Symmetric and self-adjoint matrices

A matrix A in $\mathrm{M}_n(\mathbb{F})$ is called *symmetric* if $A^T = A$, i.e. $A_{ij} = A_{ji}$ for each i, j; and *self-adjoint* if $A^* = A$, i.e. $A_{ij} = \overline{A_{ji}}$ or each i, j. Note for A in $\mathrm{M}_n(\mathbb{R})$ that $A^T = A^*$.

Notice that if $\mathbb{F} = \mathbb{R}$, then A is symmetric if and only if $(Ax, y) = (x, Ay)$ for each x, y in $\mathbb{R}^n$. Observe that the set

$$\mathrm{Sym}_n(\mathbb{R}) = \{A \in \mathrm{M}_n(\mathbb{R}) : A^T = A\}$$

is a linear subspace of $\mathrm{M}_n(\mathbb{R})$.

A matrix A in $\mathrm{M}_n(\mathbb{C})$ is called *self-adjoint* or *hermitian* if $A^* = A$. Notice that A is hermitian if and only if $(Ax, y) = (x, Ay)$ for each x, y in $\mathbb{C}^n$. Observe that the set

$$\mathrm{Herm}(n) = \{A \in \mathrm{M}_n(\mathbb{C}) : A^* = A\}$$

is a $\mathbb{R}$-linear subspace of $\mathrm{M}_n(\mathbb{C})$. Note that $\mathrm{Sym}_n(\mathbb{R}) \subset \mathrm{Herm}(n)$.

Consider the sets of *(real) orthongonal* and *unitary* matrices:

$$\mathrm{O}(n) = \{u \in \mathrm{M}_n(\mathbb{R}) : u^T u = I\} \text{ and } \mathrm{U}(n) = \{u \in \mathrm{M}_n(\mathbb{R}) : u^* u = I\}$$

Clearly, $\mathrm{O}(n) \subset \mathrm{U}(n)$.

Remark. It is easy to see that for u in $\mathrm{M}_n(\mathbb{R})$ (respectively, in $\mathrm{M}_n(\mathbb{C})$) that $u \in \mathrm{O}(n)$ (respectively, is in $\mathrm{U}(n)$) if and only if the *columns* of u:

$$u_{(1)} = \begin{bmatrix} u_{11} \\ \vdots \\ u_{1n} \end{bmatrix}, \ldots, u_{(n)} = \begin{bmatrix} u_{n1} \\ \vdots \\ u_{nn} \end{bmatrix}$$

form an orthonormal basis for $\mathbb{R}^n$ (respectively $\mathbb{C}^n$).

Lemma. *If $A \in \mathrm{M}_n(\mathbb{R})$ admits a real eigenvalue, then there is a corresponding real eigenvector.*

Proof. Let λ be a real eigenvalue of A and $z \neq 0$ in $\mathbb{C}^n$ be an eigenvector. Write $x_j = \mathrm{Re} z_j$ and $y_j = \mathrm{Im} z_j$ so $z = x + iy$ where $x, y \in \mathbb{R}^n$. Then

$$\lambda x + i\lambda y = \lambda z = Az = Ax + iAy.$$

Collecting the real and imaginary parts of each entries of each component of the above equality gives a non-zero real eigenvector: at least one of x or y. $\square$

Diagonalization Theorem (i) *If $A \in \mathrm{Herm}(n)$, then the eigenvalues of A are real. Furthermore, for any two distinct eigenvalues λ, μ of A with corresponding eigenvectors x and y in $\mathbb{C}^n$, we have $(x, y) = 0$.*

(ii) *If $A \in \mathrm{M}_n(\mathbb{R})$ (respectively, is in $\mathrm{M}_n(\mathbb{C})$) then $A \in \mathrm{Sym}_n(\mathbb{R})$ (respectively, is in $\mathrm{Herm}(n)$) if and only if there is u in $\mathrm{O}(n)$ (respectively, u in $\mathrm{U}(n)$) for which*

$$u^*Au = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \lambda_n \end{bmatrix}$$

where $\lambda_1, \dots, \lambda_n$ are the eigenvalues of A (with multiplicity).

