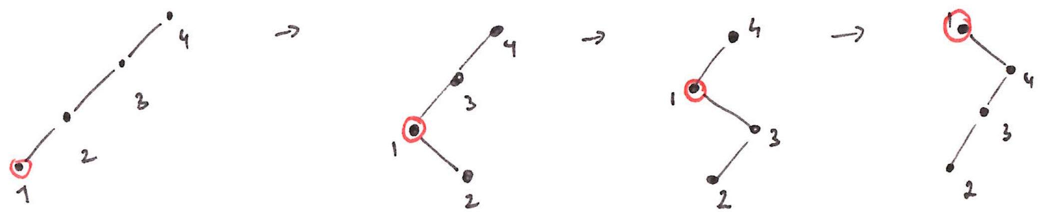
Splay stromy (varianta BVS)

<u>základní idea</u>: po přístupu k uzlu x jej pomocí rotací umístíme do kořene tak, že množina uzlů na cestě zhruba o polovinu sníží hloubku.

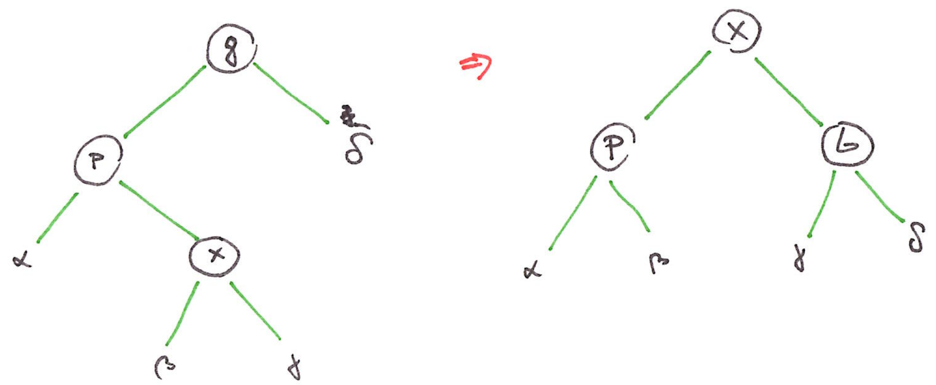
<u>obyčejné rotace nefungují:</u>



Rotace pořád provádíme od navštíveného uzlu do kořene, ale jsou složitější.

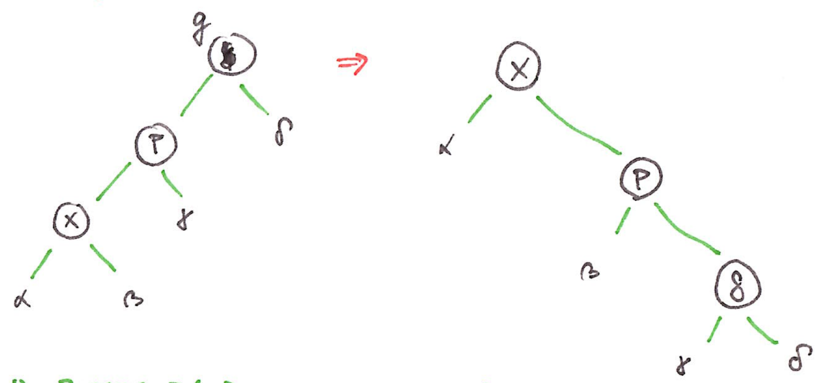
ozn: akt. uzel … x, p … x.parent, g … p.parent

zig-zag:



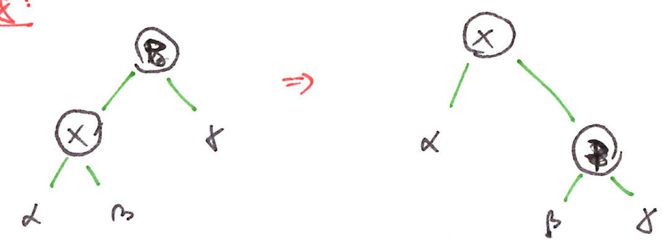
je to: 1) ROTATE-L(p), 2) ROTATE-R(g)

zig-zig:



1) ROTATE-R(g) 2) ROTATE-R(p)

zig:



