Goldbergův algoritmus

Prof. RNDr. Luděk Kučera, DrSc.

12. ledna 2004

1 Vlna

Je-li dána funkce t definovaná pro hrany sítě, pak pro vrchol v je

$$excess_t(v) = \sum_{u \to v} t(u,v) - \sum_{v \to u} t(v,u).$$

$excess_t(v)$ nazýváme přebytek vrcholu v. Řekneme, že t je *vlna* (angl. preflow), pokud $excess_t(v) \geq 0$ pro každý vrchol sítě s výjimkou zdroje (ten je jediný vrchol, ze kterého může něco vytékat, aniž by do něho přitékalo) a kromě toho $0 \leq t(h) \leq c(h)$ pro kazdou hranu h (kde $c(h)$ je kapacita hrany h).

Pokud $c(x,y)$ znamená kapacitu hrany (x,y) v síti a $t(x,y)$ tok (nebo vlna) touto hranou, pak $r(x,y)$, *rezerva* hrany (x,y), je

$$r(x,y) = c(x,y) - t(x,y) + t(y,x).$$

Poznamenejme, že předpokládáme, že je-li (v,w) hrana sítě, pak i (w,v) je hrana sítě (pokud by tam nebyla, přidáme ji tam s nulovou kapacitou). Poznamenáváme, že rezerva může být kladná i pro hrany s nulovou kapacitou.

2 Goldbergův algoritmus

Každý vrchol v má výšku, kterou budeme označovat $H(v)$. Jako N označíme počet vrcholů, jako M počet hran sítě.

Má-li vrchol v přebytek alespoň $\Delta > 0$ a hrana (v,w) má rezervu alespoň Δ, pak převést přebytek z v do w hranou (v,w) znamená:

- je-li tok opačnou hranu (w,v) nulový, pak zvýšit tok hranou (v,w) o Δ;
- je-li tok opačnou hranou (w,v) větši nebo rovný Δ, pak snížit tok hranou (w,v) o Δ;
- je-li tok opačnou hranou (w,v) kladný, ale menší než Δ, pak snížit tok hranou (w,v) na nulu a současně zvýšit tok hranou (v,w) o $\Delta - t(w,v)$, kde $t(w,v)$ byl původní tok hranou (w,v).

Je zřejmé, že tato operace sníží přebytek vrcholu v o Δ, zvýší přebytek vrcholu w o Δ, sníží rezervu hrany (v, w) o Δ a zvýší rezervu opačné hrany (w, v) o Δ.

Algoritmus

[**Inicializace**]
výška zdroje je N, výška každého dalšího vrcholu je 0;
hranami vycházejícími ze zdroje teče tok rovný kapacitě hrany;
ostatními hranami neteče nic (nulový tok)
while existuje vrchol s kladným přebytkem, různý od spotřebiče **do begin**
    v = libovolný vrchol s kladným přebytkem (ne spotřebič);
    if existuje hrana (v, w), která má kladnou rezervu a $H(v) > H(w)$
    then begin
        (v, w) = libovolná hrana vycházející z v, která má kladnou rezervu
            a splňuje $H(v) > H(w)$;
        Δ = minimum z přebytku vrcholu v a rezervy hrany (v, w);
        převést přebytek velikosti Δ z v do w;
        end
    else $H(v) := H(v) + 1$;
    end.

Budeme rozlišovat tři způsoby provedení těla **while** cyklu:

zvednutí Provedení příkazu $H(v) := H(v) + 1$

nasycené převedení přebytku Převedení přebytku velikosti Δ z v do w, jestliže Δ je rovno původní rezervě hrany (v, w) (to anuluje rezervu hrany (v, w))

nenasycené převedení přebytku Převedení přebytku velikosti Δ z v do w, jestliže Δ je menší než původní rezerva hrany (v, w) (to anuluje přebytek vrcholu v)

Lemma 1 *V žadném okamžiku od provedení inicializace neexistuje hrana $h = (v, w)$ taková, že $H(v) > H(w) + 1$, a rezerva hrany h je nenulová.*

Důkaz: Po provedení inicializace taková hrana neexistuje, jediné hrany, ktere vedou "z kopce" začínají ve zdroji, ale těmi teče plný tok a mají tedy nulovou rezervu.

