In the following are equivalent statements for a real $n \times n$ symmetric matrix $A$:

(i) $A$ is strictly positive definite

(ii) $d_i(A) > 0$ for all eigenvalues of $A$, $i = 1, \ldots, n$

(iii) The principle minors $d_i$, $i = 1, \ldots, n$ are $> 0$

(iv) $A = C^T C$ for an upper triangular matrix $C$, $\det C \neq 0$.

(v) $A = B^2$, $\det B \neq 0$, for some real symmetric matrix $B$.

**Proof:** We have already proved (i) $\Leftrightarrow$ (ii). By Sylvester's theorem $(i) \Rightarrow (iii)$.

(v) $\Rightarrow$ (i): If $A = B^2$, $B = B^T$, $\det B \neq 0$,

\[
\langle u, Au \rangle = \langle u, B^T B u \rangle = \langle Bu, Bu \rangle = \| Bu \|^2 > 0
\]

Let $\delta = \inf \{ \| Bu \| : \| u \| = 1 \}$. As $\| u \| = 1 \Rightarrow u \rightarrow \| Bu \|

is a continuous function on a compact set, it must

achieve its infimum at some point $u_0$, $\| u_0 \| = 1$.

Thus $\langle u, Au \rangle = \| Bu_0 \|^2 > 0$ for $\| u \| = 1$.

If $\| Bu_0 \| = 0$, then as our $B \neq 0$, $u_0 = 0$, which is a contradiction. Thus $\delta > 0$ and $\langle u, Au \rangle > \delta \| u \|^2$. Thus (i) is true.
(i) \Rightarrow (v) \quad A = U \Lambda U^T, \quad U \text{ real orthogonal}, \quad \Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n), \\
\lambda_i \geq 0, \quad i = 1, \ldots, n \\

Set \quad B = U \Lambda^2 U^T \quad \text{where} \quad \Lambda^2 = \text{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n}) \\
Then \quad A = B^2, \quad B = B^T \quad \text{and} \quad \det B > 0 \\
Clearly \quad B \text{ is real as } U \text{ and } \Lambda^2 \text{ are real.} \\

(iv) \Rightarrow (v) \quad (u, Au) = (u, C^T C u) = \| C u \|^2 \\
As above, \exists \text{ and } \| C u \| > 0 \quad \text{and so} \quad (u, Au) > \| C u \|^2 \| u \|^2 \\

(ii) \Rightarrow (iv) \quad \text{Do Gaussian elimination on } A \text{ as } a_{ii} = a_{ii} > 0 \\

\[
A = \begin{pmatrix}
\ddots & & \\
& a_{ii} & a_{in} \\
& \vdots & \ddots \\
& a_{ni} & \cdots & a_{nn}
\end{pmatrix} \Rightarrow \begin{pmatrix}
\ddots & & \\
& a_{ii} & a_{in} \\
& 0 & a_{ii} & \cdots & a_{in} \\
& \vdots & \ddots & \ddots & \ddots \\
& 0 & a_{ni} & \cdots & a_{nn}
\end{pmatrix}
\]

Now the sub-matrix \[
\begin{pmatrix}
a_{ii} & a_{i2} \\
0 & a_{22}
\end{pmatrix}
\]

is obtained by adding a multiple of the row \( (a_{i1}, a_{i2}) \) to \( (a_{21}, a_{22}) \)

Therefore, \[
a_{i2} = a_{i2} - (a_{i1} a_{i2}) = \text{det} \begin{pmatrix}
a_{ii} & a_{i2} \\
0 & a_{22}
\end{pmatrix} = a_{22} > 0
\]

Hence \( a_{i2} = a_{i2} / a_{ii} > 0 \)

