ME617 - Handout 9

Solving the eigenvalue problem -
Numerical Evaluation of Natural Modes and Frequencies in MDOF systems

The **standard eigenvalue problem** is

$$\mathbf{Ax} = \lambda \mathbf{x}$$

(1)

The solution of eigenvalue systems is fairly complicated. It is one of the few subjects in numerical analysis where **I do** recommend using canned routines. This handout will give you an appreciation of what goes on inside such canned routines. The knowledge below will help you to make an intelligent choice when using or selecting one of the methods detailed.

The references listed at the end of this document are the ones your lecturer is familiar with. Nowadays there are many resources available on line.

A good **eigenpackage** provides separate routines, or separate paths through sequences of routines, for the following (desired) calculations:
   * **all** eigenvalues and **no** eigenvectors (a polynomial root solver)
   * **some** eigenvalues and **some** corresponding eigenvectors
   * **all** eigenvalues and **all** corresponding eigenvectors.

Take the items above into consideration when selecting an eigenvalue solver to save computing time and storage.

- A good **eigenpackage** also provides separate paths for special forms of matrix **A.**
   * **A** is real and symmetric: $\mathbf{A} = \mathbf{A}^T$ (full, banded, tridiagonal).
   * **A** is real and not symmetric (full, banded, tridiagonal).

- **A** is complex, Hermitian or not Herminian.

Most packages solve the **generalized eigenproblem**

$$\mathbf{Ax} = \lambda \, \mathbf{Bx} \tag{2}$$

Bathe and Wilson (1976), Ref. [3], recommend:
1. The **Householder-QR inverse(HQRI)** solution is the most efficient for calculation of <u>all</u> eigenvalues and eigenvectors and if **A** is full.

2. The **determinant search technique** (polynomial root solver) is most appropriate to determine the **lowest** eigenvalue and corresponding eigenvectors of systems with a small bandwidth.
3. The **subspace iteration solution** is effective for the <u>lowest</u> eigenvalue and eigenvectors in very large systems with a large bandwidth.

In **mechanical vibrations**, the general eigenvalue problem for an undamped MDOF system must satisfy:

$$\left[-\mathbf{M}\, \lambda_i + \mathbf{K} \right] \boldsymbol{\varphi}_{(i)} = \mathbf{0}, \quad i=1,...n \tag{3}$$

where $\lambda_i = \omega_i^2$ is the square of the i_{-th} natural frequency and $\mathbf{K} = \mathbf{K}^T$, $\mathbf{M} = \mathbf{M}^T$ are symmetric stiffness and mass matrices, usually positive definite.

Eigen methods that exploit Eq. (3) are called **vector iteration or (power) methods.**

Matrix transformation methods are based on the orthogonality

properties of the **NORMALIZED** natural modes:

$$\boldsymbol{\Phi}^T \mathbf{K} \ \boldsymbol{\Phi} = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_n \end{pmatrix} ; \quad \boldsymbol{\Phi}^T \mathbf{M} \ \boldsymbol{\Phi} = \mathbf{I} \tag{4}$$

and are based on finding regular transformation to the original problem Eq. (3) until Eq. (4) becomes true. The Householder (H) QR and Jacobi's methods are among those which use orthogonal transformations.

Finally, from Eq. (3), recall that:

$$\Delta = | \mathbf{K} - \lambda \mathbf{M} | = 0 \tag{5}$$

$$\Delta = 0 = a_0 + a_1 \lambda + a_2 \lambda^2 + a_3 \lambda^3 + \dots a_n \lambda^n = \sum_{i=0}^{n} \left(a_i \lambda^i \right)$$

i.e. a characteristic equation; and by using a root solver, the set of eigenvalues $\{\lambda_i\}_{,\ i=1,\dots n}$ can be obtained.

The **HQRI method** uses a combination of iteration procedures and transformations.

Before proceeding further, let's show how Eq. (3) is brought into the standard eigenvalue problem Eq. (1), i.e. $\mathbf{Ax} = \lambda \mathbf{x}$, where $\mathbf{A}$ is a symmetric matrix. Note that the computational solution of a symmetric system is most effective in time and cost.

From Eq. (3)

$$\mathbf{K}\boldsymbol{\varphi} = \lambda \mathbf{M} \boldsymbol{\varphi} \tag{6}$$

One can define the product $\quad \mathbf{B} = \mathbf{M}^{-1}\mathbf{K} \quad$ (6a)
and write Eq. (6) as

$$\mathbf{B}\,\boldsymbol{\varphi} = \lambda\,\boldsymbol{\varphi} \qquad (6b)$$

but, unfortunately, matrix $\mathbf{B}$ is not (necessarily) symmetric, i.e. $\mathbf{B} \neq \mathbf{B}^{\mathbf{T}}$.

However, since $\mathbf{M} = \mathbf{M}^{\mathbf{T}}$ is symmetric and positive definite, there exists a decomposition of $\mathbf{M}$ as

$$\mathbf{M} = \mathbf{LL}^{\mathbf{T}} \qquad (7)$$

where $\mathbf{L}$ is a **lower triangular matrix** and $\mathbf{L}^{\mathbf{T}}$ is the **upper triangular transpose** of $\mathbf{L}$. (better known as the **Choleski decomposition**, and takes Order (*n*) operations to be performed.)

