9.2 Eigenanalysis II

Discrete Dynamical Systems

The matrix equation

$$\mathbf{y} = \frac{1}{10}\begin{pmatrix} 5 & 4 & 0 \\ 3 & 5 & 3 \\ 2 & 1 & 7 \end{pmatrix}\mathbf{x} \tag{1}$$

predicts the state $\mathbf{y}$ of a system initially in state $\mathbf{x}$ after some fixed elapsed time. The 3×3 matrix A in (1) represents the **dynamics** which changes the state $\mathbf{x}$ into state $\mathbf{y}$. Accordingly, an equation $\mathbf{y} = A\mathbf{x}$ is called a **discrete dynamical system** and A is called a **transition matrix**.

The eigenpairs of A in (1) are shown in *details* page 482 to be $(1, \mathbf{v}_1)$, $(1/2, \mathbf{v}_2)$, $(1/5, \mathbf{v}_3)$ where the eigenvectors are given by

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 5/4 \\ 13/12 \end{pmatrix}, \quad \mathbf{v}_2 = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \quad \mathbf{v}_3 = \begin{pmatrix} -4 \\ 3 \\ 1 \end{pmatrix}. \tag{2}$$

Market Shares. A typical application of discrete dynamical systems is telephone long distance company market shares x_1, x_2, x_3, which are fractions of the total market for long distance service. If three companies provide all the services, then their market fractions add to one: $x_1 + x_2 + x_3 = 1$. The equation $\mathbf{y} = A\mathbf{x}$ gives the market shares of the three companies after a fixed time period, say one year. Then market shares after one, two and three years are given by the **iterates**

$$\begin{aligned} \mathbf{y}_1 &= A\mathbf{x}, \\ \mathbf{y}_2 &= A^2\mathbf{x}, \\ \mathbf{y}_3 &= A^3\mathbf{x}. \end{aligned}$$

Fourier's eigenanalysis model gives succinct and useful formulas for the iterates: if $\mathbf{x} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + a_3\mathbf{v}_3$, then

$$\begin{aligned} \mathbf{y}_1 &= A\mathbf{x} &&= a_1\lambda_1\mathbf{v}_1 + a_2\lambda_2\mathbf{v}_2 + a_3\lambda_3\mathbf{v}_3, \\ \mathbf{y}_2 &= A^2\mathbf{x} &&= a_1\lambda_1^2\mathbf{v}_1 + a_2\lambda_2^2\mathbf{v}_2 + a_3\lambda_3^2\mathbf{v}_3, \\ \mathbf{y}_3 &= A^3\mathbf{x} &&= a_1\lambda_1^3\mathbf{v}_1 + a_2\lambda_2^3\mathbf{v}_2 + a_3\lambda_3^3\mathbf{v}_3. \end{aligned}$$

The advantage of Fourier's model is that an iterate A^n is computed directly, without computing the powers before it. Because $\lambda_1 = 1$ and $\lim_{n\to\infty} |\lambda_2|^n = \lim_{n\to\infty} |\lambda_3|^n = 0$, then for large n

$$\mathbf{y}_n \approx a_1(1)\mathbf{v}_1 + a_2(0)\mathbf{v}_2 + a_3(0)\mathbf{v}_3 = \begin{pmatrix} a_1 \\ 5a_1/4 \\ 13a_1/12 \end{pmatrix}.$$

The numbers a_1, a_2, a_3 are related to x_1, x_2, x_3 by the equations $a_1 - a_2 - 4a_3 = x_1$, $5a_1/4 + 3a_3 = x_2$, $13a_1/12 + a_2 + a_3 = x_3$. Due to $x_1 + x_2 + x_3 = 1$, the value of a_1 is known, $a_1 = 3/10$. The three market shares after a long time period are therefore predicted to be 3/10, 3/8, 39/120. The reader should verify the identity $\frac{3}{10} + \frac{3}{8} + \frac{39}{120} = 1$.

Stochastic Matrices. The special matrix A in (1) is a **stochastic matrix**, defined by the properties

$$\sum_{i=1}^{n} a_{ij} = 1, \quad a_{kj} \geq 0, \quad k, j = 1, \ldots, n.$$

The definition is memorized by the phrase *each column sum is one.* Stochastic matrices appear in **Leontief input-output models**, popularized by 1973 Nobel Prize economist Wassily Leontief.

Theorem 9 (Stochastic Matrix Properties)
Let A be a stochastic matrix. Then

(a) If $\mathbf{x}$ is a vector with $x_1 + \cdots + x_n = 1$, then $\mathbf{y} = A\mathbf{x}$ satisfies $y_1 + \cdots + y_n = 1$.

(b) If $\mathbf{v}$ is the sum of the columns of I, then $A^T\mathbf{v} = \mathbf{v}$. Therefore, $(1, \mathbf{v})$ is an eigenpair of A^T.

(c) The characteristic equation $\det(A - \lambda I) = 0$ has a root $\lambda = 1$. All other roots satisfy $|\lambda| < 1$.

Proof of Stochastic Matrix Properties:
(a) $\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_j = \sum_{j=1}^{n} \left(\sum_{i=1}^{n} a_{ij}\right) x_j = \sum_{j=1}^{n} (1) x_j = 1.$

(b) Entry j of $A^T\mathbf{v}$ is given by the sum $\sum_{i=1}^{n} a_{ij} = 1$.

(c) Apply **(b)** and the determinant rule $\det(B^T) = \det(B)$ with $B = A - \lambda I$ to obtain eigenvalue 1. Any other root λ of the characteristic equation has a corresponding eigenvector $\mathbf{x}$ satisfying $(A - \lambda I)\mathbf{x} = \mathbf{0}$. Let index j be selected such that $M = |x_j| > 0$ has largest magnitude. Then $\sum_{i \neq j} a_{ij} x_j + (a_{jj} - \lambda) x_j = 0$ implies $\lambda = \sum_{i=1}^{n} a_{ij} \dfrac{x_j}{M}$. Because $\sum_{i=1}^{n} a_{ij} = 1$, λ is a convex combination of n complex numbers $\{x_j/M\}_{j=1}^{n}$. These complex numbers are located in the unit disk, a convex set, therefore λ is located in the unit disk. By induction on n, motivated by the geometry for $n = 2$, it is argued that $|\lambda| = 1$ cannot happen for λ an eigenvalue different from 1 (details left to the reader). Therefore, $|\lambda| < 1$.

Details for the eigenpairs of (1): To be computed are the eigenvalues and eigenvectors for the 3×3 matrix

$$A = \frac{1}{10} \begin{pmatrix} 5 & 4 & 0 \\ 3 & 5 & 3 \\ 2 & 1 & 7 \end{pmatrix}.$$

Eigenvalues. The roots $\lambda = 1, 1/2, 1/5$ of the characteristic equation $\det(A - \lambda I) = 0$ are found by these details:

$$\begin{aligned}
0 &= \det(A - \lambda I) \\
&= \begin{vmatrix} .5-\lambda & .4 & 0 \\ .3 & .5-\lambda & .3 \\ .2 & .1 & .7-\lambda \end{vmatrix} \\
&= \frac{1}{10} - \frac{8}{10}\lambda + \frac{17}{10}\lambda^2 - \lambda^3 && \text{Expand by cofactors.} \\
&= -\frac{1}{10}(\lambda-1)(2\lambda-1)(5\lambda-1) && \text{Factor the cubic.}
\end{aligned}$$

The factorization was found by long division of the cubic by $\lambda - 1$, the idea born from the fact that 1 is a root and therefore $\lambda - 1$ is a factor (the Factor Theorem of college algebra). An answer check in `maple`:

```
with(linalg):
A:=(1/10)*matrix([[5,4,0],[3,5,3],[2,1,7]]);
B:=evalm(A-lambda*diag(1,1,1));
eigenvals(A); factor(det(B));
```

Eigenpairs. To each eigenvalue $\lambda = 1, 1/2, 1/5$ corresponds one **rref** calculation, to find the eigenvectors paired to λ. The three eigenvectors are given by (2). The details:

Eigenvalue $\lambda = 1$.

$$\begin{aligned}
A - (1)I &= \begin{pmatrix} .5-1 & .4 & 0 \\ .3 & .5-1 & .3 \\ .2 & .1 & .7-1 \end{pmatrix} \\
&\approx \begin{pmatrix} -5 & 4 & 0 \\ 3 & -5 & 3 \\ 2 & 1 & -3 \end{pmatrix} && \text{Multiply rule, multiplier=10.} \\
&\approx \begin{pmatrix} 0 & 0 & 0 \\ 3 & -5 & 3 \\ 2 & 1 & -3 \end{pmatrix} && \text{Combination rule twice.} \\
&\approx \begin{pmatrix} 0 & 0 & 0 \\ 1 & -6 & 6 \\ 2 & 1 & -3 \end{pmatrix} && \text{Combination rule.} \\
&\approx \begin{pmatrix} 0 & 0 & 0 \\ 1 & -6 & 6 \\ 0 & 13 & -15 \end{pmatrix} && \text{Combination rule.} \\
&\approx \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & -\frac{12}{13} \\ 0 & 1 & -\frac{15}{13} \end{pmatrix} && \text{Multiply rule and combination rule.} \\
&\approx \begin{pmatrix} 1 & 0 & -\frac{12}{13} \\ 0 & 1 & -\frac{15}{13} \\ 0 & 0 & 0 \end{pmatrix} && \text{Swap rule.} \\
&= \mathbf{rref}(A - (1)I)
\end{aligned}$$

An equivalent reduced echelon system is $x - 12z/13 = 0$, $y - 15z/13 = 0$. The free variable assignment is $z = t_1$ and then $x = 12t_1/13$, $y = 15t_1/13$. Let $x = 1$; then $t_1 = 13/12$. An eigenvector is given by $x = 1$, $y = 4/5$, $z = 13/12$.

Eigenvalue $\lambda = 1/2$.

$$
\begin{aligned}
A-(1/2)I &= \begin{pmatrix} .5-.5 & .4 & 0 \\ .3 & .5-.5 & .3 \\ .2 & .1 & .7-.5 \end{pmatrix} \\
&= \begin{pmatrix} 0 & 4 & 0 \\ 3 & 0 & 3 \\ 2 & 1 & 2 \end{pmatrix} && \text{Multiply rule, factor=10.} \\
&\approx \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} && \text{Combination and multiply rules.} \\
&= \mathbf{rref}(A-.5I)
\end{aligned}
$$

An eigenvector is found from the equivalent reduced echelon system $y = 0$, $x+z = 0$ to be $x = -1$, $y = 0$, $z = 1$.

Eigenvalue $\lambda = 1/5$.

$$
\begin{aligned}
A-(1/5)I &= \begin{pmatrix} .5-.2 & .4 & 0 \\ .3 & .5-.2 & .3 \\ .2 & .1 & .7-.2 \end{pmatrix} \\
&\approx \begin{pmatrix} 3 & 4 & 0 \\ 1 & 1 & 1 \\ 2 & 1 & 5 \end{pmatrix} && \text{Multiply rule.} \\
&\approx \begin{pmatrix} 1 & 0 & 4 \\ 0 & 1 & -3 \\ 0 & 0 & 0 \end{pmatrix} && \text{Combination rule.} \\
&= \mathbf{rref}(A-(1/5)I)
\end{aligned}
$$

An eigenvector is found from the equivalent reduced echelon system $x+4z = 0$, $y-3z = 0$ to be $x = -4$, $y = 3$, $z = 1$.

An answer check in `maple`:

```
with(linalg):
A:=(1/10)*matrix([[5,4,0],[3,5,3],[2,1,7]]);
eigenvects(A);
```

Coupled and Uncoupled Systems

The linear system of differential equations

$$
\begin{aligned}
x_1' &= -x_1 - x_3, \\
x_2' &= 4x_1 - x_2 - 3x_3, \\
x_3' &= 2x_1 - 4x_3,
\end{aligned} \tag{3}
$$

is called **coupled**, whereas the linear system of growth-decay equations

$$
\begin{aligned}
y_1' &= -3y_1, \\
y_2' &= -y_2, \\
y_3' &= -2y_3,
\end{aligned} \tag{4}
$$

is called **uncoupled**. The terminology *uncoupled* means that each differential equation in system (4) depends on exactly one variable, e.g., $y_1' = -3y_1$ depends only on variable y_1. In a *coupled* system, one of the differential equations must involve two or more variables.

