Optimization for machine learning
- Constrained Optimization -
IASD Lecture notes

Vincent Duval

November 23, 2023

Contents

Foreword

These lecture notes are intended as a course material for Constrained Optimization in the “Optimization for machine learning” class. Please note that they are at an early stage, and as a consequence, they are probably riddled with typos.

They are aimed at an active reader who is equipped with some paper and a pencil: not all the details of the proofs are given, but I have tried to provide the main clues which help completing the proofs.

On the other hand, those notes are intended to serve as a reference, and it is not necessary to learn everything that is written to complete the curriculum. The most important results and definitions, which should be known “by heart” are indicated by the symbol ♡. Conversely some parts, indicated by (♠), contain discussions or results that can be omitted in a first reading. They are not part of the curriculum, they usually give some pointers or remarks to go beyond the framework of the course.

Notation

The typical notation throughout the notes is the following. We want to optimize a function f defined on some vector space X. We shall always assume that $X = \mathbb{R}^p$ for some $p \in \mathbb{N}^*$.

Vocabulary. We say that a function $f \colon \mathbb{R} \to \mathbb{R}$ is increasing (resp. strictly increasing) if for all $x, y \in \mathbb{R}$, $x < y$ implies $f(x) \leq f(y)$ (resp. $f(x) < f(y)$). Similarly, we say that f is decreasing (resp. strictly decreasing) if for all $x, y \in \mathbb{R}$, $x < y$ implies $f(x) \geq f(y)$ (resp. $f(x) > f(y)$).

You may find other documents in the English-speaking literature in which "increasing" mean our "strictly increasing", and "non-decreasing" mean our "increasing". We adopt here the convention of [AB06], which is close to the French terminology.

Indeterminate forms. We shall avoid as much as possible indeterminate forms, but in case they happen we adopt the convention $0 \times \infty = 0 \times (-\infty) = 0$.

Moreover we define by convention

$$\inf \emptyset = +\infty, \quad \sup \emptyset = -\infty. \tag{1}$$

Line segments. Given two points $x, y \in X$, we define the *closed line segment* joining x and y as $[x, y] = \{\, \theta x + (1 - \theta) y \mid 0 \leq \theta \leq 1 \,\}$. The *open line segment* joining x and y is $]x, y[\, = [x, y] \setminus \{x, y\}$. Note that $]x, x[\, = \emptyset$ and that if $x \neq y$, $]x, y[\, = \{\, \theta x + (1 - \theta) y \mid 0 \leq \theta \leq 1 \,\}$. The intervals $]x, y]$ and $[x, y[$ are defined similarly and are left to the reader.

Interior, closure. If $A \subseteq X$, we denote by $\mathrm{int}(A)$ (resp. $\overline{A}$) the interior (resp. closure) of A. The open ball centered at x with radius $r > 0$ is denoted by

$$B(x, r) \overset{\text{def.}}{=} \{\, y \in X \mid \|x - y\| < r \,\}.$$

Chapter 1

Optimality conditions

Contents

In this chapter, we consider a general problem of the form

$$\min f(x) \quad \text{s.t.} \quad x \in A \tag{1.1}$$

where

- $f\colon X \to \mathbb{R}$ is the *objective function* (not necessarily convex in this chapter),
- $A \subseteq X$ is the set of *feasible points* (also known as admissible points). That set results from a certain number of constraints that our solution should satisfy. We may as well call it the constraint set.

For now, unless explicitely stated, we do not make any additional assumption on A. Of course, in order to ensure the existence of minimizers, it is natural to assume that A is a nonempty compact set and that f is lower semi-continuous on A, or more generally that the function $f + \chi_A$ is proper, coercive and lower semi-continuous. But we regard the existence as an issue that has been dealt with separately. We focus here on characterizing the minimizers: if they exist, *what conditions should they satisfy*?

For the sake of simplicity, we assume that f is defined on X, so that there is no ambiguity when we require that f is differentiable at $x \in A$, but everything holds if f is defined on a neighborhood of A.

1.1 Motivation: the one-dimensional case

The reader has probably known since high school how to search the minimum of functions. In dimension 1, when $A = [a, b]$ and f is differentiable, things are rather straightforward (see Figure 1.1).

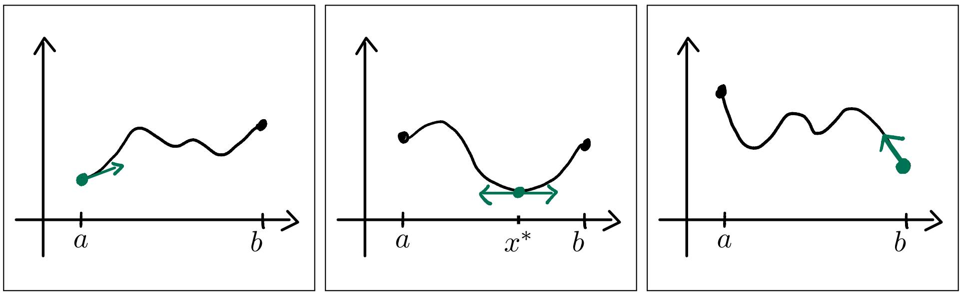

Figure 1.1: Optimization on a line segment $A = [a, b]$. *Left*: If a is a minimizer, then $f'(a) \geq 0$. *Middle*: If $x^* \in]a, b[$ is a minimizer, then $f'(x^*) = 0$. *Right*: if b is a minimizer, then $f'(b) \leq 0$.

- If $x^* \in]a, b[$, we may compare x^* with values "on the left": for all $h > 0$ small enough, $f(x^* - h) \geq f(x^*)$. Hence,

  $$\frac{f(x^* - h) - f(x^*)}{h} \geq 0,$$

  and letting $h \to 0^+$, we obtain $f'(x^*) \leq 0$.

  We may as well compare the value at x^* with values "on the right": for all $h > 0$ small enough, $f(x^* + h) \geq f(x^*)$, so that

  $$\frac{f(x^* + h) - f(x^*)}{h} \geq 0,$$

  and letting $h \to 0^+$, we obtain $f'(x^*) \geq 0$.

- For $x^* = a$, it is only possible to compare its value with the points on the right: arguing as above we deduce that $f'(x^*) \geq 0$.

- For $x^* = b$, it is only possible to compare its value with the points on the left: we deduce that $f'(x^*) \leq 0$.

When studying functions of several variables, we would like to use the same principle. We now assume that $X = \mathbb{R}^p$ and, as before, that $x^* \in \operatorname{argmin}_A f$.

Consider a direction $d \in \mathbb{R}^p \setminus \{0\}$. If for all $t > 0$ small enough, the vector $x^* + td$ is in A, we note that

$$\frac{f(x^* + td) - f(x^*)}{t} \geq 0.$$

Passing to the limit $t \to 0^+$, we obtain that $\langle \nabla f(x^*), d \rangle \geq 0$, which is equivalent to

$$\langle -\nabla f(x^*), d \rangle \leq 0.$$

In other words, we deduce that *the vector* $-\nabla f(x^*)$ *is in the half-space* $\{ s \in \mathbb{R}^p \mid \langle s, d \rangle \leq 0 \}$. If we can repeat this argument for many different vectors $d \in \mathbb{R}^p$, we deduce further that $-\nabla f(x^*)$ is in the intersection of all the corresponding half-spaces.

The difficulty when applying this argument in dimension $p \geq 1$ is to understand what comparisons are possible given the geometry of the domain A. For instance, a direction d such that $x^* + td \in A$ for all $t > 0$ small enough might not even exist (see)!

Therefore, the chapter describes how the above argument may be adapted by approaching d asymptotically.

1.2 The tangent cone and the normal cone

1.2.1 Definition and optimality condition

Our first step is to understand which directions d we can approximate when comparing some reference point $x \in A$ and some points $x_k \in A$ that converge to d, and to define the corresponding intersection of half-spaces.

Definition 1.1. *Let $A \subseteq X$ and $x \in A$. The* tangent cone *to A at x is the set*

$$\mathcal{T}_A(x) \overset{\text{def.}}{=} \Big\{ d \in \mathbb{R}^p \mid \exists \{t_k\}_{k\in\mathbb{N}} \subseteq \mathbb{R}_+^*, \exists \{x_k\}_{k\in\mathbb{N}} \subseteq A, \lim_{k\to\infty} t_k = 0, \\ \text{and } \lim_{k\to\infty} \left(\frac{x_k - x}{t_k} \right) = d \Big\}. \tag{1.2}$$

The normal cone *to A at x is the set*

$$\mathcal{N}_A(x) \overset{\text{def.}}{=} \{ s \in \mathbb{R}^p \mid \forall d \in \mathcal{T}_A(x), \langle s, d \rangle \leq 0 \}. \tag{1.3}$$

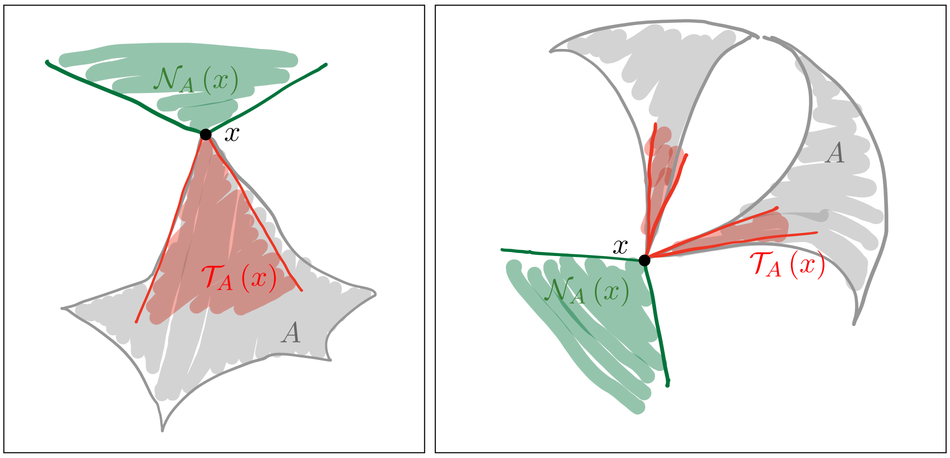

Figure 1.2: *(Left)* Some set A and its tangent (in red) and normal (in green) cones at x. Both cones have been shifted to x. *(Right)* Same representation for a different set A. The tangent cone is not necessarily convex, contrary to the normal cone.