Replacing Eq. (7) into Eq. (6) gives:

$$\mathbf{K}\,\boldsymbol{\varphi} = \lambda\left(\mathbf{LL}^{\mathbf{T}}\right)\boldsymbol{\varphi} \qquad (8a)$$

Premultiply both sides by $\mathbf{L}^{-1}$ to get:

$$\mathbf{L}^{-1}\mathbf{K}\,\boldsymbol{\varphi} = \lambda\left(\mathbf{L}^{-1}\mathbf{L}\right)\mathbf{L}^{\mathbf{T}}\boldsymbol{\varphi} = \lambda\,\mathbf{L}^{\mathbf{T}}\boldsymbol{\varphi} \qquad (8b)$$

define vector

$$\mathbf{X} = \mathbf{L}^{\mathbf{T}}\boldsymbol{\varphi} \qquad (8c)$$

so then:

$$\mathbf{L}^{-\mathbf{T}}\mathbf{X} = \mathbf{L}^{-\mathbf{T}}\mathbf{L}^{\mathbf{T}}\boldsymbol{\varphi} = \boldsymbol{\varphi} \qquad (8d)$$

Substitute Eq. (8d) into Eq. (8b) to get:

$$\mathbf{L^{-1}\,K\,L^{-T}\,X = \lambda\,X} \qquad (9)$$

and let:

$$\mathbf{A = L^{-1}\,K\,L^{-T}} \qquad (10)$$

so then Eq. (9) becomes:

$$\mathbf{A\,X = \lambda\,X} \qquad (11)$$

which is the equivalent of Eq. (6), i.e., the eigenvalues are the same (identical) but the eigenvectors $\boldsymbol{\varphi}$ are obtained from Eq. (8d) $\boldsymbol{\varphi = L^{-T}X}$.

Notice that $\mathbf{A = A^T}$, i.e. a symmetric matrix.

$$\mathbf{A^T = (\,L^{-1}\,K\,L^{-T}\,)^T \equiv (\,L^{-T})^T\,(\,L^{-1}\,K\,)^T}$$
$$\mathbf{\equiv L^{-1}\,K^T\,L^{-T} \equiv L^{-1}\,K\,L^{-T} \equiv A}$$

$$\text{since} \qquad \mathbf{K = K^T}$$

Thus, solve eigenvalue problem defined by Eq. (11), obtain the set of $\left\{\lambda_{(i)}\right\}_{i=1}^{n}$ and associated eigenvectors $\left\{\mathbf{X}_{(i)}\right\}_{i=1}^{n}$ and, then, later obtain the eigenvectors for the problem with $\boldsymbol{\varphi}_{(i)} = \mathbf{L}^{-T}\mathbf{X}_{(i)}$.

Note also that, from the eigenvectors **orthonormal property**:

$$\boldsymbol{\Phi}^{T}\mathbf{M}\,\boldsymbol{\Phi} = \mathbf{I}$$

replacing $\boldsymbol{\varphi} = \mathbf{L}^{-T}\mathbf{X}$ on the LHS of this equation:
i.e.,

$$\text{LHS} = \left(\mathbf{X}^{T}\mathbf{L}^{-1}\right)\mathbf{M}\left(\mathbf{L}^{-T}\mathbf{X}\right) = \mathbf{I} \; ; \quad \text{and since } \mathbf{M} = \mathbf{L}\,\mathbf{L}^{T}$$

$$= \mathbf{X}^{T}\left(\mathbf{L}^{-1}\mathbf{L}\right)\left(\mathbf{L}^{T}\mathbf{L}^{-T}\right)\mathbf{X} = \mathbf{X}^{T}\,\mathbf{I}\,\mathbf{I}\,\mathbf{X} = \mathbf{X}^{T}\,\mathbf{X}$$
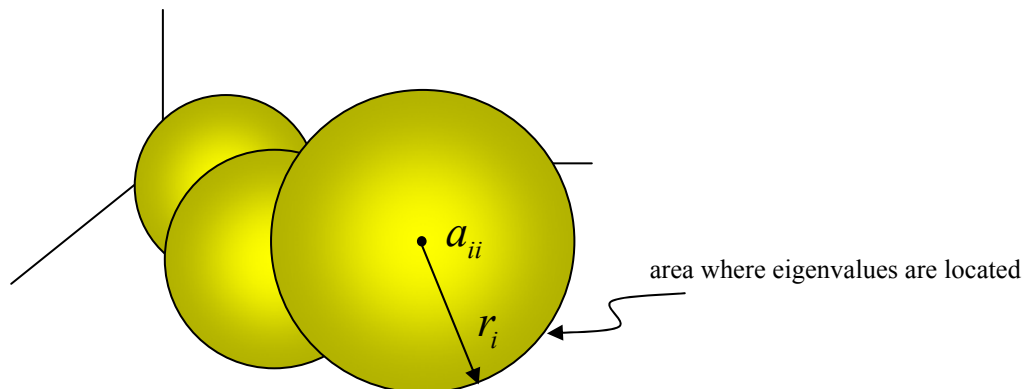
i.e., **eigenvectors X are also orthonormal!**

The following theorem due to **Gershgorin** is a helpful tool to localize eigenvalues.
Theorem:
Let the n x n matrix **A** have eigenvalues $\{\lambda_i\}$, $i = 1, 2, \ldots n$. Then, each λ_i lies in the union of the circles

$$\left|z - a_{ii}\right| \le r_i \; ; \quad r_i = \sum_{\substack{j=1 \\ j\neq i}}^{N}\left|a_{ij}\right| \tag{12}$$

area where eigenvalues are located

The knowledge above is very important to estimate the largest and smallest eigenvalues of a system.

Vector Iteration Methods (Power Methods)

When only a few eigenvalues and eigenvectors are needed, then the **power method** is the simplest to use. This method is an iterative technique which gives simultaneously eigenvalues and eigenvectors.

