zachodzi albo

$$e_i = (x_{i-1}, x_i) \text{ i } f(e_i) < c(e_i) \tag{8.11}$$

albo

$$e_i = (x_i, x_{i-1}) \text{ i } f(e_i) > 0 \tag{8.12}$$

Inaczej mówiąc: wartość przepływu jest mniejsza od przepustowości na wszystkich łukach zgodnych i dodatnia na niezgodnych łukach ścieżki P.
Gdyby t należało do W, to istniałaby ścieżka P z s do t (tj. taka, że $x_0 = s$ i $x_k = t$) spełniająca warunki (8.11)–(8.12). Niech

$$\varepsilon_1 = \min\{c(e_i) - f(e_i) | e_i - \text{ łuk zgodny } P\}$$

$$\varepsilon_2 = \min\{f(e_i) | e_i - \text{ łuk niezgodny } P\}$$

oraz

$$\varepsilon = \min\{\varepsilon_1, \varepsilon_2\}$$

Jest oczywistym, że $\varepsilon > 0$, i że dodając ε do funkcji przepływu na łukach ścieżki P zgodnych i odejmując ε na łukach ścieżki P niezgodnych, inaczej mówiąc: definiując funkcję

$$\tilde{f}(e) = \begin{cases} f(e) & \text{jeśli } e \text{ nie jest łukiem } P \\ f(e) + \varepsilon & \text{jeśli } e \text{ jest łukiem zgodnym } P \\ f(e) - \varepsilon & \text{jeśli } e \text{ jest łukiem niezgodnym } P \end{cases}$$

otrzymalibyśmy nową funkcję przepływu której wartość wynosiłaby $w(\tilde{f}) = w(f) + \varepsilon > w(f)$ co byłoby sprzeczne z maksymalnościa przepływu f.
Stąd wynika, że t nie jest elementem zbioru W i zbiór W ma następujące własności:

1. $s \in W$
2. $t \notin W$
3. jeśli $(x, y) \in A, x \in W, y \in \bar{W}$ to $f(x, t) = c(x, y)$
4. jeśli $(x, y) \in A, x \in \bar{W}, y \in W$ to $f(x, y) = 0$

Wobec (1) i (2) $(W, \bar{W})$ jest przekrojem rozdzielającym s i t. łuki spełniające (3) nazywają sie *nasyconymi* , a takie które spełniają (4) *wyschniętymi.* Wobec (3) i (4) mamy

$$c(W, \bar{W}) = \sum_{x \in W, y \in \bar{W}} c(x, y) = \sum_{x \in W, y \in \bar{W}} f(x, y)$$

$$= \sum_{x \in W, y \in \bar{W}} f(x, y) - \sum_{x \in \bar{W}, y \in W} f(x, y) = w(f)$$

co wobec twierdzenia 8.3.2 dowodzi, że $(W, \bar{W})$ jest przekrojem o przepustowości minimalnej. ■

8.4 Maksymalny przepływ a dualność

W tym podrozdziale okaże się, że *twierdzenie o maksymalnym przepływie i minimalnym przekroju* jest – w pewien sposób – zasadą dualności dla problemu maksymalnego przepływu.

Zapiszmy problem maksymalnego przepływu (8.4) – (8.6) w nieco innej postaci. Niech będzie dana sieć $S = (V, A; c)$. Przepływem w S jest funkcja $f : A \to \mathbf{R}$ spełniająca następujące warunki:

$$\begin{cases} \sum_{y:(x,y)\in A} f(x,y) - \sum_{z:(z,x)\in A} f(z,x) = \begin{cases} 0 & \text{dla } x \neq s,t \\ w & \text{dla } x = s \\ -w & \text{dla } x = t \end{cases} \\ \\ f(x,y) \leq c(x,y) \text{ dla } (x,y) \in A \\ f(x,y) \geq 0 \text{ dla } (x,y) \in A \end{cases} \tag{8.13}$$

Wartością przepływu f jest wtedy liczba w którą w naszym problemie maksymalizujemy:

$$w \to \max \tag{8.14}$$

Problemem dualnym do problemu (8.13) – (8.14) jest problem następujący[7] (por. podrozdział 4.4):

$$\begin{cases} \lambda(y) - \lambda(x) + \gamma(x,y) \geq 0 \text{ dla } (x,y) \in A \\ \gamma(x,y) \geq 0 \text{ dla } (x,y) \in A \\ \lambda(s) - \lambda(t) \geq 1 \\ \hline z = \sum_{(x,y)\in A} c(x,y)\gamma(x,y) \to \max \end{cases} \tag{8.15}$$

Zauważmy, że w problemie dualnym (8.15) nie ma ograniczeń na znak zmiennych dualnych $\lambda(v)$ (dla $v \in V$). Zmienne dualne odpowiadają poszczególnym ograniczeniom w problemie prymalnym. Tak więc

$\lambda(x)$ – są zmiennymi odpowiadającymi ograniczeniom

$$\sum_{(x,y)\in A} f(x,y) - \sum_{(z,x)\in A} f(z,x) = \begin{cases} 0 & \text{dla } x \neq s,t \\ -w & \text{dla } x = s \\ w & \text{dla } x = t \end{cases} \tag{8.16}$$

Każdemu wierzchołkowi x sieci S odpowiada zmienna dualna która nie ma ograniczenia na znak gdyż we wzorze (8.16) jest równość.

$\gamma(x,y)$ – są zmiennymi dualnymi odpowiadającymi ograniczeniu

$$f(x,y) \leq c(x,y) \tag{8.17}$$

stąd w problemie dualnym występuje ograniczenie na znak tej zmiennej[8].

[7] Fakt, że postać problemu dualnego wyraża wzór (8.15) łatwiej zrozumiemy przypominając sobie, że macierzą problemu prymalnego jest macierz incydencji grafu (plus wiersze odpowiadające ograniczeniom na przepustowości).

[8] Doradzam rozwiązanie ćwiczenia 8.9.1 właśnie teraz!

Dowolny przekrój $(W, \overline{W})$ rozdzielający s od t w sieci S definiuje pewne rozwiązanie dopuszczalne problemu (8.15) w następujący sposób:

$$\lambda(x) = \begin{cases} 1 & \text{jeśli } x \in W \\ 0 & \text{jeśli } x \notin W \end{cases} \tag{8.18}$$

$$\gamma(x,y) = \begin{cases} 1 & \text{jeśli } x \in W, y \in \overline{W} \\ 0 & \text{w przeciwnym przypadku} \end{cases} \tag{8.19}$$

Słaba zasada dualności w ogólnej postaci (twierdzenie 4.4.1) daje nam natychmiast inny dowód twierdzenia 8.3.2. Rzeczywiście, funkcje λ i γ zdefiniowane wzorami (8.18) i (8.19) przez dowolny przekrój rozdzielający s od t dają pewne rozwiązanie problemu (8.15).
Z kolei cała reszta dowodu twierdzenia 8.3.4 sprowadza się do wykazania, że pewien przekrój określa nam przy pomocy (8.18) i (8.19) rozwiązanie optymalne (8.13) (zrobili to Dantzig i Fulkerson w [6]).

8.5 Algorytm Forda–Fulkersona

Niech $S = (V, A; c)$ będzie siecią a f przepływem w S – jeśli nie mamy lepszego, może to być *przepływ zerowy*, tj. przepływ równy zero na wszystkich łukach sieci. Algorytm – wykorzystujący ideę dowodu twierdzenia 8.3.4 polega na cechowaniu wierzchołków sieci cechami postaci (x^-, ε) lub (x^+, ε) oraz modyfikacji przepływu jeśli f okazał się przepływem który nie jest maksymalny lub znalezieniu przekroju o przepustowości równej wartości przepływu f.
Cechowanie. Wierzchołkowi s nadajemy cechę $(-, \infty)$.
Niech x będzie wierzchołkiem ocechowanym. Z wierzchołka x cechujemy nieocechowane jeszcze wierzchołki w jednej z dwóch sytuacji:

(a) jeśli istnieje wierzchołek nieocechowany y taki, że $(x, y) \in A$ i $f(x, y) < c(x, y)$ to wierzchołkowi y nadajemy cechę $(x^+, \varepsilon(y))$, gdzie $\varepsilon(y) = \min\{\varepsilon(x), c(x, y) - f(x, y)\}$,

(b) jeśli istnieje nieocechowany wierzchołek y taki, że $(y, x) \in A$ i $f(y, x) > 0$ to wierzchołkowi y nadajemy cechę $(x^-, \varepsilon(y))$, gdzie $\varepsilon(y) = \min\{\varepsilon(x), f(y, x)\}$

Postępowanie cechowania kończy się jeżeli został ocechowany odpływ t lub nie ocechowano t, lecz procesu cechowania nie da się kontynuować (żadnego jeszcze nieocechowanego wierzchołka nie da się już ocechować).

(i) Jeżeli ocechowaliśmy wierzchołek t to t otrzymał cechę postaci (x^+, ε) lub (x^-, ε) (dla pewnego $x \in V$).
W pierwszym przypadku modyfikujemy przepływ na łuku (x, y) wzorem $\tilde{f}(x, t) = f(x, t) = \varepsilon$, w drugim $\tilde{f}(t, x) = f(t, x) - \varepsilon$.
Wierzchołek x jest ocechowany. Jeżeli pierwszy element jego cechy jest postaci y^+ to do przepływu na łuku (y, x) dodajemy ε, jeżeli jest to y^-

od przepływu na łuku (x, y) odejmujemy ε.
Powtarzając tę czynność odpowiednią liczbę razy dojdziemy w końcu do źródła s. Dojdziemy wzdłóż wskazanej przez pierwsze elementy cech wierzchołków ścieżkę zwaną *ścieżką powiększającą* [9]. Oczywiście wartość nowego przepływu $\tilde{f}$ wynosi $w(f) + \varepsilon > w(f)$. Proces cechowania rozpoczynamy od nowa przyjmując jako wyjściowy przepływ $\tilde{f}$.

(ii) Przypuśćmy, że w procesie cechowania nie udało się ocechować wierzchołka t, a żadnego nowego wierzchołka nie da się już ocechować. Jest oczywistym, że jeżeli przez W oznaczymy zbiór wierzchołków ocechowanych to $(W, \bar{W})$ jest przekrojem rozdzielającym s od t. Co więcej, dla wszystkich $x \in W$ i $y \in \bar{W}$ spełniających $(x, y) \in A$ zachodzi $f(x, y) = c(x, y)$, a jeśli $(y, x) \in A$ to $f(y, x) = 0$. Na podstawie uwagi 8.2 otrzymamy

$$
\begin{aligned}
w(f) &= \sum_{x \in W y \in \bar{W}} f(x, y) - \sum_{x \in W, y \in \bar{W}} f(y, x) \\
&= \sum_{x \in W, y \in \bar{W}} c(x, y) = c(W, \bar{W})
\end{aligned}
$$

Na mocy uwagi 8.1 oznacza to, że przepływ f jest przepływem maksymalnym. ■

8.6 Przepływ całkowity. Zbieżność algorytmu Forda–Fulkersona

Macierzą ograniczeń problemu (8.7) jest macierz incydencji grafu zorientowanego. Z twierdzenia 8.1.1 i wniosku 5.3.2 wynika, że istnieje taki przepływ maksymalny którego wartości na wszystkich łukach są całkowite. Co więcej, łatwo zauważyć, że całkowite rozwiązanie problemu maksymalnego przepływu można znaleźć przy pomocy algorytmu Forda–Fulkersona. W rzeczy samej, jeśli tylko przepływ początkowy jest całkowity[10] każda iteracja zwiększa wartość przepływu o liczbę całkowitą (liczba ε o którą zwiększa sie przepływ jest wtedy całkowita). Ponieważ wartość przepływu jest ograniczona przez przepustowość dowolnego przekroju, wszystkich iteracji będzie co najwyżej tyle ile wynosi przepustowość (jakiegokolwiek) przekroju.
W sieci w której przepustowości są liczbami wymiernymi, problem można łatwo sprowadzić do przypadku całkowitego. Wystarczy przemnożyć wszystkie przepustowości przez ich wspólny mianownik m. W otrzymanej sieci znajdujemy mnaksymalny przepływ a następnie wynik (t.j. zarówno wartość przepływu $v(f)$ jak i przepływy na wszystkich łukach) dzielimy przez m i otrzymujemy w ten sposób optymalne rozwiązanie problemu wyjściowego.

[9] Dokładniej ścieżka powiększająca (z którą zetknęliśmy się już w dowodzie twierdzenia 8.3.4) to ścieżka z s do t – właśnie przeszliśmy po niej *pod prąd*.
[10] Jako pierwszy przepływ można przyjąć przepływ zerowy.

Ze względu na ważne zastosowania rozwiązań całkowitych problemu maksymalnego przepływu które poznamy w dalszych podrozdziałach, zapiszmy następujący wniosek.

Wniosek 8.6.1 *Jeśli wszystkie ograniczenia przepustowości łuków w problemie (8.7) są całkowite, to istnieje rozwiązanie optymalne tego problemu, którego wartości na wszystkich łukach są całkowite. Co więcej, rozwiązanie o tej własności można otrzymać stosując algorytm Forda–Fulkersona.*

Niestety powyższe własności algorytmu Forda–Fulkersona nie przenoszą się na przypadek rzeczywisty[11]. Okazuje się, że można skonstruować prosty przykład sieci o przepustowościach rzeczywistych niewymiernych taki, że algorytm Forda–Fulkersona nie tylko nie dostarcza rozwiązania optymalnego w skończonej liczbie kroków, ale daję nieskończony ciąg przepływów których ciąg wartości nie jest zbieżny do rozwiązania optymalnego! Przykład takiej sieci podajemy za [9].

