Linear Algebra: Linear Systems and Matrices - Quadratic Forms and Definiteness - Eigenvalues and Markov Chains

Joshua Wilde, revised by Isabel Tecu, Takeshi Suzuki and María José Boccardi

August 13, 2013

1 Systems of Linear Equations

Linear Algebra is concerned with the study of systems of linear equations. A system of m linear equations in n variables has the form

$$
\begin{aligned}
b_1 &= a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\
b_2 &= a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\
&\vdots \\
b_m &= a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n
\end{aligned}
$$

Linear equations are important since non-linear, differentiable functions can be approximated by linear ones (as we have seen). For example, the behavior of a differentiable function $f : \mathbb{R}^2 \to \mathbb{R}$ around a point x^* can be approximated by the tangent plane at x^*. The equation for the tangent plane is one linear equation in two variables. In Economics, we often encounter systems of equations. Often linear equations are used since they are tractable and since they can be thought of as approximations for more complicated underlying relationships between variables.

The system **??** can be written in matrix form:

$$
\underbrace{\begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}}_{m\times 1} = \underbrace{\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}}_{m\times n} \cdot \underbrace{\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}}_{n\times 1},
$$

In short, we can write this system as $\mathbf{b} = A\mathbf{x}$ where A is an $m \times n$ matrix, $\mathbf{b}$ is an $m \times 1$ vector and $\mathbf{x}$ is an $n \times 1$ vector. A system of linear equations , also referred to as linear map, can therefore be identified with a matrix, and any matrix can be identified with ("turned into") a linear system. In order to study linear systems, we study matrices and their properties.

2 Matrices

2.1 Basic Matrix Operations and Properties

Consider two $n \times m$ matrices:

$$
A = \begin{pmatrix} a_{11} & \dots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nm} \end{pmatrix}, \; B = \begin{pmatrix} b_{11} & \dots & b_{1m} \\ \vdots & \ddots & \vdots \\ b_{n1} & \dots & b_{nm} \end{pmatrix}
$$

Then the basic matrix operations are as follows:

1. $A + B = \begin{pmatrix} a_{11} + b_{11} & \dots & a_{1m} + b_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} + b_{n1} & \dots & a_{nm} + b_{nm} \end{pmatrix}$

2. $\lambda A = \begin{pmatrix} \lambda a_{11} & \dots & \lambda a_{1m} \\ \vdots & \ddots & \vdots \\ \lambda a_{n1} & \dots & \lambda a_{nm} \end{pmatrix}$, where $\lambda \in \mathbb{R}$

Notice that the elements in the matrix are numbered a_{ij}, where i is the row and j is the column in which the element a_{ij} is found.

In order to multiply matrices CD, the number of columns in the C matrix must be equal to the number of rows in the D matrix. Say C is an $n \times m$ matrix, and D is an $m \times k$ matrix. Then multiplication is defined as follows:

$$E = \underbrace{C}_{n \times m} \underbrace{D}_{m \times k} = \underbrace{\begin{pmatrix} \sum_{q=1}^{m} c_{1,q} d_{q,1} & \dots & \sum_{q=1}^{m} c_{1,q} d_{q,k} \\ \vdots & \ddots & \vdots \\ \sum_{q=1}^{m} c_{n,q} d_{q,1} & \dots & \sum_{q=1}^{m} c_{n,q} d_{q,k} \end{pmatrix}}_{n \times k}$$

There are two notable special cases for multiplication of matrices. The first is called the inner product or dot product, which occurs when two vectors of the same length are multiplied together such that the result is a scalar:

$$v \cdot z = \underbrace{v}_{1 \times n} \underbrace{z'}_{n \times 1} = \begin{pmatrix} v_1 & \dots & v_n \end{pmatrix} \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix} = \sum_{i=1}^{n} v_i z_i$$

The second is called the outer product:

$$\underbrace{v'}_{n \times 1} \underbrace{z}_{1 \times n} = \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix} \begin{pmatrix} v_1 & \dots & v_n \end{pmatrix} = \underbrace{\begin{pmatrix} z_1 v_1 & \dots & z_1 v_n \\ \vdots & \ddots & \vdots \\ z_n v_1 & \dots & z_n v_n \end{pmatrix}}_{n \times n}.$$

Note that when we multiplied the matrices C and D together, the resulting $e_{i,j}$th element of E was just the inner product of the ith row of C and jth column of D. Also, note that even if two matrices X and Y are both $n \times n$, then $XY \neq YX$, except in special cases.

Just in case you are wondering why matrix multiplication is defined the way it is: Consider two linear maps (that is, two systems of linear equations) $f(\mathbf{x}) = A\mathbf{x}$ and $g(\mathbf{x}) = B\mathbf{x}$ where A is $m \times n$ and B is $n \times k$. We can define a new linear map h that is the composition of f after g: $h(\mathbf{x}) = f(g(\mathbf{x}))$. Then the matrix that represents the linear system h turns out to be exactly AB, that is

$$h(\mathbf{x}) = f(g(\mathbf{x})) = AB\mathbf{x}.$$

Matrix multiplication is defined to correspond to the composition of linear maps.

Definition. A mapping f of a vector space X into a vector space Y is said to be a *linear mapping* if for any vectors $\mathbf{x_1}, \dots, \mathbf{x_m} \in X$ and any scalars $c_1, \dots, c_m$,

$$f(c_1 \mathbf{x_1} + \dots + c_m \mathbf{x_m}) = c_1 f(\mathbf{x_1}) + \dots + c_m f(\mathbf{x_m}).$$

2.2 Some Special Matrices

2.2.1 Zero Matrices

A zero matrix is a matrix where each element is 0

$$\mathbf{0} = \underbrace{\begin{pmatrix} 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{pmatrix}}_{n \times k}$$

The following properties hold for zero matrices:

1. $A + \mathbf{0} = A$
2. If $AB = \mathbf{0}$, it is not necessarily the case that $A = \mathbf{0}$ or $B = \mathbf{0}$.

2.2.2 Identity Matrices

The identity matrix is a matrix with zeroes everywhere except along the diagonal. Note that the number of columns must equal the number of rows.

$$I = \underbrace{\begin{pmatrix} 1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 1 \end{pmatrix}}_{n \times n}$$

The reason it is called the identity matrix is because $AI = IA = A$.

2.2.3 Square, Symmetric, and Transpose Matrices

A square matrix is a matrix whose number of rows is the same as its number of columns. For example, the identity matrix is always square.

If a square matrix has the property that $a_{i,j} = a_{j,i}$ for all its elements, then we call it a symmetric matrix.

The transpose of a matrix A, denoted A' is a matrix such that for each element of A', $a'_{i,j} = a_{j,i}$. For example, the transpose of the matrix

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$$

is

$$\begin{pmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{pmatrix}.$$

Note that a matrix A is symmetric if $A = A'$.

The following properties of the transpose hold:

1. $(A')' = A$.
2. $(A + B)' = A' + B'$.
3. $(\alpha A)' = \alpha A'$.

4. $(AB)' = B'A'$.

5. If the matrix A is $n \times k$, then A' is $k \times n$.

2.2.4 Diagonal and Triangular Matrices

A square matrix A is diagonal if it is all zeroes except along the diagonal:

$$\begin{pmatrix} a_{11} & 0 & \dots & 0 \\ 0 & a_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{nn} \end{pmatrix}$$

Note that all diagonal matrices are also symmetric.