In other words, we may test the values of f in a neighborhood of x with

$$x_k = x + t_k d + o(t_k) \in A,$$

and the tangent cone $\mathcal{T}_A(x)$ is the collection of all the corresponding asymptotic directions d. The normal cone $\mathcal{N}_A(x)$ is the collection of directions which make a right or obtuse angle with all the directions in the tangent cone. In convex analysis, one would call it the *polar cone* of $\mathcal{T}_A(x)$. The properties of polar cones are well understood, we recall some of them in Appendix A. In particular, taking the polar is a decreasing operation for set inclusion, and it is possible to describe the polar of cones which are generated by a finite number of linear inequalities using Farkas' Lemma (see Proposition A.2).

Exercise 1.2 (Affine constraint)**.** *Let $A \subseteq X$ be an affine space (that is, a set of the form $A = x_0 + V$, where $V \subseteq \mathbb{R}^p$ is a vector space). Prove that for every $x \in A$, $\mathcal{T}_A(x) = V$ and $\mathcal{N}_A(x) = V^\perp \overset{\text{def.}}{=} \{\, s \in \mathbb{R}^p \mid \forall v \in V,\ \langle s,\, v\rangle = 0 \,\}$.*

Remark 1.3. *The tangent cone always contains* 0*. If $x \in \mathrm{int}(A)$, that is, x is in the interior of A, then $\mathcal{T}_A(x) = \mathbb{R}^p$ and $\mathcal{N}_A(x) = \{0\}$ (check this as an exercise).*

Remark 1.4. *The tangent cone $\mathcal{T}_A(x)$ is a closed cone (another exercise!), but it is not necessarily convex. On the other hand, the normal cone is always a closed convex cone (see Figure 1.2, right).*

The main reason why we are interested in the tangent and the normal cones is the following theorem.

Theorem 1.5. *Let $f\colon X \to \mathbb{R}$, $A \subseteq X$ and $x^* \in A$ such that f is differentiable at x^*. If $x^* \in \operatorname{argmin}_A f$, then $-\nabla f(x^*) \in \mathcal{N}_A(x^*)$.*

Proof. We compare the values of f at x^* and at neighboring points. Let $d \in \mathcal{T}_A(x)$ and let $(x_k)_{k\in\mathbb{N}}$, $(t_k)_{k\in\mathbb{N}}$ be sequences as in (1.2). Since, x^* is a minimizer, we have

$$0 \leq \frac{f(x_k) - f(x^*)}{t_k} = \left\langle \nabla f(x^*),\, \frac{x_k - x^*}{t_k} \right\rangle + o\left(\frac{\|x_k - x^*\|}{t_k}\right).$$

Passing to the limit as $k \to +\infty$, we get

$$0 \leq \langle \nabla f(x^*),\, d\rangle\,.$$

Since this holds for every $d \in \mathcal{T}_A(x^*)$, we obtain the claimed result. □

Remark 1.6. *If x^* is in the interior of A, $\operatorname{int}(A)$, we obtain the classical necessary condition $\nabla f(x^*) = 0$.*

1.2.2 The case of convex sets

In the case where A is convex, a different characterization can be used.

Lemma 1.7. *If $C \subseteq X$ is a convex set and $x \in C$,*

$$\mathcal{T}_C(x) = \overline{\{\,\alpha(x' - x) \mid \alpha \geq 0, x' \in C\,\}} \tag{1.4}$$
$$\mathcal{N}_C(x) = \left\{\, s \in \mathbb{R}^p \mid \forall x' \in C, \left\langle s,\, x' - x\right\rangle \leq 0\,\right\} \tag{1.5}$$

In particular, the tangent cone to C is convex.

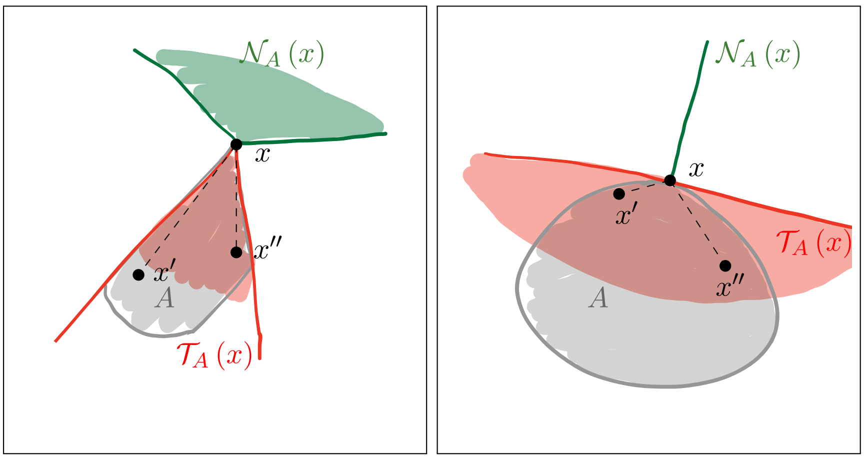

Figure 1.3: *(Left)* The tangent cone to a convex set A is obtained by considering all the vectors $(x' - x)$ for $x' \in A$, and taking the closure. *(Right)* When the convex set is smooth (with nonempty interior), the tangent cone is a half space containing A.

Proof. Let $\alpha \geq 0$, $x' \in C$, and let us prove that $\alpha(x'-x) \in \mathcal{T}_C(x)$. If $\alpha = 0$, there is nothing to prove, since $\alpha(x'-x) = 0 \in \mathcal{T}_C(x)$. If $\alpha > 0$, we let $x_k = x + \frac{1}{k+1}(x'-x)$ and $t_k = \frac{1}{\alpha(k+1)}$. Then $x_k \in C$, $t_k \to 0$ and

$$\frac{x_k - x}{t_k} = \alpha(x'-x)$$

so that $\alpha(x'-x) \in \mathcal{T}_C(x)$. Hence $\overline{\{\, \alpha(x'-x) \mid t\alpha \geq 0, x' \in C \,\}} \subseteq \mathcal{T}_C(x)$, and since $\mathcal{T}_C(x)$ is closed, the same holds for $\overline{\{\, \alpha(x'-x) \mid \alpha \geq 0, x' \in C \,\}}$.

Now, we prove the converse inclusion. Let $d \in \mathcal{T}_C(x)$ and, for $k \in \mathbb{N}$, let $x_k \in C$, $t_k > 0$ as in (1.2). Since $\frac{1}{t_k}(x_k - x) \in \{\, \alpha(x'-x) \mid \alpha \geq 0, x' \in C \,\}$, and that it converges to d, we get $d \in \overline{\{\, \alpha(x'-x) \mid \alpha \geq 0, x' \in C \,\}}$ and the claimed equality holds.

It remains to check (1.5). Let $s \in \mathcal{N}_C(x)$. For any $x' \in C$, we set $d = x' - x$ so that by (1.4), $d \in \mathcal{T}_C(x)$. The constraint $\langle s,\, d\rangle \leq 0$ exactly means that $\langle s,\, x'-x\rangle \leq 0$.

Conversely, if $s \in \mathbb{R}^p$ is such that for all $x' \in C$, $\langle s,\, x'-x\rangle \leq 0$, we consider any $d \in \mathcal{T}_C(x)$. By (1.4), there exists $d_k = t_k(x_k - x)$ such that $t_k > 0$, $x_k \in C$ and $d_k \to d$. Since

$$\langle s,\, d_k\rangle = t_k \langle s,\, x_k - x\rangle \leq 0,$$

we pass to the limit and obtain $\langle s,\, d\rangle \leq 0$, hence the converse inclusion holds. □

As an application, we deduce the following tangent cones.

Example 1.8 (The singleton)**.** *Let $A = \{x\} \subseteq \mathbb{R}^p$. Applying Lemma 1.7, we deduce that $\mathcal{T}_A(x) = \{0\}$ and $\mathcal{N}_A(x) = \mathbb{R}^p$.*

Example 1.9 (The negative cone)**.** *If $A =]-\infty, 0] \subseteq \mathbb{R}$, we obtain similarly that*

$$\mathcal{T}_A(x) = \begin{cases} \mathbb{R} & \text{if } x < 0, \\ ]-\infty, 0] & \text{if } x = 0, \end{cases} \quad \text{and} \quad \mathcal{N}_A(x) = \begin{cases} \{0\} & \text{if } x < 0, \\ [0, +\infty[& \text{if } x = 0. \end{cases} \tag{1.6}$$

Example 1.10 (The mixed case)**.** *More generally if we define*

$$Q_k \overset{\text{def.}}{=} \underbrace{\{0\} \times \cdots \times \{0\}}_{k \text{ times}} \times \underbrace{]-\infty, 0] \times \cdots \times]-\infty, 0]}_{p-k \text{ times}} \subseteq \mathbb{R}^p$$

Applying Lemma 1.7, we see that for all $d, s \in \mathbb{R}^p$,

$$d \in \mathcal{T}_{Q_k}(x) \iff \forall i \in \{1, \ldots, p\},\ d_i \in \begin{cases} \{0\} & \text{if } 1 \leq i \leq k, \\ \mathbb{R} & \text{if } k+1 \leq i \leq p \text{ and } x_i < 0, \\ ]-\infty, 0] & \text{if } k+1 \leq i \leq p \text{ and } x_i = 0, \end{cases} \tag{1.7}$$

$$s \in \mathcal{N}_{Q_k}(x) \iff \forall i \in \{1, \ldots, p\},\ s_i \in \begin{cases} \mathbb{R} & \text{if } 1 \leq i \leq k, \\ \{0\} & \text{if } k+1 \leq i \leq p \text{ and } x_i < 0, \\ [0, +\infty[& \text{if } k+1 \leq i \leq p \text{ and } x_i = 0, \end{cases} \tag{1.8}$$

(1.9)

Example 1.10 is a typical example of constraint set which involves equalities and inequalities. If $\{e_i\}_{1\leq i\leq p}$ denotes the canonical basis of $\mathbb{R}^p$, the optimality condition $-\nabla f(x^*) \in \mathcal{N}_{Q_k}(x^*)$ reformulates as

$$\begin{cases} \langle \nabla f(x^*),\, e_i\rangle = 0 & \text{if } k+1\leq i\leq p \text{ and } x_i<0,\\ \langle \nabla f(x^*),\, e_i\rangle \leq 0 & \text{if } k+1\leq i\leq p \text{ and } x_i=0. \end{cases}$$

Using Lemma 1.7, we may deduce a converse to Theorem 1.5 in the case where both A and f are convex.