Thus we can continue the reduction:

\[
\begin{pmatrix}
a_{ii} & a_{i2} & \cdots & a_{in} \\
0 & a_{i2} & \cdots & a_{in} \\
\vdots & \ddots & \ddots & \ddots \\
0 & \cdots & 0 & a_{nn}
\end{pmatrix} \Rightarrow \begin{pmatrix}
a_{ii} & a_{i2} & \cdots & a_{in} \\
0 & a_{i2} & \cdots & a_{in} \\
\vdots & \ddots & \ddots & \ddots \\
0 & \cdots & 0 & a_{nn}
\end{pmatrix}
\]
Arguing as above, we have
\[ a_{33}' = \frac{d_3}{a_{11} a_{22}} > 0 \]

etc. Thus by Gaussian elimination, we can reduce \( A \rightarrow U \), where \( U \) is real and upper triangular with positive entries on the diagonal. Now Gaussian elimination is implemented by multiplying \( A \) at each step on the left by a lower triangular matrix with 1's on the diagonal

\[
\begin{pmatrix}
1 & 0 & 0 \\
\frac{d_{11}}{a_{11}} & 1 & 0 \\
\frac{d_{31}}{a_{31}} & 0 & 1
\end{pmatrix}
\begin{pmatrix}
a_{11} & a_{12} & a_{13} \\
a_{21} & a_{22} & a_{23} \\
a_{31} & a_{32} & a_{33}
\end{pmatrix}
= 
\begin{pmatrix}
a_{11} & a_{12} & a_{13} \\
0 & a_{22} & a_{23} \\
0 & 0 & a_{33}
\end{pmatrix}
\]

where \( d_{11} = -a_{11} / a_{11}, \quad d_{31} = -a_{31} / a_{11} \)

Thus we have
\[
L A = U
\]

for some lower triangular matrix with 1's on diagonal

Clearly \( \det L = 1 \neq 0 \) and no \( L^{-1} \) exists and must also be lower triangular (why?). Hence
\[
A = SU
\]

\( S \) lower, \( S_{ii} = 1, U \) upper \( U_{ii} > 0 \). Set \( U = DV \) where
\[ D = \text{diag}(u_1, \ldots, u_n) \] Clearly \( V \) is upper unit
\[ V_{ii} = 1, \quad i = 1, \ldots, n. \] Have \( A = SDV \). But \( A = A^T \)
and so \( A = V^TDS^T \). Equating these 2 expressions
for \( A \) we find \( WD = DW^T \) where \( W = V^TS \).

A lower triangular. Thus \( WD \) is lower, but \( DW^T \) is upper. Hence \( WD \) must be diagonal, and so
\[ V^T S = \Delta \]
for some diagonal matrix \( \Delta \) i.e. \( S = V^T \Delta \).

But \( S_{ii} = V_{ii} = 1, \quad i = 1, \ldots, n \). Hence \( \Delta = I \) and
so \( V = S^T \). Thus
\[ A = V^T D V \]
Set \( C = \text{diag}(\sqrt{u_1}, \ldots, \sqrt{u_n}) \) \( V \) and so \( A = C^T C \),
which is (iv).

Thus proves the Proposition.

\[ (i) \xleftarrow{(\ast)} (i) \xrightarrow{\text{cl}} (ii) \xrightarrow{\text{cl}} (iii) \]

\[ (i) \xrightarrow{(iv)} \]

We now begin addressing the following basic questions:

How do the eigenvectors of \( A \) depend on \( A \)? How do the eigenvectors of \( A \) depend on \( A \)?
Using Rao's Theorem (see Problem #5, Problem set #7)

All eigenvalues of \( nxn \) \( A \) are continuous functions of \( A \) in the following sense: Let \( \lambda_1, \ldots, \lambda_n \) be the eigenvalues of \( A \), counting (algebraic) multiplicity. Draw disks \( D(\lambda_i, \epsilon) \) of radius \( \epsilon > 0 \) around all the distinct eigenvalues \( \lambda_i \) of \( A \), where \( \epsilon \) is sufficiently small so that each disk contains only one (distinct) eigenvalue of \( A \). For example if \( n = 4 \) and \( \lambda_1, \lambda_2, \lambda_3, \lambda_4 \) are distinct, but \( \lambda_1 = \lambda_2 \). Then we draw

\[
\begin{array}{c}
\epsilon \\
D(\lambda_1, \epsilon) \\
D(\lambda_2, \epsilon) \\
D(\lambda_3, \epsilon) \\
D(\lambda_4, \epsilon)
\end{array}
\]

Then if \( \| B - A \| \) is sufficiently small, \( \epsilon \), the eigenvalues of \( B \) lie in these disks, and there are as many eigenvalues of \( B \), counting multiplicity, in each of these disks as the multiplicity of the eigenvalue of \( A \) at the center of the disk.

