

Homework Assignment #7
Due: November 8, 2024 at 5:00 p.m.

1. Suppose you have a database S of pairs (x, y) in $\mathbb{Z}^2$. You must support the following operations on the database.

- $\text{INSERT}(x, y)$ inserts the pair (x, y) into S (if it is not already in S); returns true if the insertion succeeds.
- $\text{DELETE}(x, y)$ removes the pair (x, y) from S ; returns true if (x, y) was in S prior to performing the operation.
- $\text{SEARCH}(x, y)$ returns true if $(x, y) \in S$ and false otherwise.
- $\text{MEAN}(a)$ returns the average y -value of all points in S whose x -values are less than a .
- $\text{SD}(a)$ returns the standard deviation of the y -values of all points in S whose x -values are less than a .

The average of $y_1, \dots, y_k$ is $\bar{y} = \frac{1}{k} \sum_{i=1}^k y_i$. The standard deviation of $y_1, \dots, y_k$ is $\sqrt{\frac{1}{k} \sum_{i=1}^k (y_i - \bar{y})^2}$.

Describe how to do this using an augmented RBT. Your explanation should include the following components.

- What is stored in the RBT.
- What is used as the key for sorting the RBT.
- What field(s) you add to the nodes when augmenting the RBT and what each one represents.
- How you can compute the additional field(s) of a node using information in that node and its children.
- How you implement MEAN and SD. Provide either a detailed English description or pseudocode.

Give good bounds on the worst-case running time of all five operations in terms of $n = |S|$.