Consider the general eigenvalue value problem

$$\mathbf{K}\boldsymbol{\varphi} = \lambda \mathbf{M}\boldsymbol{\varphi} \tag{13}$$

now, assume that the fundamental frequency λ_1 is distinct from zero and that the eigenvalues are ordered as: $\lambda_n \geq \lambda_{n-1} \ldots > \lambda_2 > \lambda_1$.

An iterative procedure is set up by writing Eq. (13) in the form:

$$\mathbf{K}\mathbf{v}_{(s+1)} = \mathbf{M}\mathbf{u}_{(s)} \qquad s=1,2,\ldots \tag{14}$$

or

$$\mathbf{v}_{(s+1)} = \mathbf{D}\mathbf{u}_{(s)} \tag{15}$$

where

$$\mathbf{D} = \mathbf{K}^{-1}\mathbf{M} = \mathbf{F}\mathbf{M} \tag{16}$$

$\mathbf{D}$ is a <u>dynamical</u> matrix with $\mathbf{F} = \mathbf{K}^{-1}$ as the <u>flexibility</u> matrix (note that $\mathbf{K}$ must be non-singular, that is a zero eigenvalue is not possible).

Note that Eq. (14) can be interpreted as solving the deflection produced by the inertia force $(\mathbf{M}\mathbf{u}) \equiv \mathbf{f_i}$.

Then, the next trial vector $\mathbf{u}_{(s+1)}$ is defined to be:

$$\mathbf{u}_{(s+1)} = \lambda_{(s+1)} \mathbf{v}_{(s+1)} \qquad (17a)$$

where $\lambda_{(s+1)}$ is an appropriate scaling factor such that

$$\left| \mathbf{u}_{(s+1)} \right|_{\max} = 1, \text{ or} \qquad (17.a.1)$$

selected to make

$$\mathbf{u}^T_{(s+1)} \mathbf{M} \mathbf{u}_{(s+1)} \equiv 1 \qquad (17.a.2)$$

or, (b) by using **Rayleigh's quotient** as

$$\lambda_{(s+1)} = \frac{\mathbf{v}^T_{(s+1)} \mathbf{K} \mathbf{v}_{(s+1)}}{\mathbf{v}^T_{(s+1)} \mathbf{M} \mathbf{v}_{(s+1)}} \equiv \frac{\mathbf{v}^T_{(s+1)} \mathbf{M} \mathbf{u}_{(s)}}{\mathbf{v}^T_{(s+1)} \mathbf{M} \mathbf{v}_{(s+1)}} \qquad (17.b.1)$$

and,

$$\mathbf{u}_{(s+1)} \equiv \frac{\mathbf{v}_{(s+1)}}{\left[\mathbf{v}^T_{(s+1)} \mathbf{M} \mathbf{v}_{(s+1)} \right]^{1/2}} \qquad (17.b.2)$$

The procedure is carried out for $s = 0, 1, 2,...$ until P iterations when:
$\nu_{(p+1)} \cong \nu_{(p)}$ and $\lambda_{(p+1)} \cong \lambda_{(p)}$ within a close tolerance.

Now, let's prove that the method described above produces convergence to the first fundamental mode of the system. Recall that any vector in $\mathbf{R}^n$ can be expanded in terms of the eigenvectors. Thus, write an initial guess vector as:

$$\mathbf{u}_{(0)} = \sum_{i=1}^{n} C_i \, \boldsymbol{\varphi}_{(i)} \qquad (18)$$

From Eq. (13): $\left(\dfrac{1}{\lambda_i}\right)\boldsymbol{\varphi}_{(i)} \equiv \mathbf{K}^{-1}\mathbf{M}\boldsymbol{\varphi}_{(i)} = \mathbf{D}\boldsymbol{\varphi}_{(i)}$, with $\mathbf{D} = \mathbf{K}^{-1}\mathbf{M}$

Then:

$$\mathbf{D}\mathbf{u}_{(0)} \equiv \sum_{i=1}^{n} C_i\, \mathbf{D}\boldsymbol{\varphi}_{(i)} \equiv \sum_{i=1}^{n} \frac{C_i}{\lambda_i}\, \boldsymbol{\varphi}_{(i)} \qquad (19)$$

Assume that $\lambda_1 < \lambda_2 < \ldots$. Each iteration cycle use Eq. (15) $\mathbf{u}_{(s+1)} = \mathbf{M}\mathbf{u}_{(s+1)}$. Note that the **scaling** step is omitted without loss of generality.

After s applications of Eq. (15),

$$\mathbf{u}_1 = \mathbf{D}\mathbf{u}_0$$

$$\mathbf{u}_2 = \mathbf{D}\mathbf{u}_1 = \mathbf{D}\mathbf{D}\mathbf{u}_0 = \mathbf{D}^2\mathbf{u}_0$$

$$\mathbf{u}_{(s)} = \mathbf{D}^s\mathbf{u}_{(0)} \equiv \sum_{i=1}^{n} C_i \frac{1}{\lambda_i^s}\, \boldsymbol{\varphi}_{(i)} \qquad (20a)$$

$$\mathbf{u}_{(s)} \equiv \left(\frac{1}{\lambda_1}\right)^s \sum_{i=1}^{n} C_i \left(\frac{\lambda_1}{\lambda_i}\right)^s \boldsymbol{\varphi}_{(i)} \qquad (20b)$$

and since $\lambda_1 < \lambda_i$, $i = 2,3....n$; then

$$\left(\frac{\lambda_1}{\lambda_i}\right)^s \to 0 \ , \ i = 2,.... \text{ and } s \to \infty$$

so Eq. (20b) becomes

$$\mathbf{u}_{(s)} \equiv \left(\frac{1}{\lambda_1}\right)^s \left[C_1\boldsymbol{\varphi}_1 + C_2\left(\frac{\lambda_1}{\lambda_2}\right)^2 \boldsymbol{\varphi}_2 + ...C_n\left(\frac{\lambda_1}{\lambda_n}\right)^s \boldsymbol{\varphi}_n \right]$$