Matrix characterization. Coupled system (3) and uncoupled system (4) can be written in matrix form, $\mathbf{x}' = A\mathbf{x}$ and $\mathbf{y}' = D\mathbf{y}$, with coefficient matrices

$$A = \begin{pmatrix} -1 & 0 & -1 \\ 4 & -1 & -3 \\ 2 & 0 & -4 \end{pmatrix} \quad \text{and} \quad D = \begin{pmatrix} -3 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -2 \end{pmatrix}.$$

If the coefficient matrix is **diagonal**, then the system is **uncoupled**. If the coefficient matrix is **not diagonal**, then one of the corresponding differential equations involves two or more variables and the system is called **coupled** or **cross-coupled**.

Solving Uncoupled Systems

An uncoupled system consists of independent growth-decay equations of the form $u' = au$. The recipe solution $u = ce^{at}$ then leads to the general solution of the system of equations. For instance, system (4) has general solution

$$\begin{aligned} y_1 &= c_1 e^{-3t}, \\ y_2 &= c_2 e^{-t}, \\ y_3 &= c_3 e^{-2t}, \end{aligned} \tag{5}$$

where c_1, c_2, c_3 are **arbitrary constants**. The number of constants equals the dimension of the diagonal matrix D.

Coordinates and Coordinate Systems

If $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$ are three independent vectors in $\mathcal{R}^3$, then the matrix

$$P = \mathbf{aug}(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$$

is invertible. The columns $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$ of P are called a **coordinate system**. The matrix P is called a **change of coordinates**.

Every vector $\mathbf{v}$ in $\mathcal{R}^3$ can be uniquely expressed as

$$\mathbf{v} = t_1\mathbf{v}_1 + t_2\mathbf{v}_2 + t_3\mathbf{v}_3.$$

The values t_1, t_2, t_3 are called the **coordinates** of $\mathbf{v}$ relative to the basis $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$, or more succinctly, the **coordinates of v relative to** P.

Viewpoint of a Driver

The physical meaning of a coordinate system $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$ can be understood by considering an auto going up a mountain road. Choose orthogonal $\mathbf{v}_1$ and $\mathbf{v}_2$ to give positions in the driver's seat and define $\mathbf{v}_3$ be the seat-back direction. These are **local coordinates** as viewed from the driver's seat. The road map coordinates x, y and the altitude z define the **global coordinates** for the auto's position $\mathbf{p} = x\vec{\imath} + y\vec{\jmath} + z\vec{k}$.

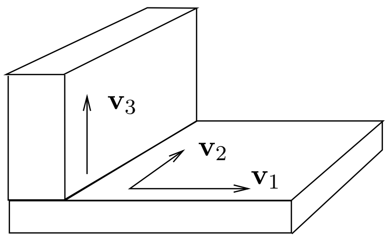


Figure 1. An auto seat.
The vectors $\mathbf{v}_1(t)$, $\mathbf{v}_2(t)$, $\mathbf{v}_3(t)$ form an orthogonal triad which is a local coordinate system from the driver's viewpoint. The orthogonal triad changes continuously in t.

Change of Coordinates

A coordinate change from $\mathbf{y}$ to $\mathbf{x}$ is a linear algebraic equation $\mathbf{x} = P\mathbf{y}$ where the $n \times n$ matrix P is required to be invertible ($\det(P) \neq 0$). To illustrate, an instance of a change of coordinates from $\mathbf{y}$ to $\mathbf{x}$ is given by the linear equations

$$\mathbf{x} = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & -1 \\ 2 & 0 & 1 \end{pmatrix} \mathbf{y} \quad \text{or} \quad \begin{cases} x_1 &= y_1 + y_3, \\ x_2 &= y_1 + y_2 - y_3, \\ x_3 &= 2y_1 + y_3. \end{cases} \tag{6}$$

Constructing Coupled Systems

A general method exists to construct rich examples of coupled systems. The idea is to substitute a change of variables into a given uncoupled system. Consider a diagonal system $\mathbf{y}' = D\mathbf{y}$, like (4), and a change of variables $\mathbf{x} = P\mathbf{y}$, like (6). Differential calculus applies to give

$$\begin{aligned} \mathbf{x}' &= (P\mathbf{y})' \\ &= P\mathbf{y}' \\ &= PD\mathbf{y} \\ &= PDP^{-1}\mathbf{x}. \end{aligned} \tag{7}$$

The matrix $A = PDP^{-1}$ is *not triangular* in general, and therefore the change of variables produces a **cross-coupled** system.

An illustration. To give an example, substitute into uncoupled system (4) the change of variable equations (6). Use equation (7) to obtain

$$\mathbf{x}' = \begin{pmatrix} -1 & 0 & -1 \\ 4 & -1 & -3 \\ 2 & 0 & -4 \end{pmatrix} \mathbf{x} \quad \text{or} \quad \begin{cases} x_1' = -x_1 - x_3, \\ x_2' = 4x_1 - x_2 - 3x_3, \\ x_3' = 2x_1 - 4x_3. \end{cases} \tag{8}$$

This **cross-coupled** system (8) can be solved using relations (6), (5) and $\mathbf{x} = P\mathbf{y}$ to give the general solution

$$(9) \qquad \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & -1 \\ 2 & 0 & 1 \end{pmatrix} \begin{pmatrix} c_1 e^{-3t} \\ c_2 e^{-t} \\ c_3 e^{-2t} \end{pmatrix}.$$

Changing Coupled Systems to Uncoupled

We ask this question, motivated by the above calculations:

> Can every coupled system $\mathbf{x}'(t) = A\mathbf{x}(t)$ be subjected to a change of variables $\mathbf{x} = P\mathbf{y}$ which converts the system into a completely uncoupled system for variable $\mathbf{y}(t)$?

Under certain circumstances, this is true, and it leads to an elegant and especially simple expression for the general solution of the differential system, as in (9):

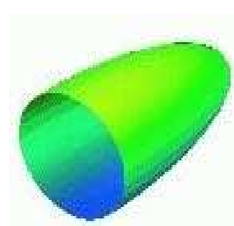
$$\mathbf{x}(t) = P\mathbf{y}(t).$$

The **task of eigenanalysis** is to simultaneously calculate from a cross-coupled system $\mathbf{x}' = A\mathbf{x}$ the change of variables $\mathbf{x} = P\mathbf{y}$ and the diagonal matrix D in the uncoupled system $\mathbf{y}' = D\mathbf{y}$

The **eigenanalysis coordinate system** is the set of n independent vectors extracted from the columns of P. In this coordinate system, the cross-coupled differential system (3) simplifies into a system of uncoupled growth-decay equations (4). Hence the terminology, *the method of simplifying coordinates*.

Eigenanalysis and Footballs

An ellipsoid or *football* is a geometric object described by its **semi-axes** (see Figure 2). In the vector representation, the **semi-axis directions** are unit vectors $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$ and the **semi-axis lengths** are the constants a, b, c. The vectors $a\mathbf{v}_1$, $b\mathbf{v}_2$, $c\mathbf{v}_3$ form an **orthogonal triad**.



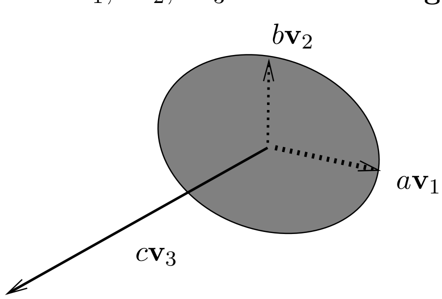


Figure 2. A football.
An ellipsoid is built from orthonormal semi-axis directions $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$ and the semi-axis lengths a, b, c. The semi-axis vectors are $a\mathbf{v}_1$, $b\mathbf{v}_2$, $c\mathbf{v}_3$.

Two vectors $\mathbf{a}$, $\mathbf{b}$ are *orthogonal* if both are nonzero and their dot product $\mathbf{a} \cdot \mathbf{b}$ is zero. Vectors are **orthonormal** if each has unit length and they are pairwise orthogonal. The orthogonal triad is an **invariant** of the ellipsoid's algebraic representations. Algebra does not change the triad: the invariants $a\mathbf{v}_1$, $b\mathbf{v}_2$, $c\mathbf{v}_3$ must somehow be **hidden** in the equations that represent the football.

Algebraic eigenanalysis finds the hidden invariant triad $a\mathbf{v}_1$, $b\mathbf{v}_2$, $c\mathbf{v}_3$ from the ellipsoid's algebraic equations. Suppose, for instance, that the equation of the ellipsoid is supplied as

$$x^2 + 4y^2 + xy + 4z^2 = 16.$$

A symmetric matrix A is constructed in order to write the equation in the form $\mathbf{X}^T A\,\mathbf{X} = 16$, where $\mathbf{X}$ has components x, y, z. The replacement equation is[4]

$$\begin{pmatrix} x & y & z \end{pmatrix} \begin{pmatrix} 1 & 1/2 & 0 \\ 1/2 & 4 & 0 \\ 0 & 0 & 4 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 16. \tag{10}$$

It is the 3×3 symmetric matrix A in (10) that is subjected to algebraic eigenanalysis. The matrix calculation will compute the unit semi-axis directions $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$, called the **hidden vectors** or **eigenvectors**. The semi-axis lengths a, b, c are computed at the same time, by finding the **hidden values**[5] or **eigenvalues** λ_1, λ_2, λ_3, known to satisfy the relations

$$\lambda_1 = \frac{16}{a^2}, \quad \lambda_2 = \frac{16}{b^2}, \quad \lambda_3 = \frac{16}{c^2}.$$

For the illustration, the football dimensions are $a = 2$, $b = 1.98$, $c = 4.17$. Details of the computation are delayed until page 490.

The Ellipse and Eigenanalysis

An ellipse equation in **standard form** is $\lambda_1 x^2 + \lambda_2 y^2 = 1$, where $\lambda_1 = 1/a^2$, $\lambda_2 = 1/b^2$ are expressed in terms of the semi-axis lengths a, b. The expression $\lambda_1 x^2 + \lambda_2 y^2$ is called a **quadratic form**. The study of the ellipse $\lambda_1 x^2 + \lambda_2 y^2 = 1$ is equivalent to the study of the quadratic form equation

$$\mathbf{r}^T D\mathbf{r} = 1, \quad \text{where} \quad \mathbf{r} = \begin{pmatrix} x \\ y \end{pmatrix}, \quad D = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}.$$

[4]The reader should pause here and multiply matrices in order to verify this statement. Halving of the entries corresponding to cross-terms generalizes to any ellipsoid.

[5]The terminology *hidden* arises because neither the semi-axis lengths nor the semi-axis directions are revealed directly by the ellipsoid equation.

Cross-terms. An ellipse may be represented by an equation in a uv-coordinate system having a cross-term uv, e.g., $4u^2+8uv+10v^2=5$. The expression $4u^2+8uv+10v^2$ is again called a quadratic form. Calculus courses provide methods to eliminate the cross-term and represent the equation in standard form, by a **rotation**

$$\begin{pmatrix} u \\ v \end{pmatrix} = R\begin{pmatrix} x \\ y \end{pmatrix}, \quad R = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}.$$

The angle θ in the rotation matrix R represents the rotation of uv-coordinates into standard xy-coordinates.