Proof. (i) Let λ be an eigenvalue of A with corresponding eigenvector $x \neq 0$ in $\mathbb{C}^n$. Then

$$\lambda(x, x) = (\lambda x, x) = (Ax, x) = (x, Ax) = (x, \lambda x) = \bar{\lambda}(x, x).$$

Dividing by (x, x) we see that $\lambda = \bar{\lambda}$.

If eigenvalues $\lambda \neq \mu$ of A correspond to eigenvectors x and y then

$$\lambda(x, y) = (Ax, y) = (x, Ay) = \mu(x, y)$$

so $(\lambda - \mu)(x, y) = 0$.

(ii) Notice that sufficiency in both the real symmetric and hermitian cases in trivial.

Let us consider necessity. Let $\mu_1, \dots, \mu_k$ denote the full set of distinct eigenvalues of A with respective eigenspaces $\mathcal{E}_j = \ker_{\mathbb{F}^n}(\mu_j I - A)$, $j = 1, \dots, k$. Since $\mathcal{E}_i \perp \mathcal{E}_j$ for $i \neq j$, as observed in (i), we may find an orthonormal basis

$$u_{(1)} = \begin{bmatrix} u_{11} \\ \vdots \\ u_{1n} \end{bmatrix}, \dots, u_{(m)} = \begin{bmatrix} u_{n1} \\ \vdots \\ u_{mn} \end{bmatrix}$$

for $\mathcal{E} = \mathcal{E}_1 + \cdots + \mathcal{E}_k$ such that $u_{(1)}, \ldots, u_{(n_1)}$ is a basis for $\mathcal{E}_1$, $u_{(n_1+1)}, \ldots, u_{(n_1+n_2)}$ is a basis for $\mathcal{E}_2$, ..., and $u_{(n_1+\cdots+n_{k-1})}, \ldots, u_{(m)}$ is a basis for $\mathcal{E}_k$. (Here each $n_j = \dim_{\mathbb{C}} \mathcal{E}_j$.) Let for $j = 1, \ldots, k$, P_j in $\mathrm{M}_n(\mathbb{F})$ be the matrix corresponding to the orthogonal projection onto $\mathcal{E}_j$, i.e. with entries

$$P_{j,i'j'} = \left(\sum_{l=0+n_1+\cdots+n_{j-1}}^{n_1+\cdots+n_j} (e_{j'}, u_{(l)}) u_{(l)}, e_{i'} \right) = \sum_{l=0+n_1+\cdots+n_{j-1}}^{n_1+\cdots+n_j} (e_{j'}, u_{(l)})(u_{(l)}, e_{i'})$$

which is easily seen to satisfy $P_j = P_j^*$. Let $B = A - \sum_{j=1}^k \mu_j P_j$, which is in $\mathrm{Herm}(n)$ (respectively, in $\mathrm{Sym}_n(\mathbb{R})$ if $\mathbb{F} = \mathbb{R}$) and satisfies $\mathcal{E} \subseteq \ker_{\mathbb{F}^n} B$. If $B \neq 0$, then it admits an eigenvalue $\mu \neq 0$ so (i) provides that its corresponding eigenvector $x \neq 0$ is in $\mathcal{E}^\perp$. But then $\mu x = Bx = Ax$, which means that μ is one of $\mu_1, \ldots, \mu_k$, above, contradicting that $x \notin \mathcal{E}$. Hence $B = 0$. Further, if $x \in \mathcal{E}^\perp$, then $Ax = Bx = 0$, so $x \in \mathcal{E} \cap \mathcal{E}^\perp$, so $x = 0$. Thus $\mathcal{E} = \mathbb{F}^n$, i.e. $m = n$.