Dokážeme, že provedení **while** cyklu takovou hranu nemůže vytvořit. Zvednutí vrcholu by takovou hranu (v, w) mohlo vytvořit, kdyby hrana měla rezervu, bylo by $H(v) = H(w) + 1$ a zvedán by byl vrchol v. Pak by ovšem v musel mít přebytek a za takové situace je zakázáno vrchol v zvedat, algoritmus musí zvýšit tok touto hranou.

Převádění přebytku takovou hranu (v, w) mohlo vytvořit, kdyby měla nulovou rezervu a ta se zvýšila na nenulovou hodnotu. To se může stát jen při převádění přebytku opačnou hranou (w, v) a v takovém případě je $H(w) = H(v) + 1$, tedy hrana (v, w) vede "do kopce". ♣

V tento okamžik jsme již snadno schopni dokázat, že pokud algoritmus skončí, nalezne největší tok (dokázat konečnost výpočtu dá větší práci):

Věta 1 *Pokud Goldbergův algoritmus skončí výpočet, nalezne největší tok.*

Důkaz: **while** cyklus v algoritmu se ukončí pouze když všechny vrcholy sítě jiné než zdroj a spotřebič mají nulový přebytek, což znamená, že konstruovaná vlna je tok. K důkazu optimality stačí dokázat, že libovolná (prostá) cesta ze zdroje do spotřebiče obsahuje alespoň jednu hranu která má nulovou rezervu.

Nechť je tedy dána nějaká prostá cesta ze zdroje do spotřebiče. Tato cesta začíná ve výšce N (výška zdroje, která se během výpočtu nemění) a končí ve výšce 0 (výška spotřebiče, která se během výpočtu také nemění), navíc tato cesta má nejvýše $N-1$ hran. Proto na ní musí být alespoň jedna hrana, překonávající výškový rozdíl alespoň 2 a taková hrana musí podle předchozího lemmatu mít nulovou rezervu. ♣

Právě provedený důkaz dává odpověď na otázku, proč se na začatku dává zdroj do výšky N. Kdyby byl v nižší výšce, mohlo by se stát, že by se zastavil v okamžiku, kdy je nalezený tok ještě neoptimální. Naopak kdyby se zdroj dal na začátku do větší výšky, algoritmus by fungoval a předcházející věta by platila, ale, jak bude vidět dále, by nejspíše pracoval déle, protože vrcholy, které svůj přebytek nemohou dopravit až do zdroje, musí vystoupat do větší výšky než zdroj a pak svůj přebytek "odlít zpět" do zdroje.

Lemma 2 *Vždy od provedení inicializace, jestliže vrchol v má kladný přebytek, pak z něj vede orientovaná cesta zpět do zdroje taková, že každá hrana této cesty má kladnou rezervu.*

Důkaz: Nechť v je vrchol s kladným přebytkem. Označme jako A množinu všech vrcholů takových, že do nich vede z v orientovaná cesta, tvořená hranami s kladnou rezervou.

Nechť (x, y) je hrana taková, že $x \notin A$, $y \in A$. Kdyby tok hranou (x, y) byl kladný, pak by opačná hrana (y, x) měla kladnou rezervu, a jelikož existuje cesta z hran kladné rezervy z v do y, přidáním této hrany bychom dostali cestu z hran kladné rezervy z v do x, takže by bylo $x \in A$, spor.

Hranami vstupujícími do A z vrcholů mimo A tedy teče nulový tok. Z toho se snadno odvodí, že součet přebytků vrcholů množiny A je menší nebo rovný nule (propočítejte si to!). Jelikož v množině A je ale jeden vrchol s kladným přebytkem (vrchol v), musí v ní být také vrchol se záporným přebytkem a jediný takovy vrchol je zdroj. Tedy zdroj leží v A a podle definice A tedy do něj vede cesta z hran kladne rezervy. ♣