Przykład 8.6.1 Zauważmy wpierw, że rozwiązaniem równania rekurencyjnego

$$a_{n+2} = a_n - a_{n+1},\ a_0 = 1,\ a_1 = \frac{-1+\sqrt{5}}{2}$$

jest $a_n = \left(\frac{-1+\sqrt{5}}{2}\right)^n$. Oznaczmy przez S sumę szeregu $\sum_{n=0}^{+\infty} a_n$,

$$S = \sum_{n=0}^{+\infty} \left(\frac{-1+\sqrt{5}}{2}\right)^n$$

(oczywiście szereg $\sum_{n=0}^{+\infty} a_n$ jest szeregiem geometrycznym zbieżnym).
Skonstruujmy sieć $(V, A; c)$ w której $V = \{s, x_1, ..., x_4, y_1, ..., y_4, t\}$, $A = \{(s, x_i) : i = 1, ..., 4\} \cup \{(x_i, y_j) : i, j = 1, ..., 4\} \cup \{(y_i, x_j) : i, j = 1, ..., 4\} \cup \{(y_i, y_i) : i \neq j, i, j = 1, ...4\} \cup \{(y_i, t) : i = 1, ..., 4\}$. Przepustowości łuków zdefiniowane są zaś następująco. Dla łuków $e_i = (x_i, y_i)(i = 1, ..., 4)$:

$$c(x_1, y_1) = a_0$$

$$c(x_2, y_2) = a_1$$

$$c(x_3, y_3) = a_2$$

$$c(x_4, y_4) = a_2$$

zaś dla pozostałych łuków przepustowości są równe S. Jest jasne, że maksymalny przepływ z s do t ma wartość $4S$. Tymczasem algorytm Forda–Fulkersona może postępować następująco:
Początkowym przepływem jest przepływ równy zero na wszystkich łukach.
Oznaczmy przez $c_n(e) = c(e) - f_n(e)$ *rezydualną przepustowość* łuku e, gdzie f_n

[11]Kłopot tylko teoretyczny. Dość trudno sobie wyobrazić zadanie praktyczne w którym niezbędne byłoby uwzględnianie ograniczeń niewymiernych na przepustowość łuków.

jest funkcją przepływu wyznaczoną po $n-$tej iteracji.
Pierwszą ścieżką zwiększającą niech będzie

$$P_1 = (s, x_1, y_1, t)$$

Wtedy

$$c_1(x_1, y_1) = 0,\ c_1(x_2, y_2) = a_1,\ c_1(x_3, y_3) = a_2,\ c_1(x_4, y_4) = a_2$$

Następne iteracje definiujemy indukcyjnie. Przypuśćmy, że

$$c_n(e_1') = 0,\ c_n(e_2') = a_n,\ c_n(e_3') = a_{n+1},\ c_n(e_4') = a_{n+1}$$

gdzie $e_1',\ e_2',\ e_3',\ e_4'$ są pewnym przenumerowaniem łuków e_1, e_2, e_3, e_4. Jako nową ścieżkę powiększającą przyjmiemy

$$s, x_2', y_2', x_3', y_3', t$$

Wtedy $\epsilon = a_{n+1}$ i o tę wartość możemy zwiększyć przepływ wzdłóż ścieżki powiększającej. Otrzymamy

$$c_{n+1}(e_1') = 0,\ c_{n+1}(e_2') = a_{n+2},\ c_{n+1}(e_3') = 0,\ c_{n+1}(e_4') = a_{n+1}$$

W następnym kroku bierzemy ścieżkę powiększającą

$$s, x_2', y_2', y_1', x_1', y_3', x_3', y_4' t$$

W tej ostatniej ścieżce łuki (x_1', y_1') i (x_3', y_3') są niezgodne (wszystkie pozostałe łuki są zgodne). Zwiększyć przepływ należy teraz o a_{n+2} i otrzymamy

$$c_{n+2}(e_1') = a_{n+2},\ c_{n+2}(e_2') = 0,\ c_{n+2}(e_3') = a_{n+2},\ c_{n+1}(e_4') = a_{n+1}$$

Po takim, przeprowadzonym w dwóch etapach, kroku algorytmu w dwóch etapach wartość przepływu zwiększa się o $a_{n+1} + a_{n+2} = a_n$. Nie tylko więc opisane operacje zwiększania przepływu możemy powtarzać w nieskończoność, ale wartość przepływu jest zbieżna do S, podczas gdy wartość przepływu maksymalnego jest równa $4S$.

8.7 Wnioski i zastosowania

Twierdzenie o maksymalnym przepływie i minimalnym przekroju i algorytm Forda–Fulkersona są źródłem wielu ważnych wniosków i prostych dowodów ważnych twierdzeń – także takich które znane były znacznie wcześniej. W bardzo wielu przypadkach kluczową rolę odgrywa tu wniosek 8.6.1.

Zastosowania w teorii grafów

Do tej pory mowa była jedynie o grafach zorientowanych. Teraz potrzebować będziemy grafu zwykłego i grafu dwudzielnego. *Grafem zwykłym* nazywamy parę $G = (V; E)$, gdzie V jest dowolnym zbiorem zwanym *zbiorem wierzchołków* grafu G, zaś E jest pewnym podzbiorem zbioru dwuelementowych podzbiorów zbioru V, czyli

$$E \subset \{\{x, y\} : x, y \in V, x \neq y\}$$

nazywanym zbiorem **krawędzi grafu**. Krawędź $\{x, y\}$ grafu G, zapisujemy krócej xy. x i y nazywamy jej **końcami**. Jeśli $x, y \in E$ to mówimy, że x i y są wierzchołkami **sąsiednimi** lub **połączonymi** . Zbiór sąsiadów wierzchołka x w grafie G znaczamy przez $N_G(x) = \{y \in V : xy \in E\}$.
Grafy zwykłe mają swoją – bardzo wygodną i sugestywną – interpretację graficzną: wierzchołki w tej reprezentacji są przedstawiane jako punkty płaszczyzny a krawędzie jako łączące je krzywe (p. rys. 8.5.1).
Zbiór krawędzi $F \subset E$ jest **niezależny** jeżeli dla dowolnych różnych krawędzi $e, f \in F$ zachodzi $e \cap f = \emptyset$. Zbiory niezalężne krawędzi grafu G nazywać będziemy **skojarzeniami** w grafie G.
Graf G nazywamy **dwudzielnym** jeżeli istnieje podział[12] zbioru $V = X \cup Y$ taki, że dla każdej krawędzi $xy \in E$ jeden z wierzchołków x, y należy do zbioru X podczas gdy drugi należy do zbioru Y. Piszemy wtedy $G = (X, Y; E)$.
Problem znajdowania skojarzeń w grafach – na przykład **pełnych skojarzeń** czyli takich zbiorów M niezależnych krawędzi, że $\bigcup_{e \in M} e = V$, lub skojarzeń o maksymalnej liczbie krawędzi, są bardzo ważnymi problemami teorii grafów, posiadającymi także zastosowania (czy też interpretacje) praktyczne.

Przykład 8.7.1 W pewnym zakładzie pracy zatrudnieni są pracownicy $p_1, ..., p_k$. W zakładzie tym są też stanowiska pracy $s_1, ..., s_l$, przy czym nie każdy pracownik może pracować przy każdym stanowisku. Dla dobrego funkcjonowania zakładu istotny jest problem jak przydzielić stanowiska pracy wszystkim zatrudnionym, by optymalnie wykorzystać pracowników i obsadzić stanowiska pracy.
Taka sytuacja łatwo daje się opisać przy pomocy grafu dwudzielnego $G = (X, Y; E)$, gdzie $X = \{p_1, ..., p_k\}, Y = \{s_1, ..., s_l\}, E = \{p_i s_j\ :\ p_i$ może obsługiwać stanowisko $j\}$.
Nasz problem w języku teorii grafów można sformułować tak: *znajdź w G skojarzenie M o maksymalnej liczbie krawędzi.* Sytuacja będzie oczywiście najlepsza wtedy, gdy takie skojarzenie będzie **pełnym skojarzeniem** X **w** Y grafu dwudzielnego $G = (X, Y; E)$, t.j. wtedy gdy dla każdego $x \in X$ istnieje krawędź $e \in E$ taka, że $x \in e$[13].

W związku z powyższym przykładem powstaje naturalny problem: *kiedy w grafie dwudzielnym $G = (X, Y; E)$ istnieje pełne skojarzenie X w Y?*
Na pytanie to odpowiada ważne twierdzenie Halla z 1935 roku.

[12]Przypomnijmy, że zbiory X i Y są **podziałem** V jeżeli $X \cup Y = V$ i $X \cap y = \emptyset$.
[13]Zwrócić tu anleży uwagę na subtelną różnicę pomiędzy *pełnym skojarzeniem* w grafie zwykłym i *pełnym skojarzeniem X w Y w grafie dwudzielnym* $(X, Y; E)$.

Twierdzenie 8.7.1 (P. Hall) *Niech $G = (X, Y; E)$ będzie grafem dwudzielnym. W G istnieje skojarzenie pełne X w Y wtedy i tylko wtedy, jeśli dla każdego podzbioru $A \subset X$ zachodzi $\mid A \mid \leq \mid N_G(A) \mid$.*

Dowód. Zacznijmy od skojarzenia z grafem dwudzielnym $G = (X, Y; E)$ sieci $S = (X' \cup Y'; A, c)$ w sposób następujący (por. rys. 8.5.1).

– zbiorem wierzchołków sieci S jest $V = X' \cup Y'$, gdzie

$$X' = X \cup \{s\}$$
$$Y' = Y \cup \{t\}$$

– s jest źródłem,

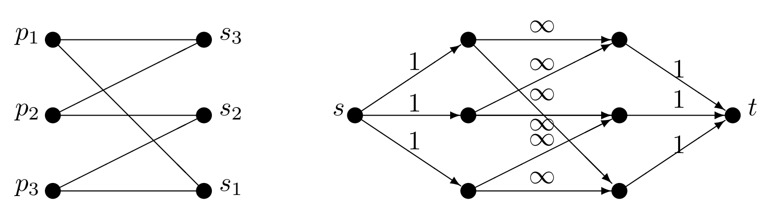
– t odpływem,

przy czym o wierzchołkach s i t zakładamy, że nie należą do $X \cup Y$ i są między sobą różne.

– Zbiorem łuków sieci S jest $A = \{(x, y) : x = s, y \in X$ lub $x \in X, xy \in E,$ lub $x \in Y, y = t\}$.

– Funkcja przepustowości $c : A \to \mathbf{R}$ dana jest wzorem

$$c(x,y) = \begin{cases} \infty & \text{gdy} \quad x \in X \text{ i } y \in Y \\ 1 & \text{gdy} \quad x = s \text{ i } y \in X \text{ lub } x \in Y \text{ i } y = t \end{cases}$$



G S

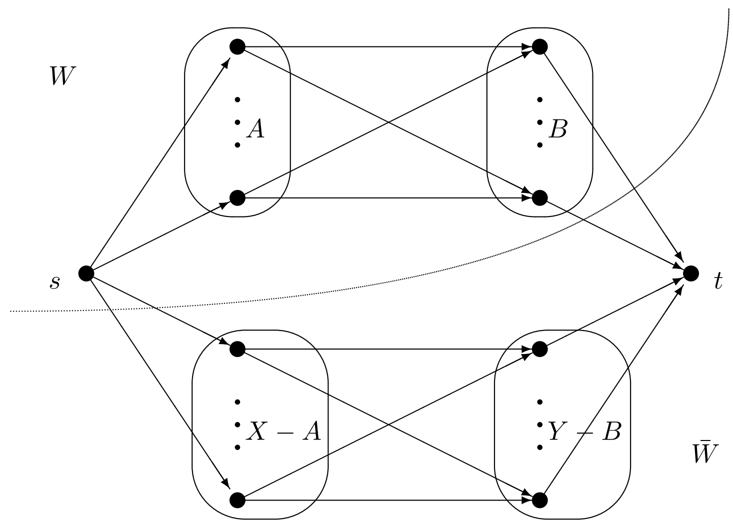
Rys. 8.5.1

Oczywistym jest, że maksymalny przepływ **całkowitoliczbowy** w sieci S wyznacza maksymalne skojarzenie w G i na odwrót: maksymalne skojarzenie w G definiuje całkowitoliczbowy przepływ w sieci S. Z wniosku 8.3.1 (lub 8.6.1) wynika, że istnieje maksymalny przepływ w S przyjmujący wartości całkowite na wszystkich łukach sieci[14].

Przypuśćmy wpierw, że dla każdego podzbioru $A \subset X$ zachodzi $\mid N_G(A) \mid \geq \mid A \mid$. Wykażemy, że wtedy istnieje w S przepływ f o wartości

[14]Zauważmy, że $+\infty$ w definicji przepustowości łuków sieci S można zastąpić wystarczająco dużymi liczbami całkowitymi (na przykład $\mid X \mid$).

$w(f) \geq\mid X \mid$.
Rzeczywiście niech f będzie całkowitoliczbowym przepływem maksymalnym w S i niech $(W, \bar{W})$ będzie przekrojem o przepustowości minimalnej rozdzielającym s od T (na przykład wyznaczonym przez algorytm Forda–Fulkersona).
Oznaczmy $A = X \cap W$ oraz $B = Y \cap W$ (p. rys. 8.5.2).



Rys. 8.5.2

Zauważmy, że nia ma łuków o początku w A i końcu w $Y-B$, w przeciwnym razie mielibyśmy $c(W, \bar{W}) = +\infty$, a więc istnieje przekrój o mniejszej przepustowości, n.p. $W = \{s\}$[15].
Stąd $B \supset N_G(A)$ i

$$c(W, \bar{W}) =\mid X - A \mid + \mid B \mid=\mid X \mid - \mid A \mid + \mid B \mid\geq$$

$$\geq\mid X \mid - \mid A \mid + \mid N_G(A) \mid\geq\mid X \mid$$

Tak więc istnieje przepływ w S o wartości $w(f) \geq\mid X \mid$, a skoro żaden przepływ nie może mieć wartości większej niż $\mid X \mid$, $w(f) =\mid X \mid$.
łuki o niezeroeym przepływie z X do Y w S wyznaczają oczywiście krawędzie pełnego skojarzenia X w Y w grafie G.

