Datové struktury 1

2. přednáška

Petr Chmel

Organizační poznámka

- Prosím vyplňujte do úkolů, jak dlouho vám zabraly

tree_successor.py

```
1  #!/usr/bin/env python3
2
3  # Time spent on this homework: TODO (please fill in)
4
```

tree_successor.h

```
1  // Time spent on this homework: TODO (please fill in)
2
```

Plán

- Připomenutí amortizace
- Líně vyvažované binární vyhledávací stromy
- Splay stromy

Zdroj: Lípa poustevník, Tomáš Kalous – Strom roku

Amortizace

- Potenciál Φ je funkce, která na základě stavu datové struktury (a případně i historie operací) přiřadí nějakou hodnotu
  - Vysoký potenciál ~ jsme připraveni na drahou operaci
  - Nízký potenciál ~ teď budou operace levné
- Definice: operace má amortizovanou složitost *A*, pokud existuje potenciál Φ takový, že skutečná cena *C* splňuje

$$A_i \geq C_i + \Phi_i - \Phi_{i-1}$$

- Potom máme $\sum_i A_i \geq \Phi_m - \Phi_0 + \sum_i C_i$

Amortizace

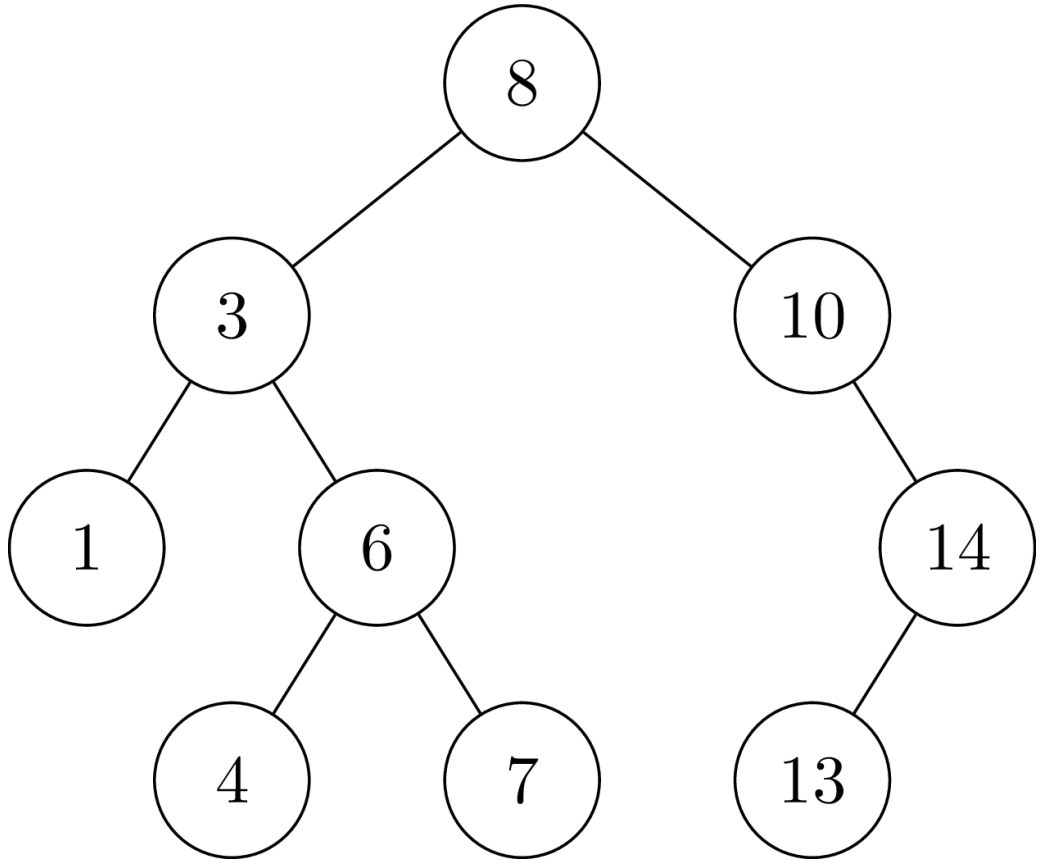
- Různá možná znění vět

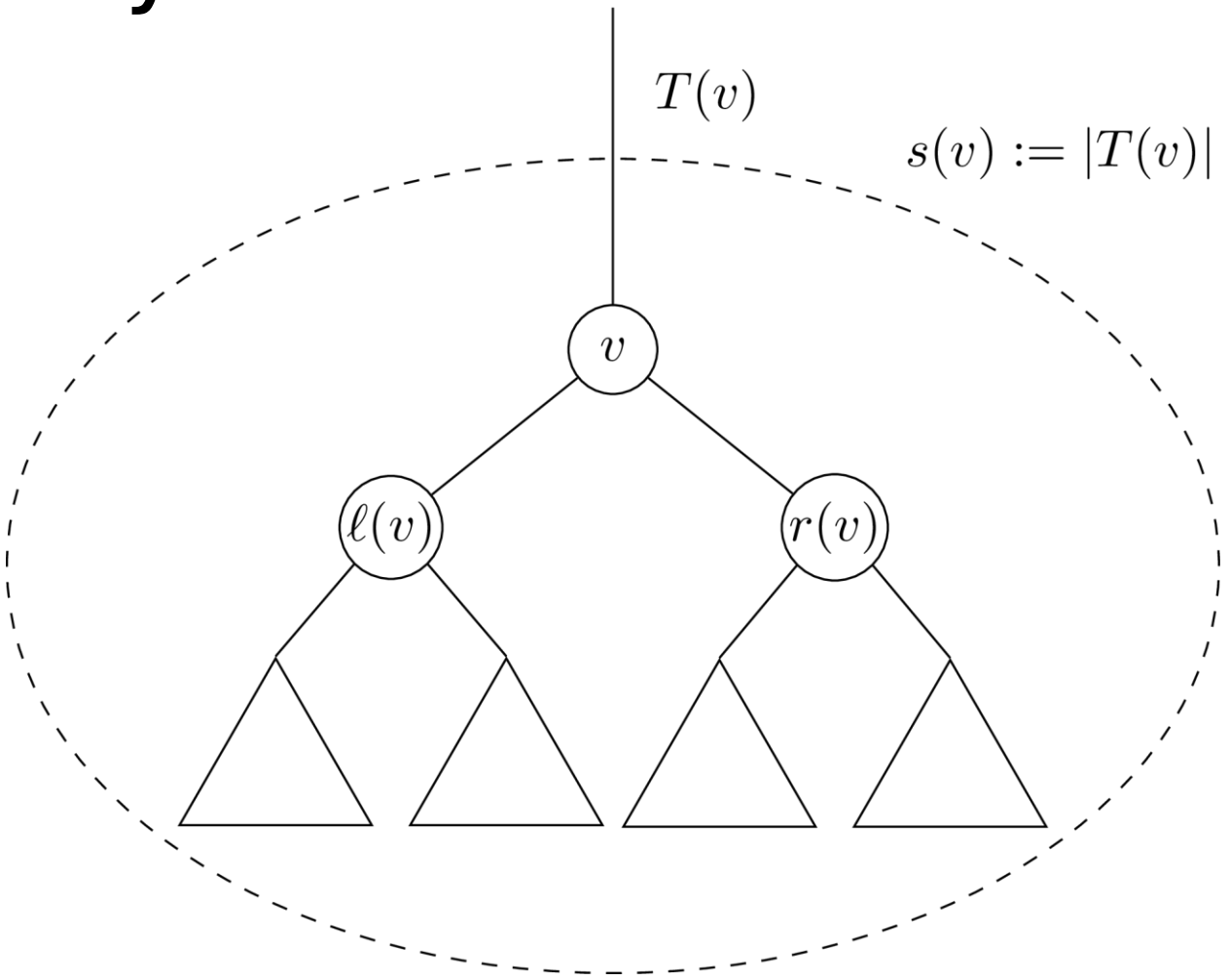
  Posloupnost m operací Find/Insert/Delete trvá celkem čas $\mathcal{O}(m \log n + n \log n)$, kde n je maximální počet prvků stromu během této posloupnosti.

  Posloupnost $m \geq n$ operací Find/Insert/Delete trvá celkem čas $\mathcal{O}(m \log n)$, kde n je maximální počet prvků stromu během této posloupnosti.

  Operace Find/Insert/Delete mají amortizovanou časovou složitost $\mathcal{O}(\log n)$

Binární vyhledávací stromy





Perfektně vyvážené stromy

- Definice

  Strom je perfektně vyvážený, pokud pro jeho každý vrchol *v* platí, že $\left|s\left(l(v)\right) - s\left(r(v)\right)\right| \leq 1$.

- Cvičení: Sestavte perfektně vyvážený strom ze setříděného pole v čase $\mathcal{O}(n)$
- Fakt: Perfektně vyvážený BVS nelze s operacemi Insert/Delete udržovat v čase lepším než $\Omega(n)$ na operaci.

