On subdifferential calculus *

Erik J. Balder

1 Introduction

The main purpose of these lectures is to familiarize the student with the basic ingredients of convex analysis, especially its subdifferential calculus. This is done while moving to a clearly discernible end-goal, the Karush-Kuhn-Tucker theorem, which is one of the main results of nonlinear programming. Of course, in the present lectures we have to limit ourselves most of the time to the Karush-Kuhn-Tucker theorem for *convex* nonlinear programming. While this is on the one hand restrictive, it is somewhat compensated for by extra structure that the Karush-Kuhn-Tucker theory gains in the presence of convexity.

The material is presented in the following way. It is assumed that several – but perhaps not all – students have already been exposed to some standard material on convex sets. *This material has been collected in the appendix*; it will be referred to during the lectures whenever the need arises. Sometimes further references will be given; as a rule these concern results that can be found in the textbooks [1] or [2]. The less standard part of the material, notably subdifferential calculus, is treated in the main part of the text.

2 Fundamental results on subdifferentials

The introduction of $+\infty$ and $-\infty$ as *extended* real numbers is an essential, simplifying ingredient of convex analysis, as we shall see below. The additional arithmetic is simple, but needs some care. Of course, one has $\alpha + (+\infty) = (+\infty) + \alpha = +\infty$ for every $\alpha \in (-\infty, +\infty]$; also, $\alpha - (+\infty) = -\infty$ for every $\alpha \in [-\infty, +\infty)$. Similar rules for adding/subtracting $-\infty$ can easily be gathered. However, neither $(+\infty) - (+\infty)$ nor $(+\infty) + (-\infty)$ is defined. This requires constant vigilance on the part of the reader: for instance, the identity $\alpha + \beta = \gamma + \beta$ can only be used to conclude that $\alpha = \gamma$ for $\alpha, \gamma \in [-\infty, +\infty]$ if $\beta \in \mathbb{R}$. For multiplication the additional rules apply: $\alpha \cdot (+\infty) = +\infty$ for every $\alpha \in (0, +\infty]$ and $\alpha \cdot (+\infty) = -\infty$ for every $\alpha \in [-\infty, 0)$. By definition, one also sets $0 \cdot (+\infty) = 0 \cdot (-\infty) = 0$. As for division, it is consistent with

*LNMB Ph.D. course "Convex Analysis for Optimization", September, 2010. These notes grew from earlier Ph.D. courses in Utrecht and Naples; the author is indebted to participants in those courses for helpful comments.

the above to have $\alpha/(+\infty) = \alpha/(-\infty) = 0$ for every $\alpha \in \mathbb{R}$, but of course fractions like $(+\infty)/(+\infty)$, etc. are undefined. Similar warnings hold: for instance, $\alpha/\beta = \gamma/\beta$ can only be used to conclude that $\alpha = \gamma$ for $\alpha, \gamma \in [-\infty, +\infty]$ if $\beta \in \mathbb{R}\backslash\{0\}$. Recall that the definition of a convex set can be found in Appendix A (Definition A.1). We now introduce a fundamental concept of this course.

Definition 2.1 A function $f : S \to (-\infty, +\infty]$, defined on a convex set $S \subset \mathbb{R}^n$, is said to be *convex on* S if for every $x_1, x_2 \in S$ and every $\lambda \in [0, 1]$

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

The same function is said to be *strictly convex* if for every $x_1, x_2 \in S$, $x_1 \neq x_2$, and for every $\lambda \in (0, 1)$

$$f(\lambda x_1 + (1-\lambda)x_2) < \lambda f(x_1) + (1-\lambda)f(x_2).$$

This definition does not take into consideration functions that can take the value $-\infty$, even though it could be expanded to include these.[1] By a "sign-mirror treatment" the above definition can be turned into the following: a function $f : S \to [-\infty, +\infty)$, defined on a convex set $S \subset \mathbb{R}^n$, is said to be *[strictly] concave on* S if the function $-f$ is [strictly] convex, as defined above. Because concave functions can always be turned into convex ones by changing the signs, this course will not consider concave functions explicitly.

Exercise 2.1 Prove the following:
a. Every linear[2] function $f(x) := a^t x + \alpha$, with $a \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$, is a convex function on $\mathbb{R}^n$ (and note that it is also concave).
b. The function $f(x) := \beta|x|^2$ is strictly convex on $\mathbb{R}^n$ if $\beta > 0$ (note that f is strictly concave if $\beta < 0$).
c. The function f defined on $\mathbb{R}_+$ by $f(x) := 1/x$ if $x > 0$ and by $f(0) := \gamma$ can only be made convex by choosing $\gamma = +\infty$.
d. The function f defined on $\mathbb{R}$ by $f(x) := 1/x$ if $x > 0$ and $f(x) := +\infty$ if $x \leq 0$ is convex.
e. The function $f(x) := -\sqrt{x}$ is convex on $\mathbb{R}_+$.
f. The function f defined on $\mathbb{R}$ by $f(x) := -\sqrt{x}$ if $x > 0$ and by defining $f(x) \in (-\infty, +\infty]$ for $x \leq 0$ can only be a convex function if one sets $f(x) := +\infty$ for every $x < 0$ and $f(0) := \gamma$ with $\gamma \in [0, +\infty]$.

Exercise 2.2 Let $S \subset \mathbb{R}^n$ be a convex set and let $f : S \to (-\infty, +\infty]$. Then f is said to be *quasiconvex* on S if for every $\alpha \in \mathbb{R}$ the so-called *lower level* set

$$S_\alpha := \{x \in S : f(x) \leq \alpha\}$$

[1] Functions that can take the value $-\infty$ are called *improper* in convex analysis. It can be shown that improper convex functions have a certain "pathological" structure, which is never encountered in realistic convex optimization problems.

[2] More accurately, such a function is called *affine*.

is convex.
a. Prove that if f is convex on S, then it is also quasiconvex on S.
b. Prove that f is quasiconvex on S if and only if for every $\alpha \in \mathbb{R}$ the set $\{x \in S : f(x) < \alpha\}$ is convex.
c. Let $g : D \to \mathbb{R}$ be a nondecreasing function on an interval $D \subset \mathbb{R}$ with $D \supset f(S)$ (note that this forces f to have values in $\mathbb{R}$). Prove that the composed function $h(x) := g(f(x))$ is also quasiconvex on S. *Hint:* Be careful: the function g is allowed to have discontinuities.

Exercise 2.3 Prove that the function $f(x) := -\exp(-x^2)$ is quasiconvex on $\mathbb{R}$, but not convex on $\mathbb{R}$. *Hint:* Prove monotonicity properties of f on respectively $\mathbb{R}_+$ and $\mathbb{R}_-$.

Exercise 2.4 For a function $f : S \to (-\infty, +\infty]$ one denotes by $\text{argmin}_{x \in S} f(x)$ the set (possibly empty) of all minimizers of f on S. That is to say

$$\text{argmin}_{x \in S} f(x) := \{z \in S : f(z) = \inf_{x \in S} f(x)\}.$$

a. Prove that the set $\text{argmin}_{x \in S} f(x)$ is convex if the function f is quasiconvex on S.
b. Prove that the set $\text{argmin}_{x \in S} f(x)$ contains at most one element if the function f is strictly convex on S.

Exercise 2.5 Prove the following *automatic* extension result for the domain of a convex function: if $f : S \to (-\infty, +\infty]$, defined on the convex set $S \subset \mathbb{R}^n$, is convex on S, then $\hat{f} : \mathbb{R}^n \to (-\infty, +\infty]$ is convex on $\mathbb{R}^n$, where $\hat{f}(x) := f(x)$ if $x \in S$ and $\hat{f}(x) := +\infty$ if $x \notin S$.

Note that this kind of extension has been practiced already in Exercise 2.1d, f above. As an important consequence of Exercise 2.5, we can often limit ourselves to the study of convex functions on the full space $\mathbb{R}^n$. This standardization can be very convenient. In the converse direction, we distinguish the subset of $\mathbb{R}^n$ on which a convex function $f : \mathbb{R}^n \to (-\infty, +\infty]$ "really matters" in the following way:

Definition 2.2 The *essential domain* of a function $f : \mathbb{R}^n \to (-\infty, +\infty]$ is the set dom f, given by

$$\text{dom } f := \{x \in \mathbb{R}^n : f(x) < +\infty\}.$$

It is clear that for every $x_0 \in \mathbb{R}^n$ the following equivalence holds: $x_0 \in \text{dom} f$ if and only if $f(x_0) \in \mathbb{R}$. Note also that if $f : \mathbb{R}^n \to (-\infty, +\infty]$ is a convex function (see Definition 2.1), then dom f is a convex set (see Definition A.1).

Next, we discuss some methods to create new convex functions from known convex functions. To begin with, it is easy to see that if $f_1, \ldots, f_m : \mathbb{R}^n \to (-\infty, +\infty]$ are convex functions, then so are their pointwise sum $f(x) := \sum_{i=1}^m f_i(x)$ and pointwise maximum $\max_{1 \leq i \leq m} f_i(x)$. More generally, if $\alpha_1, \ldots, \alpha_m$ are in $\mathbb{R}_+$, then the pointwise sum $f(x) := \sum_{i=1}^m \alpha_i f_i(x)$ is also a convex function (on $\mathbb{R}^n$). Another, more powerful device to create new convex functions out of known convex functions is composition; this is the subject of the following two exercises:

Exercise 2.6 Let $f : S \to \mathbb{R}$ be a convex function on the convex set S and let $g : D \to \mathbb{R}$ be a convex function on the convex set $D \subset \mathbb{R}$, with $D \supset f(S)$. Suppose in addition that the function g is also *nondecreasing* on D (i.e., $\xi_1 \leq \xi_2$ implies $g(\xi_1) \leq g(\xi_2)$ for all $\xi_1, \xi_2 \in D$). Demonstrate that the composed function $h(x) := g(f(x))$ is also convex on S. Prove also that if g is merely nondecreasing (but perhaps not convex), then h is a quasiconvex function on S.