Lemma 3 *Výška žádného vrcholu není nikdy větší než $2N$.*

Důkaz: Nechť je vrchol v zvedán do výšky větší než $2N$. Pak je ve výšce alespoň $2N$ a má kladný přebytek. Pak podle předchozího lemmatu vede z v do zdroje cesta složená z hran s kladnou rezervou. Tato cesta ale obsahuje nejvýše $N-1$ hran a překonává spád alespoň N (zdroj je ve výšce N) a tedy na ní leží hrana (x, y) taková, že $H(x) > H(y) + 1$, což je podle prvního lemmatu spor. ♣

Lemma 4 *Počet všech zvednutí v průběhu celého výpočtu není větší než $2N^2$.*

Důkaz: Plyne okamžitě z předchozího lemmatu, protože každý vrchol se může zvednout nejvýše $2N$-krát. ♣

Lemma 5 *Počet nasycených převedení přebytku za celý výpočet je nejvýše* NM.

Důkaz: Nechť $h = (v, w)$ je hrana. Součet výšek jejích koncových vrcholů v a w je mezi 0 a $4N$. Dojde-li k nasycenému převedení přebytku hranou h, její rezerva poklesne na nulu. V tom okamžiku také platí, že $H(v) = H(w) + 1$. Aby došlo znovu k nasycenému převádění přebytku hranou h, musí se nejprve její rezerva zvýšit na nenulovou hodnotu. K tomu může dojít jedině v případě, že dojde k převedení přebytku opačnou hranou (w, v). Před tím se ale musí zvýšit výška vrcholu w na hodnotu o 1 vyšší než je výška vrcholu v, takže výška w se musí zvětšit o alespoň 2. Jakmile se rezerva h zvýší, je výška vrcholu v menší než výška vrcholu w, ale aby mohlo dojít k dalšímu nasycenému převedení přebytku hranou h, musí se výška v zvýšit na hodnotu vyšší než je výška w, tedy alespoň o 2.

Mezi dvěma po sobě jdoucími nasycenými převáděními přebytku hranou $h = (v, w)$ se tedy $H(v) + H(w)$ zvýší o alespoň 4. Proto nasycených převedení přebytku hranou h nemůže být více než N. ♣

Lemma 6 *Počet nenasycených převedení přebytku za celý výpočet je nejvýše* $2N^2 + 2N^2M$.

Důkaz: Označme jako Σ součet výšek všech vrcholů, různých od zdroje a spotřebiče, které mají přebytek. Po inicializaci je $\Sigma = 0$ (jen zdroj má nenulovou výšku). Na konci výpočtu je Σ také rovno nule (žádný vrchol různý od zdroje a spotřebiče nemá přebytek).

Zvýšení vrcholu zvýší Σ o 1. Nasycené převedení přebytku hranou (v, w) může zvýšit Σ nejvýše o $H(w)$ (pokud w původně přebytek neměl a vrcholu v nenulový přebytek zůstane), tedy ne více než o $2N$. Celkový součet přídavků k Σ od zvedání vrcholů a nasycených převádění přebytku je $2N^2 + 2N \cdot NM$.

Nenasycené převedení přebytku hranou (v, w) *sníží* Σ nejméně o 1, neboť výšky se nemění, přebytky se mohou měnit jen u v a w a to tak, že přebytek v, původně kladný, poklesne na nulu, takže z Σ vypadne sčítanec $H(v)$, což může být částečně kompenzováno přírustkem o $H(w)$, pokud w původně neměl přebytek (po přidání jej má). Jelikož ale při převedení přebytku hranou (v, w) je $H(w) = H(v) - 1$, kompenzace není úplná a Σ alespoň o 1 poklesne. Počet nenasycených převedení přebytku ale nemůže být větší než celkový součet přídavků k Σ od zvedání vrcholů a nasycených převádění přebytku. ♣

Uvažujme nyní variantu algoritmu, která vždy vybere jako vrchol v ten vrchol s přebytkem, který má největší výsku. Pro tento algoritmus je možno odhad počtu nenasycených zvyšování podstatně zlepšit:

Lemma 7 *Počet nenasycených převedení přebytku za celý výpočet algoritmu "vždy nejvyší vrchol s přebytkem" je nejvýše* $8N^2\sqrt{M}$.