A square matrix A is upper triangular if all of its entries below the diagonal are zero

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ 0 & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{nn} \end{pmatrix},$$

and is lower triangular if all its entries above the diagonal are zero

$$\begin{pmatrix} a_{11} & 0 & \dots & 0 \\ a_{21} & a_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}.$$

2.2.5 Inverse Matrices

If there exists a matrix B such that $BA = AB = I$, then we call B the inverse of A, and denote it A^{-1}. Note that A can only have an inverse if it is a square matrix. However, not every square matrix has an inverse. The following properties of inverses hold:

1. $\left(A^{-1}\right)^{-1} = A$

2. $(\alpha A)^{-1} = \frac{1}{\alpha} A^{-1}$

3. $(AB)^{-1} = B^{-1}A^{-1}$ if B^{-1}, A^{-1} exist.

4. $(A')^{-1} = \left(A^{-1}\right)'$

2.2.6 Orthogonal and Idempotent Matrices

A matrix A is orthogonal if $A'A = I$ (which also implies $AA' = I$). In other words, a matrix is orthogonal if it is its own inverse.

A matrix is idempotent if it is both symmetric and $AA = A$.

Orthogonal and idempotent matrices are especially used in econometrics.

2.3 The Determinant

For square matrices we can define a number called the determinant of the matrix. The determinant tells us important characteristics of the matrix that we will dwell on later. Here we will simply present how it is computed.

The determinant can be defined inductively:

1. The determinant of a 1×1 matrix (a) is a, and is denoted $\det(a)$.

2. The determinant of a 2×2 matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is $ad - bc$. Notice that this is the same as $a \cdot \det(d) - b \cdot \det(c)$. The first term is the (1,1)th entry of A times the determinant of that submatrix obtained by deleting from A the row and column which contain that entry. The second term is (1,2)th entry times the determinant of the submatrix obtained by deleting A from the row and column which contain that entry. The terms alternate in sign, with the first term being added and the second term being subtracted.

3. The determinant of a 3×3 matrix $\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}$ is $aei + bfg + cdh - ceg - bdi - afh$. Notice that this can be written as $a \cdot \det\begin{pmatrix} e & f \\ h & i \end{pmatrix} - b \cdot \det\begin{pmatrix} d & f \\ g & i \end{pmatrix} + c \cdot \det\begin{pmatrix} d & e \\ g & h \end{pmatrix}$.

Can you see the pattern? In order to obtain the determinant, we multiply each element in the top row with the determinant of the matrix left when we delete the row and column in which the respective elements reside. The signs of the terms alternate, starting with positive.

In order to write the definition of the determinant of an nth order matrix, it is useful to define the (i, j)th **minor** of A and the (i, j)th **cofactor** of A:

- Let A be an $n \times n$ matrix. Let $A_{i,j}$ be the $(n-1) \times (n-1)$ submatrix obtained by deleting row i and column j from A. Then the scalar $M_{ij} = \det(A_{ij})$ is called the (i, j)th **minor** of A.

- The scalar $C_{ij} = (-1)^{i+j} M_{ij}$ is called the (i, j)th **cofactor** of A. The cofactor is merely the signed minor.

Armed with these two definitions, we notice that the determinant for the 2×2 matrix is

$$\det(A) = aM_{11} - bM_{12} = aC_{11} + bC_{12},$$

and the determinant for the 3×3 matrix is

$$\det(A) = aM_{11} - bM_{12} - cM_{13} = aC_{11} + bC_{12} + C_{13}.$$

Therefore, we can define the determinant for an $n \times n$ sqaure matrix as follows:

$$\det\left(\underbrace{A}_{n\times n}\right) = a_{11}C_{11} + a_{12}C_{12} + \cdots + a_{1n}C_{1n}.$$

Notice that the definition of the determinant uses elements and cofactors for the top row only. This is called a **cofactor expansion along the first row**. However, a cofactor expansion along any row or column will be equal to the determinant. The proof of this asssertion is left as a homework problem for the 3×3 case.

Example: Find the determinant of the upper diagonal matrix

$$\begin{pmatrix} 1 & 0 & 0 \\ 2 & 3 & 0 \\ 4 & 5 & 6 \end{pmatrix}$$

The determinant is:

$$aC_{11} + bC_{12} + C_{13} = a \cdot \det \begin{pmatrix} e & f \\ h & i \end{pmatrix} - b \cdot \det \begin{pmatrix} d & f \\ g & i \end{pmatrix} + c \cdot \det \begin{pmatrix} d & e \\ g & h \end{pmatrix} =$$

$$= 1 \cdot \det \begin{pmatrix} 3 & 0 \\ 5 & 6 \end{pmatrix} - 0 \cdot \det \begin{pmatrix} 2 & 0 \\ 4 & 6 \end{pmatrix} + 0 \cdot \det \begin{pmatrix} 2 & 3 \\ 4 & 5 \end{pmatrix} = 1 \cdot 3 \cdot 6 = 18$$

Now lets expand along the second column instead of the third row:

$$aC_{12} + bC_{22} + C_{32} = -b \cdot \det \begin{pmatrix} d & f \\ g & i \end{pmatrix} + e \cdot \det \begin{pmatrix} a & c \\ g & i \end{pmatrix} - h \cdot \det \begin{pmatrix} a & c \\ d & f \end{pmatrix} =$$

$$= -0 \cdot \det \begin{pmatrix} 2 & 0 \\ 4 & 6 \end{pmatrix} + 3 \cdot \det \begin{pmatrix} 1 & 0 \\ 4 & 6 \end{pmatrix} - 5 \cdot \det \begin{pmatrix} 1 & 0 \\ 2 & 0 \end{pmatrix} = 3 \cdot 6 = 18$$

Important properties of the determinant:

- $\det(I) = 1$ where I is the identity matrix
- $\det(AB) = \det(A)\det(B)$
- If A is invertible, $\det(A^{-1}) = \frac{1}{\det(A)}$
- $\det(A') = \det(A)$
- A is orthogonal if and only if $|\det A| = 1$

Definition The following definition will be important subsequently: An $n \times n$ matrix A is called *singular* if $\det A = 0$. It is called *non-singular* if $\det A \neq 0$.

3 Solving Systems of Linear Equations

Recall the system of equations:

$$\begin{aligned} b_1 &= a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ b_2 &= a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ &\vdots \\ b_m &= a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{aligned}$$

which can be written in matrix form by

$$\underbrace{\begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}}_{m\times 1} = \underbrace{\begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix}}_{m\times n} \cdot \underbrace{\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}}_{n\times 1},$$

or $\mathbf{b} = A\mathbf{x}$. The questions we want to answer are:

- Given a left hand side vector $\mathbf{b}$, how can we find a solution $\mathbf{x}$ to $\mathbf{b} = A\mathbf{x}$?
- Given a coefficient matrix A, what can we say about the number of solutions to $\mathbf{b} = A\mathbf{x}$, for any $\mathbf{b}$? How can we tell whether a system has one unique solution?

Let's start with the first question.

3.1 Elementary Row Operations

There are three types of elementary row operations we can perform on the matrix A and the vector b without changing the solution set to $\mathbf{b} = A\mathbf{x}$:

1. Interchanging two rows
2. Multiplying each element of a row by the same non-zero scalar.
3. Change a row by adding it to a multiple of another row.

You should convince yourself that these three operations do not change the solution set of the system.

3.2 Using Elementary Row Operations to Solve a System of Equations

A row of a matrix is said to have k **leading zeros** if the $(k+1)$th element of the row is non zero while the first k elements are zero. A matrix is in **row echelon form** if each row has strictly more leading zeros that the preceding row. For example, the matrices

$$A = \begin{pmatrix} 4 & 2 \\ 0 & 7 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 0 & 10 \\ 0 & 8 & -3 \\ 0 & 0 & 6 \end{pmatrix}$$

are in row echelon form while the matrices

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 6 & 7 \\ 0 & 0 & 8 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 1 \\ 2 & 4 \end{pmatrix}$$

are not. However, B would be in row echelon form if we performed the elementary matrix operation of switching the two rows.

