

Linear Algebra

Definition 1.1

A vector space (or linear space) V over a field \mathbb{F} consists of a set on which two operations (called addition and multiplication respectively here) are defined so that;

- (A) (V is Closed Under Addition) For all $\mathbf{x}, \mathbf{y} \in V$, there exists a unique element $\mathbf{x} + \mathbf{y} \in V$.
- (M) (V is Closed Under Scalar Multiplication) For all elements $a \in \mathbb{F}$ and elements $\mathbf{x} \in V$, there exists a unique element $a\mathbf{x} \in V$.

Such that the following properties hold:

- (VS 1) (Commutativity of Addition) For all $\mathbf{x}, \mathbf{y} \in V$, we have $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$.
- (VS 2) (Associativity of Addition) For all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$, we have $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$.
- (VS 3) (Existence of The Zero/Null Vector) There exists an element in V denoted by $\mathbf{0}$, such that $\mathbf{x} + \mathbf{0} = \mathbf{x}$ for all $\mathbf{x} \in V$.
- (VS 4) (Existence of Additive Inverses) For all elements $\mathbf{x} \in V$, there exists an element $\mathbf{y} \in V$ such that $\mathbf{x} + \mathbf{y} = \mathbf{0}$.
- (VS 5) (Multiplicative Identity) For all elements $x \in V$, we have $1\mathbf{x} = \mathbf{x}$, where 1 denotes the multiplicative identity in \mathbb{F} .
- (VS 6) (Compatibility of Scalar Multiplication with Field Multiplication) For all elements $a, b \in \mathbb{F}$ and elements $\mathbf{x} \in V$, we have $(ab)\mathbf{x} = a(b\mathbf{x})$.
- (VS 7) (Distributivity of Scalar Multiplication over Vector Addition) For all elements $a \in \mathbb{F}$ and elements $\mathbf{x}, \mathbf{y} \in V$, we have $a(\mathbf{x} + \mathbf{y}) = a\mathbf{x} + a\mathbf{y}$.
- (VS 8) (Distributivity of Scalar Multiplication over Field Addition) For all elements $a, b \in \mathbb{F}$, and elements $\mathbf{x} \in V$, we have $(a + b)\mathbf{x} = a\mathbf{x} + b\mathbf{x}$.

Theorem 1.2

Let V be a vector space and W a subset of V . Then W is a subspace of V iff the following 3 conditions hold for the operations defined in V .

- (a) $\mathbf{0} \in W$
- (b) $\mathbf{x} + \mathbf{y} \in W$ whenever $\mathbf{x} \in W$ and $\mathbf{y} \in W$.
- (c) $c\mathbf{x} \in W$ whenever $c \in \mathbb{F}$ and $\mathbf{x} \in W$.

Definition 1.3

A subset S of a vector space V *generates* (or *spans*) V iff $\text{span}(S) = V$. In this case, we also say that the vectors of S generate (or span) V .

Definition 1.4

Let V be a vector space and S a nonempty subset of V . A vector $v \in V$ is called a *linear combination* of vectors of S iff there exists a finite number of vectors u_1, u_2, \dots, u_n in S and scalars a_1, a_2, \dots, a_n in \mathbb{F} such that

$$v = \sum_{i=1}^n a_i u_i.$$

In this case we also say that v is a linear combination of u_1, u_2, \dots, u_n and call a_1, a_2, \dots, a_n the

coefficients of the linear combination

Definition 1.5

A subset S of a vector space V is called *linearly dependent* iff there exists a finite number of distinct vectors u_1, u_2, \dots, u_n in S and scalars a_1, a_2, \dots, a_n not all zero, such that

$$a_1 u_1 + a_2 u_2 + \dots + a_n u_n = \mathbf{0}.$$

Definition 1.6

A *basis* β for a vector space V is a linearly independent subset of V that generates V . If β is a basis for V , we also say that the vectors of β form a basis for V .

Theorem 1.7: The Rank-Nullity Theorem.

