1 Vector Spaces

Linear systems or systems of linear equations arise in two ways in economics. Some economic models have a natural linear structure as expenditure systems. Or, the relationships among the variables are described by a system of nonlinear equations, but one takes the derivative of these equations to approximate the nonlinear equations by a system of linear equations. Typical linear equations are

$$x_1 + 3x_2 = 2, \quad 2x_1 - x_2 = 0,$$

or in general they are of the form

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n &= b_{1n} \\ a_{m1}x_1 + a_{m2}x_2 + ... + a_{mn}x_n &= b_{mn} \end{aligned} \Leftrightarrow Ax = b$$

The letters $a_{11,...}, a_{mn}, b_1, ..., b_m$ are the parameters of the system and $x_1, ..., x_n$ are the variables to the system: exogenous vs. endogenous variables. Geometrically speaking, the problem is to find the combination of the column vectors on the left side such that they produce the vector on the right side. The representation of the linear system is simplified through the introduction of vectors and rectangular arrays of the coefficients, matrices. We start out by the defining the vector space.

Definition 1. A vector space V (over the field K) is set of objects which can be added together and multiplied by elements of K, so that the sum of two elements of V is again an element of V, the product of an elements of V by an element of K is an element of V.

Example 1. Thefinite-dimensional Euclidean space $\mathbb{R}^n$.

Example 2. The space of all functions $f(x)$ defined on the unit interval.

Definition 2. Let V be a vector space, and let W be a subset of V. W is a subspace of V if
  (i) $v, w \in W \Rightarrow v + w \in W,$
  (ii) $v \in W,\ c \in \mathbb{R} \Rightarrow \alpha v \in W,$

Definition 3. Let V be an arbitrary vector space, and let $v_1, ..., v_n$, be elements of V. Let $\alpha_1, ..., \alpha_n$ be numbers. An expression of

$$\alpha_1 v_1 + \cdots + \alpha_n v_n$$

is called a linear combination of $v_1, ..., v_n$.

Proposition 1. *Let W be the set of all linear combinations of* $v_1, ..., v_n$. *Then W is a subspace of V.*

The subspace W is called the subspace generated by $v_1, ..., v_n$.

Definition 4. The vectors a and b are perpendicular or orthogonal to each other if $a \cdot b = 0$.

Definition 5. Let V be an arbitrary vector space, and let $v_1, ..., v_n$, be elements of V. The vectors $v_1, ..., v_n$ are linearly dependent if there exist elements $\alpha_1, ..., \alpha_n$ in $\mathbb{R}$ not all equal to 0, such that

$$\alpha_1 v_1 + \cdots + \alpha_n v_n = O.$$

If there do not exist such numbers, then we say that $v_1, ..., v_n$ are linearly independent (e.g. unitary basis). If elements $v_1, ..., v_n$ of V generate V and in addition are linearly independent then $\{v_1, ..., v_n\}$ constitute or form a basis of V.

Proposition 2. *Let V be a vector space. Let* $v_1, ..., v_n$ *be linearly independent elements of V. Let* $\alpha_1, ..., \alpha_n$ *and* $\beta_1, ..., \beta_n$, *be numbers. Suppose that we have*

$$v = \alpha_1 v_1 + \cdots + \alpha_n v_n = \beta_1 v_1 + \cdots + \beta_n v_n.$$

Then $\alpha_i = \beta_i$ *for* $i = 1, ..., n$.

If $v \in V$ is written as a linear combination $v = \alpha_1 v_1 + \cdots + \alpha_n v_n$, then the n-tuple $(\alpha_1, \cdots, \alpha_n)$ is uniquely determined by v, and is called the coordinates (coordinate vector) of v with respect to the given basis.

2 Dimension of a Vector Space

The main result of this section is that any two bases of a vector space have the same number of elements.

Proposition 3. *Let V be a vector space over the field K. Let* $\{v_1, ..., v_n\}$ *be a basis of V over K. Let* $w_1, ..., w_m$ *be elements of V, and assume that* $m > n$*. Then* $w_1, ..., w_m$ *are linearly dependent.*

Proposition 4. *Let V be a vector space and suppose that one basis has* n *elements, and another basis has* m *elements. Then* $n = m$*.*

Let V be a vector space having a basis consisting of n elements. We shall say that n is the **dimension** of V. If V consists of O alone, then V does not have a basis, and we shall say that V has dimension 0.

