Lecture VI: Linear Algebra, ODE and Dynamic Systems

1. Linear Algebra

In economics, it is very common to have ***systems of equations***,

$$\begin{aligned} Q_d &= a_0 + a_1 P \\ Q_s &= b_0 + b_1 P \\ Q_d - Q_s &= 0 \end{aligned} \tag{1}$$

We have supply, demand, and equilibrium. This is oozing economics, right? This problem can be expressed in general as

$$Ax = d$$

with A some $N \times M$ matrix times a column vector $x \in \mathbb{R}^M$ equal to some other vector $d \in \mathbb{R}^N$. The above is matrix notation for the system

$$\begin{aligned} a_{11}x_1 = a_{12}x_2 + \ldots + a_{1M}x_M &= d_1 \\ &\vdots \\ a_{N1}x_1 = a_{N2}x_2 + \ldots + a_{NM}x_M &= d_N \end{aligned}$$

With some wrangling, we can see that Equation (1) can be expressed in this way:

$$\underbrace{\begin{bmatrix} a_0 \\ b_0 \\ 0 \end{bmatrix}}_{d} = \underbrace{\begin{bmatrix} -a_1 & 1 & 0 \\ -b_1 & 0 & 1 \\ 0 & 1 & -1 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} P \\ Q_d \\ Q_s \end{bmatrix}}_{x}$$

In this case, A is 3×3 and "invertible," so we can explicitly solve for equilibrium price and quantity.

$$\begin{bmatrix} P \\ Q_d \\ Q_s \end{bmatrix} = \begin{bmatrix} -a_1 & 1 & 0 \\ -b_1 & 0 & 1 \\ 0 & 1 & -1 \end{bmatrix}^{-1} \begin{bmatrix} a_0 \\ b_0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} P \\ Q_d \\ Q_s \end{bmatrix} = \frac{1}{\det(A)} \begin{bmatrix} 1 & -1 & -1 \\ b_1 & -a_1 & -a_1 \\ b_1 & -a_1 & -b_1 \end{bmatrix} \begin{bmatrix} a_0 \\ b_0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} P \\ Q_d \\ Q_s \end{bmatrix} = \frac{1}{b_1 - a_1} \begin{bmatrix} a_0 - b_0 \\ a_0 b_1 - a_1 b_0 \\ a_0 b_1 - a_1 b_0 \end{bmatrix}$$

We can check this is the same answer we'd get if we'd solve this by leveraging equilibrium:

$$a_0 + a_1 P = b_0 + b_1 P \implies P = \frac{a_0 - b_0}{b_1 - a_1}$$

Then for quantity,

$$Q^s = Q^d = a_0 + a_1 \frac{a_0 - b_0}{b_1 - a_1} = \frac{a_0 b_1 - a_0 a_1 + a_1 a_0 - a_1 b_0}{b_1 - a_1} = \frac{a_0 b_1 - a_1 b_0}{b_1 - a_1}$$

In general, we find that

- If A is invertible[1] (more on invertible matrices below) then there exists a ***unique*** solution given by $x^* = A^{-1} d$.
- If $M < N$ then in general there are no solutions (you have more equations than parameters; in general it is not possible to satisfy all the equalities).
- If $N < M$ then in general there are infinitely many solutions (you have more parameters than equations; the additional parameters give you additional degrees of freedom).

[1] Note in this case $N = M$ is necessary, but not sufficient.

Hence it will be useful to study the properties of matrices and vectors, which are the building blocks of these types of systems.

1.1. Vector Spaces.

Definition 1. A ***vector space*** V is a collection of objects called vectors endowed with

1. Addition.
2. Multiplication by scale.

Some additive properties of vector spaces:

1. ***Commutativity***: $v + w = w + v$ for any $v, w \in V$.
2. ***Associativity***: $(v + w) + u = v + (w + u)$ for any $v, w \in V$.
3. ***Additive identity***: $\exists 0 \in V$ s.t. $v + 0 = v$ for any $v \in V$ (the 0 vector or the origin).
4. ***Additive inverse***: $\forall v \in V \ \exists - v \in V$ s.t. $v + (-v) = 0$.

Some multiplicative properties of vector spaces:

1. ***Unit rule*** or ***multiplicative identity***: $1v = v$ for any $v \in V$.
2. Multiplicative ***associativity***: $(\alpha\beta)v = \alpha(\beta v)$ for any $v \in V$ and $\alpha, \beta \in \mathbb{R}$.
3. ***Distributivity***: For any $\alpha, \beta \in \mathbb{R}$, $v, w \in V$, we have

$$\alpha(u + v) = \alpha u + \alpha v \quad \text{and} \quad (\alpha + \beta)v = \alpha v + \beta v$$

Remark 1. If you are like me, you might think it's a bit odd that we are making a big deal of associativity, commutativity, and distributivity (like we did back in primary school). The reason is that a more formal treatment of linear spaces would not take any of these properties for granted, and would be very careful in discussing everything in this section (and henceforth) using just the definitions.

And if you are like me you might also wonder, why don't we do that? We've done a lot of definition-theorem-proof style lectures in this class. It is a math class after all. So why not here as well? The reason is that it turns out that for all intents and purposes, vector spaces are equivalent to $\mathbb{R}^N$ (or subsets therein). There is a formal way of defining what that means and showing it; happy to provide details for the very curious, but for this class I will simply discuss $\mathbb{R}^N$ and I will take it to have the properties we expect. □

Definition 2. Let V be a vector space and $\{v_1, \ldots v_N\}$ s.t. $v_i \in V$. A ***linear combination*** of v_i is given by

$$\sum_{i=1}^{N} \alpha_i v_i = \alpha_1 v_1 + \ldots + \alpha_N v_N$$

for some arbitrary set of coefficients $\{\alpha_1, \ldots, \alpha_N\}$ s.t. $\alpha_i \in \mathbb{R}$.

