Non-Linear Optimization

Francisco Blanco-Silva

CHAPTER 1

Background

Our starting point is, for any positive integer $d \in \mathbb{N}$, the Cartesian products:

$$\mathbb{R}^d = \mathbb{R} \times \overset{(d)}{\cdots} \times \mathbb{R} = \{\boldsymbol{x} = (x_1, \dots, x_d) : x_k \in \mathbb{R} \text{ for } 1 \leq k \leq d\}.$$

We consider in this set the following two (closed) operations:

(a) Addition: For $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d, \boldsymbol{x} + \boldsymbol{y} = (x_1 + y_1, \dots, x_d + y_d) \in \mathbb{R}^d$.
(b) Scalar multiplication: For $\boldsymbol{x} \in \mathbb{R}^d$ and $\lambda \in \mathbb{R}$, $\lambda \cdot \boldsymbol{x} = \lambda \boldsymbol{x} = (\lambda x_1, \dots, \lambda x_d) \in \mathbb{R}^d$.

With these two operations, $(\mathbb{R}^d, +, \cdot)$ turns into a vector space: Given $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathbb{R}^d$, $\lambda, \mu \in \mathbb{R}$,

(a) The addition is commutative: $\boldsymbol{x} + \boldsymbol{y} = \boldsymbol{y} + \boldsymbol{x}$.
(b) Existence of identity elements for addition: Let $\mathbf{0} = (0, \dots, 0)$. $\boldsymbol{x} + \mathbf{0} = \boldsymbol{x}$.
(c) The addition is associative: $\boldsymbol{x} + (\boldsymbol{y} + \boldsymbol{z}) = (\boldsymbol{x} + \boldsymbol{y}) + \boldsymbol{z}$.
(d) Existence of inverse elements for addition: If $\boldsymbol{x} = (x_1, \dots, x_d)$, the element $-\boldsymbol{x} = (-x_1, \dots, -x_d)$ satisfies $\boldsymbol{x} + (-\boldsymbol{x}) = \mathbf{0}$. We write $\boldsymbol{x} - \boldsymbol{y}$ instead of $\boldsymbol{x} + (-\boldsymbol{y})$.
(e) Scalar multiplication is compatible with field multiplication: $\lambda(\mu \boldsymbol{x}) = (\lambda\mu)\boldsymbol{x}$.
(f) Existence of identity for scalar multiplication: $1 \cdot \boldsymbol{x} = \boldsymbol{x}$.
(g) Scalar multiplication is distributive with respect to addition: $\lambda(\boldsymbol{x} + \boldsymbol{y}) = \lambda\boldsymbol{x} + \lambda\boldsymbol{y}$.
(h) Scalar multiplication is distributive with respect to field addition: $(\lambda + \mu)\boldsymbol{x} = \lambda\boldsymbol{x} + \mu\boldsymbol{x}$.

A *basis* of $\mathbb{R}^d$ is a finite set $\mathcal{B} = \{\boldsymbol{b}_k : 1 \leq k \leq d\}$ satisfying two properties:

(a) Spanning property: For all $\boldsymbol{x} \in \mathbb{R}^d$ there exist d scalars $\{\lambda_1, \dots, \lambda_d\}$ so that $\boldsymbol{x} = \sum_{k=1}^d \lambda_k \boldsymbol{b}_k$.
(b) Linear independence: If $\{\lambda_1, \dots, \lambda_d\}$ satisfy $\sum_{k=1}^d \lambda_k \boldsymbol{b}_k = \mathbf{0}$, then it must be $\lambda_k = 0$ for all $1 \leq k \leq d$.

PROBLEM 1.1. Define in $\mathbb{R}^d$, for each $1 \leq k \leq d$ the element $\boldsymbol{e}_k$ to be the ordered d-tuple with k-th entry equal to one, and zeros on all other entries.

(a) Prove that $\{\boldsymbol{e}_k : 1 \leq k \leq d\}$ is a basis for $\mathbb{R}^d$.

(b) Set $\boldsymbol{b}_k = \boldsymbol{e}_k - \boldsymbol{e}_{k+1}$ for $1 \leq k < d$, $\boldsymbol{b}_d = \boldsymbol{e}_d$. Is $\{\boldsymbol{b}_k : 1 \leq k \leq d\}$ a basis for $\mathbb{R}^d$?

