CHAPTER 2
BASIC MATH REVIEW FOR OPTIMIZATION

- We will review a few fundamental concepts on **linear algebra**
- We will review a few fundamental concepts on **convex analysis**

A primer on linear algebra

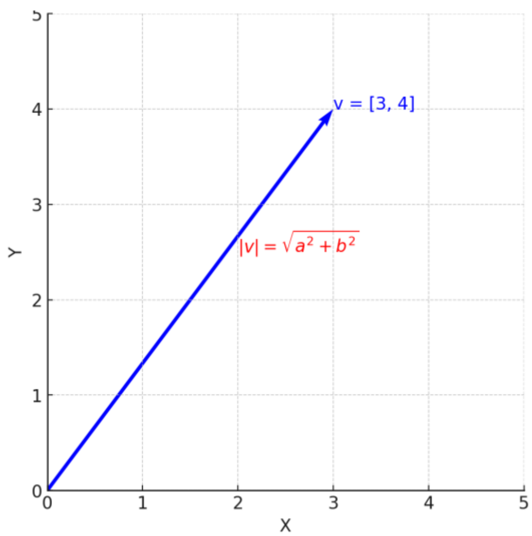
This brief review of linear algebra concepts is critical, as optimization problems and algorithms often make use of this language. First, we review the basic definitions and operations, afterwards we review the concept of eigenvalues and eigenvectors, notions that are critical in the study of convex optimization.

- *A vector* can be defined as an ordered list of numbers. In some instances, we can take advantage of the fact that a vector is a mathematical object capable of representing quantities that have magnitude and direction.

$$\boldsymbol{v} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

In general, the magnitude of a vector can be estimated using the Euclidean norm:

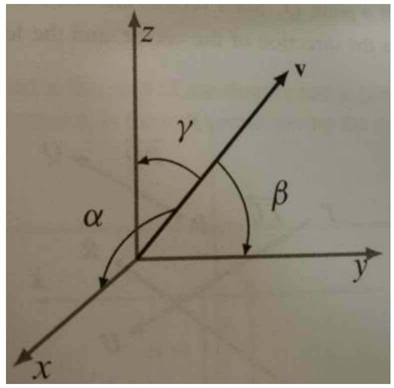
$$|\boldsymbol{v}| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$$



The direction of the vector in 2D can be estimated based on the tangent definition:

$$\theta = \tan^{-1}\left(\frac{y}{x}\right)$$

Note that for dimensions higher than 2D, there is not a single angle that we can define, but instead a collection of angles is needed. In 3D, there are 3 of these angles as shown in the figure.



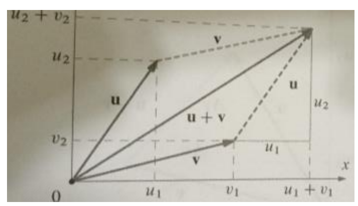
To find these angles we can use the cosine definition such that:

$$\cos(\alpha) = \frac{x}{\|\boldsymbol{v}\|}, \cos(\beta) = \frac{y}{\|\boldsymbol{v}\|}, \cos(\gamma) = \frac{z}{\|\boldsymbol{v}\|}$$

- Addition and subtraction are defined for vectors as follow:

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \pm \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 \pm y_1 \\ x_2 \pm y_2 \\ \vdots \\ x_n \pm y_n \end{bmatrix}$$

Geometrically, the addition of two vectors can be interpreted using the parallelogram law, stating that to sum two vectors we arrange them such that the head of one vector is located at the same point where the tail of the other vector is. The vector sum extends from the origin to the opposite side of the parallelogram formed.



- Scalar multiplication: Scalar multiplication is accomplished by multiplying each element of a vector by the scalar.

$$k\boldsymbol{v} = [kv_1, kv_2, kv_n]$$

- The dot product between vectors is defined as follows:

$$\boldsymbol{u}^T\boldsymbol{v} = \boldsymbol{v}^T\boldsymbol{u} = u_1v_1 + \cdots + u_nv_n$$

Also,

$$\boldsymbol{u}^T\boldsymbol{v} = [u_1 \quad u_2 \quad \dots \quad u_n] \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

Other notations for the dot product include:

$$\boldsymbol{u}^T\boldsymbol{v} = \boldsymbol{u} \cdot \boldsymbol{v} = \langle \boldsymbol{u}, \boldsymbol{v} \rangle = \langle \boldsymbol{u} | \boldsymbol{v} \rangle$$

The dot product definition it is useful too because it is related to other relevant definitions, in particular, we will say that two **vectors are orthogonal** if

$$\boldsymbol{u}^T\boldsymbol{v} = 0$$

If two vectors are orthogonal to each other, and both are of unit length, then such **vectors are called orthonormal.**

