

7. 二叉搜索树

(c) 平衡与等价

邓俊辉

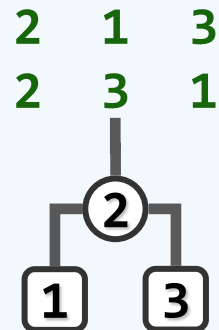
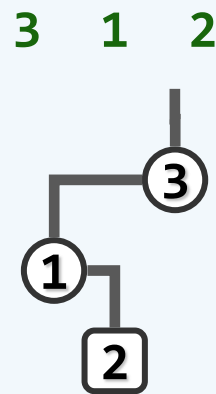
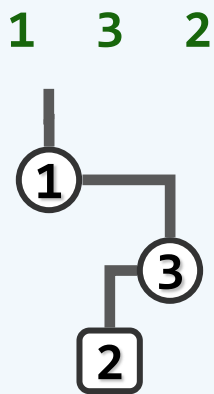
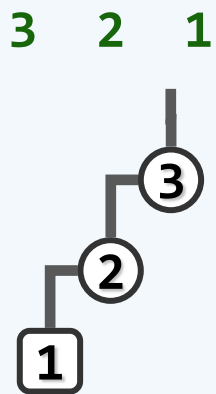
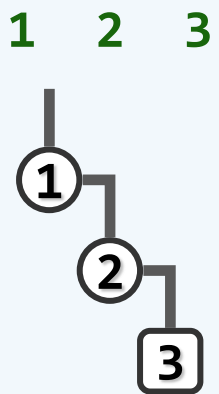
deng@tsinghua.edu.cn

树高

- ❖ 由以上的实现与分析，BST主要接口`search()`、`insert()`和`remove()`的运行时间
在最坏情况下，均线性正比于其高度 $O(h)$
- ❖ 因此，若不能有效地控制树高，则就实际的性能而言
较之此前的向量和列表等数据结构，BST将无法体现出明显优势
- ❖ 比如在最坏情况下，二叉搜索树可能彻底地退化 \square 为列表
此时的查找效率甚至会降至 $O(n)$ ，线性正比于树（列表）的规模
- ❖ 那么，出现此类最坏或较坏情况的概率有多大？
或者，从平均复杂度的角度看，二叉搜索树的性能究竟如何呢？
- ❖ 以下按两种常用的随机统计口径 \square ，就BST的平均性能 \square 做一分析和对比

随机生成

- ❖ 考查 n 个互异词条 $\{ e_1, e_2, \dots, e_n \}$ ，对任一排列 $\sigma = (e_{i_1}, e_{i_2}, \dots, e_{i_n}) \dots$
- ❖ 从空树开始，反复调用`insert()`接口将各词条依次插入，得到 $T(\sigma)$
- ❖ 与 σ 相对应的 $T(\sigma)$ ，称由 σ 随机生成 (randomly generated)



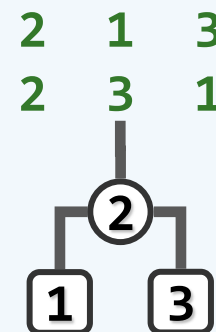
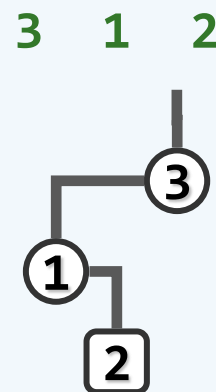
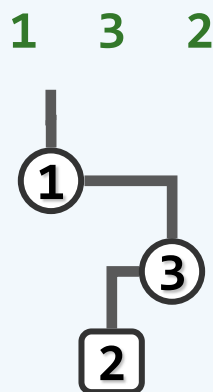
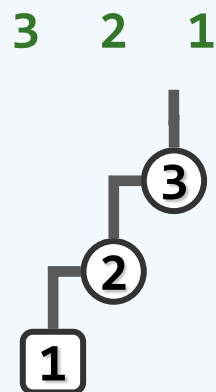
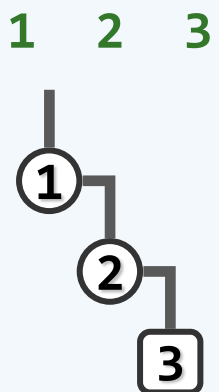
- ❖ 若假定任一排列 σ 作为输入的概率均等 $1/n!$