For any vector spaces V and W , and a linear operator $T: V \rightarrow W$, it holds that

$$\text{rank}(T) + \text{nullity}(T) = \dim(V).$$

General Information

- Let \mathbf{A} be an $m \times n$ matrix, and \mathbf{a}_j its j th column. For any $\mathbf{x} = (x_1 \ x_2 \ \dots \ x_n)^\top$,

$$\mathbf{A}\mathbf{x} = \sum_{j=1}^n x_j \mathbf{a}_j.$$

- Let \mathbf{A} and \mathbf{B} be matrices having n rows. For any matrix \mathbf{M} with n columns, we have

$$(\mathbf{A} \mid \mathbf{B}) = (\mathbf{MA} \mid \mathbf{MB}).$$

Definition 1.8

A system $\mathbf{Ax} = \mathbf{b}$ is *homogeneous* iff $\mathbf{b} = \mathbf{0}$; otherwise it is *nonhomogeneous*.

Theorem 1.9

For any matrix, its row space, column space, and rank are identical.

Theorem 1.10

A system $\mathbf{Ax} = \mathbf{b}$ of m linear equations in n unknowns has a solution space of dimension $n - \text{rank}(\mathbf{A})$.

Definition 1.11

A system $\mathbf{Ax} = \mathbf{b}$ of linear equations is *consistent* iff its solution set is nonempty; otherwise it is *inconsistent*.

Theorem 1.12: The Rouché-Capelli Theorem.

A system $\mathbf{Ax} = \mathbf{b}$ is consistent iff $\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{A}|\mathbf{b})$.

Definition 1.13

A matrix is said to be in *reduced row echelon form* iff

- Any row containing a nonzero entry precedes any row in which all the entries are zero (if any).
- The first nonzero entry in each row is the only nonzero entry in its column.
- The first nonzero entry in each row is 1 and it occurs in a column to the right of the first

nonzero entry in the preceding row.

- Gaussian elimination.
 - In the forward pass, the augmented matrix is transformed into an upper triangular matrix in which the first nonzero entry of each row is 1 and it occurs in a column to the right of the first nonzero entry of each preceding row.
 - In the backward pass, the upper triangular matrix is transformed into reduced row echelon form by making the first nonzero entry of each row the only nonzero entry of its column.
- Gaussian elimination always reduces a matrix to its rref form.

- Let \mathbf{A} be an invertible $n \times n$ matrix. Then, for some elementary row matrices \mathbf{E}_1 to \mathbf{E}_p ,

$$\mathbf{E}_p \mathbf{E}_{p-1} \dots \mathbf{E}_1 (\mathbf{A} \mid \mathbf{I}_n) = \mathbf{A}^{-1} (\mathbf{A} \mid \mathbf{I}_n) = (\mathbf{I}_n \mid \mathbf{A}^{-1}).$$

In other words, we can perform Gaussian elimination, so that $(\mathbf{A} \mid \mathbf{I}_n) \rightarrow (\mathbf{I}_n \mid \mathbf{A}^{-1})$.

- Let $\mathbf{A} := (\mathbf{a}_1 \ \mathbf{a}_2 \ \dots \ \mathbf{a}_n)$ be $m \times n$ matrix, and $\mathbf{A}' := (\mathbf{a}'_1 \ \mathbf{a}'_2 \ \dots \ \mathbf{a}'_n)$ its rref. Then, $\{\mathbf{a}_{k_1}, \mathbf{a}_{k_2}, \dots, \mathbf{a}_{k_m}\}$ is linearly independent iff $\{\mathbf{a}'_{k_1}, \mathbf{a}'_{k_2}, \dots, \mathbf{a}'_{k_m}\}$ is. Moreover, the row space of \mathbf{A} and \mathbf{A}' are clearly identical.
- Finding a basis for an intersection of subspaces. Let V and W be subspaces of \mathbb{F}^n generated by the columns of the $n \times m$ matrix \mathbf{A} and $n \times k$ matrix \mathbf{B} , respectively. Find a basis for the subspace $V \cap W$.

