Problem 1a

For example with $n = 3$, vv^T is

$$\begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix}$$

Produces a $3x3$ matrix in which the diagonal or trace is $v_1^2 + v_2^2 + v_3^2$
And $v^T v$ is $v_1^2 + v_2^2 + v_3^2$.
Any vector multiplied by its own transpose to get a matrix A will contain a square of the i^{th} term for all i up to n for every element A_{ii}. The trace will be the sum of these squares.
Any transpose of a vector multiplied by the vector will produce a scalar. The sum that produces this scalar will be $\sum_{i=1}^{n} v_i^2$ which is the same as the trace of the above matrix A.

Problem 1b

The trace of AB will be the sum of the dot product of the i^{th} row in A and the j^{th} column in B $\sum_{i=1,j=1}^{n} a_i b_j$ with $i = j$.

You can write out the entire sum for the trace as
$A_{11}B_{11} + A_{12}B_{21} + A_{13}B_{13} + ...$
$A_{21}B_{12} + A_{22}B_{22} + A_{23}B_{32} + ...$
$A_{31}B_{13} + A_{32}B_{23} + A_{33}B_{33} + ...$
...

If you look at the columns here, they would be the same terms you would get if you were calculating all the diagonal entries for BA. Therefore $tr(AB) = tr(BA)$

Problem 2

Using the formula for block matrix inversion

$$\begin{bmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{bmatrix}^{-1} = \begin{bmatrix} A_{11}^{-1} + A_{11}^{-1}0(A_{22} - A_{21}A_{11}^{-1}0)^{-1}A_{21}A_{11}^{-1} & -A_{11}^{-1}0(A_{22} - A_{21}A_{11}^{-1}0)^{-1} \\ -(A_{22} - A_{21}A_{11}^{-1}0)^{-1}A_{21}A_{11}^{-1} & (A_{22} - A_{21}A_{11}^{-1}0)^{-1} \end{bmatrix}$$

$$= \begin{bmatrix} A_{11}^{-1} & 0 \\ -(A_{22})^{-1}A_{21}A_{11}^{-1} & A_{22}^{-1} \end{bmatrix}$$

It works out so that we don't have to take the inverse of A_{21}. Now we have

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} A_{11}^{-1} & 0 \\ -(A_{22})^{-1}A_{21}A_{11}^{-1} & A_{22}^{-1} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

Problem 3

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

Q and P are inverses since they are square matrices and their product is I. So A can be expressed as $A = P\Delta P^{-1}$, which will be a 2x2 matrix. This is the eigendecomposition of A, with Δ consisting of linearly independent eigenvectors.

$$A^n = P\Delta P^{-1}P\Delta P^{-1}...$$

The inner P and P^{-1}'s all cancel so

$$A^n = P\Delta^n P^{-1} = A^n = P\Delta^n Q$$

If n is even then $\Delta^n = I$, if n is odd then $\Delta^n = \Delta$

$$A^8 = P\Delta^8 Q = I$$

$$A^9 = P\Delta^9 Q = \begin{bmatrix} 2 & -3 \\ 1 & -2 \end{bmatrix} Q = \begin{bmatrix} 7 & -12 \\ 4 & -7 \end{bmatrix}$$

Problem 4

$A^T(A^{-1})^T = (A^{-1}A)^T = I^T = I$
$(A^{-1})^T A^T = (AA^{-1})^T = I^T = I$
The first equality in each line uses $A^T B^T = (BA)^T$
This shows that the inverse of A^T is $(A^{-1})^T$ by the definition of the inverse of a matrix which is $AB = BA = I$

Problem 5

They are not linearly independent because $v_1 = -v_2 + v_3$
n vectors are said to be linearly independent if the only $c_1, c_2, .., c_n$ solving the equation $0 = c_1v_1 + c_2v_2 + ... + c_nv_n$ are $c_1 = c_2 = ... = c_n = 0$, but that is not the case here.

Problem 6

By swapping rows Matrix A can be rewritten as

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

Each row is linearly independent so the rank is 4. When in row echelon form like above the number of rows with a non zero coefficient is the number of linearly independent rows. The rank of a matrix is the number of linearly independent rows or columns.