The dot product is also related to the angle between vectors, it comes from the law of cosines which is written as:

$$||\boldsymbol{u}-\boldsymbol{v}||^2 = ||\boldsymbol{u}||^2 + ||\boldsymbol{v}||^2 - 2||\boldsymbol{u}||\,||\boldsymbol{v}||cos\theta$$

If we note that $||\boldsymbol{u}-\boldsymbol{v}||^2 = (\boldsymbol{u}-\boldsymbol{v})^T(\boldsymbol{u}-\boldsymbol{v})$, we can easily prove that:

$$||\boldsymbol{u}||\,||\boldsymbol{v}||\cos(\theta) = \boldsymbol{u}^T\boldsymbol{v}$$

$$\cos\,\theta = \frac{\boldsymbol{u}^T\boldsymbol{v}}{||\boldsymbol{u}||\,||\boldsymbol{v}||}$$

- Two critical inequalities related with vectors are
  - Cauchy-Swartz inequality: We note that $\cos\theta \leq 1$, therefore:

$$\cos\theta = \frac{\boldsymbol{u}^T\boldsymbol{v}}{||\boldsymbol{u}||\,||\boldsymbol{v}||} \leq 1$$

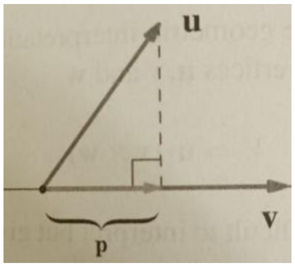
$$\frac{\boldsymbol{u}^T\boldsymbol{v}}{||\boldsymbol{u}||\,||\boldsymbol{v}||} \leq 1$$

$$\boldsymbol{u}^T\boldsymbol{v} \leq ||\boldsymbol{u}||\,||\boldsymbol{v}||$$

  - Triangle inequality:

$$||\boldsymbol{u}+\boldsymbol{v}|| \leq ||\boldsymbol{u}|| + ||\boldsymbol{v}||$$

- *Projection*: a projection of vector $\boldsymbol{u}$ into vector $\boldsymbol{v}$ is another vector $\boldsymbol{p}$**,** which can be understood as the shadow of one vector into the other one. There are two features of this projected vector $\boldsymbol{p}$ (1) vector p can be assumed to be a fraction of v, that is $\boldsymbol{p} = \alpha\boldsymbol{v}$, the perpendicular that drops from $\boldsymbol{v}$ is given by $\boldsymbol{d} = \boldsymbol{u} - \alpha\boldsymbol{v}$**,** note that by definition this vector is orthogonal to $\boldsymbol{v}$.



$$\boldsymbol{v}^T\boldsymbol{d} = 0$$

$$\boldsymbol{v}^T(\boldsymbol{u} - \alpha\boldsymbol{v}) = 0$$

$$\alpha = \frac{\boldsymbol{v}^T\boldsymbol{u}}{\boldsymbol{v}^T\boldsymbol{v}} = \frac{\boldsymbol{v}^T\boldsymbol{u}}{||\boldsymbol{v}||^2}$$

$$p = \alpha \boldsymbol{v} = \frac{\boldsymbol{v}^T \boldsymbol{u}}{\boldsymbol{v}^T \boldsymbol{v}} \boldsymbol{v}$$

- *A matrix* is a rectangular array of numbers indexed in such a way that $a_{i,j}$ refer to element in row *i* and column *j*.

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \cdots & \vdots \\ a_{mn} & \dots & a_{mn} \end{bmatrix}$$

  - *Matrix transposition:* A common operation in matrixes is transposition, the operation can be defined as follows.

$$A_{ji}^T = A_{ij}$$

  - *Addition and subtraction:*

$$\boldsymbol{A} + \boldsymbol{B} = \begin{bmatrix} a_{11} & \dots & a_{13} \\ \vdots & \vdots & \vdots \\ a_{m1} & \dots & a_{m3} \end{bmatrix} + \begin{bmatrix} b_{11} & \dots & b_{13} \\ \vdots & \vdots & \vdots \\ b_{m1} & \dots & b_{m3} \end{bmatrix} = \begin{bmatrix} a_{11} + b_{11} & \dots & a_{13} + b_{13} \\ \vdots & \vdots & \vdots \\ a_{31} + b_{m1} & \dots & a_{33} \end{bmatrix}$$

Which can be also defined using index notation as follows:

$$(\boldsymbol{A} + \boldsymbol{B})_{ij} = a_{ij} + b_{ij}$$

A relevant theorem related to the transposed sum of matrices is shown below:

$$(\boldsymbol{A} + \boldsymbol{B})^T = \boldsymbol{A}^T + \boldsymbol{B}^T$$

  - *Scalar multiplication:*

$$(\alpha A)_{i,j} = \alpha a_{i,j}$$

The transpose of a scalar product is calculated as follows:

$$(k\boldsymbol{A})^T = k\boldsymbol{A}^T$$

  - *Matrix multiplication:* the idea behind matrix multiplication comes from vector multiplication, note that in vectors we multiply a row vector times a column vector.

$$(AB)_{i,j} = x_{i1} y_{ij} + x_{i2} y_{2j} + \cdots + x_{in} y_{nj}$$

$$(m \times n)(n \times p) = (m \times p)$$

Remember that matrix multiplication is non commutative (i.e.,$AB \neq BA$).