1. First notice that $\mathbf{v} \in V \cap W$ iff

$$\mathbf{v} = \mathbf{A}\mathbf{x}_1 = \mathbf{B}\mathbf{x}_2$$

for some $\mathbf{x}_1 \in \mathbb{F}^m$ and $\mathbf{x}_2 \in \mathbb{F}^k$. That is,

$$(\mathbf{A} \ \mathbf{B}) \begin{pmatrix} \mathbf{x}_1 \\ -\mathbf{x}_2 \end{pmatrix} = \mathbf{0}.$$

So, equivalently, we write

$$(\mathbf{A} \ \mathbf{B}) \mathbf{y} = \mathbf{0}.$$

for some $\mathbf{y} \in \mathbb{F}^{m+k}$. As such, by row reducing $(\mathbf{A} \ \mathbf{B})$, we find a basis

$$\beta := \left\{ \begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}'_1 \end{pmatrix}, \begin{pmatrix} \mathbf{u}_2 \\ \mathbf{u}'_2 \end{pmatrix}, \dots, \begin{pmatrix} \mathbf{u}_r \\ \mathbf{u}'_r \end{pmatrix} \right\},$$

where $\mathbf{u}_i \in \mathbb{F}^m$ and $\mathbf{u}'_i \in \mathbb{F}^k$. Now, a generating set for $V \cap W$ is

$$\Gamma := \{\mathbf{A}\mathbf{u}_1, \mathbf{A}\mathbf{u}_2, \dots, \mathbf{A}\mathbf{u}_r\}.$$

Alternatively, another generating set for $V \cap W$ is

$$\Delta := \{\mathbf{B}\mathbf{u}'_1, \mathbf{B}\mathbf{u}'_2, \dots, \mathbf{B}\mathbf{u}'_r\}.$$

From here, it is simple to choose bases $\gamma \subseteq \Gamma$ and $\delta \subseteq \Delta$ for $V \cap W$.

(Naturally, it holds that $\mathbf{A}\mathbf{u}_i + \mathbf{B}\mathbf{u}'_i = \mathbf{0}$.)

2. An alternative method. By row reduction, we can calculate

$$\begin{aligned} r := \dim(V \cap W) &= \dim(U) + \dim(V) - \dim(U + V), \\ &= \text{rank}(\mathbf{A}) + \text{rank}(\mathbf{B}) - \text{rank}(\mathbf{A} \ \mathbf{B}), \\ &= \text{rank}(\mathbf{A}^\top) + \text{rank}(\mathbf{B}^\top) - \text{rank} \begin{pmatrix} \mathbf{A}^\top \\ \mathbf{B}^\top \end{pmatrix}. \end{aligned}$$

Then, a basis for $V \cap W$ can be formed by choosing r linearly independent columns of $(\mathbf{A} \ \mathbf{B})$, or rows of $\begin{pmatrix} \mathbf{A}^\top \\ \mathbf{B}^\top \end{pmatrix}$.

3. **Another alternative**, probably the best option! Skip the row reduction of \mathbf{A} and \mathbf{B} in the above method. We just reduce

$$(\mathbf{A} \ \mathbf{B}) \rightarrow (\mathbf{A}' \ \mathbf{B}').$$

Let \mathbf{c}_i and \mathbf{c}'_i be the i th column of $(\mathbf{A} \ \mathbf{B})$ and $(\mathbf{A}' \ \mathbf{B}')$, respectively. We compare the columns of \mathbf{A}' and \mathbf{B}' to find (with relative ease) a basis $\beta' := \{\mathbf{c}'_{i_1}, \mathbf{c}'_{i_2}, \dots, \mathbf{c}'_{i_r}\}$ for the intersection of the column spaces of \mathbf{A}' and \mathbf{B}' . Then, $\beta := \{\mathbf{c}_{i_1}, \mathbf{c}_{i_2}, \dots, \mathbf{c}_{i_r}\}$ is a basis for $V \cap W$ (the intersection of the column spaces of \mathbf{A} and \mathbf{B}).