Przypuśćmy teraz, że w G istnieje pełne skojarzenie X w Y. Wykażemy, że wtedy dla każdego $A \subset X$ zachodzi $\mid A \mid\leq\mid N_G(A) \mid$.
Rzeczywiście, gdyby istnił zbiór $A \subset X$ taki, że $\mid A \mid>\mid N_G(A) \mid$ wtedy w sieci S moglibyśmy wziąć przekrój $(W, \bar{W})$ ze zbiorem $W = \{s\} \cup A \cup N_G(A)$.
Przepustowość tego przekroju wynosi

$$c(W, \bar{W}) =\mid X \mid - \mid A \mid + \mid N_G(A) \mid<\mid X \mid$$

[15] $c(\{s\}, X' \cup Y' - \{s\}) =\mid X \mid< +\infty$.

Skoro istnieje przekrój rozdzielający o przepustowości mniejszej niż $\mid X \mid$, to w S nie ma przepływu o wartości $\mid X \mid$, czyli w G nie istnieje skojarzenie pełne X w Y – ta sprzeczność kończy dowód twierdzenia 8.7.1 ■

Dla grafów dwudzielnych prawdziwe jest następujące twierdzenie.

Zbiór różnych reprezentantów

Niech będzie dany zbiór X i pewna rodzina podzbiorów $\mathcal{X}$ tego zbioru. Kiedy istnieje zbiór $Y \subset X$ taki, że istnieje **różnowartościowa** funkcja $g : \mathcal{X} \rightarrow \mathcal{Y}$? Jeśli taki zbiór Y istnieje, nazywamy go *zbiorem różnych reprezentantów* rodziny $\mathcal{X}$.

Przykład 8.7.2 W sejmie, senacie[16] istnieją różne komisje zajmujące się różnymi problemami [17].

Komisje te kędziemy traktowali jako podzbioru zbioru posłów (lub senatorów). Każda komisja ma swojego przewodniczącego. Wygodnie jest by przewodniczącymi róznych komisji były różne osoby. Powstaje więc problem: jak powinny być dobrane komisje by mogło tak być?

Problem istnienia zbioru różnych reprezentantów można *przetłumaczyć* na język teorii grafów następująco:

Niech każdemu zbiorowi $X_i \in \mathcal{X}$ odpowiada wierzchołek x_i i niech $A = \{x_i : X_i \in \mathcal{X}\}$, oraz $B = X$. Niech teraz G bedzie grafem dwudzielnym $G = (A, B; E)$ takim, że $E = \{X_i x : x \in X_i\}$. Problem istnienia zbioru różnych reprezentantów rodziny $\mathcal{X}$ jest równoważny istnieniu w G skojarzenia pełnego zbioru A. Stąd i z twierdzenia 8.7.1 wynika bardzo łatwo następujący wniosek.

Wniosek 8.7.2 (Hall) *[12] Niech $\mathcal{X}$ będzie rodziną n podzbiorów zbioru X. Zbiór różnych reprezentantów rodziny $\mathcal{X}$ istnieje wtedy i tylko wtedy, gdy unia dowolnych $k \leq n$ zbiorów rodziny $\mathcal{X}$ zawiera co najmniej k elementów.*

Następne, słynne twierdzenie Königa-Egerváry'ego, zostało opublikowane w książce Königa [15] której ukazanie się w 1936 roku stało się, jak się powszechnie uważa, źaczynem"trwającego do dziś bardzo intensywnego rozwoju teorii grafów[18].

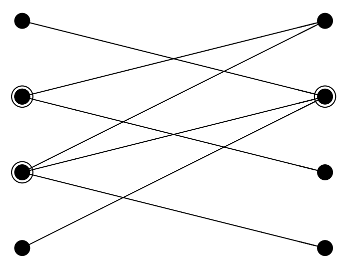
Twierdzenie 8.7.3 (Königa-Egerváry'ego) *W grafie dwudzielnym $G = (X, Y; E)$ minimalna liczba wierzchołkow w zbiorze rozdzielającym X i Y jest równa liczbie krawędzi w maksymalnym skojarzeniu.*

Przykład 8.7.3 Na rysunku 8.5.3 przedstawiono graf dwudzielny. Sprawdź, że wierzchołki wyróżnione tworzą w nim zbiór rozdzielający i wskaż skojarzenie o trzech krawędziach.

[16] Także Senacie Akademii Górniczo–Hutniczej.

[17] W Senacie AGH są to Komisje do Spraw Kształcenia, Naukowa, Budżetowa i inne.

[18] Z pewnością powodów gwałtownego rozwoju teorii grafów było z pewnością wiele (np. znany *problem czterech kolorów*, liczne zastosowania (w tym w informatyce)). Niewątpliwie jednak książka Königa odegrała w tym rozwoju dużą rolę.



Rys. 8.5.3

Dowód twierdzenia 8.7.3 jest podobny do dowodu twierdzenia 8.7.1. Z grafem G kojarzymy sieć $S = (X', Y'; E')$ w identyczny co w dowodzie twierdzenia 8.7.1 sposób i zauważamy, że każdemu przekrojowi $(W, \bar{W})$ z $W = A \cup B \cup \{s\}$, $A \subset X, B \subset Y$, o przepustowości w S równej, jak widzieliśmy, $\mid X - A \mid + \mid B \mid$, odpowiada zbiór rozdzielający $(X - A) \cup B$ który też ma $\mid X - A \mid + \mid B \mid$ wierzchołków. ■

8.8 Przepustowość wierzchołków

Zanim powiemy jak powiemy jak uogólnić twierdzenie 8.7.3 dla dowolnych grafów (niekoniecznie dwudzielnych), powiedzmy jak poradzić sobie z problemem maksymalnego przepływu w sieci w której poza ograniczeniami na przepustowości łuków dane są także przepustowości wierzchołków.
Mamy wtedy następujący problem:
Dla danej sieci $S = (X, A; c, \tilde{c})$, *gdzie* $c : A \to \mathbf{R}^+$ $\tilde{c} : X \to \mathbf{R}^+$ *o źródle* s *i odpływie* t *znaleźć funkcję* $f : A \to \mathbf{R}$ *spełniającą*

$$\sum_{(x,y)\in A} f(x,y) - \sum_{(z,x)\in A} f(z,x) = 0 \quad \textit{dla każdego } x \in V - \{s,t\}$$

$$\sum_{(z,x)\in A} f(z,x) \leq \tilde{c}(x) \quad \textit{dla każdego } x \in V$$

$$0 \leq f(x,y) \leq c(x,y) \quad \textit{dla wszystkich } (x,y) \in A$$

zmaksymalizować

$$\sum_{(s,x)\in A} f(s,x) - \sum_{(y,s)\in A} f(y,s)$$

Rozwiązanie powyższego problemu[19] otrzymujemy bardzo łatwo korzystając z poznanej już metody dla sieci bez ograniczeń przepustowości wierzchołków.
Tworzymy nową sieć S', A', c') w której każdy wierzchołek x zastępujemy dwoma wierzchołkami x' i x''.
Zbiór łuków A' i przepustowości definiujemy następująco:

1. $A' = \{(x', x'') : x \in V\} \cup \{(y'', x') : (y, x) \in A\}$,
2. $c'(x', x'') = \tilde{c}(x)$ dla każdego $x \in V$,
3. $c'(y'', x') = c(x, y)$ jeśli $(y, x) \in A$

Jest oczywiste, że wartość maksymalnego przepływu f w sieci S ze źródła s do odpływu t jest równa wartości maksymalnego f' przepływu w S' ze źródła s' do odpływu t''.
Oczywistym jest także, że f' można znaleźć przy pomocy algorytmu Forda–Fulkersona i w łatwy sposób sprowadzić potem do przepływu f w sieci S.

8.9 Twierdzenie Mengera

W niniejszym podrozdziale będzie mowa o spójności grafu i parametrze $\kappa(G)$ który jest w pewien sposób miarą spójności grafu G. Zarówno spójność jak parametr κ są podstawowymi, bardzo ważnymi pojęciami teorii grafów.
Dla grafu $G = (V; E)$ i podzbioru $S \subset V$ przez $G - S$ oznaczamy graf $G' = (V - S; E')$, gdzie $E' = \{e \in E : e \subset V - S\}$. Inaczej mówiąc, zbiór krawędzi grafu $G - S$ to te krawędzie grafu G których żaden z końców nie jest w S.
ścieżką z x do y w grafie $G = (V; E)$ nazywamy ciąg wierzchołków i krawędzi

$$P = (x_1, e_1, x_2, e_2, ..., x_k, e_k, x_{k+1})$$

$(x_i \in V, i = 1, ..., k + 1, e_j \in E, j = 1, ..., k)$ taki, że $e_i = x_i x_{i+1}$.
Mówimy, że graf G jest **spójny** jeśli dla dowolnych wierzchołków x i y grafu G istnieje ścieżka z x do y (mówimy także **ścieżka łącząca** x z y).
Minimalną liczbę k taką, że w grafie G istnieje zbiór wierzchołków S spełniający warunki

1. $\mid S \mid = k$
2. $G - S$ nie jest grafem spójnym lub ma tylko jeden wierzchołek

nazywamy **liczbą spójności grafu** G , lub **spójnością grafu** i oznaczamy przez $\kappa(G)$.

[19]Zauważmy, że ograniczenia przepustowości wierzchołków mogą występować w zagadnieniach praktycznych. W sieciach komunikacyjnych często wierzchołki oznaczają skrzyżowania ulic. Jest oczywiste, że możliwość wprowadzenia ograniczeń przepustowości wierzchołków – skrzyżowań jest tu bardzo istotna.

Przykład 8.9.1 Czytelnik zechce sprawdzić, że graf G przedstawiony na rysunku 8.5.1 ma liczbę spójności równą 2 ($\kappa(G) = 2$).

O grafie G dla którego $\kappa(G) = k$ mówimy, że jest $l-$**spójny**, dla każdego $l \leq \kappa(G)$.

Przykład 8.9.2 Graf G z rysunku 8.5.1 jest $1-$spójny i $2-$spójny.

Oczywiście grafy spójne są $1-$spójne.
Niech G będzie grafem spójnym ($\kappa(G) \geq 1$). Zbiór S taki, że $G - S$ nie jest spójny nazywamy **separatorem** grafu G. Separator S o minimalnej liczbie wierzchołków (a więc taki, że $\mid S \mid= \kappa(G)$) nazywamy separatorem minimalnym. Niech P i P' będą dwoma ścieżkami łączącymi dwa różne wierzchołki x i y. Mówimy, że P i P' są **wewnętrznie rozłączne** jeżeli jedynymi wspólnymi wierzchołkami tych ścieżek są x i y.

Twierdzenie 8.9.1 *Maksymalna liczba ścieżek wewnętrznie rozłącznych łączących dwa wierzchołki x i x grafu G jest równa liczności minimalnego separatora rozdzielającego x i y.*

Dowód. Z grafem $G = (V; E)$ skojarzmy sieć S o zbiorze wierzchołków V, oraz źródle, odpływie, zbiorze łuków A i funkcjami przepustowości łuków i wierzchołków zdefiniowanymi w sposób następujący:

1. x jest źródłem sieci S,
2. y jest odpływem S,
3. każdej krawędzi ab w grafie G odpowiadają dwa łuki w S: (a, b) i (b, a),
4. przepustowość każdego łuku wynosi $+\infty$,
5. przepustowości wszystkich wierzchołków różnych od x i y wynoszą 1,
6. przepustowości x i y wynoszą $+\infty$.

Maksymalny całkowitoliczbowy przepływ z x do y wyznaczy nam oczywiście maksymalną liczbę ścieżek wewnątrznie rozłącznych z x do y. Przekrój o minimalnej przepustowości w sieci S wyznaczy z kolei minimalny separator x od y. Z twierdzenia o maksymalnym przepływie i minimalnym przekroju wynika teraz twierdzenie 8.9.1. ■
Z twierdzenia 8.9.1 wynika natychmiast twierdzenie Mengera:

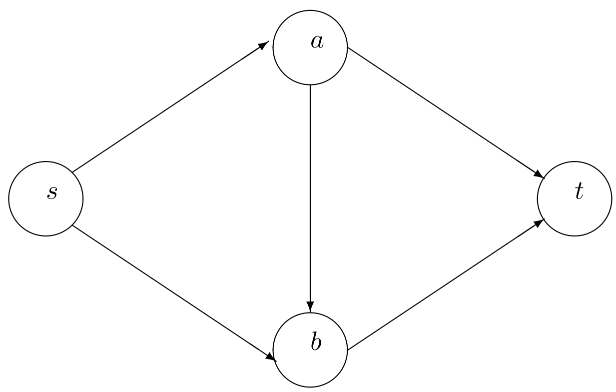
Wniosek 8.9.2 (Twierdzenie Mengera) *W dowolnym grafie G każde dwa wierzchołki są połączone $\kappa(G)$ ścieżkami wewnętrznie rozłącznymi.* ■

Czytelnik zechce samodzielnie zastanowić się jak udowodnić następny wniosek, który jest uogólnieniem twierdzenia Mengera (por. ćwiczenie 8.9.4).

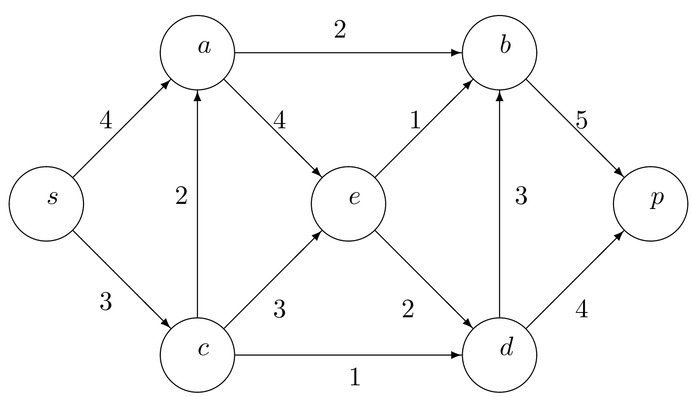
Wniosek 8.9.3 *Niech A i B będą dwoma rozłącznymi zbiorami wierzchołków grafu G takimi, że $\mid A \mid$, $\mid B \mid\geq \kappa(G)$. Wtedy istnieje w G $\kappa(G)$ wierzchołkowo rozłącznych ścieżek z A do B.*

8.9.1 ćwiczenia

Ćwiczenie 8.9.1 Napisz problem prymalny i dualny dla znajdowania maksymalnego przepływu w sieci przedstawionej na poniższym rysunku (dla dowolnej funkcji przepustowości $c : A \to \mathbf{R}$).



