Chapter 1

Review of linear algebra

- Linear functions and matrix–vector product
- Rank of a matrix and inversion
- Inner product
- Subspaces, basis, and dimension
- Eigenvalues and eigenvectors

1.1 Linear functions and matrix–vector product

Linear functions

Standard notation for a function f mapping a vector $x \in \mathbb{R}^n$ to a vector $y \in \mathbb{R}^m$ is

$$f : \mathbb{R}^n \to \mathbb{R}^m \qquad \text{or} \qquad f : x \mapsto y.$$

The value $y \in \mathbb{R}^m$ of f at $x \in \mathbb{R}^n$ is denoted by $y = f(x)$. Note that f and $f(x)$ are different objects—f is a function and $f(x)$ is a vector. Therefore, the statement "the function $f(x)$" is semantically wrong, despite the fact that its meaning is intuitively clear and is commonly used.

A function f is usually specified by an analytic expression, *e.g.*, $f(x) = x^2$, but it can be specified in other ways as well. For example, a function f can be defined by an algorithm that evaluates f for a given x or by a verbal description, *e.g.*, "$f(x)$ is the vector x rotated clockwise by α°". In system theory a function f is visualized by a box, called a *system*, that accepts as an *input* x and produces as an *output* y.

$$x \longrightarrow \boxed{\;f\;} \longrightarrow y$$

One can think of the system as a device or a signal processor that transforms energy or information. However, a system in system theory is an abstract object and is distinguished from a physical device.

By definition, f is a *linear function* if the following property holds:

$$f(\alpha x + \beta z) = \alpha f(x) + \beta f(z), \quad \text{for all } \alpha, \beta \in \mathbb{R} \text{ and } x, z \in \mathbb{R}^n.$$

An equivalent definition is that f satisfies the *homogeneity* and *superposition* properties

- homogeneity: $f(\alpha x) = \alpha f(x)$, for all $\alpha \in \mathbb{R}$, and $x \in \mathbb{R}^n$,
- superposition: $f(x+z) = f(x) + f(z)$, for all $x, z \in \mathbb{R}^n$.

Exercise problem 1. Show that rotation is a linear function.

□

Matrix–vector product

Partition a matrix $A \in \mathbb{R}^{m\times n}$ elementwise, column-wise, and row-wise, as follows

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} = \begin{bmatrix} | & & | \\ a_1 & \cdots & a_n \\ | & & | \end{bmatrix} = \begin{bmatrix} — & (a'_1)^\top & — \\ & \vdots & \\ — & (a'_m)^\top & — \end{bmatrix}.$$

The matrix–vector product $y = Ax$ can be written in three alternative ways corresponding to the three partitionings above

$$\begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^n a_{1j}x_j \\ \vdots \\ \sum_{j=1}^n a_{mj}x_j \end{bmatrix} = \sum_{j=1}^n a_j x_j = \begin{bmatrix} (a'_1)^\top x \\ \vdots \\ (a'_m)^\top x \end{bmatrix}.$$

For a given A, $y = Ax$ defines a function $f : x \mapsto y$. Matrix-vector product, however, is more than an example of a linear function. It is the only example in the sense that any linear function admits a representation in the form of a matrix times vector.

Exercise problem 2. Prove that the function $f : \mathbb{R}^n \to \mathbb{R}^m$ is linear if and only if there is a matrix $A \in \mathbb{R}^{m\times n}$, such that $f(x) = Ax$, for all $x \in \mathbb{R}^n$.

□

The matrix A is called a matrix representation of the function f, $f(x) = Ax$. Given a matrix representation A of a linear function f, the problem of evaluating the function $y = f(x)$ at a given point x is a matrix–vector multiplication $y = Ax$ problem.

Note 3. Formally one should make a distinction between a vector and a vector representation. A vector representation depends on the choice of basis and is therefore not unique. Similarly, a matrix representation of a linear function depends on the bases of the input space $\mathbb{R}^n$ and the output space $\mathbb{R}^m$ and is not unique, see (1.7) in Section 1.4.

