# Numerical Optimization

Dhyan Laad

2024ADPS0875G

## Contents

# 1  Preliminaries

## 1.1  Inner Product Spaces

**Definition 1.1.** Let $V$ be a real vector space. A function $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ is called a *real inner product* if it satisfies the following properties for $x, y, z \in V$ and $c \in \mathbb{R}$:

(a) $\langle x, y \rangle = \langle y, x \rangle$,

(b) $\langle x + z, y \rangle = \langle x, y \rangle + \langle z, y \rangle$,

(c) $\langle cx, y \rangle = c \langle x, y \rangle$, and

(d) $\langle x, x \rangle \geq 0$ and $\langle x, x \rangle = 0 \Leftrightarrow x = 0$.

The inner product is the generalization of the dot product on Euclidean vector spaces: for $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ where $x_i, y_i \in \mathbb{R}$ for $i \in 1 : n$,

$$\mathbf{x} \cdot \mathbf{y} = \sum_{i=1}^{n} x_i y_i.$$

It may also be defined on $\mathbb{C}$-spaces, replacing (a) with conjugate symmetry: $\langle x, y \rangle = \overline{\langle y, x \rangle}$, and adding conjugate linearity in the second argument: $\langle x, cy \rangle = \bar{c} \langle x, y \rangle$.

**Definition 1.2.** Let $V$ be a real vector space and $\langle \cdot, \cdot \rangle$ an inner product. Then, $(V, \langle \cdot, \cdot \rangle)$ is called a *real inner product space.*

For brevity, we may simply state that $V$ is an inner product space, with the notation for the inner product being implicit, and all future inner product spaces map into $\mathbb{R}$ unless stated otherwise.

## 1.2  Normed Linear Spaces

**Definition 1.3.** Let $V$ be a real vector space. A function $\|\cdot\| : V \to \mathbb{R}$ is called a *norm* if it satisfies the following properties for $x, y \in V$ and $c \in \mathbb{R}$:

(a) $\|cx\| = |c| \|x\|$,

(b) $\|x\| \geq 0$ and $\|x\| = 0 \Leftrightarrow x = 0$, and

(c) $\|x + y\| \leq \|x\| + \|y\|$.

This last property is referred to as the *triangle inequality.* The norm assigns a notion of length to vectors, and generalizes the standard formula for the length of a vector in a Euclidean vector space: for $\mathbf{x} = (x_1, x_2, \ldots, x_n)^\top$ where $x_i \in \mathbb{R}$ for $i \in 1 : n$.

$$\|\mathbf{x}\| = \left( \sum_{i=1}^{n} x_i^2 \right)^{1/2}.$$

**Definition 1.4.** Let $V$ be a vector space and $\|\cdot\|$ a norm. Then $(V, \|\cdot\|)$ is called a *normed linear space.*

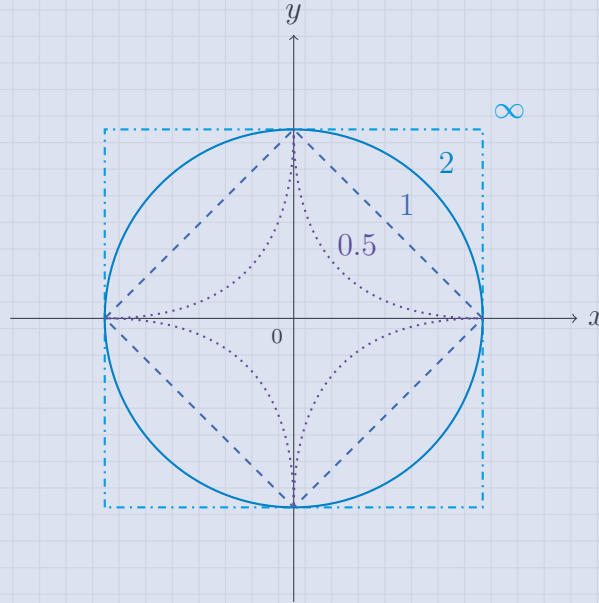Once again, we may conventionally omit the norm from notation when defining a new normed linear space.