1. Functions

Given sets X, Y, we define a *function* $f\colon X \to Y$ to be a subset of $X \times Y$ subject to the following condition: for every $\boldsymbol{x} \in X$ there is exactly one element $\boldsymbol{y} \in Y$ such that the ordered pair $(\boldsymbol{x}, \boldsymbol{y})$ is contained in the subset defining f. The sets X and Y are called respectively the *domain* and *codomain* of f.

If A is any subset of the domain X, then $f(A)$ is the subset of the codomain Y consisting of all images of elements of A. We say that $f(A)$ is the *image* of A under f. The image of f is given by $f(X)$. The *inverse image* of a subset B of the codomain Y under a function f is the subset of the domain X defined by $f^{-1}(B) = \{\boldsymbol{x} \in X : f(x) \in B\}$.

For sets X, Y, Z, the *function composition* of $f\colon X \to Y$ with $g\colon Y \to Z$ is the function $g \circ f\colon X \to Z$ defined by $\big(g \circ f\big)(\boldsymbol{x}) = g\big(f(\boldsymbol{x})\big)$.

If $Y \subset \mathbb{R}$, we say that the function f is real-valued. For a real-valued function $f\colon \mathbb{R}^d \to \mathbb{R}$, we may regard the corresponding ordered pairs $(\boldsymbol{x}, y) \in \mathbb{R}^d \times \mathbb{R}$ as points in a $(d+1)$–dimensional space. We call this set the *graph* of f.

Unless specifically stated, all functions in these notes are real-valued functions $f\colon \mathbb{R}^d \to \mathbb{R}$.

EXAMPLE 1.1 (Linear Functions). We say that a real-valued function is *linear* if it preserves the operations in $\mathbb{R}^d$:

$$f(\boldsymbol{x} + \lambda \boldsymbol{y}) = f(\boldsymbol{x}) + \lambda f(\boldsymbol{y}) \text{ for } \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d, \lambda \in \mathbb{R}.$$

With this definition, the function $f(x) = 3x$ is indeed a linear function, but $g(x) = 3x + 5$ is not!

EXAMPLE 1.2 (Inner products). We say that a function $\langle \cdot, \cdot \rangle\colon \mathbb{R}^d \to \mathbb{R}$ is an *inner product* if it satisfies the following five properties for all $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathbb{R}^d$ and $\lambda \in \mathbb{R}$.

(a) $\langle \boldsymbol{x} + \boldsymbol{y}, \boldsymbol{z} \rangle = \langle \boldsymbol{x}, \boldsymbol{z} \rangle + \langle \boldsymbol{y}, \boldsymbol{z} \rangle$.
(b) $\langle \lambda \boldsymbol{x}, \boldsymbol{y} \rangle = \lambda \langle \boldsymbol{x}, \boldsymbol{y} \rangle$.
(c) $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \langle \boldsymbol{y}, \boldsymbol{x} \rangle$.
(d) $\langle \boldsymbol{x}, \boldsymbol{x} \rangle \geq 0$.
(e) $\langle \boldsymbol{x}, \boldsymbol{x} \rangle = 0$ if and only if $\boldsymbol{x} = 0$.

PROBLEM 1.2. Consider the real-valued function $\langle \cdot, \cdot \rangle\colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ as follows: Given $\boldsymbol{x} = (x_1, \dots, x_d), \boldsymbol{y} = (y_1, \dots, y_d) \in \mathbb{R}^d$,

$$\langle x, y \rangle = \sum_{k=1}^{d} x_k y_k.$$

Prove this is a well-defined function:

(a) The domain of $\langle \cdot, \cdot \rangle$ is $\mathbb{R}^d \times \mathbb{R}^d$.

(b) For all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$, $\langle \boldsymbol{x}, \boldsymbol{y} \rangle \in \mathbb{R}$.
(c) If $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \lambda_1$ and $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \lambda_2$, then $\lambda_1 = \lambda_2$.

Prove that this function is an inner product.

PROBLEM 1.3. Prove that, if f is a linear function in the sense of Example 1.1, then there exist a unique $\boldsymbol{a}_0 \in \mathbb{R}^d$ so that $f(\boldsymbol{x}) = \langle \boldsymbol{a}_0, \boldsymbol{x} \rangle$ for all $\boldsymbol{x} \in \mathbb{R}^d$.