Ćwiczenie 8.9.2 Metodą Forda-Fulkersona znajdź maksymalny przepływ i minimalny przekrój w poniższej sieci.



Ćwiczenie 8.9.3 Skorzystaj z algorytmu Forda-Fulkersona by znaleźć maksymalne skojarzenie w grafie dwudzielnym $G = (X, Y; E)$, gdzie:

$X = \{u, v, w, x, y, z\}$

$Y = \{a, b, c, d, e, f\}$

$E = \{ub, ud, va, vb, vc, xb, xd, yb, yd, za, ze, zf, wd, we, wf\}$

Skomentuj otrzymany wynik.

Ćwiczenie 8.9.4 Wykaż, że jeśli graf jest k spójny to dla dowolnych rozłącznych zbiorów wierzchołków A i B takich, że $|A|, |B| \geq k$ istnieje k ścieżek $P_1, ..., P_k$ o rozłącznych zbiorach wierzchołków takich, że początkowy wierzchołek każdej ze ścieżek jest w zbiorze A a końcowy w B.

Rozdział 9

Problem transportowy

W problemie transportowym zajmujemy się przepływami, podobnie jak w przypadku problemu maksymalnego przepływu. Tym razem jednak zakładamy, że przepustowości łuków są nieograniczone, natomiast z przepływem związane są koszty. Mamy więc pewną liczbę magazynów o określonych zasobach, pewną liczbę odbiorców o określonych ptrzebach a także węzły pośrednie (stacje przeładunkowe) do których wpływa tyle samo transportowanych dóbr, co i wypływa. Mamy też łączące węzły łuki przez które odbywa się przepływ dóbr, który jest związany z pewnymi kosztami. Celem jest dostarczenie dóbr do odbiorców po jak najmniejszym koszcie.
Także tym razem będziemy musieli podać pewną ilość informacji z teorii grafów, konkretnie informacje o grafach zwanych drzewami, które będą nam niezbędne w dalszej części algorytmu zwanego **simpleksem sieciowym**.

9.1 Drzewa

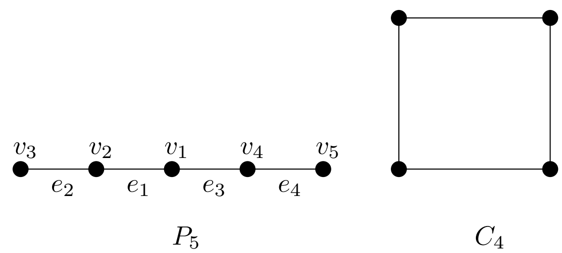
Trasą w grafie nieskierowanym $G = (V; E)$ nazywamy ciąg wierzchołków $P_k = (v_0, v_1, ..., v_{k-1})$ taki, że dowolne dwa kolejne wierzchłki trasy v_i, v_{i+1} są połączone krawędzią. **Droga** to taka trasa $(v_0, v_1, ..., v_{k-1})$, w której $v_i \neq v_{i+1}$ dla $i = 0, ..., k-2$. Jeśli dodatkowo $v_{k-1} = v_0$, wówczas drogę nazywamy **cyklem** (piszemy wtedy $C_k = (v_0, v_1, ..., v_{k-1}, v_0)$). O grafie G mówimy, że jest **spójny** jeżeli dowolne dwa wierzchołki tego grafu są połączone drogą.

Długością drogi (cyklu) nazywamy liczbę tworzących je krawędzi.

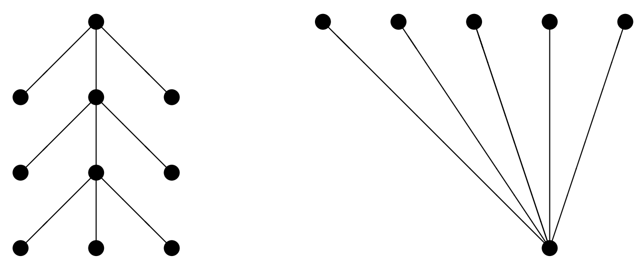
Oczywiście grafy P_5 i C_4 (przedstawione na rysunku 9.1) są spójne natomiast graf, który jest sumą P_5 i C_4, czyli całym grafem przedstawionym na tym rysunku jest niespójny. Maksymalny podgraf spójny grafu nazywamy **składową**. Graf przedstawiony na rysunku 9.1 ma dwie składowe: P_5 i C_4.

Drzewem nazywamy graf spójny nie zawierający cykli (p. rysunek 9.2).

Dla grafu $G = (V; E)$ i dwóch niepołączonych wierzchołków $u, v \in V$ przez $G + uv$ oznaczamy graf powstały przez połączenie w G wierzchołków u i v krawędzią. Inaczej $G + uv = (V; E \cup \{uv\})$.



Rysunek 9.1: Dwa grafy spójne



Rysunek 9.2: Drzewa

Twierdzenie 9.1.1 *Graf $G = (V; E)$ rzędu $n = |V| > 2$, niepełny, jest drzewem wtedy i tylko wtedy gdy dla dowolnych dwóch niepołączonych wierzchołków $u, v \in V$ graf $G + uv$ zawiera dokładnie jeden cykl.*

Dow"od.

- Przypuśćmy, że G jest drzewem a u i v są niepołączonymi wierzchołkami tego grafu. Skoro G jest spójny, istnieje w G droga łącząca u z v, powiedzmy $(u = v_1, v_2, ..., v_k = v)$. Oczywiście $(v_1, v_2, ..., v_k, v_1)$ jest cyklem grafu $G + uv$. Co więcej, cykl ten jest jedynym cyklem w $G + uv$. Gdyby bowiem istniały dwa cykle w $G + uv$, oba musiałyby zawierać krawędź uv, bowiem G jako drzewo żadnego cyklu nie zawiera. Mielibyśmy, powiedzmy, cykle $(u = u_1, u_2, ..., u_l = v, u)$ oraz $(u = w_1, w_2, ..., w_m = v, u)$. Wtedy jednak $C = (u_1, u_2, ..., u_l = w_m, w_{m-1}, ..., w_1 = u)$ byłby cyklem w G[1].

- Przypuśćmy, że graf G jest taki, że po połączeniu dowolnych dwóch niepołączonych wierzchołków istnieje w nim dokładnie jeden cykl.

[1] Zwracam uwagę na lukę w dowodzie. Zgodnie z definicją cyklu przyjętą w tym rozdziale, wierzchołki w cyklu występują co najwyżej jeden raz. W konstrukcji podanej w dowodzie otrzymujemy cykl, w którym wierzchołki mogą się powtarzać. łatwo jednak tę trudność pokonać.

G jest grafem spójnym. Rzeczywiście, jeśli $u, v \in V$ to albo te dwa wierzchołki są połączone krawędzią (drogą długości 1), albo nie, lecz wtedy w grafie $G + uv$ krawędź uv jest zawarta w cyklu, skąd łatwo wynika, że u jest w G połączone ścieżką z v. Co więcej, skoro dla dowolnych niepołączonych w G wierzchołków u, v graf $G_u v$ zawiera dokładnie jeden cykl, graf G nie może zawierać cyklu. ■

Twierdzenie 9.1.2 *Niech $T = (V; E)$ będzie drzewem rzędu n i niech v będzie dowolnym wierzchołkiem drzewa T. Wówczas istnieje takie uporządkowanie $v_1, v_2, ..., v_n$ wierzchołków grafu T, że $v = v_1$ i każdy wierzchołek v_i jest połączony z dokładnie jednym wierzchołkiem ze zbioru $\{v_1, v_2, ..., v_{i-1}\}$.*

Dow"od. Dla dowodu twierdzenia 9.1.2 podamy rekurencyjnie sposób uporządkowania $v_1, v_2, ..., v_n$ wierzchołków spełniający warunki twierdzenia.
Oczywiście musimy mieć $v_1 = v$.
Przypuśćmy teraz, że mamy uporządkowanie m wierzchołków $v_1, v_2, ..., v_m$ drzewa T spełniające warunki twierdzenia i na dodatek takie, że podgraf drzewa T indukowany przez te wierzchołki jest spójny (oczywiście wynika stąd, że podgraf ten jest drzewem). Jeśli $m = n$ wówczas nasza konstrukcja jest zakończona. Jeśli nie, wówczas ze spójności T wynika, że istnieje wierzchołek $x \in V - \{v_1, v_2, ..., v_m\}$ połączony z jednym z wierzchołków zbioru $\{v_1, v_2, ..., v_m\}$. Gdyby x był połączony z dwoma wierzchołkami spośród $\{v_1, ..., v_m\}$, na przykład gdyby $ux \in E, wx \in E$, gdzie $u, w \in \{v_1, ..., v_m\}, u \neq w$, wówczas istniałaby droga P łącząca u z w w podgrafie indukowanym przez $\{v_1, ..., v_m\}$ (wiemy, że ten podgraf jest spójny) i (xPx) byłby cyklem zawartym w T, a taki cykl nie istnieje, bowiem T jest drzewem. Stąd wynika, że x może być połączony tylko z jednym wierzchołkiem zbioru $\{v_1, ..., v_m\}$. Możemy więc przyjąc $v_{m+1} = x$. ■

Dla dowolnego grafu $G = (V; E)$ liczbę jego wierzchołków nazywamy **rzędem** grafu G i oznaczamy przez $|G|$. Liczbę krawędzi G nazywamy **rozmiarem** grafu G i oznaczamy przez $\| G \|$.

Dla dowolnego grafu G o zbiorze wierzchołków $V = \{u_1, ..., u_n\}$ i zbiorze krawędzi $E = \{e_1, ..., e_m\}$ macierz incydencji $M = M(G)$ zdefiniowana jest wzorem

$$M_{v_i, e_j} = \begin{cases} 1 & \text{jeśli krawędź } e_j \text{ jest incydentna z wierzchołkiem } v_i \\ 0 & \text{w przeciwnym przypadku} \end{cases}$$

W sytuacji takiej jak powyżej mówimy, że macierz M ma wiersze indeksowane wierzchołkami, a kolumny krawędziamu grafu G.

Przykład 9.1.1 Przykładowe uporządkowanie wierzchołków P_5 spełniające warunki twierdzenia 9.1.2 zaznaczono na rysunku 9.1. Na tym samym rysunku przez e_i oznaczono jedyną krawędź. która łączy wierzchołek v_i z jedynym wierzchołkiem o indeksie mniejszym niż i. Dla takich oznaczeń macierz incydencji jest równa

$$M(P_5) = \begin{array}{c|cccc} & e_1 & e_2 & e_3 & e_4 \\ \hline v_1 & 1 & 0 & 1 & 0 \\ v_2 & 1 & 1 & 0 & 0 \\ v_3 & 0 & 1 & 0 & 0 \\ v_4 & 0 & 0 & 1 & 0 \\ v_5 & 0 & 0 & 0 & 1 \end{array} \qquad (9.1)$$

(we wzorze (9.1) zaznaczono także wierzchołki i krawędzie indeksujące poszczególne wiersze i kolumny macierzy).

Wniosek 9.1.3 .

1. *Dla dowolnego drzewa T prawdziwy jest wzór $\| T \| = |T| - 1$.*
2. *Rząd macierzy incydencji dowolnego drzewa rzędu n jest równy $n - 1$.*

Dow"od. Niech wierzchołkami drzewa T będą $v_1, v_2, ..., v_n$ uporządkowane tak, jak w twierdzeniu 9.1.2.

Dow. 1. Dowód pierwszej części wniosku jest oczywisty: dodając do drzewa kolejny wierzchołek v_i dodajemy dokładnie jedną krawędź poza $i = 1$ kiedy to nie mamy jeszcze żadnej krawędzi.

Dow. 2. Macierz incydencji drzewa rzędu n ma $n - 1$ kolumn (tyle jest bowiem krawędzi, którymi kolumny są indeksowane). Stąd rząd macierzy M jest co najwyżej równy $n-1$. Zauważmy także, że po usunięciu pierwszego wiersza otrzymamy macierz o rozmiarze $(n - 1) \times (n - 1)$, które na przekątnej ma jedynki, natomiast pod przekątną zera. Wyznacznik tak otrzymanej macierzy kwadratowej jest równy 1, stąd jej rząd jest równy $n - 1$. Wobec tego rząd macierzy incydencji drzewa jest równy $n - 1$. ■

9.2 Drzewa skierowane

Niech $G = (V; E)$ będzie grafem skierowanym. **Szkieletem** grafu G nazywamy graf nieskierowany o zbiorze wierzchołków V (a więc tym samym, co zbiór wierzchołków grafu G, w którym każdy łuk grafu G zastępujemy krawędzią powstałą z tego łuku przez pominięcie jego orientacji. Zauważmy, że szkieletem grafu skierowanego może być graf, który nie jest grafem zwykłym, bowiem niektóre jego wierzchołki mogą być połączone krawędziami wielokrotnymi. Tak jest na przykład dla grafów przedstawionych na rysunku 9.3. Graf G' jest szkieletem grafu skierowanego G i nie jest grafem zwykłym gdyż dwa jego wierzchołki są połączone krawędzią podwójną (w ten sposób w G' istnieje cykl długości 2, co w grafach nieskierowanych zwykłych nie jest możliwe).

Drzewem skierowanym nazywamy graf skierowany, którego szkielet jest drzewem. Następujący wniosek jest dość oczywistą konsekwencją wcześniejszych twierdzeń dla drzew nieskierowanych. Przypomnijmy, że cyklem w grafie skierowanym jest ciągiem wierzchołków $C_k = (u_1, u_2, ..., u_k, u_1)$ takich, że $u_i \neq u_j$

Rysunek 9.3:

dla $i \neq j$ i każde dwa kolejne wierzchołki na cyklu są połączone (przy czym nie precyzujemy, czy $(u_i, u_{i+1}) \in E(C_k)$ czy też $(u_{i+1}, u_i) \in E(C_k)$).