Exercise 2.7 a. Let $f : \mathbb{R}^n \to [0, +\infty]$ be convex on $\mathbb{R}^n$. Prove that f^2 is also a convex function on $\mathbb{R}^n$.
b. Prove that the function $f(x) := 1 - \sqrt{1 - x^2}$ is convex on $[-1, +1]$.
c. Prove that the function $f(x) := \exp(x^2)$ is convex on $\mathbb{R}$.

Below, in Proposition 2.7, the reader will find another important tool to determine whether a given function is convex.

Definition 2.3 Given $S \subset \mathbb{R}^n$, consider the following function $\chi_S : \mathbb{R}^n \to \{0, +\infty\}$

$$\chi_S(x) := \begin{cases} 0 & \text{if } x \in S \\ +\infty & \text{if } x \notin S. \end{cases}$$

This function is called the *indicator function* of the set S.

This definition turns sets into closely related functions. It is easy to see that $S \subset \mathbb{R}^n$ is a convex set if and only if its indicator function χ_S is a convex function. In a converse direction, convex functions can also be turned into closely related convex sets:

Definition 2.4 The *epigraph* of a function $f : \mathbb{R}^n \to (-\infty, +\infty]$ is the subset epi f of $\mathbb{R}^n \times \mathbb{R}$ defined by

$$\text{epi } f := \{(x, y) \in \mathbb{R}^n \times \mathbb{R} : f(x) \leq y\}$$

Exercise 2.8 Let $f : \mathbb{R}^n \to (-\infty, +\infty]$. Prove the following: the function f is convex if and only if its epigraph epi f is a convex subset of $\mathbb{R}^n \times \mathbb{R}$.

Definition 2.5 a. A *subgradient* of a function $f : \mathbb{R}^n \to (-\infty, +\infty]$, $f \not\equiv +\infty$, at the point $x_0 \in \mathbb{R}^n$ is a vector $\xi \in \mathbb{R}^n$ such that

$$f(x) \geq f(x_0) + \xi^t(x - x_0) \text{ for all } x \in \mathbb{R}^n.$$

The set $\partial f(x_0)$ (possibly empty) of all such subgradients is called the *subdifferential* of f at the point x_0. Observe that this definition is only nontrivial if $x_0 \in \text{dom} f$: if $x_0 \in \mathbb{R}^n \backslash \text{dom } f$, then $f(x_0) = +\infty$, so $\partial f(x_0) = \emptyset$.

From now on, the trivial function $f \equiv +\infty$ is excluded from our considerations. For convex functions, subgradients form a generalization of the classical notion of gradient:

Proposition 2.6 *Let $f : \mathbb{R}^n \to (-\infty, +\infty]$ be a convex function that is differentiable at the point $x_0 \in \text{int dom } f$. Then $\partial f(x_0) = \{\nabla f(x_0)\}$.*

The proof of this proposition will be given later, because its proof uses Theorem 2.15. Observe that below this proposition applies to some points in Example 2.9(a) and also to Example 2.9(b).

Exercise 2.9 a. Consider the function $f : \mathbb{R} \to (-\infty, +\infty]$, defined by

$$f(x) := \begin{cases} 0 & \text{if } x \in [-1, +1] \\ |x| - 1 & \text{if } x \in [-2, -1) \cup (1, 2] \\ +\infty & \text{if } x \in (-\infty, -2) \cup (2, +\infty). \end{cases}$$

Demonstrate that

$$\partial f(x) = \begin{cases} \{0\} & \text{if } x \in (-1, 1) \\ [-1, 0] & \text{if } x = -1 \\ [0, 1] & \text{if } x = 1 \\ \{-1\} & \text{if } x \in (-2, -1) \\ \{1\} & \text{if } x \in (1, 2) \\ (-\infty, -1] & \text{if } x = -2 \\ [1, +\infty) & \text{if } x = 2 \\ \text{undefined} & \text{if } x \in (-\infty, -2) \cup (2, +\infty). \end{cases}$$

b. Let $f : \mathbb{R}^n \to (-\infty, +\infty]$ be given by $f(x) := 1 - \sqrt{1 - x^2}$ if $x \in [-1, +1]$ and by $f(x) := +\infty$ if $x < -1$ or $x > 1$. Demonstrate that

$$\partial f(x) = \begin{cases} \{x/\sqrt{1 - x^2}\} & \text{if } x \in (-1, +1) \\ \emptyset & \text{if } x \leq -1 \text{ or } x \geq 1. \end{cases}$$

Proposition 2.6 can be used to provide a very useful characterization of convexity for *differentiable* functions on $\mathbb{R}$:

Proposition 2.7 (i) *Let $f : S \to \mathbb{R}$ be a differentiable function on the open, convex set $S \subset \mathbb{R}^n$. Then f is convex on S if and only if the following monotonicity property holds*

$$(\nabla f(x_1) - \nabla f(x_2))^t (x_1 - x_2) \geq 0 \textit{ for every } x_1, x_2 \in S.$$

(i') *Let $f : S \to \mathbb{R}$ be a differentiable function on the open, convex set $S \subset \mathbb{R}^n$. Then f is strictly convex on S if and only if the following monotonicity property holds*

$$(\nabla f(x_1) - \nabla f(x_2))^t (x_1 - x_2) > 0 \textit{ for every } x_1, x_2 \in S,\ x_1 \neq x_2.$$

(ii) *Let $f : S \to \mathbb{R}$ be a second order continuously differentiable function on the open, convex set $S \subset \mathbb{R}^n$. Then f is convex on S if and only if its Hessian matrix*

$$H_f(x) := \left(\frac{\partial^2 f(x)}{\partial x_i \partial x_j} \right)_{i,j}$$

is positive semidefinite at every point x of S.

(ii') *Let $f : S \to \mathbb{R}$ be a second order continuously differentiable function on the open, convex set $S \subset \mathbb{R}^n$. Then f is strictly convex on S if its Hessian matrix*

$$H_f(x) := \left(\frac{\partial^2 f(x)}{\partial x_i \partial x_j}\right)_{i,j}$$

is positive definite at every point x of S.

Recall here that an $n \times n$ matrix M is *positive semidefinite* if $d^t M d \geq 0$ for all $d \in \mathbb{R}^n$. It is *positive definite* if $d^t M d > 0$ for all $d \in \mathbb{R}^n$.

PROOF. (i) If f is convex on S, then Proposition 2.6, together with the definition of subdifferential, implies

$$f(x_2) \geq f(x_1) + \nabla f(x_1)^t (x_2 - x_1) \text{ and } f(x_1) \geq f(x_2) + \nabla f(x_2)^t (x_1 - x_2).$$

This immediately gives the desired monotonicity.

Conversely, given monotonicity, fix x, x' in S and let $\phi(t) := f(tx' + (1-t)x'')$, $t \in [0,1]$. By the mean value theorem there exists $\theta \in (0,1)$ such that $\phi(1) - \phi(0) = \phi'(\theta)$, i.e., $f(x') - f(x'') = \nabla f(\tilde{x})^t (x' - x'')$, where $\tilde{x} := \theta x' + (1-\theta)x''$. Monotonicity implies $(\nabla f(\tilde{x}) - \nabla f(x''))^t (\tilde{x} - x'') \geq 0$, i.e., $\theta(\nabla f(\tilde{x}) - \nabla f(x''))^t (x' - x'') \geq 0$. Hence, $\nabla f(\tilde{x})^t (x' - x'') \geq \nabla f(x''))^t (x' - x'')$. Thus, it follows that

$$f(x') \geq f(x'') + \nabla f(x'')^t (x' - x'') \text{ for every pair } x', x'' \in S.$$

To prove that this property implies the convexity of f, let $x_1, x_2 \in S$, let $\lambda \in [0,1]$ and set $x_3 := \lambda x_1 + (1-\lambda) x_2$. By applying the previous property to $x'' := x_3$ and successively to $x' = x_1$ and $x' = x_2$, we obtain

$$f(x_1) \geq f(x_3) + \nabla f(x_3)^t (x_1 - x_3) \text{ and } f(x_2) \geq f(x_3) + \nabla f(x_3)^t (x_2 - x_3).$$

Multiplying the left hand sides by λ and $1 - \lambda$ respectively, this easily leads to $\lambda f(x_1) + (1-\lambda) f(x_2) \geq f(x_3)$.

(ii) The underlying idea is that monotonicity (as in part (i)) of the first order derivative of f can, in turn, be characterized by "nonnegativity" (i.e., positive semidefiniteness) of the second order derivative. We refer to [2] for the details. Parts (i') and (ii') go analogously (exercise). QED

Specialized to $n = 1$, Proposition 2.7 is as follows:

Corollary 2.8 (i) *Let $f : S \to \mathbb{R}$ be a differentiable function on the open, convex set $S \subset \mathbb{R}$. Then f is convex [strictly convex] on S if and only if its derivative is nondecreasing [increasing].*

(ii) *Let $f : S \to \mathbb{R}$ be a second order continuously differentiable function on the open, convex set $S \subset \mathbb{R}$. Then f is convex [strictly convex] on S if and only if [if] its second derivative is nonnegative [positive].*

Exercise 2.10 Find the smallest $\alpha \in \mathbb{R}$ for which $f(x) := x\exp(-x)$ is convex on the set $[\alpha, +\infty)$.

Exercise 2.11 Consider for $\alpha, \beta > 0$ the function $f(x_1, x_2) := -x_1^\alpha x_2^\beta$ on $\mathbb{R}^2_+$. Prove the following:
a. If $\alpha + \beta \leq 1$, then f is convex on $\mathbb{R}^2_+$.
b. If $\alpha + \beta > 1$, then f is not convex on $\mathbb{R}^2_+$, but it is still quasiconvex. *Hint:* use Exercise 2.6.