**Definition 1.5.** Let $p \geq 1$. The *p-norm* (or $\ell^p$-*norm*) of a vector $\mathbf{x} = (x_1, x_2, \ldots, x_n)^\top$ where $x_i \in \mathbb{R}$ for $i \in 1 : n$ is

$$\|\mathbf{x}\|_p := \left( \sum_{i=1}^{n} |x_i|^p \right)^{1/p}.$$

For $p = 1$, we get the *taxicab* or *Manhattan* norm, for $p = 2$, we get the standard Euclidean norm, and for $p \to \infty$, the $p$-norm approaches the *infinity* or *maximum* norm:

$$\|\mathbf{x}\|_\infty := \max_i |x_i|.$$

For $p \in (0, 1)$, the triangle inequality does not hold, and the resulting functions are called *quasinorms*. Pictured below are unit circles in the 0.5-quasinorm, and $p$-norms for $p \in \{1, 2, \infty\}$.



**Lemma 1.6.** *Let $V$ be a normed linear space. Then,*

$$\big| \|x\| - \|y\| \big| \leq \|x - y\|$$

*for every $x, y \in V$.*

**Theorem 1.7.** *Let $(V, \|\cdot\|)$ be a normed linear space. Then $\|\cdot\| : V \to \mathbb{R}$ is uniformly continuous.*

*Proof.* Let $\varepsilon > 0$, and choose $\delta = \varepsilon$. If $\|x - y\| < \delta$, then by Lemma 1.6,

$$\big| \|x\| - \|y\| \big| \leq \|x - y\| < \delta = \varepsilon.$$

Therefore, $\big| \|x\| - \|y\| \big| < \varepsilon$. $\qquad \square$

**Theorem 1.8.** *Every inner product space $V$ is naturally a normed linear space. For $x \in V$ define*

$$\|x\| := \sqrt{\langle x, x \rangle}.$$

However, the converse is not always true.

**Theorem 1.9.** *A normed linear space $V$ is an inner product space iff the norm satisfies the parallelogram law:*

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

*The inner product is given by*

$$\langle x, y \rangle = \frac{1}{4} \left( \|x + y\|^2 - \|x - y\|^2 \right)$$

*for all $x, y \in V$.*

We now look at a fundamental theorem in linear algebra.

**Theorem 1.10 (Cauchy-Schwarz Inequality).** *Let $V$ be an inner product space with an induced norm. Then for all $x, y \in V$*

$$|\langle x, y \rangle| \leq \|x\| \|y\|.$$

*Proof.* We utilize the discriminant of a quadratic equation and the nonnegativity of the norm. Firstly, if $y = 0$, then the inequality holds trivially. For $y \neq 0$, consider the vector $x + ty$ for some $t \in \mathbb{R}$. Then

$$0 \leq \|x + ty\|^2 = \langle x + ty, x + ty \rangle = \|y\|^2 t^2 + 2\langle x, y \rangle t + \|x\|^2.$$

Since the quadratic in $t$ is nonnegative, its discriminant must be nonpositive, giving us

$$4\langle x, y \rangle^2 - 4\|x\|^2 \|y\|^2 \leq 0 \Rightarrow |\langle x, y \rangle| \leq \|x\| \|y\|.$$

$\square$

## 1.3   Eigenvalues and Eigenvectors

**Definition 1.11.** Let $A$ be a square matrix of order $n$. Then, $\lambda$ is said to be an *eigenvalue* of $A$ if there exists nonzero $\mathbf{x} \in \mathbb{R}^n$ such that

$$A\mathbf{x} = \lambda \mathbf{x}.$$

Here, $\mathbf{x}$ is called an *eigenvector* corresponding to $\lambda$.

To find the eigenvalues of an arbitrary matrix, consider the following:

$$A\mathbf{x} = \lambda \mathbf{x} \Rightarrow A\mathbf{x} - \lambda \mathbf{x} = \mathbf{0} \Rightarrow (A - \lambda I)\mathbf{x} = \mathbf{0}.$$

Since $\mathbf{x}$ is nonzero, we require the matrix $A - \lambda I$ to be singular. That is,

$$\det(A - \lambda I) = 0.$$

The left side of the equation is a polynomial in $\lambda$, and is called the *characteristic polynomial*. Therefore the eigenvalues of a matrix $A$ are the *roots* to its characteristic polynomial.