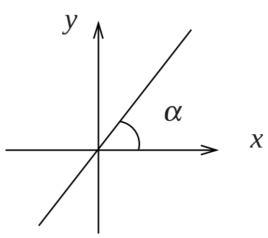
Exercise problem 4. Explain how to find a matrix representation of a linear function f, if f can be evaluated at arbitrary points $x \in \mathbb{R}^n$. Apply the procedure to the rotation function in $\mathbb{R}^n$.

□

Example 5. A scalar linear function of a scalar argument

$$y = \tan(\alpha)x, \qquad \text{where} \qquad \alpha \in [0, 2\pi)$$

is a line in the plain passing through the origin. Its matrix representation is the scalar $\tan(\alpha)$. Conversely, any line in the plain passing though the origin is a linear function.



Example 6. A scalar valued linear function of a vector argument $f : \mathbb{R}^n \to \mathbb{R}$ is given by $f(x) = a^\top x$, where $a \in \mathbb{R}^n$. (The expression $a^\top x$, *i.e.*, row vector times column vector, is called inner product, see Section 1.3.)

Example 7. The *identity function* $x = f(x)$, for all $x \in \mathbb{R}^n$, is a linear function represented by the $n \times n$ identity matrix

$$I_n := \begin{bmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{bmatrix} \in \mathbb{R}^{n\times n}.$$

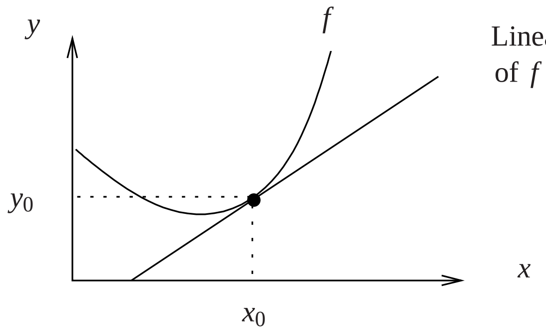
Consider a differentiable function $f : \mathbb{R}^n \to \mathbb{R}^m$. Then for a given $x_0 \in \mathbb{R}^n$

$$y = f(x_0 + \widetilde{x}) \approx f(x_0) + A\widetilde{x}, \qquad \text{where} \quad A = [a_{ij}] = \left[\left. \frac{\partial f_i}{\partial x_j} \right|_{x_0} \right].$$

($\partial f_i/\partial x_j$ is the partial derivative of f_i with respect to x_j and the matrix A of the partial derivatives is called the Jacobian of f.) When the input deviation $\widetilde{x} = x - x_0$ is "small", the output deviation

$$\widetilde{y} := y - f(x_0) =: y - y_0$$

is approximately the linear function $\widetilde{y} = A\widetilde{x}$ of $\widetilde{x}$, called the linear approximation of f at x_0.



Linear approximation of $f : x \mapsto y$ at $x = x_0$

1.2 Rank of a matrix and inversion

The set of vectors $\{a_1,\dots,a_n\}$ is *linearly independent* if the only linear combination of these vectors that is equal to the zero vector is the trivial linear combination with all weights equal to zero, *i.e.*,

$$x_1a_1+\cdots+x_na_n=0 \qquad \Longrightarrow \qquad x_1=\cdots=x_n=0.$$

Linear independence means that non of the vectors a_i, for $i=1,\dots,n$, can be expressed as a linear combination of the remaining vectors. Vice verse, in a linearly dependent set of vectors $\{a_1,\dots,a_n\}$ at least one vector is equal to a linear combination of the others. This means that in a linearly dependent set of vectors, there is redundant information.

The rank $\operatorname{rank}(A)$ of the matrix $A\in\mathbb{R}^{m\times n}$ is the number of linearly independent columns (or rows) of A and zero if A is the zero matrix. Obviously,

$$0\le \operatorname{rank}(A)\le \min(m,n)$$

The matrix A is

- *full row rank* if $\operatorname{rank}(A)=m$,
- *full column rank* if $\operatorname{rank}(A)=n$, and
- *full rank* if it is either full row rank or full column rank.

Full row and column rank of A turns out to be a necessary and sufficient condition for, respectively, existence of solution of the system $Ax=y$, for *any* given $y\in\mathbb{R}^m$, and uniqueness of a solution of $Ax=y$ for any given $y\in\mathbb{R}^m$. (Existence of solution of $Ax=y$, does depend on both A and y. However, assuming that the system $Ax=y$ is solvable, the uniqueness of a solution x depends only on A.)

