Preface: Problem dimensions & specifics will not be defined, nor will all problems be restated. All assumptions are already defined in the homework document. For example, I will not define $v \in \mathbb{R}^n$, since it is already defined in the homework.

1. $trace()$ is defined by the sum of the diagonal elements of a matrices $\in \mathbb{R}^{n \times n}$

   (a) $v^\top v$ gives us

   $$v^\top v = \sum_{i=1}^{n} v_i^2$$

   $vv^\top$ gives us

   $$vv^\top = \begin{bmatrix} v_1v_1 & \dots & v_1v_n \\ \vdots & \ddots & \vdots \\ v_nv_1 & \dots & v_nv_n \end{bmatrix}$$

   $trace(vv^\top)$ results in the summation of diagonal elements, where row indices equal column indices

   $$trace(vv^\top) = \sum_{i=1}^{n} v_i^2$$

   $$\therefore v^\top v = trace(vv^\top)$$

   (b) AB gives us

   $$AB = \begin{bmatrix} \sum_{i=1}^{n} A_{1i}B_{i1} & & \\ & \ddots & \\ & & \sum_{i=1}^{n} A_{ni}B_{in} \end{bmatrix}$$

   so

   $$trace(AB) = \sum_{j=1}^{n} \sum_{i=1}^{n} A_{ji}B_{ij}$$

   similarly, $trace(BA)$ will give us

   $$trace(BA) = \sum_{j=1}^{n} \sum_{i=1}^{n} B_{ji}A_{ij}$$

   Both variables i and j range from $1 \dots n$, so they are interchangeable. If we swap them for $trace(BA)$,

   $$\sum_{j=1}^{n} \sum_{i=1}^{n} B_{ji}A_{ij} = \sum_{i=1}^{n} \sum_{j=1}^{n} B_{ij}A_{ji}$$

   Utilizing the commutative property of multiplication & summation,

   $$\sum_{j=1}^{n} \sum_{i=1}^{n} A_{ji}B_{ij} = \sum_{i=1}^{n} \sum_{j=1}^{n} B_{ij}A_{ji}$$

   $$\therefore trace(AB) = trace(BA)$$

2. Let's say $M = I$, $v^\top M v = v^\top I v = v^\top v$. We know that

$$v^\top v = \sum_k = v_k^2$$

To get to $\sum_k \omega_k v_k^2$, we must scale up M by each element of ω. As a result, M is an $n \times n$ diagonal matrix of ω.

$$M = \begin{bmatrix} \omega_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \omega_n \end{bmatrix}$$

I don't know how to transform a $n \times 1$ vector ω into an $n \times n$ matrix M using matrix operations, but I am sure that the solution above results in the proper summation.

3.

$$A = \begin{bmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{bmatrix}, x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

so that

$$Ax = b$$

To solve for x, b is pre-multiplied with A^{-1}, so that

$$x = A^{-1} b$$

4.

$$QP = \begin{bmatrix} 2 & -3 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 2 & 3 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$A = P \Lambda Q = \begin{bmatrix} 2 & 3 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 2 & -3 \\ -1 & 2 \end{bmatrix} = \begin{bmatrix} 7 & -12 \\ 4 & -7 \end{bmatrix}$$

To solve for A^8 and A^9, lets solve for A^2:

$$A^2 = \begin{bmatrix} 7 & -12 \\ 4 & -7 \end{bmatrix} \begin{bmatrix} 7 & -12 \\ 4 & -7 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I_2$$

Given that $A^2 = I_2$, $A^{2n} = I_2^n = I_2$, so that

$$A^{2n} = I_2$$

Similarly, $A^{2n+1} = A^{2n} A = I_2 A = A$, so that

$$A^{2n+1} = A$$

meaning

$$A^8 = I_2, A^9 = A$$

and

$$A^n = \begin{cases} I_2, & n \bmod 2 = 0 \\ A, & n \bmod 2 = 1 \end{cases}$$

5.
$$\begin{aligned}(A^{-1})^\top &= (A^{-1})^\top I \\ &= (A^{-1})^\top [A^\top (A^\top)^{-1}] \\ &= [(A^{-1})^\top A^\top](A^\top)^{-1} \\ &= [AA^{-1}]^\top (A^\top)^{-1} \\ &= I^\top (A^\top)^{-1} \\ &= I(A^\top)^{-1} \\ &= (A^\top)^{-1}\end{aligned}$$

6.
$$v_1 = \begin{bmatrix}1\\1\\1\end{bmatrix}, v_2 = \begin{bmatrix}0\\2\\5\end{bmatrix}, v_3 = \begin{bmatrix}1\\3\\6\end{bmatrix}$$