PROBLEM 1.4. We say that $\tau \colon \mathbb{R}^d \to \mathbb{R}^d$ is a translation if there exist a fixed $\boldsymbol{x}_0 \in \mathbb{R}^d$ so that $\tau(\boldsymbol{x}) = \boldsymbol{x} + \boldsymbol{x}_0$ for all $\boldsymbol{x} \in \mathbb{R}^d$.

An *affine function* $h \colon \mathbb{R}^d \to \mathbb{R}$ is a composition of a linear function $f \colon \mathbb{R}^d \to \mathbb{R}$ with a translation $\tau \colon \mathbb{R} \to \mathbb{R}$.

Prove that for each affine function h there exist a unique $\boldsymbol{a}_0 \in \mathbb{R}^d$ and a unique $\lambda_0 \in \mathbb{R}$ so that $h(\boldsymbol{x}) = \lambda_0 + \langle \boldsymbol{a}_0, \boldsymbol{x} \rangle$ for all $\boldsymbol{x} \in \mathbb{R}^d$. Use this result to prove that the graph of an affine function is a hyperplane in $\mathbb{R}^{d+1}$.

EXAMPLE 1.3 (Norms). A *norm* in $\mathbb{R}^d$ is a function $\|\cdot\| \colon \mathbb{R}^d \to \mathbb{R}$ that satisfies the following properties: For all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$, and for all $\lambda \in \mathbb{R}$,

(a) $\|\boldsymbol{x}\| \geq 0$.
(b) $\|\boldsymbol{x}\| = 0$ if and only if $\boldsymbol{x} = \boldsymbol{0}$
(c) $\|\lambda \boldsymbol{x}\| = |\lambda| \|\boldsymbol{x}\|$.
(d) Triangle inequality: $\|\boldsymbol{x} + \boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$.

PROBLEM 1.5. Consider the function $\|\cdot\| \colon \mathbb{R}^d \to \mathbb{R}$ defined by

$$\|\boldsymbol{x}\| = \langle \boldsymbol{x}, \boldsymbol{x} \rangle^{1/2}.$$

(a) Prove that $\|\cdot\|$ is a norm
(b) Prove the *Cauchy-Schwartz inequality*: For all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$,

$$\big|\langle \boldsymbol{x}, \boldsymbol{y} \rangle\big| \leq \|\boldsymbol{x}\| \|\boldsymbol{y}\|.$$

2. Topology

The norm introduced in Example 1.3 induces a *metric* $\mathrm{d} \colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ on the space $\mathbb{R}^d$:

$$\mathrm{d}(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x} - \boldsymbol{y}\| \text{ for any } \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d.$$

Given $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathbb{R}^d$,

(a) Separation property: $\mathrm{d}(\boldsymbol{x}, \boldsymbol{y}) \geq 0$.
(b) Identity of indiscernibles: $\mathrm{d}(\boldsymbol{x}, \boldsymbol{y}) = 0$ if and only if $\boldsymbol{x} = \boldsymbol{y}$.
(c) Symmetry: $\mathrm{d}(\boldsymbol{x}, \boldsymbol{y}) = \mathrm{d}(\boldsymbol{y}, \boldsymbol{x})$.
(d) Triangle inequality: $\mathrm{d}(\boldsymbol{x}, \boldsymbol{z}) \leq \mathrm{d}(\boldsymbol{x}, \boldsymbol{y}) + \mathrm{d}(\boldsymbol{y}, \boldsymbol{z})$.

We say then that $\big(\mathbb{R}^d, \mathrm{d}(\cdot, \cdot)\big)$ is a *metric space*. Metric spaces inherit a *topology* in a natural manner, as explained below.

We define the *open ball* of radius $r > 0$ about $\boldsymbol{x}$ as the set $B_d(\boldsymbol{x}, r) = \big\{\boldsymbol{y} \in \mathbb{R}^d : \|\boldsymbol{x} - \boldsymbol{y}\| < r\big\}$. We say $\boldsymbol{x}$ is an interior point of $D \subset \mathbb{R}^d$ if $\boldsymbol{x} \in D$ and there exists $r > 0$ so that $B_d(\boldsymbol{x}, r) \subset D$. A subset $G \subset \mathbb{R}^d$ is said to be open if all its points are interior.