Theorem 2.9 (Moreau-Rockafellar) *Let $f, g : \mathbb{R}^n \to (-\infty, +\infty]$ be convex functions. Then for every $x_0 \in \mathbb{R}^n$*

$$\partial f(x_0) + \partial g(x_0) \subset \partial(f+g)(x_0).$$

Moreover, suppose that $\mathrm{int\,dom}\, f \cap \mathrm{dom}\, g \neq \emptyset$. *Then for every $x_0 \in \mathbb{R}^n$ also*

$$\partial(f+g)(x_0) \subset \partial f(x_0) + \partial g(x_0).$$

Proof. The proof of the first part is elementary: Let $\xi_1 \in \partial f(x_0)$ and $\xi_2 \in \partial g(x_0)$. Then for all $x \in \mathbb{R}^n$

$$f(x) \geq f(x_0) + \xi_1^t(x - x_0), \; g(x) \geq g(x_0) + \xi_2^t(x - x_0),$$

so addition gives $f(x) + g(x) \geq f(x_0) + g(x_0) + (\xi_1 + \xi_2)^t(x - x_0)$. Hence $\xi_1 + \xi_2 \in \partial(f+g)(x_0)$.

To prove the second part, let $\xi \in \partial(f+g)(x_0)$. First, observe that $f(x_0) = +\infty$ implies $(f+g)(x_0) = +\infty$, whence $f + g \equiv +\infty$, which is impossible by $\xi \in \partial(f + g)(x_0)$. Likewise, $g(x_0) = +\infty$ is impossible. Hence, from now on we know that both $f(x_0)$ and $g(x_0)$ belong to $\mathbb{R}$. We form the following two sets in $\mathbb{R}^{n+1}$.

$$\Lambda_f := \{(x - x_0, y) \in \mathbb{R}^n \times \mathbb{R} : y > f(x) - f(x_0) - \xi^t(x - x_0)\}$$

$$\Lambda_g := \{(x - x_0, y) : -y \geq g(x) - g(x_0)\}.$$

Observe that both sets are nonempty and convex (see Exercise 2.8), and that $\Lambda_f \cap \Lambda_g = \emptyset$ (the latter follows from $\xi \in \partial(f+g)(x_0)$). Hence, by the set-set-separation Theorem A.4, there exists $(\xi_0, \mu) \in \mathbb{R}^{n+1}$ and $\alpha \in \mathbb{R}$, $(\xi_0, \mu) \neq (0, 0)$, such that

$$\xi_0^t(x - x_0) + \mu y \leq \alpha \text{ for all } (x, y) \text{ with } y > f(x) - f(x_0) - \xi^t(x - x_0),$$

$$\xi_0^t(x - x_0) + \mu y \geq \alpha \text{ for all } (x, y) \text{ with } -y \geq g(x) - g(x_0).$$

By $(0, 0) \in \Lambda_g$ we get $\alpha \leq 0$. But also $(0, \epsilon) \in \Lambda_f$ for every $\epsilon > 0$, and this gives $\mu\epsilon \leq \alpha$, so $\mu \leq 0$ (take $\epsilon = 1$). In the limit, for $\epsilon \to 0$, we find $\alpha \geq 0$. Hence $\alpha = 0$ and $\mu \leq 0$. We now claim that $\mu = 0$ is impossible. Indeed, if one had $\mu = 0$, then the first of the above two inequalities would give

$$\xi_0^t(x - x_0) \leq 0 \text{ for all } (x, y) \text{ with } y > f(x) - f(x_0) - \xi^t(x - x_0),$$

which is equivalent to

$$\xi_0^t(x - x_0) \le 0 \text{ for all } x \in \text{dom } f$$

(simply note that when $f(x) < +\infty$ one can always achieve $y > f(x) - f(x_0) - \xi^t(x - x_0)$ by choosing y sufficiently large). Likewise, the second inequality would give

$$\xi_0^t(x - x_0) \ge 0 \text{ for all } x \in \text{dom } g.$$

In particular, for $\tilde{x}$ as above this would imply $\xi_0^t(\tilde{x} - x_0) = 0$. But since $\tilde{x}$ lies in the interior of dom f (so for some $\delta > 0$ the ball $N_\delta(\tilde{x})$ belongs to dom f), the preceding would imply

$$\xi_0^t u = \xi_0^t(\tilde{x} + u - x_0) \le 0 \text{ for all } u \in N_\delta(0).$$

Clearly, this would give $\xi_0 = 0$ (take $u := \delta\xi_0/2$), which would be in contradiction to $(\xi_0, \mu) \neq (0, 0)$. Hence, we conclude $\mu < 0$. Dividing the separation inequalities by $-\mu$ and setting $\bar{\xi}_0 := -\xi_0/\mu$, this results in

$$\bar{\xi}_0^t(x - x_0) \le y \text{ for all } (x, y) \text{ with } y > f(x) - f(x_0) - \xi^t(x - x_0),$$

$$\bar{\xi}_0^t(x - x_0) \ge y \text{ for all } (x, y) \text{ with } -y \ge g(x) - g(x_0).$$

The last inequality gives $-\bar{\xi}_0 \in \partial g(x_0)$ (set $y := g(x_0) - g(x)$) and the one but last inequality gives $\xi + \bar{\xi}_0 \in \partial f(x_0)$ (take $y := f(x) - f(x_0) - \xi^t(x - x_0) + \epsilon$ and let $\epsilon \downarrow 0$). Since $\xi = (\xi + \bar{\xi}_0) - \bar{\xi}_0$, this finishes the proof. QED

Exercise 2.12 Show by means of an example that the condition int dom $f \cap$dom $g \neq \emptyset$ in Theorem 2.9 cannot be omitted.

Exercise 2.13 Find and prove an version of the Moreau-Rockafellar theorem that applies to the subdifferentials of a finite sum of convex functions.

As a precursor to the Karush-Kuhn-Tucker theorem, we have now the following application of the Moreau-Rockafellar theorem.

Theorem 2.10 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function and let $S \subset \mathbb{R}^n$ be a nonempty convex set. Consider the optimization problem*

$$(P) \ \inf_{x \in S} f(x).$$

Then $\bar{x} \in S$ is an optimal solution of (P) if and only if there exists a subgradient $\bar{\xi} \in \partial f(\bar{x})$ such that

$$\bar{\xi}^t(x - \bar{x}) \ge 0 \text{ for all } x \in S. \tag{1}$$

PROOF. Recall from Definition 2.3 that χ_S is the indicator function of S. Now let $\bar{x} \in S$ be arbitrary. Then the following is trivial: $\bar{x}$ is an optimal solution of (P) if and only if

$$0 \in \partial(f + \chi_S)(\bar{x}).$$

By the Moreau-Rockafellar Theorem 2.9, we have

$$\partial(f + \chi_S)(\bar{x}) = \partial f(\bar{x}) + \partial \chi_S(\bar{x}).$$

To see that its conditions hold, observe that dom $f = \mathbb{R}^n$ and dom $\chi_S = S$. So it follows that $\bar{x}$ is an optimal solution of (P) if and only if $0 \in \partial f(\bar{x}) + \partial \chi_S(\bar{x})$. By the definition of the sum of two sets this means that $\bar{x}$ is an optimal solution of (P) if and only if $0 = \bar{\xi} + \bar{\xi}'$ for some $\bar{\xi} \in \partial f(\bar{x})$ and $\bar{\xi}' \in \partial \chi_S(\bar{x})$. Of course, the former means $\bar{\xi}' = -\bar{\xi}$, so $-\bar{\xi} \in \partial \chi_S(\bar{x})$, which is equivalent to

$$\chi_S(x) \geq \chi_S(\bar{x}) + (-\bar{\xi})^t(x - \bar{x}) \text{ for all } x \in \mathbb{R}^n,$$

i.e., to (1). QED

Remark 2.11 *As the application of the Moreau-Rockafellar theorem in the above proof shows, the sufficiency part of Theorem 2.10 remains valid for a convex function* $f : \mathbb{R}^n \to (-\infty, +\infty]$, *i.e., a function that can attain the value* $+\infty$. *In that same situation the necessity also remains valid, provided that we suppose either* int dom $f \cap S \neq \emptyset$ *or* dom $f \cap$int $S \neq \emptyset$. *In particular, this remark applies to automatic extensions of the type introduced in Exercise 2.5.*

Exercise 2.14 Show by means of an example that, without the additional condition suggested in Remark 2.11, it is essential in Theorem 2.10 to have a function f with values in $\mathbb{R}$. [*Hint:* In boundary points of dom f the subdifferential of f can be empty, as shown in certain examples above.]

Exercise 2.15 Let $S \subset \mathbb{R}^2$ be given by the following system of inequalities: $\xi_1 \geq 0, \xi_2 \geq 0, -\xi_1 + \xi_2 \leq 2, 2\xi_1 + 3\xi_2 \leq 11$. Let $f(\xi_1, \xi_2) := \xi_1^2 + \xi_2^2 - 8\xi_1 - 20\xi_2 + 89$.
a. Prove that S is a convex set and that $f : S \to \mathbb{R}$ is convex.
b. Use Theorem 2.10 to show that $\xi_1 = 1, \xi_2 = 3$ is an optimal solution for minimizing f over S.
c. Prove that, actually, f is *strictly* convex, i.e., prove that $f(\lambda x_1 + (1-\lambda)x_2) < \lambda f(x_1) + (1-\lambda)f(x_2)$ for every $x_1, x_2 \in S$, $x_1 \neq x_2$, and every $\lambda \in (0, 1)$.
d. Use part c to prove that $(1, 3)$ in part b is the only optimal solution.

Example 2.12 Let the convex set $S \subset \mathbb{R}^2$ be given by the following four inequalities: $\xi_1 \geq 0, \xi_2 \geq 0, \xi_2 \geq \xi_1^2$ and $\xi_2 \leq 4$. Let $f(\xi_1, \xi_2) := (\xi_1 - 10)^2 + (\xi_2 - 5)^2$; this measures the squared distance from (ξ_1, ξ_2) to the point $(10, 5)$. From a picture of S it would seem that $\bar{x} = (2, 4)$ is the point in S that is closest to $(10, 5)$. To check that $\bar{x} = (2, 4)$ is indeed the optimal solution of $\min_{x \in S} f(x)$, we apply Theorem 2.10: it is enough to verify that $\nabla f(2, 4)^t(\xi_1 - 2, \xi_2 - 4) \geq 0$ for every $(\xi_1, \xi_2) \in S$. Now $\nabla f(2, 4) = (-16, -2)$, so it must be verified that $-16(\xi_1 - 2) - 2(\xi_2 - 4) \geq 0$, i.e., that $8\xi_1 + \xi_2 \leq 20$ for every $(\xi_1, \xi_2) \in S$. This holds, because $(\xi_1, \xi_2) \in S$ implies directly $\xi_1 \leq 2$ and $\xi_2 \leq 4$. Since the function f is strictly convex, we conclude, moreover, from Exercise 2.4 that $(2, 4)$ is the unique point in S that is closest to $(10, 5)$.