In order to get the matrix into row echelon form, we can perform elementary matrix operations on the rows. For example, consider the matrix

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 8 & 6 \\ 1 & 1 & 1 \end{pmatrix}.$$

By taking the third row and subtracting it from the first row, we obtain the matrix

$$A_1 = \begin{pmatrix} 0 & 1 & 2 \\ 4 & 8 & 6 \\ 1 & 1 & 1 \end{pmatrix}$$

We can also subtract four times the third row from the second row

$$A_2 = \begin{pmatrix} 0 & 1 & 2 \\ 0 & 4 & 2 \\ 1 & 1 & 1 \end{pmatrix}$$

Now subtract four times the first row from the second row to obtain

$$A_3 = \begin{pmatrix} 0 & 1 & 2 \\ 0 & 0 & -6 \\ 1 & 1 & 1 \end{pmatrix}$$

Then rearrange to get the matrix in row echelon form

$$A_4 = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & -6 \end{pmatrix}$$

We can solve a system of equations by writing the matrix

$$\begin{pmatrix} a_1 & b_1 & \dots & c_1 & | & b_1 \\ a_2 & b_2 & \dots & c_2 & | & b_2 \\ \vdots & \vdots & \ddots & \vdots & | & \vdots \\ a_m & b_m & \dots & c_m & | & b_m \end{pmatrix},$$

called the **augmented matrix** of A, and use elementary row operations.
Example:
Let

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 8 & 6 \\ 1 & 1 & 1 \end{pmatrix}$$

as before. Say that the vector $\mathbf{b} = (1, 1, 1)'$. Then the augmented matrix is

$$\begin{pmatrix} 1 & 2 & 3 & | & 1 \\ 4 & 8 & 6 & | & 1 \\ 1 & 1 & 1 & | & 1 \end{pmatrix}$$

Performing the same matrix operations as before, we have

$$\begin{pmatrix} 1 & 2 & 3 & | & 1 \\ 4 & 8 & 6 & | & 1 \\ 1 & 1 & 1 & | & 1 \end{pmatrix} \Rightarrow \begin{pmatrix} 0 & 1 & 2 & | & 0 \\ 4 & 8 & 6 & | & 1 \\ 1 & 1 & 1 & | & 1 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 1 & 2 & | & 0 \\ 0 & 4 & 2 & | & -3 \\ 1 & 1 & 1 & | & 1 \end{pmatrix} \Rightarrow \begin{pmatrix} 0 & 1 & 2 & | & 0 \\ 0 & 0 & -6 & | & -3 \\ 1 & 1 & 1 & | & 1 \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & 1 & 1 & | & 1 \\ 0 & 1 & 2 & | & 0 \\ 0 & 0 & -6 & | & -3 \end{pmatrix}$$

We continue the row operations until the left hand side of the augmented matrix looks like the identity matrix:

$$\Rightarrow \begin{pmatrix} 1 & 1 & 1 & | & 1 \\ 0 & 1 & 2 & | & 0 \\ 0 & 0 & 1 & | & \frac{1}{2} \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & 1 & 1 & | & 1 \\ 0 & 1 & 0 & | & -1 \\ 0 & 0 & 1 & | & \frac{1}{2} \end{pmatrix}$$

$$\begin{pmatrix} 1 & 1 & 0 & | & \frac{1}{2} \\ 0 & 1 & 0 & | & -1 \\ 0 & 0 & 1 & | & \frac{1}{2} \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & 0 & 0 & | & \frac{3}{2} \\ 0 & 1 & 0 & | & -1 \\ 0 & 0 & 1 & | & \frac{1}{2} \end{pmatrix}$$

Notice that this implies

$$\begin{pmatrix} \frac{3}{2} \\ -1 \\ \frac{1}{2} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \Rightarrow \mathbf{x} = \left(\frac{3}{2}, -1, \frac{1}{2}\right),$$

so we have found a solution using elementary row operations.

In summary, if we form the augmented matrix of A, reduce the left hand side of the matrix to its **reduced row echelon form** (so that each row contains all zeros, except for the possibility of a one in a column of all zeros) through elementary row operations, then the remaining vector on the right hand side will be the solution to the system.

3.3 Using Cramer's Rule to Solve a System of Equations

Cramer's Rule is a theorem that yields the solutions to systems of the form $\mathbf{b} = A\mathbf{x}$ where A is a square matrix and non-singular, i.e. $\det A \neq 0$: Let A be a non-singular matrix. Then the unique solution $\mathbf{x} = (x_1, \ldots, x_n)$ of the $n \times n$ system $\mathbf{b} = A\mathbf{x}$ is:

$$x_i = \frac{\det(B_i)}{\det(A)} \text{ for } i = 1, \ldots, n,$$

where B_i is the matrix A with the right hand side $\mathbf{b}$ replacing the ith column of A.

Example:

Consider the linear IS-LM model

$$sY + ar = I^0 + G$$

$$mY - hr = M_s - M^0$$

where Y is the net national product, r is the interest rate, s is the marginal propensity to save, a is the marginal efficiency of capital, $I = I^0 - ar$ is investment, m is money balances needed per dollar of transactions, G is government spending, and M_s is the money supply. All the parameters are positive. We can rewrite the system as

$$\begin{pmatrix} I^0 + G \\ M_s - M^0 \end{pmatrix} = \begin{pmatrix} s & a \\ m & -h \end{pmatrix} \cdot \begin{pmatrix} Y \\ r \end{pmatrix}.$$

By Cramer's rule, we have

$$Y = \frac{\begin{vmatrix} I^0 + G & a \\ M_s - M^0 & -h \end{vmatrix}}{\begin{vmatrix} s & a \\ m & -h \end{vmatrix}} = \frac{(I^0 + G)h + a(M_s - M^0)}{sh + am}$$

$$r = \frac{\begin{vmatrix} s & I^0 + G \\ m & M_s - M^0 \end{vmatrix}}{\begin{vmatrix} s & a \\ m & -h \end{vmatrix}} = \frac{(I^0 + G)m - s(M_s - M^0)}{sh + am}$$

Depending on the size of A, solving a system using Cramer's Rule may be faster than solving it using elementary row operations. But be aware that *Cramer's Rule only works for non-singular square matrices.*

4 Characterization of the Solutions to a Linear System

The second question about linear system concerns the existence of solutions: How can we tell whether a system has zero, one or more solutions? In order to tackle this problem, we start by defining the concept of "linear independence".

4.1 Linear Independence

The vectors $v_1, \ldots, v_m$ are **linearly dependent** if there exist scalars $q_1, \ldots, q_m$, not all zero, such that:

$$\sum_{i=1}^{m} q_i v_i = 0.$$

This is equivalent to the statement that one of the vectors can be written as a linear combination of the other ones.

The vectors $v_1, \ldots, v_m$ are **linearly independent** if the only scalars $q_1, \ldots, q_m$ such that:

$$\sum_{i=1}^{m} q_i v_i = 0$$

are $q_1 = \cdots = q_m = 0$. This is equivalent to the statement that none of the vectors can be written as a linear combination of the other ones.

Example 1: The vectors $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 3 \\ 2 \\ 0 \end{pmatrix}$ are linearly dependent since

$$3 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + 2 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + (-1) \begin{pmatrix} 3 \\ 2 \\ 0 \end{pmatrix} = 0$$

Example 2: The vectors $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 3 \\ 0 \\ 2 \end{pmatrix}$ are linearly independent since the only scalars q_1, q_2, q_3 such that

$$q_1 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + q_2 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + q_3 \begin{pmatrix} 3 \\ 2 \\ 0 \end{pmatrix} = 0$$

are $q_1 = q_2 = q_3 = 0$.

We can use this definition of linear independence and dependence for columns as well.