Remark 2. A vector space is a space that contains all its linear combinations. That is, for any $\{v_i\}$ with

$v_i \in V$ and $\{\alpha_i\}$ with $\alpha_i \in \mathbb{R}$ we have

$$\sum_{i=1}^{N} \alpha_i v_i \in V$$

if and only if V is a vector space. This is not obvious (and in a more formal treatment we would prove it from the definitions), but it turns out one characterization of a vector space is a space that has all its linear combinations. Hence the reason why vector spaces are also known as ***linear spaces***. □

Definition 3. Let V be a vector space. We have $\{v_i\}$ s.t. $v_i \in V$ is a ***basis*** of V if for every $u \in V$ there exists a *unique* linear combination of $\{v_i\}$ s.t.

$$u = \sum_{i=1}^{N} \alpha_i v_i$$

The coefficients $\{\alpha_i\}$ are called the ***coordinates*** of u in V with respect to the basis $\{v_i\}$.

The easiest example of a basis is the standard basis in $\mathbb{R}^N$:

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \qquad e_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} \qquad \dots \qquad e_N = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

Naturally any $u \in \mathbb{R}^N$ can be uniquely represented as a linear combination of the standard basis if we take $\{\alpha_i\} = \{u_i\}$ the coefficients equal each of the entries in u.

Definition 4. Let V be a vector space; we say that $\{v_i\}$ s.t. $v_i \in V$ is a ***spanning set*** of V if for every $u \in V$ there exists a linear combination of $\{v_i\}$ s.t.

$$u = \sum_{i=1}^{N} \alpha_i v_i$$

Note that in the definition above, the linear combination does not have to be unique. Hence every basis of a vector space V is a spanning set of V, but not every spanning set of V can be a basis. Hence the question: When is a spanning set a basis?

Definition 5. A linear combination $\{\alpha_i\}$ is called ***trivial*** if $\alpha_i = 0$ for every i.

Definition 6. Let V be a vector space; we say that $\{v_i\}$ s.t. $v_i \in V$ are ***linearly independent*** if the only linear combination of v_i that is equal to 0 is trivial. That is

$$\sum_i \alpha_i v_i = 0 \implies \alpha_i = 0$$

Definition 7. Let V be a vector space; if $\{v_i\}$ s.t. $v_i \in V$ are not linearly independent, we say they are ***linearly dependent***.

Theorem 1. *Let V be a vector space; $\{v_i\}$ s.t. $v_i \in V$ are linearly dependent if for some j and some linear combination $\{\beta_{-j}\}$ we have that*

$$v_j = \sum_{i \neq j} \beta_i v_i$$

Proof. Take the linear combination $\{\alpha_i\}$ s.t. $\alpha_i = \beta_i$ for $i \neq j$ and $\alpha_j = -1$. Then

$$\sum_i \alpha_i v_i = \sum_{i \neq j} \beta_i v_i - v_j = 0$$

Since $\alpha_j \neq 0$, $\{\alpha_i\}$ is non-trivial, and so $\{v_i\}$ are not linearly independent. □

Theorem 2. *Let V be a vector space; $\{v_i\}$ s.t. $v_i \in V$ are a basis for $V \iff \{v_i\}$ are a linearly independent spanning set of V.*

Claim 1. *Let $\{v_i\}, \{u_i\}$ be any basis for a vector space V. Then $|\{v_i\}| = |\{u_i\}|$.*

That is, basis for a vector space have the same number of elements. This leads to the following:

Definition 8. A vector space V has ***dimension*** $\dim(V)$ equal to the number of elements in any basis.

1.2. Linear Transformations.

Definition 9. Let V, W be vector spaces. $T : V \to W$ is a ***linear transformation*** if

$$T(\alpha v + w) = \alpha T(v) + T(w)$$

for any $v \in V, w \in W, \alpha \in \mathbb{R}$.

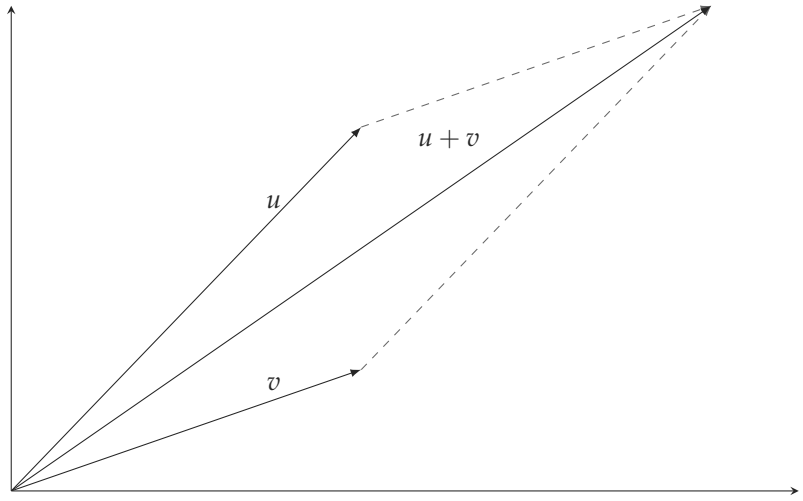


Figure 1: Example of a Linear Transformation

Theorem 3. *Let V, W be vector spaces, $T : V \to W$ a linear transformation, and $\{v_i\}$ a basis for V. Then*

$$T(v) = \sum_i \alpha_i T(v_i)$$

for $v = \sum_i \alpha_i v_i$.