Eigenanalysis computes angle θ through the columns of R, which are the unit semi-axis directions $\mathbf{v}_1$, $\mathbf{v}_2$ for the ellipse $4u^2+8uv+10v^2=5$. If the quadratic form $4u^2+8uv+10v^2$ is represented as $\mathbf{r}^T A\,\mathbf{r}$, then

$$\mathbf{r} = \begin{pmatrix} u \\ v \end{pmatrix}, \quad A = \begin{pmatrix} 4 & 4 \\ 4 & 10 \end{pmatrix}, \quad R = \frac{1}{\sqrt{5}}\begin{pmatrix} 1 & -2 \\ 2 & 1 \end{pmatrix},$$

$$\lambda_1 = 12, \quad \mathbf{v}_1 = \frac{1}{\sqrt{5}}\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad \lambda_2 = 2, \quad \mathbf{v}_2 = \frac{1}{\sqrt{5}}\begin{pmatrix} -2 \\ 1 \end{pmatrix}.$$

Rotation matrix angle θ. The components of eigenvector $\mathbf{v}_1$ can be used to determine $\theta = -63.4^\circ$:

$$\begin{pmatrix} \cos\theta \\ -\sin\theta \end{pmatrix} = \frac{1}{\sqrt{5}}\begin{pmatrix} 1 \\ 2 \end{pmatrix} \quad \text{or} \quad \left\{ \begin{array}{rcl} \cos\theta & = & \frac{1}{\sqrt{5}}, \\ -\sin\theta & = & \frac{2}{\sqrt{5}}. \end{array} \right.$$

The interpretation of angle θ: rotate the orthonormal basis $\mathbf{v}_1$, $\mathbf{v}_2$ by angle $\theta = -63.4^\circ$ in order to obtain the standard unit basis vectors $\mathbf{i}$, $\mathbf{j}$. Most calculus texts discuss only the inverse rotation, where x, y are given in terms of u, v. In these references, θ is the negative of the value given here, due to a different geometric viewpoint.[6]

Semi-axis lengths. The lengths $a \approx 1.55$, $b \approx 0.63$ for the ellipse $4u^2+8uv+10v^2=5$ are computed from the eigenvalues $\lambda_1 = 12$, $\lambda_2 = 2$ of matrix A by the equations

$$\frac{\lambda_1}{5} = \frac{1}{a^2}, \quad \frac{\lambda_2}{5} = \frac{1}{b^2}.$$

Geometry. The ellipse $4u^2+8uv+10v^2=5$ is completely determined by the orthogonal semi-axis vectors $a\mathbf{v}_1$, $b\mathbf{v}_2$. The rotation R is a rigid motion which maps these vectors into $a\vec{\imath}$, $b\vec{\jmath}$, where $\vec{\imath}$ and $\vec{\jmath}$ are the standard unit vectors in the plane.

The θ-rotation R maps $4u^2+8uv+10v^2=5$ into the xy-equation $\lambda_1 x^2 + \lambda_2 y^2 = 5$, where λ_1, λ_2 are the eigenvalues of A. To see why, let $\mathbf{r} = R\mathbf{s}$ where $\mathbf{s} = \begin{pmatrix} x & y \end{pmatrix}^T$. Then $\mathbf{r}^T A\mathbf{r} = \mathbf{s}^T(R^T AR)\mathbf{s}$. Using $R^T R = I$ gives $R^{-1} = R^T$ and $R^T AR = \mathbf{diag}(\lambda_1, \lambda_2)$. Finally, $\mathbf{r}^T A\mathbf{r} = \lambda_1 x^2 + \lambda_2 y^2$.

[6] Rod Serling, author of *The Twilight Zone*, enjoyed the view from the other side of the mirror.

Orthogonal Triad Computation

Let's compute the semiaxis directions $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$ for the ellipsoid $x^2 + 4y^2 + xy + 4z^2 = 16$. To be applied is Theorem 7. As explained on page 488, the starting point is to represent the ellipsoid equation as a quadratic form $X^TAX = 16$, where the symmetric matrix A is defined by

$$A = \begin{pmatrix} 1 & 1/2 & 0 \\ 1/2 & 4 & 0 \\ 0 & 0 & 4 \end{pmatrix}.$$

College algebra. The **characteristic polynomial** $\det(A - \lambda I) = 0$ determines the eigenvalues or hidden values of the matrix A. By cofactor expansion, this polynomial equation is

$$(4-\lambda)((1-\lambda)(4-\lambda) - 1/4) = 0$$

with roots 4, $5/2 + \sqrt{10}/2$, $5/2 - \sqrt{10}/2$.

Eigenpairs. It will be shown that three eigenpairs are

$$\lambda_1 = 4, \quad \mathbf{x}_1 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix},$$
$$\lambda_2 = \frac{5+\sqrt{10}}{2}, \quad \mathbf{x}_2 = \begin{pmatrix} \sqrt{10}-3 \\ 1 \\ 0 \end{pmatrix},$$
$$\lambda_3 = \frac{5-\sqrt{10}}{2}, \quad \mathbf{x}_3 = \begin{pmatrix} \sqrt{10}+3 \\ -1 \\ 0 \end{pmatrix}.$$

The vector norms of the eigenvectors are given by $\|\mathbf{x}_1\| = 1$, $\|\mathbf{x}_2\| = \sqrt{20 + 6\sqrt{10}}$, $\|\mathbf{x}_3\| = \sqrt{20 - 6\sqrt{10}}$. The orthonormal semi-axis directions $\mathbf{v}_k = \mathbf{x}_k/\|\mathbf{x}_k\|$, $k = 1, 2, 3$, are then given by the formulas

$$\mathbf{v}_1 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad \mathbf{v}_2 = \begin{pmatrix} \frac{\sqrt{10}-3}{\sqrt{20-6\sqrt{10}}} \\ \frac{1}{\sqrt{20-6\sqrt{10}}} \\ 0 \end{pmatrix}, \quad \mathbf{v}_3 = \begin{pmatrix} \frac{\sqrt{10}+3}{\sqrt{20+6\sqrt{10}}} \\ \frac{-1}{\sqrt{20+6\sqrt{10}}} \\ 0 \end{pmatrix}.$$

Frame sequence details.

$$\mathbf{aug}(A - \lambda_1 I, \mathbf{0}) = \left(\begin{array}{ccc|c} 1-4 & 1/2 & 0 & 0 \\ 1/2 & 4-4 & 0 & 0 \\ 0 & 0 & 4-4 & 0 \end{array}\right)$$

$$\approx \left(\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right) \qquad \text{Used combination, multiply and swap rules. Found } \textbf{rref}.$$

$$\mathbf{aug}(A - \lambda_2 I, \mathbf{0}) = \left(\begin{array}{ccc|c} \frac{-3-\sqrt{10}}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{3-\sqrt{10}}{2} & 0 & 0 \\ 0 & 0 & \frac{3-\sqrt{10}}{2} & 0 \end{array}\right)$$

$$\approx \left(\begin{array}{ccc|c} 1 & 3-\sqrt{10} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right) \quad \text{All three rules.}$$

$$\mathbf{aug}(A - \lambda_3 I, \mathbf{0}) = \left(\begin{array}{ccc|c} \frac{-3+\sqrt{10}}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{3+\sqrt{10}}{2} & 0 & 0 \\ 0 & 0 & \frac{3+\sqrt{10}}{2} & 0 \end{array}\right)$$

$$\approx \left(\begin{array}{ccc|c} 1 & 3+\sqrt{10} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right) \quad \text{All three rules.}$$

Solving the corresponding reduced echelon systems gives the preceding formulas for the eigenvectors $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$. The equation for the ellipsoid is $\lambda_1 X^2 + \lambda_2 Y^2 + \lambda_3 Z^2 = 16$, where the multipliers of the square terms are the eigenvalues of A and X, Y, Z define the new coordinate system determined by the eigenvectors of A. This equation can be re-written in the form $X^2/a^2 + Y^2/b^2 + Z^2/c^2 = 1$, provided the semi-axis lengths a, b, c are defined by the relations $a^2 = 16/\lambda_1$, $b^2 = 16/\lambda_2$, $c^2 = 16/\lambda_3$. After computation, $a = 2$, $b = 1.98$, $c = 4.17$.

9.3 Advanced Topics in Linear Algebra

Diagonalization and Jordan's Theorem

A system of differential equations $\mathbf{x}' = A\mathbf{x}$ can be transformed to an uncoupled system $\mathbf{y}' = \mathbf{diag}(\lambda_1, \ldots, \lambda_n)\mathbf{y}$ by a change of variables $\mathbf{x} = P\mathbf{y}$, provided P is invertible and A satisfies the relation

$$(1) \qquad AP = P\,\mathbf{diag}(\lambda_1, \ldots, \lambda_n).$$

A matrix A is said to be **diagonalizable** provided (1) holds. This equation is equivalent to the system of equations $A\mathbf{v}_k = \lambda_k \mathbf{v}_k$, $k = 1, \ldots, n$, where $\mathbf{v}_1$, …, $\mathbf{v}_n$ are the columns of matrix P. Since P is assumed invertible, each of its columns are nonzero, and therefore $(\lambda_k, \mathbf{v}_k)$ is an eigenpair of A, $1 \le k \le n$. The values λ_k need not be distinct (e.g., all $\lambda_k = 1$ if A is the identity). This proves:

Theorem 10 (Diagonalization)
An $n\times n$ matrix A is diagonalizable if and only if A has n eigenpairs $(\lambda_k, \mathbf{v}_k)$, $1 \le k \le n$, with $\mathbf{v}_1$, …, $\mathbf{v}_n$ independent. In this case,

$$A = PDP^{-1}$$

where $D = \mathbf{diag}(\lambda_1, \ldots, \lambda_n)$ and the matrix P has columns $\mathbf{v}_1$, …, $\mathbf{v}_n$.

Theorem 11 (Jordan's theorem)
Any $n \times n$ matrix A can be represented in the form

$$A = PTP^{-1}$$

where P is invertible and T is upper triangular. The diagonal entries of T are eigenvalues of A.

Proof: We proceed by induction on the dimension n of A. For $n = 1$ there is nothing to prove. Assume the result for dimension n, and let's prove it when A is $(n+1)\times(n+1)$. Choose an eigenpair $(\lambda_1, \mathbf{v}_1)$ of A with $\mathbf{v}_1 \ne \mathbf{0}$. Complete a basis $\mathbf{v}_1$, …, $\mathbf{v}_{n+1}$ for $\mathcal{R}^{n+1}$ and define $V = \mathbf{aug}(\mathbf{v}_1, \ldots, \mathbf{v}_{n+1})$. Then $V^{-1}AV = \left(\begin{array}{c|c} \lambda_1 & B \\ \hline \mathbf{0} & A_1 \end{array}\right)$ for some matrices B and A_1. The induction hypothesis implies there is an invertible $n \times n$ matrix P_1 and an upper triangular matrix T_1 such that $A_1 = P_1 T_1 P_1^{-1}$. Let $R = \left(\begin{array}{c|c} 1 & 0 \\ \hline 0 & P_1 \end{array}\right)$ and $\mathbf{T} = \left(\begin{array}{c|c} \lambda_1 & BT_1 \\ \hline 0 & T_1 \end{array}\right)$. Then T is upper triangular and $(V^{-1}AV)R = RT$, which implies $A = PTP^{-1}$ for $P = VR$. The induction is complete.

Jordan form J. The upper triangular matrix T in Jordan's theorem is called a **Jordan form** J of the matrix A provided

$$J = \begin{pmatrix} \lambda_1 & J_{12} & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_{n-1} & J_{n-1\,n} \\ 0 & 0 & 0 & \cdots & 0 & \lambda_n \end{pmatrix}.$$

Entries $J_{i\ i+1}$ of J along its super-diagonal are either 0 or 1, while diagonal entries λ_i are eigenvalues of A. A Jordan form is therefore a **band matrix** with zero entries off its diagonal and super-diagonal.

Coordinate system matrix P. In the equation $A = PJP^{-1}$, the columns of P are independent vectors, called **generalized eigenvectors** of A. They form a **coordinate system**. There is for each eigenvalue λ of A at least one column $\mathbf{x}$ of P satisfying $A\mathbf{x} = \lambda\mathbf{x}$. However, there may be other columns of P that *fail to be eigenvectors*, that is, $A\mathbf{x} = \lambda\mathbf{x}$ may be *false* for many columns $\mathbf{x}$ of P.

