Homework 1

2.61.)

We know that for $p(x) = a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + ... + a_1 x + a_0$ the derivative is

$$\frac{dp(x)}{dx} = (n-1)a_{n-1}x^{n-2} + (n-2)a_{n-2}x^{n-3} + ... + a_1$$

As a vector, these coefficients would be $(a_1, 2a_2, 3a_3, ..., (n-2)a_{n-2}, (n-1)a_{n-1}) \in \mathbb{R}^{n-1}$. The corresponding matrix $D \in \mathbb{R}^{(n-1)\times n}$ would thus be

$$D = \begin{bmatrix} 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 2 & 0 & \dots & 0 \\ 0 & 0 & 0 & 3 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & n-1 \end{bmatrix}$$

2.100.)

a.

To find $A \in \mathbb{R}^{2\times n}$ we first need to calculate the velocity at time n, $v(n)$, and the position at time n, $p(n)$, in terms of $x_1, x_2, .., x_n$. First, we observe that since we're working with a unit mass subject and $f(t) = ma(t)$ then $f(t) = a(t)$ which means the acceleration of the subject has the form $a(t) = x_j$ for $j-1 \leq t < j$ and $j = 1, ..., n$. Using this, we can calculate $v(n)$:

$$v(n) = \int_0^n a(t)dt = \sum_{j=1}^{n} \int_{j-1}^{j} a(t)dt = \sum_{j=1}^{n} \int_{j-1}^{j} x_j dt = \sum_{j=1}^{n} x_j \int_{j-1}^{j} dt = \sum_{j=1}^{n} x_j = x_1 + x_2 + ... + x_n$$

Calculating $p(n)$ is a bit trickier. We first solve for a formula for $v(t)$. We can think about this piecewise

$$v(t) = \begin{cases} \int_0^t a(t)dt & 0 \leq t < 1 \\ \int_0^t a(t)dt & 1 \leq t < 2 \\ \vdots & \vdots \\ \int_0^t a(t)dt & n-1 \leq t < n \end{cases} = \begin{cases} \int_0^t x_1 dt & 0 \leq t < 1 \\ x_1 + \int_1^t x_2 dt & 1 \leq t < 2 \\ \vdots & \vdots \\ x_1 + x_2 + ... + x_{n-1} + \int_{n-1}^t x_n dt & n-1 \leq t < n \end{cases}$$

$$v(t) = \begin{cases} x_1(t-0) & 0 \leq t < 1 \\ x_1 + x_2(t-1) & 1 \leq t < 2 \\ \vdots & \vdots \\ x_1 + x_2 + ... + x_{n-1} + x_n(t-(n-1)) & n-1 \leq t < n \end{cases}$$

We can now solve for $p(n)$:

$$p(n) = \int_0^n v(t)dt = \int_0^1 v(t)dt + \int_1^2 v(t)dt + ... + \int_{n-1}^n v(t)dt$$

$$= \int_0^1 x_1(t-0)dt + \int_1^2 x_1 + x_2(t-1)dt + ... + \int_{n-1}^n x_1 + x_2 + ... + x_{n-1} + x_n(t-(n-1))dt$$

$$= (0.5x_1) + (x_1 + 0.5x_2) + ... + (x_1 + x_2 + ... + x_{n-1} + 0.5x_n)$$

$$= ((n-1) + 0.5)x_1 + ((n-2) + 0.5)x_2 + ... + ((n-n) + 0.5)x_n$$

Since we've solved for $v(n)$ and $p(n)$ in terms of $x_1, ..., x_n$, we now have the matrix A:

$$A = \begin{bmatrix} ((n-1)+0.5) & ((n-2)+0.5) & ... & ((n-n)+0.5) \\ 1 & 1 & ... & 1 \end{bmatrix}$$

b.

This question is effectively asking us to find a solution for

$$\begin{bmatrix}1\\0\end{bmatrix} = A\begin{bmatrix}x_1\\x_2\\x_3\\x_4\end{bmatrix} = \begin{bmatrix}3.5 & 2.5 & 1.5 & 0.5\\1 & 1 & 1 & 1\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\\x_4\end{bmatrix}$$

One possible solution is $(x_1, x_2, x_3, x_4) = (2, -2, -1, 1)$:

$$\begin{bmatrix}3.5 & 2.5 & 1.5 & 0.5\\1 & 1 & 1 & 1\end{bmatrix}\begin{bmatrix}2\\-2\\-1\\1\end{bmatrix} = \begin{bmatrix}2(3.5) - 2(2.5) - 1(1.5) + 1(0.5)\\2 - 2 - 1 + 1\end{bmatrix} = \begin{bmatrix}1\\0\end{bmatrix}$$

2.110.)

For $B = A^k$ as defined in the problem, B_{ij} gives the number of distinct paths of length k from node i to node j. We can prove this with induction on k. This is true for $k = 1$ by construction of A, as you can have at most one distinct path of length one connecting two nodes, and $A_{ij} = 1$ when this path exists and $A_{ij} = 0$ when it doesn't.

