Further Mathematical Methods (Linear Algebra) 2002

Solutions For Problem Sheet 6

In this Problem Sheet we used the Gram-Schmidt Procedure to find an orthonormal basis, found the spectral decomposition of a matrix and investigated some of the properties of certain types of complex matrix. The solutions have been written so that the thinking involved is clear, and so they do not necessarily represent the most 'elegant' solutions.

1. Calculating the orthonormal basis using the Gram-Schmidt procedure should have posed no difficulties. *One* of the orthonormal bases that you could have found is:

$$\mathbf{e}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{e}_2 = \frac{1}{\sqrt{5}}\begin{bmatrix} 0 \\ 2 \\ 0 \\ 1 \end{bmatrix} \quad \text{and} \quad \mathbf{e}_3 = \frac{1}{\sqrt{55}}\begin{bmatrix} -5 \\ -1 \\ 5 \\ 2 \end{bmatrix}.$$

To verify that this new basis is indeed orthonormal, you need to work out the nine inner products that can be formed using these vectors. The easy way to do this is to construct a matrix P with your new basis vectors as the column vectors, for then, we can form the matrix product

$$\mathsf{P}^t\mathsf{P} = \begin{bmatrix} — & \mathbf{e}_1^t & — \\ — & \mathbf{e}_2^t & — \\ — & \mathbf{e}_3^t & — \end{bmatrix} \begin{bmatrix} | & | & | \\ \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ | & | & | \end{bmatrix} = \begin{bmatrix} \mathbf{e}_1^t\mathbf{e}_1 & \mathbf{e}_1^t\mathbf{e}_2 & \mathbf{e}_1^t\mathbf{e}_3 \\ \mathbf{e}_2^t\mathbf{e}_1 & \mathbf{e}_2^t\mathbf{e}_2 & \mathbf{e}_2^t\mathbf{e}_3 \\ \mathbf{e}_3^t\mathbf{e}_1 & \mathbf{e}_3^t\mathbf{e}_2 & \mathbf{e}_3^t\mathbf{e}_3 \end{bmatrix}.$$

Hence, as we are using the Euclidean inner product, we note that $\mathbf{x}^t\mathbf{y} = \langle \mathbf{x}, \mathbf{y} \rangle$, and so

$$\mathsf{P}^t\mathsf{P} = \begin{bmatrix} \langle \mathbf{e}_1, \mathbf{e}_1 \rangle & \langle \mathbf{e}_1, \mathbf{e}_2 \rangle & \langle \mathbf{e}_1, \mathbf{e}_3 \rangle \\ \langle \mathbf{e}_2, \mathbf{e}_1 \rangle & \langle \mathbf{e}_2, \mathbf{e}_2 \rangle & \langle \mathbf{e}_2, \mathbf{e}_3 \rangle \\ \langle \mathbf{e}_3, \mathbf{e}_1 \rangle & \langle \mathbf{e}_3, \mathbf{e}_2 \rangle & \langle \mathbf{e}_3, \mathbf{e}_3 \rangle \end{bmatrix}.$$

Thus, if the new basis is orthonormal, we should find that $\mathsf{P}^t\mathsf{P} = \mathsf{I}$. So, calculating $\mathsf{P}^t\mathsf{P}$ (this is all that *you* need to do!) we get the identity matrix, I and so the verification is complete. (Notice that as P is real and $\mathsf{P}^t\mathsf{P} = \mathsf{I}$, it is an *orthogonal* matrix.)

To calculate the [3-dimensional] hyperplane corresponding to this subspace of $\mathbb{R}^4$, you should expand the determinant equation

$$\begin{vmatrix} 1 & 0 & 1 & 0 \\ 0 & 2 & 0 & 1 \\ -5 & -1 & 5 & 2 \\ w & x & y & z \end{vmatrix} = 0.$$

Notice that to make the determinant easier to evaluate, we use the new basis as it contains more zeros and we also take the normalisation constants out of each row (indeed, you should expand along row 1 as that is easiest). Consequently, we find that the required Cartesian equation is $w - 2x - y + 4z = 0$. (Using substitution, you can check this equation by verifying that the components of your vectors satisfy it.)

2. We are asked to prove the following theorems. Firstly,

Theorem: If A is a *symmetric* matrix with real entries, then all eigenvalues of A are real.

Proof: We know from the lectures that if a matrix is Hermitian, then all eigenvalues of the matrix are real. So, if we can establish that A is Hermitian, then it will follow that all eigenvalues of A are real. To do this, we note that as A is a matrix with real entries, $\mathsf{A}^* = \mathsf{A}$ and as A is symmetric, $\mathsf{A}^t = \mathsf{A}$ too. Now, A is Hermitian if $\mathsf{A}^\dagger = \mathsf{A}$, and this is clearly the case since

$$\mathsf{A}^\dagger = (\mathsf{A}^*)^t = \mathsf{A}^t = \mathsf{A},$$

using the two properties of A noted above.

Secondly,

Theorem: If A is a *normal* matrix and all of the eigenvalues of A are real, then A is Hermitian.

Proof: We know from the lectures that if a matrix A is normal, then there is a unitary matrix P such that the matrix $\mathsf{P}^\dagger\mathsf{A}\mathsf{P} = \mathsf{D}$ is diagonal. Indeed, as P is unitary, $\mathsf{P}^\dagger\mathsf{P} = \mathsf{P}\mathsf{P}^\dagger = \mathsf{I}$ and so, $\mathsf{A} = \mathsf{P}\mathsf{D}\mathsf{P}^\dagger$. Further, the entries of D are the eigenvalues of A, and we are told that in this case they are all real, therefore $\mathsf{D}^\dagger = \mathsf{D}$. Now, to establish that A is Hermitian, we have to show that $\mathsf{A}^\dagger = \mathsf{A}$. To do this we start by noting that as $(\mathsf{P}^\dagger)^\dagger = \mathsf{P}$,

$$\mathsf{A}^\dagger = (\mathsf{P}\mathsf{D}\mathsf{P}^\dagger)^\dagger = (\mathsf{P}^\dagger)^\dagger(\mathsf{P}\mathsf{D})^\dagger = \mathsf{P}(\mathsf{P}\mathsf{D})^\dagger = \mathsf{P}\mathsf{D}^\dagger\mathsf{P}^\dagger,$$

by two applications of the $(\mathsf{A}\mathsf{B})^\dagger = \mathsf{B}^\dagger\mathsf{A}^\dagger$ rule. But, we know that $\mathsf{D}^\dagger = \mathsf{D}$ in this case, and so

$$\mathsf{A}^\dagger = \mathsf{P}\mathsf{D}^\dagger\mathsf{P}^\dagger = \mathsf{P}\mathsf{D}\mathsf{P}^\dagger = \mathsf{A},$$

and so A is Hermitian (as required).