Proof. We have that

$$T(v) = T\left(\sum_i \alpha_i v_i\right)$$

We can then proceed by induction, using the properties of a linear transformation we saw above. Since V is a linear space, it has all its linear combinations. Thus $\alpha_1 v_1 \in V$ and $\sum_{i>1} \alpha_i v_i \in V$. Applying the definition,

$$T(v) = T\left(\alpha_1 v_1 + \sum_{i>1} \alpha_i v_i\right) = \alpha_1 T(v_1) + T\left(\sum_{i>1} \alpha_i v_i\right)$$

Iterating the above:

$$T(v) = \sum_{i<k} \alpha_i T(v_i) + T\left(\alpha_k v_k + \sum_{i>k} \alpha_k v_k\right) = \sum_{i<k} \alpha_i T(v_i) + \alpha_k T(v_k) + T\left(\sum_{i>k} \alpha_k v_k\right)$$

Therefore we can simply write

$$T(v) = \sum_i \alpha_i T(v_i)$$

□

Theorem 4. *Any linear transformation $T : V \to W$ can be represented by a matrix.*

Proof. Let $\dim(V) = M \leq N$ so $V \subseteq \mathbb{R}^N$. A matrix A is a collection of vectors $\{a_1, \ldots, a_M\}$ with $a_i \in W$. Now consider the standard basis $\{e_i\}$ for $\mathbb{R}^N$ and take the subset of the standard basis that spans V. WLOG take these to be the first M elements of the sandard basis. For any vector $x \in V$ we find

$$T(x) = T\left(\sum_i x_i e_i\right) = \sum_i x_i T(e_i)$$

Let $a_i \equiv T(e_i)$ and we can see that

$$T(x) = Ax = \sum_i a_i x_i$$

Hence $T : V \to W$ can be equivalently represented by $A = [a_1 \ \cdots \ a_M]$ with $a_i \in W$. □

Example 1. Consider the following operation

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} 4 \\ 6 \\ 8 \end{bmatrix} = 4 \begin{bmatrix} 1 \\ 4 \end{bmatrix} + 6 \begin{bmatrix} 2 \\ 5 \end{bmatrix} + 8 \begin{bmatrix} 3 \\ 6 \end{bmatrix}$$

Which we express as a linear transform

$$T(e_1) = \begin{bmatrix} 1 \\ 4 \end{bmatrix} \qquad T(e_2) = \begin{bmatrix} 2 \\ 5 \end{bmatrix} \qquad T(e_3) = \begin{bmatrix} 3 \\ 6 \end{bmatrix}$$

That is,

$$T\begin{pmatrix} 4 \\ 6 \\ 8 \end{pmatrix} = 4T(e_1) + 6T(e_2) + 8T(e_3)$$

□

In general we can define matrix multiplication as linear operations on vector spaces:

$$AB = A[b_1 \dots b_N] = [Ab_1 \dots Ab_N]$$

Some properties of ***matrix multiplication***:

1. ***Associativity***: $(AB)C = A(BC)$
2. ***Distributivity***: $A(B + C) = AB + AC$

However, in general ***matrix multiplication is not commutative***. That is, $AB \neq BA$ in general (in fact, BA may not even be well-defined). Although matrix addition *is* commutative: $A + B = B + A$.

1.3. Matrix Inverse, Rank, and Determinant.

Definition 10. The vector space spanned by the columns of a matrix A is the ***column space*** of A. The ***rank*** of a matrix, $\text{rank}(A)$, is the dimension of the column space (the maximum number of linearly independent columns).[2] A square matrix $N \times N$ is called ***full-rank*** if $\text{rank}(A) = N$ and ***rank-deficient*** if $\text{rank}(A) < N$.

Definition 11. A $N \times N$ matrix A is called ***diagonal*** if all its non-diagonal entries are 0.

Definition 12. The ***identity matrix*** is an $N \times N$ diagonal matrix with all diagonal entries equal to 1 (and non-diagonal entries equal to 0). For any $N \times N$ matrix A, $AI = IA = A$.

$$I = \begin{bmatrix} 1 & 0 & 0 & & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & & 0 \\ & \vdots & & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

Definition 13. The $N \times N$ matrix A is said to be ***left invertible*** if $\exists C_L$ s.t. $C_L A = I$ the identity. We say it is ***right invertible*** if $\exists C_R$ s.t. $AC_R = I$. If A is left and right invertible with $C_L = C_R$, we say that A is ***invertible***, and we denote $C_R = C_L = A^{-1}$ the ***inverse*** of A.[3]

Definition 14. If A is not invertible, we say that A is ***singular***.

Theorem 5. *Let A be a $N \times N$ matrix. A is invertible $\iff$ $\text{rank}(A) = N$.*

Some other properties of the rank:

- $\text{rank}(A) = \text{rank}(A^T) = \text{rank}(AA^T) = \text{rank}(A^T A)$.
- $\text{rank}(AB) \leq \min\{\text{rank}(A), \text{rank}(B)\}$.
- $\text{rank}(CAB) = \text{rank}(A)$ if C, B are non-singular.

[2] It turns out that the row-rank of a matrix, the maximum number of linearly independent rows, is the same as the column-rank of a matrix. Hence we can just talk about the rank without clarifying row or column.

[3] The left and right inverses are the same for a square matrix if either exist, but this is a result, not an assumption. Further, note you can define the left and right inverses for non-square matrices, but such matrices needn't be invertible as only one of C_L or C_R might exist.