**Theorem 1.12.** *Eigenvectors corresponding to distinct eigenvalues are linearly independent.*

**Definition 1.13.** A square matrix $A$ is said to be diagonalizable if there exists an invertible matrix $P$ such that
$$P^{-1}AP = D$$
where $D$ is a diagonal matrix.

It turns out that in Definition 1.13, the entries of $D$ along the diagonal are the eigenvalues of $A$, and $P$ contains corresponding eigenvectors in the respective columns of the eigenvalues in $D$.

## 1.4   Quadratic Forms

Certain quadratic functions of several variables may be represented using matrices.

**Definition 1.14.** A *quadratic form* is a homogenous polynomial of degree 2 in multiple variables.

A quadratic form in $n$ variables may be written with a symmetric matrix $A$ of order $n$ and a vector $\mathbf{x} = (x_1, x_2, \ldots, x_n)^\top$ of unknowns:

$$Q(x) = \langle A\mathbf{x}, \mathbf{x} \rangle = \mathbf{x}^\top A\mathbf{x} = \sum_{i=1}^n a_{ii}x_i^2 + \sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^n a_{ij}x_i x_j$$

where $a_{ij}$ is the $(i,j)$th entry of $A$.

**Definition 1.15.** Let $A$ be a square matrix of order $n$, and let $p = \langle A\mathbf{x}, \mathbf{x} \rangle$. If for all nonzero $\mathbf{x} \in \mathbb{R}^n$

   (a) $p > 0$, then $A$ is said to be *positive definite*,

   (b) $p \geq 0$, *positive semidefinite*,

   (c) $p < 0$, *negative definite*, and

   (d) $p \leq 0$, *negative semidefinite*.

Otherwise, $A$ is said to be *indefinite*.

The definiteness of a matrix may also be realized through its principal minors.

**Definition 1.16.** Let $A$ be a square matrix of size $n$. A *principal minor of order $k$ of $A$* may be computed by the following.

   1. Choose any set of $k$ indices $\{i_1, i_2, \ldots, i_k\}$ from $\{1, 2, \ldots, n\}$.

   2. Keep the rows and columns corresponding to these indices, creating a $k \times k$ submatrix.

   3. Take the determinant of this resulting submatrix to yield a principal minor of order $k$.

If the first $k$ indices $\{1, 2, \ldots, k\}$ are chosen, then the resulting principal minors called *leading principal minors of order $k$*.

**Theorem 1.17.** *A square matrix is*

    (a)  *positive definite iff all leading principle minors are positive,*

    (b)  *positive semidefinite iff all principal minors are nonnegative,*

    (c)  *negative definite iff the kth leading principal minor has the sign of $(-1)^k$, and*

    (d)  *negative semidefinite iff all principal minors of order $k$ either have the sign of $(-1)^k$ or are 0.*

*A matrix that does not fall into the aforementioned categories is indefinite.*

We may also characterize definiteness using a matrix's eigenvalues.

**Theorem 1.18.** *A square matrix of order $n$ with eigenvalues $\lambda_i$ for $i \in 1 : n$ is*

    (a)  *positive definite iff $\lambda_i > 0$ for all $i$,*

    (b)  *positive semidefinite iff $\lambda_i \geq 0$ for all $i$,*

    (c)  *negative definite iff $\lambda_i < 0$ for all $i$, and*

    (d)  *negative semidefinite iff $\lambda_i \leq 0$ for all $i$.*

*A matrix that does not fall into the aforementioned categories is indefinite.*

## 1.5   Matrix Norms

**Definition 1.19.** Let $\mathbb{R}^{m \times n}$ denote the set of all matrices of size $m \times n$ with real entries. A function $\|\cdot\| : \mathbb{R}^{m \times n} \to \mathbb{R}$ is called a *real matrix norm* if it satisfies the following properties for $c \in \mathbb{R}$ and $A, B \in \mathbb{R}^{m \times n}$:

    (a)  $\|A\| \geq 0$ and $\|A\| = 0 \Leftrightarrow A = \mathbf{0}$,

    (b)  $\|cA\| = |c|\|A\|$,

    (c)  $\|A + B\| \leq \|A\| + \|B\|$, and

    (d)  $\|AB\| \leq \|A\|\|B\|$ for $m = n$.