Definition 2.13 The *directional derivative* of a convex function $f : \mathbb{R}^n \to (-\infty, +\infty]$ at the point $x_0 \in \mathrm{dom} f$ in the direction $d \in \mathbb{R}^n$ is defined as

$$f'(x_0; d) := \lim_{\lambda \downarrow 0} \frac{f(x_0 + \lambda d) - f(x_0)}{\lambda}.$$

The above limit is a well-defined number in $[-\infty, +\infty]$. This follows from the following proposition (why?), which shows that the difference quotients of a convex functions possess a monotonicity property:

Proposition 2.14 *Let $f : \mathbb{R}^n \to (-\infty, +\infty]$ be a convex function and let x_0 be a point in* $\mathrm{dom} f$. *Then for every direction $d \in \mathbb{R}^n$ and every $\lambda_1, \lambda_2 \in \mathbb{R}$ such that $\lambda_2 > \lambda_1 > 0$ we have*

$$\frac{f(x_0 + \lambda_1 d) - f(x_0)}{\lambda_1} \leq \frac{f(x_0 + \lambda_2 d) - f(x_0)}{\lambda_2}$$

PROOF. Note that

$$x_0 + \lambda_1 d = \frac{\lambda_1}{\lambda_2}(x_0 + \lambda_2 d) + (1 - \frac{\lambda_1}{\lambda_2})x_0.$$

So by convexity of f

$$f(x_0 + \lambda_1 d) \leq \frac{\lambda_1}{\lambda_2} f(x_0 + \lambda_2 d) + (1 - \frac{\lambda_1}{\lambda_2}) f(x_0).$$

Simple algebra shows that this is equivalent to the desired inequality. QED

In Appendix B the Fenchel conjugation of convex functions is studied; this tool plays a major role in the proof of the next theorem:

Theorem 2.15 *Let $f : \mathbb{R}^n \to (-\infty, +\infty]$ be a convex function and let x_0 be a point in* $\mathrm{int\, dom}\, f$. *Then*

$$f'(x_0; d) = \sup_{\xi \in \partial f(x_0)} \xi^t d \text{ for every } d \in \mathbb{R}^n.$$

Exercise 2.16 You are asked to verify the identity of Theorem 2.15 explicitly in each of the following cases (so in each case you are asked to determine both the left and right hand sides independently, and then to show that the identity holds).
a. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function which is differentiable at the point $x_0 \in \mathrm{int\, dom}\, f$.
b. Let $x_0 := 0$ and let $f : \mathbb{R}^n \to \mathbb{R}$ be the convex function given by $f(x) := |x| := (\sum_i x_i^2)^{1/2}$ (Euclidean norm). *Hint:* here you must show, among other things, that $\partial f(0) = \{x \in \mathbb{R}^n : |x| \leq 1\}$.
c. Let $x_0 := 1$ and let $f : \mathbb{R} \to \mathbb{R}$ be the convex function $f(x) := \max(1, x)$.

The proof of Theorem 2.15 uses the following lemma:

Lemma 2.16 *Let $f : \mathbb{R}^n \to (-\infty, +\infty]$ be a convex function. Then f is continuous at any point $x_0 \in \text{int dom } f$; moreover, then $\partial f(x_0)$ is nonempty and compact.*

PROOF. Continuity: Consider $g(x) := f(x_0 + x) - f(x_0)$. Then g is convex and $g(0) = 0$. Let $e_1, \ldots, e_n$ be the unit vectors in $\mathbb{R}^n$. Denote the set $\{e_1, \ldots, e_n, -e_1, \ldots, -e_n\}$ by $\{y_1, \ldots, y_{2n}\}$. Let $\alpha \in (0, 1]$ be so small that $x_0 + \alpha y_i \in \text{dom } f$ for all i. Now for every $x \in \mathbb{R}^n$ such that $|x_i| \leq \alpha/n$ one has

$$x = \sum_{i, x_i > 0} \frac{x_i}{\alpha} \alpha e_i + \sum_{i, x_i < 0} \frac{-x_i}{\alpha} \alpha(-e_i) + (1 - \sum_i \frac{|x_i|}{\alpha})0$$

so that

$$g(x) \leq \sum_{i, x_i > 0}^{n} \frac{|x_i|}{\alpha} g(\alpha e_i) + \sum_{i, x_i < 0}^{n} \frac{|x_i|}{\alpha} g(-\alpha e_i) \leq \beta \sum_i |x_i|,$$

where $\beta := \alpha^{-1} \max_{1 \leq i \leq 2n} (f(x_0 + \alpha y_i) - f(x_0)) < +\infty$. Also, for the same x one has $0 = \frac{1}{2}x + \frac{1}{2}(-x)$, so

$$0 \leq \frac{1}{2} g(x) + \frac{1}{2} g(-x),$$

Hence $g(x) \geq -g(-x) \geq -\beta \sum_i |x_i|$ holds as well. We conclude therefore that g is continuous (and even Lipschitz-continuous) at 0, i.e., f is continuous at the original point x_0.

Nonemptiness: Let $g := \chi_{x_0}$. Then by the Moreau-Rockafellar theorem $\partial(f + g)(x_0) = \partial f(x_0) + \partial g(x_0)$. But both $\partial(f + g)(x_0)$ and $\partial g(x_0)$ are equal to $\mathbb{R}^n$ in this case, so $\partial f(x_0)$ cannot be empty (because of $\emptyset + \mathbb{R}^n = \emptyset$).

Compactness: Exercise 2.17. QED

Exercise 2.17 Prove the compactness part of Lemma 2.16. *Hint:* Use the continuity part and mimic certain components of the proof of that part.

PROOF OF THEOREM 2.15. By Proposition 2.14

$$q(d) := f'(x_0; d) := \lim_{\lambda \downarrow 0} \frac{f(x_0 + \lambda d) - f(x_0)}{\lambda} = \inf_{\lambda > 0} \frac{f(x_0 + \lambda d) - f(x_0)}{\lambda}.$$

Since the pointwise limit of a sequence of convex functions is convex, it follows that $q : \mathbb{R}^n \to \mathbb{R}$ is convex (by the infimum expression for $q(d)$ the fact that $x_0 \in \text{int dom } f$ implies automatically $q(d) < +\infty$ for every d; also, $q(d) > -\infty$ for every d, because of the nonemptiness part of Lemma 2.16). Hence, q is continuous at every point $d \in \mathbb{R}^n$ (apply the continuity part of Lemma 2.16). So by the Fenchel-Moreau theorem (Theorem B.5 in the Appendix) we have for every d

$$q(d) = q^{**}(d) := \sup_{\xi \in \mathbb{R}^n} [d^t \xi - q^*(\xi)].$$

Let us calculate q^*. For any $\xi \in \mathbb{R}^n$ we have

$$q^*(\xi) := \sup_{d \in \mathbb{R}^n} [\xi^t d - q(d)] = \sup_{d, \lambda > 0} [\xi^t d - \frac{f(x_0 + \lambda d) - f(x_0)}{\lambda}] = \sup_{\lambda > 0} \sup_d [\xi^t d - \frac{f(x_0 + \lambda d) - f(x_0)}{\lambda}]$$

by the above infimum expression for $q(d)$. Fix $\lambda > 0$; then $z := x_0 + \lambda d$ runs through all of $\mathbb{R}^n$ as d runs through $\mathbb{R}^n$. Hence

$$\sup_d [\xi^t d - \frac{f(x_0 + \lambda d) - f(x_0)}{\lambda}] = \frac{f(x_0) - \xi^t x_0 + \sup_z [\xi^t z - f(z)]}{\lambda}.$$

Clearly, this gives

$$q^*(\xi) = \sup_{\lambda > 0} \frac{f(x_0) - \xi^t x_0 + f^*(\xi)}{\lambda} = \begin{cases} 0 & \text{if } \xi \in \partial f(x_0) \\ +\infty & \text{otherwise} \end{cases}$$

where we use Proposition B.4(v). Observe that in terms of the indicator function of the subdifferential this can be rewritten as $q^* = \chi_{\partial f(x_0)}$. Now that q^* has been calculated, we conclude from the above that for every $d \in \mathbb{R}^n$

$$f'(x_0; d) = q(d) = q^{**}(d) = \chi^*_{\partial f(x_0)}(d) = \sup_{\xi \in \partial f(x_0)} \xi^t d,$$

which proves the result. QED

Proof of Proposition 2.6. By Theorem 2.15 we get

$$\nabla f(x_0)^t d = \sup_{\xi \in \partial f(x_0)} \xi^t d.$$

The remainder of the proof is left as an exercise.