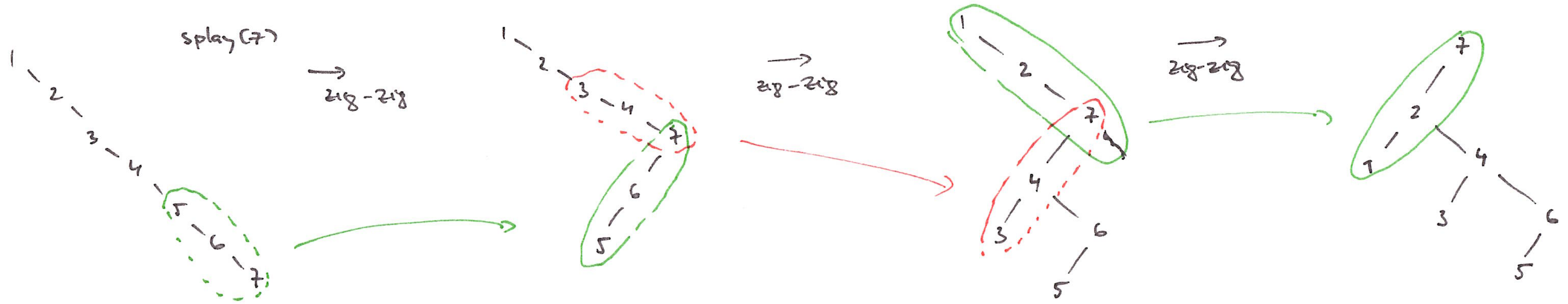
<u>Pozn:</u>

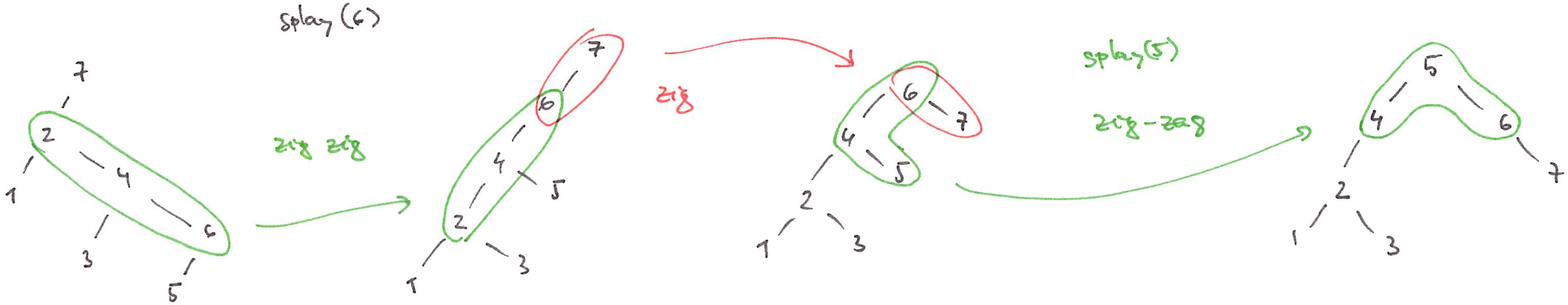
→ ke každé z rotací existují symetrické protějšky

→ zig rotace se dělá pouze pokud p.parent je nil. (tj. p je kořen).

→ operaci, která takto „probublá" x do kořene označíme splay(x)

Př:


splay(7)
zig-zig
zig-zig
zig-zig


splay(6)
zig zig
zig
splay(5)
zig-zag

Amortizovaná analýza operace splay

značení: $T(v)$ … podstrom s kořenem v (jako množina uzlů)

$m(v)$ ….. počet vrcholů ve $T(v)$

$r(v)$ ….. $\lg m(v)$ (rank vrcholu v)

potenciál stromu je roven $\sum\limits_{\substack{\text{uzel } v \\ \text{ve stromu}}} r(v)$

(tj. potenciál je součet ranků všech uzlů)

→ Idea: čím „košatější" strom v důsledku operace splay máme, tím je menší potenciál. Touto změnou platíme za provedení operace splay.

→ Složitost splay budeme vyjadřovat počtem rotací (skutečná složitost je vzhledem k tomu lineární)

Věta: Amortizovaná cena operace splay je nejvýše $3\cdot(r'(x) - r(x)) + 1$, kde $r(x)$ je rank vrcholu x před provedením operace a $r'(x)$ po něm.

Důkaz: označme $r_1(x), r_2(x), \dots, r_t(x)$ ~~ceny~~ rank uzlu x po 1., 2., … t-tém kroku (tj. provedení jedné ze zig, zig-zig, zig-zag) rotace, $r_0(x) = r(x)$.

