1 Eigen Value *(8 points)*

1.
For matrix A we have $p(\lambda) = |\lambda I - A| = \lambda^n + c_{n-1}\lambda^{n-1} + ... + c_1\lambda + c_0$
Consider the coefficient of λ^{n-1} as c_{n-1}. This can be calculated in two ways.
First off, it can be calculated by expanding $p(\lambda) = (\lambda - \lambda_1)(\lambda - \lambda_2)...(\lambda - \lambda_n)$. In order to get the λ^{n-1} term, the λ must be chosen from $n-1$ of the factors, and the constant from the other. Hence, the λ^{n-1} term will be

$$-\lambda_1\lambda^{n-1} - ... - \lambda\lambda^{n-1} = -(\lambda_1 + ... + \lambda_n)\lambda^{n-1}$$

Thus $c_{n-1} = -(\lambda_1 + ... + \lambda_n)$.
Secondly, this coefficient can be calculated by expanding $|\lambda I - A|$

$$|\lambda I - A| = \begin{bmatrix} \lambda - a_{11} & -a_{12} & ... & -a_{1n} \\ -a_{21} & \lambda - a_{22} & ... & -a_{2n} \\ . & . & & . \\ . & . & & . \\ . & . & & . \\ -a_{n1} & -a_{n2} & ... & \lambda - a_{nn} \end{bmatrix}$$

One way of calculating determinants is to multiply the elements is positions $1j_1, 2j_2, ..., nj_n$, for each possible permutation $j_1...j_n$ of $1...n$. If the permutation is odd, then the product is also multiplied by -1. Then all of these $n!$ products are added together to produce the determinant. One of these products is $(\lambda - a_{11})...(\lambda - a_{nn})$. Every other possible product can contain at most $n-2$ elements on the diagonal of the matrix, and so will contain at most $n-2$ λs. Hence, when all of these other products are expanded, they will produce a polynomial in λ of degree at most $n-2$. Denote this polynomial by $q(\lambda)$.
Hence, $p(\lambda) = (\lambda - a_{11})...(\lambda - a_{nn}) + q(\lambda)$. Since $q(\lambda)$ has degree at most $n-2$, it has no λ^{n-1} term, and so the λ^{n-1} term of $p(\lambda)$ must be the λ^{n-1} term from $(\lambda - a_{11})...(\lambda - a_{nn})$. However, the argument above for $(\lambda - \lambda_1)...(\lambda - \lambda_n)$ shows that this term must be $-(a_{11} + ... + a_{nn})\lambda^{n-1}$.
Therefore $c_{n-1} = -(\lambda_1 + ... + \lambda_n) = -(a_{11} + ... + a_{nn})$, and so $\lambda_1 + \lambda_2 + ...\lambda_n = a_{11} + ... + a_{nn}$. That is, the sum of the n eigenvalues of A is the trace of A.
2.
Suppose that $\lambda_1, \lambda_2, ..., \lambda_n$ are the eigenvalues of A. Then the λs are also the roots of the characteristic polynomial.

$$det(A - \lambda I) = p(\lambda) = (-1)^n(\lambda - \lambda_1)(\lambda - \lambda_2)...(\lambda - \lambda_n) = (-1)(\lambda - \lambda_1)(-1)(\lambda - \lambda_2)...(-1)(\lambda - \lambda_n) = (\lambda_1 - \lambda)(\lambda_2 - \lambda)...(\lambda_n - \lambda)$$

The first equality follows from the factorization of a polynomial given its roots; the leading (highest degree) coefficient $(-1)^n$ can be obtained by expanding the determinant along the diagonal.
We put λ to zero. So :

$$det(A) = \lambda_1\lambda_2...\lambda_n$$

3.
Consider the $(m+n)*(m+n)$ matrices:

$$M = \begin{bmatrix} O_{n*n} & O_{n*m} \\ A & AB \end{bmatrix}, N = \begin{bmatrix} BA & O_{n*m} \\ A & O_{m*m} \end{bmatrix}$$