则由 n 个互异词条随机生成的二叉搜索树，平均高度为 $\Theta(\log n)$

随机组成

❖ n 个互异节点，在遵守顺序性的前提下，可随机确定拓扑联接关系

❖ 如此所得的BST，称由这组节点**随机组成** (randomly composed)



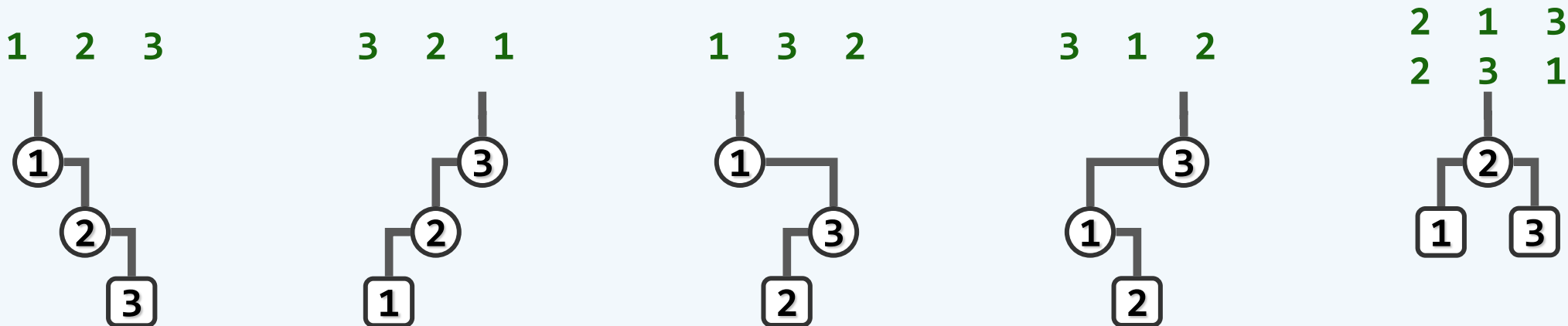
❖ 由 n 个互异节点随机组成的BST，若共计 $T(n)$ 棵，则有

$$T(n) = \sum_{k=1}^n SP(k-1) \cdot SP(n-k) = \text{Catalan}(n) = (2n)! / (n+1)! / n!$$