V následujících lemmatech ukážeme, že amort. cena i-té operace zig-zig a zig-zag ~~vždy~~ je $3r_i(x) - 3r_{i-1}(x)$,

a i-té zig je $3r_i(x) - 3r_{i-1}(x)$,

ale zig se stane max ~~jenom~~ jednou (na konci)

tj.
$$\hat{c} \le \left(\sum_{i=1}^{t} (3r_i(x) - 3r_{i-1}(x))\right) + 1$$
$$\le 3r_t(x) - 3r_0(x) + 1$$

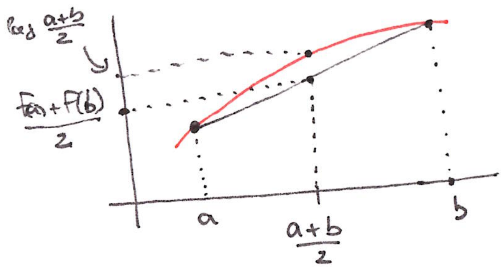
∎

Pozn:

→ protože $r(v) \le \lg n$, kde n je počet uzlů ve stromu, platí pro každý uzel v, máme ~~složit~~ amor. složitost operace splay rovnu $O(\lg n)$.

Lemma 1: Pro všechna $a, b > 0$ platí $\log\frac{a+b}{2} \ge \frac{\log a + \log b}{2}$.

Důkaz: 1) $\log$ je konkávní fce (2. derivace $\ln x$ je: $-\frac{1}{x^2}$)



(viz obrázek. tvrzení platí pro každou konkávní funkci)

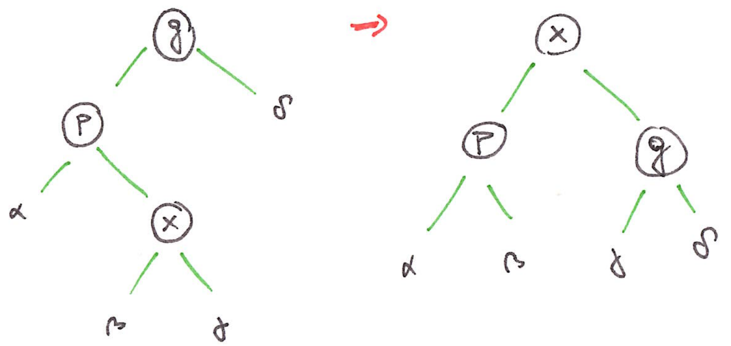
∎

jinak psáno máme $\log a + \log b \le 2\log(a+b) - 2$

protože $\log\frac{a+b}{2} = \log(a+b) - 1$

Lemma: Amortizovaná cena zig-zag je nejvýše $3r'(x) - 3r(x)$

Důkaz: připomeňme obrázek



(připomeňme že věci ze stromu před rotací značíme bez čárky, po rotaci s čárkou).

→ rank se změní jenom uzlům x, p, g

→ skutečná cena zig-zag je 2 (jsou to 2 základní rotace)

Máme tedy

$$\hat{c} = 2 + \overbrace{(r'(x) + r'(p) + r'(g)) - r(x) - r(p) - r(g)}^{\text{změna potenciálu}} \quad (0)$$

Použijeme:

$$r'(p) + r'(g) = \log m'(p) + \log m'(g)$$

$$\leq 2\log(m'(p) + m'(g)) - 2 \quad (1)$$

↑ Lemma 1

$T'(p)$ a $T'(g)$ jsou disjunktní a jsou to potomci x:

$$\log(m'(p) + m'(g)) \leq \log m'(x) = r'(x) \quad (2)$$

Spojením (1) a (2) máme

$$r'(p) + r'(g) \leq 2r'(x) - 2$$

Dosadíme do (0) a máme

$$\hat{c} \leq 3r'(x) - r(x) - r(p) - r(g)$$

Protože $T(x)$ je podstromem $T(p)$ i $T(g)$,

máme $r(x) \leq r(p)$
$r(x) \leq r(g)$

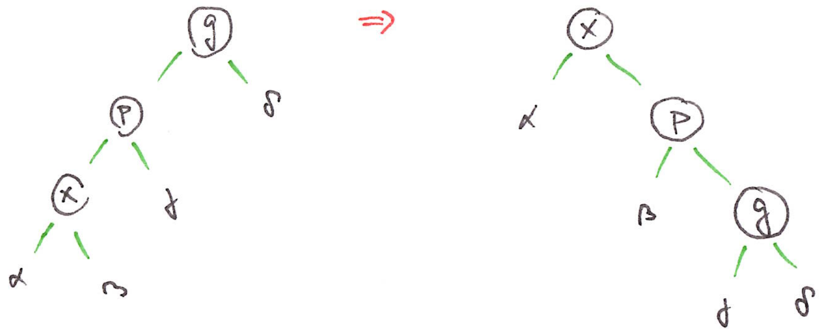
a tedy

$$\hat{c} \leq 3r'(x) - 3r(x)$$

∎

Lemma: Amortizovaná cena zig-zig je nejvýše $3r'(x) - 3r(x)$

Důkaz: připomeňme obrázek



→ rank se změní pouze uzlům x, p, g

→ skutečná cena je 2 (tj. 2 základní rotace)