4. **A fourth method** for when I learn about orthogonal complements.

Definition 1.14

Let $\mathbf{A} \in M_{n \times n}(\mathbb{F})$. If $n = 1$, so that $A = (a_1 1)$, we define $\det(\mathbf{A}) := a_1 1$. For $n \geq 2$, we define $\det(\mathbf{A})$ recursively as

$$\det(\mathbf{A}) := \sum_{j=1}^n (-1)^{1+j} \mathbf{A}_{1j} \cdot \det(\tilde{\mathbf{A}}_{1j}).$$

The scalar $\det(\mathbf{A})$ is called the *determinant* of \mathbf{A} and is also denoted by $|\mathbf{A}|$. The scalar

$$(-1)^{i+j} \det(\tilde{\mathbf{A}}_{ij})$$

is called the cofactor of the entry of \mathbf{A} in row i , column j .

- A matrix \mathbf{A} is invertible iff its determinant is nonzero.

Theorem 1.15

The determinant $\det: M_{n \times n}(\mathbb{F}) \rightarrow \mathbb{F}$ is an alternating n -linear function. The former (alternating) means that for $\mathbf{A} \in M_{n \times n}(\mathbb{F})$ and any \mathbf{B} obtained from \mathbf{A} by interchanging any two rows of \mathbf{A} ,

$$\det(\mathbf{B}) = -\det(\mathbf{A}).$$

The latter (n -linearity) means that, for any scalar $k \in \mathbb{F}$ and vectors $\mathbf{u}, \mathbf{v}, \mathbf{a}_i \in \mathbb{F}^n$,

$$\det \begin{pmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_{r-1}\mathbf{u} + k\mathbf{v} \\ \mathbf{a}_{r+1} \\ \vdots \\ \mathbf{a}_n \end{pmatrix} = \det \begin{pmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_{r-1}\mathbf{u} \\ \mathbf{a}_{r+1} \\ \vdots \\ \mathbf{a}_n \end{pmatrix} + k \det \begin{pmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_{r-1}\mathbf{v} \\ \mathbf{a}_{r+1} \\ \vdots \\ \mathbf{a}_n \end{pmatrix}.$$

(In fact, it can be shown that $\det: M_{n \times n}(\mathbb{F}) \rightarrow \mathbb{F}$ is the *unique* alternating n -linear function, such that $\det(\mathbf{I}) = 1$.)

Corollary 1.16

Let $\mathbf{A} \in M_{n \times n}(\mathbb{F})$. Then, for any matrix \mathbf{B} obtained by adding a scalar multiple of one row/column of \mathbf{A} to another, $\det(\mathbf{B}) = \det(\mathbf{A})$.

Theorem 1.17

The determinant of a square matrix can be evaluated by cofactor expansion along any row. That

is, if $\mathbf{A} \in M_{n \times n}(\mathbb{F})$, then for any integer $1 \leq i \leq n$,

$$\det(\mathbf{A}) = \sum_{j=1}^n (-1)^{i+j} \mathbf{A}_{ij} \cdot \det(\tilde{\mathbf{A}}_{ij}).$$

Here, $\tilde{\mathbf{A}}_{ij}$ is the $(n-1) \times (n-1)$ matrix obtained from \mathbf{A} by deleting its i th row and j th column.

Corollary 1.18

The determinant of any triangular matrix is the product of its diagonals.

Theorem 1.19

Let A be an $n \times n$ matrix. Then,

$$\det(\mathbf{A}) = \det(\mathbf{A}^\top).$$

So, the determinant of a square matrix can also be evaluated by cofactor expansion along any column.

Theorem 1.20

Let \mathbf{A} be an invertible $n \times n$ matrix. Then,

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

where $\text{adj}(\mathbf{A})$ is the adjugate/classical adjoint of \mathbf{A} . That is, the matrix whose (i, j) th entry is the (j, i) th cofactor $(-1)^{j+i} \det(\tilde{\mathbf{A}}_{ji})$.