Jenom vyvážené stromy

- Definice

  Vrchol *v* je vyvážený, pokud

$$s\big(l(v)\big) \leq \frac{2}{3}s(v) \text{ a } s\big(r(v)\big) \leq \frac{2}{3}s(v)$$

  Strom je vyvážený, jestliže jsou všechny jeho vrcholy vyvážené

- Lemma: Hloubka vyváženého stromu je $\mathcal{O}(\log n)$

Zdroj: Pat a Mat - Pat a Mat se vracejí: Vánoční stromeček (S04E28)

Zdroj: Pat a Mat - Pat a Mat se vracejí: Vánoční stromeček (S04E28)

Zdroj: Pat a Mat - Pat a Mat se vracejí: Vánoční stromeček (S04E28)

BB[α]-stromy (Reingold, 1972)

- Každý vrchol si navíc počítá svoje $s(v)$
- Insert:
  - Standardně vložíme do BVS a patřičně zvýšíme $s(v)$ na cestě
  - Pokud jsou všechny vrcholy na cestě vyvážené, končíme
  - Jinak vezmeme nejvyšší nevyvážený vrchol a jeho podstrom přestavíme
- Delete analogicky

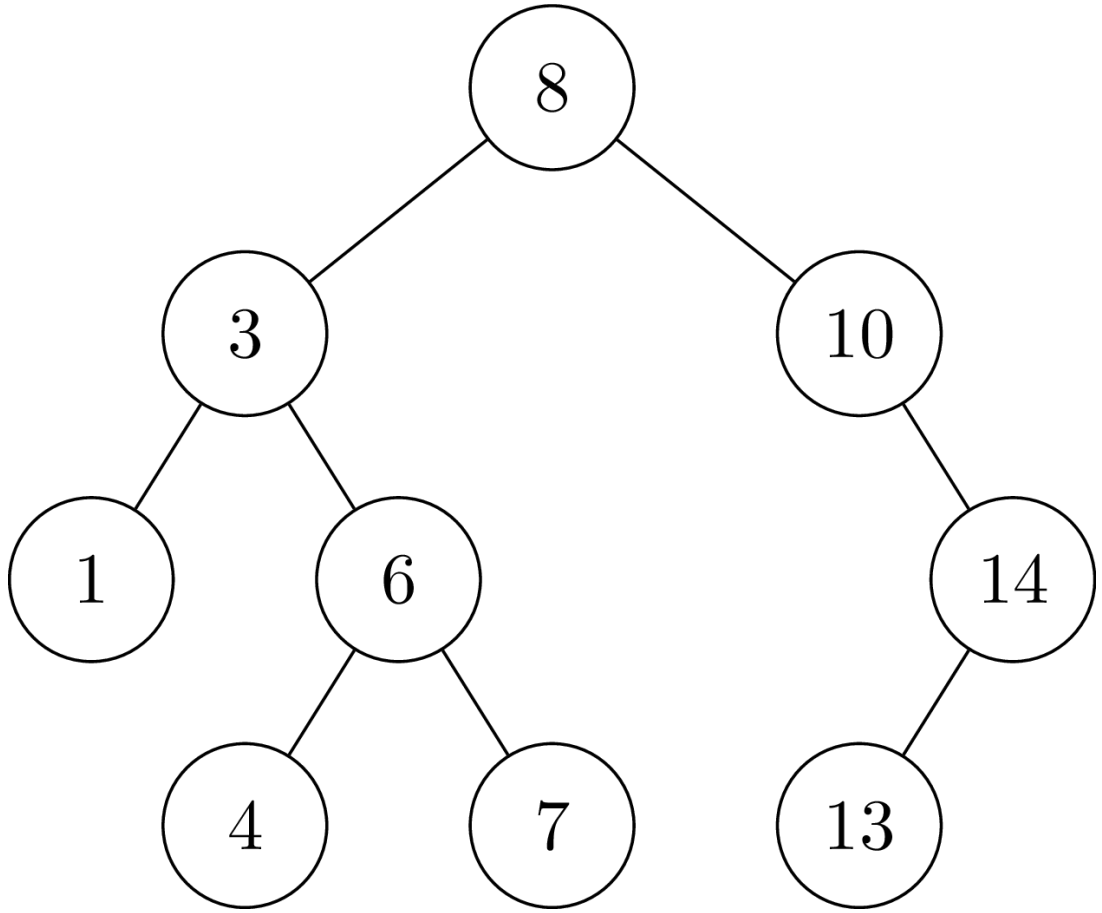
Amortizovaná analýza

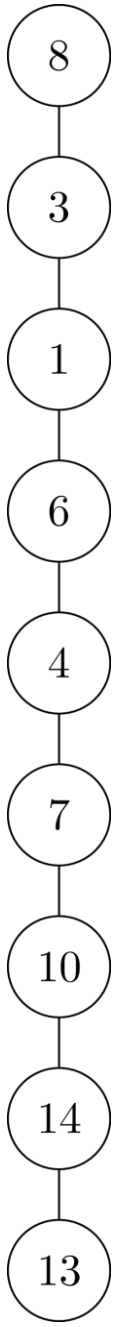
- Věta

  Operace Insert má amortizovanou cenu $\mathcal{O}(\log n)$

- Zavedeme si následující potenciál:
  - Pro vrchol v: $\varphi(v) = \begin{cases} \left|s\left(l(v)\right) - s\left(r(v)\right)\right|, \text{pokud} \geq 2 \\ 0, \text{ jinak} \end{cases}$
  - Pro celý strom: $\Phi(T) = \sum_{v \in V(T)} \varphi(v)$