Example 2. Can we find the inverse of the matrix

$$B = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$$

using elementary row operations?

$$\begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \to \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 1/2 \\ 1 & 1 \end{bmatrix} \to \begin{bmatrix} 1/2 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 1/2 \\ 0 & -1/2 \end{bmatrix} \to \begin{bmatrix} 1/2 & 0 \\ 1/2 & -1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 \\ 0 & -1/2 \end{bmatrix} \to \begin{bmatrix} 1 & -1 \\ 1/2 & -1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \to \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}$$

The steps here were:

1. Multiply the first row by $1/2$.
2. Multiply the second row by -1 and add the first row to the second row.
3. Add the second row to the first row.
4. Multiply the second row by -2.

In general, however, we have a shortcut for 2×2 matrices:

$$B = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \qquad B^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

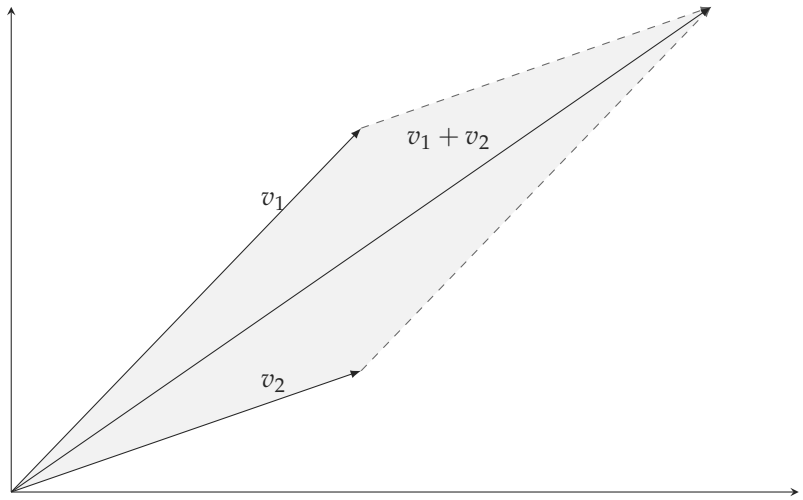
You can check this formula gives exactly what we found using row operations. □

Definition 15. The ***determinant*** of a $N \times N$ square matrix A is

$$\det(A) = \sum_{j=1}^{N} (-1)^{i+j} a_{ij} \det(A_{-i,-j})$$

with i any fixed row of A (i.e. the same i for every j).

Consider this picture:



The set

$$\left\{ v : v = \sum_{i=1}^{N} t_i v_i \quad t_i \in [0,1] \right\}$$

will give the parallelogram, and the determinant of $V = [v_1 \cdots v_N]$ will be be the area of the parallelogram. (Rather, the absolute value of the determinant; if the determinant is negative that just says something about the direction of the vectors, but the intuition remains.) Some properties of the determinant:

- $\det(AB) = \det(A)\det(B)$.
- $\det(A) = \prod_i a_{ii}$ if A is diagonal.
- $\det(I) = 1$.
- $\det(\alpha A) = \alpha^N \det(A)$ for any $\alpha \in \mathbb{R}$.
- $\det(A^{-1}) = \det(A)^{-1}$.

Another point to make about the determinant is that it is related to the rank, that is, whether the columns (or rows) of a square matrix A are linearly independent.

1. If A has a column of row that is all 0, then $\det(A) = 0$.

   To see this, note that from the formula of the determinant, if we have a row or column of all 0s, then each element of the sum will eventually be multiplied by 0, and the entire term will be 0. Alternatively, if one of the "sides" of the parallelogram is between the origin and the origin, the area will be 0.

2. If A has two columns or rows that are equal, then $\det(A) = 0$.

   If two columns of A are equal, WLOG let them be the first two columns, consider the matrix P with all

diagonal entries equal to 1 and all but one non-diagonal entries equal to 0. Let $P_{21} = -1$. That is,

$$P = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -1 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & & \vdots \\ \vdots & & & \ddots & \\ 0 & & & & 1 \end{bmatrix}$$

What is AP? In this case, with $A = [a_1 \ a_2 \ \cdots \ a_N]$, we have

$$AP = \begin{bmatrix} a_1 - a_2 & a_2 & \cdots & a_N \end{bmatrix} = \begin{bmatrix} 0 & a_2 & \cdots & a_N \end{bmatrix}$$

with the first column equal to 0 since we assumed the first two columns were equal. Finally, we found in the previous bullet that if a column was all 0s then the determinant was 0. Hence

$$0 = \det(AP) = \det(A)\det(P) \implies \det(A) = 0 \quad \text{or} \quad \det(P) = 0$$

Note $\det(P) = 1$, so it must be that $\det(A) = 0$.

3. If A has a columns is a multiple of another column, then $\det(A) = 0$.

   Again WLOG suppose these are the first two columns and let A be as above. If $a_1 = \alpha a_2$ then we can define P almost identically, except $P_{21} = -\alpha$. That is,

$$P = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -\alpha & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & & \vdots \\ \vdots & & & \ddots & \\ 0 & & & & 1 \end{bmatrix}$$

   And again we have $AP = [a_1 - \alpha a_2 \ a_2 \ \cdots \ a_N] = [0 \ a_2 \ \cdots \ a_N]$. Hence $\det(AP) = \det(A)\det(P) = 0$, but again we have $\det(P) = 1$, so $\det(A) = 0$.

