CAAM 335, Fall 2021, Homework 1 Solutions

Problem 1 (10 points).

Let $M \in \mathbb{R}^{m\times n}$ and $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$. Prove the identity

$$y^T Mx = x^T M^T y$$

(Hint: Note that $y^T Mx$ and $x^T M^T y$ are scalars, and use the identity $(AB)^T = B^T A^T$.)

Solution Let $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$ and $M \in \mathbb{R}^{m\times n}$. We will use use the identity $(AB)^T = B^T A^T$ and the fact that $y^T Mx$ and $x^T M^T y$ are scalars to prove the identity $y^T Mx = x^T M^T y$. Consider

$$\begin{aligned} y^T Mx &= \left(y^T (Mx)\right)^T && (1)\\ &= (Mx)^T (y^T)^T && (2)\\ &= x^T M^T y, && (3) \end{aligned}$$

where (1) holds since $y^T Mx$ is a scalar and (2), (3) holds because of the identity $(AB)^T = B^T A^T$.

Problem 2 (20 points). A square matrix $G \in \mathbb{R}^{n\times n}$ is called a *stochastic matrix* if its entries are nonnegative, $G_{ij} \geq 0$, $i, j = 1,\dots,n$, and its columns have entries that sum up to one, $\sum_{i=1}^n G_{ij} = 1$, $j = 1,\dots,n$.

- Let $G \in \mathbb{R}^{n\times n}$ be a stochastic matrix. Show that if $x \in \mathbb{R}^n$ is a *stochastic vector*, i.e., it has nonnegative entries, $x_i \geq 0$, and entries sum up to one, $\sum_{i=1}^n x_i = 1$, then $y = Gx$ is a *stochastic vector*, i.e., it satisfies $y_i \geq 0$ and $\sum_{i=1}^n y_i = 1$.
- Let $F, G \in \mathbb{R}^{n\times n}$ be stochastic matrices. Show that the product FG is a stochastic matrix.

Solution

- We will show that $y = Gx$ is a stochastic vector through direct computation. Here $G \in \mathbb{R}^{n\times n}$ is a stochastic matrix and x is a stochastic vector.

  First, we will shows $y_i \geq 0$ for all $i = 1,\dots,n$. Fix $i \in \{1,\dots,n\}$. By direct computation (matrix multiplication), $y_i = \sum_{j=1}^n G_{ij}x_j$. Since $x_j \geq 0$ and $G_{ij} \geq 0$ for all $i, j \in \{1,\dots,n\}$, we have $y_i \geq 0$. Since $i \in \{1,\dots,n\}$ is arbitrary, $y_i \geq 0$ for all $i \in \{1,\dots,n\}$.

Second, we will show $\sum_{i=1}^n y_i = 1$. Consider

$$\sum_{i=1}^n y_i = \sum_{i=1}^n \sum_{j=1}^n G_{ij} x_j \tag{4}$$
$$= \sum_{j=1}^n \sum_{i=1}^n G_{ij} x_j \tag{5}$$
$$= \sum_{j=1}^n x_j \sum_{i=1}^n G_{ij} \tag{6}$$
$$= \sum_{j=1}^n x_j \tag{7}$$
$$= 1, \tag{8}$$

where in (5) we interchanged the order of summation, in (6) we used distributive property, in (7) we used $\sum_{i=1}^n G_{ij} = 1$ for all $j \in \{1, \dots, n\}$ and finally (8) holds since x is a stochastic vector.

- Every column Ge_j of a stochastic matrix G is a stochastic vector. The jth column of FG is $F(Ge_j)$. Since F is a stochastic matrix and Ge_j is a stochastic vector, the previous part shows that $F(Ge_j)$ is a stochastic vector. Thus, FG is a stochastic matrix.

Problem 3 (20 points). Given vectors $x, y, z \in \mathbb{R}^n$, the identities $xy^T z = (xy^T)z = x(y^T z)$ hold. To find out which is the better formula to use in practice, implement all three formulas in MATLAB and time which one executes the fastest. MATLAB's `tic` and `toc` can be used for timing. Generate a table and a figure with timing results. Explain the difference in timing results.