Theorem 2.17 (Dubovitskii-Milyutin) *Let $f_1, \cdots, f_m : \mathbb{R}^n \to (-\infty, +\infty]$ be convex functions and let x_0 be a point in $\cap_{i=1}^m \text{int dom } f_i$. Let $f : \mathbb{R}^n \to (-\infty, +\infty]$ be given by*

$$f(x) := \max_{1 \leq i \leq m} f_i(x)$$

and let $I(x_0)$ be the (nonempty) set of all $i \in \{1, \cdots, m\}$ for which $f_i(x_0) = f(x_0)$. Then

$$\partial f(x_0) = \text{co } \cup_{i \in I(x_0)} \partial f_i(x_0).$$

Proof. For our convenience we write $I := I(x_0)$. To begin with, observe that $\xi \in \partial f_i(x_0)$ easily implies $\xi \in \partial f(x_0)$ for each $i \in I$. Since $\partial f(x_0)$ is evidently convex, the inclusion "$\supset$" follows with ease. To prove the opposite inclusion, let ξ_0 be arbitrary in $\partial f(x_0)$. If ξ_0 were not to belong to the compact set co $\cup_{i \in I} \partial f_i(x_0)$, then we could separate strictly (note that each set $\partial f_i(x_0)$ is both closed and compact (exercise)): by Theorem A.2 there would exist $d \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$ such that

$$\xi_0^t d > \alpha \geq \max_{i \in I} \sup_{\xi \in \partial f_i(x_0)} \xi^t d = \max_{i \in I} f_i'(x_0; d),$$

where the final identity follows from Theorem 2.15. But now observe that

$$f'(x_0; d) := \lim_{\lambda \downarrow 0} \max_{i \in I} \frac{f_i(x_0 + \lambda d) - f_i(x_0)}{\lambda} = \max_{i \in I} \lim_{\lambda \downarrow 0} \frac{f_i(x_0 + \lambda d) - f_i(x_0)}{\lambda} = \max_{i \in I} f_i'(x_0; d),$$

so the above gives $\xi_0^t d > f'(x_0; d)$. On the other hand, by $\xi_0 \in \partial f(x_0)$ it follows that $f(x_0 + \lambda d) \geq f(x_0) + \lambda \xi_0^t d$ for every $\lambda > 0$, whence $f'(x_0; d) \geq \xi_0^t d$. We thus have arrived at a contradiction. So the inclusion "$\subset$" must hold as well. QED

Exercise 2.18 a. In the above proof the following property is used: if $S \subset \mathbb{R}^n$ is compact, then its convex hull co S is compact. Prove this, using the following result of Carathéodory: in $\mathbb{R}^n$ every convex combination x of $p \geq n+1$ points $x_1, \ldots, x_p$ (i.e., $x = \sum_1^p \alpha_i x_i$ for $\alpha_i \geq 0$ and $\sum_1^p \alpha_i = 1$) can also be written as a convex combination of at most $n+1$ points $x_{i_1}, \ldots, x_{i_{n+1}} \subset \{x_1, \ldots, x_p\}$.
b. Give an example of a closed set $S \subset \mathbb{R}^n$ for which co S is *not* closed (conclusion: in the above proof it is essential to work with compactness).

Exercise 2.19 Let $f(x) := |x|$ on $S := \mathbb{R}$. Then $\partial f(0) = [-1, 1]$ (by Exercise 2.16(b) for $n = 1$). Demonstrate how this result can also be derived from Theorem 2.17.

Exercise 2.20 Show by means of an example that in Theorem 2.17 it is essential to have $x_0 \in \cap_i \text{int dom } f_i$.

3 The Kuhn-Tucker theorem for convex programming

We use the results of the previous section to derive the celebrated Kuhn-Tucker theorem for convex programming. Unlike its counterparts in section 4 of [1], this theorem gives necessary *and* sufficient conditions for optimality for the standard convex programming problem. First we discuss the situation with inequality constraints only.

Theorem 3.1 (Kuhn-Tucker – no equality constraints) *Let $f, g_1, \cdots, g_m : \mathbb{R}^n \to (-\infty, +\infty]$ be convex functions and let $S \subset \mathbb{R}^n$ be a convex set. Consider the* convex programming problem

$$(P) \quad \inf_{x \in S} \{f(x) : g_1(x) \leq 0, \cdots, g_m(x) \leq 0\}.$$

Let $\bar{x}$ be a feasible point of (P); denote by $I(\bar{x})$ the set of all $i \in \{1, \cdots, m\}$ for which $g_i(\bar{x}) = 0$.

(i) $\bar{x}$ is an optimal solution of (P) if there exist vectors of multipliers *$\bar{u} := (\bar{u}_1, \cdots, \bar{u}_m) \in \mathbb{R}^m_+$ and $\bar{\eta} \in \mathbb{R}^n$ such that the following three relationships hold:*

$$\bar{u}_i g_i(\bar{x}) = 0 \textit{ for } i = 1, \cdots, m \quad \textbf{(complementary slackness)},$$

$$0 \in \partial f(\bar{x}) + \sum_{i \in I(\bar{x})} \bar{u}_i \partial g_i(\bar{x}) + \bar{\eta} \quad \textbf{(normal Lagrange inclusion)},$$

$$\bar{\eta}^t (x - \bar{x}) \leq 0 \textit{ for all } x \in S \quad \textbf{(obtuse angle property)}.$$

(ii) Conversely, if $\bar{x}$ is an optimal solution of (P) and if $\bar{x} \in$ int dom $f \cap \cap_{i \in I(\bar{x})}$int dom g_i, *then there exist multipliers $\bar{u}_0 \in \{0, 1\}$, $\bar{u} \in \mathbb{R}^m_+$, $(\bar{u}_0, \bar{u}) \neq (0, 0)$, and $\bar{\eta} \in \mathbb{R}^n$ such that the complementary slackness relationship and obtuse angle property of part (i) hold, as well as the following:*

$$0 \in \bar{u}_0 \partial f(\bar{x}) + \sum_{i \in I(\bar{x})} \bar{u}_i \partial g_i(\bar{x}) + \bar{\eta} \quad \textbf{(Lagrange inclusion)}.$$

Here the normal *case is said to occur when* $\bar{u}_0 = 1$ *and the* abnormal *case when* $\bar{u}_0 = 0$.

Remark 3.2 (minimum principle) *By Theorem 2.9, the normal Lagrange inclusion in Theorem 3.1 implies*

$$-\bar{\eta} \in \partial(f + \sum_{i \in I(\bar{x})} \bar{u}_i g_i)(\bar{x}).$$

So by Theorem 2.10 and Remark 2.11 it follows that

$$\bar{x} \in \mathrm{argmin}_{x \in S}[f(x) + \sum_{i \in I(\bar{x})} \bar{u}_i g_i(x)] \textbf{(minimum principle)}.$$

Likewise, under the additional condition $\mathrm{dom}\, f \cap \cap_{i \in I(\bar{x})} \mathrm{int\ dom}\, g_i \neq \emptyset$, *this minimum principle implies the normal Lagrange inclusion by the converse parts of Theorem 2.10/Remark 2.11 and Theorem 2.9.*

Remark 3.3 (Slater's constraint qualification) *The following* Slater constraint qualification *guarantees normality: Suppose that there exists* $\tilde{x} \in S$ *such that* $g_i(\tilde{x}) < 0$ *for* $i = 1, \cdots, m$. *Then in part* (ii) *of Theorem 3.1 we have the normal case* $\bar{u}_0 = 1$.

Indeed, suppose we had $\bar{u}_0 = 0$. *For* $\bar{u}_0 = 0$ *instead of* $\bar{u}_0 = 1$ *the proof of the minimum principle in Remark 3.2 can be mimicked and gives*

$$\sum_{i=1}^{m} \bar{u}_i g_i(\bar{x}) \leq \sum_{i=1}^{m} \bar{u}_i g_i(\tilde{x}).$$

Since $(\bar{u}_1, \cdots, \bar{u}_m) \neq (0, \cdots, 0)$, *this gives* $\sum_{i=1}^{m} \bar{u}_i g_i(\bar{x}) < 0$, *in contradiction to complementary slackness.*

Proof of Theorem 3.1. Let us write $I := I(\bar{x})$. (i) By Remark 3.2 the minimum principle holds, i.e., for any $x \in S$ we have

$$f(x) + \sum_{i \in I} \bar{u}_i g_i(x) \geq f(\bar{x})$$

(observe that $\sum_{i \in I} \bar{u}_i g_i(\bar{x}) = 0$ by complementary slackness). Hence, for any *feasible* $x \in S$ we have

$$f(x) \geq f(x) + \sum_{i \in I} \bar{u}_i g_i(x) \geq f(\bar{x}),$$

by nonnegativity of the multipliers. Clearly, this proves optimality of $\bar{x}$.

(ii) Consider the auxiliary optimization problem

$$(P') \ \inf_{x \in S} \phi(x),$$

where $\phi(x) := \max[f(x) - f(\bar{x}), \max_{1 \leq i \leq m} g_i(x)]$. Since $\bar{x}$ is an optimal solution of (P), it is not hard to see that $\bar{x}$ is also an optimal solution of (P') (observe that $\phi(\bar{x}) = 0$

and that $x \in S$ is feasible if and only if $\max_{1 \le i \le m} g_i(x) \le 0$). By Theorem 2.10 and Remark 2.11 there exists $\bar{\eta}$ in $\mathbb{R}^n$ such that $\bar{\eta}$ has the obtuse angle property and $-\bar{\eta} \in \partial\phi(\bar{x})$. By Theorem 2.17 this gives

$$-\bar{\eta} \in \partial\phi(\bar{x}) = \mathrm{co}(\partial f(\bar{x}) \cup \cup_{i \in I} \partial g_i(\bar{x})).$$

Since subdifferentials are convex, we get the existence of $(u_0, \xi_0) \in \mathbb{R}_+ \times \partial f(\bar{x})$ and $(u_i, \xi_i) \in \mathbb{R}_+ \times \partial g_i(\bar{x})$, $i \in I$, such that $\sum_{i \in \{0\} \cup I} u_i = 1$ and

$$-\bar{\eta} = \sum_{i \in \{0\} \cup I} u_i \xi_i.$$

In case $u_0 = 0$, we are done by setting $\bar{u}_i := u_i$ for $i \in \{0\} \cup I$ and $\bar{u}_i := 0$ otherwise. Observe that in this case $(\bar{u}_1, \cdots, \bar{u}_m) \neq (0, \cdots, 0)$ by $\sum_{i \in I} u_i = 1$. In case $u_0 \neq 0$, we know that $u_0 > 0$, so we can set $\bar{u}_i := u_i / u_0$ for $i \in \{0\} \cup I$ and $\bar{u}_i := 0$ otherwise. QED

Example 3.4 Consider the following optimization problem:

$$(P) \text{ minimize } (x_1 - \frac{9}{4})^2 + (x_2 - 2)^2$$

over all $(x_1, x_2) \in \mathbb{R}^2_+$ such that

$$\begin{aligned} x_1^2 - x_2 &\le 0 \\ x_1 + x_2 - 6 &\le 0 \\ -x_1 + 1 &\le 0 \end{aligned}$$