A *neighborhood* of the point $\boldsymbol{x}$ is any subset of $\mathbb{R}^d$ that contains an open ball about $\boldsymbol{x}$ as subset.

A *sequence* $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ in $\mathbb{R}^d$ is an enumerated collection of elements of $\mathbb{R}^d$ in which repetitions are allowed. A sequence is said to *converge* to the limit $\boldsymbol{x} \in \mathbb{R}^d$ if and only if for every $\varepsilon > 0$ there exists $N = N(\varepsilon) \in \mathbb{N}$ so that $\|\boldsymbol{x}_n - \boldsymbol{x}\| < \varepsilon$ for all $n \geq N$. We write then

$$\boldsymbol{x} = \lim_n \boldsymbol{x}_n, \text{ or } \lim_n \|\boldsymbol{x}_n - \boldsymbol{x}\| = 0.$$

We say that a sequence $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ is a *Cauchy sequence* if for every $\varepsilon > 0$ there exists $N = N(\varepsilon) \in \mathbb{N}$ so that for any $m, n \geq N$, $\|\boldsymbol{x}_n - \boldsymbol{x}_m\| < \varepsilon$. In $\mathbb{R}^d$, all Cauchy sequences converge (this is direct consequence of the completeness of $\mathbb{R}$).

The complement of an open set is called *closed*. In $\mathbb{R}^d$, all subsets F are closed if and only if they are *sequentially closed*: If $\boldsymbol{x}_n \in F$ for all $n \in \mathbb{N}$ and $\lim_n \|\boldsymbol{x}_n - \boldsymbol{x}\| = 0$, then $\boldsymbol{x} \in F$.

We say D is *bounded* if there exists $M > 0$ so that $D \subset B_d(\mathbf{0}, M)$. A bounded and closed subset of $\mathbb{R}^d$ is called *compact*.

Theorem 2.1 (Bolzano-Weierstrass). *Every sequence in a compact subset $K \subset \mathbb{R}^d$ contains a convergent subsequence.*

3. Analysis

A real-valued function $f\colon \mathbb{R}^d \to \mathbb{R}$ is continuous at $\boldsymbol{x}_0$ if for any $\varepsilon > 0$ there exists $\delta = \delta(\varepsilon) > 0$ so that $|f(\boldsymbol{x}) - f(\boldsymbol{x}_0)| < \varepsilon$ for all $x \in B_d(\boldsymbol{x}_0, \delta)$.

Equivalently, $f\colon \mathbb{R}^d \to \mathbb{R}$ is continuous at $\boldsymbol{x}_0$ if $\lim_n f(\boldsymbol{x}_n) = f(\boldsymbol{x}_0)$ for any sequence $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ satisfying $\lim_n \boldsymbol{x}_n = \boldsymbol{x}_0$.

We say that f is continuous in $D \subset \mathbb{R}^d$ if f is continuous at all points $\boldsymbol{x} \in D$.

A real-valued function $f\colon \mathbb{R}^d \to \mathbb{R}$

Given a set $D \subset \mathbb{R}^d$, and a real-valued function $f\colon D \to \mathbb{R}$, we say that a point $\boldsymbol{x}^\star \in D$ is:

(a) A *global minimum* for f on D if $f(\boldsymbol{x}^\star) \leq f(\boldsymbol{x})$ for all $\boldsymbol{x} \in D$.
(b) A *strict global minimum* for f on D if $f(\boldsymbol{x}^\star) < f(\boldsymbol{x})$ for all $\boldsymbol{x} \in D \setminus \{\boldsymbol{x}^\star\}$.
(c) A *local minimum* for f on D if there exists $\delta > 0$ so that $f(\boldsymbol{x}^\star) \leq f(\boldsymbol{x})$ for all $\boldsymbol{x} \in B_\delta(\boldsymbol{x}^\star) \cap D$.
(d) A *local minimum* for f on D if there exists $\delta > 0$ so that $f(\boldsymbol{x}^\star) < f(\boldsymbol{x})$ for all $\boldsymbol{x} \in B_\delta(\boldsymbol{x}^\star) \cap D$, $\boldsymbol{x} \neq \boldsymbol{x}^\star$.

A real-valued function $f\colon \mathbb{R}^d \to \mathbb{R}$ is *continuous* at a point $\boldsymbol{x}$

CHAPTER 2

Unconstrained Optimization via Calculus