4. If A has a columns or rows that are linearly dependent, then $\det(A) = 0$.

   This is really the point we wanted to make: A determinant of 0 is tied to the linear dependence of the columns of A. This follows from the work we did above: In this case, $\exists\{\alpha_i\}$ s.t. $\alpha_i \neq 0$ for some i s.t.

$$a_j = \sum_{i \neq j} \alpha_i a_i$$

for some j. WLOG let $j = 1$, and the matrix P becomes

$$P = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -\alpha_2 & 1 & 0 & \cdots & 0 \\ -\alpha_3 & 0 & 1 & & \vdots \\ \vdots & & & \ddots & \\ -\alpha_N & & & & 1 \end{bmatrix}$$

So that

$$AP = \begin{bmatrix} a_1 - \alpha_2 a_2 - \alpha_3 a_3 \dots & a_2 & \cdots & a_N \end{bmatrix} = \begin{bmatrix} a_1 - \sum_{i \neq 1} \alpha_i a_i & a_2 & \cdots & a_N \end{bmatrix} = \begin{bmatrix} 0 & a_2 & \cdots & a_N \end{bmatrix}$$

So, one last time, $\det(AP) = \det(A)\det(P) = 0$ and $\det(P) = 1$ implies $\det(A) = 0$.

Theorem 6. *A is rank-deficient $\iff$ $\det(A) = 0$.. Equivalently, A is non-invertible $\iff$ $\det(A) = 0$.*

Definition 16. For a $N \times N$ matrix A, the M_{ij} ***minor*** is the *determinant* of the $(N-1) \times (N-1)$ sub-matrix obtained by deleting the ith row and jth column of A.

Definition 17. The *i, j* ***cofactor*** of the square $N \times N$ matrix A is given by

$$C_{ij} = (-1)^{i+j} M_{ij}$$

Note one way to write the determinant is as $\det(A) = \sum_{j=1}^{N} a_{ij} C_{ij}$ for any given i.

Definition 18. The ***adjoint*** of an $N \times N$ matrix A is given by

$$\text{adj}(A) = \begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1N} \\ \vdots & & \ddots & C_{1N} \\ C_{N1} & C_{N2} & \cdots & C_{NN} \end{bmatrix}$$

Note the above is a matrix of scalars, since C_{ij} is the product of a determinant, which is a scalar, and $(-1)^{i+j}$, which is a scalar as well.

Theorem 7. *If A is a non-singular $N \times N$ matrix then*

$$A^{-1} = \frac{1}{\det(A)} \, adj(A)$$

For instance, for a 2×2 matrix we have

$$B = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

Applying the theorem,

$$B^{-1} = \frac{1}{\det(B)} \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Theorem 8 (Cramer's Rule). *Let A be a $N \times N$ non-singular matrix s.t. $Ax = b$. Then*

$$x_i = \frac{\det(\widetilde{A}_i)}{\det(A)}$$

for

$$\widetilde{A}_i = \begin{bmatrix} a_1 & \cdots & a_{i-1} & b & a_{i+1} & \cdots & a_N \end{bmatrix}$$

That is, the matrix obtained by replacing the ith column of A with b.

Definition 19. The ***trace*** of an $N \times N$ matrix A is the sum of its diagonal elements,

$$tr(A) = \sum_{i=1}^{N} a_{ii}$$

Some properties of the trace:

- $\text{trace}(A + B) = \text{trace}(A) + \text{trace}(B)$
- $\text{trace}(AB) = \text{trace}(BA)$ if both products exist.
- $\text{trace}(\alpha A) = \alpha \,\text{trace}(A)$ for any $\alpha \in \mathbb{R}$.

An example which comes up in econometrics is that for any $N \times K$ matrix X s.t. $\left(X^T X\right)^{-1}$ exists, we have

$$\text{trace}\left(\underbrace{X}_{A}\overbrace{(X^TX)^{-1}X^T}^{B}\right) = \text{trace}((X^TX)^{-1}X^TX) = \text{trace}(I_{K\times K}) = K$$

Further,

$$\text{trace}\left(I_{N\times N} - X(X^TX)^{-1}X^T\right) = \text{trace}\left(I_{N\times N}\right) - \text{trace}(X(X^TX)^{-1}X^T\Big) = N - K$$

1.4. Eigenvalues and Eigenvectors. Take

$$T(v) = \begin{bmatrix} 3 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

Note that for the standard basis,

$$\begin{bmatrix} 3 & 0 \\ 0 & -1 \end{bmatrix} e_1 = 3e_1 \qquad \begin{bmatrix} 3 & 0 \\ 0 & -1 \end{bmatrix} e_2 = -e_2$$

For this particular example, we can see the standard basis is exactly scaled by 3 and -1.

Definition 20. Let A be a $N \times N$ square matrix. The $N \times 1$ vector $v \neq 0$ is an ***eigenvector*** (or characteristic vector) of A with corresponding ***eigenvalue*** or (characteristic root) λ if

$$Av = \lambda v$$

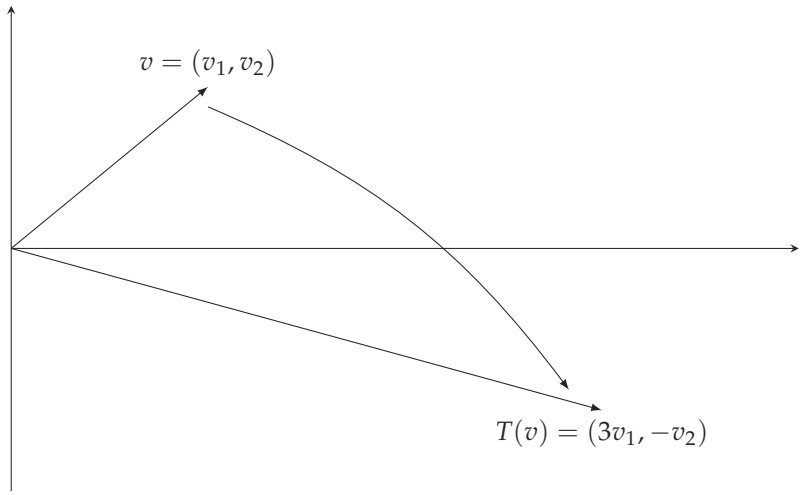


Figure 2: Visualization of a Transformation

In general we want to find a basis $\{v_i\}$ of (the column space of) A s.t.