Since Slater's constraint qualification clearly holds, we get that a feasible point $(\bar{x}_1, \bar{x}_2)$ is optimal if and only if there exists $(\bar{u}_1, \bar{u}_2, \bar{u}_3) \in \mathbb{R}^3_+$ such that

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 2(\bar{x}_1 - \frac{9}{4}) \\ 2(\bar{x}_2 - 2) \end{pmatrix} + \bar{u}_1 \begin{pmatrix} 2\bar{x}_1 \\ -1 \end{pmatrix} + \bar{u}_2 \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \bar{u}_3 \begin{pmatrix} -1 \\ 0 \end{pmatrix} + \begin{pmatrix} \bar{\eta}_1 \\ \bar{\eta}_2 \end{pmatrix}$$

for some $\bar{\eta} := (\bar{\eta}_1, \bar{\eta}_2)^t$ with

$$\bar{\eta}^t (x - \bar{x}) \le 0 \text{ for all } x \in \mathbb{R}^2_+$$

and such that

$$\begin{aligned} \bar{u}_1(\bar{x}_1^2 - \bar{x}_2) &= 0 \\ \bar{u}_2(\bar{x}_1 + \bar{x}_2 - 6) &= 0 \\ \bar{u}_3(-\bar{x}_1 + 1) &= 0 \end{aligned}$$

Let us first deal with $\bar{\eta}$: observe that the above obtuse angle property forces $\bar{\eta}_1$ and $\bar{\eta}_2$ to be nonpositive, and $\bar{x}_i > 0$ even implies $\bar{\eta}_i = 0$ for $i = 1, 2$ (this can be seen as a form of complementarity). Since $\bar{x}_1 \ge 1$, this means $\bar{\eta}_1 = 0$. Also, $\bar{x}_2 = 0$ stands no chance, because it would mean $\bar{x}_1^2 \leq 0$. Hence, $\bar{\eta} = 0$. We now distinguish the following possibilities for the set $I := I(\bar{x})$:

Case 1 ($I = \emptyset$): By complementary slackness, $\bar{u}_1 = \bar{u}_2 = \bar{u}_3 = 0$, so the Lagrange inclusion gives $\bar{x}_1 = 9/4$, $\bar{x}_2 = 2$, which violates the first constraint ($(9/4)^2 \not\leq 2$).

Case 2 ($I = \{1\}$): By complementary slackness, $\bar{u}_2 = \bar{u}_3 = 0$. The Lagrange inclusion gives $\bar{x}_1 = \frac{9}{4}(1 + \bar{u}_1)^{-1}$, $\bar{x}_2 = \bar{u}_1/2 + 2$, so, since $\bar{x}_1^2 = \bar{x}_2$, by definition of I, we obtain the equation $\bar{u}_1^3 + 6\bar{u}_1^2 + 9\bar{u}_1 = 49/8$, which has $\bar{u}_1 = 1/2$ as its only solution. It follows then that $\bar{x} = (3/2, 9/4)^t$.

At this stage we can already stop: Theorem 3.1(i) guarantees that, in fact, $\bar{x} = (3/2, 9/4)^t$ is an optimal solution of (P). Moreover, since the objective function $(x_1, x_2) \mapsto (x_1 - \frac{9}{4})^2 + (x_2 - 2)^2$ is *strictly* convex, it follows that any optimal solution of (P) must be unique. So $\bar{x} = (3/2, 9/4)^t$ is the unique optimal solution of (P).

Exercise 3.1 Consider the optimization problem

$$(P) \quad \sup_{(\xi_1,\xi_2) \in \mathbb{R}^2_+} \{\xi_1 \xi_2 : 2\xi_1 + 3\xi_2 \leq 5\}.$$

Solve this problem using Theorem 3.1. *Hint:* The set of optimal solutions does not change if we apply a monotone transformation to the objective function. So one can use $f(\xi_1, \xi_2) := \sqrt{\xi_1 \xi_2}$ to ensure convexity (see Exercise 2.11).

Exercise 3.2 Let $a_i > 0$, $i = 1, \ldots, n$ and let $p \geq 1$. Consider the optimization problem

$$(P) \quad \text{maximize} \sum_{i=1}^{n} a_i \xi_i \text{ over } (\xi_1, \ldots, \xi_n) \in \mathbb{R}^n$$

subject to $g(\xi) := \sum_{i=1}^n |\xi_i|^p = 1$.

a. Show that if the constraint $\sum_{i=1}^n |\xi_i|^p = 1$ is replaced by $\sum_{i=1}^n |\xi_i|^p \leq 1$, then this results in exactly the same optimal solutions.

b. Prove that $g : \mathbb{R}^n \to \mathbb{R}$, as defined above, is convex. Prove also that g is in fact strictly convex if $p > 1$.

c. Apply Theorem 3.1 to determine the optimal solutions of (P). *Hint:* Treat the cases $p = 1$ and $p > 1$ separately.

d. Derive from the result obtained in part (c) for $p > 1$ the following famous *Hölder inequality*, which is an extension of the Cauchy-Schwarz inequality: $|\sum_i a_i \xi_i| \leq (\sum_i a_i^q)^{1/q} (\sum_i |\xi_i|^p)^{1/p}$ for all $(\xi_1, \ldots, \xi_n) \in \mathbb{R}^n$. Here q is defined by $q := p/(p-1)$.

Corollary 3.5 (Kuhn-Tucker – general case) *Let $f, g_1, \cdots, g_m : \mathbb{R}^n \to (-\infty, +\infty]$ be convex functions, let $S \subset \mathbb{R}^n$ be a convex set. Also, let A be a $p \times n$-matrix and let $b \in \mathbb{R}^p$. Define $L := \{x : Ax = b\}$. Consider the* convex programming problem

$$(P) \quad \inf_{x \in S} \{f(x) : g_1(x) \leq 0, \cdots, g_m(x) \leq 0, Ax - b = 0\}.$$

Let $\bar{x}$ be a feasible point of (P); denote by $I(\bar{x})$ the set of all $i \in \{1, \cdots, m\}$ for which $g_i(\bar{x}) = 0$.

(i) $\bar{x}$ is an optimal solution of (P) if there exist vectors of multipliers $\bar{u} \in \mathbb{R}^m_+$, $\bar{v} \in \mathbb{R}^p$ and $\bar{\eta} \in \mathbb{R}^n$ such that the complementary slackness relationship and the obtuse angle property hold just as in Theorem 3.1(i), as well as the following version of the normal Lagrange inclusion:

$$0 \in \partial f(\bar{x}) + \sum_{i \in I(\bar{x})} \bar{u}_i \partial g_i(\bar{x}) + A^t \bar{v} + \bar{\eta}.$$

(ii) Conversely, if $\bar{x}$ is an optimal solution of (P) and if both $\bar{x} \in \operatorname{int} \operatorname{dom} f \cap \cap_{i \in I(\bar{x})} \operatorname{int} \operatorname{dom} g_i$ and $\operatorname{int} S \cap L \neq \emptyset$, then there exist multipliers $\bar{u}_0 \in \{0, 1\}$, $\bar{u} \in \mathbb{R}^m_+$, $(\bar{u}_0, \bar{u}) \neq (0, 0)$, and $\bar{v} \in \mathbb{R}^p$, $\bar{\eta} \in \mathbb{R}^n$ such that the complementary slackness relationship and obtuse angle property of part (i) hold, as well as the following Lagrange inclusion:

$$0 \in \bar{u}_0 \partial f(\bar{x}) + \sum_{i \in I(\bar{x})} \bar{u}_i \partial g_i(\bar{x}) + A^t \bar{v} + \bar{\eta}.$$

PROOF. Observe that $\partial \chi_L(\bar{x}) = \operatorname{im} A^t$. Indeed, $\eta \in \partial \chi_L(\bar{x})$ is equivalent to $\eta^t(x - \bar{x}) \leq 0$ for all $x \in L$, i.e., to $\eta^t(x - \bar{x}) = 0$ for all $x \in \mathbb{R}^n$ with $A(x - \bar{x}) = 0$. But the latter states that η belongs to the bi-orthoplement of the linear subspace $\operatorname{im} A^t$, so it belongs to $\operatorname{im} A^t$ itself. This proves the observation. Let us note that the above problem (P) is precisely the same problem as the one of Theorem 3.1, but with S replaced by $S' := S \cap L$. Thus, parts (i) and (ii) follow directly from Theorem 3.1, but now $\bar{\eta}$ as in Theorem 3.1 has to be replaced by an element (say η') in $\partial \chi_{S'}$. From Theorem 2.9 we know that

$$\partial \chi_{S'}(\bar{x}) = \partial \chi_S(\bar{x}) + \partial \chi_L(\bar{x}),$$

in view of the condition $\operatorname{int} S \cap L \neq \emptyset$. Therefore, η' can be decomposed as $\eta' = \bar{\eta} + \eta$, with $\bar{\eta} \in \partial \chi_S(\bar{x})$ (this amounts to the obtuse angle property, of course), and with $\eta \in \partial \chi_L(\bar{x})$. By the above there exists $\bar{v} \in \mathbb{R}^m$ with $\eta = A^t \bar{v}$ and this finishes the proof. QED

Example 3.6 Let $c_1, \cdots, c_n$, $a_1, \cdots, a_n$ and b be positive real numbers. Consider the following optimization problem:

$$(P) \text{ minimize } \sum_{i=1}^n \frac{c_i}{x_i}$$

over all $x = (x_1, \cdots, x_n)^t \in \mathbb{R}^n_{++}$ (the strictly positive orthant) such that

$$\sum_{i=1}^n a_i x_i = b.$$

Let us try to meet the sufficient conditions of Corollary 3.5(i). Thus, we must find a feasible $\bar{x} \in \mathbb{R}^n$ and multipliers $\bar{v} \in \mathbb{R}$, $\bar{\eta} \in \mathbb{R}^n$ such that

$$\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} -\frac{c_1}{\bar{x}_1^2} \\ \vdots \\ -\frac{c_n}{\bar{x}_n^2} \end{pmatrix} + \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \bar{v} + \bar{\eta}.$$

and such that the obtuse angle property holds for $\bar{\eta}$. To begin with the latter, since we seek $\bar{x}$ in the open set $S := \mathbb{R}^n_{++}$, the only $\bar{\eta}$ with the obtuse angle property is $\bar{\eta} = 0$. The above Lagrange inclusion gives $\bar{x}_i = (c_i/(\bar{v}a_i))^{1/2}$ for all i. To determine $\bar{v}$, which must certainly be positive, we use the constraint: $b = \sum_i a_i\bar{x}_i = \sum_i (a_i c_i/\bar{v})^{1/2}$, which gives $\bar{v} = (\sum_i (a_i c_i)^{1/2}/b)^2$. Thus, all conditions of Corollary 3.5(i) are seen to hold: an optimal solution of (P) is $\bar{x}$, given by

$$\bar{x}_i = \sqrt{\frac{c_i}{a_i}} \frac{b}{\sum_{j=1}^n \sqrt{a_j c_j}},$$

and it is implicit in our derivation that this solution is unique (exercise).