Let u denote the matrix with columns $u_{(1)}, \ldots, u_{(n)}$, which, by the remark above, is unitary; and let $\lambda_1, \ldots, \lambda_n$ the respective eigenvalues μ_1 (n_1 times), ..., μ_k (n_k times). Let $e_1, \ldots, e_n$ denote the standard basis for $\mathbb{F}^n$. Then $ue_j = u_{(j)}$ and we have that

$$(u^* A u e_j, e_i) = (A u_{(j)}, u e_i) = \lambda_j (u_{(j)}, u_{(i)}) = \lambda_j (e_j, e_i)$$

for each $i, j = 1, \ldots, n$, so $u^* A u$ admits the claimed diagonal form. □

A representation of a symmetric/hermitian matrix. The proof above tells us that if $\mu_1, \ldots, \mu_k$ are the distinct eigenvalues of a symmetric (respectively, hermitian matrix) A and $P_1, \ldots, P_n$ are the matrices representing the orthogonal projections onto the respective eigenspaces $\mathcal{E}_1, \ldots, \mathcal{E}_k$ (which span all of $\mathbb{F}^n$), then

$$A = \sum_{j=1}^k \mu_j P_j \text{ where } I = \sum_{j=1}^k P_j. \qquad (\dagger)$$

Since $\mathcal{E}_i \perp \mathcal{E}_j$ if $i \neq j$, $P_i P_j = 0 = P_j P_i$.

Hence if $p(X) = \sum_{l=0} a_l X^l$ is any polynomial, we have

$$p(A) = \sum_{l=0} a_l A^l = \sum_{j=1}^k p(\mu_j) P_j$$

where $A^0 = I$, by convention.

Lemma. *Given A and $P_1, \ldots, P_k$ as above, another matrix B commutes with A, i.e. $[A, B] = AB - BA = 0$, if and only if $[P_j, B] = 0$ for each j.*

Proof. Sufficiency is evident from (†).

To see necessity, let for each j

$$p_j(A) = \frac{(A - \mu_1 I)\ldots(A - \mu_{j-1} I)(A - \mu_{j+1} I)\ldots(A - \mu_k I)}{(\mu_j - \mu_1)\ldots(\mu_j - \mu_{j-1})(\mu_j - \mu_{j+1})\ldots(\mu_j - \mu_k)}.$$

Clearly $[p_j(A), B] = 0$. Let $x \in \mathcal{E}_j = \ker_{\mathbb{F}^n}(A - \mu_j I)$. Then

$$p_j(A)x = \begin{cases} 0 & \text{if } i \neq j \\ x & \text{if } i = j. \end{cases}$$

Hence, since $\mathbb{F}^n = \mathcal{E}_1 + \cdots + \mathcal{E}_k$, $p_j(A)^2 = p_j(A)$. Also, it is obvious that $p_j(A)^* = p_j(A)$. Thus $p_j(A) = P_j$. □

Simultaneous Diagonalization Theorem. *If $A, B \in \mathrm{Sym}_n(\mathbb{R})$ (or in $\mathrm{Herm}(n)$), and $[A, B] = 0$, then there is v in $\mathrm{O}(n)$ (respectively, in $\mathrm{U}(n)$) for which*

$$v^*Av = \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & \lambda_n \end{bmatrix} \text{ and } v^*Bv = \begin{bmatrix} \nu_1 & 0 & \ldots & 0 \\ 0 & \nu_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & \nu_n \end{bmatrix}$$

where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A and $\nu_1, \ldots, \nu_n$ are the eigenvalues of B (each with multiplicity).

