Homework Assignment #8
Due: July 21, 2023 at 10:00 p.m.

1. Consider a B-tree with parameter t. Assume that INSERT operations split eagerly: while going down the path to the leaf, the INSERT splits any node containing $2t-1$ keys before entering it. Although an INSERT into a tree containing n keys may have to update $\Omega(\log_t n)$ nodes in the worst case (if it splits every node along the search path from the root to a leaf), our goal is to show that the amortized number of nodes an INSERT updates is $O(1)$. In other words, we wish to prove that a sequence of m INSERTS (starting from an empty tree) has to update m nodes in total. (We count the creation of a new node or any modification to an existing node as a node update.)

[2] **(a)** Eric wants to design a potential function to do the amortized analysis. The idea is that each node split should reduce the potential by 3 to cover the cost of updating three nodes (the two new children created by the split and the parent, which loses a key). Since the potential is supposed to measure how close we are to expensive INSERTS that do lots of splits, Eric figures that if there are more full nodes (i.e., nodes that contain $2t-1$ keys) then there is more potential for split operations in the future, so he tries designing a potential function Φ_1 that depends (only) on the number of nodes that are full (i.e., contain $2t-1$ keys).

Show that an INSERT can do many splits that do not change the number of full nodes in the tree. (This means that if Φ_1 depends only on the number of full nodes, it will not be useful for proving the desired amortized bound.)

[2] **(b)** Eric tries again. This time, he figures vacant nodes (i.e., nodes containing only $t-1$ keys so that they have lots of vacant space for more keys) are good since they do not have to be split. So, he defines $\Phi_2 = (\#\text{ nodes}) - (\#\text{ vacant nodes})$. The negative sign in front of the number of vacant nodes ensures that the potential is higher if there are fewer vacant nodes, and adding the number of nodes ensures that Φ_2 is non-negative. Moreover, a split creates two new vacant nodes so he hopes Φ_2 will decrease each time a split is done.

Show that Φ_2 is not useful for the desired amortized analysis either, by giving an example where an INSERT does many splits that do *not* cause Φ_2 to decrease.

[2] **(c)** Let $\Phi_3 =$ ______ $-$ (# vacant nodes). Fill in the blank to ensure that

- Φ_3 is always non-negative,
- Φ_3 is 0 when the B-tree is initially empty, and
- Φ_3 decreases by at least 1 every time a split is done.

[3] **(d)** Use $3 \cdot \Phi_3$ to show that the amortized number of node updates per INSERT is $O(1)$.