Exercise problem 8. Prove that the matrix A being full row rank is equivalent to the system of equations $Ax=y$ having a solution for *any* $y\in\mathbb{R}^m$.

Exercise problem 9. Prove that the matrix A being full column rank is equivalent to uniqueness of a solution x to the system $y=Ax$, where $y=A\bar{x}$ for some $\bar{x}\in\mathbb{R}^n$.

A system theoretic interpretation of A being full row rank is that the system defined by $y=Ax$ has no redundant outputs, *i.e.*, none of the components of y can be inferred from the others. An interpretation of A being full column rank is that the there exists an inverse system, *i.e.*, a mechanism (which is also a system) to infer the the input from the output.

Next we consider the inversion problem: given $y\in\mathbb{R}^m$, find x, such that $y=Ax$. We distinguish three cases depending on the shape of the matrix A (square, more rows than columns, or more columns than rows) and in all cases we assume that A is full rank.

- If $m=n=\operatorname{rank}(A)$, then there exists a matrix A^{-1}, such that

  $$AA^{-1}=A^{-1}A=I_m. \tag{1.1}$$

  Then for all $y\in\mathbb{R}^m$

  $$y=\underbrace{(AA^{-1})}_{I_m}y=A\underbrace{(A^{-1}y)}_{x}=Ax.$$

  In this case, the inversion problem is solvable and the solution is unique.

  Exercise problem 10. Prove the fact that $m=n=\operatorname{rank}(A)$ implies existance of a matrix A^{-1}, such that (1.1).

  Exercise problem 11. Find a matrix representation of a linear function f, from given values $y_1,\dots,y_n$ of f at given points $x_1,\dots,x_n$. When is this problem solvable?

  □

- If $m \geq n = \operatorname{rank}(A)$, *i.e.*, A is full column rank, the inversion problem may have no solution. In such cases, an approximate solution may be desirable. The least-squares approximate solution minimizes the 2-norm

$$\|e\|_2 := \sqrt{e^\top e} = \sqrt{e_1^2 + \cdots + e_n^2}, \tag{1.2}$$

of the *approximation error* (or *residual*)

$$e := y - Ax.$$

The least-squares approximation problem is

$$\text{minimize } \|e\|_2 \quad \text{subject to} \quad Ax = y + e$$

and the solution is given by the famous formula

$$x_{\text{ls}} = (A^\top A)^{-1} A^\top y =: A_{\text{ls}}^{\text{L}} y. \tag{1.3}$$

Note that x_{ls} is a linear function $A_{\text{ls}}^{\text{L}} y$ of y. It is called the least squares approximate solution of the system of equations $Ax = y$. If $y = A\bar{x}$, for some $\bar{x} \in \mathbb{R}^n$, x_{ls} is an exact solution, *i.e.*, $x_{\text{ls}} = \bar{x}$.

Note 12 (Left inverse). Any matrix A^{L}, satisfying the property $A^{\text{L}} A = I_n$ is called a *left inverse* of A. Left inverse of A exists if and only if A is full column rank. If $m > n$, the left inverse is nonunique. If $m = n$, the left inverse is unique and is equal to the inverse. The matrix A_{ls}^{L} is a left inverse of A. Moreover, it is the smallest left inverse, in the sense that it minimizes the Frobenius norm

$$\|A^{\text{L}}\|_{\text{F}} := \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} (a_{ij}^{\text{L}})^2}$$

over all left inverses A^{L} of A.

Exercise problem 13. Prove that $A_{\text{ls}}^{\text{L}} = \arg\min_{A^{\text{L}}} \|A^{\text{L}}\|_{\text{F}}$ subject to $A^{\text{L}} A = I$.