Then, as $s \to \infty$

$$\mathbf{u}_{(s)} = \mathbf{D}^s \mathbf{u}_{(0)} \to \left(\frac{1}{\lambda_1}\right)^S C_1 \boldsymbol{\varphi}_1 \qquad (21)$$

So upon convergence, the method returns a vector proportional to $\boldsymbol{\varphi}_{(1)}$. It is also obvious that the convergence rate depends on the ratio$\left(\dfrac{\lambda_1}{\lambda_i}\right)$ and in particular$\left(\dfrac{\lambda_1}{\lambda_2}\right)$, and on the C_i's that make up the starting (guess) vector.

The method can also be used to calculate the other natural frequencies ω_2 ,ω_3 , etc., provided that at each time, previous eigenvectors $\boldsymbol{\varphi}_2, \boldsymbol{\varphi}_3$, etc., are removed from the iteration procedure. This method is called **MATRIX DEFLATION** or **GRAMSCHMIDT ORTHOGONALIZATION PROCEDURE**, and it is explained as follows:

To determine the second natural frequency, ω_2 ($> \omega_1$), let the trial (guess) vector be determined as:

$$\hat{\mathbf{u}}_{(S)} = \mathbf{u}_{(S)} - \alpha_1 \boldsymbol{\varphi}_1 \qquad (22)$$

where the coefficient α_1 is chosen so that $\hat{\mathbf{u}}$ is orthogonal to $\boldsymbol{\varphi}_{(1)}$, i.e.,

$$\boldsymbol{\varphi}_1^T \mathbf{M} \hat{\mathbf{u}}_{(S)} = 0 \qquad (23)$$

$$\boldsymbol{\varphi}_1^T \mathbf{M} \hat{\mathbf{u}}_{(S)} = \boldsymbol{\varphi}_1^T \mathbf{M} \mathbf{u}_{(S)} - \alpha_1 \boldsymbol{\varphi}_1^T \mathbf{M} \boldsymbol{\varphi}_1 \equiv 0 \quad \text{then} \quad \alpha_1 \equiv \frac{\boldsymbol{\varphi}_1^T \mathbf{M} \mathbf{u}_{(S)}}{\boldsymbol{\varphi}_1^T \mathbf{M} \boldsymbol{\varphi}_1} \qquad (24)$$

Or,

$$\hat{\mathbf{u}}_{(S)} = \mathbf{S_1}\mathbf{u}_{(S)} \equiv \mathbf{u}_{(S)} - \frac{\boldsymbol{\varphi}_1\boldsymbol{\varphi}_1^T\mathbf{M}}{\boldsymbol{\varphi}_1^T\mathbf{M}\boldsymbol{\varphi}_1}\mathbf{u}_{(S)} \tag{25}$$

where

$$\mathbf{S_1} = \mathbf{I} - \frac{\boldsymbol{\varphi}_1\boldsymbol{\varphi}_1^T\mathbf{M}}{\boldsymbol{\varphi}_1^T\mathbf{M}\boldsymbol{\varphi}_1} \tag{26}$$

To produce convergence to $\boldsymbol{\varphi}_{(2)}$, use the fundamental relation:

$$\mathbf{v}_{(s+1)} = \mathbf{D}\hat{\mathbf{u}}_{(s)} = \mathbf{D}\mathbf{S_1}\mathbf{u}_{(s)} = \mathbf{D_2}\mathbf{u}_{(s)} \tag{27}$$

where:

$$\mathbf{D_2} = \mathbf{D}\mathbf{S_1} = \mathbf{D} - \frac{\mathbf{D}\boldsymbol{\varphi}_1\boldsymbol{\varphi}_1^T\mathbf{M}}{\boldsymbol{\varphi}_1^T\mathbf{M}\boldsymbol{\varphi}_1} \tag{28}$$

with $\mathbf{D} = \mathbf{K}^{-1}\mathbf{M} = \mathbf{F}\mathbf{M}$

and from the fundamental relation: $\mathbf{D}\boldsymbol{\varphi}_i = \dfrac{1}{\lambda_i}\boldsymbol{\varphi}_i$

then,

$$\mathbf{D_2} = \mathbf{D} - \frac{1}{\lambda_1}\frac{\boldsymbol{\varphi}_1\boldsymbol{\varphi}_1^T\mathbf{M}}{\boldsymbol{\varphi}_1^T\mathbf{M}\boldsymbol{\varphi}_1} \tag{29}$$

A procedure analogous to Eqs. (22) to (29) can be used to define $\mathbf{D_3}$, $\mathbf{D_4}$, etc. for removing higher modes. This procedure is called **MATRIX DEFLATION.**

Rather lengthy algebraic and matrix manipulation (left as an important exercise for the reader) lead to expressions of the form:

$$\mathbf{D}_3 = \mathbf{D} - \frac{1}{\lambda_1}\frac{\boldsymbol{\varphi}_1\boldsymbol{\varphi}_1^T\mathbf{M}}{\boldsymbol{\varphi}_1^T\mathbf{M}\boldsymbol{\varphi}_1} - \frac{1}{\lambda_2}\frac{\boldsymbol{\varphi}_2\boldsymbol{\varphi}_2^T\mathbf{M}}{\boldsymbol{\varphi}_2^T\mathbf{M}\boldsymbol{\varphi}_2} = \mathbf{D}_2 - \frac{1}{\lambda_2}\frac{\boldsymbol{\varphi}_2\boldsymbol{\varphi}_2^T\mathbf{M}}{\boldsymbol{\varphi}_2^T\mathbf{M}\boldsymbol{\varphi}_2}$$