Also let $X = \begin{bmatrix} I_{n*n} & B \\ O_{m*n} & I_{m*m} \end{bmatrix}$

So, it is obvious that:

$$XM = NX = \begin{bmatrix} BA & BAB \\ A & AB \end{bmatrix}$$

Now X is an upper triangular matrix with every entry on the diagonal equal to 1. Therefor, it is invertible. Hence, we can multiply both sides of this equation by X^{-1} to get $M = X^{-1}NX$. Thus M & N are similar, and so have the same characteristic polynomial.
Consider the characteristic polynomial of each :

$$|\lambda I - M| = |\begin{bmatrix} \lambda I_{n*n} & O_{n*m} \\ -A & \lambda I_{m*m} - AB \end{bmatrix}| = |\lambda I_{n*n}||\lambda I_{m*m} - AB| = \lambda^n |\lambda I_{m*m} - AB|$$
$$|\lambda I - N| = |\begin{bmatrix} \lambda I_{n*n} - AB & O_{n*m} \\ -A & \lambda I_{m*m} \end{bmatrix}| = |\lambda I_{n*n} - BA||\lambda I_{m*m}| = \lambda^m |\lambda I_{m*m} - BA|$$

Since M and N have the same characteristic polynomial,

$$|\lambda I - M| = |\lambda I - N|$$
$$\lambda^n |\lambda I_{m*m} - AB| = \lambda^m |\lambda I_{m*m} - BA|$$
$$\lambda^{n-m} |\lambda I_{m*m} - AB| = |\lambda I_{m*m} - BA|$$

So, the characteristic polynomial of BA is the same as the characteristic polynomial of AB, but multiplied by λ^{n-m}. Hence, BA has all of the eigenvalues of AB, but with $n - m$ extra zeros. 4.
We know that we should find eigen values by solving $det(A - \lambda I) = 0$.

$$(A - \lambda I)^T = A^T - \lambda I$$

Also, we know that :

$$det(A) = det(A^T)$$

So :

$$det(A - \lambda I) = det(A^T - \lambda I)$$

Two matrices have the same characteristic polynomial, hence the same eigenvalues.

2 Covariance & Expectation

1.

$$Var(X) = cov(X, X) = E[cov(X, X|Z)] + cov(E[X|Z], E[X|Z]) = E[Var(X|Z)] + Var[E[X|Z]]$$

2.

$$cov(X, Y|Z) = E[(X - E[X|Z](Y - E[Y|Z]))|Z] = E[(XY - XE[Y|Z] - YE[X|Z] + E[X|Z]E[Y|Z])|Z] =$$
$$E[XY|Z] - E[XE[Y|Z]|Z] - E[YE[X|Z]|Z] + E[E[X|Z]E[Y|Z]|Z] =$$
$$E[XY|Z] - E[X|Z]E[Y|Z] - E[Y|Z]E[X|Z] + E[X|Z]E[Y|Z] = E[XY|Z] - E[X|Z]E[Y|Z]$$

3. From 2 we know that:

$$E[cov(X, Y|Z)] = E[E[XY|Z]] - E[E[X|Z]E[Y|Z]] = E[XY] - E[E[X|Z]E[Y|Z]]$$
$$cov(E[X|Z], E[Y|Z]) = E[E[X|Z]E[Y|Z]] - E[E[X|Z]]E[E[Y|Z]] = E[E[X|Z]E[Y|Z]] - E[X]E[Y]$$
$$E[cov(X, Y|Z)] + cov(E[X|Z], E[Y|Z]) = E[XY] - E[X]E[Y] = cov(X, Y)$$

3 Matrix Derivative

1. Assume that :

$$x = \begin{bmatrix} x_1 \\ x_2 \\ . \\ . \\ . \\ x_n \end{bmatrix}, A = \begin{bmatrix} a_{11} & a_{12} & ... & a_{1n} \\ a_{21} & a_{22} & ... & a_{2n} \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ a_{n1} & a_{n2} & ... & a_{nn} \end{bmatrix}$$

So, we will have:

$$x^T Ax = \begin{bmatrix} \Sigma_{i=1}^n a_{i1} x_i & \Sigma_{i=1}^n a_{i2} x_i & ... & \Sigma_{i=1}^n a_{in} x_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ . \\ . \\ . \\ x_n \end{bmatrix} = \Sigma_{i=1}^n \Sigma_{j=1}^n a_{ij} x_i x_j$$

Therefore,

$$\frac{\partial x^T Ax}{\partial x_i} = \Sigma_{j=1}^{n}(a_{ij}+a_{ji})x_j$$
$$\frac{\partial x^T Ax}{\partial x} = x^T(A+A^T) = 2x^TA$$