Theorem 1.11. *Let $f\colon X\to\mathbb{R}$ convex, $C\subseteq X$ convex and $x^*\in C$ such that f is differentiable at x^*. Then, $x^*\in \operatorname{argmin}_C f$ if and only if $-\nabla f(x^*)\in \mathcal{N}_C(x^*)$.*

Proof. The "only if" part follows from Theorem 1.5. We prove the converse. If $-\nabla f(x^*)\in \mathcal{N}_A(x^*)$, then by (1.4),

$$\begin{aligned} \forall x'\in C,\ \langle -\nabla f(x^*),\, x-x^*\rangle &\leq 0,\\ \text{in other words,}\quad \langle \nabla f(x^*),\, x-x^*\rangle &\geq 0. \end{aligned}$$

But the convexity of f implies that

$$\forall x'\in C,\quad f(x')\geq f(x^*)+\left\langle \nabla f(x^*),\, x'-x^*\right\rangle \geq f(x^*), \tag{1.10}$$

hence $x^*\in \operatorname{argmin}_C f$.

□

1.2.3 Evolution with diffeomorphisms

The tangent and the normal cones are geometric objects. Their evolution when the space is deformed is given by the following proposition.

Proposition 1.12. *Let $X=\mathbb{R}^p$, $\tilde{X}=\mathbb{R}^p$, and $\varphi\colon X\to\tilde{X}$ be a $\mathscr{C}^1$-diffeomorphism. Let $A\subseteq X$ and $x\in A$, and define $\tilde{x}\stackrel{\text{def.}}{=}\varphi(x)$ and $\tilde{A}\stackrel{\text{def.}}{=}\varphi(A)$. Then,*

$$\mathcal{T}_{\tilde{A}}(\tilde{x}) = (\mathrm{D}\,\varphi(x))(\mathcal{T}_A(x)), \tag{1.11}$$

$$\mathcal{N}_{\tilde{A}}(\tilde{x}) = (\mathrm{D}\,\varphi(x))^{\top,-1}(\mathcal{N}_A(x)), \tag{1.12}$$

where $\mathrm{D}\,\varphi(x) = (\frac{\partial\varphi_i}{\partial x_j}(x))_{1\leq i,j\leq p}$ is the Jacobian matrix of φ at x.

Proof. We begin by proving (1.11). Let $d\in\mathcal{T}_A(x)$ and let $(x_k)_{k\in\mathbb{N}}$, $(t_k)_{k\in\mathbb{N}}$ be sequences as in (1.2). We consider $\tilde{x}_k\stackrel{\text{def.}}{=}\varphi(x_k)\in\tilde{A}$, and we note that $\tilde{x}_k\to\tilde{x}$ and

$$\frac{x_k-x}{t_k} = \frac{\varphi(x_k)-x_k}{t_k} = \frac{(D\varphi(x))(x_k-x)}{t_k} + o\left(\frac{\|x_k-x\|}{t_k}\right) \to (\mathrm{D}\,\varphi(x))d.$$

Therefore, $(\mathrm{D}\,\varphi(x))d\in\mathcal{T}_{\tilde{A}}(\tilde{x})$, and $(\mathrm{D}\,\varphi(x))\mathcal{T}_A(x)\subseteq\mathcal{T}_{\tilde{A}}(\tilde{x})$. Swapping the roles of A and $\tilde{A}$, we obtain similarly that $(\mathrm{D}\,\varphi^{-1}(\tilde{x}))\mathcal{T}_{\tilde{A}}(\tilde{x})\subseteq\mathcal{T}_A(x)$, which eventually yields (1.11), since $(\mathrm{D}\,\varphi^{-1}(\tilde{x})) = (\mathrm{D}\,\varphi(x))^{-1}$.

Now, we focus on (1.12). Let $\tilde{s} \in \mathbb{R}^p$. We note that $\tilde{s} \in \mathcal{N}_{\tilde{A}}(\tilde{x})$ if and only if

$$\begin{aligned}\left(\forall \tilde{d} \in \mathcal{T}_{\tilde{A}}(\tilde{x}),\ \left\langle \tilde{s},\, \tilde{d} \right\rangle \leq 0\right) &\iff (\forall d \in \mathcal{T}_A(x),\ \langle \tilde{s},\, (\mathrm{D}\,\varphi(x))d \rangle \leq 0) \\ &\iff \left(\forall d \in \mathcal{T}_A(x),\ \left\langle (\mathrm{D}\,\varphi(x))^\top \tilde{s},\, d \right\rangle \leq 0\right)\end{aligned}$$

which means that $(\mathrm{D}\,\varphi(x))^\top \tilde{s} \in \mathcal{N}_A(x)$, that is $\tilde{s} \in (\mathrm{D}\,\varphi(x))^{\top,-1}(\mathcal{N}_A(x))$. □

1.3 Sets defined by constraint functions

To make the most of Theorem 1.5, it becomes crucial to describe the normal cone $\mathcal{N}_A(x^*)$ as accurately as possible. In practical cases, the set A is usually not given as a collection of points, but as the set of points which should satisfy some equality or inequality constraints,

$$A = \left\{ x \in X \mid \begin{array}{c} g_1(x) = 0, \ldots, g_n(x) = 0, \\ h_1(x) \leq 0, \ldots, h_m(x) \leq 0 \end{array} \right\}. \tag{1.13}$$

Unless otherwise stated, we assume in the rest of this chapter that the constraint functions are smooth, $g_i \in \mathscr{C}^1(X)$, for $1 \leq i \leq n$ and $h_j \in \mathscr{C}^1(X)$ for $1 \leq j \leq m$.

How can we deduce $\mathcal{N}_A(x)$ from the functions $\{g_i\}_{i=1}^n$ and $\{h_j\}_{j=1}^m$?

1.3.1 Constraint functions yield a representation of A

A key idea to describe the normal cone is the following remark (well-known to physicists):

the gradient of a function is orthogonal to its level lines.

However, a given constraint set may be described in many different ways (*i.e.* different choices of functions g_i, h_j), which are not equally convenient.

Abstract or geometric description. For instance, consider the set $A = S(0,1)$ (the Euclidean unit sphere). A drawing suggests that the tangent cone is given by the tangent (hyper)plane to the sphere, $\mathcal{N}_A(x) = \{x\}^\perp$, while the normal cone is given by $\mathcal{N}_A(x) = \mathrm{Vect}\{x\}$. That turns out to be true, of course! You may check it using the tools given below.

Analytic description. A simple way to describe A would be $A = \{\, x \in X \mid g(x) = 0 \,\}$ where $g(x) = 1 - \|x\|^2$. For every $x \in A$, we note that $\nabla g(x) = -2x$ so that it spans the normal cone, $\mathrm{Vect}\{\nabla g(x)\} = \mathcal{N}_A(x)$.

Analytic description (bis). We may as well write $A = \{\, x \in X \mid g(x) = 0 \,\}$ where $g(x) = \left(1 - \|x\|^2\right)^2$. The function g is smooth and yields exactly the same set. However, for every $x \in A$,

$$\nabla g(x) = 2\left(1 - \|x\|^2\right)(-2x) = 0,$$

so that $\mathrm{Vect}\{\nabla g(x)\} = \{0\} \subsetneq \mathcal{N}_A(x)$.

Exercise 1.13. *Consider the set $A = \overline{B(0,1)}$ and describe its normal cone. Compare its representation using the constraint functions $h(x) = \|x\|^2 - 1$ and $h(x) = \left[\max(\left(\|x\|^2 - 1, 0\right)\right]^2$.*

As we see, writing the set A using (1.13) yields *a representation* of A, but that representation is certainly not unique. In the first description above, the gradient of the constraint provides enough information to determine the normal cone at each point, contrary to the second description. In the following, our main concern is to understand whether a certain representation provides enough information to fully describe this normal cone using the gradients of the constraints. That is usually called *constraint qualification.*

1.3.2 Constraint gymnastics

Since some constraints are better than others for representing the same set, it is important to know how to switch from one representation to another.

Active constraints. A first remark is that inactive constraints have no influence on the normal cone.

Definition 1.14 (Active constraints)**.** *Let A as in* (1.13) *and let $x \in A$. We say that the constraint $\{h_j = 0\}$ is* active *at x if $h_j(x) = 0$. Otherwise, $h_j(x) < 0$, and we say that the constraint is* inactive*. The collection of active constraint is indexed by $\mathcal{A}(x) \overset{\text{def.}}{=} \{\, j \in \{1, \ldots, m\} \mid h_j(x) = 0 \,\}$.*

If $h_j(x) < 0$ (*inactive constraint)*, the continuity of h_j implies that $h_j(x') < 0$ for x' in a neighborhood of x. In other words, there exists $r > 0$ such that

$$A \cap B(x,r) = \left\{ x' \in B(x,r) \mid \begin{array}{r} g_1(x') = 0, \ldots, g_n(x') = 0, \\ \forall j \in \mathcal{A}(x),\ h_j(x') \leq 0 \end{array} \right\}.$$

and so that, locally, the inactive contraints do not matter.

As a result, *the inactive constraints can be omitted when computing the tangent cone* (hence the normal cone),

$$\forall d \in \mathbb{R}^p, \quad d \in \mathcal{T}_A(x) \iff \exists \{t_k\}_{k\in\mathbb{N}} \subseteq \mathbb{R}_+^*, \exists \{x_k\}_{k\in\mathbb{N}} \subseteq X \text{ such that,} \\ \begin{cases} \forall i \in \{1,\dots,n\},\ g_i(x_k) = 0, \\ \forall j \in \mathcal{A}(x),\ h_j(x_k) \leq 0, \\ \lim_{k\to\infty} \left(\frac{x_k - x}{t_k}\right) = d. \end{cases} \tag{1.14}$$

The inactive constraints have *locally* no effect, but they may have a strong *global influence*, on the set of minimizers. Compare for instance the problem

$$\min_{x\geq 0}(x^2-1)^2 \tag{1.15}$$

with

$$\min_{x\in\mathbb{R}}(x^2-1)^2. \tag{1.16}$$

The solution set to (1.15) is obviously $\{1\}$, and the constraint is inactive at the minimizer. However if we remove it (as in (1.16)), the solution set is $\{-1, 1\}$. It is even possible to modify the objective so that 1 is not even a minimizer of (1.16).

Rescaling. Additionnally, it is possible to reformulate a constraint $\{h_j \leq 0\}$ by a change of variable: if $\varphi\colon \mathbb{R} \to \mathbb{R}$ is strictly increasing with $\varphi(0) = 0$, then

$$h_j(x) \leq 0 \iff \varphi(h_j(x)) \leq 0. \tag{1.17}$$

Eventually, it is possible to reformulate the constraints, to switch from one type to another. This may make a lot of difference when formulating algorithms.