The **rank** of a matrix, $\mathrm{rk}(A)$, is the number of linearly independent rows or columns in a matrix. (Note that the number of linearly independent rows is the same as the number of linearly independent columns). Therefore $\mathrm{rk}(A) \leq \min\{\text{number of rows of A},\ \text{number of columns of A}\}$. A matrix is said to have full rank if $\mathrm{rk}(A) = \min\{\text{number of rows of A},\ \text{number of columns of A}\}$.

When a matrix is in row echelon form, it is easy to check whether all the rows are linearly independent. For linear independence, we must have that all the rows in the row echelon form are non-zero. If not, then the matrix will have linearly dependant rows. For example, in the matrix

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 8 & 6 \\ 1 & 1 & 1 \end{pmatrix}$$

in the example above, all rows are linearly independent because its row echelon form

$$A_4 = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & -6 \end{pmatrix}$$

contains no zero rows.

4.2 Numbers of Solutions

Let's build some intuition by looking at a system of two variables and two equations. For a given $\mathbf{b} = (b_1, b_2)$, we can view the two equations $y_1 = a_1x_1 + b_1x_2$ and $y_1 = a_2x_1 + b_2x_2$ as two lines in $\mathbb{R}^2$. A solution $\mathbf{x} = (x_1, x_2)$ is a point that lies on both of these lines at the same time. Now, two lines can intersect once, be parallel to each other or be identical to each other. In the first case, there will be one solution (one point of intersection), in the second case there will be no solution (no point of intersection)and in the third case there will be infinitely many solutions (infinitely many points of intersection). Therefore, a system of two equations and two unknowns can have zero, one or infinitely many solution, depending on the vector $\mathbf{b}$ and the matrix A. This result generalizes to a linear system of any size: It can have either zero, one or infinitely many solutions.

Fortunately we can say a bit more about the possible numbers of solutions when we look at the dimension of a system and the rank of its coefficient matrix A. Remember than m is the number of equations in the system and the number of rows of A and that n is the number of variables in the system and the number of columns of A.

- If $m < n$, then
  1. for any given $\mathbf{b}$, $\mathbf{b} = A\mathbf{x}$ has zero or infinitely many solutions
  2. if $\text{rk}A = m = \text{rk}(A|\mathbf{b})$, $\mathbf{b} = A\mathbf{x}$ has infinitely many solutions for every $\mathbf{b}$

  A system with more unknowns than equation can never have one unique solution. (Example: Two planes in $\mathbb{R}^3$ cannot intersect in only one point. They are either parallel (no solution) or intersect in a line (infinitely many solutions)). If we know that A has maximal rank, namely m, then we know that the system has infinitely many solutions. (Continuing the example above, we know then that the planes are not parallel and therefore have to intersect in at least a line. We also know then that the planes are not all identical, but that does not help us in narrowing down the number of solutions in this case.)

- If $m > n$, then
  1. for any given $\mathbf{b}$, $\mathbf{b} = A\mathbf{x}$ has zero, one or infinitely many solutions
  2. if $\text{rk}A = n$, $\mathbf{b} = A\mathbf{x}$ has zero or one solution for every $\mathbf{b}$

  A system with more equations than unknowns may have zero, one or infinitely many solutions. (Example: Three lines in $\mathbb{R}^2$ can either not all intersect in the same point (no solution), two of them can be identical and intersect with the third one in one point (one solution), or all three of them can be identical (infinitely many solutions). If we know that A has maximal rank, namely n, then we know that the system cannot have infinitely many solutions. (Continuing the example above, we know then that the lines are not all identical and therefore cannot intersect in infinitely many points. We also know then that the lines are not all parallel, but that does not narrow down the number of solutions in this case.)

- If $m = n$, then
  1. for any given $\mathbf{b}$, $\mathbf{b} = A\mathbf{x}$ has zero, one or infinitely many solutions

2. if $\text{rk}A = m = n$, $\mathbf{b} = A\mathbf{x}$ has exactly one solution for every $\mathbf{b}$

A system with as many equations as unknowns may have zero, one or infinitely many solutions. (Example: Two lines in $\mathbb{R}^2$ can be parallel (zero solutions), intersect in one point (one solution), or be identical (infinitely many solutions). Same for three planes in $\mathbb{R}^3$.) If we know that A has maximal rank, namely $n = m$, then we know that the system has exactly one solution. (Continuing the example above, we know then that the lines are not identical and therefore cannot intersect in infinitely many points. We also know then that the lines are not all parallel and therefore have to intersect in at least one point. Two lines that are neither parallel nor identical intersect in exactly one point.)

Whether a square matrix has full rank can be checked by looking at the determinant: *An $n \times n$ square matrix A has full rank n if and only if* $\det A \neq 0$. Therefore the last statement above translates into: *The $n \times n$ system* $\mathbf{b} = A\mathbf{x}$ *has exactly one solution for each* $\mathbf{b}$ *if and only if* $\det A \neq 0$. This is an extremely important and useful property of the determinant. So, saying that a square matrix A is non-singular is the same as saying that $\text{rk}A = n$ is the same as saying $\mathbf{b} = A\mathbf{x}$ has exactly one solution for each $\mathbf{b}$. (Note also how this explains why Cramer's Rule fails when $\det A = 0$.)

4.3 Inverse Matrices

Suppose we are given an $n \times n$ matrix A. The question whether its inverse exists is connected to whether a unique solution to $\mathbf{b} = A\mathbf{x}$ exists for each choice of $\mathbf{b}$.

To see why, suppose the inverse A^{-1} exists. Then we can multiply both sides of $\mathbf{b} = A\mathbf{x}$ by A^{-1}:

$$A^{-1}\mathbf{b} = A^{-1}A\mathbf{x} = I\mathbf{x} = \mathbf{x}$$

Therefore $\mathbf{x} = A^{-1}\mathbf{b}$ has to be the unique solution to $\mathbf{b} = A\mathbf{x}$. Conversely, it can also be shown that if there is a unique solution to $\mathbf{b} = A\mathbf{x}$, then the inverse A^{-1} has to exist (take $\mathbf{b_1} = (1, 0, \ldots, 0)'$, $\mathbf{b_2} = (0, 1, \ldots, 0)'$, etc., and let the solution to $\mathbf{b_i} = Ax$ be $\mathbf{x_i}$. Then $A^{-1} = (\mathbf{x_1}, \mathbf{x_2}, \ldots \mathbf{x_n})$.)

Therefore, a square matrix A is invertible if and only if it is non-singular, that is, if and only if $\det A \neq 0$.

The following statements are all equivalent for a square matrix A:

1. A is non-singular.
2. All the columns and rows in A are linearly independent.
3. A has full rank.
4. Exactly one solution X^* exists for each vector Y^*.
5. A is invertible.
6. $\det(A) \neq 0$.
7. The row-echelon form of the matrix is upper triangular.
8. The reduced row echelon form is the identity matrix.

Now that we know how to check whether the inverse matrix exists (i.e. by looking at the determinant), how do we compute it? Here are two strategies:

4.3.1 Calculating the Inverse Matrix by Elementary Row Operations

To calculate the inverse matrix, form the augmented matrix

$$\left(\begin{array}{cccc|cccc} a_{11} & a_{12} & \dots & a_{1n} & 1 & 0 & \dots & 0 \\ a_{21} & a_{22} & \dots & a_{2n} & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} & 0 & 0 & \dots & 1 \end{array}\right),$$

where the left hand side is the matrix A and the right hand side is the identity matrix. Reduce the left hand side to the reduced row echelon form, and what remains on the right hand side will be the inverse matrix of A. In other words, by elementary row operations, you can transform the matrix $(A|I)$ to the matrix $(I|A^{-1})$.