….. etc.

$$\mathbf{D}_i = \mathbf{D} - \sum_{j=1}^{i-1}\frac{1}{\lambda_j}\frac{\boldsymbol{\varphi}_j\boldsymbol{\varphi}_j^T\mathbf{M}}{\boldsymbol{\varphi}_j^T\mathbf{M}\boldsymbol{\varphi}_j} \qquad (30)$$

$i \equiv 2,3,.....n.$

Note that the products ($\underset{n\times 1}{\boldsymbol{\varphi}}\ \underset{1\times n}{\boldsymbol{\varphi}^T}\ \underset{n\times n}{\mathbf{M}}$) give a matrix of n x n elements.

The matrix deflation technique is used by more advanced methods (HQRI, for example) to determine eigenvectors once a sequence $\{\boldsymbol{\varphi}_i\}_{i=1}^{j}$ has been obtained.

Exercises (homework):

1. Determine an iterative scheme to find the largest eigenvalue and corresponding eigenvector for the problem

$$\mathbf{K}\boldsymbol{\varphi} = \lambda\mathbf{M}\boldsymbol{\varphi}$$

based on the power iteration scheme. Show that the procedure developed indeed converges to the largest eigenvalue. Hint: Assume $\mathbf{M}^{-1}$ exists.

2. The method of **INVERSE Iteration with SPECTRUM SHIFT** allows to determine the eigenvalue closest to a shift number $\mu \in \mathbb{R}_e$. The method is based on the Eq.:

$$\left[\left(\mathbf{K} - \mu\mathbf{M}\right) - \left(\lambda_i - \mu\right)\mathbf{M}\right]\boldsymbol{\varphi}_1 \equiv 0$$

Show that the scheme: $\mathbf{v}_{(s+1)} = \mathbf{D}\mathbf{u}_{(s+1)}$

where $\mathbf{D} = \hat{\mathbf{K}}^{-1}\mathbf{M}$; $\hat{\mathbf{K}} = \mathbf{K} - \mu\mathbf{M}$ converges to $\hat{\lambda}_i \equiv \lambda_i - \mu$.

Methods Based on Similarity Transformations
(From Numerical Methods by Dahlquist and Bjork, Prentice Hall, pp. 208-*ff.*)

The fundamental eigenvalue problem

$$\left[\mathbf{K} - \lambda_i\mathbf{M}\right]\boldsymbol{\varphi}_{(i)} = 0 \qquad (3)\to(31)$$

is recast in the general form:
$$\left[\mathbf{A} - \lambda_i\mathbf{I}\right]\mathbf{X} = 0 \qquad (32)$$

where $\mathbf{A} = \mathbf{A}^{\mathbf{T}} = \mathbf{L}^{-1}\mathbf{K}\mathbf{L}^{-\mathbf{T}}$ (symmetric)

$\qquad \mathbf{M} = \mathbf{L}\mathbf{L}^{\mathbf{T}}$ $\qquad$ (Choleski decomposition of $\mathbf{M}$) (33)

and $\mathbf{X} = \mathbf{L}^{-T}\boldsymbol{\varphi}$ $\qquad$ (Eigenvector relationship).

Note that eigenvalues of a real symmetric matrix are always real and if $\mathbf{A}$ is positive-definite, then all eigenvalues are greater than zero, i.e. $\lambda_i > 0$.

Eqs. (31) and (32) are the same <u>problem</u>: i.e., they have the same eigenvalues and with eigenvectors related by $\mathbf{X} = \mathbf{L}^{-T}\boldsymbol{\varphi}$.

If $\mathbf{P}$ is any non-singular matrix, then the matrices $\mathbf{A}$ and $\mathbf{PAP}^{-1}$ have the same eigenvalues. Indeed, they have the same characteristic equation, since $\mathbf{PP}^{-1}=\mathbf{I}$, and

$$\det\left(\lambda\mathbf{I} - \mathbf{PAP}^{-1}\right) = \det\left(\lambda\mathbf{PP}^{-1} - \mathbf{PAP}^{-1}\right)\det(\mathbf{P})\det\left(\lambda\mathbf{I} - \mathbf{A}\right)\det\left(\mathbf{P}^{-1}\right) \quad (34)$$
$$\equiv \det\left(\lambda\mathbf{I} - \mathbf{A}\right)$$

since $\qquad \det(\mathbf{P})\det(\mathbf{P^{-1}}) \equiv 1 \qquad$ from $\quad \mathbf{PP^{-1}} = \mathbf{I}$

Also since $\quad \mathbf{Ax} = \lambda\mathbf{x}$, it implies that

$$\mathbf{PAx} = \lambda\mathbf{Px} \text{ , and with } \quad \mathbf{PP^{-1}} = \mathbf{I} = \mathbf{P^{-1}P}$$

$$(\mathbf{PAP^{-1}})(\mathbf{Px}) = \lambda(\mathbf{Px}) \tag{35}$$

where it follows that if:

$\quad$ $\mathbf{x}$ is an eigenvector of $\mathbf{A}$, then $\mathbf{Px}$ is an eigenvector of $(\mathbf{PAP^{-1}})$.