$$\hat{c} = 2 + r'(x) + r'(p) + r'(g) - r'(x) - r'(p) - r'(g).$$

$$r(x) + r'(g) = \log m(x) + \log m'(z)$$

$$\leq 2\log(m(x) + m'(z)) - 2 \quad \text{// Lemma 1}$$

$$\leq 2\log m'(x) - 2 = 2r'(x) - 2$$

$T(x)$ a $T'(g)$ jsou disjunktní

a $T'(x) = T(x) \cup T'(g) \cup \{p\}$

úpravou máme

$$r'(g) \leq 2r'(x) - r(x) - 2$$

a tedy

$$\hat{c} \leq 3r'(x) + r'(p) - 2r(x) - r(g) - r(p)$$

z obrázku vidíme

$r(g) = r'(x)$ $\quad (T(g) = T'(x))$

$r(p) \geq r(x)$ $\quad (T(p) \supseteq T(x))$

$r'(p) \leq r'(x)$ $\quad (T'(p) \subseteq T'(x))$

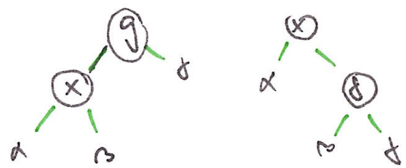
Po dosazení máme

$$\hat{c} \leq 3r'(x) - 3r(x)$$

∎

Lemma: Amortizovaná cena zig je nejvýše $3r'(x) - 3r(x) + 1$

Důkaz: Obrázek



$$\hat{c} = 1 + r'(x) + r'(g) - r(x) - r(g)$$

z obrázku vidíme

$r'(g) \leq r'(x)$ a $r(g) \geq r(x)$

Tj. $\hat{c} \leq 1 + 2r'(x) - 2r(x)$

Ale $r'(x) - r(x) \geq 0$ a proto

$$\hat{c} \leq 1 + 3r'(x) - 3r(x).$$

∎

SEARCH (k)

1) najdeme uzel s hledaným klíčem jako v BVS a potom na něj uděláme splay
2) pokud se k ve stromu nenachází, uděláme splay na poslední navštívený uzel během hledání k.

Složitost:

→ slož. splay je lin. závislá na výšce

→ slož. search v BVS je také lin. závislá na výšce

⇒ slož. search ve splay-stromu ~~je~~ má amort. slož. $O(\lg n)$

INSERT

1) INSERT jako u BVS
2) splay pro vložený uzel

Složitost.

→ podobné jako u search

→ dojde ale ke zvýšení potenciálu, to omezíme na $O(\log n)$ a tím pádem zůstane amort. slož $O(\log n)$.

Lemma: Přidání listu zvýší potenciál o $O(\log n)$

Důkaz: vkládáme uzel x, uzly na cestě z x do kořene ozn.

$$v_t, v_{t-1}, \dots v_1 \quad (= \text{kořen})$$

Změna potenciálu je

$$\Delta = \sum_{i=1}^{t} (r'(v_i) - r(v_i)) + r'(x) \qquad (*)$$

→ x je list, proto $r'(x) = 0$

→ $r'(v_i) = \log(m(v_i)+1)$ ← do podstromu přibyl nový uzel.

Máme

$$m(v_{i-1}) \geq m(v_i+1) = m'(v_i)$$

a tedy

$$r(v_{i-1}) \geq r'(v_i).$$

Dosadíme do $(*)$ a máme

$$\Delta \leq (r'(v_1) - r(v_1)) + \underbrace{\sum_{i=2}^{t} (r(v_{i-1}) - r(v_i))}$$

$$= (r'(v_1) - r(v_1)) + (r(v_1) - r(v_t))$$

$$= \underbrace{r'(v_1) - r(v_t)}_{\parallel \atop O(\lg n).}$$

$\square$

DELETE (x)

1) splay (x)

2) odebereme {x} a získáme 2 podstromy, levý a pravý

3) najdeme maximum y v levém podstromu a zavoláme na něj splay → ~~výsledný strom nemá pravéh~~ y potom nemá pravého potomka

4) jako pravého potomka y připojíme ⊛

Složitost:

2 → tady snížíme potenciál o $O(\log n)$, protože odebereme kořen (potenciál teď chápeme jako součet potenciálů levého a pravého podstromu).

4 → kořeni zvedneme potenciál o $O(\lg n)$.

3) → hledání maxima má amort. slož $O(\lg n)$, podobně jako INSERT.

Celkově tedy máme $O(\lg n)$