4.3.2 Calculating the Inverse Matrix by the Adjoint Matrix

The **adjoint matrix** of a square matrix A is the transposed matrix of cofactors of A, or

$$\text{adj}(A) = \begin{pmatrix} C_{11} & C_{21} & \dots & C_{n1} \\ C_{12} & C_{22} & \dots & C_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ C_{1n} & C_{2n} & \dots & C_{nn} \end{pmatrix}$$

Notice that the adjoint of a 2×2 matrix is $\begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix}$.

The inverse of the matrix A can be found by

$$A^{-1} = \frac{1}{\det(A)} \cdot \text{adj}(A).$$

Therefore, the inverse of a 2×2 matrix is

$$A^{-1} = \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \cdot \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix}.$$

5 Homework

Do the following:

1. Let $A = \begin{pmatrix} 2 & 0 \\ 3 & 8 \end{pmatrix}$ and $B = \begin{pmatrix} 7 & 2 \\ 6 & 3 \end{pmatrix}$. Find $A - B$, $A + B$, AB, and BA.

2. Let $v = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ and $u = \begin{pmatrix} 3 \\ 5 \end{pmatrix}$. Find $u \cdot v$, $u'v$ and $v'u$.

3. Prove that the multiplication of any matrix with its transpose yields a symmetric matrix

4. Prove that A only has an inverse if it is a square matrix.

5. Prove the first four properties of transpose matrices above.

6. In econometrics, we deal with a matrix called the projections matrix: $A = I - X(X'X)^{-1}X'$. Must A be square? Must $X'X$ be square? Must X be square?

7. Show that the projection matrix in 6 is idempotent.

8. Calculate the determinants of the following matrices:

   (a) $\begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}$

   (b) $\begin{pmatrix} -3 & 4 \\ 4 & -5 \end{pmatrix}$

   (c) $\begin{pmatrix} -3 & 4 \\ 4 & -6 \end{pmatrix}$

   (d) $\begin{pmatrix} 2 & 4 \\ 4 & 8 \end{pmatrix}$

   (e) $\begin{pmatrix} 1 & 2 & 0 \\ 2 & 4 & 5 \\ 0 & 5 & 6 \end{pmatrix}$

   (f) $\begin{pmatrix} -1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & -2 \end{pmatrix}$

   (g) $\begin{pmatrix} 1 & 0 & 3 & 0 \\ 0 & 2 & 0 & 5 \\ 3 & 0 & 4 & 0 \\ 0 & 5 & 0 & 6 \end{pmatrix}$

9. Evaluate the following determinants:

   (a) $\begin{vmatrix} 1 & 1 & 4 \\ 8 & 11 & -2 \\ 0 & 4 & 7 \end{vmatrix}$

   (b) $\begin{vmatrix} 1 & 2 & 0 & 9 \\ 2 & 3 & 4 & 6 \\ 1 & 6 & 0 & -1 \\ 0 & -5 & 0 & -8 \end{vmatrix}$

10. Find the inverse of the matrix $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$

11. Prove that for a 3×3 matrix, one may find the determinant by a cofactor expansion along any row or column in the matrix.

12. Determine the ranks of the matrices below. How many linearly independent rows are in each? Which have inverses?

    (a) $\begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}$

    (b) $\begin{pmatrix} -3 & 4 \\ 4 & -5 \end{pmatrix}$

    (c) $\begin{pmatrix} -3 & 4 \\ 4 & -6 \end{pmatrix}$

    (d) $\begin{pmatrix} 2 & 4 \\ 4 & 8 \end{pmatrix}$

    (e) $\begin{pmatrix} 1 & 2 & 0 \\ 2 & 4 & 5 \\ 0 & 5 & 6 \end{pmatrix}$

    (f) $\begin{pmatrix} -1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & -2 \end{pmatrix}$

    (g) $\begin{pmatrix} 1 & 0 & 3 & 0 \\ 0 & 2 & 0 & 5 \\ 3 & 0 & 4 & 0 \\ 0 & 5 & 0 & 6 \end{pmatrix}$

6 Quadratic Forms

Quadratic forms are the next simplest functions after linear ones. Like linear functions they have a matrix representation, so that studying quadratic forms reduces to studying symmetric matrices. This is what this section is about.

Consider the function $F : \mathbb{R}^2 \to \mathbb{R}$, where $F = a_{11}x_1^2 + a_{12}x_1x_2 + a_{22}x_2^2$. We call this a **quadratic form** in $\mathbb{R}^2$. Notice that this can be expressed in matrix form as

$$F(\mathbf{x}) = \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} a_{11} & \frac{1}{2}a_{12} \\ \frac{1}{2}a_{12} & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \mathbf{x}'A\mathbf{x},$$

where $\mathbf{x} = (x_1, x_2)$, and A is unique and symmetric.

The quadratic form in $\mathbb{R}^n$ is

$$F(\mathbf{x}) = \sum_{i,j=1}^{n} a_{ij}x_ix_j,$$

where $\mathbf{x} = (x_1, \ldots, x_n)$, and A is unique and symmetric. This can also be expressed in matrix form:

$$F(\mathbf{x}) = \begin{pmatrix} x_1 & x_2 & \ldots & x_n \end{pmatrix} \begin{pmatrix} a_{11} & \frac{1}{2}a_{21} & \ldots & \frac{1}{2}a_{1n} \\ \frac{1}{2}a_{12} & a_{22} & \ldots & \frac{1}{2}a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{2}a_{n1} & \frac{1}{2}a_{n2} & \ldots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \mathbf{x}'A\mathbf{x}.$$

A quadratic form has a critical point at $\mathbf{x} = 0$, where it takes on the value 0. Therefore we can classify quadratic forms by whether $\mathbf{x} = 0$ is a maximum, minimum or neither. This is what definiteness is about.

7 Definiteness

Let A be an $n \times n$ symmetric matrix. Then A is:

1. **positive definite** if $\mathbf{x}'A\mathbf{x} > 0 \ \forall \ \mathbf{x} \neq \mathbf{0}$ in $\mathbb{R}^n$. That is, $\mathbf{x} = 0$ is a unique global minimum of the quadratic form given by A.
2. **positive semidefinite** if $\mathbf{x}'A\mathbf{x} \geq 0 \ \forall \ \mathbf{x} \neq \mathbf{0}$ in $\mathbb{R}^n$. That is, $\mathbf{x} = 0$ is a global minimum, but not a unique global one, of the quadratic form given by A.
3. **negative definite** if $\mathbf{x}'A\mathbf{x} < 0 \ \forall \ \mathbf{x} \neq \mathbf{0}$ in $\mathbb{R}^n$. That is, $\mathbf{x} = 0$ is a unique global maximum of the quadratic form given by A.
4. **negative semidefinite** if $\mathbf{x}'A\mathbf{x} \leq 0 \ \forall \ \mathbf{x} \neq \mathbf{0}$ in $\mathbb{R}^n$. That is, $\mathbf{x} = 0$ is a global maximum, but not a unique global one, of the quadratic form given by A.
5. **indefinite** if $\mathbf{x}'A\mathbf{x} > 0$ for some $\mathbf{x} \in \mathbb{R}^n$, and < 0 for some other $\mathbf{x} \in \mathbb{R}^n$. That is, $\mathbf{x} = 0$ is neither a maximum nor a minimum of the quadratic form given by A.

The definiteness of a matrix plays an important role. For example, for a function $f(x)$ of one variable, the sign of the second derivative $f''(x_0)$ at a critical point x_0 gives a sufficient condition for determining whether x_0 is a maximum, minimum or neither. This test generalizes to more than one variable using the definiteness of the Hessian matrix H (the matrix of the second order derivatives). (More on this when we get to optimization.) Similarly, a function $f(x)$ in one variable is concave if its second derivative is ≤ 0. A function of more than one variable is concave if its Hessian matrix is negative semidefinite.