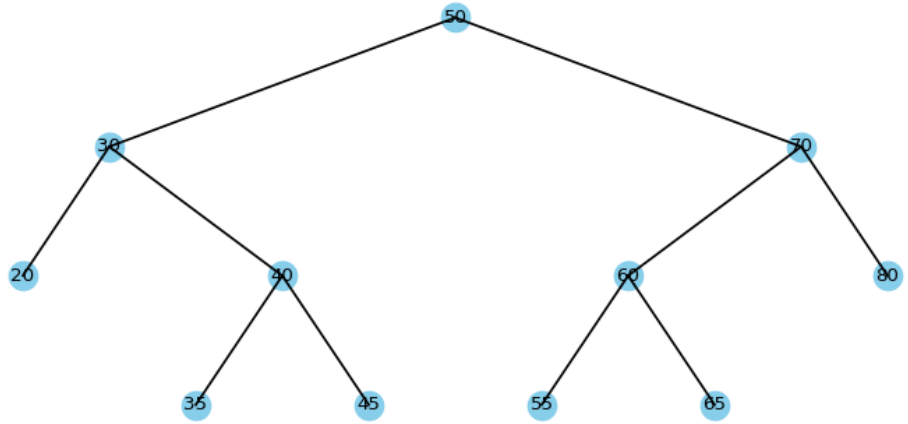
Motivace splay stromů



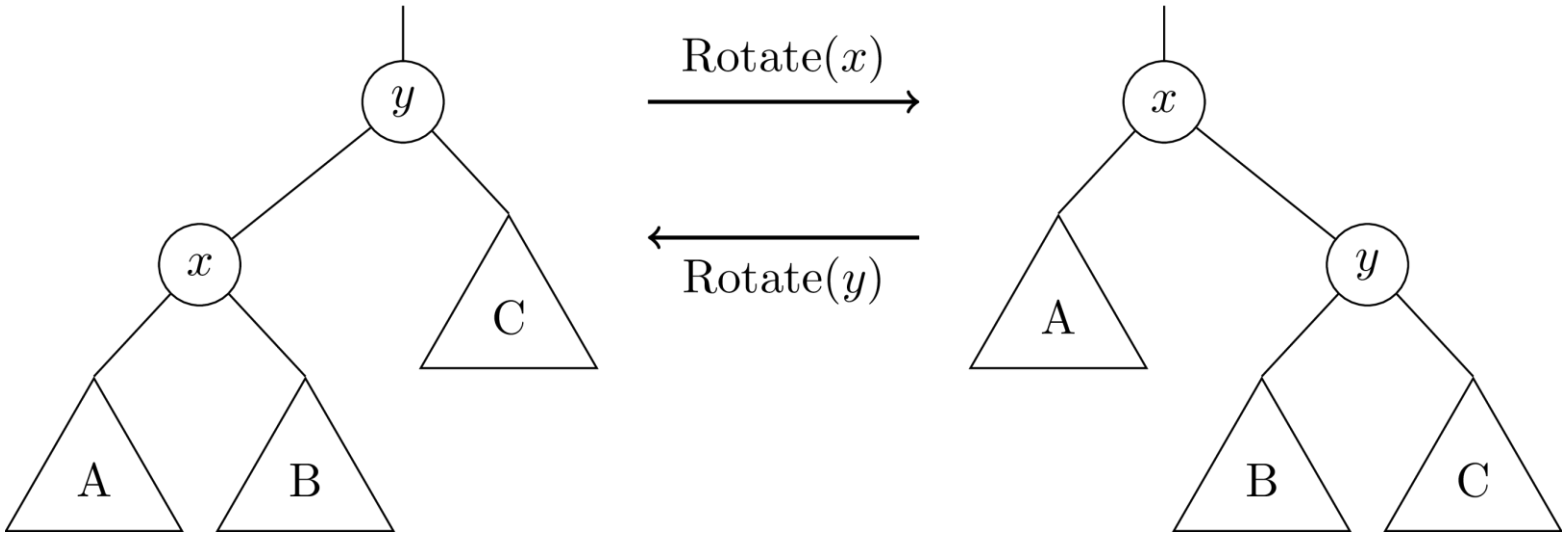


Rotace

Oh nein!