For B add $-2R_2$ to R_3, then add $-1.5R_2$ to R_1, then add $-2R_1$ to R_3, then swap R_1 and R_2 to get

$$\begin{bmatrix} 2 & -1 & 3 & 1 & -3 \\ 0 & 3.5 & -5.5 & -4.5 & 2.5 \\ 0 & 0 & 0 & 1 & 2 \end{bmatrix}$$

It is now in row echelon form with 3 linearly independent rows, so the rank is 3. The max rank of a m x n matrix is $min(m, n)$

Problem 7a

$\sum_i x_i^2 = x^T x$. Let $f(x) = x^T x$, using the product rule you get

$$\frac{d}{d_{x_1}}x, \frac{d}{d_{x_2}}x, ... = x$$

for the derivative of x^T wrt x multiplied by x. Because $\frac{d}{d_{x_1}} = (1, 0, 0, ...)$, so $\frac{d}{d_{x_1}}x = x_1$. You get the same when treating x^T as a constant so $\frac{df}{dx} = 2x$
Now with $g(x) = \sqrt{f(x)}$ by the chain rule you get

$$\frac{dg}{dx} = \frac{x}{\sqrt{x^T x}}$$

Problem 7b

Let $f(x) = x^T Ax + b$, the gradient of $g(x) = e^{f(x)}$ will be $\frac{df}{dx} e^{f(x)}$
$\frac{df}{dx} = \begin{bmatrix} \frac{\partial f}{\partial x^T} & \frac{\partial f}{\partial x} \end{bmatrix} \begin{bmatrix} \frac{\partial x^T}{\partial x^T} \\ \frac{\partial x}{\partial x} \end{bmatrix}$ by the chain rule
$\frac{\partial f}{\partial x^T}$ treats Ax as a constant and a rule of vector differentiation is $\frac{d(x^T a)}{dx} = a^T$, and $(Ax)^T = x^T A^T$
So $\frac{\partial f}{\partial x^T} = x^T A^T$
$\frac{\partial f}{\partial x} = x^T A$, treating $x^T A$ as a constant.
The partial derivatives in the column vector above are both 1. Therefore $\frac{df}{dx} = x^T A^T + x^T A = x^T(A^T + A)$

So $\frac{dg}{dx} = x^T(A^T + A)e^{x^T Ax+b}$

Problem 8a

Let $A = P^T P$. The scalar product of a real matrix A and 2 real column vectors x and y can be transposed without affecting its value.

$$g(x, y) = y^T Ax = g^T = x^T A^T y$$

Using the product rule gives

$$dg = y^T Adx + x^T A^T dy$$

Now set $y = x$, $A = P^T P$ and add the $q^T x$ term, and use the fact that $(P^T P)^T = P^T P$

$$dg = x^T P^T Pdx + x^T P^T Pdx + q^T dx$$
$$\frac{dg}{dx} = 2x^T P^T P + q^T$$

Problem 8b

$$2x^T P^T P + q^T = 0$$
$$x^T P^T P = -.5q^T$$
$$x^T = -.5q^T (P^T P)^{-1} = -.5q^T P^{-T} P^{-1}$$

If $P^T P$ is positive definite then $-.5q^T P^{-T} P^{-1}$ is the global minimum.

Problem 9

Working from the inside out, let $h(x) = a_i^T x + b_i$, $h(x) \epsilon R$
$\frac{dh}{dx} = \begin{bmatrix} \frac{\partial h}{\partial x_1} & \frac{\partial h}{\partial x_2} & ... \end{bmatrix} = a_i^T$
Now with $f(x) = \sum_{i=1}^m e^{a_i^T x + b_i} = e^{a_1^T x + b_1} + e^{a_2^T x + b_2} + .. + e^{a_m^T x + b_m}$
Using $\frac{dh}{dx}$ we have $\frac{df}{dx} = a_1^T e^{a_1^T x + b_1} + a_2^T e^{a_2^T x + b_2} + .. + a_m^T e^{a_m^T x + b_m} = \sum_{i=1}^m a_i^T e^{a_i^T x + b_i}$

Now the outer function is $g(x) = ln(f(x))$. Using the chain rule we get

$$\frac{dg}{dx} = \frac{\sum_{i=1}^m a_i^T e^{a_i^T x + b_i}}{\sum_{i=1}^m e^{a_i^T x + b_i}}$$