Four of the most widely used ones are detailed below for a square matrix $A = [a_{ij}]_n$.

**1-norm.** The 1-norm is the *maximum absolute column sum*:

$$\|A\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^{n} |a_{ij}|.$$

**$\infty$-norm.** The $\infty$-norm is the *maximum absolute row sum*:

$$\|A\|_\infty = \max_{1 \leq i \leq n} \sum_{j=1}^{n} |a_{ij}|.$$

**2-norm.** The 2-norm is also called the *spectral norm* and is given by

$$\|A\|_2 = \sqrt{\lambda_{\max}(A^\top A)}$$

where $\lambda_{\max}(M)$ is the largest eigenvalue of $M$.

**Frobenius norm.** The Frobenius norm is the sum of the squared entries of the matrix:

$$\|A\|_F = \left( \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}^2 \right)^{1/2} = \sqrt{\operatorname{tr}(A^\top A)}$$

where $\operatorname{tr}(M)$ is the trace of $M$.

In general, a specific matrix norm can be induced from an existing vector norm by defining

$$\|A\| = \sup_{\mathbf{x} \neq 0} \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|}.$$

# 2   Calculus

## 2.1   Extreme Points

We start by reviewing and defining the concepts of a minimum and maximum from elementary calculus. Let $I = [a, b] \subset \mathbb{R}$ and $f : I \to \mathbb{R}$.

**Definition 2.1.** $x^* \in I$ is said to be a *local minimum* of $f$ if there exists $\delta > 0$ such that $f(x^*) \le f(x)$ for all $x \in B(x^*, \delta)$.

**Definition 2.2.** $x^* \in I$ is said to be a *local maximum* of $f$ if there exists $\delta > 0$ such that $f(x^*) \ge f(x)$ for all $x \in B(x^*, \delta)$.

**Definition 2.3.** $x^* \in I$ is said to be a *global minimum* of $f$ if $f(x^*) \le f(x)$ for all $x \in I$.

**Definition 2.4.** $x^* \in I$ is said to be a *global maximum* of $f$ if $f(x^*) \ge f(x)$ for all $x \in I$.

The definitions for the corresponding strict versions of the maxima and minima may be obtained by simply replacing the inequality with a strict one. The point $x^*$ is called a *minimizer* of *maximizer*.

**Theorem 2.5.** *Let $x^*$ be a local minimizer of $f$ and let $x^*$ be an interior point of $I = [a, b]$. If $f$ is differentiable at $x^*$, then $f'(x^*) = 0$.*

*Proof.* Since $x^*$ is a local minimizer and an interior point of $I$, there exists $\delta > 0$ such that $f(x^*) \le f(x)$ for all $x \in B(x^*, \delta)$. For $h > 0$ such that $h < \delta$,

$$f(x^*) \le f(x^* + h) \quad \text{and} \quad f(x^*) \le f(x^* + h).$$

We know that $f$ is differentiable at $x^*$. As such,

$$f'(x^*) = f'_+(x^*) = \lim_{h \to 0^+} \frac{f(x^* + h) - f(x^*)}{h},$$
$$f'(x^*) = f'_-(x^*) = \lim_{h \to 0^+} \frac{f(x^*) - f(x^* - h)}{h}.$$

As such, $f'(x^*) = f'_+(x^*) \geq 0$ and $f'(x^*) = f'_-(x^*) \leq 0$, and therefore $f'(x^*) = 0$. $\qquad \square$

An analogous result holds for a maximizer.

**Theorem 2.6.** *Let $x^*$ be a local maximizer of $f$ and let $x^*$ be an interior point of $I = [a, b]$. If $f$ is differentiable at $x^*$, then $f'(x^*) = 0$.*

**Definition 2.7.** Let $f : [a, b] \to \mathbb{R}$, and $x^* \in (a, b)$. If $f$ is differentiable at $x^*$ and $f'(x^*) = 0$, or if $f$ is not differentiable at $x^*$, it is said to be a *critical point of $f$*.

**Definition 2.8.** Let $f : [a, b] \to \mathbb{R}$, and $x^* \in (a, b)$. If $f$ is differentiable at $x^*$ and $f'(x^*) = 0$, then it is said to be a *stationary point of $f$*.