The transpose of the product of two or more matrices is equal to the product of their transposes in reverse order.

$$(\boldsymbol{ABC})^T = \boldsymbol{C}^T \boldsymbol{B}^T \boldsymbol{A}^T$$

  - *Inverse of a matrix:* If $\boldsymbol{A}$ is and nxn matrix, and there exists a matrix $\boldsymbol{A}^{-1}$ such that then $\boldsymbol{A}^{-1}$ is called the inverse of $\boldsymbol{A}$.

$$\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}$$

The matrix can be defined only for square matrices. Even if a matrix is a square matrix, it may not have an inverse, in such case we say that the matrix is singular. When the inverse exists, we call it non-singular. Inverse properties include:

$$(\boldsymbol{AB})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$$

$$(\boldsymbol{B}^{T})^{-1} = (\boldsymbol{B}^{-1})^{T}$$

$$(\boldsymbol{A}^{-\mathbf{1}})^{-\mathbf{1}} = \boldsymbol{A}$$

- *Orthonormal matrices:* A matrix is said to be orthonormal if its columns or rows are orthogonal with each other and of unit length.

$$\boldsymbol{Q}^{T}\boldsymbol{Q} = \boldsymbol{I}$$

$$\boldsymbol{Q}^{T} = \boldsymbol{Q}^{-1}$$

- *Matrix partitioning:* A matrix can be often partitioned into submatrices, and many of the previously presented results hold.

$$\begin{bmatrix} \boldsymbol{A}_1 \\ \boldsymbol{B}_2 \end{bmatrix} + \begin{bmatrix} \boldsymbol{A}_2 \\ \boldsymbol{B}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}_1 + \boldsymbol{A}_2 \\ \boldsymbol{B}_1 + \boldsymbol{B}_2 \end{bmatrix}$$

$$\begin{bmatrix} \boldsymbol{A} \\ \boldsymbol{B} \end{bmatrix}^{T} = [\boldsymbol{A}^{\boldsymbol{T}} \quad \boldsymbol{B}^{\boldsymbol{T}}]$$

$$\boldsymbol{A}[\boldsymbol{B}_{\mathbf{1}} \quad \boldsymbol{B}_{\mathbf{2}}] = [\boldsymbol{AB}_{\mathbf{1}} \quad \boldsymbol{AB}_{\mathbf{2}}]$$

- *Determinants:* The determinant of a matrix is a scalar number that can be calculated from a square matrix. Its calculation is done recursively. For a 2x2 matrix

$$\det(\boldsymbol{A}) = \det\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = ad - bc$$

For an arbitrary nxn matrix the determinant can be calculated as follows:

$$\det(\boldsymbol{A}) = a_{11}C_{11} + a_{12}C_{12} + \cdots + a_{1n}C_{1n}$$

Where C_{i1} is a cofactor. To define the cofactor let's consider A to be a nxn matrix with entries i,j. We define a minor M_{rc} to be an (n-1)x(n-1) submatrix that is obtained by deleting the rth row and the cth column. The cofactor can be calculated using the determinant of a minor as follows:

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

The determinant of a matrix can be also estimated as the product of its eigenvalues.

Importantly, if the determinant of a matrix is zero then the matrix is singular. The following properties hold true:

$$\det(I) = I$$

$$\det(\boldsymbol{A}) = \det(\boldsymbol{A}^{T})$$

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

$$\det(\boldsymbol{AB}) = \det(\boldsymbol{A})\det(\boldsymbol{B})$$

- *Vector spaces:* A vector space is a non-empty set of objects called vectors on which there are two operations defined: addition and multiplication by scalars. These operations are subject to the following properties:

$$\boldsymbol{u} + \boldsymbol{v} \in \mathbf{V}$$

$$\boldsymbol{u} + (\boldsymbol{v} + \boldsymbol{w}) = (\boldsymbol{u} + \boldsymbol{v}) + \boldsymbol{w}$$
$$\boldsymbol{u} + \boldsymbol{v} = \boldsymbol{v} + \boldsymbol{u}$$
$$\exists \mathbf{0}: \boldsymbol{v} + 0 = 0 + \boldsymbol{v} = \boldsymbol{v} \ \forall \ \boldsymbol{v} \in \mathbf{V}$$
$$\exists -\boldsymbol{v}: \boldsymbol{v} + (-\boldsymbol{v}) = 0 \ \forall \ \boldsymbol{v} \in \mathbf{V}$$
$$\alpha \boldsymbol{u} \in \mathbf{V}$$
$$(\alpha + \beta)\boldsymbol{u} = \alpha \boldsymbol{u} + \beta \boldsymbol{u}$$
$$\beta(\boldsymbol{u} + \boldsymbol{v}) = \beta \boldsymbol{u} + \beta \boldsymbol{v}$$
$$(\alpha\beta)\boldsymbol{u} = \alpha(\beta)\boldsymbol{u}$$
$$1\boldsymbol{v} = \boldsymbol{v}$$

An example of a vector space is R^2, note that all ordered pairs (x, y) satisfy the properties that were previously outlined.

A non-example of a vector space is the set of positive real numbers, note that this is not a vector space because:

- It is not closed under scalar multiplication, that is $-1 \times v \notin R^+$
- There is no zero element (Additive identity), note that zero by definition is not part of R^+
- The additive inverse does not exist, that is, $-v$ is not part of R^+

- *Linear combinations and span:* given a set of vectors $\boldsymbol{u}_1$, $\boldsymbol{u}_2$,... $\boldsymbol{u}_n$, a linear combination of these vectors is given by:

$$k_1\boldsymbol{u}_1 + k_2\boldsymbol{u}_2 + \cdots + k_n\boldsymbol{u}_n$$

Let $\boldsymbol{u}_1$, $\boldsymbol{u}_2$,... $\boldsymbol{u}_n$ be any set of vectors in a vector space $\boldsymbol{V}$. The set of all combinations of these vectors is called the span and it is denoted by

$$span(\boldsymbol{u}_1, \boldsymbol{u}_2, \dots \boldsymbol{u}_n)$$

- *Linear independence:* given a collection of vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \dots \boldsymbol{v}_n$ if one or more of the vectors can be written as a combination of the others, then the collection is called linearly dependent. Otherwise, the collection is linearly independent.

  A more formal definition is the following, given a set of vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \dots \boldsymbol{v}_n$ in a subspace $\mathbf{V}$ and a set of scalars $c_1, c_2, \dots c_n$, then the set of vectors is linearly independent if the only solution to the following equation is $c_i = 0$. Otherwise, the set of vectors is linearly dependent.

$$c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_n\boldsymbol{v}_n$$

  When a set of vectors is arranged into a matrix as a series of columns, and the rank of the matrix is determined, if the rank is less than the number of vectors, we can say the vectors are linearly dependent.

- *Basis and dimension:* The basis of a vector space $\mathbf{V}$ is the smallest set of linearly independent vectors that spans $\mathbf{V}$.

- *Eigenvalues and eigenvectors:* for an $n \times n$ matrix $\boldsymbol{A}$, a scalar λ is called an eigenvalue of $\boldsymbol{A}$ if there is a non-zero vector $\boldsymbol{v}$, called an eigenvector, such that:

$$\boldsymbol{Av} = \lambda\boldsymbol{v}$$

To illustrate this concept, let's analyze the following example; think about the following matrix:

$$\begin{bmatrix} 4 & -2 \\ 1 & 1 \end{bmatrix}$$

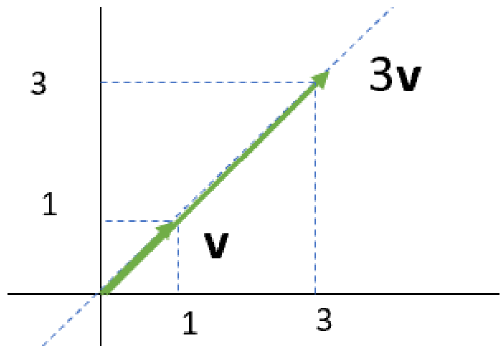
This matrix can be applied to transform any vector with adequate dimensions, that is we can multiply the matrix by the vector and obtain a new vector (this is a transformation):

$$\begin{bmatrix} 4 & -2 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 2 \\ 1 \end{bmatrix} = 3\begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

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

$$\begin{bmatrix} 4 & -2 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 4 \\ 1 \end{bmatrix} = \begin{bmatrix} 14 \\ 5 \end{bmatrix}$$

For the first two vectors, we have that the matrix has stretched the vectors by a factor of 2 and 3, respectively.



More generally, we see that in our case the matrix will stretch the following two vectors by factors of 3 and 2. Vectors that exhibit this behavior are called **eigenvectors**, and the **stretching factor is the eigenvalue**.

$$\begin{bmatrix} 2\alpha \\ \alpha \end{bmatrix} \text{ and } \begin{bmatrix} \alpha \\ \alpha \end{bmatrix}$$

For a matrix of dimension n, there will be at most n eigenvalues.