Notice that A is symmetric.
2. We Know that :

$$x^T Ax = \Sigma_{j=1}^{n}\Sigma_{i=1}^{n}a_{ij}x_ix_j$$

Differentiating with respect to the kth element of x we have :

$$\frac{\partial x^T Ax}{x_k} = \Sigma_{j=1}^{n}a_{kj}x_j + \Sigma_{i=1}^{n}a_{ik}x_i$$

for all $k = 1, 2, ..., n$, and so

$$\frac{\partial x^T Ax}{x} = x^TA^T + x^TA = x^T(A^T + A)$$

4 Random Variable

Assume $Y_1 = X^2$, $Y_2 = \sqrt{X}$:

$$F_{Y_1}(y) = P(Y_1 <= Y) = P(X_2 <= Y) = P(X <= \sqrt{Y}) = F_X(\sqrt{y})$$
$$\frac{d}{dy}F_{Y_1}(y) = \frac{d}{dy}F_X(\sqrt{y}) = \frac{1}{\sqrt{2y}}f_X(\sqrt{y}) = \frac{1}{\sqrt{2y}}$$
$$F_{Y_2}(y) = P(Y_2 \le y) = P(\sqrt{X} \le y) = P(X \le y^2) = F_X(y^2)$$
$$\frac{d}{dy}F_{Y_2}(y) = \frac{d}{dy}F_X(y^2) = 2yf_X(y^2) = 2y$$

5 Rank

Assume λ as the eigen value of $P^{-1}MP$ and v as its eigen vector. so :

$$P^{-1}MPv = \lambda v \; MPv = \lambda(Pv)$$

So, λ is an eigen value for M , and the eigen vector for this eigne value will be Pv.

Therefore :

$$Mu = \lambda u \; M(PP^{-1})u = \lambda u \; P^{-1}MP(P^{-1}u) = \lambda(P^{-1}u)$$

6 Matrix Factorization

1. We use inductive reasoning in this case. So :

for n = 1 we have $A = [a]$, and so $L = \sqrt{L}$

for n = 2 assume that $A = \begin{pmatrix} a & b \\ b & d \end{pmatrix}$ & $L = \begin{pmatrix} \lambda & 0 \\ x & \delta \end{pmatrix}$. So, we will have: $A = LL^T$

2.
Assume that the SVD decomposition of A be $A = U\Sigma V^T$. So, $QR = U\Sigma V^T$. We know that $Q^{-1} = Q^T$ so $R = (Q^TU)\Sigma V^T$. Also :

$$(Q^TU)^T(Q^TU) = U^TQQ^TU = U^TU = I$$

So,

$$R = (Q^TU)\Sigma V^T$$

7 Nilpotent Matrix

$$A^k = 0$$
$$I - A^k = I$$
$$I - A^k = (I-A)(A^{k-1} + A^{k-2} + ... + A + I) \; A^{-1} = A^{k-1} + A^{k-2} + ... + A + I$$

8 MAP

1.

$$f_{X,Y,\theta}(x,y,t) = f_{X,Y|\theta}(x,y|t)f_\theta(t)$$

$$= \begin{cases} \frac{1}{2\pi}\exp\left(-\frac{(x-t)^2+(y-t)^2}{2}\right), & 0 \le t \le 1 \\ 0, & O.W. \end{cases}$$

$$= \begin{cases} \frac{1}{2\pi}\exp\left(-\frac{(x)^2+(y)^2}{2}\right)\exp\frac{x+y}{2}t - t^2, & 0 \le t \le 1 \\ 0, & O.W. \end{cases}$$

2. Notice that for fixed x, y , we need to maximize the term $\exp\frac{x+y}{2}t - t^2$ subject to $t \in [0,1]$. This is equivalent to minimizizing $t^2 - \frac{x+y}{2}t$ with respect to $tin[0,1]$. But this is a quadratic with a minimum at $\frac{x+y}{2}$. Therefore, the MAP estimate is :

$$\hat{t} = \begin{cases} 0, & \frac{x+y}{2} < 0 \\ \frac{x+y}{2}, & 0 \le \frac{x+y}{2} \le 1 \\ 1, & 1 < \frac{x+y}{2} \end{cases}$$