Proposition 5. *Let V be a vector space, and* $\{v_1, ..., v_n\}$ *a maximal set of linearly independent elements of V. Then* $\{v_1, ..., v_n\}$ *is a basis of V.*

Let U, W be subspaces of V. We define the sum of U and W to be the subset of V consisting of all sums $u + w$ with $u \in U,\ w \in W$. V is the direct sum of U and W if for every element $v \in V$, there exists a unique elements $u \in U,\ w \in W$, such that $u + w = v$.

3 Matrices

Let m,n be two integers ≥ 1. An array of numbers A,

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & & a_{2n} \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

is called a $(m \times n)$ matrix with entry or component, $(a_{ij})\,, i = 1, ..., m,\ j = 1, ..., n$; and i-th row $\mathrm{A}_i = (a_{i1}, ..., a_{in})$ and j-th column $\mathrm{A}^j = (a_{1j}, ..., a_{mj})$. If $m = n$, we say it is a square matrix.
Let $A = (a_{ij})\,,\ i = 1, ..., m$ and $j = 1, ..., n$, be an $m \times n$ matrix. Let $B = (b_{jk})\,,\ j = 1, ..., n$ and $k = 1, ..., s$, be an $n \times s$ matrix. We define the product AB to be the $m \times s$ matrix, whose ik-coordinate is

$$(AB)_{ik} = \sum_{j=1}^{n} a_{ij} b_{jk} = a_{i1} b_{1k} + \cdots + a_{in} b_{nk}$$

and

$$AB = \begin{pmatrix} A_1B^1 & \cdots & A_1B^s \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ A_mB^1 & \cdots & A_mB^s \end{pmatrix}.$$

Multiplication of matrices is therefore a generalization of the dot product. Hence matrix multiplication is associative and distributive, but not commutative. Special matrices: transpose, symmetric, zero, diagonal, unit matrix, lower and upper triangular matrices.

Definition 6. The matrix A is invertible if there exists a matrix B such that $BA = I$ and $AB = I$.

There is at most one such B, called the inverse of A and denoted by A^{-1}. Moreover $AA^{-1} = A^{-1}A = I$. A product AB of invertible matrices has an inverse. It is found by multiplying the individual inverses in reverse order: $(AB)^{-1} = B^{-1}A^{-1}$.We also say for an **invertible** A that A is **non-singular.** The transpose matrix of A is A^T and is found by

$$\left(A^T\right)_{ij} = A_{ji}.$$

The transpose of AB is $(AB)^T = B^T A^T$. The transpose of A^{-1} is $\left(A^{-1}\right)^T = \left(A^T\right)^{-1}$.

4 Linear Equations

The analysis of many economic models reduces to the study of systems of linear equations. There are essentially three ways of solving such systems: (i) substitution, (ii) elimination of variables (Gaussian and Gaussian-Jordan elimination), (iii) Cramer's rule.

Let $A = (a_{ij})$, let $b_1, ..., b_m$ be elements of $\Re$. Equations like

$$\begin{array}{ccccc} a_{11}x_1 & + & \cdots & a_{1n}x_n & = & b_1 \\ & & \cdots & & & \\ a_{m1}x_1 & + & \cdots & a_{mn}x_n & = & b_m \end{array}$$

are called linear equations. The system is said to be homogeneous if all the number $b_1, ..., b_m$ are equal to 0. The number n is called the number of unknowns and m is the number of equations:

$$x_1A^1 \quad + \quad \cdots \quad x_nA^n \quad = \quad b \;\Leftrightarrow Ax = b,$$

where A^i denotes the $i-th$ column. Hence the system is solvable if and only if the vector b can be expressed as a linear combination of the columns of A. We can describe this result geometrically: $Ax = b$ can be solved if and only if b lies in the subspace that is spanned by the column vectors: the column space, denoted by $\mathfrak{R}\,(A)$. For $b = 0$, the set of solutions to $Ax = 0$ is also a vector space of A, it is the nullspace of A, denoted by $\mathfrak{N}\,(A)$. We shall see that the nullspace is a subspace of the same "dimension" as the number of free variables.