So in the example above for \( \| B - A \| \) sufficiently small, there are 2 eigenvalues of \( B \), counting multiplicity, in the disk around \( \lambda_1 = \lambda_2 \), and one eigenvalue in the other 2 disks.
We want to ask further questions: For example, are the eigenvalues and eigenvectors analytic functions of (the entries) of A?

Note: But in general the eigenvalues are not analytic functions of A. For example, for \( n = 2 \) with \( A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \),

\[ \lambda^2 - \nu \lambda + \Delta = 0 \]

is the eigenvalue equation, where \( \nu = a + d, \Delta = ad - bc. \)

Then

\[ \lambda = \frac{\nu \pm \sqrt{\nu^2 - 4\Delta}}{2} \]

But \( \nu^2 - 4\Delta = (a + d)^2 - 4(ad - bc) = (a - d)^2 + 4bc. \)

Now \( \sqrt{\nu^2 - 4\Delta} = \sqrt{(a - d)^2 + 4bc} \), \( \nu \) is an analytic function of \((a, b, c, d)\) in the neighborhood of any point \((a_0, b_0, c_0, d_0)\) where \( (a_0 - d_0)^2 + 4b_0c_0 \neq 0 \). But if \( (a_0 - d_0)^2 + 4b_0c_0 = 0 \) then \( \lambda \) is clearly not analytic.

In fact, its derivative blows up at this point. The takeaway trouble occurs at points at which \( \lambda_1 = \lambda_2 \), and only at such points.

We consider first the case where \( A(\rho) = (A_\nu(\rho)) \) depends analytically on one complete variable \( \rho \) lying in some region \( \mathbb{C} \). The standard example is \( A(\rho) = A + \rho B \).
where $A$ and $B$ are given $n \times n$ matrices. To see what can happen, consider the following examples (taken from T. Kato, Perturbation Theory for Linear Operators, Chap. II — this is a basic reference for perturbation theory)

c) $A(\beta) = \begin{pmatrix} 1 & \beta \\ \beta & -1 \end{pmatrix}$

For the eigenvalues, we have $\lambda = \pm \sqrt{1+\beta^2}$.

For $\beta = 0$, the eigenvalues are $\pm 1$ so $\lambda_+(0) = \lambda_-(0)$, i.e., the spectrum is simple, and for $|\beta|$ small, we see that $A(\beta)$ and $\lambda(\beta)$ are analytic functions of $\beta$. The spectrum remains simple as long as $\beta \neq \pm i$. The eigenvectors are $v_\pm(\beta) = \begin{pmatrix} -\beta \\ 1 + \sqrt{1+\beta^2} \end{pmatrix}$

As long as $\beta \neq \pm i$, the eigenvectors are analytic and independent. However, if $\beta = \pm i$, then the eigenvalues and eigenvectors are no longer analytic at $\beta$, and moreover, the eigenvectors are no longer independent. Note also
That at \( \beta = \pm i \), \( A(\beta) = A(i) = \begin{pmatrix} i & 1 \\ i & -1 \end{pmatrix} \) is not diagonalizable: indeed as \( \lambda_{\pm}(\pm i) = \lambda_{\pm}(i) = 0 \), \( A(i) \) must be zero if it is diagonalizable, \( A(\pm i) = U \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} U^{-1} = 0 \), which is a contradiction.

If \( \beta \) is real, the eigenvector can be normalized

\[
(126.1) \quad \hat{\mathbf{v}}_{\pm}(\beta) = \left( \frac{\beta}{\sqrt{2(1 + \beta^2)}}, \frac{1 - \beta}{\sqrt{2(1 + \beta^2)}} \right)
\]

with \( \| \hat{\mathbf{v}}_{\pm}(\beta) \| = 1 \), in such a way that \( \hat{\mathbf{v}}_{\pm}(\beta) \) depends real analytically on \( \beta \).