Remark 3.7 *By using the* relative interior *(denoted as "ri") of a convex set, i.e., the interior relative to the linear variety spanned by that set, one can obtain the following improvement of the nonempty intersection condition in Theorem 2.9: it is already enough that* ri dom $f \cap$ dom g *is nonempty. Since one can also prove that* $A(\text{ri } S) = \text{ri } A(S)$ *for any convex set* $S \subset \mathbb{R}^n$ *and any linear mapping* $A : \mathbb{R}^n \to \mathbb{R}^p$ *[2, Theorem 4.9], it follows that the nonempty intersection condition in Corollary 3.5 can be improved considerably into* ri $S \cap L \neq \emptyset$ *or,* equivalently*, into* $b \in A(\text{ri } S)$.

Exercise 3.3 In the above proof of Corollary 3.5 the fact was used that for a linear subspace M of $\mathbb{R}^n$ the following holds: let

$$M^\perp := \{x \in \mathbb{R}^n : x^t\xi = 0 \text{ for all } \xi \in M\},$$

This is a linear subspace itself (prove this), so $M^{\perp\perp} := (M^\perp)^\perp$ is well-defined. Prove that $M = M^{\perp\perp}$. *Hint:* This identity can be established by proving two inclusions; one of these is elementary and the other requires the use of projections.

Exercise 3.4 What becomes of Corollary 3.5 in the situation where there are no inequality constraints (i.e., just equality constraints)? Derive this version.

Exercise 3.5 Use Corollary 3.5 to prove the following famous theorem of Farkas. Let A be a $p \times n$-matrix and let $c \in \mathbb{R}^n$. Then precisely one of the following is true:

$$(1)\ \exists_{x\in\mathbb{R}^n} Ax \leq 0 \text{ (componentwise) and } c^t x > 0,\ \ (2)\ \exists_{y\in\mathbb{R}^p_+} A^t y = c.$$

Hint: Show first, by elementary means, that validity of (2) implies that (1) cannot hold. Next, apply Corollary 3.5 to a suitably chosen optimization problem in order to prove that if (1) does not hold, then (2) must be true.

A Standard material on convexity

Definition A.1 A set S in $\mathbb{R}^n$ is said to be *convex* if for every $x_1, x_2 \in S$ the line segment $\{\lambda x_1 + (1-\lambda)x_2 : 0 \leq \lambda \leq 1\}$ belongs to S.

For instance, a hyperplane $S = \{x \in \mathbb{R}^n : p^t x = \alpha\}$ or a ball $S = \{x \in \mathbb{R}^n : |x - x_0| \leq \beta\}$ are examples of convex sets. However, the sphere $S = \{x \in \mathbb{R}^n : |x - x_0| = \beta\}$ provides an example of a set that is not convex ($\beta > 0$). It is easy to see that arbitrary intersections of convex sets are again convex; also finite sums of convex sets are convex again.

Theorem A.2 (strict point-set separation [1, Thm. 2.4.4]) *Let S be a nonempty closed convex subset of $\mathbb{R}^n$ and let $y \in \mathbb{R}^n \backslash S$. Then there exists $p \in \mathbb{R}^n$, $p \neq 0$, such that*

$$\sup_{x \in S} p^t x < p^t y.$$

PROOF. It is a standard result that there exists $\hat{x} \in S$ such that $\sup_{s \in S} |y - s| = |y - \hat{x}|$ (consider a suitable closed ball around y and apply the theorem of Weierstrass [1, Thm. 2.3.1]). By convexity of S, this means that for every $x \in S$ and every $\lambda \in (0, 1]$

$$|y - (\lambda x + (1-\lambda)\hat{x})|^2 \geq |y - \hat{x}|^2.$$

Obviously, the expression on the left equals

$$|y - \hat{x} - \lambda(x - \hat{x})|^2 = |y - \hat{x}|^2 - 2\lambda(y - \hat{x})^t(x - \hat{x}) + \lambda^2|x - \hat{x}|^2,$$

so the above inequality amounts to

$$2\lambda(y - \hat{x})^t(x - \hat{x}) \leq \lambda^2|x - \hat{x}|^2$$

for every $x \in S$ and every $\lambda \in (0, 1]$. Dividing by $\lambda > 0$ and letting λ go to zero then gives

$$(y - \hat{x}) \cdot (x - \hat{x}) \leq 0 \text{ for all } x \in S.$$

Set $p := y - \hat{x}$; then $p \neq 0$ (note that $p = 0$ would imply $y \in S$). We clearly have $p^t x \leq p^t \hat{x}$. Also, we have now $p^t \hat{x} > p^t y$, for otherwise $(y - \hat{x})^t(\hat{x} - y) \geq 0$ would imply $y = \hat{x} \in S$, which is impossible. QED

For our next result, recall that $\partial S := \text{cl} S \cap \text{cl}(\mathbb{R}^n \backslash S) = \text{cl} S \backslash \text{int S}$ denotes the boundary of a set $S \subset \mathbb{R}^n$.

Theorem A.3 (supporting hyperplane [1, Thm. 2.4.7]) *Let S be a nonempty convex subset of $\mathbb{R}^n$ and let $y \in \partial S$. Then there exists $q \in \mathbb{R}^n$, $q \neq 0$, such that*

$$\sup_{x \in cl\ S} q^t x \leq q^t y.$$

In geometric terms, $H := \{x \in \mathbb{R}^n : q^t x = q^t y\}$ is said to be a supporting hyperplane *for S at y: the hyperplane H contains the point y and the set S (as well as* cl S*) is contained the halfspace $\{x \in \mathbb{R}^n : p^t x \leq p^t y\}$.*

PROOF. Let $Z := \mathrm{cl}\ S$; then $\partial S \subset \partial Z$ (exercise). Of course, Z is closed and it is easy to show that Z is convex (use limit arguments). So there exists a sequence (y_k) in $\mathbb{R}^n \backslash Z$ such that $y_k \to y$. By Theorem A.2 there exists for every k a nonzero vector $p_k \in \mathbb{R}^n$ such that

$$\sup_{x \in Z} p_k^t x < p_k^t y_k.$$

Division by $|p_k|$ turns this into

$$\sup_{x \in Z} q_k^t x < q_k^t y_k,$$

where $q_k := p_k/|p_k|$ belongs to the unit sphere of $\mathbb{R}^n$. This sphere is compact (Bolzano-Weierstrass theorem), so we can suppose without loss of generality that (q_k) converges to some q, $|q| = 1$ (so q is nonzero). Now for every $x \in Z$ the inequality $q_k^t x < q_k^t y_k$, which holds for all k, implies

$$q^t x = \lim_k q_k^t x \leq \lim_k q_k^t y_k = q^t y,$$

and the proof is finished. QED

Theorem A.4 (set-set separation [1, Thm. 2.4.8]) *Let S_1, S_2 be two nonempty convex sets in $\mathbb{R}^n$ such that $S_1 \cap S_2 = \emptyset$. Then there exist $p \in \mathbb{R}^n$, $p \neq 0$, and $\alpha \in \mathbb{R}$ such that*

$$\sup_{x \in S_1} p^t x \leq \alpha \leq \inf_{y \in S_2} p^t y.$$

In geometric terms, $H := \{x \in \mathbb{R}^n : p^t x = \alpha\}$ is said to be a separating hyperplane *for S_1 and S_2: each of the two convex sets is contained in precisely one of the two halfspaces $\{x \in \mathbb{R}^n : p^t x \leq \alpha\}$ and $\{x \in \mathbb{R}^n : p^t x \geq \alpha\}$.*

PROOF. It is easy to see that $S := S_1 - S_2$ is convex. Now $0 \notin S$, for otherwise we get an immediate contradiction to $S_1 \cap S_2 = \emptyset$. W distinguish now two cases: (i) $0 \in \mathrm{cl}\ S$ and (ii) $0 \notin \mathrm{cl}\ S$.

In case (i) we have $0 \in \partial S$, so by Theorem A.3 we then have the existence of a nonzero $p \in \mathbb{R}^n$ such that

$$p^t z \leq 0 \text{ for every } z \in S = S_1 - S_2, \tag{2}$$

i.e., for every $z = x - y$, with $x \in S_1$ and $y \in S_2$. This gives $p^t x \leq p^t y$ for all $x \in S_1$ and $y \in S_2$, whence the result.

In case (ii) we apply Theorem A.2 to get immediately (2) as well. The result follows just as in case (i). QED

Theorem A.5 (strong set-set separation [1, Thm. 2.4.10]) *Let S_1, S_2 be two nonempty closed convex sets in $\mathbb{R}^n$ such that $S_1 \cap S_2 = \emptyset$ and such that S_1 is bounded. Then there exist $p \in \mathbb{R}^n$, $p \neq 0$, and $\alpha \in \mathbb{R}$, $\beta \in \mathbb{R}$ such that*

$$\sup_{x \in S_1} p^t x \leq \alpha < \beta \leq \inf_{y \in S_2} p^t y.$$

PROOF. As in the previous proof, it is easy to see that $S := S_1 - S_2$ is convex. Now S is also seen to be closed (exercise). As in the previous proof, we have $0 \notin S$. We can now apply Theorem A.2 to get the desired result, just as in case (ii) of the previous proof. QED

B Fenchel conjugation

Definition B.1 For a function $f : \mathbb{R}^n \to (-\infty, +\infty]$ the *(Fenchel) conjugate* function of f is $f^* : \mathbb{R}^n \to [-\infty, +\infty]$, given by

$$f^*(\xi) := \sup_{x \in \mathbb{R}^n} [\xi^t x - f(x)].$$

By repeating the conjugation operation one also defines the *(Fenchel) biconjugate* of f, which is simply given by $f^{**} := (f^*)^*$.