❖ 假定所有BST等概率出现，则其平均高度为 $\Theta(\sqrt{n})$

$\Theta(\log n)$ vs. $\Theta(\sqrt{n})$

❖ 按两种口径所估计的平均性能，差异极大——谁更可信？谁更接近于真实情况？



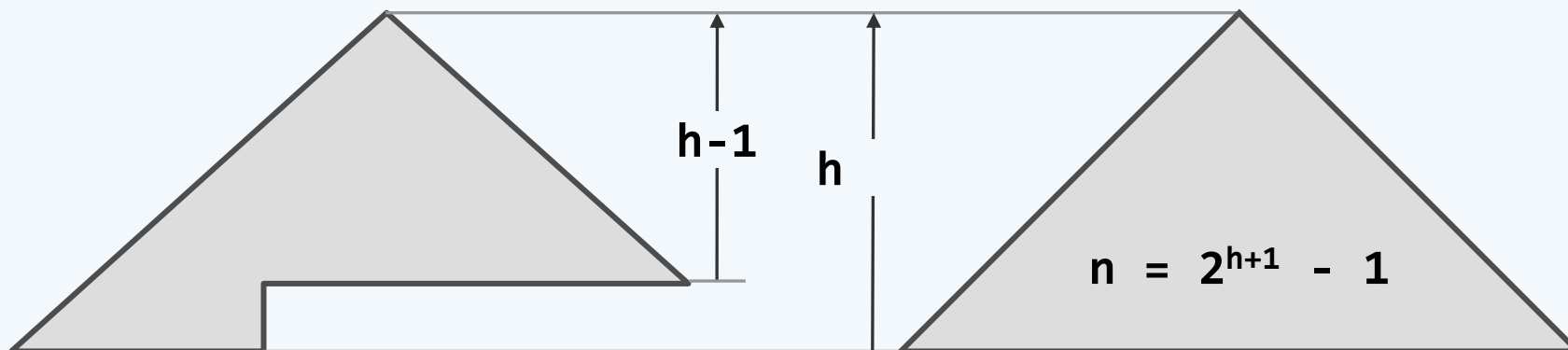
❖ 前一口径中，越低的BST被重复统计更多次——故嫌过于乐观

❖ 若删除算法固定使用 `succ()`，则每棵BST都有越来越左倾的趋势

❖ 理想随机在实际中并不常见，关键码往往按单调甚至线性的次序出现
极高

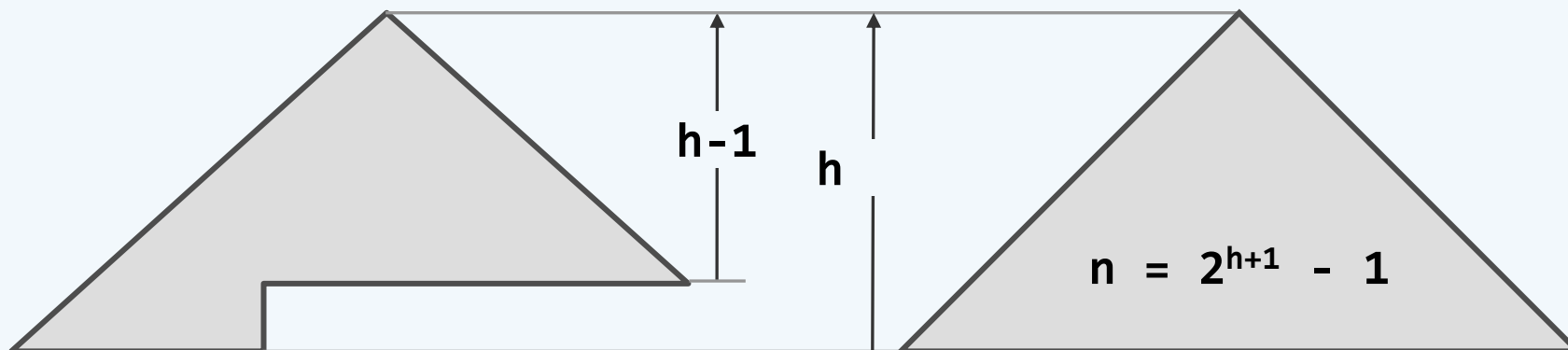
理想平衡

- ❖ 节点数目固定时，兄弟子树高度越接近（平衡），全树也将倾向于更低
- ❖ 由 n 个节点组成的二叉树，高度不低于 $\lfloor \log_2 n \rfloor$ ——恰为 $\lfloor \log_2 n \rfloor$ 时，称作理想平衡
- ❖ 大致相当于完全树甚至满树：叶节点只能出现于最底部的两层——条件过于苛刻



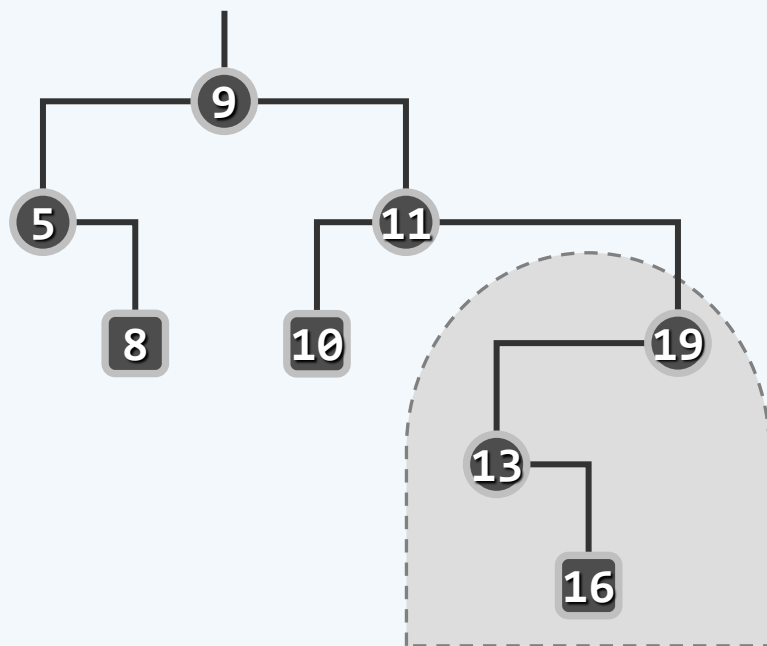
适度平衡

- ❖ 理想平衡出现**概率**极低、维护**成本**过高，故须适当地放松标准
- ❖ **退一步海阔天空**：高度**渐进地**不超过 $O(\log n)$ ，即可称作**适度平衡**
- ❖ 适度平衡的BST，称作**平衡二叉搜索树**（BBST）