$\quad$ The matrices $\mathbf{A}$ and $(\mathbf{PAP^{-1}})$ are said to be **SIMILAR** and the transformation $(\mathbf{PAP^{-1}})$ of $\mathbf{A}$ is called a **similarity transformation**. If the matrix $\mathbf{P}$ is orthogonal, $\mathbf{P^{-1}} = \mathbf{P^T}$, then the condition of the eigenproblem is not affected.

Many methods for solving the eigenvalue problems are based on a **sequence of SIMILARITY TRANSFORMATIONS with ORTHOGONAL MATRICES.** A sequence of matrices $\mathbf{A_0}$, $\mathbf{A_1}$, $\mathbf{A_2}$, . . . is formed by:

$$\mathbf{A_k} = \mathbf{Q_k^T A_{k-1} Q_k} \; ; \; \mathbf{Q_k^T Q_k} \equiv \mathbf{I} \; ; \; k = 1, 2, ... \tag{36}$$

$\quad$ The matrix $\mathbf{A_k}$ is similar to $\mathbf{A}$ and the corresponding eigenvectors are related by

$$\mathbf{X} = \mathbf{Q_1 Q_2 ...... Q_k X_k} \tag{37}$$

(The ultimate goal is to have $\mathbf{A_k}$ in diagonal form after a finite number k of transformations; then, the eigenvalues are just the elements in the diagonal of $\mathbf{A}_k$).

Note that the sequence of transformations, Eq. (36), preserves symmetry since

$$A_k^T = \left(Q_k^T A_{k-1} Q_k\right)^T \equiv Q_k^T \left(Q_k^T A_{k-1}\right)^T$$

$$A_k^T \equiv Q_k^T A_{k-1}^T Q_k \qquad ; \; k = 1, 2, \ldots \qquad (38)$$

and if $A_{k-1} = A_{k-1}^T$ so does $A_k^T = A_k$.

The orthogonal transformations used in practice are of two different types: **PLANE ROTATIONS and/or REFLECTIONS**.

A plane rotation in the (p, q) plane is defined by the matrix $R_{pq(\phi)}$; $|\phi| \le \pi$, which is equal to the unit matrix except for the elements

$$r_{pp} = r_{qq} = \cos\phi , \quad r_{pq} = -r_{qp} = \sin\phi \qquad (39)$$

note that $\mathbf{R}^T{}_{pq} = \mathbf{R}_{pq(-\phi)}$

if the product $\mathbf{A'} = \mathbf{A} \ \mathbf{R}_{pq(\phi)}$ is formed, then only the elements in columns p and q will be changed so then

$$a_{ij} = a_{ij} \ \text{ if } \ j \neq p \ \text{ and } j \neq q, \text{ and}$$
$$a'_{ip} = a_{ip} \cos\phi - a_{iq} \sin\phi \qquad (40)$$
$$a'_{iq} = a_{ip} \sin\phi + a_{iq} \cos\phi$$

In the same way, when completing the similarity transformation by forming
$\mathbf{A''} = \mathbf{R}_{pq(-\phi)} \mathbf{A'}$, then only rows p and q will change.

The ϕ values are obtained so as to make the off-diagonal values equal to zero. This is the basis for the so-called **Jacobi method**. It is easily shown that in order to do this, the angle φ needs to satisfy

$$\text{cotangent } \phi \ = \frac{a_{pp} - a_{qq}}{2a_{pq}} \qquad (41)$$

This reduces the element a''_{pq} equal to zero. When a_{pq} is the largest nondiagonal element ϕ is usually chosen in the range $|\phi| \leq \pi/4$.

The convergence of **Jacobi's method** is quadratic, and taking symmetry into account, the total number of operations for computing all the eigenvalues of a **symmetric** matrix $\mathbf{A}$ are of the order
$$5\tfrac{1}{2}n(n-1)4n \simeq 10n^3 \qquad (42)$$

If the eigenvectors are also desired, then compute

$$\mathbf{X}_k = \mathbf{Q}_1\mathbf{Q}_1....\mathbf{Q}_k = \mathbf{X}_{k-1}\mathbf{Q}_k \ ; \qquad k = 1, 2,... \qquad (43)$$

where $\mathbf{X}_k = (x_1\, x_2 \ldots x_n)_k^T$ will approach to the eigen-matrix $\mathbf{X}$ even when $\mathbf{A}$ has multiple or closely spaced eigenvalues. Since only 2 columns in $\mathbf{X}_k$ will change in each step, then only $4n$ operations per step are required. This procedure will exactly double the operation count for the Jacobi method.

In practice it is not possible to reduce a given matrix to diagonal form in a <u>FINITE</u> sequence number of similarity transformations. Many methods therefore start with an initial transformation of the matrix $\mathbf{A}$ to form other compact forms, which can be attained in a finite number of steps. For symmetric matrices, a suitable form is **symmetric tridiagonal**, while for non-symmetric matrices <u>is **almost triangular**</u> or of **Hessemberg form**.

Once a compact form is obtained, any simplified numerical method could be used to determine the eigenvalues. In particular **Gerschgorin theorem** is extremely helpful to determine starting guess values for the calculations.