Solution Students code structure should look like

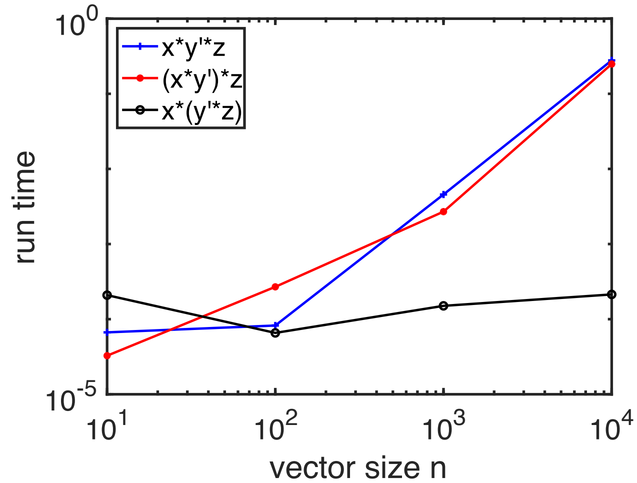
```
n = 1;
while n < 1.e4
    % vectors
    n        = 10*n;
    x = rand(n,1);   y = rand(n,1);   z = rand(n,1);

    result1 = x*y'*z;
    time1 = toc;
    :
    table = [table;
             n, time1, time2, time3];
end
```

See the MATLAB code `matrix_mult_hw.m`
The timing table should look like

n	time1	time2	time3
10	7.51e-05	3.77e-05	1.93e-04
100	1.06e-04	2.49e-04	4.88e-05
1000	1.05e-03	1.21e-03	8.94e-05
10000	2.42e-01	2.20e-01	1.86e-04

The timing figure should look like



Matlab executes from left to right. Thus, in Matlab $xy^Tz = (xy^T)z$. The large difference in timing between $(xy^T)z$ and $x(y^Tz)$ is due to the fact that xy^T is an $n \times n$ matrix and $(xy^T)z$ requires a matrix-vector multiplication at $O(n^2)$ operations (multiplications, additions). On the other hand, y^Tz is a scalar (its computation requires $O(n)$ operations) and $x(y^Tz)$ is a scalar multiplication of a vector, which are requires $O(n)$ operations.

For small sizes n execution times are so small that other tasks impact the time needed to compute xy^Tz, etc.

Computer software such as MATLAB and Wolfram Alpha are not permitted on this part.

Problem 4 (50 points).
Suppose

$$u = \begin{pmatrix} 1 \\ a \end{pmatrix}, \quad v = \begin{pmatrix} 2 \\ 3 \end{pmatrix}, \quad A = \begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} -1 & b \\ 0 & -1 \end{pmatrix}$$

1. Since $u+v = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$, we have $a+3=4$ and $a=1$.

2. Since $u^T v = 5$, we have $2+3a=5$ and $a=1$.

3. Since $\|u\| = \sqrt{10}$, we have $\sqrt{1+a^2} = \sqrt{10}$ and $a^2 = 9$. Note that $a \geq 0$ by assumption, so $a = 3$.

4. Since $Au = \begin{pmatrix} -5 \\ 3 \end{pmatrix}$, we have $\begin{pmatrix} 1-2a \\ a \end{pmatrix} = \begin{pmatrix} -5 \\ 3 \end{pmatrix}$ and $a=3$.

5. Since $uu^T = \begin{pmatrix} 1 & -2 \\ -2 & 4 \end{pmatrix}$, we have $\begin{pmatrix} 1 & a \\ a & a^2 \end{pmatrix} = \begin{pmatrix} 1 & -2 \\ -2 & 4 \end{pmatrix}$ and $a=-2$.

6. Since $AB = \begin{pmatrix} -1 & b+2 \\ 0 & -1 \end{pmatrix}$, $\begin{pmatrix} -1 & b+2 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} -1 & 3 \\ 0 & -1 \end{pmatrix}$ implies $b=1$.

7. Since $B^2 = \begin{pmatrix} 1 & -2b \\ 0 & 1 \end{pmatrix}$, $\begin{pmatrix} 1 & -2b \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 4 \\ 0 & 1 \end{pmatrix}$ and $b=-2$.

8. Since $Bv = \begin{pmatrix} -2+3b \\ -3 \end{pmatrix}$, $\begin{pmatrix} -2+3b \\ -3 \end{pmatrix} = \begin{pmatrix} 4 \\ -3 \end{pmatrix}$ and $b=2$.

9. Since $BB^T = \begin{pmatrix} -1 & b \\ 0 & -1 \end{pmatrix}\begin{pmatrix} -1 & 0 \\ b & -1 \end{pmatrix} = \begin{pmatrix} 1+b^2 & -b \\ -b & 1 \end{pmatrix}$, we have $\operatorname{tr}(BB^T) = 2+b^2$. It is given that $\operatorname{tr}(BB^T) = 3$ and $b \geq 0$, so $b=1$.

10. Since $b = \sqrt{2}$, we have $\|B\|_F = \sqrt{\sum_{i=1}^2 \sum_{j=1}^2 B_{ij}} = \sqrt{1+2+1} = 2$.