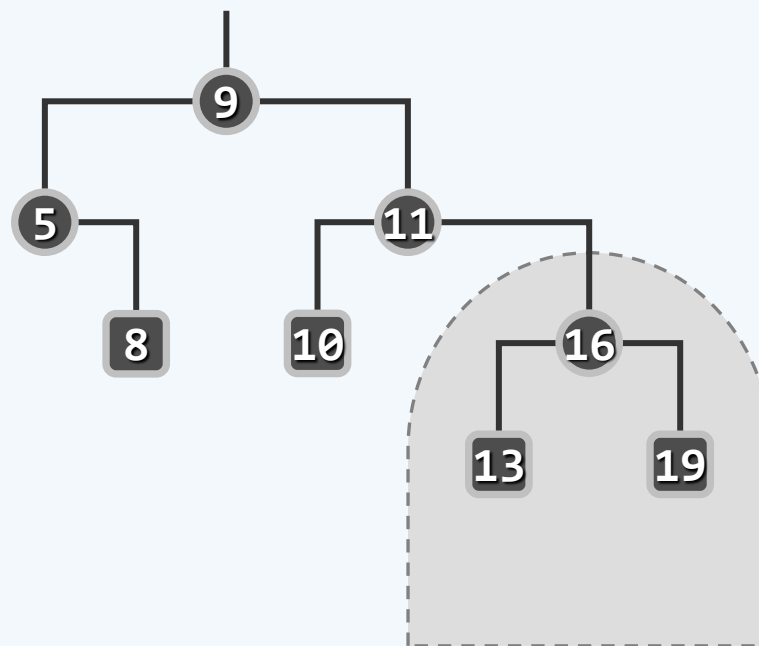


等价BST

- ❖ **上下可变**：联接关系不尽相同，承袭关系可能颠倒
- 左右不乱**：中序遍历序列完全一致，全局单调非降



5 8 9 10 11 13 16 19

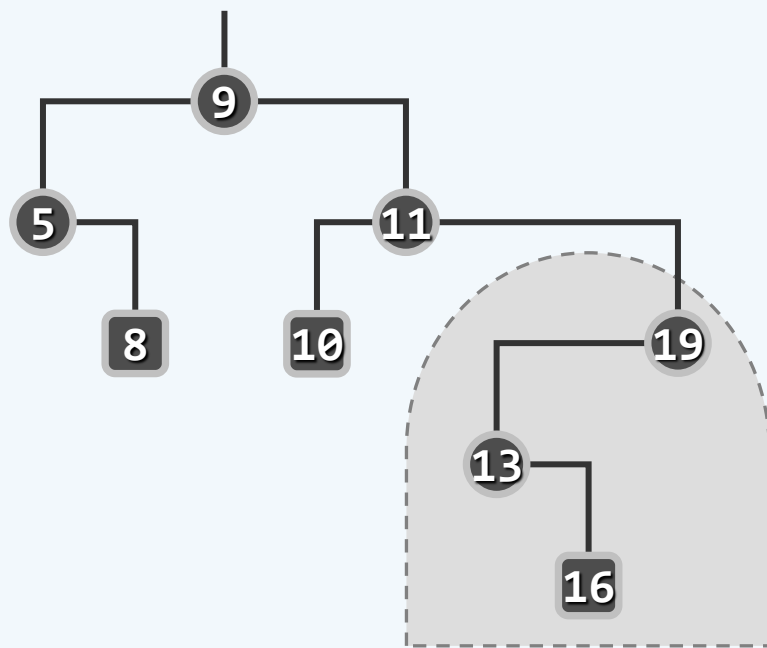


5 8 9 10 11 13 16 19

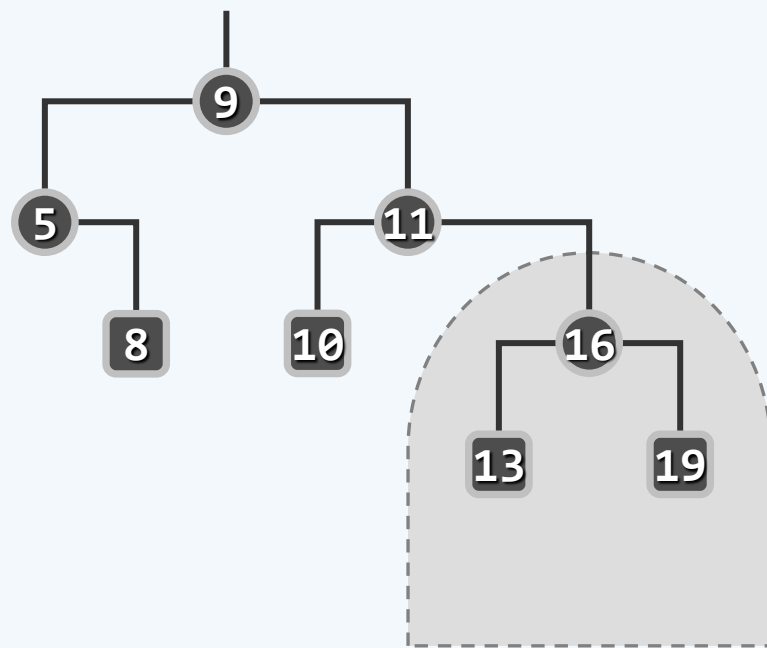
限制条件 + 局部性

❖ 各种BBST都可视作BST的某一子集，相应地满足精心设计的限制条件

- 1) 单次动态修改操作后，至多 $O(1)$ 处局部不再满足限制条件