Equality to inequalities. For all $x \in X$,

$$g_i(x) = 0 \iff g_i(x) \leq 0 \quad \text{and} \quad (-g_i(x)) \leq 0. \tag{1.18}$$

Inequality to equality. For all $x \in X$,

$$h_j(x) \leq 0 \iff \exists s_j \in \mathbb{R},\ h_j(x) + s_j^2 = 0. \tag{1.19}$$

Inequality to equality and elementary inequality. For all $x \in X$,

$$h_j(x) \leq 0 \iff \exists s_j \geq 0,\ h_j(x) + s_j = 0. \tag{1.20}$$

In the last two cases, we need to introduce a new variable s_j. It is sometimes called a *slack variable*.

1.3.3 Gradients, constraints and qualification

The following result is central to our discussion. It shows that the gradients of the constraints at $x \in A$ are tightly connected to the tangent and normal cones at x. Under some conditions, they are even sufficient to describe the normal cone.

Proposition 1.15. *Let A as in (1.13), with $g_1, \dots, g_n, h_1, \dots, h_m \colon X \to \mathbb{R}$ all differentiable at $x \in A$. Then*

$$\left\{ d \in \mathbb{R}^p \mid \begin{array}{r} \forall i \in \{1, \dots, n\},\ \langle \nabla g_i(x),\, d \rangle = 0, \\ \forall j \in \mathcal{A}(x),\ \langle \nabla h_j(x),\, d \rangle \le 0. \end{array} \right\} \supseteq \mathcal{T}_A\left(x\right), \tag{1.21}$$

$$\left\{ \sum_{i=1}^{n} \lambda_i \nabla g_i(x) + \sum_{j \in \mathcal{A}(x)} \mu_j \nabla h_j(x) \mid \lambda \in \mathbb{R}^n,\ \mu \in (\mathbb{R}_+)^{\mathcal{A}(x)} \right\} \subseteq \mathcal{N}_A\left(x\right). \tag{1.22}$$

Proof. Let $d \in \mathcal{T}_A(x)$, and let $\{x_k\}_{k\in\mathbb{N}} \subseteq A$, $\{t_k\}_{k\in\mathbb{N}} \subseteq \mathbb{R}_+^*$ such as in (1.2). For each $k \in \mathbb{N}$ and each $i \in \{1, \dots, n\}$, since x_k is feasible,

$$0 = \frac{g_i(x_k) - g_i(x)}{t_k} = \left\langle \nabla g_i(x),\, \frac{x_k - x}{t_k} \right\rangle + o\left(\frac{\|x_k - x\|}{t_k}\right).$$

Passing to the limit $k \to +\infty$, we obtain $\langle \nabla g_i(x),\, d \rangle = 0$.

Moreover, for $j \in \mathcal{A}(x)$, we have

$$0 \ge \frac{h_j(x_k)}{t_k} = \frac{h_j(x_k) - h_j(x)}{t_k} = \left\langle \nabla h_j(x),\, \frac{x_k - x}{t_k} \right\rangle + o\left(\frac{\|x_k - x\|}{t_k}\right),$$

and we obtain similarly that $0 \ge \langle \nabla h_j(x),\, d \rangle$. Thus, we have obtained (1.21).

The dual result, on the normal cones, follows from Farkas' Lemma (Lemma A.1) and the fact that the polar is a nonincreasing operation. $\square$

Nevertheless, as we have seen in Section 1.3.1, those inclusions may be strict. Our main focus is on the equality case.

Definition 1.16 (Qualified constraints)**.** *We say that the constraints are qualified at $x \in A$ if (1.22) is an equality.*

1.3.4 Three constraint qualification properties

A quick way to assert that the constraints are qualified is to check that they are *affine*.

Proposition 1.17 (Affine constraints are qualified)**.** *Let A as in (1.13), where the constraint functions are affine, i.e.*

$$\forall i \in \{1, \dots, n\}, \quad g_i(x) = \langle a_i,\, x \rangle - b_i \tag{1.23}$$

$$\forall j \in \{1, \dots, m\}, \quad h_j(x) = \langle p_j,\, x \rangle - q_j, \tag{1.24}$$

with $a_i, p_j \in \mathbb{R}^p$*,* $b_i, q_j \in \mathbb{R}$ *for* $1 \leq i \leq n$*,* $1 \leq j \leq m$*. Then, for all* $x \in A$*,*

$$\mathcal{T}_A(x) = \left\{ d \in \mathbb{R}^p \mid \begin{array}{r} \forall i \in \{1, \dots, n\},\ \langle a_i, d \rangle = 0, \\ \forall j \in \mathcal{A}(x),\ \langle p_j, d \rangle \leq 0 \end{array} \right\}, \tag{1.25}$$

$$\mathcal{N}_A(x) = \left\{ \sum_{i=1}^n \lambda_i a_i + \sum_{j \in \mathcal{A}(x)} \mu_j p_j \in \mathbb{R}^p \mid \lambda \in \mathbb{R}^n, \mu \in (\mathbb{R}_+)^{|\mathcal{A}(x)|} \right\}. \tag{1.26}$$

Proof. By Proposition 1.15 we know that $\mathcal{T}_A(x)$ is included in the right-hand side of (1.25), hence we only need to prove the converse inclusion.

Let $d \in \mathbb{R}^p$ such that $\langle a_i, d \rangle = 0$ for all $i \in \{1, \dots, n\}$, $\langle p_j, d \rangle \leq 0$ for all $j \in \mathcal{A}(x)$. We set $t_k = 1/(k+1)$ and $x_k = x + t_k d$. We see that $t_k \to 0^+$, $\lim_{k \to \infty}(x_k - x)/t_k = d$, and

$$\begin{aligned} \forall i \in \{1, \dots, n\}, &\quad < a_i, x_k > - b_i = 0, \\ \forall j \in \mathcal{A}(x), &\quad \langle p_j, x_k \rangle - q_j = 0 + t_k \langle p_j, d \rangle \leq 0, \\ \forall j \in \{1, \dots, m\} \setminus \mathcal{A}(x), &\quad \langle p_j, x_k \rangle - q_j = (\langle p_j, x \rangle - q_j) + t_k \langle d, p_j \rangle < 0 \end{aligned}$$

where the last inequality holds for k large enough, since $\langle p_j, x \rangle - q_j < 0$.

As a result, $d \in \mathcal{T}_A(x)$, and (1.25) is an equality.

We deduce the expression for $\mathcal{N}_A(x)$ in (1.26) by applying Farkas' Lemma (Lemma A.1) and its consequence Proposition A.2.

□

In the more general case of a convex constraint set A, Slater's conditions are an alternative way to ensure the constraint qualification.

Proposition 1.18 (Slater's conditions)**.** *Let* A *as in* (1.13)*, where* $h_j \colon X \to \mathbb{R}$ *is* convex*, differentiable at* $x \in A$*, and where the equality constraints are affine, i.e.*

$$\forall i \in \{1, \dots, n\},\ \forall x' \in X, \quad g_i(x') = \langle a_i, x' \rangle - b_i, \tag{1.27}$$

with $a_i \in \mathbb{R}^p$*,* $b_i \in \mathbb{R}$ *for* $1 \leq i \leq n$*.*

Assume moreover, that there exists a strict feasible point, i.e. some point $x_0 \in A$ *such that* $h_j(x_0) < 0$ *for all* $j \in \{1, \dots, m\}$ *such that* h_j is not affine.

Then, for all $x \in A$*,*

$$\mathcal{T}_A(x) = \left\{ d \in \mathbb{R}^p \mid \begin{array}{r} \forall i \in \{1, \dots, n\},\ \langle a_i, d \rangle = 0, \\ \forall j \in \mathcal{A}(x),\ \langle \nabla h_j(x), d \rangle \leq 0 \end{array} \right\}, \tag{1.28}$$

$$\mathcal{N}_A(x) = \left\{ \sum_{i=1}^n \lambda_i a_i + \sum_{j \in \mathcal{A}(x)} \mu_j \nabla h_j(x) \in \mathbb{R}^p \mid \lambda \in \mathbb{R}^n, \mu \in (\mathbb{R}_+)^{|\mathcal{A}(x)|} \right\}. \tag{1.29}$$

Remark 1.19. *The set* A *in Proposition 1.18 is convex. Would it still be the case if we had only assumed that each* g_i *is convex?*

Proof of Proposition 1.18. By Proposition 1.15 we know that the right-hand side of (1.28) is included in $\mathcal{T}_A(x)$, hence we only need to prove the converse inclusion. Let $d \in \mathbb{R}^p$ such that $\langle a_i, d\rangle = 0$ for all $i \in \{1, \ldots, n\}$ and $\nabla h_j(x) \leq 0$ for all $j \in \mathcal{A}(x)$.

Let $\varepsilon > 0$, and define $d_\varepsilon \stackrel{\text{def.}}{=} d + \varepsilon(x_0 - x)$. First, we prove that $d_\varepsilon \in \mathcal{T}_A(x)$. Let $t_k = 1/(k+1)$ and $x_k = x + t_k d_\varepsilon$. We see that $t_k \to 0^+$, $(x_k - x)/t_k \to d_\varepsilon$ and, by construction, for all $i \in \{1, \ldots, n\}$,

$$\langle a_i, x_k\rangle = \langle a_i, x\rangle + t_k \left(\langle a_i, d\rangle + \varepsilon \langle a_i, x_0 - x\rangle\right) = 0.$$

Now, let $j \in \mathcal{A}(x)$. If h_j is affine,

$$\begin{aligned} h_j(x_k) &= h_j(x + t_k\varepsilon(x_0 - x)) + t_k \langle \nabla h_j(x), d\rangle \\ &= (1 - t_k\varepsilon)\underbrace{h_j(x)}_{\leq 0} + t_k\varepsilon \underbrace{h_j(x_0)}_{\leq 0} + t_k \underbrace{\langle \nabla h_j(x), d\rangle}_{\leq 0} \\ &\leq 0 \quad \text{for } k \text{ large enough.} \end{aligned}$$

On the other hand, if h_j is not affine, the convexity of h_j implies

$$\begin{aligned} 0 > h_j(x_0) &\geq \underbrace{h_j(x)}_{=0} + \langle \nabla h_j(x), x_0 - x\rangle, \\ \text{hence} \quad h_j(x_k) &= \underbrace{h_j(x)}_{=0} + \langle \nabla h_j(x), x_k - x\rangle + o\left(\|x_k - x\|\right) \\ &= t_k \left(\underbrace{\langle \nabla h_j(x), d\rangle}_{\leq 0} + \varepsilon \underbrace{\langle \nabla h_j(x), x_0 - x\rangle}_{<0} + o(1) \right) \end{aligned}$$

As a result, for k large enough, $h_j(x_k) < 0$ and we obtain that $x_k \in A$. Thus, $d_\varepsilon \in \mathcal{T}_A(x)$, and since $\mathcal{T}_A(x)$ is closed, we see that $d \in \mathcal{T}_A(x)$ by letting $\varepsilon \to 0^+$. Thus, we have proved (1.28).