Thirdly,

Theorem: If P is a *unitary* matrix, then all eigenvalues of P have a modulus of one.

Proof: Let λ be any eigenvalue of P, and let $\mathbf{x}$ be an eigenvector of P corresponding to λ, i.e. $\mathsf{P}\mathbf{x} = \lambda\mathbf{x}$. As P is unitary, $\mathsf{P}^\dagger\mathsf{P} = \mathsf{I}$, and so

$$\mathbf{x}^\dagger\mathsf{P}^\dagger\mathsf{P}\mathbf{x} = \mathbf{x}^\dagger\mathsf{I}\mathbf{x} = \mathbf{x}^\dagger\mathbf{x}.$$

But, using the $(\mathsf{A}\mathsf{B})^\dagger = \mathsf{B}^\dagger\mathsf{A}^\dagger$ rule, we can also see that

$$\mathbf{x}^\dagger\mathsf{P}^\dagger\mathsf{P}\mathbf{x} = (\mathsf{P}\mathbf{x})^\dagger(\mathsf{P}\mathbf{x}) = (\lambda\mathbf{x})^\dagger(\lambda\mathbf{x}) = \lambda^*\lambda\mathbf{x}^\dagger\mathbf{x} = |\lambda|^2\mathbf{x}^\dagger\mathbf{x}.$$

Equating these two expressions we find

$$|\lambda|^2\mathbf{x}^\dagger\mathbf{x} = \mathbf{x}^\dagger\mathbf{x} \implies (|\lambda|^2 - 1)\mathbf{x}^\dagger\mathbf{x} = 0.$$

But, as $\mathbf{x}$ is an eigenvector, $\mathbf{x}^\dagger\mathbf{x} = \|\mathbf{x}\|^2 \neq 0$,[1] this gives $|\lambda|^2 = 1$, and so $|\lambda| = 1$ (as required).

3. For the matrix A you should have found that the eigenvalues were 2, -2, 16, and that the corresponding eigenvectors were of the form $[0,1,0]^t$, $[-1,0,1]^t$, $[1,0,1]^t$ respectively. We want an orthogonal matrix P, and so we require that the eigenvectors form an orthonormal set. They are obviously orthogonal, and so normalising them we find the matrices

$$\mathsf{P} = \frac{1}{\sqrt{2}}\begin{bmatrix} 0 & -1 & 1 \\ \sqrt{2} & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix} \quad \text{and} \quad \mathsf{D} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 16 \end{bmatrix},$$

where $\mathsf{P}^t\mathsf{A}\mathsf{P} = \mathsf{D}$ (Verify this!).[2]

[1] For those of you who are pedantic, you may care to notice that when we write $\mathbf{x}^\dagger\mathbf{x} = 0$ (say) the right-hand side of this equality is not a scalar but a 1×1 matrix, i.e. $\mathbf{x}^\dagger\mathbf{x} = [0]$. We have implicitly adopted the convention that such 1×1 matrices can be equated with the single scalar which they contain as, in practice, no confusion results. Really, we should say $\mathbf{x}^\dagger\mathbf{x} = \left[\|\mathbf{x}\|^2\right]$ which is not equal to $[0]$ in this case, and so $\|\mathbf{x}\|^2 \neq 0$. Similar remarks will apply whenever we use this trick, although I shall not mention it again. Probably.

[2] Notice that as A is symmetric, it is a normal matrix (as $\mathsf{A}^t\mathsf{A} = \mathsf{A}^2 = \mathsf{A}\mathsf{A}^t$) and so you should have expected it to be orthogonally diagonalisable!

Now, as P is an orthogonal matrix, $\mathsf{P}^t\mathsf{P} = \mathsf{I} = \mathsf{P}\mathsf{P}^t$ and so we can see that $\mathsf{A} = \mathsf{P}\mathsf{D}\mathsf{P}^t$. Then, writing the columns of P (i.e. our orthonormal eigenvectors) as $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_3$, we have

$$\mathsf{A} = \mathsf{P}\mathsf{D}\mathsf{P}^t = \begin{bmatrix} | & | & | \\ \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 \\ | & | & | \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 16 \end{bmatrix} \begin{bmatrix} — & \mathbf{x}_1^t & — \\ — & \mathbf{x}_2^t & — \\ — & \mathbf{x}_3^t & — \end{bmatrix}$$

$$= \begin{bmatrix} | & | & | \\ \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 \\ | & | & | \end{bmatrix} \begin{bmatrix} — & 2\,\mathbf{x}_1^t & — \\ — & -2\,\mathbf{x}_2^t & — \\ — & 16\,\mathbf{x}_3^t & — \end{bmatrix},$$

which on expanding gives us the required result, i.e.

$$\mathsf{A} = 2\mathbf{x}_1\mathbf{x}_1^t + (-2)\mathbf{x}_2\mathbf{x}_2^t + 16\mathbf{x}_3\mathbf{x}_3^t = 2\mathsf{E}_1 + (-2)\mathsf{E}_2 + 16\mathsf{E}_3,$$

where $\mathsf{E}_i = \mathbf{x}_i\mathbf{x}_i^t$ for $i = 1, 2, 3$. Multiplying these out, we find that the appropriate matrices are:

$$\mathsf{E}_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \mathsf{E}_2 = \tfrac{1}{2}\begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \mathsf{E}_3 = \tfrac{1}{2}\begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix}.$$

A quick calculation should then convince you that these matrices have the property that

$$\mathsf{E}_i\mathsf{E}_j = \begin{cases} \mathsf{E}_i & \text{for} \quad i = j \\ \mathsf{0} & \text{for} \quad i \neq j \end{cases},$$

for $i, j = 1, 2, 3$.