Proposition 6. *Let*

$$x_1A^1 \quad + \quad \cdots \quad x_nA^n \quad = \quad 0.$$

Assume that $n > m$. Then the system has a nontrivial solution.

Every solution to $Ax = b$ is the sum of one particular solution and a solution to $Ax = 0$ with $n - r$ free variables as independent parameters. The number r is called the rank of the matrix A. The general solution therefore has the same dimension as the nullspace.

Proposition 7. *Let*

$$x_1A^1 \quad + \quad \cdots \quad x_nA^n \quad = \quad b.$$

Assume than $m = n$ and that the vectors $A^1, ..., A^n$ are linearly independent. Then the system has a unique solution.

The column space is often called the range of A, which is consistent with the usual idea of the range of a function f. Notice that the number of independent columns equals the number of independent rows: "row rank" ="column rank".

Definition 7. If A is an $n \times m$ matrix, then the columns $A^1, ..., A^n$ generate a subspace, whose dimension is called the column rank of A. The rows $A_1, ..., A_m$ generate a subspace whose dimension is called the row rank of A.

5 Linear Mappings

Definition 8. A function $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is called linear if $f(\alpha x + \beta y) = \alpha f(x) + \beta f(y)$ for all $x, y \in \mathbb{R}^n$ and $\alpha, \beta \in \mathbb{R}$.

Example 3. Let V be a finite dimensional space and let $\{v_1, ..., v_n\}$ be a basis of N. We define a map $F : V \rightarrow \mathbb{R}^n$ by associating to each element $v \in V$ its coordinate vector x with respect to the basis. This is a linear map.

5.1 Kernel and Image of a Linear Map

Definition 9. Let V, W be vector spaces , and let $F : V \rightarrow W$ be a linear map. We define the kernel of F to be the set of elements $v \in V$ such that $F(v) = O$.

We denote the kernel of F by $KerF$. The kernel of a linear map $F : V \rightarrow W$ is a subspace of V.

Proposition 8. *The following two conditions are equivalent:*
(i) the kernel of F is equal to $\{O\}$,
(ii) if v, w *are elements of V such that* $F(v) = F(w)$ *,then* $v = w$*, i.e. F is injective.*

Proposition 9. *Let* $F : V \rightarrow W$ *be a linear map whose kernel is* $\{O\}$*. If* $v_1, ..., v_n$ *are linearly independent elements of V, then* $F(v_1), ..., F(v_n)$ *are linearly independent elements of W.*

Definition 10. Let $F : V \rightarrow W$ be a linear map. The image of F is the set of elements $w \in W$ such that there exists an element $v \in V$ such that $F(v) = w$. The image of F is a subspace of W.

Let

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & & a_{2n} \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

be a $(m \times n)$ matrix. We can then associate with A a map

$$L_A : K^n \rightarrow K^m,$$

by letting

$$L_A(x) = Ax.$$

A set of solutions to a system of homogeneous linear equations

$$Ax = O,$$

is the kernel of the linear map L_A.

Proposition 10. *Let $L : K^n \to K^m$ be a linear map. Then there exists a unique matrix A such that $L = L_A$.*

Proposition 11. *Let A be an $n \times n$ matrix, and let $A^1, ..., A^n$ be its columns. Then A is invertible if and only if $A^1, ..., A^n$ are linearly independent.*

The solution sets to non-homogeneous equations are not subspaces. Not surprisingly, however, they do have a linear structure. they are affine subspaces, which is to say that they are translates of a subspace.

Definition 11. Let V be a subspace of $\mathbb{R}^n$ and let $y \in I + \mathbb{R}^n$ be a fixed vector. The set

$$\{x \in \mathbb{R}^n : x = y + v \text{ for some } v \in V\}$$

is called the set of translates of V by c and it is written by $y+V$. Subsets of $\mathbb{R}^n$ of the form $y+V$, where V is a subspace of $\mathbb{R}^n$, are called affine supspaces of $\mathbb{R}^n$.