Theorem 1.21

For any $\mathbf{A}, \mathbf{B} \in M_{n \times n}(\mathbb{F})$, we have $\det(\mathbf{AB}) = \det(\mathbf{A}) \cdot \det(\mathbf{B})$.

Definition 1.22

A linear operator T on a finite-dimensional vector space V is called *diagonalisable* iff there is an ordered basis β for V such that $[T]_\beta$ is a diagonal matrix. A square matrix \mathbf{A} is called diagonalisable iff $L_{\mathbf{A}}$ is diagonalisable.

Definition 1.23

Let T be a linear operator on a vector space V . A nonzero vector $\mathbf{v} \in V$ is called an *eigenvector* of T iff there exists a scalar λ such that $T(\mathbf{v}) = \lambda \mathbf{v}$. The scalar λ is called the *eigenvalue* corresponding to the eigenvector \mathbf{v} .

Let \mathbf{A} be in $M_{n \times n}(\mathbb{F})$. A nonzero vector $v \in \mathbb{F}^n$ is called an *eigenvector* of \mathbf{A} iff v is an eigenvector of $L_{\mathbf{A}}$; that is, iff $\mathbf{A}v = \lambda v$ for some scalar λ . The scalar λ is called the eigenvalue of \mathbf{A} corresponding to the eigenvector v .

Definition 1.24

Let $\mathbf{A} \in M_{n \times n}(\mathbb{F})$. The polynomial $f(t) = \det(\mathbf{A} - t\mathbf{I}_n)$ is called the *characteristic polynomial* of \mathbf{A} .

- A matrix $\mathbf{A} \in M_{n \times n}(\mathbb{F})$ is diagonalizable iff there exists an ordered basis $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ for \mathbb{F}^n consisting of eigenvectors of \mathbf{A} , i.e. a eigenbasis. Furthermore, if \mathbf{Q} is the $n \times n$ matrix whose j th column is \mathbf{v}_j , then $\mathbf{A} = \mathbf{Q}^{-1} \mathbf{D} \mathbf{Q}$ is a diagonal matrix such that d_{jj} is the eigenvalue of A corresponding to \mathbf{v}_j . The matrix \mathbf{Q} is said to *diagonalise* \mathbf{A} .
- Hence, we obtain the following procedure to diagonalise a 3×3 matrix \mathbf{A} with three distinct

eigenvalues.

1. Find the eigenvalues λ_1 , λ_2 , and λ_3 of \mathbf{A} . They are just the roots of the characteristic polynomial of \mathbf{A} . [This can be done using the GC.](#)
2. Find an eigenvector \mathbf{v}_j corresponding to each eigenvalue λ_j by finding the nullspace of $\mathbf{A} - \lambda_j \mathbf{I}$.
3. Let $\mathbf{Q} = (\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$. Then,

$$\mathbf{D} := \mathbf{Q}^{-1} \mathbf{A} \mathbf{Q}$$

is a diagonal matrix.

Note

Let \mathbf{A} be a 3×3 real matrix with the eigenvalue λ . Then, the cross product of two nonzero rows/columns of $\mathbf{A} - \lambda \mathbf{I}$ is an eigenvector of \mathbf{A} .

Theorem 1.25: The Cayley-Hamilton Theorem.

Let T be a linear operator on a finite dimensional vector space V , and let $f(t)$ be the characteristic polynomial of T . Then $f(T) = T_0$, the zero transformation. That is, T “satisfies” its characteristic equation.

Corollary 1.26: The Cayley-Hamilton Theorem for Matrices.

Let A be an $n \times n$ matrix, and let $f(t)$ be the characteristic polynomial of A . Then, $f(A) = O$, the $n \times n$ zero matrix.

G.C. Skills

Finding eigenvalues of a matrix \mathbf{A} using the GC.