$$Av_i = \lambda_i v_i$$

that is, a basis of eigenvectors. Note that

$$Av = \lambda v \iff (A - \lambda I)v = 0$$

This means that for any $v \neq 0$, we have that

$$p(\lambda) \equiv \det(A - \lambda I) = 0$$

$p(\lambda)$ is called the ***characteristic polynomial*** of A. (Recall that a matrix has a zero determinant if it is singular; in this case if$(A - \lambda I)v = 0$ for any non-zero vector then the columns of $A - \lambda I$ are not linearly independent, and thus the matrix is singular and has a zero determinant.)

Example 3. Consider

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \qquad \det \begin{bmatrix} 1-\lambda & 2 \\ 2 & 1-\lambda \end{bmatrix} = 0 \iff (1-\lambda)^2 - 4 = 0 \iff (\lambda+1)(\lambda-3) = 0$$

Hence $\lambda = \{-1, 3\}$. These are the ***eigenvalues***. Now we find the ***eigenvectors***. For $\lambda = -1$

$$(A - \lambda I)v = 0 \iff \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix} v = 0$$

Which gives that v must be s.t. $v_1 = -v_2$. For $\lambda = 3$

$$(A - \lambda I)v = 0 \iff \begin{bmatrix} -2 & 2 \\ 2 & -2 \end{bmatrix} v = 0$$

Which gives that v must be s.t. $v_1 = v_2$. It is often useful for the norm of the basis to be 1. Hence the

eigenvectors are given by

$$v = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \end{bmatrix} \quad \text{and} \quad u = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Note that we can express the above as a system:

$$AP = P\Lambda$$

where

$$P = \begin{bmatrix} v & u \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \qquad \Lambda = \begin{bmatrix} -1 & 0 \\ 0 & 3 \end{bmatrix}$$

In this case, note $v^T u = 0$ and $v^T v = u^T u = 1$. Hence

$$P^T P = \begin{bmatrix} v^T \\ u^T \end{bmatrix} \begin{bmatrix} v & u \end{bmatrix} = \begin{bmatrix} v^T v & v^T u \\ u^T v & u^T u \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

and if we pre-multiply the eigen-system by P^T we get

$$P^T AP = \Lambda$$

□

Definition 21. A an $N \times N$ matrix is ***diagonalizable*** if $\exists P$ and diagonal matrix Λ s.t.

$$P^{-1}AP = \Lambda$$

Note we can equivalently write $A = P\Lambda P^{-1}$.

It will be useful to note if A is diagonalizable

- A^m for any $m \in \mathbb{N}$ can computed as $P\Lambda^m P^{-1}$.

  $$A^m = A \times \ldots \times A = P\Lambda P^{-1} \times \ldots \times P\Lambda \times \ldots \times \Lambda P^{-1} = P\Lambda^m P^{-1}$$

  since each of the $P^{-1}P$ cancel and $\Lambda \times \Lambda = \Lambda^2$ and so on.

- $A^{1/m}$ for any $m \in \mathbb{N}$ can computed as $P\Lambda^{1/m}P^{-1}$.

  $$\left(P\Lambda^{1/m}P^{-1}\right)^m = P\Lambda^{1/m}P^{-1} \times \ldots \times P\Lambda^{1/m}P^{-1} = P\Lambda^{1/m} \times \ldots \times \Lambda^{1/m}P^{-1} = P\Lambda P^{-1} = A$$

- If Λ is invertible, then A^{-m} for any $m \in \mathbb{N}$ can computed as $P\Lambda^{-m}P^{-1}$.

  $$A^{-m} = (A^m)^{-1} = \left(P\Lambda^m P^{-1}\right)^{-1} = P\Lambda^{-m}P^{-1}$$

  and we can check

  $$A^m A^{-m} = P\Lambda^m P^{-1} P\Lambda^{-m} P^{-1} = P\Lambda^m \Lambda^{-m} P^{-1} = PP^{-1} = I$$

- Similarly, if Λ is invertible $A^{-1/m}$ for any $m \in \mathbb{N}$ can computed as $P\Lambda^{-1/m}P^{-1}$.

- Let $q \in \mathbb{Q}$ (so $q = m/n$ for $m, n \in \mathbb{Z}$). Then $A^q = (A^m)^{1/n}$ is $P(\Lambda^m)^{1/n}P^{-1} = P\Lambda^q P^{-1}$.
- Finally, $A^r = P\Lambda^r P^{-1}$ for any $r \geq 0$ (if in addition Λ is invertible, this holds for any $r \in \mathbb{R}$).

Claim 2. *If A is diagonalizable then P is a matrix of eigenvectors and Λ is a matrix of eigenvalues.*

Proof. If A is diagonalizable then

$$AP = P\Lambda$$

Since the kth column of P is s.t. $Av_k = \lambda_k v_k$ we can see v_k is an eigenvector and λ_k an eigenvalue. □

Theorem 9. *Let A be a $N \times N$ matrix with eigenvalues $\lambda_1, \ldots, \lambda_N$.*

- $trace(A) = \sum_{i=1}^N \lambda_i$
- $\det(A) = \prod_{i=1}^N \lambda_i$.