Notice that the affine space $y+V$ has the same dimension as the vector space V.

Proposition 12. *Let $Ax = b$ be an $m \times n$ system of linear equations. Let $y \in \mathbb{R}^n$ ba a particular solution of this system. Then, every other solution y' of $Ax = b$ can be written as $y' = y + w$, where w is a vector in the nullspace of A.*

Theorem 1. ***(Fundamental Theorem of Linear Algebra).***

Let V be a vector space. Let $F : V \to W$ be a linear map of V into another space W. Let n be the dimension of V, q the dimension of the kernel of F, and s the dimension of the image of L. Then $n = q + s$:

$$\dim V = \dim KerF + \dim ImF.$$

Since we stated the theorem in an abstract metric space, it might be useful to restate it in directly in terms of the matrix which generates the linear map F, F_A. If $A_{m\times n}$ and $x_{n\times 1}$, then a typical element of the image set is $y_{m\times 1}$. Then we can restate Theorem 1 as:

Theorem 2. *Let A be an $m \times n$ matrix. Then*

$$\dim \mathfrak{N}(A) = n - \mathfrak{R}(A) = n - rank(A).$$

Notice that we defined $\mathfrak{R}(A)$ to be the column space of the matrix A, but we saw that it is also equal to the rank of the matrix. This theorem is one of the central theorems of linear algebra, because it determines the dimensions of the solution set of a linear system. We summarize the result of this section for a system $Ax = b$ of m linear equations in n unknowns. The columns space of the matrix A tells us whether this system of equations has a solution. The nullspace tells us how large the solution set is. In particular:

1. if rank$(A) = m$, the number of rows (=equations), then $Ax = b$ has a solution to every b,
2. if rank $(A) < m$, then $Ax = b$ will have a solution only for those $b \in \mathfrak{R}(A)$,
3. if rank$(A) = n$, then $\mathfrak{N}(A) = \{0\}$, and $Ax = b$ will have at most one solution for any b,
4. if rank$(A) < n$, then if $Ax = b$ has any solution at all, it will have an affine subspace of solutions of dimension $n - rank(A)$.

Anticipating the main result in orthogonality which will be presented in Theorem 3, it seems tempting to conclude that in the case of a square matrix, we could have

$$(KerF)^{\perp} = ImF,$$

however this result is only true if we require in addition to the squareness of the matrix it to be symmetric.

6 Orthogonality

A basis is algebraically a set of independent vectors that span the space. Geometrically, it is the set of coordinate axis. In $\mathbb{R}^n$ say are usually perpendicular. We need a basis to convert geometric constructions into algebraic calculations, and we need orthogonal basis to make these calculations simple. A furhter specialization makes the basis vectors to have length one.

Definition 12. Let V be a vector space over a field K. A scalar product or dot product on V is a function which to any pair $v, w \in V$ associates a scalar, denoted by $\langle v, w\rangle$ satisfying the following properties:
(i) $\langle v, w\rangle = \langle w, v\rangle$ for all $v, w \in V$,
(ii) $\langle u, v + w\rangle = \langle u, v\rangle + \langle u, w\rangle$ for all $u, v, w \in V$,
(iii) if $x \in K$, then $\langle xu, v\rangle = x\langle u, v\rangle$ and $\langle u, xv\rangle = x\langle u, v\rangle$.

Definition 13. Let V be a vector space with a scalar product. The elements $v, w \in V$ are orthogonal or perpendicular, $v \perp w$, if $\langle v, w\rangle = 0$.

If S is a subset of V, then we denote by $S^{\perp}$ the set of all elements $w \in V$, which are perpendicular to all elements of S.

Let $w \in V$, such that $\|w\| \neq 0$. For any v there exists a number c, such that $v - cw$ is perpendicular to w, with

$$c = \frac{\langle v, w\rangle}{\langle w, w\rangle}.$$

We call c the **component** of v along w, We call cw the **projection** of v along w.

Example 4. Let $V = \mathbb{R}^n$ with the usual scalar product, i.e. the dot product. If E_i is the $i - th$ unit vector, and $x = (x_1, ..., x_n)$ then the component of X along E_i is simply

$$xe_i = x_i,$$

that is the $i - th$ component of X.