□

- If $n \geq m = \operatorname{rank}(A)$, *i.e.*, A is full row rank, the inversion problem has infinitely many solutions. The set of all solutions is

$$\{\, x \mid Ax = y \,\} = \{\, x_{\text{p}} + z \mid Az = 0 \,\}, \qquad \text{where} \quad Ax_{\text{p}} = y,$$

i.e., x_{p} is a particular solution of $Ax = y$ and z is a parameter describing the nonuniquness of the solution. The *least-norm solution* is

$$\text{minimize } \|x\|_2 \quad \text{subject to} \quad Ax = y.$$

It is given by the following closed form expression

$$x_{\text{ln}} = A^\top (AA^\top)^{-1} y =: A_{\text{ln}}^{\text{R}} y. \tag{1.4}$$

Note 14 (Right inverse). Any matrix A^{R}, satisfying the property $AA^{\text{R}} = I_m$ is called a *right inverse* of A. Right inverse of A exists if and only if A is full row rank. If $m < n$, the right inverse is nonunique. If $m = n$, the right inverse is unique and is equal to the inverse. The matrix A_{ln}^{R} is a right inverse of A. Moreover, it can be shown that it is the smallest right inverse, in the sense that it minimizes the Frobenius norm $\|A^{\text{R}}\|_{\text{F}}$ over all right inverses A^{R} of A.

Exercise problem 15. Prove that $A_{\text{ln}}^{\text{R}} = \arg\min_{A^{\text{R}}} \|A^{\text{R}}\|_{\text{F}}$ subject to $AA^{\text{R}} = I$.

□

Note 16 (Inversion problem in the singular case). If $A \in \mathbb{R}^{m \times n}$ is rank deficient (or almost rank deficient), the inversion problem is called *ill-posed* (or *ill-conditioned*). In this case, the inverse (assuming A is square) does not exist. Also the least-squares (1.3) (assuming $m > n$) and the least-norm (1.4) (assuming $m < n$) formulas make no sense (because the indicated inverses do not exist). A general solution to the inversion problem, which is independent of size and rank assumptions on A, is given by $\hat{x} = A^+ y$, what $A^+ \in \mathbb{R}^{n \times m}$ is the *pseudo-inverse* of A. A related approach for solving ill-posed and ill-conditioned inverse problems is *regularization*.

1.3 Inner product

The inner product $\langle a,b\rangle = \langle b,a\rangle$ of two vectors $a,b \in \mathbb{R}^m$ is defined by

$$\langle a,b\rangle := a^\top b = \sum_{i=1}^m a_i b_i.$$

The matrix–matrix product BA, where $B : \mathbb{R}^{p\times m}$ and $A : \mathbb{R}^{m\times n}$, can be viewed as a collection of pn inner products (between the rows of B and the columns of A)

$$BA = \begin{bmatrix} — & (b_1')^\top & — \\ & \vdots & \\ — & (b_p')^\top & — \end{bmatrix} \begin{bmatrix} | & & | \\ a_1 & \cdots & a_n \\ | & & | \end{bmatrix} = \begin{bmatrix} \langle b_1' a_1\rangle & \cdots & \langle b_1', a_n\rangle \\ \vdots & & \vdots \\ \langle b_p', a_1\rangle & \cdots & \langle b_p', a_n\rangle \end{bmatrix}.$$

The *Gram matrix* of the vectors $a_1,\dots,a_n$ is defined by

$$\begin{bmatrix} a_1^\top \\ \vdots \\ a_n^\top \end{bmatrix} \begin{bmatrix} a_1 & \cdots & a_n \end{bmatrix} = A^\top A.$$

The Gram matrix $H := A^\top A$ is *symmetric, i.e.*, $H = H^\top$ and *positive semidefinite, i.e.*, $x^\top Hx \geq 0$, for all $x \in \mathbb{R}^n$. A matrix H is called *positive definite* if $x^\top Hx > 0$, for all $x \in \mathbb{R}^m$.

Exercise problem 17. The Gram matrix $A^\top A$ is positive definite if and only if A is full column rank.

□

The Cauchy-Schwarz inequality relates the inner product with the product of the 2-norms

$$\langle a,b\rangle \leq \|a\|\|b\|. \tag{1.5}$$

Equality holds in (1.5) if and only if $b = \alpha a$, for some $\alpha \in \mathbb{R}$ or $b = 0$.

Exercise problem 18. Prove (1.5).