A method for computing eigenvalues and eigenvectors can be devised if we rearrange the eigenvalue definition as follows:

$$\boldsymbol{Av} = \lambda \boldsymbol{v}$$

$$(\boldsymbol{A} - \lambda \boldsymbol{I})\boldsymbol{v} = \boldsymbol{0}$$

For this equation to be true, λ must be selected such that the matrix $(\boldsymbol{A} - \lambda \boldsymbol{I})$ is singular. We also know that the determinant for a singular matrix must be zero. Therefore, we can use the following equation to find the eigenvalues.

$$\det(\boldsymbol{A} - \lambda \boldsymbol{I}) = \boldsymbol{0}$$

Example: Calculate the Eigenvalues and Eigenvectors of the following matrix.

$$A = \begin{bmatrix} 13 & -4 \\ -4 & 7 \end{bmatrix}$$

To calculate the eigenvalues we use the definition, as follows:

$$\det(A-\lambda I)=0=\begin{vmatrix}13-\lambda & -4\\ -4 & 7\end{vmatrix}=(13-\lambda)(7-\lambda)-16=\lambda^2-20\lambda+75=0$$

$$\lambda_1=15$$

$$\lambda_2=5$$

To calculate the eigenvectors then we use the definition as well

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

$$\left(\begin{bmatrix}13 & -4\\ -4 & 7\end{bmatrix}-15\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\right)\begin{pmatrix}v_1\\ v_2\end{pmatrix}=\begin{bmatrix}-2 & -4\\ -4 & -8\end{bmatrix}\begin{pmatrix}v_1\\ v_2\end{pmatrix}=\begin{pmatrix}0\\ 0\end{pmatrix}$$

If we diagonalize this matrix, we find that:

$$\begin{bmatrix}-2 & -4\\ 0 & 0\end{bmatrix}\begin{pmatrix}v_1\\ v_2\end{pmatrix}=\begin{pmatrix}0\\ 0\end{pmatrix}\rightarrow v_1=-2v_2$$

Then the first eigenvector is given by

$$v=\begin{pmatrix}-2a\\ a\end{pmatrix}$$

For the second eigenvectors we apply the same approach such that:

$$\begin{bmatrix}8 & -4\\ -4 & 2\end{bmatrix}\begin{pmatrix}w_1\\ w_2\end{pmatrix}=\begin{pmatrix}0\\ 0\end{pmatrix}$$

$$\begin{bmatrix}8 & -4\\ 0 & 0\end{bmatrix}\begin{pmatrix}w_1\\ w_2\end{pmatrix}=\begin{pmatrix}0\\ 0\end{pmatrix}\rightarrow w_1=-\frac{1}{2}w_2$$

$$v=\begin{pmatrix}-\frac{a}{2}\\ a\end{pmatrix}$$

Depending on the eigenvalues matrices can be classified as follows:

- If $\lambda_i>0$ for all *i* we say that the matrix is positive definite, meaning that $y^TAy>0$ for all $y\neq 0$
- If $\lambda_i<0$ for all *i* we say that the matrix is negative definite, meaning that $y^TAy<0$ for all $y\neq 0$
- If $\lambda_i\geq 0$ for all *i* we say that the matrix is positive semidefinite, meaning that $y^TAy\geq 0$ for all $y\neq 0$
- If $\lambda_i\leq 0$ for all *i* we say that the matrix is negative semidefinite, meaning that $y^TAy<0$ for all $y\neq 0$
- Matrices with both positive and negative eigenvalues are called indefinite

A primer on convex analysis

- Convex sets
  - Definition of a line: Let the vectors x_1 and $x_2\in R^n$. The line through x_1 and x_2 is defined as the set

    $$x|x=(1-\lambda)x_1+\lambda x_2,\lambda\in R$$

    We say that this segment is closed if:

    $$x|x=(1-\lambda)x_1+\lambda x_2,0\leq\lambda\leq 1$$

Note that this is an alternative way of writing a straight-line equation ($y = mx + b$ in 2D). Yet, both formulations are completely equivalent.