There is a convenient way to test for the definiteness of a matrix. Before we can formulate this test we first need to define the concept of principal minors of a matrix.

7.1 Principal Minors and Leading Principal Minors

7.1.1 Principal Minors

Let A be an $n \times n$ matrix. A $k \times k$ submatrix of A obtained by deleting any $n-k$ columns and the same $n-k$ rows from A is called a kth-order **principal submatrix** of A. The determinant of a $k \times k$ principal submatrix is called a kth order **principal minor** of A.

Example List all the principal minors of the 3×3 matrix:

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}.$$

Answer: There is one third order principal minor of A, $\det(A)$. There are three second order principal minors:

1. $\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$, formed by deleting the third row and column of A.
2. $\begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix}$, formed by deleting the second row and column of A.
3. $\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix}$, formed by deleting the first row and column of A.

There are also three first order principal minors: a_{11}, by deleting the last two rows and columns; a_{22}, by deleting the first and last rows and columns; and a_{33}, by deleting the first two rows and columns.

7.1.2 Leading Principal Minors

The **leading principal minor** is the determinant of the **leading principal submatrix** obtained by deleting the last $n-k$ rows and columns of an $n \times n$ matrix A.

Example List the first, second, and third order leading principal minors of the 3×3 matrix:

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}.$$

Answer: There are three leading principal minors, one of order 1, one of order 2, and one of order 3:

1. $|a_{11}|$, formed by deleting the last two rows and columns of A.
2. $\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$, formed by deleting the last row and column of A.
3. $\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$, formed by deleting the no rows or columns of A.

Why in the world do we care about principal and leading principal minors? We need to calculate the signs of the leading principal minors in order to determine the definiteness of a matrix. We need definiteness to check second-order conditions for maxima and minima. We also need definiteness of the Hessian matrix to check to see whether or not we have a concave function.

7.2 Testing for Definiteness

We can test for the definiteness of the matrix in the following fashion:

1. A is positive definite iff all of its n leading principal minors are strictly positive.
2. A is negative definite iff all of its n leading principal minors alternate in sign, where $|A_1| < 0$, $|A_2| > 0$, $|A_3| < 0$, etc.
3. If some kth order leading principal minor of A is nonzero but does not fit either of the above sign patterns, then A is indefinite.

If the matrix A would meet the criterion for positive or negative definiteness if we relaxed the strict inequalities to weak inequalities (i.e. we allow zero to fit into the pattern), then although the matrix is not positive or negative definite, it may be positive or negative semidefinite. In this case, we employ the following tests:

1. A is positive semidefinite iff *every* principal minor of A is ≥ 0.
2. A is negative semidefinite iff *every* principal minor of A of odd order is ≤ 0 and every principal minor of even order is ≥ 0.

Notice that for determining semidefiniteness, we can no longer check just the leading principal minors, but we must check *all* principal minors. What a pain!

8 Homework

1. Express the quadratic form as a matrix product involving a symmetric coefficient matrix.

   (a) $Q = 8x_1x_2 - x_1^2 - 31x_2^2$

   (b) $Q = 3x_1^2 - 2x_1x_2 + 4x_1x_3 + 5x_2^2 + 4x_3^2 - 2x_2x_3$

2. List all the principal minors of the 4×4 matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix}$$

3. Prove that:

   (a) Every diagonal matrix whose diagonal elements are all positive is positive definite.

   (b) Every diagonal matrix whose diagonal elements are all negative is negative definite.

   (c) Every diagonal matrix whose diagonal elements are all positive or zero is positive semidefinite.

   (d) Every diagonal matrix whose diagonal elements are all negative or zero is negative semidefinite.

   (e) All other diagonal matrices are indefinite.

4. Determine the definiteness of the following matrices:

   (a) $\begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}$

   (b) $\begin{pmatrix} -3 & 4 \\ 4 & -5 \end{pmatrix}$

   (c) $\begin{pmatrix} -3 & 4 \\ 4 & -6 \end{pmatrix}$

   (d) $\begin{pmatrix} 2 & 4 \\ 4 & 8 \end{pmatrix}$

   (e) $\begin{pmatrix} 1 & 2 & 0 \\ 2 & 4 & 5 \\ 0 & 5 & 6 \end{pmatrix}$

   (f) $\begin{pmatrix} -1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & -2 \end{pmatrix}$

   (g) $\begin{pmatrix} 1 & 0 & 3 & 0 \\ 0 & 2 & 0 & 5 \\ 3 & 0 & 4 & 0 \\ 0 & 5 & 0 & 6 \end{pmatrix}$

9 Eigenvalues, Eigenvectors and Diagonalizability

9.1 Eigenvalues

Let A be a square matrix. An **eigenvalue** is a number λ such that

$$(A - \lambda I)\,\mathbf{x} = 0, \text{ for some } \mathbf{x} \neq 0$$

Notice also that this implies that an eigenvalue is a number λ such that $A\mathbf{x} = \lambda\mathbf{x}$.

Note that $x = 0$ is always a solution to $(A - \lambda I)\,\mathbf{x} = 0$. The definition of an eigenvalue requires that there is another solution to $(A - \lambda I)\,\mathbf{x} = 0$, which is different from 0. The system $(A - \lambda I)\,\mathbf{x} = 0$ has therefore more than one solution and $(A - \lambda I)$ is a singular matrix. In other words, and eigenvalue is a number λ which when subtracted from the diagonal elements of the matrix A creates a singular matrix. Also, a matrix A is non-singular if and only if 0 is not an eigenvalue of A.

Example: Find the eigenvalues of the matrix

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 4 & -17 & 8 \end{pmatrix}.$$

Assume that λ is an eigenvalue of A. Remember that if the matrix resulting when we subtract λI from A is singular, then its determinant must be zero. Then λ solves the equation

$$\left| \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 4 & -17 & 8 \end{pmatrix} - \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{pmatrix} \right| = 0$$

$$\left| \begin{pmatrix} -\lambda & 1 & 0 \\ 0 & -\lambda & 1 \\ 4 & -17 & 8-\lambda \end{pmatrix} \right| = 0$$

$$-\lambda^3 + 8\lambda^2 - 17\lambda + 4 = 0$$

$$-(\lambda - 4)(\lambda^2 - 4\lambda + 1) = 0$$

So one eigenvalue is $\lambda = 4$. To solve the quadratic, we use the quadratic formula:

$$\lambda = 4 \pm \frac{\sqrt{(-4)^2 - 4 \cdot 1 \cdot 1}}{2 \cdot 1} = 2 \pm \sqrt{3}$$

Therefore, the eigenvalues are $\lambda = \{2 - \sqrt{3}, 2 + \sqrt{3}, 4\}$.

9.2 Eigenvectors

Definition Given an eigenvalue λ, and **eigenvector** associated with λ is a non-zero vector $\mathbf{x}$ such that $(A - \lambda I)\,\mathbf{x} = 0$.

Notice that eigenvectors are only defined up to a scalar: If $(A - \lambda I)\,\mathbf{x} = 0$ then also $(A - \lambda I)\,2\mathbf{x} = 0$ etc., so all multiples of $\mathbf{x}$ are also eigenvectors associated with λ.

Example: Find the eigenvectors of the matrix

$$A = \begin{pmatrix} 0 & 0 & -2 \\ 1 & 2 & 1 \\ 1 & 0 & 3 \end{pmatrix}.$$

We must first find the eigenvalues of the matrix. The determinant of $(A - \lambda I)$ is

$$-\lambda (2-\lambda)(3-\lambda) + 2(2-\lambda) = -\lambda \left(6 - 5\lambda + \lambda^2\right) + 4 - 2\lambda = -8\lambda + 5\lambda^2 - \lambda^3 + 4$$

Therefore, we must find

$$\lambda^3 - 5\lambda^2 + 8\lambda - 4 = (\lambda - 1)(\lambda - 2)(\lambda - 2) = 0.$$

We can see that $\lambda = 1$ and $\lambda = 2$, where $\lambda = 2$ is a repeated root.