(Exercise: Check that \( \hat{\mathbf{v}}_{\pm}(\beta) \) are indeed eigenvector, \( \| \hat{\mathbf{v}}_{\pm}(\beta) \| = 1 \) and that \( \hat{\mathbf{v}}_{\pm}(\beta) \) are analytic \( \forall \beta \in \mathbb{R} \), and \( \langle \hat{\mathbf{v}}_{\mp}(\beta), \hat{\mathbf{v}}_{\pm}(\beta) \rangle = 0 \).)

Note that \( \lambda_{+}(\beta) \neq \lambda_{-}(\beta) \) if \( \beta \neq 0 \).

(b) \( A(\beta) = \begin{pmatrix} \beta & 1 \\ 0 & \beta \end{pmatrix} \). Here the eigenvalues are \( \lambda_{\pm}(\beta) = \pm \beta \).

For \( \beta = 0 \), both eigenvalues are 0 and the spectrum is not simple. Nevertheless, the eigenvectors are \( \mathbf{v}_{\pm}(\beta) = (1, \pm 1)^T \), which can be normalized (trivially) for \( \beta \) real (in this case \( \forall \beta \)).
$$\hat{v}_\pm(\beta) = (\frac{1}{\sqrt{2}}, \pm \frac{1}{\sqrt{2}})^T$$ such that \(\hat{v}_\pm(\beta)\) is analytic for all real \(\beta\) (cf. (26.1)) (in fact for all \(\beta\) in this case). Note that at \(\beta = 0\), \(A(0) = 0\), so that every vector is an eigenvector for \(A(0)\). To ensure analytic behavior of \(v_\pm(\beta)\) as \(\beta \to 0\), however, we cannot choose the eigenvector of \(A(0)\) freely; we must choose special vectors in \(N(A(0))\), viz.,

\[(27.1) \quad v_\pm(0) = (\frac{1}{\sqrt{2}}, \pm \frac{1}{\sqrt{2}})^T\]

This is an illustrative example of singular perturbation theory.

Here we are trying to perturb the eigenvalues of \(A(0) = 0\) where \(A(0) = \lambda(0)\). The problem is singular because the eigenvectors of \(A(0)\) are not well-determined. It is the nature of the perturbation \(A(0) = \lambda(\beta)\) that determines which eigenvector \(v_\pm(0)\) of \(A(0)\) extends smoothly to \(v_\pm(\beta)\).

For example, consider \(\hat{A}(\beta) = \begin{pmatrix} \beta & 0 \\ 0 & \beta \end{pmatrix}\) with eigenvalues
$\lambda \pm (\beta) = \beta \left( \frac{1 \pm \sqrt{5}}{2} \right)$ and associated eigenvectors

$\mathbf{v}_\pm (\beta) = \left( 1, \ -1 \pm \frac{\sqrt{5}}{2} \right) \ T$  \ 

So again,

we are dealing with a singular problem $\mathbf{A}(0) = 0$.

and the eigenvectors of $\mathbf{A}(0)$ that extend analytically are

$\left( 127.1 \right) \quad \mathbf{v}_\pm (0) = \left( 1, \ -1 \pm \frac{\sqrt{5}}{2} \right) \ T$

This should be compared with (127.1). The takeaway is the following: the ambiguity in the eigenvectors at $\beta = 0$ is resolved by the nature of the perturbation.

$\mathbf{A}(0) \rightarrow \mathbf{A}(\beta)$. Singular perturbation problems are very common in physics and applied mathematics: the unperturbed problem is singular in some sense, and the ambiguity in the problem is resolved by the specific nature of the perturbation.