The next theorem shows that $c_1v_1 + ... + c_nv_n$ gives the closest approximation to v as a linear combination of $v_1, ..., v_n$.

Proposition 13. *Let $v_1, ..., v_n$ be vectors which are mutually perpendicular and such that $\|v_i\| \neq 0$ for all i. Let $v \in V$, and let c_i be the component of v along v_i. Let $a_1, ..., a_n$ be numbers, then*

$$\left\| v - \sum_{k=1}^{n} c_k v_k \right\| \leq \left\| v - \sum_{k=1}^{n} a_k v_k \right\| .$$

Proposition 14. ***(Bessel's Inequality)*** *If $v_1, ..., v_n$ are mutually perpendicular unit vectors, and if c_i is the component of v along v_i, then*

$$\sum_{i=1}^{n} c_i^2 \leq \|v\|^2 .$$

Definition 14. A basis $\{v_1, ..., v_n\}$ of V is said to be orthogonal if its elements are mutually perpendicular. If in addition each element of the basis has norm 1, then the basis is called orthonormal.

Example 5. The standard unit vectors of $\mathbb{R}^n$ form an orthonormal basis of $\mathbb{R}^n$.

Proposition 15. *Let V be a vector space and W be a subspace of W, where $\{w_1, ..., w_m\}$ is an orthogonal basis of W. If W$\neq V$, then there exist elements of $w_{m+1}, ..., w_n$ of V such that $\{w_1, ..., w_n\}$ is an orthogonal basis of V.*

Theorem 3. *Let V be a vector space of dimension n. Let W be a subspace of V of dimension r. Then V is the direct sum of W and $W^{\perp}$, $W^{\perp}$ has dimension n-r:*

$$\dim W + \dim W^{\perp} = \dim V.$$

The space $W^{\perp}$ is called the orthogonal complement of W.

7 Determinants

Let

$$A = (a_{ij}) = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} .$$

We define its determinant according to the formula known as the expansion by a row. Let A_{ij} be the matrix obtained from A by deleting the $i-th$ row and the $j-th$ column, where $\det A_{ij}$ is called the $(i,j)-$minor of A, and $(-1)^{i+j}\det A_{ij}$ is the (i,j)th cofactor of A. Then

$$Det\,(A) = a_{11}Det\,(A_{11}) - a_{12}Det\,(A_{12}) + a_{13}Det\,(A_{13})\,,$$

and in general

Proposition 16. *Determinants satisfy the rule for expansion according to rows and columns. For any row A_i of the matrix A,*

$$D\,(A) = (-1)^{i+1}\,a_{i1}Det\,(A_{i1}) + \cdots + (-1)^{i+n}\,a_{in}Det\,(A_{in})\,,$$

or for any column A^j of the matrix A

$$D\,(A) = (-1)^{1+j}\,a_{1j}Det\,(A_{1j}) + \cdots + (-1)^{n+j}\,a_{nj}Det\,(A_{nj})\,,$$

We next list some useful properties of determinants for general square matrices A.

Proposition 17. *The determinant satisfies the following properties:*
(i) *as a function of each column vector, the determinant is linear,*
(ii) *if two columns, A^i, A^j are equal, then $Det\,(A) = 0$,*
(iii) $Det\,(A) = Det\left(A^T\right)$,
(iv) *if the $i-th$ and the $j-th$ columns are interchanged, then the determinant changes by a sign,*
(v) *if one adds a scalar multiple of one column to another, then the value of the determinant does not change,*
(vi) *the determinant of A is equal to the determinant of its transpose A^T.*

For arbitrary square matrices we have $\det(AB) = \det A \cdot \det B$. Furthermore if A is invertible, then $\det A^{-1} = \frac{1}{\det A}$.

Theorem 4. ***(Cramer's Rule).*** *Let $A^1, ..., A^n$ be column vectors such that*

$$D\left(A^1, ..., A^n\right) \neq 0.$$

Let B be a column vector. If $x_1, ..., x_n$ are numbers such that

$$x_1A^1 + \cdots + x_nA^n = B,$$

then for each $j = 1, ..., n$, we have

$$x_j = \frac{D\,(A^1, ..., B, ...A^n)}{D\,(A^1, ..., A^n)},$$

where B occurs in the j-th column instead of A^j.