□

Exercise problem 19 (Optimization of a linear function over the unit ball). Show that the solution of the problem, given $a \in \mathbb{R}^n$,

$$\text{maximize} \quad (\text{over } x) \quad a^\top x \quad \text{subject to} \quad \|x\| \leq 1$$

is $x_{\text{opt}} = a/\|a\|$.

□

The angle between the vectors $a,b \in \mathbb{R}^n$ is defined as

$$\angle(a,b) = \cos^{-1}\frac{a^\top b}{\|a\|\|b\|}.$$

- $a \neq 0$ and b are *aligned* if $b = \alpha a$, for some $\alpha \geq 0$ (in this case, $\angle(a,b) = 0$).
- $a \neq 0$ and b are *opposite* if $b = -\alpha a$, for some $\alpha \geq 0$ (in this case, $\angle(a,b) = \pi$).
- a and b are *orthogonal*, denoted $a \perp b$, if $a^\top b = 0$ (in this case, $\angle(a,b) = \pi/2$).

1.4 Subspace, basis, and dimension

The set $\mathscr{A} \subset \mathbb{R}^n$ is a *subspace* of a vector space $\mathbb{R}^n$ if $\mathscr{A}$ is a vector space itself, *i.e.*,

$$a, b \in \mathscr{A} \quad \Longrightarrow \quad \alpha a + \beta b \in \mathscr{A}, \quad \text{for all } \alpha, \beta \in \mathbb{R}.$$

The set $\{a_1, \ldots, a_n\}$ is a *basis* for the subspace $\mathscr{A}$ if the following hold:

- $a_1, \ldots, a_n$ *span* $\mathscr{A}$, *i.e.*,

$$\mathscr{A} = \operatorname{span}(a_1, \ldots, a_n) := \{x_1 a_1 + \cdots + x_n a_n \mid x_1, \ldots, x_n \in \mathbb{R}\} = \{ \begin{bmatrix} a_1 & \cdots & a_n \end{bmatrix} x \mid x \in \mathbb{R}^n \}$$

- $\{a_1, \ldots, a_n\}$ is an independent set of vectors.

Exercise problem 20. The number of basis vectors does not depend on the choice of the basis

□

The number of basis vectors of a subspace is invariant of the choice of the basis and is called the *dimension* of the subspace. The dimension of $\mathscr{A}$ is denoted by $\dim(\mathscr{A})$.

The *kernel* (also called *null space*) of the matrix $A \in \mathbb{R}^{m \times n}$ is the set of vectors mapped to zero by $f(x) := Ax$, *i.e.*,

$$\ker(A) := \{x \in \mathbb{R}^n \mid Ax = 0\}.$$

Adding a vector z in the kernel of A to a solution x_{p} of the system $Ax = y$ produces another solution of the system, *i.e.*, if $Ax_{\mathrm{p}} = y$, then $y = A(x_{\mathrm{p}} + z)$, for all $z \in \ker(A)$. From a parameter estimation point of view, $\ker(A)$ is the uncertainty in finding the parameter x, given the observation y. From a control point of view, $\ker(A)$ is the freedom in the control x that achieves the desired output y. If $\ker(A) = \{0\}$, the function $f(x) := Ax$ is called *one-to-one* .

Exercise problem 21. Show that $\ker(A) = \{0\}$ if and only if A is full column rank.

□

The image (also called column span or range) of the matrix $A^{m \times n}$ is the set of vectors that can be obtained as an output of the function $f(x) := Ax$, *i.e.*,

$$\operatorname{image}(A) := \{Ax \in \mathbb{R}^m \mid x \in \mathbb{R}^n\}.$$

Obviously, $\operatorname{image}(A)$ is the span of the columns of A. Alternatively, $\operatorname{image}(A)$ is the set of vectors y for which the system $Ax = y$ has a solution. If $\operatorname{image}(A) = \mathbb{R}^m$, the function $f(x) := Ax$ is called *onto*.

Exercise problem 22. Show that $\operatorname{image}(A) = \mathbb{R}^m$ if and only if A is full row rank.

□

For a matrix $A \in \mathbb{R}^{m \times n}$, $\ker(A)$ is a subspace of $\mathbb{R}^n$ and $\operatorname{image}(A)$ is a subspace of $\mathbb{R}^n$.