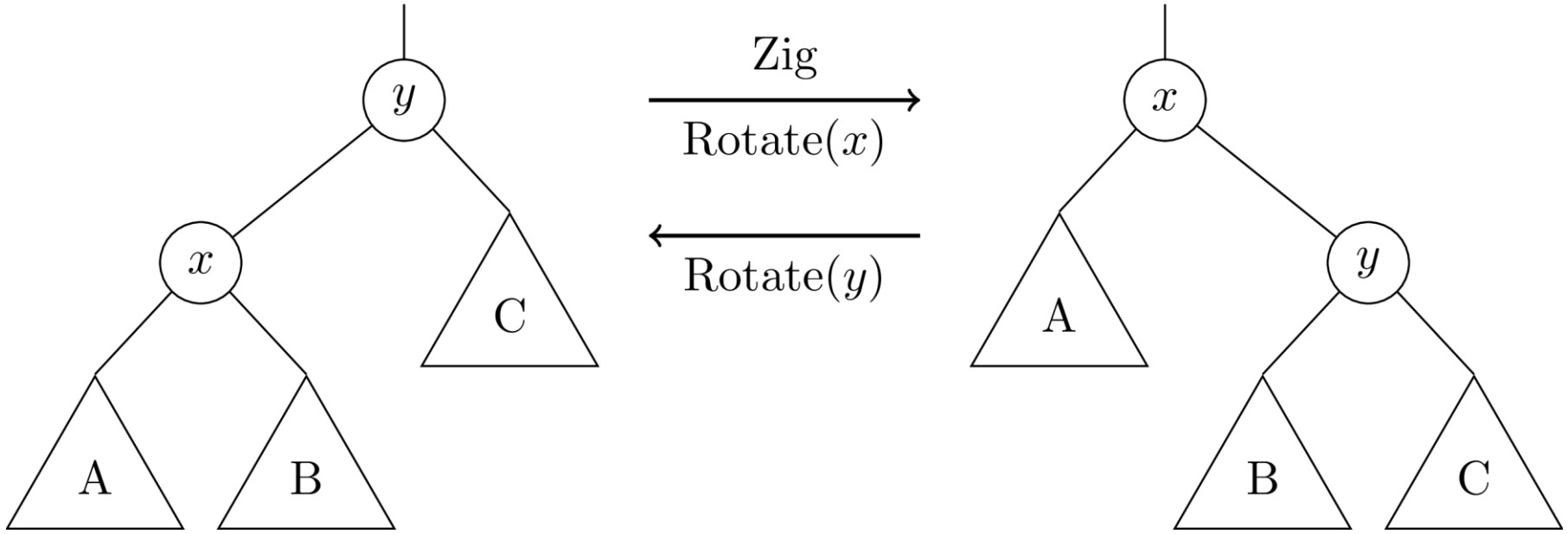


Rotierender binärer Suchbaum

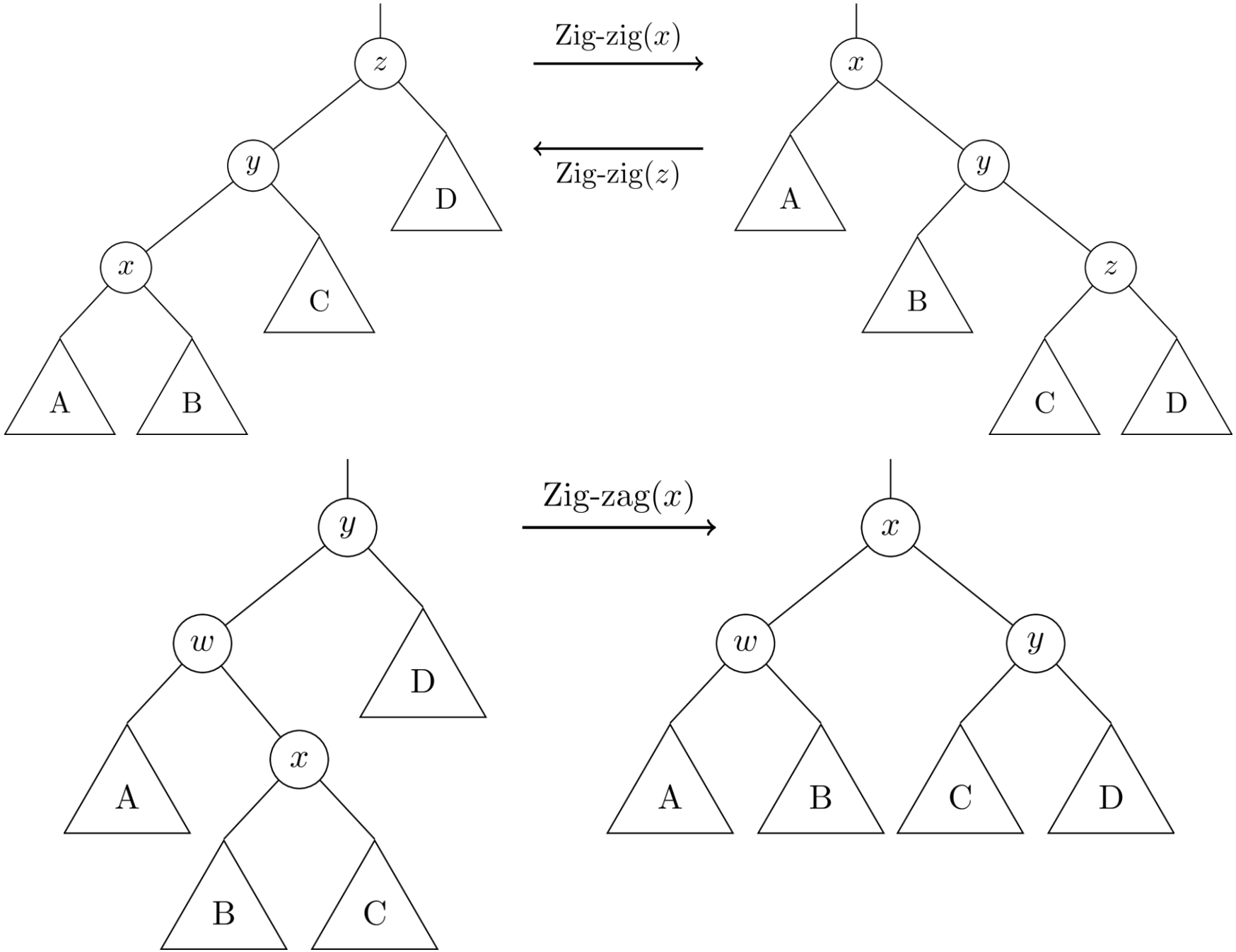


Splay operace (Sleator & Tarjan 1985)