Wniosek 9.2.1
1. *Graf skierowany $G = (V; E)$ rzędu $n = |V| > 2$, niepełny, jest drzewem wtedy i tylko wtedy gdy dla dowolnych dwóch niepołączonych wierzchołków $u, v \in V$ graf $G + (u, v)$ zawiera dokładnie jeden cykl.*

2. *Niech T będzie drzewem skierowanym rzędu n i niech v będzie dowolnym wierzchołkiem drzewa T. Wówczas istnieje takie uporządkowanie $v_1, v_2, ..., v_n$ wierzchołków grafu T, że $v = v_1$ i każdy wierzchołek v_i jest połączony, powiedzmy łukiem e_i, z dokładnie jednym wierzchołkiem ze zbioru $\{v_1, v_2, ..., v_{i-1}\}$.*

3. *Rozmiar dowolnego drzewa skierowanego rzędu n jest równy $n - 1$.*

4. *Rząd macierzy incydencji dowolnego grafu skierowanego rzędu n wynosi $n - 1$.*

9.3 Problem transportowy - simpleks sieciowy

Celem naszych rozważań będzie przypisanie przepływów łukom zadanej sieci (zwanej siecią transportową) w taki sposób aby spełnione były oczekiwania **odbiorców** (potrzeby). Transport tych dóbr odbywa się ze zbioru **magazynów**, których zapasy nie mogą zostać przekroczone i odpywa się w sieci, w której znany jest koszt jednostkowego (jednostki dóbr) przepływu przez każdy łuk. Przepustowość łuków nie jest ograniczona. Tym razem podejście do problemu jest takie, żeprzez każdy łuk da się przetransportować dowolną ilość dóbr, natomiast istotne są koszty, które minimalizujemy.
Nasza sieć jest więc grafem skierowanym $G = (V, A)$, przy czym $V = M \cup O \cup P$, gdzie

- M - to magazyny,
- O - odbiorcy,
- P - węzły pośrednie (możemy też interpretować te wierzchołki grafu G jako stacje przeładunkowe).

- Funkcja
$$c : E \to \mathbb{R}^+$$
jest funkcją kosztów - interpretujemy ja tak, że przepływ przez łuk e α jednostek dóbr kosztuje $\alpha \cdot c(e)$.

- W każdym magazynie $m \in M$ mamy określony zapas (powiedzmy) surowca. Podobnie, każdy odbiorca $o \in O$ ma pewne zapotrzebowanie na transportowany w sieci surowiec. Dla naszego problemu wygodniej jednak będzie nam mówić wyłącznie o **zapotrzebowaniach** w wierzchołkach. Zapotrzebowania w wierzchołkach, które nazwaliśmy **odbiorcami** będą dodatnie, zapotrzebowania w wierzchołkach **pośrednich** będą równe zeru, natomiast w wierzchołkach, które nazwaliśmy **magazynami** zapotrzebowania będą ujemne. W takiej konwencji funkcję $f : V \to \mathbb{R}$ będziemy nazywali **przepływem dopuszczalnym** jeżeli dla każdego wierzchołka $v \in V$ prawdziwa jest równość
$$\sum_{x:(x,v)\in A} f(x,v) - \sum_{x:(v,x)\in A} f(v,x) = b(v) \tag{9.2}$$
przy czym pamiętamy, że
$$b(v) \quad \begin{cases} < 0 & \text{jeśli } v \in M \\ = 0 & \text{jeśli } v \in P \\ > 0 & \text{jeśli } v \in O \end{cases}$$

- Będziemy zakładać, jakkolwiek by to było nieralne, że suma wszystkich zapotrzebowań jest równa sumie wszystkich zapasów[2], czyli
$$\sum_{v\in V} b(v) = 0 \tag{9.3}$$

- Funkcją celu, którą będziemy minimalizować jest koszt całkowity transportu, czyli
$$k(f) = \sum_{e\in A} c(e) \cdot f(e) \to \min \tag{9.4}$$

Z przepływem f można skojarzyć wektor $\mathbf{x}$ o $|E|$ współrzędnych $x_e = f(e)$, gdzie $e \in E$. Jeśli f jest przepływem dopuszczalnym (spełniającym (9.2) to $\mathbf{x}$ także będziemy nazywać przeplywem dopuszczalnym. Podobnie, koszty jednostkowe przepływów przez łuki $e \in E$ tworzą wektor $\mathbf{c}$ o współrzędnych c_e, zwany **wektorem kosztów jednostkowych**. Oznaczmy przez $\mathbf{A}$ macierz incydencji grafu

[2] To jest oczywiście założenie zupełnie nierealne, taka sytuacja nigdy nie zdarza się w rzeczywistości. Później jednak zobaczymy, że z sytuacją kiedy suma zapasów jest większa od sumy zapotrzebowań (to znaczy gdy $-\sum_{v\in M} b(v) \geq \sum_{v\in O} b(v)$) bardzo łatwo sobie poradzić. W przypadku przeciwnym, to znaczy gdy suma zapasów jest mniejsza od sumy zapotrzebowań, sytuacja jest beznadziejna: tego nie da się rozwiązać. Potrzebna rewolucja.

G i przez $\mathbf{b}$ wektor o współrzędnych $b_v = b(v)$. Nasz problem możemy teraz zapisać w postaci wektorowej

$$\begin{array}{ll} \text{zminimalizuj} & \mathbf{c}\cdot\mathbf{x} \\ \text{przy warunku:} & \mathbf{A}\mathbf{x} = \mathbf{b} \end{array} \tag{9.5}$$

Oczywiście i tym razem mamy do czynienia z problemem programowania liniowego.

Dla tak postawionego zadania pokażemy metodę, znacznie lepiej dopasowaną do tego zagadnienia niż poznany już simpleks, zwaną **simpleksem sieciowym**. Na metodę simpleksu siecioweg składają się dwa elementy: inicjalizacja i iteracje.

1. **Inicjalizacja.** Będziemy musieli znaleźć jakiekolwiek rozwiązanie dopuszczalne, to znaczy funkcję $f : E \to \mathbb{R}^+$ spełniającą układ równań (9.2) lub, co na jedno wychodzi, wektor $\mathbf{x}$ spełniający równanie $\mathbf{A}\mathbf{x} = \mathbf{b}$. Otrzymane na etapie inicjalizacji rozwiązanie będziemy w następnym etapie algorytmu polepszać aż do momentu, kiedy otrzymamy rozwiązanie optymalne.
2. **Iteracje.** Będziemy polepszać kolejno otrzymywane rozwiązania dopuszczalne aż do momentu kiedy rozważane rozwiązanie będzie optymalne.

W dalszym ciągu opisu metody wykażemy, że jeśli aktualnego rozwiązania dopuszczalnego nie da się polepszyć metodą simpleksu sieciowego, to rozwiązanie to jest optymalne oraz pokażemy jak rozwiązać problem, który nie spełnia nierealistycznego założenia (9.3).

9.3.1 Iteracje

Przypuśćmy, że mamy rozwiązanie dopuszczalne f. Co więcej, przypuśćmy, że to rozwiązanie jest generowane przez pewne drzewo zawierające wszystkie wierzchołki grafu G (to znaczy, że wszystkie łuki dla których przepływ dopuszczalny f jest niezerowy tworzą pewne drzewo zawarte w G[3]).
To drzewo nazywać będziemy krótko **drzewem dopuszczalnym**.
Wykażemy, że albo funkcja przepływu dopuszczalnego f jest optymalna, albo istnieje inne rozwiązanie dopuszczalne f' takie, że

- $k(f') \leq k(f)$ oraz
- f' także jest generowana przez pewne drzewo.

Oznaczmy przez T drzewo dopuszczalne generowane przez rozwiązanie dopuszczalne f.

[3]W rzeczywistości będzie mogło być tak, że podgraf grafu G generowany przez łuki na których dopuszczalna funkcja przepływu f jest niezerowa nie jest drzewem, lecz nie posiada cykli. To już nam wystarczy. łatwo bowiem sprawdzić, że wtedy można zawsze dodać takie łuki na których jest przepływ zerowy, dodanie ich do rozwiązania niewiele zmienia, a w rezultacie otrzymamy drzewo.

Krok 1. Uporządkujmy wierzchołki zbioru V wszystkich wierzchołków grafu G w taki sposób, by wierzchołek v_{i+1} był połączony ze zbiorem $\{v_1, ..., v_i\}$ dokładnie jednym łukiem e_i.

Kluczową rolę w metodzie odgrywają tak zwane **uczciwe ceny w wierzchołkach**.

Liczby $y_1, y_2, ..., y_n \in \mathbb{R}$ nazywamy **uczciwymi cenami** jeśli spełniają układ równań

$$y_i + c(v_i, v_j) = y_j \quad \text{dla } (v_i, v_j) \in A(T) \tag{9.6}$$

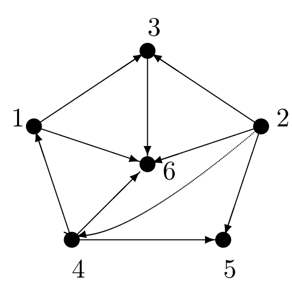
Nazwa *uczciwe ceny* jest w pełni uzasadniona. Wiadomo bowiem skąd się te ceny biorą: w wierzchołku v_j na cenę uczciwą składają się ceny dojścia do tego wierzchołka (przy czym pamiętamy, że używamy wyłącznie łuków naszego drzewa T).

Macierz układu $|V| - 1 = n - 1$ (rozmiar drzewa) równań (9.6) jest macierzą incydencji drzewa T, a więc jest rzędu $n - 1$ (na mocy wniosku 9.2.1). Skoro zaś zmiennych jest n układ ten jest niesprzeczny i jego zbiór rozwiązań zależy od jednego parametru, oznaczmy go przez d.

Co więcej, łatwo zauważyć, że jeśli $y_1, ..., y_n$ jest rozwiązaniem układu (9.6), wówczas także $y_1 + d, ..., y_n + d$ jest rozwiązaniem (9.6). Tak więc zawsze można znaleźć rozwiązanie (9.6), w którym wszystkie uczciwe ceny są nieujemne (co prawda ceny ujemne nie przeszkadzałyby nam w poprawnych rachunkach, ale byłoby to mało realistyczne[4]).

Od tego momentu prezentacji algorytmu będzie towarzyszył przykład.

Przykład 9.3.1 Niech graf G będzie taki, jak na rysunku 9.4. Numerki przy wierzchołkach oznaczają ich początkowe etykiety, których będziemy używać także w przyszłości).



Rysunek 9.4: G

Wierzchołki 1 i 2 są magazynami o zapotrzebowaniach 6 (każdy z nich),

[4]Uczciwość też musi mieć jakieś granice. Na przykład rozsądek.

odbiorcami są wierzchołki 5 i 6, zapotrzebowanie Te informacje wygodniej zapisć w postaci tabelek.

magazyn	zapas
1	6
2	6

odbiorca	zapotrzebowanie
5	4
6	8

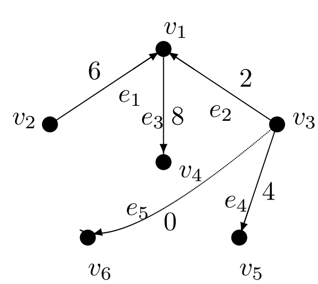
Równania układu (9.2) przyjmą postaci następujące:

$$x_{41} - x_{13} - x_{16} = -6$$
$$-x_{23} - x_{24} - x_{25} - x_{26} = -6$$
$$x_{36} - x_{13} - x_{23} = 0$$
$$x_{41} + x_{45} + x_{46} - x_{24} = 0$$
$$x_{25} + x_{45} = 4$$
$$x_{16} + x_{26} + x_{36} + x_{46} = 8$$

gdzie, dla wygody, przepływ z i do j oznaczono przez x_{ij} (zamiast $f(i,j)$). Koszty jednostkowe przewozu przez poszczególne łuki obrazuje poniższa tabelka.

łuk	$(1,3)$	$(1,6)$	$(2,3)$	$(2,4)$	$(2,5)$	$(2,6)$	$(3,6)$	$(4,1)$	$(4,5)$	$(4,6)$
koszt	4	3	3	1	1	3	5	2	1	1

Na rysunku następnym (9.5) przedstawiono drzewo i rozwiązanie dopuszczalne f. Liczby wpisane obok łuków są wielkościami przepływów na tych łukach natomiast obok wierzchołków są teraz wpisane etykiety wierzchołków spełniające warunek drugi wniosku 9.2.1. Obok łuków wpisano także ich etykiety nadane jak o tym mówi druga część wniosku 9.2.1. Przypomnijmy, że magazynami są wierzchołki oznaczone teraz przez v_2, v_3, oba o zapasie 6, natomiast odbiorcami v_4 i v_5 o zapotrzebowaniach, odpowiednio 8 i 4.



Rysunek 9.5: Pierwsze dopuszczalne drzewo i przepływ: $f(v_2,v_1) = 6, f(v_3,v_1) = 2, f(v_1,v_4) = 8, f(v_3,v_5) = 4, f(v_3,v_6) = 0$

Koszt przepływu wynosi

$$k(f) = 6 \cdot 4 + 2 \cdot 3 + 8 \cdot 5 + 4 \cdot 1 + 0 \cdot 1 = 74$$

Macierz incydencji jest postaci

$$M_1 = \begin{array}{c|ccccc} & e_1 & e_2 & e_3 & e_4 & e_5 \\ \hline v_1 & -1 & -1 & 1 & 0 & 0 \\ v_2 & 1 & 0 & 0 & 0 & 0 \\ v_3 & 0 & 1 & 0 & 1 & 1 \\ v_4 & 0 & 0 & -1 & 0 & 0 \\ v_5 & 0 & 0 & 0 & -1 & 0 \\ v_6 & 0 & 0 & 0 & 0 & -1 \end{array}$$

Układ równań definiujący uczciwe ceny jest następujący (z lewej strony zaznaczono łuki indeksujące kolejne równania).

$$\begin{array}{ll} e_1: & y_2 + 4 = y_1 \\ e_2: & y_3 + 3 = y_1 \\ e_3: & y_1 + 5 = y_4 \\ e_4: & y_3 + 1 = y_5 \\ e_5: & y_3 + 1 = y_6 \end{array}$$

Rozwiązaniem tego układu (na dodatek takim, w którym wszystkie uczciwe ceny są nieujemne) jest na przykład

$$y_1 = 4,\ y_2 = 0,\ y_3 = 1,\ y_4 = 9,\ y_5 = 2,\ y_6 = 2$$

Do przykładu wrócimy, teraz następny krok algorytmu.