To establish the next result, we consider any three matrices E_1, E_2 and E_3 with this property, and three arbitrary real numbers α_1, α_2 and α_3. Now, observe that:

$$\begin{aligned} (\alpha_1\mathsf{E}_1 + \alpha_2\mathsf{E}_2 + \alpha_3\mathsf{E}_3)^2 &= (\alpha_1\mathsf{E}_1 + \alpha_2\mathsf{E}_2 + \alpha_3\mathsf{E}_3)(\alpha_1\mathsf{E}_1 + \alpha_2\mathsf{E}_2 + \alpha_3\mathsf{E}_3) \\ &= \alpha_1^2\mathsf{E}_1\mathsf{E}_1 + \alpha_2^2\mathsf{E}_2\mathsf{E}_2 + \alpha_3^2\mathsf{E}_3\mathsf{E}_3 && :\text{as } \mathsf{E}_i\mathsf{E}_j = \mathsf{0} \text{ for } i \neq j. \\ &= \alpha_1^2\mathsf{E}_1 + \alpha_2^2\mathsf{E}_2 + \alpha_3^2\mathsf{E}_3 && :\text{as } \mathsf{E}_i\mathsf{E}_j = \mathsf{E}_i \text{ for } i = j. \end{aligned}$$

Consequently, using a similar argument, i.e.

$$\begin{aligned} (\alpha_1\mathsf{E}_1 + \alpha_2\mathsf{E}_2 + \alpha_3\mathsf{E}_3)^3 &= (\alpha_1\mathsf{E}_1 + \alpha_2\mathsf{E}_2 + \alpha_3\mathsf{E}_3)^2(\alpha_1\mathsf{E}_1 + \alpha_2\mathsf{E}_2 + \alpha_3\mathsf{E}_3) \\ &= (\alpha_1^2\mathsf{E}_1 + \alpha_2^2\mathsf{E}_2 + \alpha_3^2\mathsf{E}_3)(\alpha_1\mathsf{E}_1 + \alpha_2\mathsf{E}_2 + \alpha_3\mathsf{E}_3) && :\text{from above.} \\ &= \alpha_1^3\mathsf{E}_1\mathsf{E}_1 + \alpha_2^3\mathsf{E}_2\mathsf{E}_2 + \alpha_3^3\mathsf{E}_3\mathsf{E}_3 && :\text{as } \mathsf{E}_i\mathsf{E}_j = \mathsf{0} \text{ for } i \neq j. \\ &= \alpha_1^3\mathsf{E}_1 + \alpha_2^3\mathsf{E}_2 + \alpha_3^3\mathsf{E}_3 && :\text{as } \mathsf{E}_i\mathsf{E}_j = \mathsf{E}_i \text{ for } i = j. \end{aligned}$$

we obtain the desired result.

To find a matrix B such that $\mathsf{B}^3 = \mathsf{A}$, we use the above result to see that

$$\mathsf{B}^3 = 2\mathsf{E}_1 + (-2)\mathsf{E}_2 + 16\mathsf{E}_3 = \mathsf{A},$$

implies that $\alpha_1^3 = 2$, $\alpha_2^3 = -2$ and $\alpha_3^3 = 16$, i.e.

$$\mathsf{B} = \sqrt[3]{2}\mathsf{E}_1 + \sqrt[3]{-2}\mathsf{E}_2 + \sqrt[3]{16}\mathsf{E}_3.$$

Thus, as $\sqrt[3]{-2} = -\sqrt[3]{2}$ and $\sqrt[3]{16} = 2\sqrt[3]{2}$, this gives us

$$\mathsf{B} = \frac{\sqrt[3]{2}}{2}\begin{bmatrix} 1 & 0 & 3 \\ 0 & 2 & 0 \\ 3 & 0 & 1 \end{bmatrix}.$$

If you are very keen you can check that this is correct by multiplying it out!

Other Problems.

Here are the solutions for the other problems. As these were not covered in class the solutions will be a bit more detailed.

4. We are given that the Taylor expansion of the exponential function, e^x, is

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$$

for all $x \in \mathbb{R}$. As we saw in lectures, we can define the exponential of A, an $n \times n$ Hermitian matrix, by replacing each x in the above expression by A, i.e.

$$e^{\mathsf{A}} = \mathsf{I} + \mathsf{A} + \frac{\mathsf{A}^2}{2!} + \frac{\mathsf{A}}{3!} + \cdots$$

Now, as A is Hermitian, we can find its spectral decomposition which will be of the form

$$\mathsf{A} = \sum_{i=1}^{n} \lambda_i \mathsf{E}_i,$$

where λ_i is an eigenvalue with corresponding eigenvector $\mathbf{x}_i$. The set of all such eigenvectors $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ is taken to be orthonormal and so the matrices $\mathsf{E}_i = \mathbf{x}_i \mathbf{x}_i^\dagger$ for $1 \leq i \leq n$ have the property that

$$\mathsf{E}_i \mathsf{E}_j = \begin{cases} \mathsf{E}_i & \text{for} \quad i = j \\ 0 & \text{for} \quad i \neq j \end{cases},$$

which means that for any integer $k \geq 1$, we can write[3]

$$\mathsf{A}^k = \sum_{i=1}^{n} \lambda_i^k \mathsf{E}_i.$$

(By the way, notice that if $k = 0$, we can use the theory given in the lectures to get

$$\mathsf{A}^0 = \sum_{i=1}^{n} \lambda_i^0 \mathsf{E}_i = \sum_{i=1}^{n} \mathsf{E}_i = \mathsf{I},$$