Cayley-Hamilton Identity

A celebrated and deep result for powers of matrices was discovered by Cayley and Hamilton (see [B-M]), which says that an $n \times n$ matrix A satisfies its own characteristic equation. More precisely:

Theorem 12 (Cayley-Hamilton)
Let $\det(A - \lambda I)$ be expanded as the nth degree polynomial

$$p(\lambda) = \sum_{j=0}^{n} c_j \lambda^j,$$

for some coefficients c_0, …, c_{n-1} and $c_n = (-1)^n$. Then A satisfies the equation $p(\lambda) = 0$, that is,

$$p(A) \equiv \sum_{j=0}^{n} c_j A^j = 0.$$

In factored form in terms of the eigenvalues $\{\lambda_j\}_{j=1}^n$ (duplicates possible),

$$(-1)^n (A - \lambda_1 I)(A - \lambda_2 I) \cdots (A - \lambda_n I) = 0.$$

Proof: If A is diagonalizable, $AP = P\,\mathbf{diag}(\lambda_1, \ldots, \lambda_n)$, then the proof is obtained from the simple expansion

$$A^j = P\,\mathbf{diag}(\lambda_1^j, \ldots, \lambda_n^j) P^{-1},$$

because summing across this identity leads to

$$\begin{aligned} p(A) &= \textstyle\sum_{j=0}^{n} c_j A^j \\ &= P\left(\textstyle\sum_{j=0}^{n} c_j\, \mathbf{diag}(\lambda_1^j, \ldots, \lambda_n^j)\right) P^{-1} \\ &= P\,\mathbf{diag}(p(\lambda_1), \ldots, p(\lambda_n)) P^{-1} \\ &= P\,\mathbf{diag}(0, \ldots, 0) P^{-1} \\ &= 0. \end{aligned}$$

If A is not diagonalizable, then this proof fails. To handle the general case, we apply **Jordan's theorem**, which says that $A = PTP^{-1}$ where T is *upper triangular* (instead of *diagonal*) and the not necessarily distinct eigenvalues λ_1, ..., λ_n of A appear on the diagonal of T. Using Jordan's theorem, define

$$A_\epsilon = P(T + \epsilon\,\mathbf{diag}(1, 2, \ldots, n))P^{-1}.$$

For small $\epsilon > 0$, the matrix A_ϵ has distinct eigenvalues $\lambda_j + \epsilon j$, $1 \le j \le n$. Then the diagonalizable case implies that A_ϵ satisfies its characteristic equation. Let $p_\epsilon(\lambda) = \det(A_\epsilon - \lambda I)$. Use $0 = \lim_{\epsilon\to 0} p_\epsilon(A_\epsilon) = p(A)$ to complete the proof.

Solving Triangular Differential Systems

A matrix differential system $\mathbf{y}'(t) = T\mathbf{y}(t)$ with T upper triangular splits into scalar equations which can be solved by elementary methods for first order scalar differential equations. To illustrate, consider the system

$$\begin{aligned} y_1' &= 3y_1 + x_2 + y_3, \\ y_2' &= 3y_2 + y_3, \\ y_3' &= 2y_3. \end{aligned}$$

The techniques that apply are the growth-decay recipe for $u' = ku$ and the integrating factor method for $u' = ku + p(t)$. Working backwards from the last equation, using back-substitution, gives

$$\begin{aligned} y_3 &= c_3 e^{2t}, \\ y_2 &= c_2 e^{3t} - c_3 e^{2t}, \\ y_1 &= (c_1 + c_2 t)e^{3t}. \end{aligned}$$

What has been said here applies to any triangular system $\mathbf{y}'(t) = T\mathbf{y}(t)$, in order to write an exact formula for the solution $\mathbf{y}(t)$.

If A is an $n \times n$ matrix, then Jordan's theorem gives $A = PTP^{-1}$ with T upper triangular and P invertible. The change of variable $\mathbf{x}(t) = P\mathbf{y}(t)$ changes $\mathbf{x}'(t) = A\mathbf{x}(t)$ into the triangular system $\mathbf{y}'(t) = T\mathbf{y}(t)$.

There is no special condition on A, to effect the change of variable $\mathbf{x}(t) = P\mathbf{y}(t)$. The solution $\mathbf{x}(t)$ of $\mathbf{x}'(t) = A\mathbf{x}(t)$ is a product of the invertible matrix P and a column vector $\mathbf{y}(t)$; the latter is the solution of the triangular system $\mathbf{y}'(t) = T\mathbf{y}(t)$, obtained by growth-decay and integrating factor methods.

The *importance of this idea* is to provide a solid method for solving any system $\mathbf{x}'(t) = A\mathbf{x}(t)$. In later sections, we outline how to find the matrix P and the matrix T, in Jordan's theorem $A = PTP^{-1}$. The additional theory provides efficient matrix methods for solving any system $\mathbf{x}'(t) = A\mathbf{x}(t)$.

Symmetric Matrices and Orthogonality

Described here is a process due to Gram-Schmidt for replacing a given set of independent eigenvectors by another set of eigenvectors which are of unit length and **orthogonal** (dot product zero or 90 degrees apart). The process, which applies to any independent set of vectors, is especially useful in the case of eigenanalysis of a **symmetric** matrix: $A^T = A$.

Unit eigenvectors. An eigenpair $(\lambda, \mathbf{x})$ of A can always be selected so that $\|\mathbf{x}\| = 1$. If $\|\mathbf{x}\| \neq 1$, then replace eigenvector $\mathbf{x}$ by the scalar multiple $c\mathbf{x}$, where $c = 1/\|\mathbf{x}\|$. By this small change, it can be assumed that the eigenvector has unit length. If in addition the eigenvectors are *orthogonal*, then the eigenvectors are said to be **orthonormal**.

Theorem 13 (Orthogonality of Eigenvectors)
Assume that $n \times n$ matrix A is **symmetric**, $A^T = A$. If $(\alpha, \mathbf{x})$ and $(\beta, \mathbf{y})$ are eigenpairs of A with $\alpha \neq \beta$, then $\mathbf{x}$ and $\mathbf{y}$ are orthogonal: $\mathbf{x} \cdot \mathbf{y} = 0$.

Proof: To prove this result, compute $\alpha\mathbf{x} \cdot \mathbf{y} = (A\mathbf{x})^T\mathbf{y} = \mathbf{x}^T A^T \mathbf{y} = \mathbf{x}^T A\mathbf{y}$. Also, $\beta\mathbf{x} \cdot \mathbf{y} = \mathbf{x}^T A\mathbf{y}$. Subtracting the relations implies $(\alpha - \beta)\mathbf{x} \cdot \mathbf{y} = 0$, giving $\mathbf{x} \cdot \mathbf{y} = 0$ due to $\alpha \neq \beta$. The proof is complete.

Theorem 14 (Real Eigenvalues)
If $A^T = A$, then all eigenvalues of A are real. Consequently, matrix A has n real eigenvalues counted according to multiplicity.

Proof: The second statement is due to the fundamental theorem of algebra. To prove the eigenvalues are real, it suffices to prove $\lambda = \overline{\lambda}$ when $A\mathbf{v} = \lambda\mathbf{v}$ with $\mathbf{v} \neq \mathbf{0}$. We admit that $\mathbf{v}$ may have complex entries. We will use $\overline{A} = A$ (A is real). Take the complex conjugate across $A\mathbf{v} = \lambda\mathbf{v}$ to obtain $A\overline{\mathbf{v}} = \overline{\lambda}\overline{\mathbf{v}}$. Transpose $A\mathbf{v} = \lambda\mathbf{v}$ to obtain $\mathbf{v}^T A^T = \lambda\mathbf{v}^T$ and then conclude $\mathbf{v}^T A = \lambda\mathbf{v}^T$ from $A^T = A$. Multiply this equation by $\overline{\mathbf{v}}$ on the right to obtain $\mathbf{v}^T A\overline{\mathbf{v}} = \lambda\mathbf{v}^T\overline{\mathbf{v}}$. Then multiply $A\overline{\mathbf{v}} = \overline{\lambda}\overline{\mathbf{v}}$ by $\mathbf{v}^T$ on the left to obtain $\mathbf{v}^T A\overline{\mathbf{v}} = \overline{\lambda}\mathbf{v}^T\overline{\mathbf{v}}$. Then we have

$$\lambda\mathbf{v}^T\overline{\mathbf{v}} = \overline{\lambda}\mathbf{v}^T\overline{\mathbf{v}}.$$

Because $\mathbf{v}^T\overline{\mathbf{v}} = \sum_{j=1}^n |v_j|^2 > 0$, then $\lambda = \overline{\lambda}$ and λ is real. The proof is complete.

The Gram-Schmidt process. The eigenvectors of a symmetric matrix A may be constructed to be orthogonal. First of all, observe that eigenvectors corresponding to distinct eigenvalues are orthogonal by Theorem 13. It remains to construct from k independent eigenvectors $\mathbf{x}_1, \ldots, \mathbf{x}_k$, corresponding to a single eigenvalue λ, another set of independent eigenvectors $\mathbf{y}_1, \ldots, \mathbf{y}_k$ for λ which are pairwise orthogonal. The idea, due to Gram-Schmidt, applies to any set of k independent vectors.

Application of the Gram-Schmidt process can be illustrated by example: let $(-1, \mathbf{v}_1)$, $(2, \mathbf{v}_2)$, $(2, \mathbf{v}_3)$, $(2, \mathbf{v}_4)$ be eigenpairs of a 4×4 symmetric matrix A. Then $\mathbf{v}_1$ is orthogonal to $\mathbf{v}_2$, $\mathbf{v}_3$, $\mathbf{v}_4$. The vectors $\mathbf{v}_2$, $\mathbf{v}_3$, $\mathbf{v}_4$ belong to eigenvalue $\lambda = 2$, but they are not necessarily orthogonal. The Gram-Schmidt process replaces these vectors by $\mathbf{y}_2$, $\mathbf{y}_3$, $\mathbf{y}_4$ which are pairwise orthogonal. The result is that eigenvectors $\mathbf{v}_1$, $\mathbf{y}_2$, $\mathbf{y}_3$, $\mathbf{y}_4$ are pairwise orthogonal.

Theorem 15 (Gram-Schmidt)
Let $\mathbf{x}_1, \ldots, \mathbf{x}_k$ be independent n-vectors. The set of vectors $\mathbf{y}_1, \ldots, \mathbf{y}_k$ constructed below as linear combinations of $\mathbf{x}_1, \ldots, \mathbf{x}_k$ are pairwise orthogonal and independent.

$$\begin{aligned}
\mathbf{y}_1 &= \mathbf{x}_1 \\
\mathbf{y}_2 &= \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{y}_1}{\mathbf{y}_1 \cdot \mathbf{y}_1}\mathbf{y}_1 \\
\mathbf{y}_3 &= \mathbf{x}_3 - \frac{\mathbf{x}_3 \cdot \mathbf{y}_1}{\mathbf{y}_1 \cdot \mathbf{y}_1}\mathbf{y}_1 - \frac{\mathbf{x}_3 \cdot \mathbf{y}_2}{\mathbf{y}_2 \cdot \mathbf{y}_2}\mathbf{y}_2 \\
&\vdots \\
\mathbf{y}_k &= \mathbf{x}_k - \frac{\mathbf{x}_k \cdot \mathbf{y}_1}{\mathbf{y}_1 \cdot \mathbf{y}_1}\mathbf{y}_1 - \cdots - \frac{\mathbf{x}_k \cdot \mathbf{y}_{k-1}}{\mathbf{y}_{k-1} \cdot \mathbf{y}_{k-1}}\mathbf{y}_{k-1}
\end{aligned}$$

Proof: Let's begin with a lemma: *Any set of nonzero orthogonal vectors* $\mathbf{y}_1, \ldots, \mathbf{y}_k$ *are independent.* Assume the relation $c_1\mathbf{y}_1 + \cdots + c_k\mathbf{y}_k = \mathbf{0}$. Take the dot product of this relation with $\mathbf{y}_j$. By orthogonality, $c_j\,\mathbf{y}_j \cdot \mathbf{y}_j = 0$, and since $\mathbf{y}_j \neq \mathbf{0}$, cancellation gives $c_j = 0$ for $1 \le j \le k$. Hence $\mathbf{y}_1, \ldots, \mathbf{y}_k$ are independent.