- Half-space: let the $\boldsymbol{c}$ be a vector in R^n and z a scalar value in R, the **open** half-space is defined as the following set:

$$\{x|\boldsymbol{c}^T\boldsymbol{x} < z, x \in R^n\}$$

The **closed** half space is similarly defined as follows

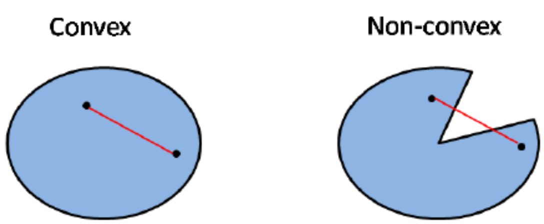
$$\{x|\boldsymbol{c}^T\boldsymbol{x} \leq z, x \in R^n\}$$

A hyperplane in R^n is defined as the set

$$\{x|\boldsymbol{c}^T\boldsymbol{x} = z, x \in R^n\}$$

The interception of a finite number of closed half-spaces in R^n is defined as a polytope. A bounded polytope is called a polyhedron.

- Convex-set: a set $S \in R^n$ is said to be convex if the closed line segment joining any two points x_1 and x_2 of the set S (that is, $(1-\lambda)x_1 + \lambda x_2$) belongs to the set S for each $0 \leq \lambda \leq 1$.



If S_1 and S_2 are convex sets, then:

  - The intersection $S_1 \cap S_2$ is convex
  - The Minkowski sum $S_1 + S_2$ is convex (the sum of sets is defined as $S_1 + S_2 = s_1 + s_2, s_1 \in S_1 \text{ and } s_2 \in S_2$).
  - The union of two convex sets is not necessarily convex.

- Extreme points: Let S be a convex set in R^n. The point $x \in S$ for which there exist no two distinct x_1 and $x_2 \in S$ different from x such that $x \in [x_1, x_2]$ is called a vertex or extreme point.
  - This is a point in S which does not lie in any open line segment joining two points.
  - Another way of saying this is, let $C \subseteq R^n$ be a convex set. Let $z \in R^n$. Then z is an extreme point of C if $z \in C$ and there are no x and $y \in C$ and $0 < \lambda < 1$ such that $z = (1-\lambda)x + \lambda y$

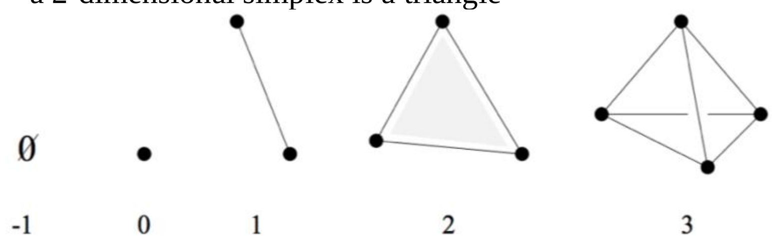
A convex set may have no vertices, a finite number of vertices, or an infinite number of vertices.

- Convex combination: let $\{x_1, \dots, x_r\}$ be any finite set of points in R^n. A convex combination of this set is a point of the form

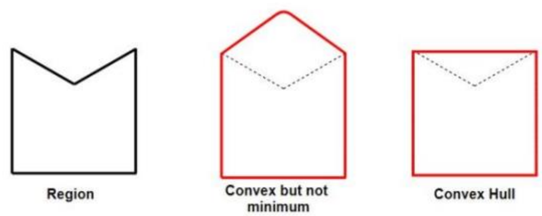
$$\lambda_1 x_1 + \cdots + \lambda_r x_r$$
$$\lambda_1 + \cdots + \lambda_r = 1$$
$$\lambda_1, \dots, \lambda_r \geq 0$$

Note that this is different from a linear combination because the values of $\lambda \geq 0$.

- Simplex: Let $\{x_o, x_1, \dots, x_r\}$ be r+1 distinct points in R^n $(r \leq n)$ and the vectors$\{x_1 - x_o, \dots, x_r - x_o\}$ be linearly independent. An r simplex in R^n is the set of all convex combinations of $\{x_o, x_1, \dots, x_r\}$.
    - The simplex is so named because it represents the simplest possible polytope in any given dimension (Wikipedia). For example,
        - a 0-dimensional simplex is a point
        - a 1-dimensional simplex is a line segment
        - a 2-dimensional simplex is a triangle



- Convex hull: Let S be a set (convex or nonconvex) in R^n. The convex hull, $H(S)$, is defined as the intersection of all convex sets in R^n which contain S as a subset.



    - *Caratheodory theorem*: let S be a set in R^n. If $x \in H(S)$, then it can be expressed as:

$$x = \sum_{i=1}^{n+1} \lambda_i x_i$$

$$\sum_{i=1}^{n+1} \lambda_i = 1$$

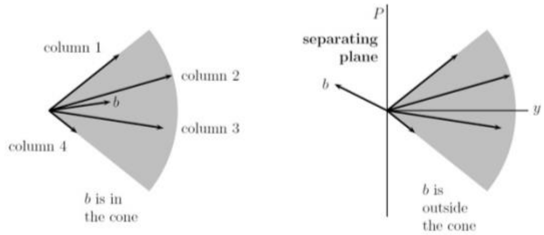
$$\lambda_i \geq 0, i = 1, \dots, n+1$$

$$x_i \in S, i = 1, \dots, n+1$$

    This theorem can be interpreted as follows, any point in the convex hull, can be expressed as a convex combination of at most $n + 1$ elements within the set.