$\left( 128 \right) \quad \mathbf{A}(\beta) = \left( \begin{array}{cc} 0 & \beta \\ 0 & 0 \end{array} \right)$, Again the problem is singular at $\beta = 0$; $\mathbf{A}(0) = 0$ and so the eigenvectors of $\mathbf{A}(0)$ are not uniquely
determined. However, for $\beta \neq 0$, we still have

$$\lambda_{\pm}(\beta) = 0,$$

but $N(A(\beta)) = \langle (0) \rangle$ and thus

in no choice of a basis $(v_+(0), v_-(0))$ for $N(A(0))$

which continues analytically $v_{\pm}(0) \to v_{\pm}(\beta)$ as $A(0) \to A(\beta)$.

Only $v_{\pm}(\beta) = (0)$ continues. We note that the difference

between (b) and (c) is that in case (b), $A(\beta)$ is Hermitian

for $\beta$ real, but not in case (c).

(1) $A(\beta) = \begin{pmatrix} 0 & 1 \\ -\beta & 0 \end{pmatrix}$. Here $\lambda_{\pm}(\beta) = \pm i \beta$. We see that

in this case there is no way to choose the branches of

$\sqrt{\beta}$ so that $\lambda_{\pm}(\beta)$ are analytic in $\beta$ in a neighborhood

d of $\beta = 0$. Note that here the problem is singular in

the sense that $\lambda_{\pm}(0) = 0$ but $A(0)$ has only 1

eigenvalue $v(0) = (1, 0)^T$. For $\beta \neq 0$, $v_{\pm}(\beta) = (1, \pm i \beta)^T$

Thus the single eigenvalue $v(0) = (1, 0)^T$ differentiates to two

eigenvalues $v_{\pm}(\beta) = (1, \pm \sqrt{\beta})^T$ as $0 \to \beta$.

Again we note
That even for $\beta$ real, $\beta \neq 1$, $A(\beta)$ is not Hermitian.

Contrast the singular perturbation problems (c) and (d), $A(\beta)$ has two eigenvectors, but we see that in case (c), only one of the eigenvectors

continues as $A(0) \rightarrow A(\beta)$. However, in case (d), $A(0)$ has only 1 eigenvector, which then bifurcates, non-analytically, to two eigenvectors as $A(0) \rightarrow A(\beta)$.

(c) $A(\beta) = \begin{pmatrix} 1 & \beta \\ 0 & 0 \end{pmatrix}$. Here $\lambda_+(\beta) = 1$ and $\lambda_-(\beta) = 0$,

which are clearly analytic (in fact constant) functions of $\beta$. The eigenvectors are $v_+(\beta) = (1, 0)$, $v_-(\beta) = (-\beta, 1)$,

which can be normalised as $\tilde{v}_+(\beta) = (1, 0)$, $\tilde{v}_-(\beta) = \left( \frac{-\beta}{\sqrt{1+\beta^2}} \right, 1)\sqrt{1+\beta^2}$

for $\beta$ real. Note that this is true even as $A(\beta)$ is not Hermitian, $\beta \in \mathbb{R}$.

(d) $A(\beta) = \begin{pmatrix} 0 & \beta \\ 0 & 0 \end{pmatrix}$. Here $\lambda_+(\beta) = \beta$, $\lambda_-(\beta) = 0$, which are clearly analytic in $\beta$.

Again the problem is singular at $\beta = 0$, as $\lambda_+ / \lambda_1 = 0$.

Here the eigenvectors are $v_+(\beta) = (0, 1)$ and $v_-(\beta) = (-1, 0)$, which
are analytic in $\beta$. Note that, as opposed to case (d), which also has a simple eigenvalue at $\beta(0)$, and for which the eigenvector bifurcates non-analytically, as $\epsilon \to 0^-$, in this case the eigenvector at $\beta = 0$, bifurcates analytically as $\epsilon \to 0^-$. Note also that for $\beta$ real, the eigenvector can be normalized in an analytic fashion:

$$\hat{\psi}_+ (\beta) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \hat{\psi}_- (\beta) = \begin{pmatrix} \frac{1 + \beta \epsilon}{\sqrt{1 + \beta^2}} \\ \frac{-\beta}{\sqrt{1 + \beta^2}} \end{pmatrix},$$

In summary, what we observe/guess from these examples is that everything is "nice", i.e., all the eigenvalues and eigenvectors depend nicely on $\beta$ in a neighborhood of a simple eigenvalue, when the eigenvalue is multiple, then things are still "nice" in the self-adjoint case $A(\beta) = A(\beta^*)$, $\beta \in \mathbb{R}$, but, in general, not in the non-self-adjoint case (but there are
In order to develop perturbation theory in full for n x n matrices, we need to introduce some concepts/definitions.