Induction will be applied on k to show that $\mathbf{y}_1, \ldots, \mathbf{y}_k$ are nonzero and orthogonal. If $k = 1$, then there is just one nonzero vector constructed $\mathbf{y}_1 = \mathbf{x}_1$. Orthogonality for $k = 1$ is not discussed because there are no pairs to test. Assume the result holds for $k - 1$ vectors. Let's verify that it holds for k vectors, $k > 1$. Assume orthogonality $\mathbf{y}_i \cdot \mathbf{y}_j = 0$ and $\mathbf{y}_i \neq \mathbf{0}$ for $1 \le i, j \le k - 1$. It remains to test $\mathbf{y}_i \cdot \mathbf{y}_k = 0$ for $1 \le i \le k - 1$ and $\mathbf{y}_k \neq \mathbf{0}$. The test depends upon the identity

$$\mathbf{y}_i \cdot \mathbf{y}_k = \mathbf{y}_i \cdot \mathbf{x}_k - \sum_{j=1}^{k-1} \frac{\mathbf{x}_k \cdot \mathbf{y}_j}{\mathbf{y}_j \cdot \mathbf{y}_j}\, \mathbf{y}_i \cdot \mathbf{y}_j,$$

which is obtained from the formula for $\mathbf{y}_k$ by taking the dot product with $\mathbf{y}_i$. In the identity, $\mathbf{y}_i \cdot \mathbf{y}_j = 0$ by the induction hypothesis for $1 \le j \le k-1$ and $j \neq i$. Therefore, the summation in the identity contains just the term for index $j = i$, and the contribution is $\mathbf{y}_i \cdot \mathbf{x}_k$. This contribution cancels the leading term on the right in the identity, resulting in the orthogonality relation $\mathbf{y}_i \cdot \mathbf{y}_k = 0$. If $\mathbf{y}_k = \mathbf{0}$, then $\mathbf{x}_k$ is a linear combination of $\mathbf{y}_1, \ldots, \mathbf{y}_{k-1}$. But each $\mathbf{y}_j$ is a linear combination of $\{\mathbf{x}_i\}_{i=1}^{j}$, therefore $\mathbf{y}_k = \mathbf{0}$ implies $\mathbf{x}_k$ is a linear combination of $\mathbf{x}_1, \ldots, \mathbf{x}_{k-1}$, a contradiction to the independence of $\{\mathbf{x}_i\}_{i=1}^{k}$. The proof is complete.

Orthogonal Projection.

Reproduced here is the basic material on shadow projection, for the convenience of the reader. The ideas are then extended to obtain the orthogonal projection onto a subspace V of $\mathcal{R}^n$. Finally, the orthogonal projection formula is related to the Gram-Schmidt equations.

The **shadow projection** of vector $\vec{X}$ onto the direction of vector $\vec{Y}$ is the number d defined by

$$d = \frac{\vec{X} \cdot \vec{Y}}{|\vec{Y}|}.$$

The triangle determined by $\vec{X}$ and $d\dfrac{\vec{Y}}{|\vec{Y}|}$ is a right triangle.

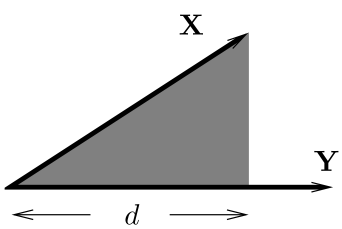


Figure 3. Shadow projection d of vector X onto the direction of vector Y.

The **vector shadow projection** of $\vec{X}$ onto the line L through the origin in the direction of $\vec{Y}$ is defined by

$$\mathbf{proj}_{\vec{Y}}(\vec{X}) = d\frac{\vec{Y}}{|\vec{Y}|} = \frac{\vec{X} \cdot \vec{Y}}{\vec{Y} \cdot \vec{Y}}\vec{Y}.$$

Orthogonal projection for dimension 1. The extension of the shadow projection formula to a subspace V of $\mathcal{R}^n$ begins with unitizing $\vec{Y}$ to isolate the vector direction $\mathbf{u} = \vec{Y}/\|\vec{Y}\|$ of line L. Define the subspace $V = \mathbf{span}\{\mathbf{u}\}$. Then V is identical to L. We define the **orthogonal projection** by the formula

$$\mathbf{Proj}_V(\mathbf{x}) = (\mathbf{u} \cdot \mathbf{x})\mathbf{u}, \quad V = \mathbf{span}\{\mathbf{u}\}.$$

The reader is asked to verify that

$$\mathbf{proj}_{\vec{Y}}(\mathbf{x}) = \mathbf{Proj}_V(\mathbf{x}) = d\mathbf{u}.$$

These equalities mean that the orthogonal projection is the vector shadow projection when V is one dimensional.

Orthogonal projection for dimension k. Consider a subspace V of $\mathcal{R}^n$ given as the span of orthonormal vectors $\mathbf{u}_1, \ldots, \mathbf{u}_k$. Define the **orthogonal projection** by the formula

$$\mathbf{Proj}_V(\mathbf{x}) = \sum_{j=1}^{k} (\mathbf{u}_j \cdot \mathbf{x})\mathbf{u}_j, \quad V = \mathbf{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}.$$

Orthogonal projection and Gram-Schmidt. Define $\mathbf{y}_1, \ldots, \mathbf{y}_k$ by the Gram-Schmidt relations on page 496. Let $\mathbf{u}_j = \mathbf{y}_j/\|\mathbf{y}_j\|$ for $j = 1, \ldots, k$. Then $V_{j-1} = \mathbf{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_{j-1}\}$ is a subspace of $\mathcal{R}^n$ of dimension $j-1$ with orthonormal basis $\mathbf{u}_1, \ldots, \mathbf{u}_{j-1}$ and

$$\begin{aligned} \mathbf{y}_j &= \mathbf{x}_j - \frac{\mathbf{x}_j \cdot \mathbf{y}_1}{\mathbf{y}_1 \cdot \mathbf{y}_1}\mathbf{y}_1 - \cdots - \frac{\mathbf{x}_k \cdot \mathbf{y}_{j-1}}{\mathbf{y}_{j-1} \cdot \mathbf{y}_{j-1}}\mathbf{y}_{j-1} \\ &= \mathbf{x}_j - \mathbf{Proj}_{V_{j-1}}(\mathbf{x}_j). \end{aligned}$$

In remembering the Gram-Schmidt formulas, and in the use of the orthogonal projection in proofs and constructions, the following key theorem is useful.

Theorem 16 (Orthogonal Projection Properties)
Let V be the span of orthonormal vectors $\mathbf{u}_1, \ldots, \mathbf{u}_k$.

(a) Every vector in V has an orthogonal expansion $\mathbf{v} = \sum_{j=1}^{k} (\mathbf{v} \cdot \mathbf{u}_j)\mathbf{u}_j$.

(b) The vector $\mathbf{Proj}_V(\mathbf{x})$ is a vector in the subspace V.

(c) The vector $\mathbf{w} = \mathbf{x} - \mathbf{Proj}_V(\mathbf{x})$ is orthogonal to every vector in V.

(d) Among all vectors $\mathbf{v}$ in V, the minimum value of $\|\mathbf{x} - \mathbf{v}\|$ is uniquely obtained by the orthogonal projection $\mathbf{v} = \mathbf{Proj}_V(\mathbf{x})$.

Proof:
(a): Every element $\mathbf{v}$ in V is a linear combination of basis elements:

$$\mathbf{v} = c_1\mathbf{u}_1 + \cdots + c_k\mathbf{u}_k.$$

Take the dot product of this relation with basis element $\mathbf{u}_j$. By orthogonality, $c_j = \mathbf{v} \cdot \mathbf{u}_j$.

(b): Because $\mathbf{Proj}_V(\mathbf{x})$ is a linear combination of basis elements of V, then (b) holds.

(c): Let's compute the dot product of $\mathbf{w}$ and $\mathbf{v}$. We will use the orthogonal expansion from (a).

$$\begin{aligned} \mathbf{w} \cdot \mathbf{v} &= (\mathbf{x} - \mathbf{Proj}_V(\mathbf{x})) \cdot \mathbf{v} \\ &= \mathbf{x} \cdot \mathbf{v} - \left(\sum_{j=1}^{k} (\mathbf{x} \cdot \mathbf{u}_j)\mathbf{u}_j \right) \cdot \mathbf{v} \\ &= \sum_{j=1}^{k} (\mathbf{v} \cdot \mathbf{u}_j)(\mathbf{u}_j \cdot \mathbf{x}) - \sum_{j=1}^{k} (\mathbf{x} \cdot \mathbf{u}_j)(\mathbf{u}_j \cdot \mathbf{v}) \\ &= 0. \end{aligned}$$

(d): Begin with the Pythagorean identity

$$\|\mathbf{a}\|^2 + \|\mathbf{b}\|^2 = \|\mathbf{a}+\mathbf{b}\|^2$$

valid exactly when $\mathbf{a}\cdot\mathbf{b} = 0$ (a right triangle, $\theta = 90°$). Using an arbitrary $\mathbf{v}$ in V, define $\mathbf{a} = \mathbf{Proj}_V(\mathbf{x}) - \mathbf{v}$ and $\mathbf{b} = \mathbf{x} - \mathbf{Proj}_V(\mathbf{x})$. By (b), vector $\mathbf{a}$ is in V. Because of (c), then $\mathbf{a}\cdot\mathbf{b} = 0$. This gives the identity

$$\|\mathbf{Proj}_V(\mathbf{x}) - \mathbf{v}\|^2 + \|\mathbf{x} - \mathbf{Proj}_V(\mathbf{x})\|^2 = \|\mathbf{x}-\mathbf{v}\|^2,$$

which establishes $\|\mathbf{x} - \mathbf{Proj}_V(\mathbf{x})\| < \|\mathbf{x}-\mathbf{v}\|$ except for the unique $\mathbf{v}$ such that $\|\mathbf{Proj}_V(\mathbf{x}) - \mathbf{v}\| = 0$.

The proof is complete.

Theorem 17 (Near Point to a Subspace)
Let V be a subspace of $\mathcal{R}^n$ and $\mathbf{x}$ a vector not in V. The **near point** to $\mathbf{x}$ in V is the orthogonal projection of $\mathbf{x}$ onto V. This point is characterized as the minimum of $\|\mathbf{x}-\mathbf{v}\|$ over all vectors $\mathbf{v}$ in the subspace V.

Proof: Apply **(d)** of the preceding theorem.

Theorem 18 (Cross Product and Projections)
The cross product direction $\mathbf{a}\times\mathbf{b}$ can be computed as $\mathbf{c} - \mathbf{Proj}_V(\mathbf{c})$, by selecting a direction $\mathbf{c}$ not in $V = \mathbf{span}\{\mathbf{a},\mathbf{b}\}$.

Proof: The cross product makes sense only in $\mathcal{R}^3$. Subspace V is two dimensional when $\mathbf{a}$, $\mathbf{b}$ are independent, and Gram-Schmidt applies to find an orthonormal basis $\mathbf{u}_1$, $\mathbf{u}_2$. By (c) of Theorem 16, the vector $\mathbf{c} - \mathbf{Proj}_V(\mathbf{c})$ has the same or opposite direction to the cross product.

Theorem 19 (The QR-Decomposition)
Let the $m\times n$ matrix A have independent columns $\mathbf{x}_1, \ldots, \mathbf{x}_n$. Then there is an upper triangular matrix R with positive diagonal entries and an orthonormal matrix Q such that

$$A = QR.$$

Proof: Let $\mathbf{y}_1, \ldots, \mathbf{y}_n$ be the Gram-Schmidt orthogonal vectors given by relations on page 496. Define $\mathbf{u}_k = \mathbf{y}_k/\|\mathbf{y}_k\|$ and $r_{kk} = \|\mathbf{y}_k\|$ for $k = 1,\ldots,n$, and otherwise $r_{ij} = \mathbf{u}_i\cdot\mathbf{x}_j$. Let $Q = \mathbf{aug}(\mathbf{u}_1,\ldots,\mathbf{u}_n)$. Then

$$(2)\qquad \begin{aligned} \mathbf{x}_1 &= r_{11}\mathbf{u}_1, \\ \mathbf{x}_2 &= r_{22}\mathbf{u}_2 + r_{21}\mathbf{u}_1, \\ \mathbf{x}_3 &= r_{33}\mathbf{u}_3 + r_{31}\mathbf{u}_1 + r_{32}\mathbf{u}_2, \\ &\vdots \\ \mathbf{x}_n &= r_{nn}\mathbf{u}_n + r_{n1}\mathbf{u}_1 + \cdots + r_{nn-1}\mathbf{u}_{n-1}. \end{aligned}$$

It follows from (2) and matrix multiplication that $A = QR$. The proof is complete.