1. `2nd \Rightarrow x^{-1} (matrix) \Rightarrow Key in the matrix $\mathbf{A} - t\mathbf{I}$, e.g. into $[\mathbf{A}]$.`
2. `Plot $Y_1 = \det([\mathbf{A}])$.`
3. `2nd \Rightarrow trace \Rightarrow 2:zero \Rightarrow Find the roots.`

Numerical Methods

General Information

- The parity of the degree of a real polynomial is the same as that of its number of real roots.
- Let the real polynomial p given by $p(x) = a_{2n}x^{2n} + a_{2n-1}x^{2n-1} + \dots + a_0$ have coefficients $a_n > 0$ and $a_0 < 0$. Then, it has at least one positive and one negative root.
- Suppose we have some function $f: \mathbb{R} \rightarrow \mathbb{R}$ with a root α , whose value we want to approximate. There are three ways to obtain this approximation.

1. Linear interpolation on an interval $[a, b]$ containing α . We let $x_0 := b$ and

$$x_{n+1} := \frac{a|f(x_n)| + x_n|f(a)|}{|f(x_n)| + |f(a)|}.$$

- The sequence $\{x_n\}$ of approximations *always* converges to α .
- The smaller $f''(x)$ is (i.e. the slower the gradient $f'(x)$ changes) near α , the faster the rate of convergence.
- Error:

Concave/Gradient	Positive	Negative
Upwards \cap	underestimation	overestimation
Downwards \cup	overestimation	underestimation

Table 2.1: Approximation errors when using linear interpolation.

- See Figure 2.1 for an illustration.

2. Fixed-point Iteration. First select a function $F: \mathbb{R} \rightarrow \mathbb{R}$, such that $F(\alpha) = \alpha$, and choose some initial approximation x_0 to α . Then, we recursively define $x_{n+1} := F(x_n)$. The desired convergence behavior is for x_n to approach α .

- Convergence behavior

Behavior of $ F'(x) $	Converges?	Rate of convergence
$ F'(x) < 1$ and is small near α	✓	fast
$ F'(x) < 1$ but is close to 1 near α	✓	slow
$ F'(x) \geq 1$ near α	✗	-

Table 2.2: Convergence behavior of fixed-point iteration.

- See Figure 2.2 for an illustration.

3. The Newton-Raphson Method. For a function $f: \mathbb{R} \rightarrow \mathbb{R}$ with a root α , the Newton-Raphson formula is

$$x_{n+1} := x_n - \frac{f(x_n)}{f'(x_n)}.$$

- The Newton-Raphson method fails in the following cases.
 - (a) The gradient at x_0 is too gentle.
 - (b) The gradient changes too rapidly.
 - (c) The initial approximation x_0 is too far from the root α .
 - (d) There is a turning point between the initial approximation x_0 and the root α .
 - (e) There is a point of inflection.

– Error:

Concave/Gradient	Positive	Negative
Upwards \cap	overestimation	underestimation
Downwards \cup	underestimation	overestimation

Table 2.3: Approximation errors when using the Newton-Raphson method.

– See Figure 2.3 for an illustration.

Screw trying to make nice diagonal cells. Pain. Suffering.

Note

At every iteration of linear interpolation, we must ensure that $\alpha \in [a, x_n]$. Otherwise x_n may not approximate α . If $\alpha \notin [a, x_n]$, simply consider $\alpha \in [x_n, b]$ (or any other suitable interval) instead.

G.C. Skills

Linear interpolation: finding an approximation to a root in $[a, b]$ up to n decimal places.

1. $Y_1 = f(x)$,
2. $a \rightarrow A$ and $b \rightarrow B$,
3. $\frac{B|Y_1(A)| + A|Y_1(B)|}{|Y_1(A)| + |Y_1(B)|}$,
4. Ans $\rightarrow A$ or B (choose the one that has the opposite sign to Ans),
5. Repeat steps 4 to 5,
6. Terminate this process when the approximations are consistent up to n decimal places.

You can freely enter any function and shift the initial values in the Desmos graphs below!