The expression for $\mathcal{N}_A(x)$ in (1.29) follows by applying Farkas' Lemma (Lemma A.1) and its consequence Proposition A.2.

□

Remark 1.20. *It is important to note that the strict inequality $h_j(x_0) < 0$ is not required if h_j is affine. As a consequence, Proposition 1.18 (Slater's conditions) truly is a generalization of Proposition 1.17 (affine case).*

Eventually, the linear independence criterion may be applied both in the convex and nonconvex cases.

Proposition 1.21 (Linear Independence Constraint Qualification (LICQ))**.** *Let A as in (1.13) with $g_i \in \mathscr{C}^1(X)$ for $1 \leq i \leq n$, $h_j \in \mathscr{C}^1(X)$ for $1 \leq j \leq m$, and let $x \in A$.*

If the family $\{\nabla g_i(x)\}_{i=1}^n \cup \{\nabla h_j(x)\}_{j \in \mathcal{A}(x)}$ is linearly independent, then the constraints are qualified at x,

$$\mathcal{T}_A(x) = \left\{ d \in \mathbb{R}^p \mid \begin{array}{r} \forall i \in \{1, \ldots, n\},\ \langle \nabla g_i(x), d\rangle = 0, \\ \forall j \in \mathcal{A}(x),\ \langle \nabla h_j(x), d\rangle \leq 0 \end{array} \right\}, \tag{1.30}$$

$$\mathcal{N}_A(x) = \left\{ \sum_{i=1}^n \lambda_i \nabla g_i(x) + \sum_{j \in \mathcal{A}(x)} \mu_j \nabla h_j(x) \in \mathbb{R}^p \mid \lambda \in \mathbb{R}^n, \mu \in (\mathbb{R}_+)^{|\mathcal{A}(x)|} \right\}. \tag{1.31}$$

Proof (♠). Let $x \in A$. Since the inactive constraints have no influence on the tangent cone, we may ignore them and assume that $\mathcal{A}(x) = \{1, \ldots, m\}$.

We introduce the mapping

$$\varphi\colon \begin{array}{l} X \longrightarrow \mathbb{R}^{n+m} \\ x' \longmapsto \left(g_1(x'), \cdots, g_n(x'), h_1(x'), \cdots, h_m(x')\right)^\top \end{array}$$

We note that $x' \in A$ if and only if $\varphi(x') \in Q$,where Q is the set studied in Example 1.10,

$$Q \overset{\text{def.}}{=} \underbrace{\{0\} \times \cdots \times \{0\}}_{n \text{ times}} \times \underbrace{]-\infty, 0] \times \cdots \times]-\infty, 0]}_{m \text{ times}}.$$

Moreover, $\varphi(x) = (0, \ldots, 0)$, and $\mathcal{T}_Q(0) = Q$ (see Example 1.10).

First, in the case where $n + m = p$, the linear independence assumption ensures that $\mathrm{D}\,\varphi(x)$ is invertible. Hence, by the local inversion theorem, φ is locally a diffeomorphism which maps (a neighborhood of x in) A to (a neighborhood of 0 in) Q. By Proposition 1.12, we deduce that $\mathcal{T}_Q(0) = (\mathrm{D}\,\varphi(x))(\mathcal{T}_A(x))$ and $\mathcal{N}_Q(0) = (\mathrm{D}\,\varphi(x))^{-1,\top}\mathcal{N}_A(x)$. In other words, $\mathcal{T}_A(x) = (\mathrm{D}\,\varphi(x))^{-1}\mathcal{T}_Q(0)$ and $\mathcal{N}_A(x) = (\mathrm{D}\,\varphi(x))^\top\mathcal{N}_Q(0)$, which precisely mean (1.30) and (1.31) respectively.

Then, in the case $n + m < p$ (the converse inequality is impossible, by the linear independence assumption), we complete φ to make it an isomorphism and conclude similarly. Indeed, if P denotes the orthogonal projector onto $\ker \mathrm{D}\,\varphi(x)$, the first isomorphism theorem ensures that the linear map $\begin{pmatrix} P & \mathrm{D}\,\varphi(x) \end{pmatrix} : \mathbb{R}^p \to (\ker \mathrm{D}\,\varphi(x) \oplus \operatorname{Im} \mathrm{D}\,\varphi(x)) \approx \mathbb{R}^p$ is an isomorphism. As a result, the mapping $\tilde{\varphi}\colon x' \mapsto (P(x' - x), \varphi(x'))$ is locally an isomorphism which maps a neighborhood of x in A to a neighborhood of 0×0 in $\mathbb{R}^{p-(n+m)} \times Q$. We deduce as above that

$$\mathcal{T}_A(x) = \left\{ d \in \mathbb{R}^p \mid (\mathrm{D}\,\tilde{\varphi}(x))d \in \mathbb{R}^{p-(n+m)} \times Q \right\} = \{ d \in \mathbb{R}^p \mid (\mathrm{D}\,\varphi(x))d \in \times Q \},$$

which is precisely (1.30). Using Farkas' Lemma (Lemma A.1), we deduce (1.31) as well. □

It is worth noting the differences between the three qualification criteria:

- The constraint qualification provided by Slater's conditions (as well as the affine criterion) is *global*: checking these condistions ensures that the constraints are qualified at every point of A.

- On the contrary, the linear independence criterion (LICQ) only ensures the qualification of the constraint at the point x. But it may be applied to nonconvex sets A. Therefore, it is adapted to subtle situations where the constraints are not qualified at every point of A, but only on a small subset.

- However, LICQ is not more general than Slater's conditions, since Slater's conditions may hold with linearly dependent gradients.

1.4 The Karush, Kuhn, and Tucker (KKT) conditions

1.4.1 The KKT conditions in the general case

Motivated by the description of the normal cone in the qualified case (equality in (1.22)), we introduce the so-called *Lagrangian* function,

$$\forall(x,\lambda,\mu)\in X\times\mathbb{R}^n\times\mathbb{R}^m,\ \mathcal{L}(x,\lambda,\mu)\overset{\text{def.}}{=}f(x)+\sum_{i=1}^n\lambda_i g_i(x)+\sum_{j=1}^m\mu_j h_j(x). \tag{1.32}$$

Reformulating Theorem 1.5 we obtain:

Theorem 1.22 (Karush, Kuhn, Tucker)**.** *Let A as in (1.13) and let $x^*\in A$ such that the constraints are qualified at x^*. Let $f\colon X\to\mathbb{R}$, differentiable at x^* and assume that x^* is a (local) minimizer of f over A.*

Then there exists $\mu^*\in\mathbb{R}^n$, $\lambda^*\in\mathbb{R}^m$ such that

$$\nabla_x\mathcal{L}(x^*,\lambda^*,\mu^*)=0, \tag{1.33}$$

$$\forall i\in\{1,\dots,n\},\quad g_i(x^*)=0, \tag{1.34}$$

$$\forall j\in\{1,\dots,m\},\quad h_j(x^*)\leq 0, \tag{1.35}$$

$$\forall j\in\{1,\dots,m\},\quad \mu_j^*\geq 0, \tag{1.36}$$

$$\forall j\in\{1,\dots,m\},\quad \mu_j^*h_j(x^*)=0. \tag{1.37}$$

The coefficients $(\lambda_i^*)_{i=1}^n$ and $(\mu_j^*)_{j=1}^m$ are called the *Lagrange multipliers* associated to μ^*. The equations (1.33) to (1.37) are called the *KKT conditions.* They are necessary conditions for f to have a minimizer on A.

Note that (1.34) and (1.35) simply express that x^* is an admissible point.

The last equation, (1.37) is called the *complementary slackness* condition. It expresses that the Lagrange multiplier corresponding to an inequality constraint *vanishes when the constraint is inactive.*

Proof. By Theorem 1.5, if x^* is a (local) minimizer, $-\nabla f(x^*)\in\mathcal{N}_A(x^*)$. Equivalently, since the constraints are qualified, there exists $(\lambda_i^*)_{i=1}^n\in\mathbb{R}^n$ and $(\mu_j^*)_{j\in\mathcal{A}(x^*)}\in(\mathbb{R}_+)^{|\mathcal{A}(x^*)|}$ such that

$$-\nabla f(x^*)=\sum_{i=1}^n\lambda_i^*\nabla g_i(x^*)+\sum_{j\in\mathcal{A}(x^*)}\mu_j^*\nabla h_j(x^*)$$

We complete the vector μ^* by setting $\mu_j^*=0$ for $j\in\{1,\dots,m\}\setminus\mathcal{A}(x^*)$, and we see that Equations (1.33) to (1.37) hold. □

Remark 1.23. *The Lagrangian function is designed so that*

$$\forall(x,\lambda,\mu)\in A\times\mathbb{R}^n\times(\mathbb{R}_+)^m,\quad\mathcal{L}(x,\lambda,\mu)\leq f(x).$$

Yet, at the optimum, $\mathcal{L}(x^*,\lambda^*,\mu^*)=f(x^*)$.

1.4.2 Convex problems and Lagrangian duality

If both f and A are convex, we know from Theorem 1.11 that the condition $-\nabla f(x^*) \in \mathcal{N}_A(x^*)$ is necessary and sufficient for x^* to be a global minimizer. We therefore obtain:

Proposition 1.24. *Let A as in (1.13) where $h_j : X \to \mathbb{R}$ is* convex*, differentiable at $x \in A$, and where the equality constraints are affine, i.e.*

$$\forall i \in \{1, \dots, n\},\ \forall x' \in X, \quad g_i(x') = \langle a_i, x' \rangle - b_i, \tag{1.38}$$

with $a_i \in \mathbb{R}^p$, $b_i \in \mathbb{R}$ for $1 \le i \le n$. Assume that the constraints are qualified. Let $f : X \to \mathbb{R}$, differentiable at x^* and convex.

Then $x^* \in A$ is a global minimizer of f over A if and only if the KKT conditions (1.33) to (1.37) hold.