Proof. Consider the representations of A and B as in (†):

$$A = \sum_{j=1}^{k} \mu_j P_j \text{ and } B = \sum_{i=1}^{k'} \mu'_i P'_i.$$

Since $[A, B] = 0$, the lemma above provides that $[A, P'_i] = 0 = [P_j, B]$ for any i, j, and the lemma again provides that $[P_j, P'_i] = 0$ for any i, j. Hence each $P_j P'_i$ is self-adjoint and squares to itself, and is hence the orthogonal projection onto $\mathcal{E}_{ij} = \ker_{\mathbb{F}^n}(A - \mu_j I) \cap \ker_{\mathbb{F}^n}(B - \mu'_i I)$. Further (†) provides that

$$\sum_{j=1}^{k}\sum_{i=1}^{k'} P_j P'_i = \left(\sum_{j=1}^{k} P_k\right)\left(\sum_{i=1}^{k'} P'_i\right) = I \qquad (*)$$

so

$$A = AI = \sum_{j=1}^{k}\sum_{i=1}^{k'} \mu_j P_j P'_i \text{ and } B = IB = \sum_{j=1}^{k}\sum_{i=1}^{k'} \mu'_i P_j P'_i.$$

Take orthonormal bases for each of the non-zero spaces $\mathcal{E}_{ij}$ and combine them into an orthonormal basis

$$v_{(1)} = \begin{bmatrix} v_{11} \\ \vdots \\ v_{1n} \end{bmatrix}, \ldots, v_{(n)} = \begin{bmatrix} v_{n1} \\ \vdots \\ v_{nn} \end{bmatrix}$$

for $\mathbb{F}^n$ (this is possible by $(*)$). Let v be the matrix with columns $v_{(1)}, \ldots, v_{(n)}$, and we obtain the desired diagonal forms. □

Corollary. *A in $\mathrm{M}_n(\mathbb{C})$ is* normal, *i.e.* $[A, A^*] = 0$, *if and only if there is v in $\mathrm{U}(n)$ for which*

$$v^* A v = \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & \lambda_n \end{bmatrix}$$

where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A (with multiplicity).

Proof. Sufficiency being evident, we show only necessity. Let

$$\mathrm{Re}A = \frac{1}{2}(A + A^*) \text{ and } \mathrm{Im}A = \frac{1}{2i}(A - A^*)$$

so $\mathrm{Re}A, \mathrm{Im}A \in \mathrm{Herm}(n)$ and $A = \mathrm{Re}A + i\mathrm{Im}A$. It is easy to verify that A is normal if and only if $[\mathrm{Re}A, \mathrm{Im}A] = 0$. Hence simultaneous diagonalization, above, provides the necessary unitary diagonalizing matrix v. □

No real analogue. The real matrix $J_2 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ is normal, but admits only purely imaginary eigenvalues, and hence cannot be diagonalized by orthogonal matrices (i.e. unitary matrices with real entries).

Real skew-symmetric matrices. A matrix B in $\mathrm{M}_n(\mathbb{R})$ is *skew-symmtric* if $B^T = -B$.

Real Skew-symmetric Block Diagonalization Theorem. *If $B^T = -B$ in $\mathrm{M}_n(\mathbb{R})$ then there is u in $\mathrm{O}(n)$ and $\lambda_1, \ldots, \lambda_m > 0$ in $\mathbb{R}$ such that*

$$u^T B u = \begin{bmatrix} \lambda_1 J_2 & 0 & \ldots & \ldots & \ldots & 0 \\ 0 & \ddots & & & & \vdots \\ \vdots & & \lambda_m J_2 & & & \vdots \\ \vdots & & & 0 & & \vdots \\ \vdots & & & & \ddots & \vdots \\ 0 & \ldots & \ldots & \ldots & \ldots & 0 \end{bmatrix}$$

where $J_2 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$, *i.e.* $2m \leq n$.

Proof. First, notice that $iB \in \mathrm{Herm}(n)$ and hence has real eigenvalues so B has purely imaginary eigenvalues (including, possibly 0). In particular, the only real eigenvectors may be in $\ker B$.