Fix $z \in \mathbb{N}$ (n is number of rows of A so choosing z instead). Assume that for $B = A^z$ that B_{ij} gives the number of distinct paths of length z from node i to node j. We want to show that for $B' = A^{z+1}$ that B'_{ij} gives the number of distinct paths of length $z+1$ from node i to node j. Since $B' = A^{z+1} = A^z A = BA$, we have

$$B'_{ij} = B_{i1}A_{1j} + B_{i2}A_{2j} + ... + B_{in}A_{nj}$$

Every path of length $z+1$ from node i to node j can be split into a path of length z from node i to some intermediate node, node t, and a path of length 1 from node t to node j. In the equation above, B_{it} (for $t = 1, ..., n$) gives the number of distinct paths of length z from node i to node t and A_{tj} gives the number of distinct paths of length 1 from node t to node j. Multiplying them together gives the number of distinct paths of length $z+1$ from node i to node j where node t is the immediate predecessor to node j in the path. Each summand in the equation represents one of these sets of paths, and adding all of them together gives the number of distinct paths of length $z+1$ from node i to node j where any node can be the immediate predecessor to node j in the path. Since any node can precede node j, this means B'_{ij} gives the number of paths of length $z+1$ from node i to node j. We've proven the base case and inductive step. Hence, for $B = A^k$ as defined in the problem, B_{ij} gives the number of distinct paths of length k from node i to node j.

2.150.)

a.

We know that

$$f(x) = a^T x + b = \begin{bmatrix}a_1\\a_2\\a_3\\\vdots\\a_n\end{bmatrix}^T \begin{bmatrix}x_1\\x_2\\x_3\\\vdots\\x_n\end{bmatrix} + b = a_1x_1 + a_2x_2 + ... + a_nx_n + b$$

We can then calculate the gradient below

$$\nabla f(x) = \begin{bmatrix}\frac{\partial f(x)}{\partial x_1}\\\frac{\partial f(x)}{\partial x_2}\\\frac{\partial f(x)}{\partial x_3}\\\vdots\\\frac{\partial f(x)}{\partial x_n}\end{bmatrix} = \begin{bmatrix}a_1\\a_2\\a_3\\\vdots\\a_n\end{bmatrix}$$

b.

We know that

$$f(x) = x^T A x = \begin{bmatrix} x_1 & x_2 & x_3 & ... & x_n \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} & ... & a_{1n} \\ a_{21} & a_{22} & a_{23} & ... & a_{2n} \\ a_{31} & a_{32} & a_{33} & .. & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & ... & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix}$$

$$= \begin{bmatrix} x_1 & x_2 & x_3 & ... & x_n \end{bmatrix} \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + ... + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + ... + a_{2n}x_n \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + ... + a_{3n}x_n \\ \vdots \\ a_{n1}x_1 + a_{n2}x_2 + a_{n3}x_3 + ... + a_{nn}x_n \end{bmatrix} = a_{11}x_1x_1 + a_{12}x_1x_2 + a_{13}x_1x_3 + ... + a_{1n}x_1x_n$$

$$+ a_{21}x_2x_1 + a_{22}x_2x_2 + a_{23}x_2x_3 + ... + a_{2n}x_2x_n + ... + a_{n1}x_nx_1 + a_{n2}x_nx_2 + a_{n3}x_nx_3 + ... + a_{nn}x_nx_n$$

We can then calculate the gradient below

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \\ \frac{\partial f(x)}{\partial x_3} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix} = \begin{bmatrix} 2a_{11}x_1 + (a_{12} + a_{21})x_2 + (a_{13} + a_{31})x_3 + ... + (a_{1n} + a_{n1})x_n \\ (a_{21} + a_{12})x_1 + 2a_{22}x_2 + (a_{23} + a_{32})x_3 + ... + (a_{2n} + a_{n2})x_n \\ (a_{31} + a_{13})x_1 + (a_{32} + a_{23})x_2 + 2a_{33}x_3 + ... + (a_{3n} + a_{n3})x_n \\ \vdots \\ (a_{n1} + a_{1n})x_1 + (a_{n2} + a_{2n})x_2 + (a_{n3} + a_{3n})x_3 + ... + 2a_{nn}x_n \end{bmatrix}$$

c.

This question is the same as the last except for all i, j s.t. $1 \leq i, j \leq n$, $a_{ij} = a_{ji}$. This lets us simplify the answer for part b a bit. The gradient using this new assumption is

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \\ \frac{\partial f(x)}{\partial x_3} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix} = \begin{bmatrix} 2a_{11}x_1 + 2a_{12}x_2 + 2a_{13}x_3 + ... + 2a_{1n}x_n \\ 2a_{21}x_1 + 2a_{22}x_2 + 2a_{23}x_3 + ... + 2a_{2n}x_n \\ 2a_{31}x_1 + 2a_{32}x_2 + 2a_{33}x_3 + ... + 2a_{3n}x_n \\ \vdots \\ 2a_{n1}x_1 + 2a_{n2}x_2 + 2a_{n3}x_3 + ... + 2a_{nn}x_n \end{bmatrix}$$

2.210.)

a.