Proposition 1.24 has an interesting consequence called *Lagrangian duality*. We assume from now on that $f, h_j \in \mathscr{C}^1(X)$ and they are convex, and each g_i is affine. By general properties of the infimum and the supremum, we always have

$$\inf_{x \in X} \sup_{(\lambda,\mu) \in \mathbb{R}^n \times (\mathbb{R}_+)^m} \mathcal{L}(x, \lambda, \mu) \ge \sup_{(\lambda,\mu) \in \mathbb{R}^n \times (\mathbb{R}_+)^m} \inf_{x \in X} \mathcal{L}(x, \lambda, \mu), \tag{1.39}$$

which is known as the weak duality inequality. In general, that inequality is strict.

Now, observe that for fixed $x \in X$,

$$\sup_{(\lambda,\mu) \in \mathbb{R}^n \times (\mathbb{R}_+)^m} \mathcal{L}(x, \lambda, \mu) = \begin{cases} f(x) & \text{if } x \in A, \\ +\infty & \text{otherwise.} \end{cases}$$

As a result, the left-hand side of(1.39) is equal to $\min_{x \in A} f(x) = f(x^*)$.

On the other hand, let us fix $(\lambda, \mu) \in \mathbb{R}^n \times (\mathbb{R}_+)^m$. By convexity of f and h_j for $1 \le j \le m$, the function $\mathcal{L}(\cdot, \lambda, \mu)$ is convex. It has a minimizer at $x \in X$ if and only if $\nabla_x \mathcal{L}(\cdot, \lambda, \mu) = 0$. If x^* is a minimizer of f over A, Proposition 1.24 precisely ensures that $\nabla_x \mathcal{L}(x^*, \lambda^*, \mu^*) = 0$, so that $\mathcal{L}(x^*, \lambda^*, \mu^*) = \min_{x \in X} \mathcal{L}(x, \lambda^*, \mu^*)$. As a result

$$\sup_{(\lambda,\mu) \in \mathbb{R}^n \times (\mathbb{R}_+)^m} \inf_{x \in X} \mathcal{L}(x, \lambda, \mu) \ge \min_{x \in X} \mathcal{L}(x, \lambda^*, \mu^*) = \mathcal{L}(x^*, \lambda^*, \mu^*).$$

Since $\mathcal{L}(x^*, \lambda^*, \mu^*) = f(x^*)$, we deduce that (1.39) is an equality, that is,

$$\inf_{x \in X} \sup_{(\lambda,\mu) \in \mathbb{R}^n \times (\mathbb{R}_+)^m} \mathcal{L}(x, \lambda, \mu) = \sup_{(\lambda,\mu) \in \mathbb{R}^n \times (\mathbb{R}_+)^m} \inf_{x \in X} \mathcal{L}(x, \lambda, \mu). \tag{1.40}$$

We say that *strong duality* holds.

Remark 1.25. *It is sometimes easier to study the dual problem, that is*

$$\sup_{(\lambda,\mu)\in\mathbb{R}^n\times(\mathbb{R}_+)^m} \left(\inf_{x\in X} \mathcal{L}(x,\lambda,\mu) \right), \tag{1.41}$$

rather than the primal one (1.1)*, The KKT conditions relate the solutions of the dual and primal problems.*

Chapter 2

Algorithms for constrained optimization

This chapter is devoted to the resolution of constrained optimization problems. The considered algorithms can mainly be divided into two classes.

First, algorithms which only produce *feasible iterates*, such as the projected gradient descent and the conditional gradient (also known as Frank-Wolfe) algorithm. In some situations it is more important to return a vector which approximately minimizes the objective, but which exactly satisfies the constraints. For instance, imagine that you try to recover a probability vector $(p_i)_{i=1}^n$ such that $p_i \geq 0$ for $1 \leq i \leq n$, and $\sum_{i=1}^n p_i = 1$. How would you interpret the result if the algorithm returned a vector with $p_j < 0$ for some $j \in [\![1\,;n]\!]$, or $\sum_{i=1}^n p_i \neq 1$?

Second, there are algorithms which produce iterates which only *approximately satisfy the contraint*. Even though it is less satisfying, the

2.1 Methods with feasible iterates

2.1.1 Projected gradient descent

The projected gradient descent is a straightforward adaptation of the classical gradient descent. It consists in projecting back onto the feasible set after performing each step of the gradient descent.

For a given sequence of stepsizes $(\tau_k)_{k\in\mathbb{N}} \in (\mathbb{R}_+)^{\mathbb{N}}$, the algorithm is given in Algorithm 1.

While we have not set any stopping criterion, one may use the standard empirical criteria, such as $f(x_k) - f(x_{k+1}) \leq \varepsilon$ or $|x_{k+1} - x_k| \leq \varepsilon$.

On the projection step. When $A = X$ (unconstrained problem), the algorithm amounts to the classical gradient descent. In the general case, it is crucial to be able to project onto A efficiently. If A is not convex (or not

Algorithm 1 Projected gradient descent algorithm

```
Choose x_0 ∈ A, k ← 0,
while True do
    x_{k+1} ← Proj_A(x_k − τ_k ∇f(x_k))
    k ← k + 1
end while
```

closed), that projection might not even exist! In fact, even if A is closed and convex (the projection is then well-defined), the projection might not be straightforward to compute. It is an optimization problem in itself.

Nevertheless, for several classical sets (convex or not), the projection operator is well known.

Euclidean ball If $A = \overline{B(0, R)} = \{ x \in X \mid \|x\|_2 \leq R \}$, the projection is given by

$$\mathrm{Proj}_A(x) = \frac{R}{\max\left(\|x\|_2, R\right)} x. \tag{2.1}$$

Cuboid If $A = [a_1, b_1] \times \cdots \times [a_p, b_p]$, then

$$\mathrm{Proj}_A(x) = \left(\mathrm{Proj}_{[a_1,b_1]}(x_1), \cdots, \mathrm{Proj}_{[a_p,b_p]}(x_p)\right)$$

$$\text{where} \quad \mathrm{Proj}_{[a,b]}(t) = \max\left(\min\left(t, b\right), a\right) = \begin{cases} a & \text{if } t \leq a \\ t & \text{if } a \leq t \leq b \\ b & \text{if } t \geq b. \end{cases}$$

Unit simplex If $A = \{ x \in X \mid \forall i \in [\![1\,;p]\!],\ x_i \geq 0, \text{and } \sum_{i=1}^{p} x_i = 1 \}$, the projection can be computed with sorting operations, with complexity $O\left(p \log(p)\right)$, see [Con16].

Euclidean sphere Let $A = S(0, 1) = \{ x \in X \mid \|x\|_2 = 1 \}$. That set is not convex. However

$$\forall x \neq 0, \quad \mathrm{Proj}_A(x) = \frac{1}{\|x\|_2} x. \tag{2.2}$$

For $x = 0$, the projection is ill-defined.

Just like the standard gradient descent, the projected gradient descent has many variants and may be applied to both convex or nonconvex functions f. Still, the convex case is the easiest to study, and we provide an example of convergence result below.

Theorem 2.1. *Let $A \subseteq X$ be a nonempty closed convex set, and let $f \in \mathscr{C}^{1,1}(X)$ with $L \stackrel{\text{def.}}{=} \operatorname{Lip}(\nabla f)$ such that $\operatorname{argmin}_A f \neq \emptyset$. Set $\tau_k = \tau$ for all $k \in \mathbb{N}$, with $0 < \tau < 1/L$.*

Then Algorithm 1 converges towards some solution of $\min_A f$.

Moreover,

$$f(x_k) - \min_A f \leq \frac{1}{2\tau k} \left(\operatorname{dist}(x_0, \operatorname{argmin}_A f)\right)^2. \tag{2.3}$$

In this convex case, the projected gradient descent can be seen as a proximal method known as *forward-backward splitting.* Set

$$g(x) = \chi_A(x) \stackrel{\text{def.}}{=} \begin{cases} 0 & \text{if } x \in A \\ +\infty & \text{otherwise.} \end{cases}. \tag{2.4}$$

The problem (1.1) can be reformulated as

$$\min_{x \in X} f(x) + g(x). \tag{2.5}$$

The function f is smooth ($\mathscr{C}^1(X)$). On the contrary, g is not smooth but it is possible to compute its proximity operator

$$\forall \tau > 0,\ \operatorname{prox}_{\tau g}(x) \stackrel{\text{def.}}{=} \operatorname{argmin}_{y \in X} \left(g(y) + \frac{1}{2\tau} \|x - y\|^2 \right) = \operatorname{Proj}_A(x). \tag{2.6}$$

The projected gradient algorithm Algorithm 1 may then be reformulated as alternating between a gradient descent step and an application of a proximity operator,

$$\forall k \in \mathbb{N},\ x_{k+1} = \operatorname{prox}_{\tau g}\left(x_k - \tau \nabla f(x_k)\right). \tag{2.7}$$

The convergence claimed in Theorem 2.1 follows from general results on the forward-backward splitting method, which will be studied later. Therefore, we omit its proof.

2.1.2 The conditional gradient algorithm (Frank-Wolfe)

We assume now that $A \subseteq X$ is a nonempty *compact* convex set, and that $f \in \mathscr{C}^1(X)$. At each iteration, instead of directly minimizing f over A, we minimize its *linearization*, $s \mapsto (f(x_k) + \langle \nabla f(x_k), s - x_k \rangle)$. Since A is compact, this surrogate problem has a solution.

The corresponding point s_{k+1} gives us a descent direction, and the next iterate is given by a convex combination. The weight of this convex combination can be set using different rules, for instance

$$\theta_k \in \operatorname{argmin}_{\theta \in [0,1]} f(\theta_k s_{k+1} + (1 - \theta_k) x_k) \qquad \textit{(exact linesearch)} \tag{2.8}$$

$$\text{or} \quad \theta_k = \frac{2}{k+2} \qquad \textit{(prescribed stepsize)}. \tag{2.9}$$

Algorithm 2 Frank-Wolfe / Conditional gradient algorithm

Choose $x_0 \in A$, $k \leftarrow 0$,
while True **do**
  Choose $s_{k+1} \in \operatorname{argmin}_{s \in A} (f(x_k) + \langle \nabla f(x_k),\, s - x_k \rangle)$
  if $\langle \nabla f(x_k),\, s_{k+1} - x_k \rangle \geq 0$ **then**
    return x_k
    break
  end if
  Choose $\theta_k \in [0, 1]$ according to the update rule (see (2.8) or (2.9))
  $x_{k+1} \leftarrow \theta_k s_{k+1} + (1 - \theta_k) x_k$
  $k \leftarrow k + 1$
end while

The algorithm is summarized in Algorithm 2. Since it always constructs convex combinations of points in A, all the iterates are in A, that is, they are feasible.