as one might expect.) So, using this formula to substitute for [integer] powers of A in our expression for e^{A}, we get

$$e^{\mathsf{A}} = \sum_{i=1}^{n} \mathsf{E}_i + \sum_{i=1}^{n} \lambda_i \mathsf{E}_i + \frac{1}{2!} \sum_{i=1}^{n} \lambda_i^2 \mathsf{E}_i + \frac{1}{3!} \sum_{i=1}^{n} \lambda_i^3 \mathsf{E}_i + \cdots$$

and gathering up the coefficients of each matrix E_i it should be clear that

$$e^{\mathsf{A}} = \sum_{i=1}^{n} \left[1 + \lambda_i + \frac{\lambda_i^2}{2!} + \frac{\lambda_i^3}{3!} + \cdots \right] \mathsf{E}_i.$$

Thus, as the eigenvalues of an Hermitian matrix are all real, we can use the Taylor expansion of e^x given above to deduce that

$$e^{\mathsf{A}} = \sum_{i=1}^{n} e^{\lambda_i} \mathsf{E}_i,$$

as required.

[3] If you don't believe me see the Aside below.

To verify that this function has the property that $e^{2\mathsf{A}} = e^{\mathsf{A}}e^{\mathsf{A}}$ (which is analogous to the property that $e^{2x} = e^x e^x$ when $x \in \mathbb{R}$), we write

$$e^{\mathsf{A}}e^{\mathsf{A}} = \left[\sum_{i=1}^{n} e^{\lambda_i}\mathsf{E}_i\right]\left[\sum_{i=1}^{n} e^{\lambda_i}\mathsf{E}_i\right],$$

which on expanding yields

$$e^{\mathsf{A}}e^{\mathsf{A}} = \sum_{i=1}^{n} e^{2\lambda_i}\mathsf{E}_i,$$

where we have used the fact that $e^{2x} = e^x e^x$ when $x \in \mathbb{R}$.[4] Now, clearly, if A is Hermitian with eigenvalues λ_i and corresponding eigenvectors given by $\mathbf{x}_i$, then

$$\mathsf{A}\mathbf{x}_i = \lambda_i \mathbf{x}_i,$$

for $1 \leq i \leq n$. But, this implies that the matrix $2\mathsf{A}$ is Hermitian with eigenvalues $2\lambda_i$ and corresponding eigenvectors given by $\mathbf{x}_i$, as

$$(2\mathsf{A})\mathbf{x}_i = (2\lambda_i)\mathbf{x}_i,$$

for $1 \leq i \leq n$. Consequently, the right-hand-side of our expression for $e^{\mathsf{A}}e^{\mathsf{A}}$ is just the spectral decomposition of the matrix $e^{2\mathsf{A}}$ and so

$$e^{\mathsf{A}}e^{\mathsf{A}} = e^{2\mathsf{A}},$$

as required.

Aside: We want to prove that the spectral decomposition of an $n \times n$ normal matrix, A, has the property that

$$\mathsf{A}^k = \sum_{i=1}^{n} \lambda_i^k \mathsf{E}_i,$$

where, as usual we take

$$\mathsf{A} = \sum_{i=1}^{n} \lambda_i \mathsf{E}_i,$$

and the matrices E_i for $1 \leq i \leq n$ have the property that

$$\mathsf{E}_i\mathsf{E}_j = \begin{cases} \mathsf{E}_i & \text{for} \quad i = j \\ 0 & \text{for} \quad i \neq j \end{cases}.$$

To do this, we can use induction on k:

- **Induction Hypothesis:** If A is an $n \times n$ normal matrix and $k \geq 1$, then

  $$\mathsf{A}^k = \sum_{i=1}^{n} \lambda_i^k \mathsf{E}_i,$$

  where the relationship between A, λ_i and the matrices E_i for $1 \leq i \leq n$ is as described above.

- **Basis:** In the case where $k = 1$, the Induction Hypothesis just gives the spectral decomposition, and so we have established the basis case.

- **Induction Step:** Suppose that the Induction Hypothesis is true for k. We want to show that

  $$\mathsf{A}^{k+1} = \sum_{i=1}^{n} \lambda_i^{k+1} \mathsf{E}_i.$$

[4] Recall that Hermitian matrices have real eigenvalues!

To do this, we write A^{k+1} as $\mathsf{A}^k\mathsf{A}$ and apply the Induction Hypothesis to get

$$\mathsf{A}^{k+1} = \mathsf{A}^k\mathsf{A} = \left[\sum_{i=1}^{n} \lambda_i^k \mathsf{E}_i\right]\left[\sum_{i=1}^{n} \lambda_i \mathsf{E}_i\right],$$

which, on expanding the right-hand-side and using the standard properties of the matrices E_i gives

$$\mathsf{A}^{k+1} = \sum_{i=1}^{n} \lambda_i^{k+1} \mathsf{E}_i,$$

which is what we were after.

Thus, by the Principle of Induction, if A is an $n \times n$ normal matrix as described above and $k \geq 1$, then

$$\mathsf{A}^k = \sum_{i=1}^{n} \lambda_i^k \mathsf{E}_i,$$

as required.

5. We are asked to establish that for any two matrices A and B, $(\mathsf{AB})^\dagger = \mathsf{B}^\dagger\mathsf{A}^\dagger$. To do this, we refer to the (i,j)th element of the $m \times n$ matrix A (where $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$) as a_{ij}, and the (i,j)th element of the $n \times r$ matrix B (where $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, r$) as b_{ij}. This means that the ith row of the matrix A is given by the vector $[a_{i1}, a_{i2}, \ldots, a_{in}]$ and the jth column of the matrix B is given by the vector $[b_{1j}, b_{2j}, \ldots, b_{nj}]^t$. Thus, the (i,j)th element of the matrix AB (for $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, r$) will be given by

$$a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj}.$$

Consequently, the (j,i)th element of the complex conjugate transpose of the matrix AB, i.e. $(\mathsf{AB})^\dagger$, will be $a_{i1}^*b_{1j}^* + a_{i2}^*b_{2j}^* + \cdots + a_{in}^*b_{nj}^*$.