- 2) 可在 $O(\log n)$ 时间内，使这些局部（以至全树）重新满足



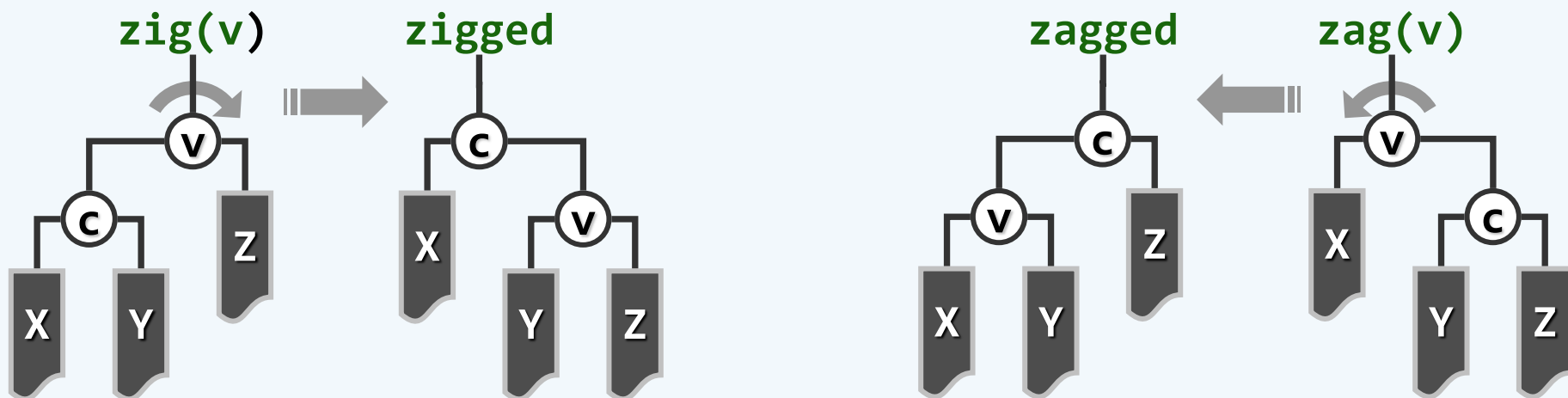
5 8 9 10 11 13 16 19



5 8 9 10 11 13 16 19

等价变换 + 旋转调整

❖ 刚刚失衡的BST，必可**迅速**转换为一棵等价的BBST——为此，只需 $O(\log n)$ 甚至 $O(1)$ 次旋转



❖ zig和zag：仅涉及**常数**个节点，只需调整其间的联接关系；均属于**局部**操作、**基本**操作

❖ 调整之后：v/c深度加/减1，子（全）树高度的变化幅度，上下**不超过1**

❖ 实际上，经过不超过 $O(n)$ 次旋转，等价的BST均可相互转化 **习题解析[7-15]**