Krok 2. Zauważmy teraz, że jeśli w sieci istnieje łuk $(v_k, v_l) \in A(G)$ taki, który nie należy do drzewa dopuszczalnego T oraz

$$y_k + c(v_k, v_l) < y_l \tag{9.7}$$

wówczas możemy zaopatrzyć wierzchołek v_l w przesyłane dobra po niższej cenie[5] wykorzystując łuk (v_k, v_l) (ten łuk nie jest w rozwiązaniu dopuszczalnym f wykorzystywany, bo nie należy do drzewa dopuszczalnego).
łuk (v_k, v_l) **nazywamy łukiem wchodzącym**.

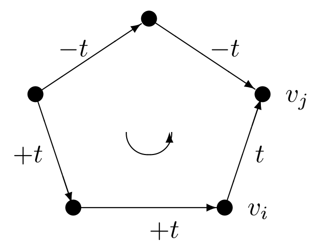
Krok 3. Do grafu T dodajemy łuk (v_k, v_l) (graf otrzymany z grafu T przez dodanie łuku (v_k, v_l) oznaczamu przez $T + (v_k, v_l)$).
Na mocy wniosku 9.2.1 w grafie grafu $T + (v_k, v_l)$ zawarty jest dokładnie jeden cykl C.

łuk (v_k, v_l) wyznacza pewną orientację cyklu C (por. rys. 9.6). Wszystkie łuki cyklu C są albo **zgodne** albo **niezgodne** z tą orientacją.
Modyfikujemy przepływ f

dodając t na łukach zgodnych oraz

odejmując t na łukach niezgodnych (na pozostałych łukach, tych, które nie nałeżą do cyklu C przepływ nie ulega zmianie).

[5]Formalnie udowodnimy ten fakt w twierdzeniach 9.3.1 i 9.3.2.



Rysunek 9.6: Krok 2: dodajemy t na łukach zgodnych cyklu C i odejmujemy na niezgodnych

Zdefiniujmy nową funkcję $\bar{f}_t$ wzorem

$$\bar{f}_t = \begin{cases} f(v_i, v_j) & \text{jeśli } (v_i, v_j) \text{ nie jest łukiem cyklu } C, \\ f(v_i, v_j) + t & \text{jeśli } (v_i, v_j) \text{ jest łukiem} \\ & \text{zgodnym cyklu } C, \\ f(v_i, v_j) - t & \text{jeśli } (v_i, v_j) \text{ jest łukiem} \\ & \text{niezgodnym cyklu } C \end{cases} \tag{9.8}$$

Twierdzenie 9.3.1 *Jeśli f jest przepływem dopuszczalnym generowanym przez drzewo T, $y_1, ..., y_n$ są uczciwymi cenami wyznaczonymi z układu równań (9.6), istnieje łuk (v_k, v_l) spełniający nierówność (9.7), C jest cyklem grafu $G+(v_k, v_l)$ oraz*

$$t_0 = \min\{f(v_i, v_j) : (v_i, v_j) \text{ jest łukiem niezgodnym cyklu } C\}$$

wówczas funkcja $\bar{f}_{t_0}$ jest także przepływem dopuszczalnym. Co więcej, $\bar{f}_{t_0}$ jest także przepływem generowanym przez pewne drzewo T'.

Twierdzenie 9.3.1 w zasadzie nie wymaga dowodu. Rzeczywiście, przepływ netto przez wierzchołki nie zmienia się wraz ze zmianą funkcji f na $\bar{f}_t$ dokładnie tak jak to miało miejsce dla ścieżek powiększających w dowodzie twierdzenia 8.3.4, tak więc $\bar{f}_t$ jest przepływem dopuszczalnym, skoro f jest przepływem dopuszczalnym. Co więcej, skoro na pewnym łuku niezgodnym cyklu[6] C wartość przepływu f jest równa t_0 (minimum wartości przepływu na łukach niezgodnych cyklu C jest przyjęta dla pewnego łuku), funkcja przepływu $\bar{f}_0$ przyjmuje na tym łuku wartość zero. Usuwając z grafu $T+(v_i, v_j)$ ten właśnie łuk[7] niszczymy w tym grafie jedyny cykl i otrzymujemy drzewo T'.

Pozostaje nam teraz udowodnić, że drzewo T' otrzymane w wyniku iteracji jest lepsze od poprzedniego drzewa T w tym sensie, że koszt przekływu $\bar{f}_{t_0}$ generowanego przez drzewo T' jest mniejsze niż koszt przepływu f.

[6]Wykażemy później, w twierdzeniu 9.3.2, że łuk niezgodny w cyklu C istnieje

[7]Oczywiście może być więcej niż jeden łuk, na którym wartość przepływu f jest równa t_0. Wtedy wybieramy jeden z nich, dowolnie.

Udowodnimy nieco więcej mianowicie nie tylko, że gdy istnieje łuk (v_k, v_l) nienależący do drzewa T taki, że

$$y_k + c(v_k, v_l) < y_l \tag{9.9}$$

to istnieje przepływ lepszy (a przynajmniej nie gorszy[8]) od f od przepływu f, natomiast jeśli łuk (v_k, v_l) spełniający (9.17), nie istnieje to rozwiązanie dopuszczalne f jest optymalne.

Twierdzenie 9.3.2 *1. Jeśli nie istnieje łuk (v_k, v_l) nie należący do drzewa T dla którego spełniona jest nierówność (9.17) wówczas rozwiązanie dopuszczalne f jest optymalne.*

2. Jeżeli (v_k, v_l) jest łukiem nie należącym do drzewa dopuszczalnego T takim, że spełniona jest nierówność (9.17), wtedy dla rozwiązania dopuszczalnego $\bar{f}_{t_0}$ spełniona jest nierówność $k(\bar{f}_{t_0}) \leq k(f)$.

Dow"od. Pamiętamy, że $y_1, ..., y_n$ są rozwiązaniami układu równań (9.6), które nazwaliśmy uczciwymi cenami w wierzchołkach $v_1, ..., v_n$ ponumerowanymi dla drzewa T tak, jak we wniosku 9.2.1. Przez $\mathbf{y}$ oznaczmy wektor o współrzędnych $y_1, ..., y$.
Niech $\mathbf{x}$ będzie rozwiązaniem dopuszczalnym generowanym przez T a więc takim, że

$$\mathbf{Ax} = \mathbf{b}$$

oraz $x_{ij} = 0$ dla wszystkich i, j takich, że $(v_i, v_j) \notin E(T)$ i niech $\bar{\mathbf{x}}$ będzie dowolnym rozwiązaniem dopuszczalnym czyli takim wektorem, że

$$\mathbf{A\bar{x}} = \mathbf{b}$$

Zdefiniujmy wektor

$$\bar{\mathbf{c}} = \mathbf{c} - \mathbf{yA} \tag{9.10}$$

Mamy wtedy

$$\bar{\mathbf{c}}(v_i, v_j) = \mathbf{c}(v_i, v_j) + y_i - y_j$$

i

$$\mathbf{c}(v_i, v_j) = 0 \quad \text{gdy } (v_i, v_j) \in E(T)$$

oraz

$$x_{ij} = 0 \quad \text{dla } i, j \text{ takich, że } (v_i, v_j) \notin E(T)$$

Stąd

$$\bar{\mathbf{c}}\mathbf{x} = 0$$

Z równości (9.10) definiującej $\bar{\mathbf{c}}$ wynika, że $\mathbf{c} = \bar{\mathbf{c}} + \mathbf{yA}$ i w konsekwencji

$$\mathbf{c\bar{x}} = \bar{\mathbf{c}}\bar{\mathbf{x}} + \mathbf{yA\bar{x}}$$

[8]Problemem, kiedy $k(\bar{f}_{t_0}) = k(f)$ zajmiemy się później. A problem jest poważny bo dla simpleksu sieciowego, podobnie jak dla simpleksu, może w takiej sytuacji wystąpić zjawisko cykliczności.

a wobec tego

$$\mathbf{c}\bar{\mathbf{x}} = \bar{\mathbf{c}}\bar{\mathbf{x}} + \mathbf{y}\mathbf{b} \tag{9.11}$$

Równość (9.11) jest spełniona dla każdego rozwiązania dopuszczalnego $\bar{\mathbf{x}}$, a więc także dla $\bar{\mathbf{x}} = \mathbf{x}$, czyli

$$\mathbf{c}\mathbf{x} = \bar{\mathbf{c}}\mathbf{x} + \mathbf{y}\mathbf{b} = \mathbf{y}\mathbf{b}$$

Korzystając z ostatniej równości i z (9.11) otrzymujemy

$$\mathbf{c}\bar{\mathbf{x}} = \bar{\mathbf{c}}\bar{\mathbf{x}} + \mathbf{c}\mathbf{x} \tag{9.12}$$

(przypomnijmy: równość (9.12) spełniona jest dla każdego rozwiązania dopuszczalnego $\mathbf{x}$ generowanego przez drzewo T i dla dowolnego rozwiązania dopuszczalnego $\mathbf{x}$).

1. Jeśli nie istnieje łuk $(v_k, v_l) \notin E(T)$ taki, że $y_k + c(v_k, v_l) < y_l$ to oczywiście mamy

   $$\bar{c}(v_k, v_l) = c(v_k, v_l) + y_k - y_l \geq 0$$

   dla wszystkich $(v_k, v_l) \in E(G)$. Skoro na dodatek $\bar{\mathbf{x}} \geq 0$, mamy $\bar{\mathbf{c}}\bar{\mathbf{x}} \geq 0$ i wobec tego (biorąc pod uwagę (9.12))

   $$\mathbf{c}\bar{\mathbf{x}} \geq \mathbf{c}\mathbf{x} \tag{9.13}$$

   dla dowolnego rozwiązania dopuszczalnego $\bar{\mathbf{x}}$, co oznacza, że rozwiązanie $\mathbf{x}$ jest optymalne. Dowód pierwszej części twierdzenia jest zakończony.

2. Przypuśćmy teraz, że dla pewnego łuku $v_k, v_l)$ mamy

   $$y_k + c(v_k, v_l) < y_l$$

   Niech $\bar{\mathbf{x}}$ będzie teraz rozwiązaniem dopuszczalnym wyznaczonym przez funkcję przepływu $\bar{f}_t$ (zdefiniowaną wzorem (9.8). łuk (v_k, v_l) jest jedynym dla którego równocześnie zachodzą związki

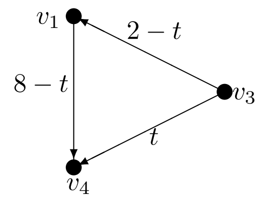
   $$\bar{\mathbf{c}}(v_k, v_l) \neq 0 \quad \text{i} \quad \bar{x}_{v_k v_l} \neq 0$$

   Stąd $\bar{\mathbf{c}}\bar{\mathbf{x}} = \bar{c}(v_k, v_l)\bar{x}_{v_k v_l} = \bar{c}(v_k, v_l) \cdot t$.
   Co więcej, $\bar{c}(v_k, v_l) < 0$ a więc dla $t \geq 0$ mamy $k(\bar{f}_t) = \mathbf{c}\bar{\mathbf{x}} \leq \mathbf{c}\mathbf{x}$. Ta ostatnia nierówność kończy dowód twierdzenia. ■

Przykład 9.3.1 c.d. W naszym przykładzie (str. 123) mamy

$$y_3 + c(v_3, v_4) = 1 + 3 = 4 < 9 = y_4$$

Po dodaniu do drzewa dopuszczalnego łuku (v_3, v_4) otrzymujemy cykl (trójkąt) przedstawiony na rysunku 9.7.

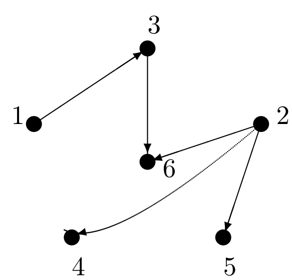


Rysunek 9.7: Krok 2: Cykl w $T + (v_3, v_4)$

Modyfikujemy funkcję przepływu otrzymując nowe wartości dla łuków znajdujących się na cyklu dodając lub odejmując największe t możliwe, to znaczy nie dające przepływów ujemnych na łukach niezgodnych (w naszym przypadku $t_0 = 2$).:

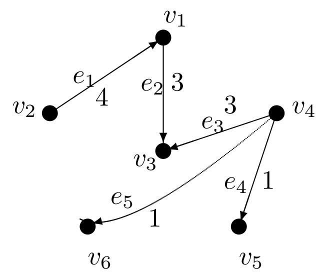
$$\begin{array}{rcl} \bar{f}(v_3, v_4) & = & 2 \\ \bar{f}(v_3, v_4) & = & 0 \\ \bar{f}(v_1, v_4) & = & 6 \end{array}$$

Wartości przepływu na pozostałych łukach pozostawiamy bez zmian.
Po powrocie do wyjściowych oznaczeń dla wierzchołków (z rysunku 9.4) otrzymujemy przepływ $f(1,3) = 6, f(3,6) = 6, f(2,6) = 2, f(2,4) = 0,\ f(2,5) = 4$.
Koszt nowego przepływu wynosi $k = 6 \cdot 4 + 0 \cdot 1 + 4 \cdot 1 + 2 \cdot 3 + 6 \cdot 5 = 70$.
Teraz tworzymy nowe drzewo dopuszczalne $T' = T - (v_3, v_1) + (v_3, v_4)$. To nowootrzymane drzewo jest teraz na następnym rysunku 9.8, w którym wierzchołki są oznaczone tak, jak na rysunku 9.4.