Theorem 20 (Matrices Q and R in $A = QR$)
Let the $m \times n$ matrix A have independent columns $\mathbf{x}_1, \ldots, \mathbf{x}_n$. Let $\mathbf{y}_1, \ldots, \mathbf{y}_n$ be the Gram-Schmidt orthogonal vectors given by relations on page 496. Define $\mathbf{u}_k = \mathbf{y}_k/\|\mathbf{y}_k\|$. Then $AQ = QR$ is satisfied by $Q = \mathbf{aug}(\mathbf{u}_1, \ldots, \mathbf{u}_n)$ and

$$R = \begin{pmatrix} \|y_1\| & \mathbf{u}_1 \cdot \mathbf{x}_2 & \mathbf{u}_1 \cdot \mathbf{x}_3 & \cdots & \mathbf{u}_1 \cdot \mathbf{x}_n \\ 0 & \|y_2\| & \mathbf{u}_2 \cdot \mathbf{x}_3 & \cdots & \mathbf{u}_2 \cdot \mathbf{x}_n \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & \|y_n\| \end{pmatrix}.$$

Proof: The result is contained in the proof of the previous theorem.
Some references cite the diagonal entries as $\|\mathbf{x}_1\|$, $\|\mathbf{x}_2^\perp\|, \ldots, \|\mathbf{x}_n^\perp\|$, where $\mathbf{x}_j^\perp = \mathbf{x}_j - \mathbf{Proj}_{V_{j-1}}(\mathbf{x}_j)$, $V_{j-1} = \mathbf{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_{j-1}\}$. Because $\mathbf{y}_1 = \mathbf{x}_1$ and $\mathbf{y}_j = \mathbf{x}_j - \mathbf{Proj}_{V_{j-1}}(\mathbf{x}_j)$, the formulas for R are identical.

Theorem 21 (Uniqueness of Q and R)
Let $m \times n$ matrix A have independent columns and satisfy the decomposition $A = QR$. If Q is $m \times n$ orthogonal and R is $n \times n$ upper triangular with positive diagonal elements, then Q and R are uniquely determined.

Proof: The problem is to show that $A = Q_1R_1 = Q_2R_2$ implies $R_2R_1^{-1} = I$ and $Q_1 = Q_2$. We start with $Q_1 = Q_2R_2R_1^{-1}$. Define $P = R_2R_1^{-1}$. Then $Q_1 = Q_2P$. Because $I = Q_1^TQ_1 = P^TQ_2^TQ_2P = P^TP$, then P is orthogonal. Matrix P is the product of square upper triangular matrices with positive diagonal elements, which implies P itself is square upper triangular with positive diagonal elements. The only matrix with these properties is the identity matrix I. Then $R_2R_1^{-1} = P = I$, which implies $R_1 = R_2$ and $Q_1 = Q_2$. The proof is complete.

Theorem 22 (Orthonormal Diagonal Form)
Let A be a given $n \times n$ real matrix. Then $A = QDQ^{-1}$ with Q orthogonal and D diagonal if and only if $A^T = A$.

Proof: The reader is reminded that Q orthogonal means that the columns of Q are **orthonormal**. The equation $A = A^T$ means A is **symmetric**.

Assume first that $A = QDQ^{-1}$ with $Q = Q^T$ orthogonal ($Q^TQ = I$) and D diagonal. Then $Q^T = Q = Q^{-1}$. This implies $A^T = (QDQ^{-1})^T = (Q^{-1})^TD^TQ^T = QDQ^{-1} = A$.

Conversely, assume $A^T = A$. Then the eigenvalues of A are real and eigenvectors corresponding to distinct eigenvalues are orthogonal. The proof proceeds by induction on the dimension n of the $n \times n$ matrix A.

For $n = 1$, let Q be the 1×1 identity matrix. Then Q is orthogonal and $AQ = QD$ where D is 1×1 diagonal.

Assume the decomposition $AQ = QD$ for dimension n. Let's prove it for A of dimension $n + 1$. Choose a real eigenvalue λ of A and eigenvector $\mathbf{v}_1$ with $\|\mathbf{v}_1\| = 1$. Complete a basis $\mathbf{v}_1, \ldots, \mathbf{v}_{n+1}$ of $\mathcal{R}^{n+1}$. By Gram-Schmidt, we assume as well that this basis is orthonormal. Define $P = \mathbf{aug}(\mathbf{v}_1, \ldots, \mathbf{v}_{n+1})$.

Then P is orthogonal and satisfies $P^T = P^{-1}$. Define $B = P^{-1}AP$. Then B is symmetric ($B^T = B$) and $\mathbf{col}(B, 1) = \lambda\,\mathbf{col}(I, 1)$. These facts imply that B is a block matrix

$$B = \left(\begin{array}{c|c} \lambda & 0 \\ \hline 0 & C \end{array}\right)$$

where C is symmetric ($C^T = C$). The induction hypothesis applies to C to obtain the existence of an orthogonal matrix Q_1 such that $CQ_1 = Q_1D_1$ for some diagonal matrix D_1. Define a diagonal matrix D and matrices W and Q as follows:

$$\begin{aligned} D &= \left(\begin{array}{c|c} \lambda & 0 \\ \hline 0 & D_1 \end{array}\right), \\ W &= \left(\begin{array}{c|c} 1 & 0 \\ \hline 0 & Q_1 \end{array}\right), \\ Q &= PW. \end{aligned}$$

Then Q is the product of two orthogonal matrices, which makes Q orthogonal. Compute

$$W^{-1}BW = \left(\begin{array}{c|c} 1 & 0 \\ \hline 0 & Q_1^{-1} \end{array}\right) \left(\begin{array}{c|c} \lambda & 0 \\ \hline 0 & C \end{array}\right) \left(\begin{array}{c|c} 1 & 0 \\ \hline 0 & Q_1 \end{array}\right) = \left(\begin{array}{c|c} \lambda & 0 \\ \hline 0 & D_1 \end{array}\right).$$

Then $Q^{-1}AQ = W^{-1}P^{-1}APW = W^{-1}BW = D$. This completes the induction, ending the proof of the theorem.

Theorem 23 (Eigenpairs of a Symmetric A)
Let A be a symmetric $n \times n$ real matrix. Then A has n eigenpairs $(\lambda_1, \mathbf{v}_1)$, ..., $(\lambda_n, \mathbf{v}_n)$, with independent eigenvectors $\mathbf{v}_1$, ..., $\mathbf{v}_n$.

Proof: The preceding theorem applies to prove the existence of an orthogonal matrix Q and a diagonal matrix D such that $AQ = QD$. The diagonal entries of D are the eigenvalues of A, in some order. For a diagonal entry λ of D appearing in row j, the relation $A\,\mathbf{col}(Q, j) = \lambda\,\mathbf{col}(Q, j)$ holds, which implies that A has n eigenpairs. The eigenvectors are the columns of Q, which are independent because Q is orthogonal. The proof is complete.

Theorem 24 (Diagonalization of Symmetric A)
Let A be a symmetric $n \times n$ real matrix. Then A has n eigenpairs. For each distinct eigenvalue λ, replace the eigenvectors by orthonormal eigenvectors, using the Gram-Schmidt process. Let $\mathbf{u}_j$, ..., $\mathbf{u}_n$ be the orthonormal vectors so obtained and define

$$Q = \mathbf{aug}(\mathbf{u}_1, \ldots, \mathbf{u}_n), \quad D = \mathbf{diag}(\lambda_1, \ldots, \lambda_n).$$

Then Q is orthogonal and $AQ = QD$.

Proof: The preceding theorem justifies the eigenanalysis result. Already, eigenpairs corresponding to distinct eigenvalues are orthogonal. Within the set of eigenpairs with the same eigenvalue λ, the Gram-Schmidt process produces a replacement basis of orthonormal eigenvectors. Then the union of all the eigenvectors is orthonormal. The process described here does not disturb the ordering of eigenpairs, because it only replaces an eigenvector. The proof is complete.

The Singular Value Decomposition

Theorem 25 (Positive Eigenvalues of A^TA)
Given an $m \times n$ real matrix A, then A^TA is a real symmetric matrix whose eigenpairs $(\lambda, \mathbf{v})$ satisfy

$$\lambda = \frac{\|A\mathbf{v}\|^2}{\|\mathbf{v}\|^2} \geq 0. \tag{3}$$

Proof: Symmetry follows from $(A^TA)^T = A^T(A^T)^T = A^TA$. An eigenpair $(\lambda, \mathbf{v})$ satisfies $\lambda\overline{\mathbf{v}}^T\mathbf{v} = \overline{\mathbf{v}}^TA^TA\mathbf{v} = (\overline{A\mathbf{v}})^T(A\mathbf{v}) = \|A\mathbf{v}\|^2$, hence (3).

Definition 4 (Singular Values of A)
Let the real symmetric matrix A^TA have real eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0 = \lambda_{r+1} = \cdots = \lambda_n$. The numbers $\sigma_k = \sqrt{\lambda_k}$ $(1 \leq k \leq n)$ are called the **singular values** of the matrix A. The ordering of the singular values is always with decreasing magnitude.

Theorem 26 (Orthonormal Set $\mathbf{u}_1$, . . . , $\mathbf{u}_m$)
Let the real symmetric matrix A^TA have real eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0 = \lambda_{r+1} = \cdots = \lambda_n$ and corresponding orthonormal eigenvectors $\mathbf{v}_1$,. . . ,$\mathbf{v}_n$, obtained by the Gram-Schmidt process. Define the vectors

$$\mathbf{u}_1 = \frac{1}{\sigma_1}A\mathbf{v}_1, \ldots, \mathbf{u}_r = \frac{1}{\sigma_r}A\mathbf{v}_r.$$

Because $\|A\mathbf{v}_k\| = \sigma_k$, then $\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ is orthonormal. Gram-Schmidt can extend this set to an orthonormal basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ of $\mathcal{R}^m$.

Theorem 27 (The Singular Value Decomposition (svd))
Let A be a given real $m \times n$ matrix. Let $(\lambda_1, \mathbf{v}_1)$,. . . ,$(\lambda_n, \mathbf{v}_n)$ be a set of orthonormal eigenpairs for A^TA such that $\sigma_k = \sqrt{\lambda_k}$ $(1 \leq k \leq r)$ defines the positive singular values of A and $\lambda_k = 0$ for $r < k \leq n$. Complete $\mathbf{u}_1 = (1/\sigma_1)A\mathbf{v}_1$, . . . , $\mathbf{u}_r = (1/\sigma_r)A\mathbf{v}_r$ to an orthonormal basis $\{\mathbf{u}_k\}_{k=1}^m$for $\mathcal{R}^m$. Define

$$U = \mathbf{aug}(\mathbf{u}_1, \ldots, \mathbf{u}_m), \quad \Sigma = \left(\begin{array}{c|c} \mathbf{diag}(\sigma_1, \ldots, \sigma_r) & 0 \\ \hline 0 & 0 \end{array}\right),$$
$$V = \mathbf{aug}(\mathbf{v}_1, \ldots, \mathbf{v}_n).$$

Then the columns of U and V are orthonormal and

$$\begin{aligned} A &= U\Sigma V^T \\ &= \sigma_1\mathbf{u}_1\mathbf{v}_1^T + \cdots + \sigma_r\mathbf{u}_r\mathbf{v}_r^T \\ &= A(\mathbf{v}_1)\mathbf{v}_1^T + \cdots + A(\mathbf{v}_r)\mathbf{v}_r^T \end{aligned}$$

Proof of Theorem 26: Because $A^TA\mathbf{v}_k = \lambda_k\mathbf{v}_k \neq \mathbf{0}$ for $1 \leq k \leq r$, the vectors $\mathbf{u}_k$ are nonzero. Given $i \neq j$, then $\sigma_i\sigma_j\mathbf{u}_i \cdot \mathbf{u}_j = (A\mathbf{v}_i)^T(A\mathbf{v}_j) =$

$\lambda_j \mathbf{v}_i^T \mathbf{v}_j = 0$, showing that the vectors $\mathbf{u}_k$ are orthogonal. Further, $\|\mathbf{u}_k\|^2 = \mathbf{v}_k \cdot (A^T A \mathbf{v}_k)/\lambda_k = \|\mathbf{v}_k\|^2 = 1$ because $\{\mathbf{v}_k\}_{k=1}^n$ is an orthonormal set.