Theorem 4 gives us an explicit way of finding the coordinates of B with respect to $A^1, ..., A^n$. In the language of linear equations, Theorem 4 allows us to solve explicitly in terms of determinants the system of n linear equations in n unknowns.

Proposition 18. *Let A be a square matrix of dimension n. The following conditions are equivalent:*
(i) The columns $A^1, ..., A^n$ are linearly dependent,
(ii) A is invertible,
(iii) $D(A^1, ..., A^n) \neq 0$.

A square matrix whose determinant is not equal to zero, or equivalently which admits an inverse, is called non-singular.

Proposition 19. *Let $A = (a_{ij})$ be an $n \times n$ matrix, and assume that $D(A) \neq 0$. Then A is invertible. Let E^j be the $j-th$ column unit vector, and let*

$$b_{ij} = \frac{D(A^1, ..., E^j, ..., A^n)}{D(A)},$$

where E^j occurs in the i-th place. Then the matrix $B = (b_{ij})$ is an inverse for A.

Based on Proposition 19 one can also show that

$$b_{ij} = \frac{(-1)^{i+j} D(A_{ij})}{D(A)}.$$

8 Eigenvector and Eigenvalues

Let V be a vector space and let

$$A : V \rightarrow V$$

be a linear map of V into itself.

Definition 15. An element $v \in V$ is called an eigenvector of A if there exists a number λ such that

$$Av = \lambda v,$$

and λ is an eigenvalue of A belonging to the eigenvector v.

Eigenvector and eigenvalue are sometimes called characteristic vector and characteristic value.

Proposition 20. *Let V be a vector space and let $A : V \to V$ be a linear map. Let $\lambda \in K$. Let V_λ be the subspace of V generated by all eigenvectors of A having λ as eigenvalue. Then every non-zero element of V_λ is an eigenvector of A having λ as eigenvalue.*

The subspace V_λ is called the eigenspace of A belonging to λ. If v_1, v_2 are eigenvectors of A with different eigenvalues $\lambda_1 \neq \lambda_2$ then of course $v_1 + v_2$ is not an eigenvector of A. In fact we have the following

Proposition 21. *Let V be a vector space and let $A : V \to V$ be a linear map. Let $v_1, ..., v_n$ be eigenvectors of A, with eigenvalues $\lambda_1, ..., \lambda_n$ respectively. Assume that these eigenvalues are distinct, $\lambda_i \neq \lambda_j$ if $i \neq j$. Then $v_1, ..., v_n$ are linearly independent.*

Definition 16. The sum of the diagonal entries of a $n \times n$ matrix A,

$$\sum_{i=1}^{n} a_{ii}$$

is called the trace of the matrix A.

Proposition 22. *The sum of n eigenvalues equals the trace of the matrix:*

$$\lambda_1 + ... + \lambda_n = a_{11} + ... + a_{nn}.$$

The product of n eigenvalues equals the determinant of A.

Proposition 23. *Let V be a finite dimensional vector space, and let λ be a number. Let $A : V \to V$ be a linear map. Then λ is an eigenvalue of A if and only if $A - \lambda I$ is not invertible.*

Definition 17. The characteristic polynomial P_A is the determinant

$$P_A(t) = Det\,(tI - A).$$

We say that $P_A(t)$ is the characteristic polynomial of the linear map A.

Proposition 24. *Let A be an $n \times n$ matrix. A number λ is an eigenvalue of A if and only if λ is a root of the characteristic polynomial of A.*

Proposition 24 gives us an explicit way of determining the eigenvalues of a matrix, provided that we can determine explicitly the roots of its characteristic polynomial.

8.1 Diagonalization

We start with the central theorem and then go beyond the case of n distinct eigenvectors to the Jordan normal form. We omit the subject of complex eigenvalues and simply mention that the theory developed for real eigenvalues applies exactly to complex eigenvalues.