Rysunek 9.8: Drugie drzewo dopuszczalne

Etykiety wierzchołków i łuków spełniające warunki wniosku 9.2.1 zaznaczono na rysunku 9.10



Rysunek 9.9: Drugie drzewo dopuszczalne z nowymi etykietami wierzchołków i łuków

Układ równań dla uczciwych cen w wierzchołkach przyjmie postać

$$\begin{array}{rrcl} e_1: & y_2+4 & = & y_1 \\ e_2: & y_1+3 & = & y_3 \\ e_3: & y_4+3 & = & y_3 \\ e_4: & y_4+1 & = & y_5 \\ e_5: & y_2+1 & = & y_6 \end{array}$$

Jednym z rozwiązań tego układu (o cenach nieujemnych) jest $y_1 = 4, y_2 = 0, y_3 = 7, y_4 = 4, y_5 = 5, y_6 = 1$.
Teraz szukamy takich k, l, żeby $y_k + c(v_k, v_l) < y_l$. Można przyjąć $k = 6, l = 3$ io wtedy otrzymamy $y_6 + c(v_6, v_3) = 2 < 7 = y_3$.
W grafie $T+(v_6, v_3)$ (gdzie T jest teraz drzewem z rysunku 9.10 jedynym cyklem jest cykl utworzony przez łuki $(v_6, v_3), (v_4, v_3)$ i (v_4, v_6). Modyfikujemy funkcję przepływu otrzymując

$$\begin{array}{rcccl} \bar{f}(v_6, v_3) & = & \bar{f}(4,6) & = & 2 \\ \bar{f}(v_4, v_3) & = & \bar{f}(2,6) & = & 0 \\ \bar{f}(v_4, v_6) & = & \bar{f}(2,4) & = & 2 \\ \bar{f}(v_2, v_1) & = & \bar{f}(1,3) & = & 6 \\ \bar{f}(v_1, v_3) & = & \bar{f}(3,6) & = & 6 \\ \bar{f}(v_4, v_5) & = & \bar{f}(2,5) & = & 4 \end{array}$$

Wartości nowego przepływu na pozostałych łukach pozostają nie zmienione (i są równe zeru). Można sprawdzić, że koszt nowego rozwiązania dopuszczalnego wynosi 62.
Nowym drzewem jest powstałe z poprzedniego przez dodanie łuku $(1, 6)$ i usunięcie $(3, 6)$.
Powtarzamy kolejny raz iterację by w kolejnej otrzymać rozwiązanie o wartościach przepływów

$$f(1,3) = 0, f(1,6) = 6, f(4,6) = 2, f(2,4) = 2, f(2,5) = 4$$

i koszcie całkowitym równym 26. Wykorzystujemy teraz twierdzenie 9.3.2 by wykazać, że to rozwiązanie jest już optymalne.

9.3.2 Inicjalizacja

Znalezienie jakiegokolwiek rozwiązania dopuszczalnego generowanego przez pewne drzewo nie jest problemem banalnym. Można go jednak rozwiÄ... zać wykorzystując samą metodę simpleksu sieciowego.
Powiedzmy, że dany jest graf G i w nim określony problem transportowy tak, jak na stronie 120. Chcemy znaleźć rozwiązanie dopuszczalne, na dodatek generowane przez pewne drzewo. Problem znalezienia takiego rozwiązania (czyli inicjalizacji algorytmu) można rozwiązać następująco.

Wpierw zdefiniujmy nowy graf G'.
Wybierzmy w zbiorze wierzchołków grafu G wierzchołek w. Wierzchołek ten będzie od tego momentu do końca działania algorytmu wierzchołkiem wyróżnionym (taki wierzchołek w grafie nazywamy **korzeniem**).Zbiorem wierzchołków grafu G' jest zbiór V wierzchołków grafu G. Zbiorem łuków G' jest zbiór wszystkich łuków grafu G oraz zbiór łuków postaci

- (m, w) jeśli $m \in M$ i $(m, w) \notin E(G)$,
- (p, w) jeśli $p \in P$ i $(p, w) \notin E(G)$,
- (w, o) jeśli $o \in O$ i $(w, o) \notin E(G)$.

Każdy z łuków postaci (w, x) lub (x, w), który należy do $E(G')$ a nie należy do $E(G)$, a więc jest nowym łukiem dodanym do grafu na potrzebę inicjalizacji algorytmu nazywamy **łukiem sztucznym** (dla odróżnienia łuki grafu G będziemy nazywać **łukami oryginalnymi**).
W tak skonstruowanym grafie będziemy poszukiwali funkcji $f : E(G') \ni ij \to \mathbb{R}$ spełniającej warunki

$$\sum_{i:ik \in E(G')} f(ik) - \sum_{i:ki \in E(G')} f(ki) = b_k$$

dla dowolnego $k \in V$. W zapisie wektorowym, chodzić będzie o wyznaczenie wektora $\mathbf{x}$ spełniającego

$$\sum_{i:ik \in E(G')} x_{ik} - \sum_{i:ki \in E(G')} x_{ki} = b_k$$

przy czym spełniony jest warunek zgodności

$$\sum_k b_k = 0$$

Ceny jednostkowe przepływu przez krawędzie definiujemy wzorem

$$p_{ij} = \begin{cases} 1 & \text{gdy } ij \text{ jest łukiem sztucznym,} \\ 0 & \text{gdy } ij \text{ jest łukiem grafu } G \end{cases}$$

Funkcją kosztów jest więc $\sum_{ij\in E(G')} p_{ij}x_{ij}$.

W tak zdefiniowanej sieci znajdujemy przepływ o minimalnym koszcie, przy pomocy simpleksu sieciowego i przyjmując jako pierwsze drzewo dopuszczalne T drzewo o zbiorze łuków

$$E(T) = \{xw : x \in M \cup P\} \cup \{wo : o \in O\}$$

przy czym przepływ generowany przez T jest równy

- $-b_k$ na każdym łuku postaci kw, gdy $k \in M$,
- 0 na łukach postaci kw, gdy $k \in P$,
- b_k gdy $k \in O$.

Ten nowy problem transportowy nazwiemy **pomocniczym** (z problemem oryginalnym Przypuśćmy, że rozwiązanie optymalne tak sformułowanego problemu transportowego jest generowane przez pewne drzewo T'. Moż się zdarzyć jeden z trzech przypadków.

1. W T' jeden (lub więcej) z łuków jest sztuczny i przepływ przez ten łuk jest niezerowy. **Wówczas rozwiązanie dopuszczalne problemu oryginalnego nie istnieje**.
   Rzeczywiście, gdyby istniało, to miałoby koszt zerowy (w problemie pomocniczym), tymczasem rozwiązanie generowane przez T' ma oczywiście koszt dodatni, bowiem zawiera łuk sztuczny z przepływem niezerowym).
2. T' nie zawiera sztucznych łuków. **Wtedy T' jest drzewem dopuszczalnym problemu oryginalnego, a rozwiązanie optymalne problemu pomocniczego jest rozwiązaniem dopuszczalnym problemu oryginalnego.** To rozwiązanie możemy więc przyjąć jako pierwsze w serii iteracji dla tego problemu.
3. T' zawiera sztuczne łuki, ale na każdym z nich jest zerowy przepływ (w rozwiązaniu optymalnym problemu pomocniczego).
   Ten przypadek jest nieco bardziej skomplikowany i musimy się mu przyjrzeć nieco dłużej.

9.3.3 Dekompozycja problemu transportowego

Zauważmy wpierw, że dla dowolnego rozwiązania $\mathbf{x} = (x_{ij})$ i podzbioru $S \subset V(G)$ zachodzi następująca tożsamość

$$\sum_{i\notin S, j\in S} x_{ij} - \sum_{i\in S, j\notin S} x_{ij} = \sum_{k\in S} b_k \tag{9.14}$$

Formalny dowód tożsamości (9.14) jest prosty: po jej lewej i prawej stronie są te równania które w postaci macierzowej $\mathbf{Ax} = \mathbf{b}$ odnoszą się do wierzchołków zbioru S. Interpretacją ekonomiczną tożsamości (**??**) jet następująca:

$$\text{import} - \text{eksport dóbr} = \text{zapotrzebowanie zbioru } S$$

Oznaczmy przez R dopełnienie zbioru S do $V(G)$, czyli $R = V(G) - S$. Jeśli zbiory R i S są niepuste i w grafie G spełnione są dwa warunki:

(W1) Nie ma łuków ze zbioru R do zbioru S oraz

(W2) $\sum_{k \in S} b_k = 0$

to zbiór S niczego nie importuje, swoje potrzeby zaspokaja własnymi zapasami. W związku zaś z warunkiem (W2) zbiór S niczego nie może eksportować.
W takiej sytuacji mówimy, że problem ma **podział** lub **dekompozycję**.
Oczywiście jeśli istnieje dekompozycja problemu transportowego wówczas oznacza to, że problem rozpada się na dwa mniejsze problemy, które można rozwiązywać niezależnie.

Lemat 9.3.3 *W przypadku 3, a więc gdy w optymalnym rozwiązaniu problemu pomocniczego optymalne drzewo T' zawiera sztuczne łuki, lecz dla każdego takiego łuku przepływ jest zerowy, istnieje dekompozycja problemu oryginalnego.*

Dow"od. Niech $y_1, ..., y_n$ będą uczciwymi cenami w wierzchołkach wyznaczonymi dla optymalnego drzewa T' problemu pomocniczego zaś $\mathbf{x}^*$ rozwiązaniem optymalnym tego problemu.
Niech uv będzie dowolnie wybranym łukiem sztucznym tego drzewa (wtedy, zgodnie z założeniem lematu, $x^*_{uv} = 0$). Zdefiniujmy teraz następujące dwa zbiory wierzchołków.

- $R = \{k : y_k \leq y_u\}$
- $S = \{k : y_k > y_u\}$

Zbiory R i S dają pewien podział zbioru $V(G)$, bowiem

- $R \cap S = \emptyset$
- Mamy także $u \in R$ oraz $y_v = y_u + p_{uv} = y_u + 1$, a więc $v \in S$ skąd wynika że i zbiór S jest niepusty.
- Prawdą jest także, że $R \cup S = V(G)$.

Pozostaje więc do wykazania, że $\sum_{k \in S} b_k = 0$.
Dla łuków oryginalnych ij nie możemy mieć $i \in R, j \in S$ bo wtedy mielibyśmy

$$y_i + p_{ij} = y_j > y_u \geq y_i$$

czyli $y_i > y_i$, a więc sprzeczność. Nie ma więc oryginalnych łuków skierowanych ze zbioru R do zbioru S (czyli o początku w R i końcu w S).
Pamiętamy (na mocy (9.14)), że

$$\sum_{i \in R, j \in S} x^*_{ij} - \sum_{i \in S, j \in R} x^*_{ij} = \sum_{k \in S} b_k \qquad (9.15)$$

Jeśli teraz wiemy, na mocy założenia, że dla rozwiązania $\mathbf{x}^*$ problemu pomocniczego $x^*_{ij} = 0$ dla wszystkich łuków sztucznych, zaś łuków oryginalnych ij z $i \in R, j \in S$ w ogóle nie ma, z równości (9.15) wynika, że

$$- \sum_{i \in S, j \in R} x^*_{ij} = \sum_{k \in S} b_k \tag{9.16}$$

żaden z łuków oryginalnych ij takich, że $i \in S, j \in R$ nie należy do drzewa T' (w przeciwnym przypadku mielibyśmy $y_j = y_i + p_{ij} = y_i$ ($p_{ij} = 0$ dla łuków oryginalnych) podczas gdy $y_i > y_u \geq y_j$, skoro $i \in S, j \in R$). Stą zaś i z (refjuz blisko) wynika, że

$$\sum_{k \in S} b_k = 0$$

a więc spełnione są warunki dekompozycji (W1) i (W2), co było do okazania.■

9.3.4 Modyfikacja uczciwych cen w wierzchołkach.

Przypomnijmy, że dla danego drzewa T o zbiorze wierzchołków V, $|V| = n$, *uczciwymi cenami w wierzchołkach* są liczby $y_i, ..., y_n$ spełniające układ równań

$$y_i + c_{ij} = y_j \text{ dla } ij \in E(T) \tag{9.17}$$

Układ równań (9.17) jest rzędu $n-1$ i dowolne jego dwa rozwiązania różnią się o stałą. Przekonamy się, że przechodząc w procesie iteracji simpleksu sieciowego od drzewa dopuszczalnego T do drzewa $T + e - f$ nowe uczciwe ceny w wierzchołkach (ceny uczciwe dla drzewa $T + e - f$) można łatwo obliczyć bez rozwiązywania układu równań typu (9.17) dla nowego drzewa.

Dla łuku wchodzącego $e = uv$ i wychodzącego f oznaczmy przez T_u składową lasu $T - f$ zawierającą wierzchołek u zaś przez T_v składową zawierającą wierzchołek v.

Twierdzenie 9.3.4 *Liczby $\bar{y}_1, ..., \bar{y}_n$ postaci*

$$\bar{y}_k = \begin{cases} y_k & \text{gdy } k \in T_u \\ y_k + \bar{c}_e & \text{gdy } k \in T_v \end{cases} \tag{9.18}$$

gdzie $\bar{c}_e = c_e + y_u - y_v$ spełniają równanie (9.17) dla drzewa $T + e - f$ (to znaczy spełniają układ równań $\bar{y}_i + c_{ij} = \bar{y}_j$ dla $ij \in E(T + e - f)$).