References for Perturbation Theory
1. T. Kato, as above, p. 225
2. Reed-Simon, Methods of Modern Math. Physics, Vol IV
3. F. Rellich, Perturbation Theory of Eigenvalue Problems

We say that a square matrix \( P \) is a projection if \( P^2 = P \). For a projection \( P \) we have:

\[
M_P = R(P), \quad N_P = N(P).
\]

Clearly \( R(P) \) and \( N(P) \) are subspaces.

Note that if \( x \in R(P) \), then \( x = Py \) for some \( y \) and no \( Px = P^2 y = Py = x \). Thus

\[
R(P) = \{ x : x = Px \}.
\]

**Proposition.** Let \( V = \mathbb{R}^n \) on \( \mathbb{R}^n \) and let \( P : V \to V \) be a projection. Then \( V \) has a direct decomposition \( V = R(P) \oplus N(P) \) if each \( x \in V \) has a unique representation

\[
x = r + n
\]

where \( r \in R(P) \) and \( n \in N(P) \).
Conversely, if 
\[ V = X \oplus Y \]
and a direct decomposition of \( V \), then
\[ X = R(P) \quad \text{and} \quad Y = N(P) \]
for some projection \( P : V \to V \).

\textbf{Proof:} Let \( x \in V \), then 
\[ x = Px + (1-P)x \]
(\( Px \in R(P) \) and \( (1-P)x \in N(P) \))
we see that \( (1-P)x \in N(P) \).

Now suppose \( x' = x' + n' \) where \( x', r' \in R(P) \) and \( n, n' \in N(P) \). Then \( r - r' = n - n' \)
so \( P(x-r') = P(n') - P(n) = 0 \). Hence \( r = P r = P r' = 5' \).
But then \( n = n' \). Thus \( R(P) \oplus N(P) \) is indeed \( a \) direct decomposition.

Conversely, suppose 
\[ V = X \oplus Y \]
is a direct decomposition of \( V \) into subspaces \( x \oplus y \).

Set 
\[ Px = x \quad \text{if} \quad x \in X \]
\[ Py = 0 \quad \text{if} \quad y \in Y \]
\[ P' : \]
Then \( P' (x + y) = P(Px + Py) = Px \) \( \quad \) as \( Px \in X \) is \( (P'(Px + Py) = P'Px = Px \). \] But \( Py = 0 \) and \( Py = 0 \).
\[ p^2(x + y) = p(x + y) \quad \text{if} \quad p^2 = p. \]

Now suppose \( x \in \mathbb{X} \). Then \( x = px \in R(p) \), so \( x \subseteq R(p) \). On the other hand, if \( x \subseteq R(p) \), then \( x = pu \) for some \( u \in U \). But \( u = x' + v' \) for some \( x' \in \mathbb{X}, v' \in \mathbb{Y} \).

Hence \( x = pu = px' + py' = px' = x' \in \mathbb{X} \).

Thus \( R(p) \subseteq \mathbb{X} \) and so \( \mathbb{X} = R(p) \).

If \( y \in \mathbb{Y} \), then \( py = 0 \) and so \( y \in N(p) \).

Thus \( N(p) \subseteq \mathbb{X} \). On the other hand, if for some \( y \), \( py = 0 \), then \( y = x' + v' \), \( x' \in \mathbb{X}, v' \in \mathbb{Y} \).

Hence \( 0 = py = px' + py' = px' = x' \). So \( x' = 0 \) and hence \( y = y' \in \mathbb{Y} \). Thus \( \mathbb{Y} = N(p) \), which proves the result. \( \square \).