The extension of the $\mathbf{u}_k$ to an orthonormal basis of $\mathcal{R}^m$ is not unique, because it depends upon a choice of independent spanning vectors $\mathbf{y}_{r+1}, \ldots, \mathbf{y}_m$ for the set $\{\mathbf{x} \ : \ \mathbf{x} \cdot \mathbf{u}_k = 0, \quad 1 \le k \le r\}$. Once selected, Gram-Schmidt is applied to $\mathbf{u}_1, \ldots, \mathbf{u}_r, \mathbf{y}_{r+1}, \ldots, \mathbf{y}_m$ to obtain the desired orthonormal basis.

Proof of Theorem 27: The product of U and Σ is the $m \times n$ matrix

$$\begin{aligned} U\Sigma &= \mathbf{aug}(\sigma_1 \mathbf{u}_1, \ldots, \sigma_r \mathbf{u}_r, \mathbf{0}, \ldots, \mathbf{0}) \\ &= \mathbf{aug}(A(\mathbf{v}_1), \ldots, A(\mathbf{v}_r), \mathbf{0}, \ldots, \mathbf{0}). \end{aligned}$$

Let $\mathbf{v}$ be any vector in $\mathcal{R}^n$. It will be shown that $U\Sigma V^T \mathbf{v}$, $\sum_{k=1}^r A(\mathbf{v}_k)(\mathbf{v}_k^T \mathbf{v})$ and $A\mathbf{v}$ are the same column vector. We have the equalities

$$\begin{aligned} U\Sigma V^T \mathbf{v} &= U\Sigma \begin{pmatrix} \mathbf{v}_1^T \mathbf{v} \\ \vdots \\ \mathbf{v}_n^T \mathbf{v} \end{pmatrix} \\ &= \mathbf{aug}(A(\mathbf{v}_1), \ldots, A(\mathbf{v}_r), \mathbf{0}, \ldots, \mathbf{0}) \begin{pmatrix} \mathbf{v}_1^T \mathbf{v} \\ \vdots \\ \mathbf{v}_n^T \mathbf{v} \end{pmatrix} \\ &= \sum_{k=1}^r (\mathbf{v}_k^T \mathbf{v}) A(\mathbf{v}_k). \end{aligned}$$

Because $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is an orthonormal basis of $\mathcal{R}^n$, then $\mathbf{v} = \sum_{k=1}^n (\mathbf{v}_k^T \mathbf{v}) \mathbf{v}_k$. Additionally, $A(\mathbf{v}_k) = \mathbf{0}$ for $r < k \le n$ implies

$$\begin{aligned} A\mathbf{v} &= A\left(\sum_{k=1}^n (\mathbf{v}_k^T \mathbf{v}) \mathbf{v}_k\right) \\ &= \sum_{k=1}^r (\mathbf{v}_k^T \mathbf{v}) A(\mathbf{v}_k) \end{aligned}$$

Then $A\mathbf{v} = U\Sigma V^T \mathbf{v} = \sum_{k=1}^r A(\mathbf{v}_k)(\mathbf{v}_k^T \mathbf{v})$, which proves the theorem.

Standard equation of an ellipse. Calculus courses consider ellipse equations like $85x^2 - 60xy + 40y^2 = 2500$ and discuss removal of the cross term $-60xy$. The objective is to obtain a **standard** ellipse equation $\dfrac{X^2}{a^2} + \dfrac{Y^2}{b^2} = 1$. We re-visit this old problem from a different point of view, and in the derivation establish a connection between the ellipse equation, the symmetric matrix $A^T A$, and the singular values of A.

9 Example (Image of the Unit Circle) Let $A = \begin{pmatrix} -2 & 6 \\ 6 & 7 \end{pmatrix}$.

Then the invertible matrix A maps the unit circle K into the ellipse

$$85x^2 - 60xy + 40y^2 = 2500.$$

Verify that after a rotation to remove the xy-term, in the new XY-coordinates the equation is $\dfrac{X^2}{a^2} + \dfrac{Y^2}{b^2} = 1$, where $a = 10$ and $b = 5$.

Solution: The Pythagorean identity $\cos^2\theta + \sin^2\theta = 1$ will be used together with the parameterization $\theta \to (\cos\theta, \sin\theta)$ of the unit circle K, $0 \le \theta \le 2\pi$. Mapping K by the matrix A is formally the set of dual relations

$$\begin{pmatrix} x \\ y \end{pmatrix} = A \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}, \quad \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} = A^{-1} \begin{pmatrix} x \\ y \end{pmatrix}.$$

The Pythagorean identity used on the second relation implies

$$85x^2 - 60xy + 40y^2 = 2500.$$

This ellipse equation can be represented by the vector-matrix identity

$$\begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} 85 & 30 \\ 30 & 40 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 2500.$$

The symmetric matrix $A^TA = \begin{pmatrix} 85 & 30 \\ 30 & 40 \end{pmatrix}$ has eigenpair packages

$$P = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix}, \quad D = \begin{pmatrix} 25 & 0 \\ 0 & 100 \end{pmatrix}.$$

In the coordinate system $\begin{pmatrix} x \\ y \end{pmatrix} = P \begin{pmatrix} X \\ Y \end{pmatrix}$ of the orthogonal matrix P, the ellipse vector-matrix identity becomes

$$\begin{pmatrix} X & Y \end{pmatrix} P^T \begin{pmatrix} 85 & 30 \\ 30 & 40 \end{pmatrix} P \begin{pmatrix} X \\ Y \end{pmatrix} = 2500.$$

Because $P^T \left(A^TA\right) P = D = \textbf{diag}(25, 100)$, then the ellipse equation has the standard form

$$25X^2 + 100Y^2 = 2500.$$

The semi-axis lengths for this ellipse are $a = \sqrt{\frac{2500}{25}} = 10$ and $b = \sqrt{\frac{2500}{100}} = 5$, which are precisely the singular values $\sigma_1 = 10$ and $\sigma_2 = 5$ of matrix A.

Singular values and geometry.

The preceding example is typical for all invertible 2×2 matrices A. Described here is the geometrical interpretation for the singular value decomposition $A = U\Sigma V^T$, shown in Figure 4.

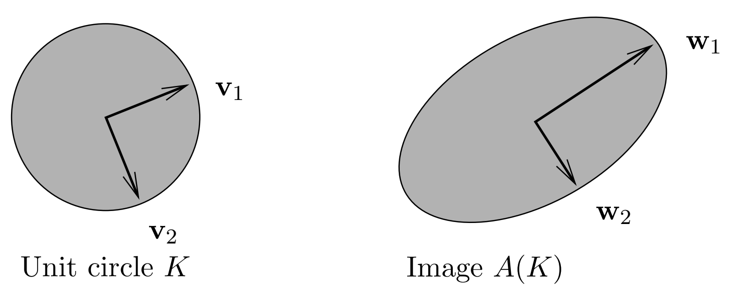


Figure 4. Mapping the unit circle.
Invertible matrix A maps the unit circle K into the ellipse $A(K)$. Orthonormal vectors $\mathbf{v}_1$, $\mathbf{v}_2$ are mapped by matrix $A = U\Sigma V^T$ into orthogonal vectors $\mathbf{w}_1 = A\mathbf{v}_1$, $\mathbf{w}_2 = A\mathbf{v}_2$, which are the semi-axes vectors of the ellipse. The semi-axis lengths $\|\mathbf{w}_1\|$, $\|\mathbf{w}_2\|$ equal the singular values σ_1, σ_2.

A summary of the example $A = \begin{pmatrix} -2 & 6 \\ 6 & 7 \end{pmatrix}$:

> A 2×2 invertible matrix A maps the unit circle K into an ellipse $A(K)$. Decomposition $A = U\Sigma V^T$ means A maps the columns of V into re-scaled columns of U. These vectors, $\sigma_1 \mathbf{u}_1$ and $\sigma_2 \mathbf{u}_2$, are the semi-axis vectors of the ellipse $A(K)$, whose lengths σ_1, σ_2 are the singular values.

The columns of V are orthonormal vectors $\mathbf{v}_1$, $\mathbf{v}_2$, computed as eigenpairs $(\lambda_1, \mathbf{v}_1)$, $(\lambda_2, \mathbf{v}_2)$ of A^TA. Then $A\mathbf{v}_1 = U\Sigma V^T \mathbf{v}_1 = U \begin{pmatrix} \sigma_1 \\ 0 \end{pmatrix} = \sigma_1 \mathbf{u}_1$. Similarly, $A\mathbf{v}_2 = U\Sigma V^T \mathbf{v}_2 = U \begin{pmatrix} 0 \\ \sigma_2 \end{pmatrix} = \sigma_2 \mathbf{u}_2$.

11.6 Jordan Form and Eigenanalysis

Generalized Eigenanalysis

The main result of generalized eigenanalysis is **Jordan's theorem**

$$A = PJP^{-1},$$

valid for any real or complex square matrix A. We describe here how to compute the invertible matrix P of generalized eigenvectors and the upper triangular matrix J, called a Jordan form of A:

$$J = \begin{pmatrix} \lambda_1 & J_{12} & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_{n-1} & J_{n-1\,n} \\ 0 & 0 & 0 & \cdots & 0 & \lambda_n \end{pmatrix}.$$

Entries $J_{i\,i+1}$ of J along its super-diagonal are either 0 or 1, while diagonal entries λ_i are eigenvalues of A. A Jordan form is therefore a **band matrix** with zero entries off its diagonal and super-diagonal.

An $m \times m$ matrix $B(\lambda, m)$ is called a **Jordan block** provided it is a Jordan form, all m diagonal elements are the same eigenvalue λ and all super-diagonal elements are one:

$$B(\lambda, m) = \begin{pmatrix} \lambda & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda & 1 \\ 0 & 0 & 0 & \cdots & 0 & \lambda \end{pmatrix} \quad (m \times m \text{ matrix})$$

The Jordan block form of J. Given a square matrix A, a Jordan form J for A is built from Jordan blocks, more precisely, J is a block diagonal matrix

$$J = \textbf{diag}(B(\lambda_1, m_1), B(\lambda_2, m_2), \ldots, B(\lambda_k, m_k)),$$

where $\lambda_1, \ldots, \lambda_k$ are eigenvalues of A and $m_1 + \cdots + m_k = n$. If eigenvalues appear in magnitude order, then Jordan blocks with equal diagonal entries will be adjacent.

Zeros can appear on the super-diagonal of J, because adjacent Jordan blocks join on the super-diagonal with a zero. A complete specification of how to build J from A is done in *generalized eigenanalysis*.

Geometric and algebraic multiplicity. The **geometric multiplicity** $\textbf{GeoMult}(\lambda)$ is the nullity of $A - \lambda I$, which is the number of basis vectors in a solution to $(A - \lambda I)\mathbf{x} = \mathbf{0}$, or, equivalently, the number of free variables. The **algebraic multiplicity** $\textsf{\textbf{AlgMult}}(\lambda)$ is the largest integer k such that $(r - \lambda)^k$ divides the characteristic polynomial $\det(A - rI)$.

Theorem 20 (Algebraic and Geometric Multiplicity)
Let A be a square real or complex matrix. Then

$$1 \le \textsf{\textbf{GeoMult}}(\lambda) \le \textsf{\textbf{AlgMult}}(\lambda). \tag{1}$$

In addition, there are the following relationships between the Jordan form J and algebraic and geometric multiplicities.

$\textsf{\textbf{GeoMult}}(\lambda)$	Equals the number of Jordan blocks $B(\lambda, m)$ that appear in J,
$\textsf{\textbf{AlgMult}}(\lambda)$	Equals the number of times λ is repeated along the diagonal of J.