On the other hand, the jth row of the matrix $\mathsf{B}^\dagger$ is given by the complex conjugate transpose of the jth column vector of B, i.e. $[b_{1j}^*, b_{2j}^*, \ldots, b_{nj}^*]$, and the ith column of the matrix $\mathsf{A}^\dagger$ is given by the complex conjugate transpose of the ith row vector of A, i.e. $[a_{i1}^*, a_{i2}^*, \ldots, a_{in}^*]^t$. Thus, the (j,i)th element of the matrix $\mathsf{B}^\dagger\mathsf{A}^\dagger$ will be given by

$$b_{1j}^*a_{i1}^* + b_{2j}^*a_{i2}^* + \cdots + b_{nj}^*a_{in}^*,$$

which is the same as above. Thus, as the (j,i)th elements of the matrices $(\mathsf{AB})^\dagger$ and $\mathsf{B}^\dagger\mathsf{A}^\dagger$ are equal (for all $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, r$), $(\mathsf{AB})^\dagger = \mathsf{B}^\dagger\mathsf{A}^\dagger$ as required.

In particular, notice that if A and B are real matrices, then $\mathsf{A}^* = \mathsf{A}$, $\mathsf{B}^* = \mathsf{B}$ and $(\mathsf{AB})^* = \mathsf{AB}$. This means that using this rule we have $(\mathsf{AB})^\dagger = \mathsf{B}^\dagger\mathsf{A}^\dagger$, implying $[(\mathsf{AB})^*]^t = (\mathsf{B}^*)^t(\mathsf{A}^*)^t$, and hence $(\mathsf{AB})^t = \mathsf{B}^t\mathsf{A}^t$ as required.

6. We are allowed to assume that if A is an $n \times n$ matrix with complex entries, then $\det(\mathsf{A}^*) = \det(\mathsf{A})^*$. For such a matrix, we are then asked to prove that $\det(\mathsf{A}^\dagger) = \det(\mathsf{A})^*$. This is easy because we know that $\det(\mathsf{A}) = \det(\mathsf{A}^t)$[5] and so using these two results:

$$\det(\mathsf{A}^\dagger) = \det((\mathsf{A}^t)^*) = \det(\mathsf{A}^t)^* = \det(\mathsf{A})^*,$$

as required.

To establish the other two results, we use this new result. Firstly, if A is Hermitian, we want to show that $\det(\mathsf{A})^*$ is real. So, as A is Hermitian,

$$\mathsf{A} = \mathsf{A}^\dagger \implies \det(\mathsf{A}) = \det(\mathsf{A}^\dagger) \implies \det(\mathsf{A}) = \det(\mathsf{A})^*,$$

[5] This is *obvious* as we can expand a determinant along any row *or* column to get the answer.

but this means that $\det(\mathsf{A})$ is equal to its complex conjugate and so it is real (as required). Secondly, if A is unitary, we want to show that $|\det(\mathsf{A})| = 1$. So, as A is unitary, $\mathsf{A}\mathsf{A}^\dagger = \mathsf{I}$ and consequently,

$$\det(\mathsf{A}\mathsf{A}^\dagger) = \det(\mathsf{I}) \Rightarrow \det(\mathsf{A})\det(\mathsf{A}^\dagger) = 1 \Rightarrow \det(\mathsf{A})\det(\mathsf{A})^* = 1 \Rightarrow |\det(\mathsf{A})|^2 = 1 \Rightarrow |\det(\mathsf{A})| = 1,$$

as required.

When A is a real $n \times n$ matrix, we note that $\mathsf{A} = \mathsf{A}^*$ and the determinant of a real matrix is obviously real, i.e. $\det(\mathsf{A}) = \det(\mathsf{A})^*$. Thus, the result at the beginning is trivial as $\det(\mathsf{A}^*) = \det(\mathsf{A}) = \det(\mathsf{A})^*$, and so the corresponding result is $\det(\mathsf{A}) = \det(\mathsf{A})$ [!]. The first theorem then gives

$$\det(\mathsf{A}^\dagger) = \det(\mathsf{A})^* \Longrightarrow \det([\mathsf{A}^*]^t) = \det(\mathsf{A}) \Longrightarrow \det(\mathsf{A}^t) = \det(\mathsf{A}),$$

[which we knew already]. Now, as an Hermitian matrix, A with real entries is symmetric (i.e. $\mathsf{A} = \mathsf{A}^\dagger = (\mathsf{A}^*)^t = \mathsf{A}^t$) and $\det(\mathsf{A})^* = \det(\mathsf{A})$, the second theorem becomes: if A is symmetric, then $\det(\mathsf{A})$ is real [obvious as $\det(\mathsf{A})$ is always real if A is real]. Whilst, as a unitary matrix, A with real entries is orthogonal (i.e. $\mathsf{I} = \mathsf{A}\mathsf{A}^\dagger = \mathsf{A}(\mathsf{A}^*)^t = \mathsf{A}\mathsf{A}^t$), the third theorem becomes: if A is orthogonal, then $|\det(\mathsf{A})| = 1$ [which is not unexpected, but at least we have seen it now].