Since each eigenvector corresponds to an eigenvalue, let us consider the eigenvalue $\lambda = 1$. The matrix $A - \lambda I$ is then

$$A - \lambda I = \begin{pmatrix} -1 & 0 & -2 \\ 1 & 1 & 1 \\ 1 & 0 & 2 \end{pmatrix}.$$

This implies that the system of equations can be written

$$x_1 = -2x_2$$

$$x_1 = -x_2 - x_3$$

$$x_1 = -2x_2$$

when $A - \lambda I = \mathbf{0}$. Combining the second and third equation, we have

$$-2x_2 = -x_2 - x_3 \Rightarrow x_2 = x_3$$

Therefore, we have

$$x_1 = -2x_2$$

$$x_2 = x_2$$

$$x_3 = x_2$$

Which implies

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix} x_2.$$

Therefore, we can let x_2 be anything we want, say, $x_2 = s$. As we increase and decrease s, we trace out a line in 3-space, the direction of which is the eigenvector of $\lambda = 1$. Each point along the line has $\det(A - \lambda I) = 0$.

As a check, we can see whether $(A - \lambda I)\mathbf{x} = 0$. We then have

$$\begin{pmatrix} -1 & 0 & -2 \\ 1 & 1 & 1 \\ 1 & 0 & 2 \end{pmatrix} \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2+0-2 \\ -2+1+1 \\ -2+0+2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = \mathbf{0}.$$

We now know one eigenvector is $\begin{pmatrix} -2 & 1 & 1 \end{pmatrix}'$. We can find the others in a similar manner.

Let $\lambda = 2$. Then

$$(A - \lambda I) = \begin{pmatrix} -2 & 0 & -2 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix},$$

which implies

$$x_1 = -x_3,$$

$$x_1 = -x_3,$$

$$x_1 = -x_3.$$

Therefore, we can write the system as

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 \\ ??? \\ -1 \end{pmatrix} x_1.$$

Notice that we cannot write x_2 as a function of either x_1 or x_3. Therefore, there is no single eigenvector which corresponds to $\lambda = 2$. However, notice that since the system is independent of x_2 when $\lambda = 2$, we can let x_1 and x_2 be anything we want. Say $x_1 = s$ and $x_2 = t$. Then we can write the solution to the system as

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} s + \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} t.$$

Therefore, corresponding eigenvectors are $\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$, and their linear combinations. To check that this is the case, notice

$$(A - \lambda I)\,\mathbf{x} = \begin{pmatrix} -2 & 0 & -2 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} = \begin{pmatrix} -2+0+2 \\ 1+0-1 \\ 1+0-1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = \mathbf{0}$$

and

$$(A - \lambda I)\,\mathbf{x} = \begin{pmatrix} -2 & 0 & -2 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0+0+0 \\ 0+0+0 \\ 0+0+0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = \mathbf{0}.$$

Generally, if the matrix is an $n \times n$ matrix, we have n eigenvalues (not necessarily real) and n eigenvectors (up to scalars). This is because there are n roots (not necessarily real) to an nth order polynomial. Also, if an eigenvalue is repeated k times, then there are k corresponding eigenvectors to the repeated root.

9.3 Diagonalizability

A square matrix A is called **diagonalizable** if there exists an invertible matrix P such that $P^{-1}AP$ is a diagonal matrix. We state without proof the following theorem:

> Let A be $n \times n$, and assume A has n linearly independent eigenvectors. Define the matrices
>
> $$\Lambda = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix}, \quad P = \begin{pmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \dots & \mathbf{p}_3 \end{pmatrix},$$
>
> where $\mathbf{p}_i$ is the ith linearly independent eigenvector. Then $\Lambda = P^{-1}AP$.

Example: Consider again the matrix

$$A = \begin{pmatrix} 0 & 0 & -2 \\ 1 & 2 & 1 \\ 1 & 0 & 3 \end{pmatrix},$$

and remember that the corresponding eigenvalues were $\lambda = \{1, 2, 2\}$ and $\left\{ \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right\}$.

Therefore, the matrix P can be expressed as

$$P = \begin{pmatrix} -2 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix}$$

Notice that

$$P^{-1} = \begin{pmatrix} -1 & 0 & -1 \\ -1 & 0 & 2 \\ 1 & 1 & 1 \end{pmatrix}.$$

Then we can see that

$$P^{-1}AP = \begin{pmatrix} -1 & 0 & -1 \\ 1 & 0 & 2 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & -2 \\ 1 & 2 & 1 \\ 1 & 0 & 3 \end{pmatrix} \begin{pmatrix} -2 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix} =$$

$$= \begin{pmatrix} -1 & 0 & -1 \\ 2 & 0 & 4 \\ 2 & 2 & 2 \end{pmatrix} \begin{pmatrix} -2 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

Example: Consider the matrix

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

$det(A - \lambda I) = (1 - \lambda)^2$ and so the repeated eigenvalue of this matrix is $\lambda = 1$. To find the associated eigenvector , we solve

$$(A - I)x = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} x = \mathbf{0}$$

We get $x = \begin{pmatrix} x_1 \\ 0 \end{pmatrix}$, which can be written as $x = \begin{pmatrix} 1 \\ 0 \end{pmatrix} s$. Although λ has multiplicity 2, there is only one linearly independent eigenvector associated with it. We cannot find two linearly independent eigenvectors, and therefore A is not diagonalizable.
Notice however that A is invertible ($\det(A) = 1$). This is an example of a matrix that is invertible but not diagonalizable.

Real Symmetric Matrices Let A be $n \times n$ real symmetric matrix. Then all eigenvalues of A are real (but not necessarily distinct) and A is diagonalizable. Furthermore, we can find a matrix P of eigenvectors such that P is orthogonal, i.e. $P^{-1} = P^T$.

Eigenvalues & Definiteness For a diagonalizable matrix A, we can easily read of its definiteness from the eigenvalues (compare to homework question 3 in LA II notes):

1. A diagonalizable matrix A is positive (semi)definite if all eigenvalues are (weakly) positive.
2. A diagonalizable matrix A is negative (semi)definite if all eigenvalues are (weakly) negative.
3. A diagonalizable matrix A is indefinite if at least one eigenvalue is strictly positive and at least one eigenvalue is strictly negative.

We will need to diagonalize matrices when solving systems of differential equations.

10 Markov Chains

Assume k states of the world, which we label $1, 2, \ldots, k$. Let the probability that we move to state j from state i p_{ij}, and call it the **transition probability**. Then the **transition probability matrix** of the Markov chain is the matrix

$$P = \begin{pmatrix} p_{11} & \cdots & p_{1k} \\ \vdots & \ddots & \vdots \\ p_{k1} & \cdots & p_{kk} \end{pmatrix}.$$

Notice that the rows are the current state, and the columns are the states into which we can move. Also notice that the rows of the transition probability matrix must add up to one (otherwise the system would have positive probability of moving into an undefined state).

Example: Say that there are three states of the world: rainy, overcast, and sunny. Let state 1 be rainy, 2 be overcast, and 3 be sunny. Consider the transition probability matrix

$$P = \begin{pmatrix} .8 & .1 & .1 \\ .3 & .2 & .5 \\ .2 & .6 & .2 \end{pmatrix}.$$

For example, if it is rainy, the probability it remains rainy is .8. Also, if it is sunny, the probability that it becomes overcast is .6.