- *Farkas theorem*: Let A be an $m \times n$ matrix and b be an m dimensional vector. Exactly one of the following two propositions is true
    - $Ax = b$ and $x \geq 0$ for some $x \in R^n$. This means that b is in the cone spanned by the columns of A
        - In a cone we have a linear combination where we insist that the values of the coefficients (x_i in the equation) are positive.

$$b = x_1 \boldsymbol{a}_1 + x_2 \boldsymbol{a}_2 + \cdots + x_n \boldsymbol{a}_n$$



- $A^T y \geq 0$ and $b^T y \leq 0$ for some $y \in R^m$. "The angle between y and b is greater than 90 degrees, so $y^T b \leq 0$. The angle between y and every column of A is less than 90 degrees, so $y^T A \geq 0$." Remember from the dot product definition that

$$\cos\ \theta = \frac{\boldsymbol{u}^T \boldsymbol{v}}{\|\boldsymbol{u}\|\|\boldsymbol{v}\|}$$

Farkas theorem can be understood as a certificate of feasibility, that is of the possibility that the system of equations is true.

- Convex functions:
  - *Basic notions*:
    - *Gradient of a function:* For a function with n variables the gradient can be expressed as follows.

$$\nabla f(x) = \left\{\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}\right\}$$

    - *Directional derivative:* the directional derivative can be defined as follows

$$\lim_{\alpha \to 0} \frac{f(x + \alpha d) - f(x)}{\alpha} = \nabla f(x)^T d$$

      - *Let $f: R^n \to R$ be a differentiable function. Consider x, d* $\in R^n$. The direction d is a descent direction in x if

$$d^T \nabla f(x) < 0$$

    - *Jacobian matrix:* consider $f: R^n \to R^m$. The function $J(x): R^n - R^{mxn}$ is alled a Jacobian matrix and it is defined as.

$$J(x) = \nabla f(x)^T = \begin{pmatrix} - & \nabla f_1(x)^T & - \\ \vdots & \vdots & \vdots \\ - & \nabla f_m(x)^T & - \end{pmatrix}$$

    - *Hessian of a function:* For a function with n the Hessian of a function is defined as follows.

$$H(f(x)) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

- *Definition:* Let S be a convex subset of R^n, and $f(x)$ be a real valued function defined on S. The function $f(x)$ is said to be convex if for any $x_1, x_2 \in S$, and $0 \leq \lambda \leq 1$, we have:

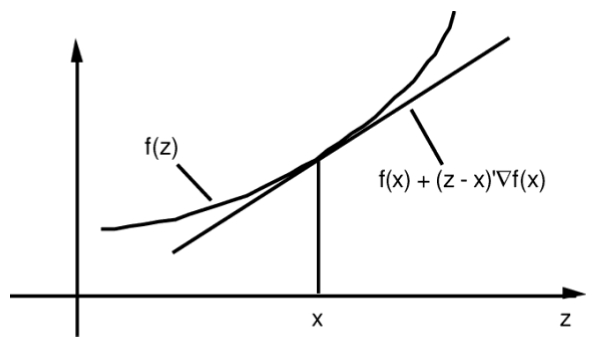
$$f\big((1-\lambda)x_1 + \lambda x_2\big) \leq (1-\lambda)f(x_1) + \lambda f(x_2)$$

This is called Jensen's inequality. If we have that the inequality is strictly satisfied (i.e., <) then we call the function strictly convex.

- Let S be a nonempty open set in R^n and $f(x)$ a differentiable function at $x_0 \in S$, if $f(x)$ is convex at x_o, then:

$$f(x) - f(x_0) \geq \nabla f(x_o)(x - x_o)$$

This means that the linear approximation of a convex function is always an under estimator.



- Let S be a nonempty open set in R^n, and $f(x)$ be a twice differentiable function in the domain S. If $f(x)$ is convex on S, then

$$\nabla^2 f(x) \text{ is positive semidefinite}$$

Remember that positive semidefinite means that all eigenvalues are greater or equal than zero, or that $z^T \nabla^2 f(x) z \geq 0$

- Let S be a nonempty open set in R^n, and $f(x)$ be a twice differentiable function at $x_0 \in S$. If $f(x)$ is convex at x_o, then

$$\nabla^2 f(x_o) \text{ is positive semidefinite}$$

Example 1: Consider the following function

$$f(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2 + 2x_1x_2$$

$$H(f(\boldsymbol{x})) = \begin{bmatrix} 2 & 2 & 0 \\ 2 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

$$|H - \lambda I| = \begin{vmatrix} 2-\lambda & 2 & 0 \\ 2 & 2-\lambda & 0 \\ 0 & 0 & 2-\lambda \end{vmatrix} = 0$$