Exercise problem 23. Show that

$$\dim\big(\operatorname{image}(A)\big) = \operatorname{rank}(A) \qquad \text{and} \qquad \operatorname{col}\dim(A) - \dim\big(\ker(A)\big) = \operatorname{rank}(A). \tag{1.6}$$

□

A direct consequence of (1.6) is the so called preservation of dimensions theorem (in $\mathbb{R}^n$)

$$\dim\big(\ker(A)\big) + \dim\big(\operatorname{image}(A)\big) = \operatorname{col}\dim(A).$$

Note that

$$\operatorname{rank}(A) = \dim\big(\operatorname{image}(A)\big) = \operatorname{rank}(A^\top) = \dim\big(\operatorname{image}(A^\top)\big).$$

$\operatorname{image}(A)$ is the span of the columns of A and $\operatorname{image}(A^\top)$ is the span of the rows of A. The former is a subspace of $\mathbb{R}^m$ and the latter is a subspace of $\mathbb{R}^n$, they are equal (to the rank of A). By defining the *left kernel* of A,

$$\operatorname{left}\ker(A) := \{y \in \mathbb{R}^m \mid y^\top A = 0\},$$

we have a preservation of dimensions theorem for $\mathbb{R}^m$

$$\dim\left(\text{left}\ker(A)\right) + \dim\left(\text{image}(A^\top)\right) = \text{row}\dim(A).$$

The *standard basis* vectors in $\mathbb{R}^n$ are the vectors

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \dots, \quad e_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}.$$

Note that $e_1, \dots, e_n$ are the columns of the identity matrix I_n. The elements of a vector $x \in \mathbb{R}^n$ are the coordinates of x with respect to a basis understood from the context. The default basis is the standard basis $e_1, \dots, e_n$. Suppose that a new bases is given by the columns $t_1, \dots, t_n$ of a matrix $T \in \mathbb{R}^{n\times n}$. Since $\{t_1, \dots, t_n\}$ is a basis, the set is linearly independent. Therefore, the matrix T is nonsingular. Vice verse, any nonsingular matrix $T \in \mathbb{R}^{n\times n}$ defines a basis for $\mathbb{R}^n$. Let the coordinates of x in the basis T be $\widetilde{x}_1, \dots, \widetilde{x}_n$. Then

$$x = \widetilde{x}_1 t_1 + \cdots + \widetilde{x}_n t_n = T\widetilde{x} \quad \implies \quad \widetilde{x} = T^{-1}x,$$

i.e., the inverse matrix T^{-1} transforms the standard basis coordinates x into the T-basis coordinates $\widetilde{x}$.

Consider, now a linear operator $f : \mathbb{R}^n \to \mathbb{R}^n$ (*i.e.*, a function mapping from a space to the same space), given by $f(x) = Ax$, $A \in \mathbb{R}^{n\times n}$. The matrix A is a representation of f in a basis that is understood from the context. By default this is the standard basis. Changing the standard basis to a basis defined by the columns of a nonsingular matrix $T \in \mathbb{R}^{n\times n}$, the matrix representation of f changes to $T^{-1}AT$, *i.e.*,

$$\widetilde{y} = (T^{-1}AT)\widetilde{x}. \tag{1.7}$$

The mapping $A \mapsto T^{-1}AT$, defined by T, is called a *similarity transformation* of A.

1.5 Eigenvalues and eigenvectors

The (complex) number $\lambda \in \mathbb{C}$ is an *eigenvalue* of the square matrix $A \in \mathbb{R}^{n\times n}$ if there is a (complex) nonzero vector $v \in \mathbb{C}^n$, called an *eigenvector* associated to λ, such that $Av = \lambda v$. Equivalently, λ is an eigenvalue of A if the matrix $\lambda I_n - A$ is singular. If (λ, v) is an eigenpair of A, the action of A on vectors in $\text{span}(v)$ is equivalent to a scalar multiplication by λ.

The *characteristic polynomial* of A is

$$p_A(\lambda) := \det(\lambda I_n - A).$$

The degree of p_A, denoted $\deg(p_A)$, is equal to n and p is *monic*, *i.e.*, the coefficient of the highest order term is one.

Exercise problem 24. Prove that the scalar $\lambda \in \mathbb{C}$ is an eigenvalue of A if and only if λ is a root of p_A.