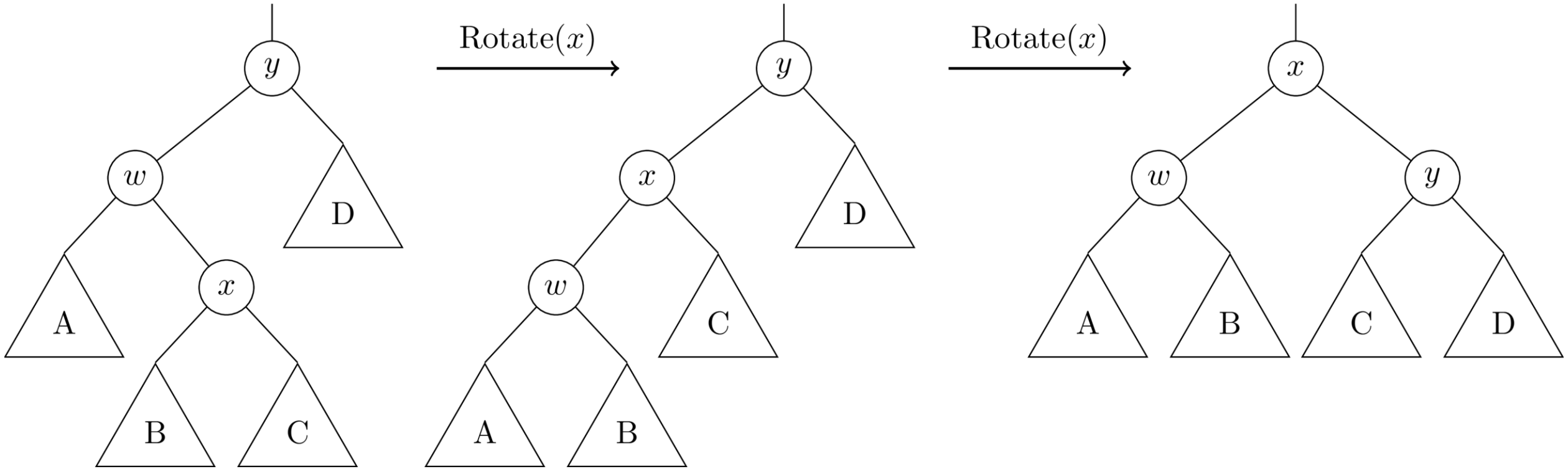
- Vezmeme vrchol, a dokud můžeme, provádíme dvourotace (zig-zig, nebo zig-zag), na konci možná provedeme jednu jednoduchou rotaci (zig)