Using an analytical approach, the vectors above are linearly dependent since $v_1 + v_2 = v_3$, meaning a vector space can be described by only using vectors v_1 and v_2 since v_3 is dependent on v_1 and v_2. Numeric approach:

$$\left[\begin{array}{ccc|c} 1 & 0 & 1 & 0 \\ 1 & 2 & 3 & 0 \\ 1 & 5 & 6 & 0 \end{array}\right] \rightarrow \begin{bmatrix}1 & 0 & 1\\0 & 2 & 2\\0 & 5 & 5\end{bmatrix} \rightarrow \begin{bmatrix}1 & 0 & 1\\0 & 1 & 1\\0 & 0 & 0\end{bmatrix}$$

7. I will be using elementary row operations on A and B to reduce them to row-echelon form in order to find $rank(A)$ and $rank(B)$.

$$A = \begin{bmatrix}1 & 2 & 3 & 4\\0 & 0 & -1 & -2\\0 & 0 & 0 & 4\\0 & 1 & 2 & -1\end{bmatrix}$$

$$\begin{bmatrix}1 & 2 & 3 & 4\\0 & 0 & -1 & -2\\0 & 0 & 0 & 4\\0 & 1 & 2 & -1\end{bmatrix} \rightarrow \begin{bmatrix}1 & 2 & 3 & 4\\0 & 1 & 2 & -1\\0 & 0 & -1 & -2\\0 & 0 & 0 & 4\end{bmatrix} \rightarrow \begin{bmatrix}1 & 0 & -1 & 6\\0 & 1 & 2 & -1\\0 & 0 & 1 & 0\\0 & 0 & 0 & 1\end{bmatrix} \rightarrow \begin{bmatrix}1 & 0 & 0 & 0\\0 & 1 & 0 & 0\\0 & 0 & 1 & 0\\0 & 0 & 0 & 1\end{bmatrix}$$

$$\therefore rank(A) = 4$$

$$B = \begin{bmatrix}3 & 2 & -1 & -3 & -2\\2 & -1 & 3 & 1 & -3\\4 & 5 & -5 & -6 & 1\end{bmatrix}$$

$$\begin{bmatrix}3 & 2 & -1 & -3 & -2\\2 & -1 & 3 & 1 & -3\\4 & 5 & -5 & -6 & 1\end{bmatrix} \rightarrow \begin{bmatrix}1 & 3 & -4 & -4 & 1\\2 & -1 & 3 & 1 & -3\\0 & 7 & -11 & -8 & 5\end{bmatrix} \rightarrow \begin{bmatrix}1 & 3 & -4 & -4 & 1\\0 & -7 & 11 & 9 & -5\\0 & 7 & -11 & -8 & 5\end{bmatrix} \rightarrow$$
$$\begin{bmatrix}1 & 3 & -4 & 0 & 1\\0 & 7 & -11 & 0 & 5\\0 & 0 & 0 & 1 & 0\end{bmatrix} \rightarrow \begin{bmatrix}1 & 0 & \frac{5}{7} & 0 & \frac{-8}{7}\\0 & 1 & \frac{-11}{7} & 0 & \frac{5}{7}\\0 & 0 & 0 & 1 & 0\end{bmatrix}$$

$$\therefore rank(B) = 3$$

8. Gradient with respect to x of $f(x) = e^{x^\top Ax+b}$ is $\frac{df(x)}{d\vec{x}} = 2x^\top Ae^{x^\top Ax+b}$. In question 10, these matrix derivative properties are somewhat derived.

9. $f(x) = \frac{1}{2}x^\top P^\top Px + q^\top x + r$

   (a) $\frac{df}{d\vec{x}}(x) = x^\top P^\top P + q^\top$

   (b) As the problem states, we can show that the Hessian of the matrix is $P^\top P$, therefore all positive values. This means the function's slope is increasing at a constant rate, meaning that there exists a minimum. That minimum is when the first derivative equals 0.

$$\frac{df}{d\vec{x}}(x) = x^\top P^\top P + q^\top = 0$$
$$x^\top P^\top P = -q^\top$$
$$P^\top Px = -q$$
$$Px = -(P^\top)^{-1}q$$
$$x = -(P^\top P)^{-1}q$$

10. Given $f(x) = x^\top Ax$, we must show $\frac{df}{d\vec{x}}(x) = 2x^\top A$

$$f(x) = x^\top Ax = x^\top \begin{bmatrix} \sum_i^n A_{1i}x_i \\ \vdots \\ \sum_i^n A_{ni}x_i \end{bmatrix}$$

$$f(x) = \sum_j^n \sum_i^n A_{ji}x_ix_j$$

So let's find the partial derivative with respect to a general index k

$$\frac{\partial f}{\partial x_k} = \frac{\partial}{\partial x_k}\sum_j^n \sum_i^n A_{ji}x_ix_j$$

