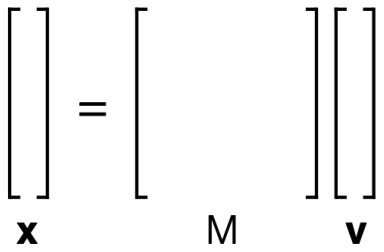
Take-Home Quiz 5 (Big Data)

1. Use the Map Reduce algorithm to solve the following problems and define exactly Map and Reduce functions.

   - Matrix-Vector Multiplication
     Solution:



     - assumption : each mapper can load vector **v**
     - map function : maps $((i, j), M_{ij})$ to $((i, M_{ij}\, v_j)$
     - reduce function : receives $(i, [M_{i1}\, v_1, \ldots M_{in}\, v_n])$
       sums all values of the list of a key i, i.e., $x_i = \sum_{j=1}^{n} M_{ij} v_j$
       produces (i, x_i)

   - Matrix Multiplication
     Solution:
     `http://www.mathcs.emory.edu/~cheung/Courses/554/Syllabus/9-parallel/matrix-mult.html`

2. In which cases Map Reduce algorithm cant solve the problem efficiently? (At least two cases should be mentioned and the cause of each should be briefly explained).

   Solution:

   `https://www.andrew.cmu.edu/course/14-848/applications/ln/graphlab.pdf`

3. Choose which framework is better in each of tasks below, Spark or Hadoop? Why? Solution:

   - Getting monthly report from all sale data in Digikala website : Hadoop, because it is batch processing.

- Processing big social networks like twitter: Saprk, because it has good graph libraries.
- Real-time fraud detection with all camera data in the city: Spark, because it can do near real-time processing.