Using a cofactor expansion

$$(2-\lambda)(-1)^2 \begin{vmatrix} 2-\lambda & 0 \\ 0 & 2-\lambda \end{vmatrix} + 2(-1)^3 \begin{vmatrix} 2 & 0 \\ 0 & 2-\lambda \end{vmatrix} + 0(-1)^4 \begin{vmatrix} 2 & 2-\lambda \\ 0 & 0 \end{vmatrix}$$

the eigenvalues are $\lambda = 2, \lambda = 4,\ \lambda = 0$, thus this equation is positive semidefinite. Meaning that the function is convex.

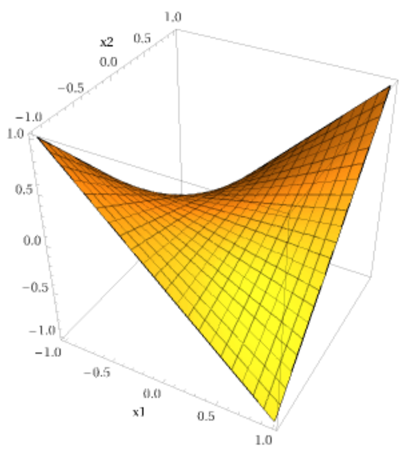
Example 2: consider the following function, is it convex?

$$f(x_1, x_2) = x_1 x_2$$

In this case the resulting Hessian is as follows:

$$H\big(f(x_1, x_2)\big) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

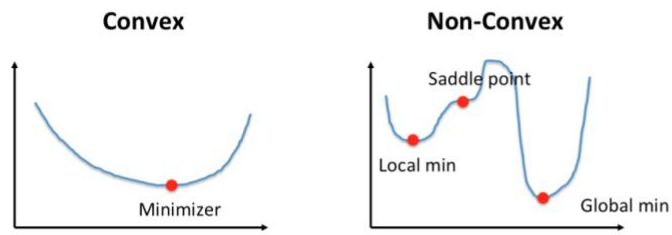
With Eigen values -1 and 1. Meaning that the equation is indefinite, if we check the plot, we can see that it looks like a saddle.



- *Properties of convex functions:* Convex functions can be combined in a number of ways to produce a new convex function
  - If $\{f_1(x), \dots, f_n(x)\}$ are convex functions on a convex subset $S \subseteq R^n$. Then their summation $f_1(x) + \cdots + f_n(x)$ is convex.
  - If $f_i(x)$ is convex/strictly convex on a convex subset $S \subseteq R^n$, and λ is a positive number, then $\lambda f_i(x)$ is convex/strictly convex.
  - Let $f_i(x)$ be convex/strictly convex on a convex subset $S \subseteq R^n$, and $g(y)$ be an increasing convex function defined on the range of $f(x)$ in R. Then the composite function $g[f(x)]$ is convex/strictly convex.

Consider for example the function $g(x) = e^{x^2}$ this function is convex because in this case $f(x) = x^2$ is strictly convex, and $g(x) = e^x$ is increasing and convex.

- *Relation between convexity and optimality:* one of the reasons that explains our interest in convexity is that it is related to the existence of a global optima vs a local optima. This relation is summarized in the following theorem:
  - Let S be a nonempty convex set in R^n and $x^* \in S$ be a local minimum:
    - If $f(x)$ is convex, then x^* is a global minimum
    - If $f(x)$ is strictly convex, then x^* is the unique global minimum



- Quadratic forms: Quadratic functions are critical in the development of non-linear programing algorithms, in general que can show that a quadratic function can be written in the following general way (a quadratic form)

$$f(x) = c + a^T x + \frac{1}{2} x^T B x$$

Where B is a simetric square matrix.

For example, if

$$B = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, a = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, c = 0$$

Then the function is given by:

$$f(x_1, x_2) = [1,1] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \frac{1}{2} [x_1, x_2] \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$f(x_1, x_2) = x_1 + x_2 + \frac{1}{2} [x_1, x_2] \begin{bmatrix} 2x_1 + x_2 \\ x_1 + 2x_2 \end{bmatrix} = x_1 + x_2 + \frac{1}{2} (2x_1^2 + 2x_1 x_2 + 2x_2^2)$$

- *Taylor series expansion:* when you have multiple variables the second order Taylor series expansion can be written as follows:

$$f(x) \approx f(a) + \nabla f(a)^T (x - a) + \frac{1}{2} (x - a)^T H(a)(x - a) + Error$$