Decoding the equation $A = PJP^{-1}$. The relation $A = PJP^{-1}$, equivalent to $AP = PJ$, can be expressed in terms of the columns of the matrix P. If J is a single Jordan block $B(\lambda, m)$, then the columns $\mathbf{v}_1$, ..., $\mathbf{v}_m$ of P satisfy

$$\begin{array}{rcl} A\mathbf{v}_1 & = & \lambda\mathbf{v}_1, \\ A\mathbf{v}_2 & = & \lambda\mathbf{v}_2 + \mathbf{v}_1, \\ \vdots & \vdots & \vdots \\ A\mathbf{v}_m & = & \lambda\mathbf{v}_m + \mathbf{v}_{m-1}. \end{array}$$

Chains of generalized eigenvectors. Given an eigenvalue λ of the matrix A, the topic of generalized eigenanalysis determines a Jordan block $B(\lambda, m)$ in J by finding an m**-chain** of generalized eigenvectors $\mathbf{v}_1$, ..., $\mathbf{v}_m$, which appear as columns of P in the relation $A = PJP^{-1}$. The very first vector $\mathbf{v}_1$ of the chain is an eigenvector, $(A - \lambda I)\mathbf{v}_1 = \mathbf{0}$. The others $\mathbf{v}_2$, ..., $\mathbf{v}_k$ are not eigenvectors but satisfy

$$(A - \lambda I)\mathbf{v}_2 = \mathbf{v}_1, \quad \ldots \quad , \quad (A - \lambda I)\mathbf{v}_m = \mathbf{v}_{m-1}.$$

Implied by the term m**-chain** is insolvability of $(A - \lambda I)\mathbf{x} = \mathbf{v}_m$. The chain size m is subject to the inequality $1 \le m \le \textsf{\textbf{AlgMult}}(\lambda)$.

The Jordan form J may contain several Jordan blocks for one eigenvalue λ. To illustrate, if J has only one eigenvalue λ and $\textsf{\textbf{AlgMult}}(\lambda) = 3$, then J might be constructed as follows:

$$J = \mathbf{diag}(B(\lambda,1), B(\lambda,1), B(\lambda,1)) \quad \text{or} \quad J = \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{pmatrix},$$

$$J = \mathbf{diag}(B(\lambda,1), B(\lambda,2)) \quad \text{or} \quad J = \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix},$$

$$J = B(\lambda,3) \quad \text{or} \quad J = \begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix}.$$

The three generalized eigenvectors for this example correspond to

$$J = \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{pmatrix} \quad \leftrightarrow \quad \text{Three 1-chains,}$$

$$J = \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix} \quad \leftrightarrow \quad \text{One 1-chain and one 2-chain,}$$

$$J = \begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix} \quad \leftrightarrow \quad \text{One 3-chain.}$$

Computing m-chains. Let us fix the discussion to an eigenvalue λ of A. Define $N = A - \lambda I$ and $p = \mathbf{AlgMult}(\lambda)$.

To compute an m-chain, start with an eigenvector $\mathbf{v}_1$ and solve recursively by `rref` methods $N\mathbf{v}_{j+1} = \mathbf{v}_j$ until there fails to be a solution. This must seemingly be done for *all possible choices* of $\mathbf{v}_1$! The search for m-chains terminates when p independent generalized eigenvectors have been calculated.

If A has an essentially unique eigenpair $(\lambda, \mathbf{v}_1)$, then this process terminates immediately with an m-chain where $m = p$. The chain produces one Jordan block $B(\lambda, m)$ and the generalized eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_m$ are recorded into the matrix P.

If $\mathbf{u}_1$, $\mathbf{u}_2$ form a basis for the eigenvectors of A corresponding to λ, then the problem $N\mathbf{x} = \mathbf{0}$ has 2 free variables. Therefore, we seek to find an m_1-chain and an m_2-chain such that $m_1 + m_2 = p$, corresponding to two Jordan blocks $B(\lambda, m_1)$ and $B(\lambda, m_2)$.

To understand the logic applied here, the reader should verify that for $\mathcal{N} = \mathbf{diag}(B(0, m_1), B(0, m_2), \ldots, B(0, m_k))$ the problem $\mathcal{N}\mathbf{x} = \mathbf{0}$ has k free variables, because $\mathcal{N}$ is already in `rref` form. These remarks imply that a k-dimensional basis of eigenvectors of A for eigenvalue λ

causes a search for m_i-chains, $1 \le i \le k$, such that $m_1 + \cdots + m_k = p$, corresponding to k Jordan blocks $B(\lambda, m_1), \ldots, B(\lambda, m_k)$.

A common naive approach for computing generalized eigenvectors can be illustrated by letting

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{u}_1 = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}, \quad \mathbf{u}_2 = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}.$$

Matrix A has one eigenvalue $\lambda = 1$ and two eigenpairs $(1, \mathbf{u}_1)$, $(1, \mathbf{u}_2)$. Starting a chain calculation with $\mathbf{v}_1$ equal to either $\mathbf{u}_1$ or $\mathbf{u}_2$ gives a 1-chain. This naive approach leads to only two independent generalized eigenvectors. However, the calculation must proceed until three independent generalized eigenvectors have been computed. To resolve the trouble, keep a 1-chain, say the one generated by $\mathbf{u}_1$, and start a new chain calculation using $\mathbf{v}_1 = a_1\mathbf{u}_1 + a_2\mathbf{u}_2$. Adjust the values of a_1, a_2 until a 2-chain has been computed:

$$\mathbf{aug}(A - \lambda I, \mathbf{v}_1) = \begin{pmatrix} 0 & 1 & 1 & a_1 \\ 0 & 0 & 0 & -a_1 + a_2 \\ 0 & 0 & 0 & a_1 - a_2 \end{pmatrix} \approx \begin{pmatrix} 0 & 1 & 1 & a_1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

provided $a_1 - a_2 = 0$. Choose $a_1 = a_2 = 1$ to make $\mathbf{v}_1 = \mathbf{u}_1 + \mathbf{u}_2 \neq \mathbf{0}$ and solve for $\mathbf{v}_2 = \begin{pmatrix} 0, 1, 0 \end{pmatrix}$. Then $\mathbf{u}_1$ is a 1-chain and $\mathbf{v}_1$, $\mathbf{v}_2$ is a 2-chain. The generalized eigenvectors $\mathbf{u}_1$, $\mathbf{v}_1$, $\mathbf{v}_2$ are independent and form the columns of P while $J = \mathbf{diag}(B(\lambda, 1), B(\lambda, 2))$ (recall $\lambda = 1$). We justify $A = PJP^{-1}$ by testing $AP = PJ$, using the formulas

$$J = \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix}, \quad P = \begin{pmatrix} 1 & 1 & 0 \\ -1 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}.$$

Theorem 21 (Exponential of a Jordan Block Matrix)
If λ is real and

$$B = \begin{pmatrix} \lambda & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda & 1 \\ 0 & 0 & 0 & \cdots & 0 & \lambda \end{pmatrix} \quad (m \times m \text{ matrix})$$

then

$$e^{Bt} = e^{\lambda t} \begin{pmatrix} 1 & t & \frac{t^2}{2} & \cdots & \frac{t^{m-2}}{(m-2)!} & \frac{t^{m-1}}{(m-1)!} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & t \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{pmatrix}.$$

The equality also holds if λ is a complex number, in which case both sides of the equation are complex.

The Real Jordan Form of A

Given a real matrix A, generalized eigenanalysis seeks to find a *real* invertible matrix $\mathcal{P}$ and a *real* upper triangular block matrix R such that $A = \mathcal{P}R\mathcal{P}^{-1}$. This requirement leads to a *real* equation for e^{At}, appropriate if A itself is real.

If λ is a real eigenvalue of A, then a **real Jordan block** is a matrix

$$B = \mathbf{diag}(\lambda, \ldots, \lambda) + N, \quad N = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}.$$

If $\lambda = a + ib$ is a complex eigenvalue of A, then symbols λ, 1 and 0 are replaced respectively by 2×2 real matrices $\Lambda = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}$, $\mathcal{I} = \mathbf{diag}(1,1)$ and $\mathcal{O} = \mathbf{diag}(0,0)$. The corresponding $2m\times 2m$ real Jordan block matrix is given by the formula

$$B = \mathbf{diag}(\Lambda, \ldots, \Lambda) + \mathcal{N}, \quad \mathcal{N} = \begin{pmatrix} \mathcal{O} & \mathcal{I} & \mathcal{O} & \cdots & \mathcal{O} & \mathcal{O} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathcal{O} & \mathcal{O} & \mathcal{O} & \cdots & \mathcal{O} & \mathcal{I} \\ \mathcal{O} & \mathcal{O} & \mathcal{O} & \cdots & \mathcal{O} & \mathcal{O} \end{pmatrix}.$$

Nilpotent matrices. The matrix N satisfies $N^m = 0$. Similarly, $\mathcal{N}^m = 0$. Such matrices are called **nilpotent block matrices**. The least integer m for which $N^m = 0$ is called the **nilpotency** of N. A nilpotent matrix N has a finite exponential series:

$$e^{Nt} = I + Nt + N^2\frac{t^2}{2!} + \cdots + N^{m-1}\frac{t^{m-1}}{(m-1)!}.$$

Computing $\mathcal{P}$ and R. Generalized eigenvectors for a real eigenvalue λ are placed into the matrix $\mathcal{P}$ in the same order as specified in R by the corresponding real Jordan block. In the case when $\lambda = a+ib$ is complex, $b > 0$, the real and imaginary parts of each generalized eigenvector are entered pairwise into $\mathcal{P}$; the conjugate eigenvalue $\overline{\lambda} = a - ib$ is skipped. The result is a *real* matrix $\mathcal{P}$ and a *real* upper triangular block matrix R which satisfy $A = \mathcal{P}R\mathcal{P}^{-1}$.

Theorem 22 (Real Block Diagonal Matrix, Eigenvalue $a + ib$)

Let $\Lambda = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}$, $\mathcal{I} = \mathbf{diag}(1,1)$ and $\mathcal{O} = \mathbf{diag}(0,0)$. Consider a real

Jordan block matrix of dimension $2m \times 2m$ given by the formula

$$B = \begin{pmatrix} \Lambda & \mathcal{I} & \mathcal{O} & \cdots & \mathcal{O} & \mathcal{O} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathcal{O} & \mathcal{O} & \mathcal{O} & \cdots & \Lambda & \mathcal{I} \\ \mathcal{O} & \mathcal{O} & \mathcal{O} & \cdots & \mathcal{O} & \Lambda \end{pmatrix}.$$

If $\mathcal{R} = \begin{pmatrix} \cos bt & \sin bt \\ -\sin bt & \cos bt \end{pmatrix}$, then

$$e^{Bt} = e^{at} \begin{pmatrix} \mathcal{R} & t\mathcal{R} & \frac{t^2}{2}\mathcal{R} & \cdots & \frac{t^{m-2}}{(m-2)!}\mathcal{R} & \frac{t^{m-1}}{(m-1)!}\mathcal{R} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathcal{O} & \mathcal{O} & \mathcal{O} & \cdots & \mathcal{R} & t\mathcal{R} \\ \mathcal{O} & \mathcal{O} & \mathcal{O} & \cdots & \mathcal{O} & \mathcal{R} \end{pmatrix}.$$

Solving $\mathbf{x}' = A\mathbf{x}$. The solution $\mathbf{x}(t) = e^{At}\mathbf{x}(0)$ must be real if A is real. The real solution can be expressed as $\mathbf{x}(t) = \mathcal{P}\mathbf{y}(t)$ where $\mathbf{y}'(t) = R\mathbf{y}(t)$ and R is a real Jordan form of A, containing real Jordan blocks B down its diagonal. Theorems above provide explicit formulas for e^{Bt}, hence the resulting formula

$$\mathbf{x}(t) = \mathcal{P}e^{Rt}\mathcal{P}^{-1}\mathbf{x}(0)$$

contains only real numbers, real exponentials, plus sine and cosine terms, which are possibly multiplied by polynomials in t.