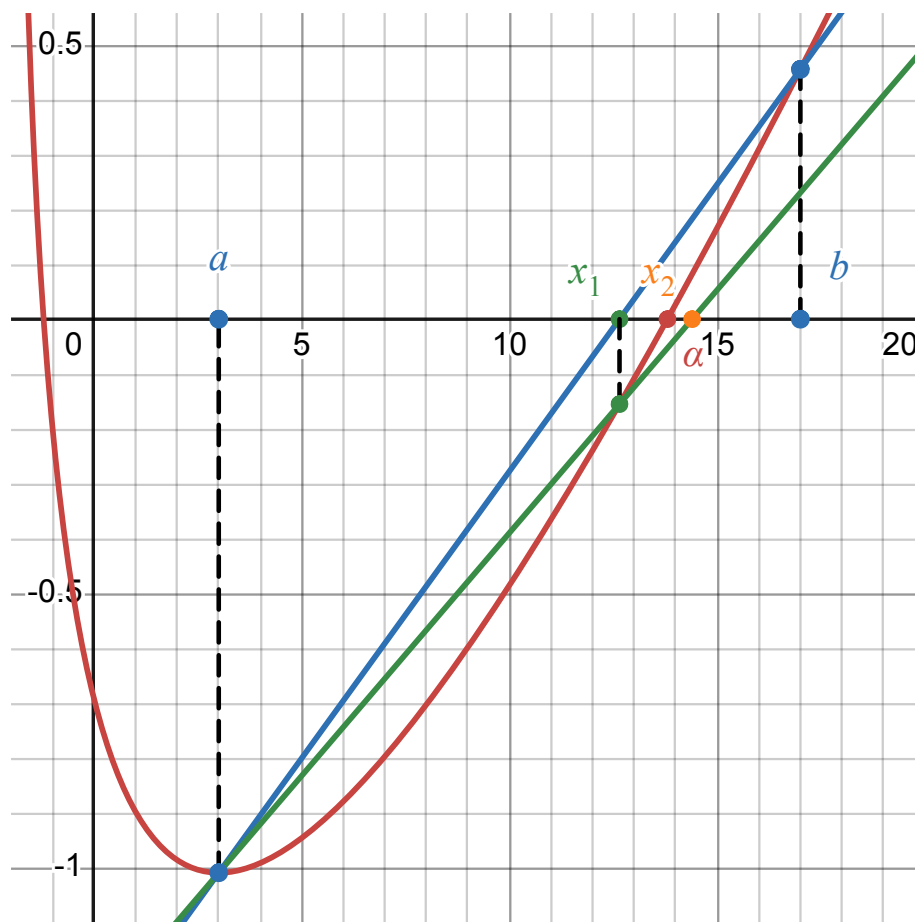


Figure 2.1: An illustration of linear interpolation ([Desmos](#)).