Stopping criterion. The Frank-Wolfe algorithm has a natural stopping criterion in view of Chapter 1. Indeed, since s_{k+1} minimizes $s \mapsto \langle \nabla f(x_k),\, s - x_k \rangle$, the stopping criterion is equivalent to

$$\forall s \in A, \quad \langle \nabla f(x_k),\, s - x_k \rangle \geq 0.$$

In other words, $-\nabla f(x_k) \in \mathcal{N}_A(x_k)$ and the necessary condition for optimality (Theorem 1.5) is satisfied. If f is convex, it even certifies that $x_k \in \operatorname{argmin}_A f$.

Taking advantage of the extreme points of A. Since minimizing $s \mapsto (f(x_k) + \langle \nabla f(x_k),\, s - x_k \rangle)$ amounts to minimizing the linear form $\langle \nabla f(x_k),\, \cdot \rangle$ over A, it is possible to choose a minimizer s_{k+1} among the *extreme points* of A. An extreme point of A is a point $x \in A$ that cannot be written as a convex combination of two other points of A:

$$(x = \lambda y + (1 - \lambda) z \quad \text{with } \lambda \in]0, 1[\, , y \in A, z \in A) \Longrightarrow y = z = x. \tag{2.10}$$

The extreme points of a compact convex set A are particularly important. They contain all the necessary information to reconstruct A, since A is the closed convex hull of its extreme points (*i.e.* the smallest closed set which contains the extreme points of A).

For some specific convex sets, the collection of their extreme points is well-known and easy to work with. Below are a few instances.

Euclidean ball The collection of the extreme points of $B(0, 1)$ is the Euclidean unit sphere $S(0, 1)$. More generally, the collection of extreme points of the ℓ^q-unit ball $B_q(0,1)$ for $1 < q < +\infty$ is its boundary $\partial B_q(0,1)$. To minimize a linear form $s \mapsto \langle a,\, s\rangle$ on such sets, it suffices to recall that the Hölder inequality yields (for $1/q + 1/q' = 1$)

$$\left|\sum_{i=1}^{p} a_i s_i\right| \leq \left(\sum_{i=1}^{p} |a_i|^{q'}\right)^{1/q'} \left(\sum_{j=1}^{p} |s_j|^{q}\right)^{1/q}$$

with equality if and only if $s_i \propto (\mathrm{sign}(a_i))\, |a_i|^{q'-1}$. Taking into account that $s \in \partial B_q(0,1)$, one may check that the minimizer s is characterized by

$$\forall j \in [\![1\,;p]\!], \quad s_j = -\frac{(\mathrm{sign}(a_j))\, |a_j|^{q'-1}}{\left(\sum_j |a_j|^{q'}\right)^{1/q}}$$

ℓ^1- unit ball The extreme points of $B_1(0,1)$ are the vectors $\pm e_k$, for $k \in [\![1\,;p]\!]$, where the e_k's are the canonical basis vectors,

$$e_k = (0, 0, \cdots, 0, 1, 0, \cdots, 0).$$

To minimize $\langle a_i,\, \cdot\rangle$ over such vectors, it suffices to take $-\,\mathrm{sign}(a_{k_0})e_{k_0}$, where

$$k_0 = \mathrm{argmax}_{1\leq k\leq p}\, |a_k|$$

ℓ^∞-unit ball The extreme points of $B_\infty(0,1)$ are the sign vectors, $(\varepsilon_1, \cdots, \varepsilon_p)$ where $\varepsilon_i \in \{-1, +1\}$, for all $i \in [\![1\,;p]\!]$. To minimize $\langle a_i,\, \cdot\rangle$ over such vectors, it suffices to take

$$\forall j \in [\![1\,;p]\!], \quad s_j = -(\mathrm{sign}(a_j))$$

Many more convex sets may be considered. We refer to [Jag13] for more examples, including unit balls for matrix norms.

With a suitable initialization (*e.g.* x_0 is an extreme point), the conditional gradient algorithm produces by induction iterates x_k that are convex combinations of at most $k+1$ extreme points. That property is interesting when sparsity is an expected property of the solutions.

Convergence results. Just like the projected gradient descent, the conditional gradient algorithm may be applied to nonconvex problems (still, A is assumed to be convex), see [DR70] for a detailed analysis. Nevertheless, the case where f is convex yields more guarantees.

Theorem 2.2. *Assume that $f \in \mathscr{C}^{1,1}(X)$ is convex, and that $A \subseteq X$ is nonempty compact convex. Then, Algorithm 2 yields a minimizing sequence (i.e. $f(x_k) \to \min_A f$) provided the update rule is either the line search (2.8) or (2.9).*

Moreover, if $L \overset{\text{def.}}{=} \operatorname{Lip}(\nabla f)$, both update rules yield

$$f(x_k) - \min_A f \leq 2L \frac{(\operatorname{diam} A)^2}{k+2}. \tag{2.11}$$

where $\operatorname{diam} A = \max_{x,y \in A} |x - y|$ is the diameter of A.

Proof. We only deal with the case where ∇f is Lipschitz, and we refer to [DR70] for the proof the first item.

Possibly replacing f with $f - \min_A f$, we assume without loss of generality that $\min_A f = 0$. Since ∇f is L-Lipschitz, we note that for all $x, y \in A$, and all $\theta \in]0, 1[$,

$$\begin{aligned} f(x + \theta(y - x)) - f(x) - \theta \langle \nabla f(x), y - x \rangle &= \theta \int_0^1 \langle \nabla f(x + t\theta(y - x)) - \nabla f(x), y - x \rangle \, \mathrm{d}t \\ &\leq \theta^2 L \|y - x\|^2 \int_0^1 t \mathrm{d}t \\ &\leq L(\operatorname{diam} A)^2 \frac{\theta^2}{2}. \end{aligned}$$

We set $C \overset{\text{def.}}{=} L(\operatorname{diam} A)^2/2$ and $\theta_k = 2/(k+2)$ and we let $x^* \in \operatorname{argmin}_A f$. Whether the update rule is the linesearch or the prescribed value, we have

$$\begin{aligned} f(x_{k+1}) &\leq f(\theta_k s_{k+1} + (1 - \theta_k) x_k) \\ &\leq f(x_k) + \theta_k \langle \nabla f(x_k), s_{k+1} - x_k \rangle + C\theta_k^2 \end{aligned}$$

By optimality of s_{k+1} and convexity of f,

$$\langle \nabla f(x_k), s_{k+1} - x_k \rangle \leq \langle \nabla f(x_k), x^* - x_k \rangle \leq f(x^*) - f(x_k).$$

As a result,

$$(f(x_{k+1}) - f(x^*)) \leq (f(x_k) - f(x^*)) - \theta_k (f(x_k) - f(x^*)) + C\theta_k^2,$$

or equivalently

$$\ell_{k+1} \leq (1 - \theta_k)\ell_k + C\theta_k^2 \tag{2.12}$$

where we have set $\ell_k \overset{\text{def.}}{=} f(x_k) - f(x^*)$.

Next, we prove by induction that $\ell_k \leq 2C\theta_k$. For $k = 0$, this follows from (2.12) since $\theta_0 = 1$. Now, we assume the property holds for some $k \in \mathbb{N}$, and we observe:

$$\begin{aligned} \ell_{k+1} \leq 2C(1 - \theta_k)\theta_k - C(\theta_k)^2 &= C\left[2\theta_k - (\theta_k)^2\right] \\ &= \frac{4C}{k+2}\left[1 - \frac{1}{k+2}\right] \\ &= \frac{4C}{k+2}\left[\frac{k+1}{k+2}\right] \\ &\leq \frac{4C}{k+2}\left[\frac{k+2}{k+3}\right] = 2C\theta_{k+1} \end{aligned}$$

since $x \mapsto \frac{k+x}{k+1+x}$ is increasing on $\mathbb{R}_+$. As a result, the property holds for all $k \in \mathbb{N}$, that is,

$$f(x_k) - \min_A f \leq 2C\theta_k = L(\operatorname{diam} A)^2 \frac{2}{k+2}$$

which is precisely (2.11) □

Since A is assumed to be compact the sequence $(x_k)_{k\in\mathbb{N}}$ has cluster points, and the above theorem ensures that they are all minimizers of f over A.

Remark 2.3. *In fact, the Frank-Wolfe algorithm does not really require the existence of the gradient of f, but merely directional derivatives. Therefore it can be applied in in very general settings (e.g. topological vector spaces) on functions that are not necessarily (Fréchet) differentiable.*

Remark 2.4. *Moreover, several convergence results hold under weak assumptions on f. For instance, if f is merely $\mathscr{C}^1(X)$, the linesearch update rule yields a minimizing sequence (without any guaranteed rate).*

If f is not convex, a different argument may be used to prove

$$\max_{y\in A} \left(- \langle \nabla f(x_k), y - x_k\rangle\right) \to 0$$

(hence the necessary condition for optimality is satisfied at any cluster point of the sequence). See [DR70] for such variants.

2.2 Methods which approximately satisfy the constraints

2.2.1 Penalty methods

The idea is to replace the constrained optimization problem (1.1) with some unconstrained optimization problem by introducing a penalty function, *i.e.* a function that vanishes on the feasible set A and which is strictly positive outside A.

Assume that the problem has been reformulated with equality constraints only (for instance, replacing the inequality constraint $h_j(x) \leq 0$ with $\max(h_i(x), 0) = 0$), that is

$$\min_{x\in X} f(x) \quad \text{s.t.} \quad g_1(x) = \cdots = g_n(x) = 0. \tag{2.13}$$

A standard way is the quadratic penalty method, which consists, given $\rho > 0$, in solving the problem

$$\min_{x\in X} f(x) + \frac{\rho}{2}\sum_{i=1}^{n}(g_i(x))^2. \tag{2.14}$$

Under a few more assumptions, the methods produces solutions that converge, as $\rho \to +\infty$, to a solution of (2.13). One may use standard algorithms for unconstrained optimization to solve (2.14), hence find an approximate solution to (2.13). However, a limitation of the method is that, numerically, using large values of ρ yields ill-conditionning.