Consider $B^T B$, which is in $\mathrm{Sym}_n(\mathbb{R})$. Any eigenvalue μ of $B^T B$ with eigenvector x in $\mathbb{R}^n \setminus \{0\}$ satisfies

$$\mu(x, x) = (B^T B x, x) = (Bx, Bx) \geq 0$$

so $\mu \geq 0$. Let $\mu_1, \ldots, \mu_l$ denote the distinct non-zero eigenvalues of $B^T B$ and $\mathcal{E}_j$ the eigenspace of μ_j, so $\mathcal{E}_i \perp \mathcal{E}_j$ for $i \neq j$ and each $\mathcal{E}_j \perp \ker B^T B$. Let $\mathcal{E} = \mathcal{E}_1 + \cdots + \mathcal{E}_l$.

Let $x \in \mathcal{E}_j \setminus \{0\}$ and $\mathcal{V}_x = \mathbb{R}x + \mathbb{R}Bx$. Then

$$B(Bx) = B^2 x = -B^T B x = -\mu_j B x \in \mathcal{V}_x$$

and it follows that $\mathcal{V}_x$ is B-invariant, i.e. $B\mathcal{V}_x \subseteq \mathcal{V}_x$. Furthermore $\dim_{\mathbb{R}} \mathcal{V}_x = 2$ as B admits no non-zero real eigenvalues and $\ker B = \ker B^T B$. Now if $y \in \mathcal{E}_j \setminus \{0\}$ with $y \perp x$ then

$$\begin{aligned} (By, x) &= (y, B^T x) = -(y, x) = 0 \text{ and} \\ (By, Bx) &= (y, B^T B x) = -\mu_j (y, Bx) = 0 \end{aligned}$$

so $\mathcal{V}_y \perp \mathcal{V}_x$. Hence we may build an orthonormal basis $u_1^{(j)}, \dots, u_{l_j}^{(j)}$ for $\mathcal{E}_j$ such that each $u_{2i}^{(j)} \in \mathcal{V}_{u_{2i-1}^{(j)}}$ and $\mathcal{V}_{u_{2i-1+2i'}^{(j)}} \perp \mathcal{V}_{u_{2i-1}^{(j)}}$ for $i' = 1, \dots, \lfloor l_j/2 \rfloor$, and, in particular, l_j is even.

Putting everything together we have that $\dim \mathcal{E} = \sum_{j=1}^{l} \dim \mathcal{E}_j$ is even, and we can find an orthonormal basis $u_1, \dots, u_{2m}, u_{2m+1}, \dots, u_n$ for $\mathbb{R}^n$ for which the spaces $\mathcal{V}_j = \mathcal{V}_{u_{2j-1}}$ are pairwise orthogonal and span $\mathcal{E}$, and $u_{2m+1}, \dots, u_n \in \ker B$. Letting u be the matrix whose rows $u_1, \dots, u_n$ we find that

$$u^T B u = \begin{bmatrix} B_1 & 0 & \dots & \dots & \dots & 0 \\ 0 & \ddots & & & & \vdots \\ \vdots & & B_m & & & \vdots \\ \vdots & & & 0 & & \vdots \\ \vdots & & & & \ddots & \vdots \\ 0 & \dots & \dots & \dots & \dots & 0 \end{bmatrix}$$

where each $B_j \in \mathrm{M}_2(\mathbb{R})$. Since $(u^T B u)^T = u^T B^T u = -u^T B u$ we find that each block must have the form $B_j = \lambda_j J_2$ with λ_j in $\mathbb{R} \setminus \{0\}$, and by applying block permutations we may assume $\lambda_j > 0$. (One may further check that the values $\lambda_1, \dots, \lambda_m$ are the values $\sqrt{\mu_1}, \dots, \sqrt{\mu_k}$ with multiplicities.) $\square$

Remark. The complex analogue of this result is much easier. If $B \in \mathrm{M}_n(\mathbb{C})$ with $B^* = -B$ we can B *skew-hermitian*. Notice that $iB \in \mathrm{Herm}(n)$ and is hence unitarily diagonalizable with real eigenvalues, so B too is unitarily diagonalizable but with purely imaginary eigenvalues.