Proof. Let us show this assuming A is diagonalizable.[4]

$$\text{trace}(A) = \text{trace}(P^{-1}\Lambda P) = \text{trace}(\Lambda P P^{-1}) = \text{trace}(\Lambda) = \sum_i \lambda_i$$

$$\det(A) = \det(P^{-1}\Lambda P) = \det(P^{-1})\det(\Lambda)\det(P) = \det(\Lambda) = \prod_i \lambda_i$$

□

Theorem 10. *If A is symmetric then A is diagonalizable and P can be chosen to be orthonormal ($P^{-1} = P^T$ with each column of P equal to a unit vector). This is known as the* ***spectral decomposition****.*

Recall in **Example 3**, the matrix A we considered was symmetric, and we found a diagonalization where $P^{-1} = P^T$. This theorem states that we can do this for any symmetric matrix. Below I present an example for a matrix that is not symmetric.

Remark 3. I am not altogether in the know about linear algebra terminology. I have heard this type of diagonalization be referred to as eigendecomposition, which is intuitive, and as spectral decomposition, as it apparently comes from the spectral theorem. However, a spectre sounds too spooky for this class. □

Theorem 11. *Let A be a symmetric $N \times N$ matrix. Then*

1. *A is PD (ND) $\iff$ $\lambda_i > 0$ (< 0) for all eigenvalues λ_i of A.*

[4] For the general proof, recall $\lambda_1, \ldots, \lambda_N$ are the roots of the polynomial given by

$$\det(A - \lambda I) = (\lambda_1 - \lambda) \times (\lambda_2 - \lambda) \times \ldots \times (\lambda_N - \lambda)$$

For the determinant, set $\lambda = 0$ and find $\det(A) = \prod_{i=1}^N \lambda_i$. For the trace, we need to leverage something known as "Vieta's formulas." The relevant result is that for a polynomial of order N, the $N-1$ coefficient is the sum of the roots of the polynomial (this sounds esoteric, but think about how you learned to expand formulas like $(\lambda_1 - \lambda)(\lambda_2 - \lambda)$; the "middle" term is $-(\lambda_1 + \lambda_2)$, and this is just the generalization). If we can show $\text{trace}(A)$ is the $N-1$ coefficient of the characteristic polynomial we'd be done. To see it, note that λ^{N-1} will only appear if all the diagonal elements are multiplied (any other permutation will give at most a polynomial of order $N-2$). Hence λ^{N-1} only appears as part of the term

$$\prod_i (a_{ii} - \lambda)$$

Here the roots are a_{ii}, so the coefficient on λ^{N-1} is $\sum_i a_{ii} = \text{trace}(A)$.

2. *A is PSD (NSD)* $\iff$ $\lambda_i \geq 0$ (≤ 0) *for all eigenvalues* λ_i *of* A.

Proof. If A is symmetric, then

$$v'Av = v'P\Lambda P^T v = u^T \Lambda u \tag{2}$$

for $u = P^T v$ (since P is non-singular, $u \neq 0$ if $v \neq 0$). Hence the definiteness of A is the same as the definiteness of Λ. From here we can see that

$$u^T \Lambda u = \sum_{i=1}^{N} \lambda_i u_i^2 \tag{3}$$

where $u_i^2 \geq 0$ (strict for at least one i). If $\lambda_i > 0$ for all i, $v'Av = \sum_{i=1}^{N} \lambda_i u_i^2 > 0$, so it is PD. Conversely, if A is PD, we know $v'Av > 0$ for any $v \neq 0$. In particular, it must be true for v equal to each of the columns of P. Since the columns of P are orthonormal, for $v = p_i$ the ith column of P, $P^T p_i = e_i$ the ith standard vector (i.e. 1 in the ith entry and 0 elsewhere). Hence $v'Av = \lambda_i$ when $v = p_i$; since all such quadratic forms are positive, $\lambda_i > 0$. The logic for < 0 and ND, ≥ 0 and PSD, and ≤ 0 and NSD is completely analogous. □

EigenExample

$$A = \begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix}$$

We can find that $\det(A - \lambda I) = 0$ gives

$$(1-\lambda)^2 + 4 = 0 \qquad \lambda = \frac{2 \pm \sqrt{4 - 4 \cdot 5}}{2}$$

Hence $\lambda = 1 \pm 2i$. Solving for v, we get

$$v_1 = (\alpha, \alpha i) v_2 = (\alpha, -\alpha i)$$

Let $\alpha = 1$. Then

$$P = \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix} \qquad P^{-1} = \begin{bmatrix} -i & -1 \\ -i & 1 \end{bmatrix}$$

Hence

$$A = \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix} \begin{bmatrix} 1+i & 0 \\ 0 & 1-i \end{bmatrix} \begin{bmatrix} -i & -1 \\ -i & 1 \end{bmatrix}$$

2. Introduction to Ordinary Differential Equations (ODE)

2.1. Dynamic Equations. The basic idea is that we want to understand how a variable changes over time. If time is continuous (like it is IRL), then we say that

$$\dot{y} \equiv \frac{dy}{dt}$$

In general, we have an nth order linear differential equation as

$$y^{(n)} + \alpha_{n-1}(t)y^{(n-1)} + \ldots + \alpha_0(t)y = f(t)$$

where $\alpha_i(t)$ are constants that depend only on t and $y^{(n)}$ is the nth derivative of y with respect to t. We say it is ***homogeneous*** if $f(t) = 0$.

The equation is called ***autonomous*** if t only enters through $y(t)$; that is, $\alpha_i(t) \equiv \alpha_i$. For example,

$$\dot{y} = ky \iff \dot{y} - ky = 0$$

is an autonomous *and* homogeneous equation. By contrast,

$$\dot{y} = ty \iff \dot{y} - ty = 0$$

is homogeneous *but not* autonomous.