Důkaz: Za fázi označme souvislý úsek výpočtu algoritmu, během kterého je stálá maximální výska vrcholu s přebytkem, tedy

$$\mathcal{H} = \max\{H(v) \quad : \quad v \text{ má kladný přebytek}\}.$$

Po celou dobu fáze se jako vrchol v v algoritmu vybírají jen vrcholy výšky $\mathcal{H}$. Fáze končí, jakmile

- buď došlo ke zvednutí vrcholu ve vrstvě výšky $\mathcal{H}$,
- nebo se přebytky všech vrcholů výšky $\mathcal{H}$ snížily na nulu.

Je tedy zřejmé, že v průběhu fáze nemohlo dojít k více nenasyceným převedením přebytku, než kolik bylo na jejím začátku vrcholů s přebytkem a s výškou $\mathcal{H}$.

Číslo $\mathcal{H}$ roste jen při zvednutí vrcholu a to o 1, na začátku je nula a stále je nezáporné. Počet klesnutí $\mathcal{H}$ je tedy nejvýše tolik, kolik je zvýšení $\mathcal{H}$. Jelikož počet zvýšení $\mathcal{H}$ je nejvýše rovný počtu zvednutí vrcholu, tedy maximálně $2N^2$, je celkový počet změn $\mathcal{H}$ a tedy i počet fází nejvýše $4N^2$.

Zvolme (zatím libovolně) číslo K. Pro vrchol v označme jako $\eta(v)$ počet vrcholů, které mají výšku nejvýše $H(v)$ a položme

$$\Psi = \sum_{excess(v)>0} \frac{\eta(v)}{K}.$$

Nazvěme fázi *lacinou*, dojde-li v ní k nejvýše K nenasyceným převedení přebytku a *drahou* jinak.

Laciných fází není více než fází všech, tedy maximálně $4N^2$ a tedy nenasycených převedení přebytku v laciných fázích je nejvýše $4N^2K$.

Zvednutí vrcholu v zvýší $\eta(v)$, ale nikoli více než o N, tedy $\eta(v)/K$ se zvýší o nejvýše N/K. Zvednutím vrcholu v mohou hodnoty $\eta(w)$ pro ostatní vrcholy w jen klesnout, nikdy ne vzrůst. Zvednutí vrcholu tedy nemůže zvýšit Ψ o více než o N/K.

Při změnách toku se η vrcholů nemění. Jakékoli, tedy např. nasycené, převedení přebytku hranou (v, w) může z Ψ ubrat sčítance $\eta(v)/K$ a může přidat sčítanec $\eta(w)/K \leq N/K$, takže hodnota Ψ se nemůže zvýšit o víc než N/K.

Nenasycené převedení přebytku hranou (v, w) sníží přebytek v na nulu, což z Ψ odstraní člen $\eta(v)$, do Ψ může ještě přibýt člen $\eta(w)$, takže změna Ψ je nejvýše $(\eta(w) - \eta(v))/K$. Vrchol v přitom leží ve výšce $\mathcal{H}$ a vrchol w ve výšce $\mathcal{H} - 1$, tedy $\eta(v) - \eta(w)$ je roven počtu vrcholů ve výšce $\mathcal{H}$. V drahé fázi je tento rozdíl větší než K, což vede k tomu, že v drahé fází se Ψ *sníží* alespoň o 1.

Celkový součet přírustků Ψ je tedy nejvýše

$$(2N^2 + NM)\frac{N}{K},$$

a jelikož Ψ je stále kladné a na počátku je nejvýše N^2/K, je počet všech nenasycených převedení přebytku v drahých fázích nejvýše

$$\frac{N^2}{K} + (2N^2 + NM)\frac{N}{K} \leq 4N^2M/K,$$

kde jsme použili nerovnost $1 \leq N \leq M$.

Dohromady tedy počet všech nenasycených převedení přebytku v laciných a drahých fázích dohromady je nejvýše

$$4N^2K + 4N^2\frac{M}{K} \leq 4N^2\left(K + \frac{M}{K}\right) = 8N^2\sqrt{M} \qquad \text{pro } K = \sqrt{M}.$$

♣