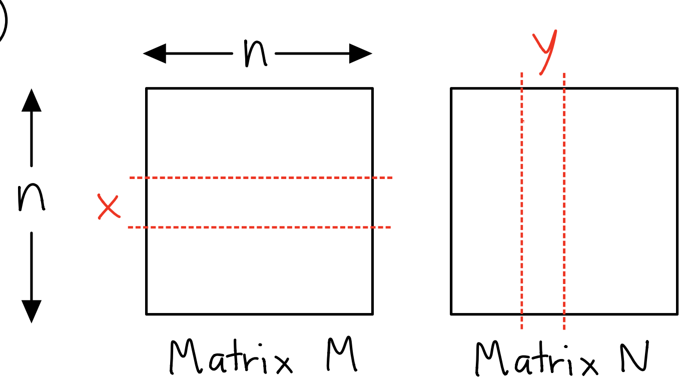
9



Assuming a reducer receives x rows of M and y columns of N, then $q = n(x + y)$, and xy outputs are covered.

Let p be covered outputs, so

$$q = n(x+y) \quad \text{— ①}$$

$$p = xy \quad \text{— ②}$$

From ①, we get $x = \frac{q}{n} - y$. Then, substitute x into ②.

$$p = y\left(\frac{q}{n} - y\right)$$

$$p = \frac{q}{n}y - y^2$$

Next, take derivative with respect to y and set it to zero to find the maximum number of covered outputs.

$$p' = \frac{q}{n} - 2y$$

$$0 = \frac{q}{n} - 2y$$

$$y = \frac{q}{2n} \quad \text{and} \quad x = \frac{q}{n} - y = \frac{q}{2n}$$

$$\therefore x = y \ \#$$

So, the maximum number of outputs are covered by a reducer when that reducer receives an equal number of rows of M and columns of N.

In order to find lower bounds on replication rate, there is inequality stating that "Suppose there are k reducers, and the i-th reducer has $q_i < q$ inputs. Observe that $\sum_{i=1}^{k} g(q_i)$ must be no less than the total number of outputs, where $g(q)$ is the number of outputs covered by the reducer."

From ②, we get $g(q) = xy = \left(\frac{q}{2n}\right)^2$, and the total number of outputs is n^2.

So, the inequality is

$$\sum_{i=1}^{k} \left(\frac{q_i}{2n}\right)^2 \geq n^2$$

$$\sum_{i=1}^{k} \left(\frac{q_i^2}{4n^2}\right) \geq n^2$$

$$\therefore \sum_{i=1}^{k} q_i^2 \geq 4n^4 \quad \# \quad —③$$

From replication rate formula,

$$r = \frac{\text{total inputs to all the reducers}}{\text{total inputs}}$$

$$= \frac{\sum_{i=1}^{k} q_i}{n^2 + n^2}$$

$$r = \frac{\sum_{i=1}^{k} q_i}{2n^2} \quad —④$$

From ③ and the fact that $q_i < q$, we can rearrange ③.

$$\sum_{i=1}^{k} q_i q_i \geq 4n^4$$

$$\sum_{i=1}^{k} q_i q \geq 4n^4$$

$$\sum_{i=1}^{k} q_i \geq 4n^4/q \quad \text{substitute into ④}$$

$$r \geq \frac{4n^4/q}{2n^2}$$

$$\therefore r \geq \frac{2n^2}{q} \quad \#$$