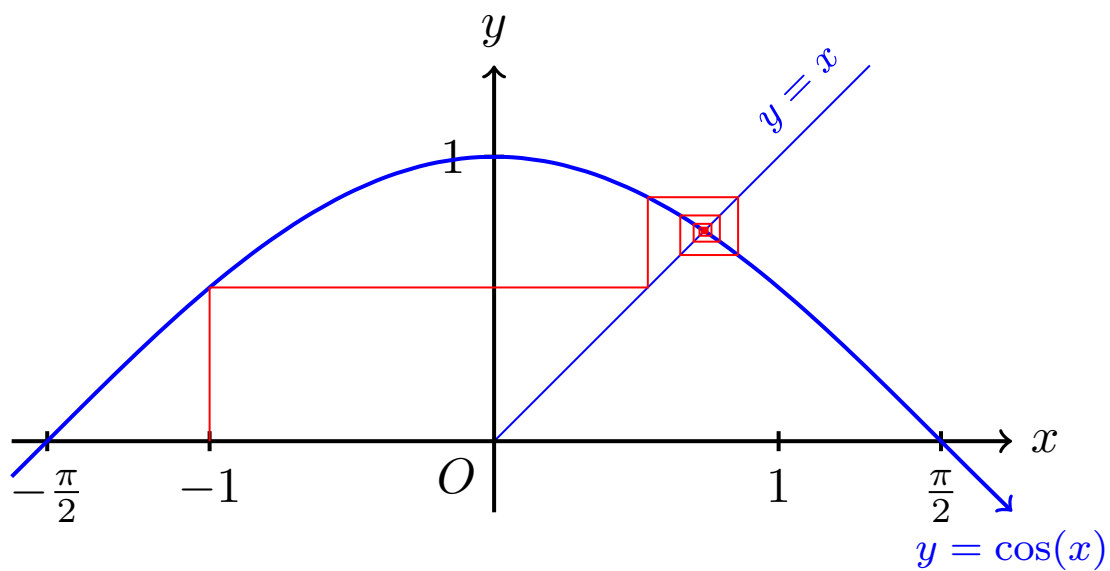


Figure 2.2: An illustration of fixed-point iteration.

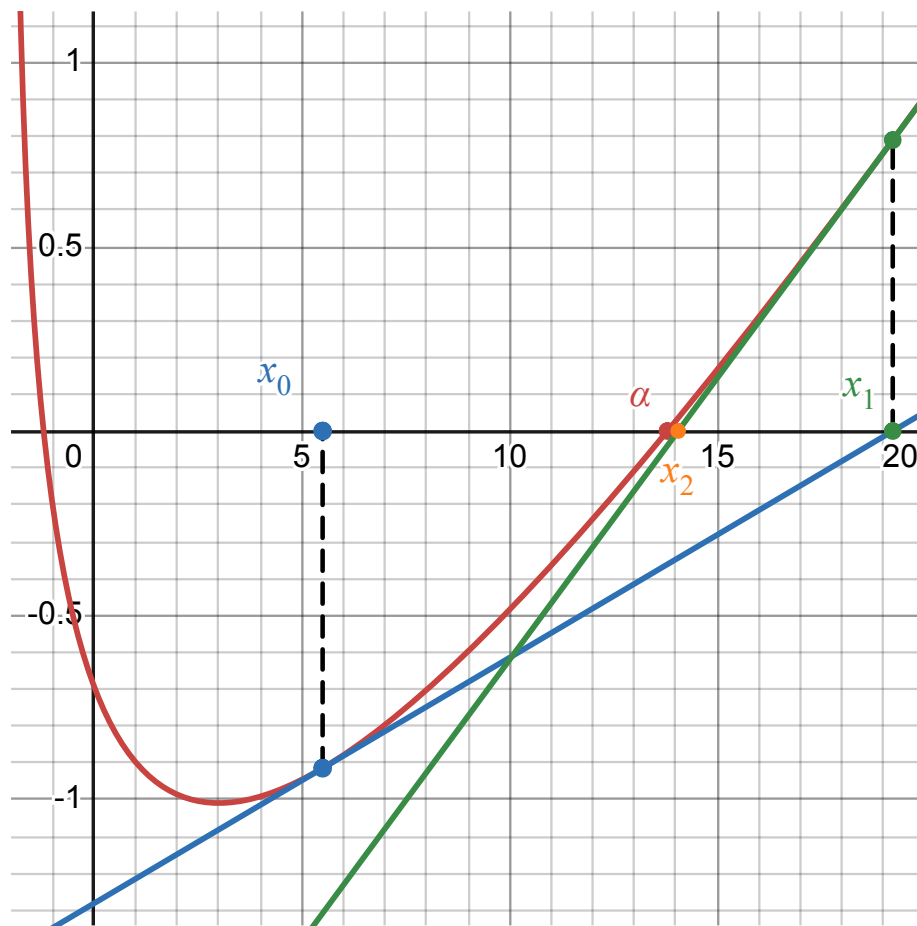


Figure 2.3: An illustration of Newton's Method ([Desmos](#)).