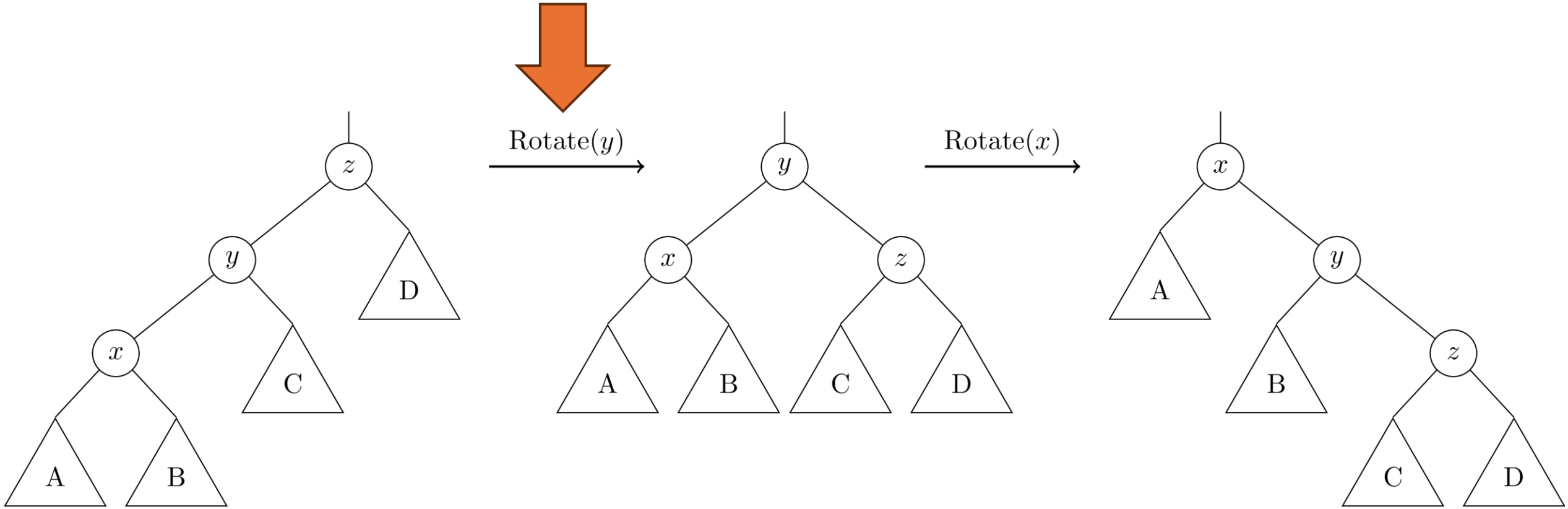
Dvojrotace



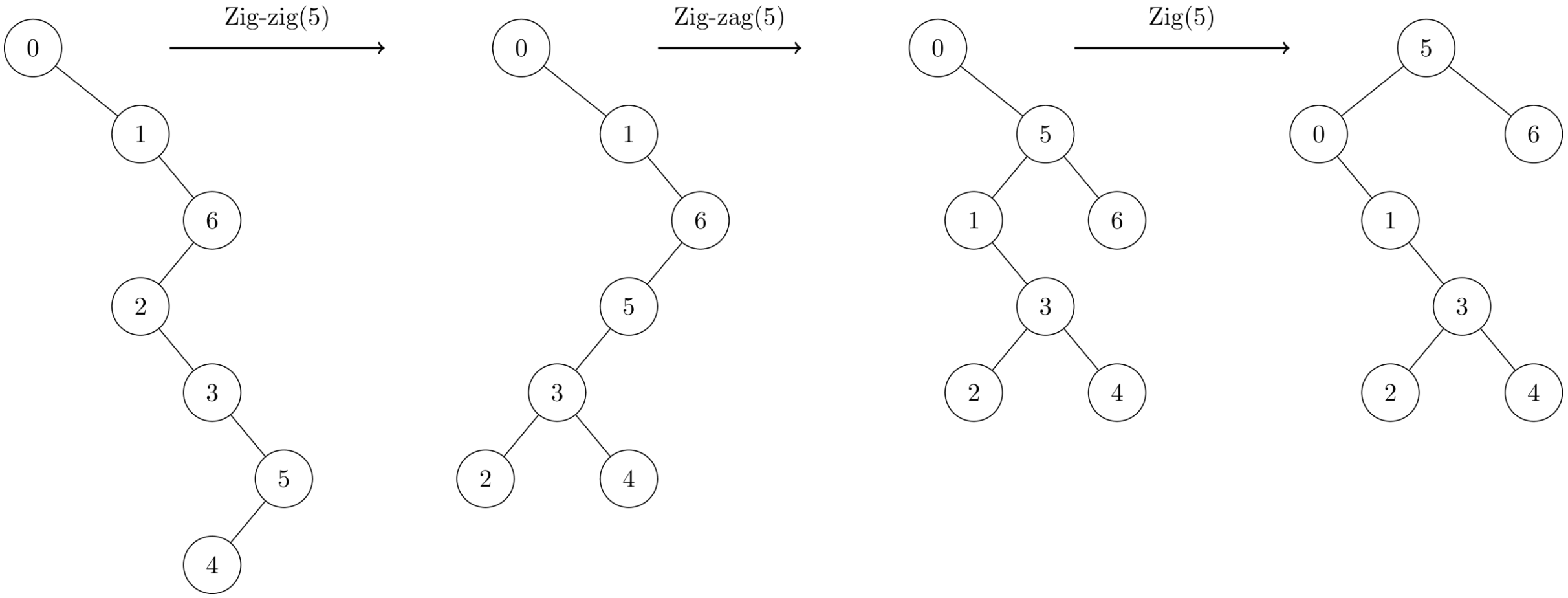
Zig-zag jednoduchými rotacemi



Zig-zig jednoduchými rotacemi



Příklad



Tajná myškavěc

- Lemma: $\forall \alpha, \beta > 0: \log_2 \alpha + \log_2 \beta \leq 2\log_2(\alpha + \beta) - 2$
- Důkaz:

  Víme $4\alpha\beta = (\alpha + \beta)^2 - (\alpha - \beta)^2 \leq (\alpha + \beta)^2$, protože $(\alpha - \beta)^2 \geq 0$.

  Zlogaritmujeme, a máme $\log_2 4 + \log_2 \alpha + \log_2 \beta \leq 2 \log_2(\alpha + \beta)$.

Cena operace Splay

- Cíl: operace Splay má amortizovanou cenu $\mathcal{O}(\log n)$
- Pro vrchol *v* zadefinujeme jeho rank $r(v) \coloneqq \log_2 s(v)$
- Pak definujeme potenciál stromu $\Phi(T) = \sum_{v \in T} r(v)$
- Věta

  Amortizovaná cena operace Splay(*x*) je $3(r'(x) - r(x)) + 1$, kde $r'(x)$ je rank po provedení splaye a $r(x)$ je rank před provedením splaye

- Zpozorujeme $r(x) \leq \log_2 n$, tedy cena Splay(x) je $\mathcal{O}(\log n)$

Find, Insert, Delete na splay stromech

- Neformálně: operaci provedeme stejně jako na standardních BVS, a pak vysplayujeme nejhlubší prvek, kterého jsme se dotkli
- Find
  - Úspěšný: vysplayujeme nalezený vrchol
  - Neúspěšný: vysplayujeme poslední navštívený vrchol
- Insert
  - Úspěšný: vysplayujeme vložený vrchol
  - Neúspěšný: vysplayujeme už existující vrchol
- Delete
  - Úspěšný: vysplayujeme rodiče smazaného vrcholu
  - Neúspěšný: vysplayujeme poslední navštívený vrchol

Amortizovaná složitost Find/Insert/Delete

- Tvrzení