The transition probability matrix tells us the likelihood of the state of the system next period. However, what if we want to know the likelihood of the state in two periods? Call this probability transition matrix P_2. It can be shown that $PP = P_2$. Similarly, it can be shown that

$$P_n = P^n.$$

Example: If it is rainy today, what is the probability that is will be cloudy the day after tomorrow?

$$P_2 = \begin{pmatrix} .8 & .1 & .1 \\ .3 & .2 & .5 \\ .2 & .6 & .2 \end{pmatrix} \begin{pmatrix} .8 & .1 & .1 \\ .3 & .2 & .5 \\ .2 & .6 & .2 \end{pmatrix} = \begin{pmatrix} .69 & .16 & .15 \\ .40 & .37 & .23 \\ .38 & .26 & .36 \end{pmatrix}$$

Therefore, there is a 0.16 chance that if it is rainy today, then it will be cloudy the day after tomorrow.

Implicitly we have been assuming that we know the vector of probabilities today. For example, we say "assume it is cloudy". But what if we don't know what the probability will be today? We can define an initial state of probabilities as $\mathbf{x} = (p_1, p_2, \ldots, p_n)$. Then the probability that it will be rainy tomorrow is $\mathbf{x}_0 P = \mathbf{x}_1$.
A vector $\mathbf{x}$ with $0 \leq x_i \leq 1$ for all i and $\sum_{i=1}^n x_i = 1$ is called a *probability vector*.

Example: If there is a 30% chance of rain today, and a 10% chance of sun, what is the probability that is will be cloudy the day after tomorrow?

$$\mathbf{x}_2 = \mathbf{x}_1 P = \mathbf{x}_0 P^2 = \begin{pmatrix} .3 & .6 & .1 \end{pmatrix} \begin{pmatrix} .8 & .1 & .1 \\ .3 & .2 & .5 \\ .2 & .6 & .2 \end{pmatrix} \begin{pmatrix} .8 & .1 & .1 \\ .3 & .2 & .5 \\ .2 & .6 & .2 \end{pmatrix} =$$

$$= \begin{pmatrix} .3 & .6 & .1 \end{pmatrix} \begin{pmatrix} .69 & .16 & .15 \\ .40 & .37 & .23 \\ .38 & .26 & .36 \end{pmatrix} = \begin{pmatrix} .485 & .296 & .219 \end{pmatrix}$$

Therefore, there is a 0.296 chance that it will be cloudy the day after tomorrow.

A transition matrix P is called *regular* if P^n has only strictly positive entries for some integer n.
A *steady-state* vector q of a transition matrix P is a probability vector that satisfies the equation $\mathbf{q}P = \mathbf{q}$. Notice that this corresponds to $\lambda = 1$ as an eigenvector of P'.
If P is a regular transition matrix the following holds true:

- 1 is an eigenvalue of P.
- There is a unique probability vector q which is an eigenvector associated with the eigenvalue 1.
- $\mathbf{x}P^n \to q$ as $n \to \infty$

Example: What is the probability it will be cloudy as $n \to \infty$?

The matrix P is regular since P has only strictly positive entries. In order to find the long-run probabilities, we must realize that

$$\mathbf{q}P = \mathbf{q} \Rightarrow P'\mathbf{q}' = \mathbf{q}'.$$

Notice that if $\lambda = 1$, then we have

$$\mathbf{q}P = \lambda\mathbf{q} \Rightarrow P'\mathbf{q}' = \lambda\mathbf{q}'.$$

Therefore, $\mathbf{q}'$ is the eigenvector of P' which corresponds to the eigenvalue of $\lambda = 1$. Therefore, it suffices to find this eigenvalue.

$$(P' - \lambda I)\mathbf{q}' = \begin{pmatrix} -.2 & .3 & .2 \\ .1 & -.8 & .6 \\ .1 & .5 & -.8 \end{pmatrix} \mathbf{q}' = 0 \Rightarrow$$

$$\left(\begin{array}{ccc|c} -.2 & .3 & .2 & 0 \\ .1 & -.8 & .6 & 0 \\ .1 & .5 & -.8 & 0 \end{array}\right)$$

$$\left(\begin{array}{ccc|c} 1 & 0 & -\frac{34}{13} & 0 \\ 0 & 1 & -\frac{14}{13} & 0 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

$$q_1 = \frac{34}{13} q_3$$

$$q_2 = \frac{14}{13} q_3$$

$$q_3 = q_3.$$

Therefore, $\mathbf{q} = \left(\frac{34}{13}, \frac{14}{13}, 1\right) q_3$, where q_3 can be whatever we want. Since we must have that the sum of the elements of $\mathbf{q}$ be equal to 1, we choose q_3 accordingly.

$$\frac{34}{13} + \frac{14}{13} + \frac{13}{13} = \frac{61}{13} \Rightarrow q_3 = \frac{13}{61} \Rightarrow \mathbf{q} = \left(\frac{34}{61}, \frac{14}{61}, \frac{13}{61}\right).$$

11 Homework

1. Find the eigenvalues and the corresponding eigenvectors for the following matrices:

   (a) $\begin{pmatrix} 3 & 0 \\ 8 & -1 \end{pmatrix}$

   (b) $\begin{pmatrix} 10 & -9 \\ 4 & -2 \end{pmatrix}$

   (c) $\begin{pmatrix} 0 & 3 \\ 4 & 0 \end{pmatrix}$

   (d) $\begin{pmatrix} -2 & -7 \\ 1 & 2 \end{pmatrix}$

   (e) $\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$

   (f) $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

2. Determine whether the following matrices are diagonalizable:

   (a) $\begin{pmatrix} 2 & 0 \\ 1 & 2 \end{pmatrix}$

   (b) $\begin{pmatrix} 2 & -3 \\ 1 & -1 \end{pmatrix}$

   (c) $\begin{pmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 1 & 2 \end{pmatrix}$

   (d) $\begin{pmatrix} -1 & 0 & 1 \\ -1 & 3 & 0 \\ -4 & -13 & -1 \end{pmatrix}$

   (e) $\begin{pmatrix} 2 & -1 & 0 & 1 \\ 0 & 2 & 1 & -1 \\ 0 & 0 & 3 & 2 \\ 0 & 0 & 0 & 3 \end{pmatrix}$

3. Diagonalize the following matrices, if possible:

   (a) $\begin{pmatrix} -14 & 12 \\ -20 & 17 \end{pmatrix}$

   (b) $\begin{pmatrix} 1 & 0 \\ 6 & -1 \end{pmatrix}$

   (c) $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix}$

   (d) $\begin{pmatrix} 2 & 0 & -2 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{pmatrix}$

4. (midterm exam) You have the following transition probability matrix of a discrete state Markov chain:
$$\begin{pmatrix} \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{2} \\ 0 & 1 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{pmatrix}$$
What is the probability that the system is in state 2 at time $n+2$ given that it was in state 3 at time n?

5. (Homework problem) Suppose that the weather can be sunny or cloudy and the weather conditions on succesive mornings form a Markov chain with stationary transition probabilities. Suppose that the transition matrix is as follows
$$\begin{pmatrix} .7 & .3 \\ .6 & .4 \end{pmatrix}$$
where sunny is state 1 and cloudy is state 2. If it is cloudy on a given day, what is the probability that it will also be cloudy the next day?

6. (Homework problem) Suppose that three boys 1, 2, and 3 are throwing the ball to one another. Whenever 1 has the ball, he throws it to 2 with a probability of 0.2. Whenever 2 has the ball, he will throw it to 1 with probability 0.6. Whenever 3 has the ball, he is equally likely to throw it to 1 or 2.

   (a) Construct the transition probability matrix

   (b) If each of the boys is equally likely to have the ball at a certain time n, which boy is most likely to have the ball at time $n+2$?