If the final form is tridiagonal and symmetric then one can perform a Choleski decomposition such as:
$$\mathbf{A}_k = \mathbf{L}\mathbf{L}^{\mathbf{T}} \tag{44}$$

where the eigenvalues are obtained from the diagonal elements of $\mathbf{L}$.
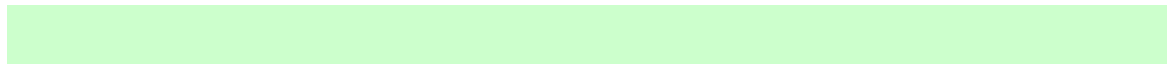
Householder Method: (from Ref. [1])

This algorithm, based on reflection transformations, reduces an n x n symmetric matrix **A** to tridiagonal form by n-2 orthogonal transformations. Each transformation reduces to zero the required part of a whole column and whole corresponding row. The basic ingredient is a **Householder matrix P** which has the form:

$$\mathbf{P} = \mathbf{I} - 2 \underset{n \times 1}{\boldsymbol{\omega}} \times \underset{1 \times n}{\boldsymbol{\omega}^T} \tag{45}$$

where **ω** is a real vector such that $|\boldsymbol{\omega}|^2 = \boldsymbol{\omega}^T \boldsymbol{\omega} = 1$.

The matrix **P** is orthogonal and symmetric since

$$\mathbf{P}^\mathbf{T}\mathbf{P} = \mathbf{P}^2 = \left(\mathbf{I} - 2\boldsymbol{\omega}\boldsymbol{\omega}^\mathbf{T}\right)\left(\mathbf{I} - 2\boldsymbol{\omega}\boldsymbol{\omega}^\mathbf{T}\right)$$
$$\equiv \mathbf{I} - 4\boldsymbol{\omega}\boldsymbol{\omega}^\mathbf{T} + 4\boldsymbol{\omega}\boldsymbol{\omega}^\mathbf{T}\boldsymbol{\omega}\boldsymbol{\omega}^\mathbf{T} \equiv \mathbf{I} \tag{46}$$

Hence:

$$\mathbf{P}^{-1} = \mathbf{P}^\mathbf{T} = \mathbf{P} \tag{47}$$

Rewrite **P** as:

$$\mathbf{P} = \mathbf{I} - \frac{\mathbf{u}\mathbf{u}^\mathbf{T}}{H} \tag{48}$$

where scalar H is
$$H = \frac{1}{2}|\mathbf{u}|^2 \equiv \frac{1}{2}\left(u_1^2 + u_2^2 + \ldots\ldots + u_n^2\right) \tag{49}$$

and **u** can now be any vector.

Suppose **x** is the vector composed of the first column of **A** , i.e.
$$\mathbf{x}^\mathbf{T} = \left\{a_{1,1} \ a_{1,1} \ \ldots\ldots \ a_{n,1}\right\} \tag{50}$$

with $|\mathbf{x}| \equiv \sqrt{x_1^2 + x_2^2 + ...x_n^2}$ as its **L₂** norm

and choose

$$\mathbf{u} = \mathbf{x} \mp |\mathbf{x}| \mathbf{e}_1 \qquad\qquad (51)$$

where $e_1^T = \{1, 0, 0, ...0_n\}$.

The choice of signs in Eq. (51) will be discussed later. Then:

$$\mathbf{Px} = \left(\mathbf{I} - \frac{\mathbf{uu^T}}{H} \right) \mathbf{x} = \mathbf{x} - \frac{\mathbf{u}(\mathbf{u^T x})}{H}$$

$$\equiv \mathbf{x} - \frac{\mathbf{u}\left(\left[\mathbf{x^T} \mp |\mathbf{x}| e_1^T \right] \mathbf{x} \right)}{H} \quad ; \quad \begin{array}{l} \mathbf{x^T x} = |\mathbf{x}|^2 \\[2mm] \mathbf{e_1^T x} = \mathbf{x_1} = a_{1,1} \end{array}$$

$$\equiv \mathbf{x} - \frac{\mathbf{u}\left(|\mathbf{x}|^2 \mp |\mathbf{x}| \mathbf{x_1} \right)}{\frac{1}{2}|\mathbf{u}|^2} \qquad\qquad H = \frac{1}{2}|\mathbf{u}|^2$$

Recall that $\mathbf{u} = \mathbf{x} \mp |\mathbf{x}| \mathbf{e}_1 \equiv \{\mathbf{x_1} \mp |\mathbf{x}|, \mathbf{e}\, \mathbf{x_2}, ...\mathbf{x}_n \}$.

$$|\mathbf{u}|^2 = \left[\mathbf{x_1} \mp |\mathbf{x}| \right]^2 + \mathbf{x_2^2} + ...\mathbf{x}_n^2$$

$$\equiv \mathbf{x_1^2} \mp 2\mathbf{x_1}|\mathbf{x}| \pm |\mathbf{x}|^2 + \mathbf{x_2^2} + ...\mathbf{x}_n^2$$

$$\equiv \left(\mathbf{x_1^2} + \mathbf{x_2^2} + ...\mathbf{x}_N^2 \right) + |\mathbf{x}|^2 \mp 2\mathbf{x_1}|\mathbf{x}|$$

$$\equiv |\mathbf{x}|^2 + |\mathbf{x}|^2 \mp 2\mathbf{x_1}|\mathbf{x}|$$

$$\equiv 2\left[|\mathbf{x}|^2 \mp \mathbf{x_1}|\mathbf{x}| \right]$$

Then

$$\mathbf{P\,x = x - u} \frac{\left(\left|\mathbf{x}\right|^2 \mp \mathbf{x_1}\left|\mathbf{x}\right|\right)}{\left(\left|\mathbf{x}\right|^2 \mp \mathbf{x_1}\left|\mathbf{x}\right|\right)} \equiv \mathbf{x - u} = \pm\left|\mathbf{x}\right|\mathbf{e}_1$$

$$\equiv \pm\left|\mathbf{x}\right| \begin{Bmatrix} 1 \\ 0 \\ 0 \\ . \\ . \\ . \\ 0 \end{Bmatrix} \tag{52}$$

which shows that the Householder matrix **P** acts on a given vector **x** to zero all its elements except the first one.

