Dualność LP a problem maksymalnego przepływu

Tekst może nieco odstraszać swoją długością, ale starałem się nie dokonywać zbyt szybkich przejść. Pytania i uwagi proszę śmiało przysyłać do mnie na maila. Dziękuję Radosławowi Burnemu za pomocne uwagi i znalezienie błędu we wcześniejszym rozumowaniu.

Dla danego programu liniowego:

$$\begin{array}{ll} \text{zminimalizuj} & c_1x_1 + c_2x_2 + \cdots + c_nx_n (= c^Tx) \\ \text{z zachowaniem warunków} & Ax \geq b \\ & x \geq 0 \end{array}$$

napiszemy najpierw program dualny.

Intuicja jest następująca. Szukamy jak najlepszego ograniczenia dolnego na wartość funkcji celu. Każda z nierówności określonych przez $Ax \geq b$ to pewne ograniczenie dolne na wektor x. Chcemy zsumować te ograniczenia, nadając każdemu pewną nieujemną wagę y_i, tj. bierzemy $y^TAx \geq y^Tb$. Dodatkowo, dbamy o to, by żadnej zmiennej x_i nie wziąć z większym współczynnikiem niż w funkcji celu. W wyrażeniu y^TAx zmienna x_i będzie wzięta ze współczynnikiem równym i-tej współrzędnej wektora powstałego z wymnożenia y^TA. Wektor ten jest poziomy, dlatego transponujemy go: $(y^TA)^T = A^Ty$. Zatem, naszym celem jest takie dobranie y, by $A^Ty \leq c$. Przy takim dobraniu y, dostaniemy ograniczenie dolne na wartość funkcji celu równe b^Ty.

Aby znaleźć *jak najlepsze* ograniczenie dolne zapiszmy program:

$$\begin{array}{ll} \text{zmaksymalizuj} & b^Ty \\ \text{z zachowaniem warunków} & A^Ty \leq c \\ & y \geq 0 \end{array}$$

Powyższy program nazywamy programem dualnym do pierwotnego (zwanego również prymalnym). Zauważmy, że proces zamiany programu na dualny jest odwracalny, tj. program dualny do dualnego jest taki sam jak program prymalny. Twierdzenie o dualności mówi, że ograniczenie dolne jest optymalne. Dokładniej mówiąc, dla programu prymalnego i dualnego zawsze zachodzi jeden z poniższych warunków:

- obydwa programy są ograniczone i ich optymalne rozwiązania są sobie równe,
- jeden z programów ma nieograniczone rozwiązanie, a drugi nie ma rozwiązań dopuszczalnych,
- obydwa programy nie mają rozwiązań dopuszczalnych.

Zastosujemy powyższą technikę do problemu maksymalnego przepływu, danego programem:

$$\begin{array}{lll} \text{zmaksymalizuj} & x_{ts} & \\ \text{z zachowaniem warunków} & \sum_{w \in V}(x_{vw} - x_{wv}) = 0 & \text{dla każdego } v \in V \\ & x_{vw} \leq c(v,w) & \text{dla każdego } v, w \in V \\ & x_{wv} \geq 0 \text{ dla wszystkich } v, w \in V & \end{array}$$

W programie tym zmienna x_{ab} odpowiada wartości przepływu z a do b. Przyjmujemy tu definicję przepływu bez skośnej symetryczności.

Naszym priorytetem nie jest minimalizowanie czasu działania, dlatego dla uproszczenia zakładamy, że mamy do czynienia z pełnym grafem, w którym niektóre krawędzie mają przepustowość 0. Przypominam, że powyższy program liniowy stosuje nieco inne podejście do przepływu. Szukamy w grafie takiej cyrkulacji, by przepływ od ujścia do źródła (x_{ts}) był jak największy. Przepływ po tej krawędzi musi być równy sumie przepływów na krawędziach wychodzących z s, zatem tym sposobem znajdziemy maksymalny przepływ.

Pierwszy warunek w programie liniowym mówi, że do każdego wierzchołka musi wpływać tyle samo, ile z niego wypływa. Możemy tę równość zamienić na nierówność, nie zmieniając wyniku [1].

[1] Da się całość rozumowania przeprowadzić bez tej operacji, ale byłoby ono nieco bardziej skomplikowane.

zmaksymalizuj	x_{ts}	
z zachowaniem warunków	$\sum_{w \in V}(x_{vw} - x_{wv}) \leq 0$	dla każdego $v \in V$
	$x_{vw} \leq c(v,w)$	dla każdego $v, w \in V$
	$x_{wv} \geq 0$ dla wszystkich $v, w \in V$	

Teraz może się zdarzyć, że do wierzchołka wpłynie więcej niż z niego wypłynie. Intuicja jest taka, że w niektórych wierzchołkach woda może nam wyciekać. Zatem taka zmiana na pewno nie powiększy wartości maksymalnego przepływu i oczywiście stara wartość będzie nadal osiągalna (bo w szczególności dopuszczamy brak przecieków).

Można też przeprowadzić bardziej formalny dowód, na przykład pokazać indukcyjnie dla coraz większych zbiorów S (spełniających $s \in S, t \notin S$), że ze zbioru wypływa nie więcej niż do niego wpływa. Na końcu okaże się nawet, że w każdym rozwiązaniu dopuszczalnym programu liniowego i tak w pierwszym warunku otrzymamy równości, ale dla nas nie jest to potrzebne.

Zastanówmy się teraz, jak wygląda program dualny, tzn. jak nasz zbiór nierówności przekłada się na zapis $Ax \leq b$.