Proposition 25. *Suppose that the $n \times n$ matrix A has n linearly independent eigenvectors. Then if these vectors are choosen to be the columns of the matrix P, it follows that*

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

where Λ is a diagonal matrix having the eigenvalues of A along its diagonal:

$$P^{-1}AP = \Lambda = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}.$$

If the matrix A has no repeated eigenvalues, the numbers $\lambda_1, ..., \lambda_n$ are distinct, then the eigenvectors are automatically independent. therefore any matrix with distinct eigenvalues can be diagonalized. Notice that the diagonalizing matrix P is not unique. Finally a word of caution. Not all matrices possess n linearly indepedent eigenvectors, and therefore not all matrices are diagonalizable. The following example illustrates this:

$$A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

Since $\lambda = 0$ is a double eigenvalue, we say its *algebraic multiplicity* is 2, however it has only a one-dimensional space of eigenvectors, hence its *geometric multiplicity* is 1.

Proposition 26. *The eigenvalues of A^k are $\lambda_1^k, ..., \lambda_n^k$, the $k-th$ powers of the eigenvalues of A. Each eigenvector of A is still an eigenvectro of A^k, and if P diagonalizes A it also diagonalizes A^k:*

$$\Lambda^k = \left(P^{-1}AP\right)\left(P^{-1}AP\right)\cdots\left(P^{-1}AP\right) = P^{-1}A^kP.$$

Any symmetric real matrix allows a particularly simple procedure of diagonalization

Proposition 27. ***(Spectral Theorem).*** *Any real symmetric matrix can be diagonalized by*

$$Q^T A Q = \Lambda$$

with the orthonormal eigenvectors in Q and the eigenvalues in Λ.

9 Minimum norm problems and the projection theorem

We turn now to our first optimization problem. This section is devoted to the problem of finding the best approximation of a vector x in $\mathbb{R}^n$ by a member m of a subspace of $\mathbb{R}^n$. Specifically, let M be a subspace of $\mathbb{R}^n$ and let x be a vector not in M. We are going to find a way to select a vector m^* in M such that

$$\|x - m^*\| \leq \|x - m\|$$

for all other m in M.

Lemma 1. ***(Pythagorean theorem).*** *For two vectors $x, y \in \mathbb{R}^n$, if $x \perp y$, then*

$$\|x + y\|^2 = \|x\|^2 + \|y\|^2 .$$

Lemma 2. ***(The parallelogram law).*** *For two vectors $x, y \in \mathbb{R}^n$,*

$$\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2 .$$

Theorem 5. ***(Projection theorem).*** *If M is a subspace of $\mathbb{R}^n$ and if $x \in IR^n$, then there is a unique $m^* \in M$ such that*

$$\|x - m^*\| \leq \|x - m\|$$

for all $m \in M$. The point m^* of M is characterized by $x - m^* \perp M$.

In view of the above result, given a vector x and a closed subspace M, the vector m^* such that $x - m^* \perp M$ is called the **orthogonal projection of x onto M.** We now generalize the results for minimum norm problems from linear varieties to convex sets. The first step is to extend to convex sets our earlier characterization of closest vectors to subspaces.

Theorem 6. *Suppose that C is a convex set in $\mathbb{R}^n$ and that y is a vector in $\mathbb{R}^n$ that is not in C. Then $x^* \in C$ is the closest vector in C to y (that is, $\|y - x^*\| \leq \|y - x\|$ for all $x \in C$) if and only if*

$$(y - x^*)(x - x^*) \leq 0$$

for all $x \in C$.

Corollary 1. *Suppose that M is a subspace of $\mathbb{R}^n$ and that $y \in \mathbb{R}^n$ is a vector not in M. Then $x^* \in M$ is the closest vector in M to y if and only if $y - x^* \perp M$.*

Theorem 7. *If C is a closed (convex or not) subset of $\mathbb{R}^n$ and if $y \in \mathbb{R}^n$ does not belong to C, then there is a vector $x^* \in C$ that is closest to y, that is*

$$\|y - x^*\| \leq \|y - x\|$$

for all $x \in C$.

Corollary 2. *Suppose that C is a closed convex subset of $\mathbb{R}^n$ and that y is a vector in $\mathbb{R}^n$ that is not in C.Then there is one and only one vector that is closest to y in C.*