Dow"od. Dla dowolnego $ij \in E(T_u) \cup E(T_v)$ bardzo łatwo sprawdzić, że $\bar{y}_i + c_{ij} = \bar{y}_j$, stąd jedyne co pozostaje, to sprawdzenie równości odpowiadającej łukowi $e = uv$. Wtedy

$$\bar{y}_u + c_{uv} = y_u + c_{uv} = y_u + (\bar{c}_{uv} - y_u + y_v) = y_v + \bar{c}_{uv} = \bar{y}_v$$

. ■

9.3.5 Procedura unikania zapętlenia

Jak już zostało zasygnalizowane wcześniej, w pewnych sytuacjach simpleks sieciowy może się zapętlić. Może tak się zdarzyć, gdy dla kolejnych drzew dopuszczalnych otrzymanych w procesie iteracji $T_i, T_{i+1}, ..., T_{i+l}$ wartości funkcji celu są równe. Wtedy mogłoby się zdarzyć, że $T_i = T_{i+l}$, czyli mamy zapętlenie algorytmu. W niniejszym podrozdziale pokażemy procedurę Cunninghama pozwalającą zapętlenia uniknąć[9].

Dla danego drzewa T, łuku $uv \in E(T)$ i korzenia $w \in V(T)$ mówimy, że

- łuk uv **jest skierowany do korzenia** w jeśli $w \in V(T_v)$,
- uv **jest skierowany od korzenia** w jeśli $w \in V(T_u)$.

Korzeń w to dowolnie ustalony na cały proces optymalizacji wierzchołek grafu G, ten sam, który wybraliśmy już podczas inicjalizacji algorytmu.

Krok iteracji polegający na przejściu od drzewa dopuszczalnego T_i do drzewa dopuszczalnego T_{i+1} nazywamy **zdegenerowanym** jeżeli wartość funkcji celu dla obu tych drzew jest taka sama.

Twierdzenie 9.3.5 *Jeśli dla każdego zdegenerowanego kroku iteracji simpleksu sieciowego prowadzącego od drzewa T do drzewa $T + e - f$ łuk wchodzący e jest skierowany od korzenia w drzewa $T + e - f$, to algorytm simpleksu sieciowego nie zapętla się.*

Dow"od. Dla danego dopuszczalnego drzewa T zdefiniujmy dwie funkcje.

$$g(T) = \mathbf{c} \cdot \mathbf{x}$$

gdzie $\mathbf{x}$ jest rozwiązaniem zdeterminowanym przez drzewo T[10], oraz

$$h(T) = \sum_{k=1}^{n} (y_k - y_w)$$

Ze względu na wspomniane wyżej własności uczciwych cen w wierzchołkach wiemy, że nie są one wyznaczone jednoznacznie, jednak każde dwa rozwiązania różnią się o stałą. Stąd wartość funkcji h dla drzewa T jest już wyznaczona jednoznacznie.

Niech $T_1, T_2, ...$ będzie ciągiem kolejnych drzew dopuszczalnych otrzymanych w wyniku iteracji i przypuśćmy, że $g(T_i) = g(T_{i+1})$, czyli przejście od drzewa T_i do T_{i+1} jest zdegenerowane. Powiedzmy, że $T_{i+1} = T_i + e - f$. Z założenia łuk e jest skierowany od korzenia w w drzewie T_{i+1}, to znaczy $w \in (T_{i+1})_u$. Mamy wtedy

$$\sum_{k=1}^{n} (\bar{y}_l - \bar{y}_w) = \sum_{k=1}^{n} (y_k - y_w) + \bar{c}_e \cdot |(T_{i+1})_v|$$

[9] Jak niemal we wszystkich zgadnieniach przedstawionych tutaj, opiera/lem się na książce Chvátala [4]. W przypadku unikania zapętlenia wykorzystałem także pracę magisterską Elżbiety Pelczar [18]

[10] Funkcja g to oczywiście koszt rozwiązania dopuszczalnego generowanego przez drzewo T.

gdzie $\bar{c}_e = c_e + y_u - y_v$. Ponieważ $\bar{c}_e < 0$ jako, że e jest łukiem wchodzącym, prawdziwa jest nierówność

$$h(T_i) > h(T_{i+1})$$

Stąd już wynika twierdzenie. Jeśli bowiem ciąg kroków o drzewach $T_i, T_{i+1}, ...$ jest zdegenerowany, to znaczy $g(T_i) = g(T_{i+1})$ dla pewnego i, to wówczas $h(T_i) > h(T_{i+1})$ a stąd już wynika, że żadne dwa drzewa w ciągu $T_1, T_2, ...$ nie powtarzają się[11]. ■

Pozostaje nam teraz wykazanie, że dla dowolnego przejścia zdegenerowanego algorytmu istnieje łuk uv skierowany od korzenia.

Definicja 9.3.1 *Drzewo T nazywamy* **silnie dopuszczalnym** *jeśli każdy łuk $uv \in E(T)$, dla którego $x_{uv} = 0$ jest skierowany od korzenia (czyli $w \in V(T_u)$).*

Zauważmy, że jeżeli wszystkie drzewa $T_1, T_2, ...$ otrzymane w kolejnych iteracjach są silnie dopuszczalne, to spełnione są założenia twierdzenia 9.3.5 i w konsekwencji algorytm simpleksu sieciowego nie zapętla się.

Drzewo początkowe $T = T_1$ utworzone podczas inicjalizacji algorytmu jest drzewem silnie dopuszczalnym bowiem wszystkie łuki postaci tego drzewa to:

- łuki postaci mw gdzie $m \in M$. Dla tych łuków $x_{mw} = -b_m \neq 0$.
- Pozostałe łuki, które są postaci wx, gdzie $x \in P \cup O$, a więc **skierowane od korzenia** w. To wśród tych łuków będą łuki wchodzące o przepływie zero.

Wystarczy więc udowodnić, że da się wybierać łuki wchodzące i wychodzące tak, by z drzewa silnie dopuszczalnego otrzymywać inne silne dopuszczalne drzewo lub dekompozycję problemu na mniejsze, dla których da się przeprowadzić podobną manipulację.

Definicja 9.3.2 *łuk $uv \in E(T)$ nazywamy* **złym** *jeśli*

- *jest skierowany do korzenia oraz*
- $x_{uv} = 0$.

Inaczej mówiąc, zły łuk to taki, którego istnienie powoduje, że drzewo T nie jest silnie dopuszczalne.

Przypuśćmy, że uv jest złym łukiem drzewa T. Oznaczmy przez $R = V(T_v)$ i przez $S = V(T_u)$. Oczywiście nie ma łuków w drzewie T, o jednym końcu w R a drugim w S. Z wykorzystywanej już wcześniej tożsamości

$$\sum_{i \in R, j \in S} x_{ij} - \sum_{i \in S, j \in R} x_{ij} = \sum_{k \in S} b_k$$

[11] Warto sobie zdać sprawę dlaczego rzeczywiście tak jest. Jeśli w ciągu (T_i) nie ma przejść zdegenerowanych sprawa jest oczywista. Jeśli dla pewnych i, l mamy $g(T_i) = g(T_{i+1}) = ... = g(T_{i+l})$ wówczas $h(T_i) > h(T_{i+1}) > ... > h(T_{i+l})$, drzewa T_j, dla $1 \leq j \leq i + l$ są między sobą różne. Ważne jednak jest i to, że funkcja g jest słabo malejąca, dzięki czemu $T_p \neq T_q$ dla dowolnych $p \neq q$.

wynika więc, że

$$\sum_{k \in S} b_k = 0$$

(bowiem na wszystkich łukach ij takich, że $i \in R, j \in S$ lub $i \in S, j \in R$ mamy $x_{ij} = 0$). Jeśli więc nie istnieje w grafie G żaden łuk ij: $i \in R, j \in S$ to spełnione są warunki (W1) i (W2) i problem dekomponuje się na dwa mniejsze.

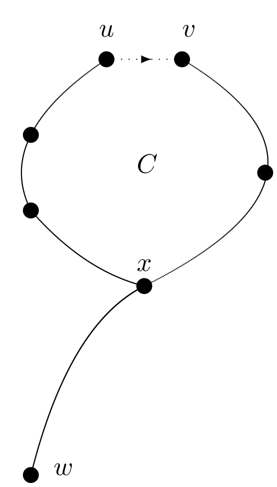
Jeśli zaś istnieje łuk $ij : i \in R, j \in S$ to drzewo T możemy zastąpić drzewem $T + ij - uv$, które ma mniej złych łuków niż drzewo T (rzeczywiście, $T + ij - uv$ powstaje przez zastąpienia złego łuku uv łukiem ij, który jest skierowany od w, a więc nie jest łukiem złym).

Aby mieć pewność, że w każdym przejściu otrzymujemy z drzewa silnie dopuszczalnego inne drzewo silnie dopuszczalne (dzięki czemu algorytm nie będzie się zapętlał) przyjmiemy specjalną zasadę wyboru łuków wychodzących, natomiast zachowamy dowolność wyboru łuków wchodzących.

Definicja 9.3.3 *Niech T będzie drzewem, podgrafem grafu G, $e = uv \in V(G) - V(T)$. Pierwszy wierzchołek, w którym spotykają się ścieżki*

- *z u do korzenia*
  ze ścieżką
- *z v do korzenia*

nazywamy **spinaczem**.



Rysunek 9.10: Cykl C ze spinaczem x

Niech C będzie jedynym cyklem w grafie $T + uv$ $uv = e$.
Zasada Cunninghama

- Jako łuk wychodzący przyjmij pierwszego kandydata napotkanego na cyklu C podczas przechodzenia po cyklu zgodnie z orientacją wyznaczoną przez łuk e, począwszy od spinacza.

Zauważmy, że wszystkie łuki - kandydaci na łuki wychodzące na cyklu C dalej będą mieć orientację **od** korzenia. Prawdziwe jest więc twierdzenie następujące.

Twierdzenie 9.3.6 *Jeśli wybór łuku wychodzącego jest dokonywany według zasady Cunninghama, wówczas simpleks sieciowy nie zapętla się.*

9.4 Zapotrzebowanie mniejsze od zasobów

Nierealistycznego w praktyce założenia, że

$$\sum_{v \in M} b(v) = \sum_{v \in O} b(v)$$

czyli, że suma zapasów magazynów jest dokładnie równa zapotrzebowań odbiorców można się pozbyć bardzo łatwo. Oczywiście, tak jak wspomniane zostało wcześniej, zakładać będziemy spełnienie warunku

$$\sum_{v \in M} b(v) \geq \sum_{v \in O} b(v)$$

Dodajmy do naszej sieci, dla której zakładamy spełnienie warunków

$$\sum_{v:(v,x) \in E} f(v,x) - \sum_{x:(x,v) \in E} f(x,v) = \begin{cases} -b(x) & \text{jeśli } x \in M \\ 0 & \text{jeśli } x \in P \\ b(x) & \text{jeśli } x \in O \end{cases}$$

dodatkowy wierzchołek, powiedzmy w, oraz wszystkie łuki postaci (m, w), gdzie $m \in M$. Koszty tych wszystkich nowych łuków będą równe zeru. Inaczej mówiąc, do sieci dodajemy dodatkowego, fikcyjnego odbiorcę i fikcyjne łuki łączące go z magazynami, który będzie miał zapotrzebowanie równe sumie wszystkich nadwyżek, czyli

$$b(w) = \sum_{v \in M} b(v) - \sum_{v \in O} b(v)$$

Optymalne rozwiązanie w tak otrzymanej sieci jest równe optymalnemu rozwiązaniu w sieci wyjściowej (wystarczy pominąć przepływy na fikcyjnych łukach).

Bibliografia

[1] ALEKSIEWICZ A., Analiza funkcjonalna, PWN, Warszawa 1969.

[2] BLAND R.G., New finite pivoting rule for the simplex method, Mathematics of Operations Research 2 (1977) 103-107.

[3] CHARON I., GERMA A. I HUDRY O., Méthodes d'optimisation combinatoire, Masson, Paris 1996.

[4] CHVÁTAL V., Linear Programming, W.H. Freeman and Comp., N. Y. 1984.

[5] DANTZIG G.B, Upper bounds, secondary constrains and block triangularity in linear programming, Econometrica 23, str. 174-183, 1955.

[6] DANTZIG G.B. I FULKERSON D.R., On the max-flow min-cut theorem of networks, w: Linear Inequalities and Related Systems, Princeton, New Jersey 1956 (215-221).

[7] ELIAS P., FEINSTEIN A. I SHANNON C.A., A note on the maximum flow through a network, IRE Transactions on Information Theory IT 2 (1956) 117-119.

[8] FORD JR. L.R. I FULKERSON D.R., Maximal flow through network, Canadian Journal of Mathematics 8 (1956) 399-404.

[9] FORD JR. L.R. I FULKERSON D.R., Przepływy w sieciach, PWN, Warszawa 1969.

[10] GASS S.I., Programowanie liniowe, PWN, Warszawa 1973.

[11] GRABOWSKI W., Programowanie matematyczne, PWE, Warszawa 1980.

[12] HALL P., On representatives sets, J. London Math. Soc. 10 (1935) 26-30.

[13] HARRISON, JR. J.O. Programming and Operations Research, w: Research for Menagement, wyd. J.F. McCloskey i F.N. Trefethen, vol. 1, str. 217 – 337, Baltimore, Md., 1954.

[14] KARLOFF H., Linear Programming, Birghäuser, Boston 1991.

[15] König D., Theorie der Endlichen und Unendlichen Graphen, New York 1950.

[16] Ligęza E., Wodociągi dawnego Krakowa - do połowy XVII wieku, Towarzystwo Miłośników Historii i Zabytków Krakowa, red. J. Mitkowski, Kraków 1971.

[17] Nykowski I., Programowanie liniowe, PWE, Warszawa 1984.

[18] Pelczar E., Metoda simpleksu sieciowego, praca magisterska, Wydział Matematyki Stosowanej AGH, Kraków, 2013.

[19] Simmonard M., Programowanie liniowe, PWN, Warszawa 1969.

[20] Schrijvers A., Theory of linear and integer programming, Wiley 1998.

[21] Sysło M.M., Deo N. i J.S. Kowalik, Algorytmy optymalizacji dyskretnej, PWN, Warszawa 1999.

[22] Wagner H.M., Badania operacyjne, PWE, Warszawa 1980.

[23] Zawistowska Z., żywienie dietetyczne, PZWL, Warszawa 1973.

Indeks