7. We are asked to prove the following theorems. Firstly,

> **Theorem:** If A is invertible, then so is $\mathsf{A}^\dagger$. In particular, $(\mathsf{A}^\dagger)^{-1} = (\mathsf{A}^{-1})^\dagger$.
>
> **Proof:** As the matrix A is invertible, there is a matrix A^{-1} such that $\mathsf{A}\mathsf{A}^{-1} = \mathsf{I}$. Taking the complex conjugate transpose, this gives
>
> $$(\mathsf{A}\mathsf{A}^{-1})^\dagger = \mathsf{I}^\dagger \Rightarrow (\mathsf{A}^{-1})^\dagger\mathsf{A}^\dagger = \mathsf{I},$$
>
> and so there exists a matrix, namely $(\mathsf{A}^{-1})^\dagger$, which acts as the inverse of $\mathsf{A}^\dagger$ and so $\mathsf{A}^\dagger$ is invertible, as required. In particular, as $\mathsf{A}^\dagger$ is invertible, there is a matrix $(\mathsf{A}^\dagger)^{-1}$ such that $(\mathsf{A}^\dagger)^{-1}\mathsf{A}^\dagger = \mathsf{I}$. Taking this and the last part of the previous calculation, on subtracting we get:
>
> $$[(\mathsf{A}^{-1})^\dagger - (\mathsf{A}^\dagger)^{-1}]\mathsf{A}^\dagger = \mathsf{0},$$
>
> where the right-hand-side is the zero matrix. Now, multiplying both sides by $(\mathsf{A}^\dagger)^{-1}$ (say) and rearranging, we get $(\mathsf{A}^{-1})^\dagger = (\mathsf{A}^\dagger)^{-1}$ as required.

Secondly,

> **Theorem:** If A is a unitary matrix, then $\mathsf{A}^\dagger$ is unitary too.
>
> **Proof:** As the matrix A is unitary, $\mathsf{A}\mathsf{A}^\dagger = \mathsf{I}$. But, we know that $(\mathsf{A}^\dagger)^\dagger = \mathsf{A}$ and so, on substitution, we get $(\mathsf{A}^\dagger)^\dagger(\mathsf{A}^\dagger) = \mathsf{I}$ which entails that $\mathsf{A}^\dagger$ is unitary, as required.

Harder Problems.

Here are the solutions for the harder problems. Again, as these were not covered in class the solutions will be a bit more detailed.

8. We are asked to prove that an $n \times n$ matrix, A with complex entries is unitary iff its column vectors form an orthonormal set in $\mathbb{C}^n$ with the [complex] Euclidean inner product (i.e. the inner product where for two vectors $\mathbf{x} = [x_1, x_2, \ldots, x_n]^t$ and $\mathbf{y} = [y_1, y_2, \ldots, y_n]^t$ we have $\langle \mathbf{x}, \mathbf{y} \rangle = x_1y_1^* + x_2y_2^* + \cdots + x_ny_n^*$). To do this, we denote the column vectors of the matrix A by $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$, and follow the observation in Question 1, i.e.

$$\mathsf{A}^\dagger\mathsf{A} = \begin{bmatrix} — & \mathbf{e}_1^\dagger & — \\ — & \mathbf{e}_2^\dagger & — \\ & \vdots & \\ — & \mathbf{e}_n^\dagger & — \end{bmatrix} \begin{bmatrix} | & | & & | \\ \mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} \mathbf{e}_1^\dagger\mathbf{e}_1 & \mathbf{e}_1^\dagger\mathbf{e}_2 & \cdots & \mathbf{e}_1^\dagger\mathbf{e}_n \\ \mathbf{e}_2^\dagger\mathbf{e}_1 & \mathbf{e}_2^\dagger\mathbf{e}_2 & \cdots & \mathbf{e}_2^\dagger\mathbf{e}_n \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{e}_n^\dagger\mathbf{e}_1 & \mathbf{e}_n^\dagger\mathbf{e}_2 & \cdots & \mathbf{e}_n^\dagger\mathbf{e}_n \end{bmatrix}.$$

Then, as we are using the (complex) Euclidean inner product, we note that $\mathbf{x}^\dagger\mathbf{y} = \langle \mathbf{x}, \mathbf{y}\rangle^*$, and so

$$\mathsf{A}^\dagger\mathsf{A} = \begin{bmatrix} \langle \mathbf{e}_1, \mathbf{e}_1\rangle^* & \langle \mathbf{e}_1, \mathbf{e}_2\rangle^* & \cdots & \langle \mathbf{e}_1, \mathbf{e}_n\rangle^* \\ \langle \mathbf{e}_2, \mathbf{e}_1\rangle^* & \langle \mathbf{e}_2, \mathbf{e}_2\rangle^* & \cdots & \langle \mathbf{e}_2, \mathbf{e}_n\rangle^* \\ \vdots & \vdots & \ddots & \vdots \\ \langle \mathbf{e}_n, \mathbf{e}_1\rangle^* & \langle \mathbf{e}_n, \mathbf{e}_2\rangle^* & \cdots & \langle \mathbf{e}_n, \mathbf{e}_n\rangle^* \end{bmatrix}.$$

Thus, clearly, $\mathsf{A}^\dagger\mathsf{A} = \mathsf{I}$ iff the vectors $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ form an orthonormal set, i.e. A is unitary iff the column vectors of A form an orthonormal set (as required).

9. We are asked to prove that if $\mathsf{A} = \mathsf{A}^\dagger$, then for every vector in $\mathbb{C}^n$, the entry in the 1×1 matrix $\mathbf{x}^\dagger\mathsf{A}\mathbf{x}$ is real. To do this, we let P be the 1×1 matrix given by $\mathbf{x}^\dagger\mathsf{A}\mathbf{x}$. This means that

$$\mathsf{P}^\dagger = (\mathbf{x}^\dagger\mathsf{A}\mathbf{x})^\dagger \implies \mathsf{P}^\dagger = \mathbf{x}^\dagger(\mathbf{x}^\dagger\mathsf{A})^\dagger \implies \mathsf{P}^\dagger = \mathbf{x}^\dagger\mathsf{A}^\dagger\mathbf{x}.$$

But, as A is Hermitian, $\mathsf{A} = \mathsf{A}^\dagger$ and so $\mathsf{P} = \mathsf{P}^\dagger$. Then, as P is a 1×1 matrix we have $\mathsf{P} = \mathsf{P}^t$, which means that $\mathsf{P} = \mathsf{P}^\dagger$ implies $\mathsf{P} = \mathsf{P}^*$. Consequently, the single element in $\mathsf{P} = \mathbf{x}^\dagger\mathsf{A}\mathbf{x}$ is real (as required).