Example 4. Consider the autonomous and homogeneous system

$$\dot{y} = ky$$

This translates to

$$\frac{dy}{dt} = ky$$

which means that

$$\int \frac{dy}{dt}\frac{1}{y}dt = \int k dt \iff \log(y) + C_1 = kt + C_2$$

Let $C_3 \equiv \exp(C_2 - C_1)$, and we have that

$$y = C_3 e^{kt}$$

This equation is not quite defined, since the behavior of y depends on the constant C_3. Hence we need some initial conditions to pin down the path of y with respect to t. For example,

$$y(0) = 5 \qquad 5 = C_3 e^0 \implies C_3 = 5 \implies y(t) = 5e^{kt}$$

□

Example 5. Consider the homogeneous but not autonomous system

$$\dot{y} = ty$$

Again,

$$\frac{dy}{dt} = ty$$

which means that

$$\int \frac{dy}{dt}\frac{1}{y}dt = \int t dt \iff \log(y) + C_1 = \frac{t^2}{2} + C_2$$

Let $C_3 \equiv \exp(C_2 - C_1)$, and we have that

$$y = C_3 e^{\frac{t^2}{2}}$$

Note that $y(0) = C_3$. Hence we can write

$$y(t) = y(0)e^{\frac{t^2}{2}}$$

□

We will not always use the separation of variables method. In general, if we have an equation of the form

$$\dot{y} + p(t)y = q(t)$$

that will be neither autonomous nor homogeneous. One method to solve this system is to use the ***integrating factor***. Multiply each term by $\exp(\int p(t)dt)$:

$$\exp\left\{\int p(t)dt\right\}\left(\dot{y} + p(t)y\right) = \exp\left\{\int p(t)dt\right\}q(t)$$

Why? Because we have that

$$\frac{d}{dt}\left(\exp\left\{\int p(t)dt\right\}y\right) = \exp\left\{\int p(t)dt\right\}\dot{y} + \exp\left\{\int p(t)dt\right\}p(t)y$$

Hence

$$\int\left[\frac{d}{dt}\left(\exp\left\{\int p(t)dt\right\}y\right)\right]dt = \int\left[\exp\left\{\int p(t)dt\right\}q(t)\right]dt$$

gives

$$y(t) = \frac{\int\left[\exp\left\{\int p(t)dt\right\}q(t)\right]dt + C}{\exp\left\{\int p(t)dt\right\}}$$

for some constant C. Let's see an example:

Example 6. Consider the system

$$\dot{y} = ay + b$$

The issue is that this is not homogeneous. How to proceed in this case? $p(t) = -a$ and $q(t) = b$; the integrating factor is

$$\exp\left\{\int p(t)dt\right\} = \exp\left\{\int -adt\right\} = \exp\{-at\}$$

Plugging the formula,

$$y(t) = \frac{\int \exp\{-at\}bdt + C}{\exp\{-at\}} = -\frac{b}{a} + C\exp\{at\}$$

□

2.2. Dynamic Systems. Consider an equation in discrete time

$$y_t = \alpha y_{t-1} + x_t$$

An nth order ***linear difference*** equation is

$$y_{t-n} + \alpha_{(n-1)}(t)y_{t-(n-1)} + \ldots + \alpha_0(t)y_t = f(t)$$

We say that this is ***homogeneous*** if $f(t) = 0$ and ***autonomous*** if $\alpha_i(t) \equiv \alpha_i$ do not depend on t. The way to solve this is by recursive computation:

$$y_t = \frac{1}{k} y_{t-1}$$

In general,

$$y_t = \frac{1}{k^2} y_{t-2} = \frac{1}{k^3} y_{t-3} = \ldots = \frac{1}{k^t} y_0$$

For example take the 2nd order homogeneous and autonomous difference equation

$$y_{t-2} + \alpha y_{t-1} + \beta y_t = 0$$

We assume that $y_t = Ca^t$, and we check against that. So

$$Ca^{t-2} + \alpha Ca^{t-1} + \beta Ca^t = 0 \iff a^{-2} + \alpha a^{-1} + \beta = 0 \iff \beta a^2 + \alpha a + 1 = 0$$

Hence

$$a = \frac{-\alpha \pm \sqrt{\alpha^2 - 4\beta}}{2\beta}$$

Considera first-order ***dynamic system***:

$$x_t = Ax_{t-1}$$

If A is diagonalizable, then

$$x_t = P^{-1}\Lambda P \left(P^{-1}\Lambda P x_{t-2}\right) = P^{-1}\Lambda^2 P x_{t-2} = \ldots = P^{-1}\Lambda^t P x_0$$

Hence the dynamics of the system can be determined by analyzing the matrix of eigenvectors Λ.

- If all eigenvalues have absolute value < 1 then the system is globally stable ($\Lambda^t \to 0$).
- If all eigenvalues have absolute value > 1 then the system is globally unstable ($\Lambda^t \to \pm\infty$)..

In continuous time, consider

$$\dot{x} = Ax$$

Again, assuming A id diagonalizable,

$$P\dot{x} = \Lambda P x$$

Let $y = Px$, so $\dot{y} = P\dot{x}$, and we have

$$\dot{y} = \Lambda y$$

with Λ diagonal. In this case, the solution is given by

$$y = \exp(\Lambda t)C$$

for some constant vector C and $\exp(\Lambda t)$ denoting the diagonal matrix diag $\{\lambda_1 t, \ldots, \lambda_N t\}$. Note

$$\dot{y} = \Lambda \underbrace{\exp(\Lambda t)C}_{y} = \Lambda y$$