$Z = RY$ where R is an upper triangular matrix ($r_{ij} = 0$ if $i > j$). To prove this is equivalent, first consider an arbitrary entry of Z, z_{ij}. From our formula for Z, we have

$$z_{ij} = r_{i1}y_{1j} + r_{i2}y_{2j} + ... + r_{in}y_{nj}$$

Notice that if we fix i, then this holds true for $j = 1, ..., n$. Then we have

$$\begin{bmatrix} z_{i1} \\ z_{i2} \\ \vdots \\ z_{in} \end{bmatrix} = r_{i1} \begin{bmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1n} \end{bmatrix} + r_{i2} \begin{bmatrix} y_{21} \\ y_{22} \\ \vdots \\ y_{2n} \end{bmatrix} + ... + r_{in} \begin{bmatrix} y_{n1} \\ y_{n2} \\ \vdots \\ y_{nn} \end{bmatrix}$$

All of the vectors above are transposed row vectors of either Z or Y. In other words, the row vectors of Z are linear combinations of the row vectors of Y. Also note that $r_{ij} = 0$ for $i > j$. We can then simplify the above formula to

$$\begin{bmatrix} z_{i1} \\ z_{i2} \\ \vdots \\ z_{in} \end{bmatrix} = r_{ii} \begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{in} \end{bmatrix} + r_{i(i+1)} \begin{bmatrix} y_{(i+1)1} \\ y_{(i+1)2} \\ \vdots \\ y_{(i+1)n} \end{bmatrix} + ... + r_{in} \begin{bmatrix} y_{n1} \\ y_{n2} \\ \vdots \\ y_{nn} \end{bmatrix}$$

Thus, row i of Z is a linear combination of rows $i, ..., n$ of Y. As i was arbitrary, this holds for each row i of Z, proving the matrix language is equivalent.

b.

I'm assuming this question means if $V = \begin{bmatrix} v_1 & v_2 & v_3 & v_4 & ... \end{bmatrix}$ then $W = \begin{bmatrix} v_2 & v_1 & v_4 & v_3 & ... \end{bmatrix}$. To get this, we have $W = VA$ where

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & ... \\ 1 & 0 & 0 & 0 & 0 & 0 & ... \\ 0 & 0 & 0 & 1 & 0 & 0 & ... \\ 0 & 0 & 1 & 0 & 0 & 0 & ... \\ 0 & 0 & 0 & 0 & 0 & 1 & ... \\ 0 & 0 & 0 & 0 & 1 & 0 & ... \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

This is because each column of A corresponds to a column of W and each column of A is a unit vector, so multiplying V by an individual column of A returns an individual column of V. In other words,

$$V \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ \vdots \end{bmatrix} = \begin{bmatrix} v_1 & v_2 & v_3 & v_4 & ... \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ \vdots \end{bmatrix} = v_2$$

So the alternating pattern in A permutes the columns of V for W.

c.

Every entry of $P^T Q$ is positive. This is because for every (P_i, Q_j) pair of columns, where P_i is an arbitrary column of P and Q_j is an arbitrary column of Q, there is a corresponding entry in $P^T Q$ which is the dot product of the two columns. When the dot product of two vectors is positive, that means the angle between them is acute. Hence, all of the matrix entries being positive means every (P_i, Q_j) pair of columns has a positive dot product, so every (P_i, Q_j) pair of columns has an acute angle.

d.

The diagonal entries of $P^T Q$ are positive. This is very similar logic to part c, except the diagonal entries are representative of the corresponding column pairs (P_i, Q_i) which is what this question wants. Note that $P^T Q$ isn't necessarily a square matrix as P and Q can have a different number of columns, so there might not be a corresponding column of Q for every column of P or vice versa. Regardless, the diagonal entries being positive (diagonal starting from the top left entry) is still sufficient.

e.

For $B = A^T A$, $B_{ij} = 0$ if $i \leq k$ and $j > k$ or if $i > k$ and $j \leq k$. Once again, this uses similar logic to part c with entries in $A^T A$ being dot products of column pairs of A (this time both columns being from A). B_{ij} is the dot product between column i of A and column j of A. So if $i \leq k$ and $j > k$ or if $i > k$ and $j \leq k$ then we have one column from the first k columns of A and one column from the remaining columns. When the dot product of two vectors is 0, those vectors are orthogonal to each other. Hence, $B_{ij} = 0$ if $i \leq k$ and $j > k$ or if $i > k$ and $j \leq k$ means that every pair of columns of A, where one of the columns if from the first k columns and the other column is from the remaining columns, is orthogonal to each other. Thus, this condition holding means the first k columns of A are orthogonal to the remaining columns of A.