Każdy wiersz macierzy A odpowiada jednemu ograniczeniu. Pierwsze $|V|$ wierszy to (rozluźnione) warunki zachowania przepływu, w kolejnych wierszach mamy $|E|$ warunków na przepustowość.

Kolumny macierzy A odpowiadają zmiennym lub, inaczej mówiąc, krawędziom grafu. Operujemy na pełnym grafie skierowanym, ma on $|V|(|V|-1)$ krawędzi. Załóżmy, że pierwsza kolumna macierzy odpowiada krawędzi z t do s (zmiennej x_{ts}). Funkcja celu to x_{ts}, więc wektor c składa się z jedynki na pierwszej współrzędnej i samych zer na reszcie współrzędnych.

Teraz chcemy napisać zbiór ograniczeń dany przez $A^T y \geq c$, czyli musimy przeczytać macierz A kolumnami. Dla każdej kolumny (czyli każdej krawędzi) dostaniemy jedno ograniczenie. Zmienne y odpowiadają z kolei wierszom pierwotnej macierzy A. Oznacza to, że mamy najpierw po jednej zmiennej y_v dla $|V|$ pierwszych wierszy macierzy, a następnie po jednej zmiennej y_{ab} dla kolejnych wierszy (tj. dla każdej krawędzi).

Jeśli teraz spojrzymy na dowolną kolumnę w macierzy A (niech odpowiada ona krawędzi ab), to znajdziemy tam:

- 1 w wierszu odpowiadającym wierzchołkowi a,
- -1 w wierszu dla wierzchołka b,
- 1 w wierszu dla krawędzi ab.

Stąd, dla $ab \neq ts$ otrzymamy:

$$y_a - y_b + y_{ab} \geq 0$$

Zaś dla ts:

$$y_t - y_s + y_{ts} \geq 1$$

Ostatecznie nasz program liniowy będzie wyglądać następująco:

zminimalizuj	$\sum_{uv} c(u,v) y_{uv}$	
z zachowaniem warunków	$y_t - y_s + y_{ts} \geq 1$	
	$y_a - y_b + y_{ab} \geq 0$	dla $ab \neq ts$
	$y \geq 0$	

W funkcji celu zmienna y_{ts} jest brana z wagą $c(t,s)$. W oryginalnym programie nie mieliśmy żadnego ograniczenia na przepływ z t do s, więc możemy przyjąć, że $c(t,s)$ było jakąś dowolnie dużą liczbą. Ograniczmy się teraz do przypadków, gdy $y_{ts} = 0$. Jeśli tylko uda nam się znaleźć rozwiązanie, to wiemy, że będzie ono lepsze od jakiegokolwiek, które nie spełnia tego warunku.

Zapiszmy program nieco inaczej:

zminimalizuj	$\sum_{uv} c(u,v) y_{uv}$	
z zachowaniem warunków	$y_t - y_s \geq 1$	
	$y_{ab} \geq y_b - y_a$	dla $ab \neq ts$
	$y \geq 0$	

Zauważmy, że na każde y_{ab} mamy dokładnie dwa ograniczenia. Jasne jest, że chcemy by każde y_{ab} było jak najmniejsze. Zatem możemy równoważnie napisać:

$$
\begin{array}{lll}
\text{zminimalizuj} & \sum_{uv} c(u,v) y_{uv} & \\
\text{z zachowaniem warunków} & y_t - y_s \geq 1 & \\
 & y_{ab} = \max(0, y_b - y_a) & \text{dla } ab \neq ts \\
 & y \geq 0 &
\end{array}
$$

Przyjrzyjmy się temu programowi. Musimy znaleźć takie wagi wierzchołków, że waga t jest co najmniej o 1 większa od wagi s i żeby zminimalizować pewne różnice wag. Kluczowe spostrzeżenie jest takie, że dla dowolnego rozwiązania możemy zmniejszyć wszystkie wagi większe od $y_s + 1$ do $y_s + 1$ oraz powiększyć wszystkie wagi mniejsze od y_s do y_s. To nie pogorszy wartości funkcji celu, bo jeśli y_b było mniejsze lub równe od y_a (wtedy możemy położyć $y_{ab} = 0$), to nadal tak będzie, zaś jeśli y_b było większe od y_a (wówczas $y_{ab} = y_b - y_a$), to różnica $y_b - y_a$ może tylko zmaleć.

Dostaliśmy rozwiązanie, w którym wagi są z zakresu od y_s do $y_s + 1$. Macierz programu dualnego jest transpozycją macierzy A, wyznaczającej program prymalny. Na ćwiczeniach pokazaliśmy, że macierz ta jest całkowicie unimodularna, więc algorytmy rozwiązywania programowania liniowego znajdą nam całkowitoliczbowe rozwiązanie.

Pozostaje podsumować wszystko, co wiemy. Nasz program liniowy przydziela wagi y_s i $y_s + 1$ do wierzchołków. Wiemy ponadto, że $y_t = y_s + 1$. Oznacza to, że program szuka podziału zbioru wierzchołków na dwa zbiory, z których jeden zawiera źródło, a drugi ujście, innymi słowy — przekroju w grafie. Jeśli spojrzymy na postać funkcji celu, widzimy, że znajduje on przekrój, którego przepustowość jest minimalna.

Z twierdzenia o dualności LP, wiemy, że rozwiązania optymalne programu prymalnego i dualnego mają równe wartości funkcji celu. W ten sposób udowodniliśmy twierdzenie o minimalnym przepływie i maksymalnym przekroju.