Discretely, there are 4 different partial derivatives segments that must added together. If i & j both equal k, then

$$\frac{\partial}{\partial x_k}\sum_j^n \sum_i^n A_{ji}x_ix_j = 2A_{kk}x_k \mid i,j = k$$

If i equals k but j does not, then

$$\frac{\partial}{\partial x_k}\sum_j^n \sum_i^n A_{ji}x_ix_j = \sum_{j\neq k}^n A_{jk}x_j \mid i = k, j \neq k$$

Likewise, if i does not equal k but j does, then

$$\frac{\partial}{\partial x_k}\sum_j^n \sum_i^n A_{ji}x_ix_j = \sum_{i\neq k}^n A_{ki}x_i \mid i \neq k, j = k$$

And if neither i nor j equal k, then

$$\frac{\partial}{\partial x_k}\sum_j^n\sum_i^n A_{ji}x_ix_j = 0 \mid i,j \neq k$$

Together,

$$\forall i,j,k : \frac{\partial}{\partial x_k}\sum_j^n\sum_i^n A_{ji}x_ix_j = 2A_{kk}x_k + \sum_{j\neq k}^n A_{jk}x_j + \sum_{i\neq k}^n A_{ki}x_i + 0$$

Let's ease the restriction on the summations, so that both sets can add from 1 to n, while keeping the solution equivalent

$$= 2A_{kk}x_k + (\sum_j^n A_{jk}x_j - A_{kk}x_x) + (\sum_i^n A_{ki}x_i - A_{kk}x_x)$$

$$= \sum_j^n A_{jk}x_j + \sum_i^n A_{ki}x_i$$

Since both i and j range from 1 to n, we can reduce the amount of variables

$$= \sum_i^n A_{ik}x_i + \sum_i^n A_{ki}x_i$$

$$= \sum_i^n A_{ik}x_i + A_{ki}x_i$$

$$= \sum_i^n (A_{ik} + A_{ki})x_i$$

This was the case for any index k in the range 1 to n. But to take every partial derivative into account, we must sum all partial derivatives together and format it into the following form

$$\frac{df}{d\vec{x}}(x) \triangleq \begin{bmatrix} \frac{\partial f}{\partial x_1} & \dots & \frac{\partial f}{\partial x_n} \end{bmatrix}$$

So

$$\sum \frac{df}{d\vec{x}}(x) = \sum_k^n\sum_i^n (A_{ik} + A_{ki})x_i$$

And since

$$x^\top A = \begin{bmatrix} \sum_i^n A_{i1}x_1 & \dots & \sum_i^n A_{in}x_n \end{bmatrix}$$

then

$$x^\top(A + A^\top) = \begin{bmatrix} \sum_i^n (A_{i1} + A_{1i})x_1 & \dots & \sum_i^n (A_{in} + A_{ni})x_n \end{bmatrix}$$

$$\sum x^\top(A + A^\top) = \sum_k^n\sum_i^n (A_{ik} + A_{ki})x_i$$

$$\therefore \frac{df}{d\vec{x}}(x) = x^\top(A + A^\top)$$

And since A is a symmetric matrix, $\frac{df}{d\vec{x}}(x) = 2x^\top A$

11. $\frac{df(x(s))}{d\vec{s}}$ is equivalent to $\frac{df}{d\vec{x}}\frac{d\vec{x}}{d\vec{s}}$ due to the chain rule.
First solve for $\frac{df}{d\vec{x}}$

$$f(x) = x_1 x_2^2$$

$$\frac{df}{d\vec{x}} = \begin{bmatrix} x_2^2 & 2x_1x_2 \end{bmatrix}$$

Then solve for $\frac{d\vec{x}}{d\vec{s}}$

$$x(s) = \begin{bmatrix} 3s_1 + s_2 \\ s_1 + 5s_2 \end{bmatrix} = \begin{bmatrix} 3 & 1 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix}$$

$$\frac{d\vec{x}}{d\vec{s}} = \begin{bmatrix} 3 & 1 \\ 1 & 5 \end{bmatrix}$$

$$\therefore \frac{df(x(s))}{d\vec{s}} = \begin{bmatrix} x_2^2 & 2x_1x_2 \end{bmatrix} \begin{bmatrix} 3 & 1 \\ 1 & 5 \end{bmatrix} = \begin{bmatrix} 3x_2^2 + 2x_1x_2 & x_2^2 + 10x_1x_2 \end{bmatrix}$$

where

$$x_1 = 3s_1 + s_2$$

$$x_2 = s_1 + 5s_2$$