  Všechny tyto operace mají amortizovanou časovou složitost $\mathcal{O}(\log n)$

- Lemma (pro Insert)

  Přidání listu zvýší potenciál o $\mathcal{O}(\log n)$

Vlastnosti splay stromů

- Věta (sequential access theorem, Tarjan 1985)

  Vyhledávání prvků $1, \dots, n$ v tomto pořadí má celkovou cenu $\mathcal{O}(n)$.

- Věta (working set theorem, Sleator & Tarjan 1985)

  Buď $x_1, \dots, x_m$ posloupnost vyhledávání prvků ve splay stromě. Označme jako z_i počet různých prvků vyhledaných od posledního hledání prvku x_i. Pak celková složitost hledání prvků je $\mathcal{O}(n \log n + m + \sum_i \log(1 + z_i))$.

- Důsledek (o vyhledávání podmnožiny prvků)

  Pokud provedeme m vyhledání prvků z podmnožiny $X \subseteq [n]: |X| = s$, pak celková cena vyhledávání bude $\mathcal{O}(n \log n + m + m \log s)$.

Vlastnosti splay stromů II

- Věta (statická optimalita, Sleator & Tarjan 1985)

  Buď $x_1, \ldots, x_m$ posloupnost vyhledávání prvků z množiny X, kde každý prvek je vyhledán aspoň jednou. Buď T statický BVS na množině X s celkovou cenou vyhledání posloupnosti f. Pak celková cena téže posloupnosti ve splay stromě na X je $\mathcal{O}(f)$.

- Domněnka (dynamická optimalita, Sleator & Tarjan 1985)

  Buď $x_1, \ldots, x_m$ posloupnost vyhledávání prvků z množiny X, kde každý prvek je vyhledán aspoň jednou. Buď T dynamický BVS na množině X s celkovou cenou vyhledání posloupnosti f. Pak celková cena téže posloupnosti ve splay stromě na X je $\mathcal{O}(f)$.

Příště

- Ještě více stromů: (a,b)-stromy a amortizace