10. Suppose that λ and μ are distinct eigenvalues of a Hermitian matrix A. We are asked to prove that if $\mathbf{x}$ is an eigenvector corresponding to λ and $\mathbf{y}$ is an eigenvector corresponding to μ, then

$$\mathbf{x}^\dagger\mathsf{A}\mathbf{y} = \lambda\mathbf{x}^\dagger\mathbf{y} \quad \text{and} \quad \mathbf{x}^\dagger\mathsf{A}\mathbf{y} = \mu\mathbf{x}^\dagger\mathbf{y}.$$

To do this,[6] we note that, by stipulation,

$$\mathsf{A}\mathbf{x} = \lambda\mathbf{x} \quad \text{and} \quad \mathsf{A}\mathbf{y} = \mu\mathbf{y},$$

and so, clearly, taking the second of these and multiplying both sides by $\mathbf{x}^\dagger$, we get $\mathbf{x}^\dagger\mathsf{A}\mathbf{y} = \mu\mathbf{x}^\dagger\mathbf{y}$ (as required). Also, we can see that taking the first of these and multiplying both sides by $\mathbf{y}^\dagger$ we get

$$\mathbf{y}^\dagger\mathsf{A}\mathbf{x} = \lambda\mathbf{y}^\dagger\mathbf{x}.$$

Thus, taking the complex conjugate transpose of this expression we get

$$(\mathbf{y}^\dagger\mathsf{A}\mathbf{x})^\dagger = (\lambda\mathbf{y}^\dagger\mathbf{x})^\dagger \implies \mathbf{x}^\dagger(\mathbf{y}^\dagger\mathsf{A})^\dagger = \lambda^*(\mathbf{y}^\dagger\mathbf{x})^\dagger \implies \mathbf{x}^\dagger\mathsf{A}^\dagger\mathbf{y} = \lambda^*\mathbf{x}^\dagger\mathbf{y},$$

and so as A is Hermitian, $\mathsf{A}^\dagger = \mathsf{A}$ and $\lambda^* = \lambda$,[7] which means that $\mathbf{x}^\dagger\mathsf{A}\mathbf{y} = \lambda\mathbf{x}^\dagger\mathbf{y}$ (as required).

We are then asked to use this result to prove that if A is a normal matrix, then the eigenvectors from different eigenspaces are orthogonal. But, clearly, we are not entitled to use *this* result because not all normal matrices are Hermitian.[8] Consequently, the book I stole this question from is being a bit optimistic, as although the previous result *does* hold for normal matrices, we have yet to establish it. So, let us start by proving this result for normal matrices. We shall then discuss what is meant by 'the eigenvectors from different eigenspaces are orthogonal' as this was not really mentioned in the lectures and you may not know what you are being asked to prove [!]. Then we shall prove the result in question[9]

So, let us start by showing that if A is a normal matrix with eigenvectors $\mathbf{x}$ and $\mathbf{y}$ corresponding to distinct eigenvalues λ and μ, then

$$\mathbf{x}^\dagger\mathsf{A}\mathbf{y} = \lambda\mathbf{x}^\dagger\mathbf{y} \quad \text{and} \quad \mathbf{x}^\dagger\mathsf{A}\mathbf{y} = \mu\mathbf{x}^\dagger\mathbf{y}.$$

[6] Some of you may think that there is an error in this question as these formulae seem to imply that $\lambda = \mu$ contrary to the assumption that $\lambda \neq \mu$. But, this is clearly not the case since distinct eigenvalues of a Hermitian matrix have orthogonal eigenvectors and so, as this entails that $\mathbf{x}^\dagger\mathbf{y} = 0$ we should *not* conclude that $\lambda = \mu$.

[7] Recall that Hermitian matrices have real eigenvalues.

[8] For example, anti-Hermitian matrices (i.e. matrices such that $\mathsf{A}^\dagger = -\mathsf{A}$) are normal (as $\mathsf{A}\mathsf{A}^\dagger = -\mathsf{A}^2 = \mathsf{A}^\dagger\mathsf{A}$) but are clearly not Hermitian.

[9] Although once we have done the other two things this is *very* quick!

To do this, we [again] note that, by stipulation,

$$\mathsf{A}\mathbf{x} = \lambda\mathbf{x} \quad \text{and} \quad \mathsf{A}\mathbf{y} = \mu\mathbf{y},$$

and so, clearly, taking the second of these and multiplying both sides by $\mathbf{x}^\dagger$, we get $\mathbf{x}^\dagger\mathsf{A}\mathbf{y} = \mu\mathbf{x}^\dagger\mathbf{y}$ (as required).[10] However, the other result is harder to prove because, in general, normal matrices do not have real eigenvalues[11] and so we cannot proceed as we did in the earlier proof. Indeed, to prove the remaining result we have to establish the following two results:

Lemma: If A is a normal matrix, then $\mathsf{A} - \lambda\mathsf{I}$ is a normal matrix too.

Proof: We need to establish that $\mathsf{A} - \lambda\mathsf{I}$ is a normal matrix, to do this we observe that

$$(\mathsf{A} - \lambda\mathsf{I})^\dagger(\mathsf{A} - \lambda\mathsf{I}) = (\mathsf{A}^\dagger - \lambda^*\mathsf{I})(\mathsf{A} - \lambda\mathsf{I}) = \mathsf{A}^\dagger\mathsf{A} - \lambda^*\mathsf{A} - \lambda\mathsf{A}^\dagger + \lambda^*\lambda\mathsf{I},$$

and

$$(\mathsf{A} - \lambda\mathsf{I})(\mathsf{A} - \lambda\mathsf{I})^\dagger = (\mathsf{A} - \lambda\mathsf{I})(\mathsf{A}^\dagger - \lambda^*\mathsf{I}) = \mathsf{A}\mathsf{A}^\dagger - \lambda\mathsf{A}^\dagger - \lambda^*\mathsf{A} + \lambda\lambda^*\mathsf{I}.$$