To reduce $\mathbf{A} = \mathbf{A^T}$ to tridiagonal form, select vector **x** for the first Householder matrix to be the lower $(n-1)$ elements of the first column. Then, the lower $(n-2)$ elements will be <u>zeroed</u>:

$$\mathbf{P_1 A} = \begin{bmatrix} 1 & 0 & 0 & . & . & 0 \\ 0 & & & & & \\ 0 & & & & & \\ . & & & & & \\ . & & (n-1)P_1 & & & \\ 0 & & & & & \end{bmatrix} \begin{bmatrix} a_{11} & a_{11} & a_{11} & . & . & a_{1n} \\ a_{21} & & & & & \\ a_{31} & & & & & \\ . & & & & & \\ . & & & irrelevant & & \\ a_{n1} & & & & & \end{bmatrix} \equiv \begin{bmatrix} a_{11} & a_{12} & . & . & . & a_{1n} \\ k & & & & & \\ 0 & & & & & \\ 0 & & & & & \\ 0 & & & irrelevant & & \\ 0 & & & & & \end{bmatrix} \tag{53}$$

where the quantity $k \equiv \pm\left|a_{21}.......a_{2N}\right|$ and the complete orthogonal transformation is:

$$
\mathbf{A^1} = \begin{bmatrix} a_{11} & k & 0 & & 0 & & 0 & 0 \\ k & & & & & & & \\ 0 & & & & & & & \\ 0 & & & \mathit{irrelevant} & & & & \\ 0 & & & & & & & \\ 0 & & & & & & & \end{bmatrix} \quad ; \qquad \mathbf{P^T = P} \qquad (54)
$$

Now choose the vector **x** for the second Householder matrix to be the bottom (n-2) terms of the second column, and from it construct

$$
\mathbf{P_2} = \begin{bmatrix} 1 & 0 & 0 & . & & . & & 0 \\ 0 & 1 & 0 & . & & . & & 0 \\ 0 & 0 & & & & & & \\ . & . & & & & & & \\ . & . & & & (n-2)P_2 & & & \\ 0 & 0 & & & & & & \end{bmatrix} \qquad (55)
$$

The identity block on the upper left corner insures that the tridiagonalization vector will not be spoiled in this case, while the (n - 2) Householder matrix $^{(n\,-\,2)}\mathbf{P_2}$ creates one additional column of the tridiagonal output. Clearly, a sequence of (n-2) such transformations will reduce **A** to tridiagonal form.

In practice, instead of performing the matrix multiplication **PAP**, compute vector

$$
\mathbf{p} = \frac{\mathbf{Au}}{H} \qquad (56)
$$

Then $\mathbf{AP} = \mathbf{A}\left(\mathbf{I} - \dfrac{\mathbf{uu^T}}{H} \right) \equiv \mathbf{A} - \dfrac{\mathbf{pu^T}}{H}$

and

$$\mathbf{A}^1 = \mathbf{PAP} = \left(\mathbf{I} - \frac{\mathbf{uu}^\mathsf{T}}{H}\right)\mathbf{AP} = \mathbf{A} - \mathbf{pu}^\mathsf{T} - \mathbf{up}^\mathsf{T} + 2k\mathbf{uu}^\mathsf{T} \qquad (57)$$

where the scalar

$$k \equiv \frac{\mathbf{u}^\mathsf{T}\mathbf{p}}{2H} \qquad (58)$$

and if

$$\mathbf{q} = \mathbf{p} - k\mathbf{u} \qquad (59)$$

then

$$\mathbf{A}^1 = \mathbf{A} - \mathbf{qu}^\mathsf{T} - \mathbf{uq}^\mathsf{T} \qquad (60)$$

The reduction to tridiagonal symmetry takes $2n^3/3$ operations if original matrix $\mathbf{A}$ is symmetric.

The QR Algorithm

The basic idea behind the **QR** algorithm is that **any** non-singular real matrix can be decomposed in the form

$$\mathbf{A} = \mathbf{QR} \qquad \overset{where}{\longrightarrow} \qquad \mathbf{R} = \mathbf{Q}^\mathsf{T}\mathbf{A} \qquad (61)$$

and $\mathbf{Q}$ is orthogonal, i.e., $\mathbf{Q}^\mathsf{T}\mathbf{Q} = \mathbf{I}$ and $\mathbf{R}$ is upper-triangular. For a general matrix, the decomposition is constructed by applying the Householder transformation to zero-the successive columns of $\mathbf{A}$ below its diagonal.

Now consider

$$\mathbf{A}' = \mathbf{QR}$$

$$\mathbf{A}' = \mathbf{Q}^\mathsf{T}\mathbf{AQ}$$

thus $\mathbf{A}'$ is an orthogonal transformation of $\mathbf{A}$.

Exercise: Show that the decomposition **A** = **QR** for **A** symmetric and positive definite IS UNIQUE !

References

1. Numerical Recipes, The Art of Scientific Computing, by W. Press, B. Flannery, S. Teukolsky, W. Vetterling, Cambridge Press, 1989.

2. Wilkinson, J.H. and Reinsch, Linear Algebra, Vol. II, Handbook for Automatic Computation, N.Y., Springer-Verlag, 1971.

3. K. J. Bathe, and E. L. Wilson, Numerical Methods in Finite Element Analysis, Englewood Cliffs, N.J., 1976.

4. Dahlquist and Bjork, Numerical Methods, Prentice Hall, pp. 208-*ff*