□

The *geometric multiplicity* of λ is the dimension of the kernel of $\lambda I_n - A$. The *algebraic multiplicity* of λ is the multiplicity of the root λ of p_A. A matrix that has an eigenvalue for which the geometric and algebraic multiplicities do not coincide is called *defective*.

Suppose that $\{v_1, \dots, v_n\}$ is a linearly independent set of eigenvectors of $A \in \mathbb{R}^{n\times n}$, *i.e.*,

$$Av_i = \lambda_i v_i, \qquad \text{for } i = 1, \dots, n.$$

Written in a matrix form, the above set of equations is

$$A\underbrace{\begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}}_{V} = \underbrace{\begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}}_{V} \underbrace{\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}}_{\Lambda}.$$

The matrix V has as columns the eigenvectors and is nonsingular since by assumption $\{v_1,\ldots,v_n\}$ is a linearly independent set. Then

$$AV = V\Lambda \quad \Longrightarrow \quad V^{-1}AV = \Lambda,$$

i.e., we obtain a similarity transformation (defined by the matrix $T := V^{-1}$) that *diagonalizes* A. Conversely, if there is a nonsingular $V \in \mathbb{C}^{n\times n}$, such that

$$V^{-1}AV = \Lambda$$

then $Av_i = \lambda_i v_i$, for $i = 1,\ldots,n$, and therefore $\{v_1,\ldots,v_n\}$ is a linearly independent set of eigenvectors.

The matrix $A \in \mathbb{R}^{n\times n}$ is *diagonalizable* if

- there is a nonsingular matrix T, such that TAT^{-1} is diagonal, or
- there is a set of n linearly independent eigenvectors of A.

The set of *defective* matrices corresponds to the set of matrices that are not diagonalizable. The eigenvalues of a matrix being distinct implies that the matrix is diagonalizable, however, the converse is not true (consider, for example, the the identity matrix). A prototypical example of a defective matrix is what is called the *Jordan block*

$$J_\lambda := \begin{bmatrix} \lambda & 1 & & \\ & \lambda & \ddots & \\ & & \ddots & 1 \\ & & & \lambda \end{bmatrix}.$$

A generalization of the eigenvalue decomposition $TAT^{-1} = \Lambda$ for defective matrices is the Jordan canonical form

$$TAT^{-1} = \operatorname{diag}(J_{\lambda_1},\ldots,J_{\lambda_k}),$$

where $\lambda_1,,\ldots,\lambda_k$ are the distinct eigenvalues of A.

Exercise problem 25. Show that the eigenvalues of a symmetric matrix $A \in \mathbb{R}^{n\times n}$ are real and the eigenvectors can be chosen to form an orthonormal set, *i.e.*, be orthogonal to each other and have unit norm.

□

1.6 Summary

- f is *linear* if homogeneity and superposition holds, *i.e.*, $f(\alpha x + \beta v) = \alpha f(x) + \beta f(v)$
- f is linear if and only if there is a matrix A, such that $f(x) = Ax$
- *image* (column span or range) of $A \in \mathbb{R}^{m\times n}$ is the set $\operatorname{image}(A) := \{Ax \mid x \in \mathbb{R}^n\}$
- *kernel* (or null space) of $A \in \mathbb{R}^{m\times n}$ is the set $\ker(A) := \{x \in \mathbb{R}^m \mid Ax = 0\}$
- $\mathscr{A} \subset \mathbb{R}^n$ is a *subspace* if $\alpha a + \beta b \in \mathscr{A}$ for all $a, b \in \mathscr{A}$
- a *basis* of a subspace is a set of linearly independent vectors that span the subspace
- the *dimension* of a subspace is the number of basis vectors
- the $\operatorname{image}(A)$ and the $\ker(A)$ of any matrix A are subspaces
- the *rank* of A is the number of linearly independent rows (or columns)
- $\dim\big(\operatorname{image}(A)\big) = \operatorname{rank}(A)$ and $\operatorname{col\,dim}(A) - \dim\big(\ker(A)\big) = \operatorname{rank}(A)$
- A is *full row rank* if $\operatorname{rank}(A) = \operatorname{row\,dim}(A)$