2.2.2 Augmented Lagrangian

Heuristics. We consider the problem (2.13) again, with f convex. If the KKT conditions hold, we should have $\nabla_x \mathcal{L}(x^*, \lambda^*) = 0$. We do not know λ^* in advance, but we know that (x^*, λ^*) should be a saddle point, *i.e.* solve (1.40). In order to improve the convergence to the feasible set, we add the quadratic penalty term, and we consider the *augmented Lagrangian*,

$$\mathcal{L}^a(x, \lambda, \rho) = f(x) + \sum_{i=1}^{n} \lambda_i g_i(x) + \frac{\rho}{2} \sum_{i=1}^{n} (g_i(x))^2, \tag{2.15}$$

and we iteratively perform a descent in x and an ascent in λ, while progressively increasing ρ. This is summarized in Algorithm 3.

In fact, it is not necessary to take $\rho \to +\infty$ to reach a solution. Indeed, notice that for $x \in A$, the quadratic term and its gradient vanish. As a result, a point (x^*, λ^*) is a saddle point to $\mathcal{L}$ if and only if it is a saddle point to $\mathcal{L}^a(\cdot, \cdot, \rho)$.

Algorithm 3 Augmented Lagrangian algorithm

```
Choose x_0 ∈ A, k ← 0,
while True do
    x_{k+1} ← argmin_x L^a(x, λ_k, ρ_k) (starting from x_k as an intial guess)
    λ_{k+1} ← λ_k + ρ_k g(x_{k+1})
    Choose ρ_{k+1} ≥ ρ_k
    k ← k + 1
end while
```

Appendix A

Cones and their polar

This chapter is devoted to cones and a notion of polarity.

A *cone*[1] $K \subseteq X$ is a set which stable by positive scalings:

$$\forall t > 0, \forall x \in K,\ tx \in K. \tag{A.1}$$

In short, K is a cone if and only if $tK \subseteq K$ for all $t > 0$.

One may check that a set K is a convex cone if and only

$$\forall t, t' > 0,\ \forall x, x' \in K,\ (tx + tx') \in K. \tag{A.2}$$

To each cone, it is possible to associate another cone, called its *polar cone*

$$K^\circ \stackrel{\text{def.}}{=} \{\, s \in X \mid \forall d \in K, \langle s,\, d\rangle \leq 0 \,\}\,. \tag{A.3}$$

Even though K is not convex nor a cone, K° is always a closed convex cone.

Observe that the polarity is a decreasing operation: if $A \subseteq B$ then $B^\circ \subseteq A^\circ$. Moreover, $K \cap K^\circ = \{0\}$.

A.1 The polar cone of finitely-generated convex cones (Farkas' lemma)

The following result is known as Farkas' Lemma. It is the key to describing the polars of finitely-generated convex cones.

Lemma A.1 (Farkas). *Let $v_1, \ldots, v_m \in \mathbb{R}^p$ and $x \in \mathbb{R}^p$. Then*

$$\exists \lambda_1 \geq 0, \ldots, \lambda_m \geq 0, \quad x = \sum_{j=1}^{m} \lambda_j v_j \tag{A.4}$$

if and only if

$$\forall u \in \mathbb{R}^p,\ [(\forall j \in \{1, \ldots, m\},\ \langle v_j,\, u\rangle \geq 0) \Rightarrow \langle x,\, u\rangle \geq 0]\,. \tag{A.5}$$

[1] Unless explicitely stated, all the cones we consider have apex 0. General cones K' result from the translation of cones with apex 0 by some vector x_0, *i.e.* $K' = x_0 + K$.

Proof. The "only if" part is immediate. We focus on the converse implication. We consider the set $K \stackrel{\text{def.}}{=} \left\{ \sum_{j=1}^{m} \lambda_j v_j \mid \lambda \in (\mathbb{R}_+)^p \right\}$. One may check that it is a convex cone. Moreover, that set is *closed*.

➢ For all $J \subseteq \{1, \ldots, m\}$, define $K_J \stackrel{\text{def.}}{=} \left\{ \sum_{j \in J} \lambda_j v_j \mid \forall j \in J,\ \lambda_j \geq 0 \right\}$. We note that if $\{v_j\}_{j \in J}$ is linearly independent, then K_J is closed. Indeed, if $(w_n)_{n \in \mathbb{N}}$ is a sequence in K_J which converges to some $w \in \mathbb{R}^p$, writing $w_n = \sum_{j \in J} \lambda_{j,n} v_j$, we have

$$\left(\lambda_{j,n}\right)_{j \in J} = (V_J)^{-1} w_n,$$

where the matrix V_J has rows $(v_j)_{j \in J}$. The convergence of w_n as $n \to +\infty$ implies the convergence of $\lambda_{j,n}$ to some λ_j, and $\lambda_{j,n} \geq 0$ implies $\lambda_j \geq 0$. Therefore $w = \sum_{j \in J} \lambda_j v_j \in K_J$ and K_J is closed.

Now, we claim that K is the (finite) union of cones K_J where $\{v_j\}_{j \in J}$ is linearly independent. Indeed, let $v \in K \setminus \{0\}$, and let

$$\mathcal{F} = \{ J \subseteq \{1, \ldots, m\} \mid v \in K_J \}.$$

The collection $\mathcal{F}$ is nonempty, therefore it has an element with minimal cardinality, say $J_0 \in \mathcal{F}$. By contradiction, if $\{v_j\}_{j \in J_0}$ is linearly dependent, there is some combination $\sum_{j \in J_0} \mu_j v_j = 0$ where the μ_j's are not all zero. We may even assume $\mu_j < 0$ for some index j. Then for all $t \geq 0$,

$$v = \sum_{j \in J_0} (\lambda_j + t\mu_j) v_j$$

and we may choose t such that $\lambda_j + t\mu_j \geq 0$ and $\lambda_{j_0} + t\mu_{j_0} = 0$ for some index j_0. Thus $v \in K_{J_0 \setminus \{j_0\}}$, which contradicts the minimality of J_0. Therefore $\{v_j\}_{j \in J_0}$ is linearly independent, which proves the claim.

As a result, K is closed as the finite union of closed sets K_J.

Assume by contradiction that (A.5) holds but (A.4) does not, *i.e.* $x \notin K$. It is thus possible to strictly separate[2] x and K: there exists $w \in \mathbb{R}^p$, such that

$$\inf_{v \in K} \langle w, v \rangle > \langle w, x \rangle . \tag{A.6}$$

Note that the left-hand side must be equal to 0, since for all $v \in K$ and $t > 0$, we have $tv \in K$, hence

$$t \langle w, v \rangle = \langle w, tv \rangle > \langle w, x \rangle .$$

Letting $t \to +\infty$ we see note that necessarily $\langle w, v \rangle \geq 0$. Moreover, equality holds for $v = 0 \in K$.

As a result, we have found a vector $w \in \mathbb{R}^p$ such that $\langle w, v_j \rangle \geq 0$ for $1 \leq j \leq m$ but $\langle w, x \rangle < 0$. That contradicts (A.5). □

Our main interest in Farkas'Lemma lies in the following consequence.

[2]That result is known as the Hahn-Banach separation theorem. In our setting (finite dimension) we may simply project x onto K and consider $w = p - x$.

Proposition A.2. *Let $a_1, \ldots, a_m \in \mathbb{R}^p$ and let K be the closed convex cone*

$$K \stackrel{\text{def.}}{=} \{\, d \in \mathbb{R}^p \mid \forall i \in \{1, \ldots, m\},\ \langle a_i,\, d\rangle \leq 0 \,\}\,. \tag{A.7}$$

Then, its polar cone $K^\circ \stackrel{\text{def.}}{=} \{\, s \in \mathbb{R}^p \mid \forall d \in K,\ \langle s,\, d\rangle \leq 0 \,\}$ is given by

$$K^\circ = \left\{ \sum_{i=1}^{m} \lambda_i a_i \mid \lambda \in (\mathbb{R}_+)^m \right\}. \tag{A.8}$$

Proof. It is clear that for any $\lambda \in (\mathbb{R}_+)^m$ and any $d \in K$,

$$\left\langle \sum_{j=1}^{m} \lambda_i a_i,\, d \right\rangle = \sum_{j=1}^{m} \lambda_i \langle a_j,\, d\rangle \leq 0,$$

hence the set in (A.8) is contained in K°.

Conversely, let $s \in K^\circ$, then for any $u \in \mathbb{R}^p$,

$$(\forall j \in [[1, m]]\,, \langle a_i,\, u\rangle \leq 0) \Rightarrow (u \in \mathcal{T}) \Rightarrow (\langle s,\, u\rangle \leq 0)$$

□

A.2 Polarity for closed convex cones

Another interesting property of polar cones is that it is an involution on the collection of closed convex sets.

Proposition A.3. *Let $K \subseteq X$ be a closed convex cone. Then $(K^\circ)^\circ = K$.*

We leave the proof to the reader, as an exercise. In particular, if $K \subseteq X$ is any set (not necessarily convex), then $((K^\circ)^\circ)^\circ = K^\circ$.

Appendix B

References

To prepare these lecture notes, we have relied on several references, in particular on the lecture notes by Clément Royer [1] which have been of considerable help. The derivation of optimality conditions using the tangent and normal cone is inspired from [RW98], but the theory developed therein is much more elaborate than ours. The book [BV04] is also a comprehensive and pedagogical reference, and is illustrated with modern problems from signal processing or statistics. It is available on the website of its authors[2].

For numerical aspects, the book by Nocedal and Wright [NW06] is a comprehensive reference, with many illustrations.

[1] https://www.lamsade.dauphine.fr/%7Ecroyer/ensdocs/LectureNotes-OML-ConOpt.pdf

[2] https://web.stanford.edu/~boyd/cvxbook/

Bibliography

[AB06] Charalambos D. Aliprantis and Kim C. Border. *Infinite dimensional analysis: a hitchhiker's guide*. Springer, Berlin ; New York, 3rd [rev. and enl.] ed edition, 2006. OCLC: ocm69983226.

[BV04] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[Con16] Laurent Condat. Fast Projection onto the Simplex and the l1 Ball. *Mathematical Programming, Series A*, 158(1):575–585, July 2016.

[DR70] Vladimir Fedorovich Demyanov and Aleksandr Moiseevich Rubinov. *Approximate methods in optimization problems*, volume 32. Elsevier Publishing Company, 1970.

[Jag13] Martin Jaggi. Revisiting Frank-Wolfe: Projection-free sparse convex optimization. In *International Conference on Machine Learning*, pages 427–435. PMLR, 2013.

[NW06] Jorge Nocedal and Stephen J. Wright. *Numerical Optimization*. Springer, New York, NY, USA, 2e edition, 2006.

[RW98] R. Tyrrell Rockafellar and Roger J.-B. Wets. *Variational Analysis*. Grundlehren der mathematischen Wissenschaften. Springer-Verlag, Berlin Heidelberg, 1998.