Then, on subtracting these two results we find that

$$(\mathsf{A} - \lambda\mathsf{I})^\dagger(\mathsf{A} - \lambda\mathsf{I}) - (\mathsf{A} - \lambda\mathsf{I})(\mathsf{A} - \lambda\mathsf{I})^\dagger = \mathsf{A}^\dagger\mathsf{A} - \mathsf{A}\mathsf{A}^\dagger = 0,$$

because the matrix A is normal. Consequently, rearranging this we get

$$(\mathsf{A} - \lambda\mathsf{I})^\dagger(\mathsf{A} - \lambda\mathsf{I}) = (\mathsf{A} - \lambda\mathsf{I})(\mathsf{A} - \lambda\mathsf{I})^\dagger,$$

and so $\mathsf{A} - \lambda\mathsf{I}$ is normal, as required.

and

Lemma: If A is a normal matrix and $\mathbf{x}$ is an eigenvector corresponding to the eigenvalue λ, then $\mathsf{A}^\dagger\mathbf{x} = \lambda^*\mathbf{x}$.

Proof: As $\mathbf{x}$ is an eigenvector of the normal matrix A corresponding to the eigenvalue λ, we know that

$$(\mathsf{A} - \lambda\mathsf{I})\mathbf{x} = \mathbf{0},$$

and multiplying both sides of this expression by $\mathbf{x}^\dagger(\mathsf{A} - \lambda\mathsf{I})^\dagger$ we get

$$\mathbf{x}^\dagger(\mathsf{A} - \lambda\mathsf{I})^\dagger(\mathsf{A} - \lambda\mathsf{I})\mathbf{x} = \mathbf{0}.$$

But, by the previous lemma, the matrix $\mathsf{A} - \lambda\mathsf{I}$ is normal too, and so this is equivalent to

$$\mathbf{x}^\dagger(\mathsf{A} - \lambda\mathsf{I})(\mathsf{A} - \lambda\mathsf{I})^\dagger\mathbf{x} = \mathbf{0},$$

which, on using the $(\mathsf{AB})^\dagger = \mathsf{B}^\dagger\mathsf{A}^\dagger$ rule becomes

$$\left[(\mathsf{A} - \lambda\mathsf{I})^\dagger\mathbf{x}\right]^\dagger \left[(\mathsf{A} - \lambda\mathsf{I})^\dagger\mathbf{x}\right] = \mathbf{0}.$$

This, in turn, entails that[12]

$$(\mathsf{A} - \lambda\mathsf{I})^\dagger\mathbf{x} = \mathbf{0},$$

i.e. $\mathsf{A}^\dagger\mathbf{x} = \lambda^*\mathbf{x}$, as required.[13]

---

[10] This obviously holds for all square matrices!

[11] For instance, recall that unitary matrices, which are themselves normal, generally have complex eigenvalues (although as we saw in Question 2, these eigenvalues have a modulus of one).

[12] Bear in mind that $(\mathsf{A} - \lambda\mathsf{I})^\dagger\mathbf{x}$ will be a vector, and so the left-hand-side of the previous expression is effectively the inner product of this vector with itself.

[13] That is, when considering the complex conjugate transpose of a normal matrix A, the eigenvectors of A remain unchanged, but the eigenvalues are complex conjugated. Thus, our theorem about the eigenvalues of an Hermitian matrix being real follows almost immediately from this result.

Now that we have the result of this lemma, we multiply both sides by $\mathbf{y}^\dagger$ and take the complex conjugate transpose of both sides as before, i.e.

$$(\mathbf{y}^\dagger \mathsf{A}^\dagger \mathbf{x})^\dagger = (\lambda^* \mathbf{y}^\dagger \mathbf{x})^\dagger \implies \mathbf{x}^\dagger (\mathbf{y}^\dagger \mathsf{A}^\dagger)^\dagger = \lambda (\mathbf{y}^\dagger \mathbf{x})^\dagger \implies \mathbf{x}^\dagger \mathsf{A} \mathbf{y} = \lambda \mathbf{x}^\dagger \mathbf{y},$$

as required.

Now, the *eigenspace* of A corresponding to the eigenvalue λ is the vector space spanned by the eigenvectors corresponding to λ. But, before we can proceed, we must bear in mind that the eigenspace in question may not be one-dimensional. That is, if we have an eigenvalue of A which is of multiplicity r, then we will find at most r linearly independent eigenvectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_r$ corresponding to this eigenvalue. Thus, in this case, the eigenspace of A corresponding to the eigenvalue λ will be given by $\mathrm{Lin}\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_r\}$.[14] However, no matter how many eigenvectors are associated with a given eigenvalue, they must all satisfy the equation $\mathsf{A}\mathbf{x}_i = \lambda \mathbf{x}_i$ (for $i = 1, 2, \ldots, r$) and be non-zero.

Thus, as you may have suspected, the theorem that we have been asked to prove is really just: if A is a normal matrix with distinct eigenvalues λ and μ, then eigenvectors corresponding to these eigenvalues are orthogonal. So, to prove this, we just subtract the two results that we proved above to get

$$(\lambda - \mu)\mathbf{x}^\dagger \mathbf{y} = 0,$$

and as $\lambda \neq \mu$, $\mathbf{x}^\dagger \mathbf{y} = \langle \mathbf{x}, \mathbf{y} \rangle = 0$, which implies that these eigenvectors are orthogonal, as required.[15]

[14] As mentioned in the question, eigenspaces are subspaces of $\mathbb{C}^n$ (if A is an $n \times n$ matrix). This can be seen either by noting that the eigenspace corresponding to the eigenvalue λ is the null-space of the matrix $\mathsf{A} - \lambda \mathsf{I}$, or directly using Theorem 2.4. Alternatively, you can amuse yourself by showing that the eigenspace is closed under scalar multiplication and vector addition, as demanded by Theorem 1.4.

[15] I told you that it would be quick! Incidentally, this is what the question I told you about hinted at. Obviously, it was being *very* optimistic...