Example B.2 Consider $f : \mathbb{R} \to \mathbb{R}$, given by

$$f(x) := \begin{cases} x \log x & \text{if } x > 0, \\ 0 & \text{if } x = 0, \\ +\infty & \text{if } x < 0. \end{cases}$$

Observe that this function is convex. Then (counting $0 \log 0$ as 0) we clearly have $f^*(\xi) = \sup_{x \geq 0} \xi x - x \log x$ for the conjugate. For an *interior* maximum in $\mathbb{R}_+$ (by concavity of the function to be maximized) the necessary and sufficient condition is $\xi - \log x - 1 = 0$, i.e., $x = \exp(\xi - 1)$, which gives the value $\xi x - x \log x = \exp(\xi - 1)$. Since this value is positive, we conclude that the point $x = 0$ stands no chance for the maximum, i.e., the maximum is always interior, as calculated above, giving $f^*(\xi) = \exp(\xi - 1)$ for the conjugate function. We can also determine the biconjugate function: by definition, $f^{**}(x) = \sup_{\xi \in \mathbb{R}} x\xi - \exp(\xi - 1)$. If $x < 0$, then, by $\exp(\xi - 1) \to 0$ as $\xi \to -\infty$, the supremum value is clearly $+\infty$. Hence, $f^{**}(x) = +\infty$ for $x < 0$. If $x > 0$, then setting the derivative of the concave function $\xi \mapsto x\xi - \exp(\xi - 1)$ equal to zero gives a solution (whence a global maximum) for $\xi = \log x + 1$. Hence $f^{**}(x) = x \log x$ for $x > 0$. Finally, if $x = 0$, then the supremum of $-\exp(\xi - 1)$ is clearly the limit value 0. So $f^{**}(0) = 0$. We conclude that $f^{**} = f$ in this example. The Fenchel-Moreau theorem below will support this observation.

Exercise B.1 Determine for each of the following functions f the conjugate function f^* and verify also explicitly if $f = f^{**}$ holds.
a. $f(x) = ax^2 + bx + c$, $a \geq 0$,
b. $f(x) = |x| + |x - 1|$,
c. $f(x) = x^a/a$ for $x \geq 0$ and $f(x) = +\infty$ for $x < 0$ (here $a \geq 1$).
d. $f = \chi_B$, where B is the closed unit ball in $\mathbb{R}^n$.

Example B.3 Let K be a nonempty convex cone in $\mathbb{R}^n$ (recall that a *cone* (at zero) is a set such that $\alpha x \in K$ for every $\alpha > 0$ and $x \in K$; cf. Definition 2.5.1 in [1]). Let $f := \chi_K$. Then

$$f^*(\xi) = \sup_{x \in K} \xi^t x = \begin{cases} 0 & \text{if } \xi \in K^*, \\ +\infty & \text{otherwise.} \end{cases}$$

Recall here that K^*, the *polar cone* of K, is defined by $K^* := \{\xi \in \mathbb{R}^n : \xi^t x \leq 0 \text{ for all } x \in K\}$. Hence, we conclude that $(\chi_K)^* = \chi_{K^*}$.

Denote the closure of K by $\bar{K}$. We also observe that $\xi \in \partial \chi_{\bar{K}}(0)$ is equivalent to $\xi^t x \leq 0$ for all $x \in \bar{K}$, i.e., to $\xi^t x \leq 0$ for all $x \in K$, i.e., to $\xi \in K^*$.

Proposition B.4 *Let $f,\ g : \mathbb{R}^n \to (-\infty, +\infty]$.*
(i) If $f \geq g$ then $f^* \leq g^*$.
(ii) If $f^*(x) = -\infty$ for some $x \in \mathbb{R}^n$, then $f \equiv +\infty$.
(iii) For every $x_0,\ \xi \in \mathbb{R}^n$

$$f^*(\xi) \geq \xi^t x_0 - f(x_0) \ \textit{(Young's inequality).}$$

(iv) $f \geq f^{**}$.
(v) For every $x_0,\ \xi \in \mathbb{R}^n$

$$f^*(\xi) = \xi^t x_0 - f(x_0) \text{ if and only if } \xi \in \partial f(x_0).$$

Exercise B.2 Give a proof of Proposition B.4.

Theorem B.5 (Fenchel-Moreau) *Let $f : \mathbb{R}^n \to (-\infty, +\infty]$ be convex. Then*

$$f(x_0) = f^{**}(x_0) \text{ if and only if } f \text{ is lower semicontinuous at } x_0.$$

Proof. One implication is very simple: if $f(x_0) = f^{**}(x_0)$, and if $x_n \to x_0$ then $\liminf_n f(x_n) \geq \liminf_n f^{**}(x_n)$ by Proposition B.4(*iv*). Also, $\liminf_n f^{**}(x_n) \geq f^{**}(x_0)$ because every conjugate, being the supremum of a collection of continuous functions, is automatically lower semicontinuous. So we conclude that $\liminf_n f(x_n) \geq f^{**}(x_0) = f(x_0)$, i.e., f is lower semicontinuous at x_0.

In the converse direction, by Proposition B.4(*iv*) it is enough to prove $f^{**}(x_0) \geq r$ for an arbitrary $r < f(x_0)$, both when $f(x_0) < +\infty$ and when $f(x_0) = +\infty$.

Case 1: $f(x_0) < +\infty$. It is easy to check that $C := \text{epi } f := \{(x, r) \in \mathbb{R}^n \times \mathbb{R} : r \geq f(x)\}$, the epigraph of f, is a convex set in $\mathbb{R}^{n+1}$ (this is Theorem 3.2.2 in [1] – as can be seen immediately from its proof, it continues to hold for functions with values in $(-\infty, +\infty]$ and we know already that this theorem also holds for sets with empty interior). Hence, the closure cl C is also convex. We claim now that $(x_0, r) \notin \text{cl } C$. For suppose (x_0, r) would be the limit of a sequence of points $(x_n, y_n) \in C$. Then $y_n \geq f(x_n)$ for each n, and in the limit this would give $r \geq \liminf_n f(x_n) \geq f(x_0)$ by lower semicontinuity of f at x_0. This contradiction proves that the claim holds. We may now apply separation [1, Theorem 2.4.10]: there exist $\alpha \in \mathbb{R}$ and $p =: (\xi_0, \mu) \neq (0, 0)$, with $\xi_0 \in \mathbb{R}^n$ and $\mu \in \mathbb{R}$, such that

$$\xi_0^t x + \mu y \leq \alpha < \xi_0^t x_0 + \mu r \text{ for all } (x, y) \in C. \tag{3}$$

It is clear that $\mu \leq 0$ by the definition of C. Also, it is obvious that $\mu \neq 0$ (just consider what happens if we take $(x, y) = (x_0, f(x_0))$ in (3) – and we may do this by virtue of $f(x_0) \in \mathbb{R}$). Hence, we can divide by $-\mu$ in (3) and get

$$\xi_1^t x - f(x) \leq \xi_1^t x_0 - r \text{ for all } x \in \text{dom } f.$$

Notice that this inequality continues to hold outside dom f as well; thus, $f^*(\xi_1) \leq \xi_1^t x_0 - r$, which implies the desired inequality $f^{**}(x_0) \geq r$.

Case 2a: $f \equiv +\infty$. In this case, the desired result is trivial, for $f^* \equiv -\infty$, so $f^{**} \equiv +\infty$.

Case 2b: $f(x_1) < +\infty$ for some $x_1 \in \mathbb{R}^n$. We can repeat the proof of Case 1 until (3). If μ happens to be nonzero, then of course we finish as in Case 1. However, if $\mu = 0$ we only get

$$\xi_0^t x \leq \alpha < \xi_0^t x_0 \text{ for all } x \in \text{dom } f$$

from (3). We then repeat the full proof of Case 1, but with x_0 replaced by x_1 and r by $f(x_1) - 1$. This gives the existence of $\xi \in \mathbb{R}^n$ such that

$$\xi^t x - f(x) \leq \xi^t x_1 - f(x_1) + 1 \text{ for all } x \in \text{dom } f.$$

Now for any $\lambda > 0$, observe that by the two previous inequalities

$$f(x) \geq (\xi + \lambda\xi_0)^t x - \xi^t x_1 + f(x_1) - 1 - \alpha\lambda \text{ for all } x \in \mathbb{R}^n,$$

which implies $f^*(\xi + \lambda\xi_0) \leq \xi^t x_1 - f(x_1) + 1 + \lambda\alpha$. By definition of $f^{**}(x_0)$, this gives

$$f^{**}(x_0) \geq \lambda(\xi_0^t x_0 - \alpha) + \xi^t x_0 - \xi^t x_1 + f(x_1) - 1,$$

which implies $f^{**}(x_0) = +\infty$, by letting λ go to infinity (note that $\xi_0^t x_0 - \alpha > 0$ by the above). QED

Corollary B.6 (bipolar theorem for cones) *Let K be a closed convex cone in $\mathbb{R}^n$. Then $K = K^{**} := (K^*)^*$.*

PROOF. Observe that $f := \chi_K$ is a lower semicontinuous convex function. Hence, $f^{**} = f$ by Theorem B.5. By Example B.3 we know that $f^* = \chi_{K^*}$, so $f^{**} = \chi_{K^{**}}$ follows by another application of this fact. Hence $\chi_K = \chi_{K^{**}}$. QED

Exercise B.3 Prove Farkas' theorem (see Exercise 3.5) by means of Corollary B.6.

Exercise B.4 Redo Exercise 3.3 by making it a special case of Corollary B.6.

References

[1] M.S. Bazaraa, H.D. Sherali and C.M. Shetty, *Nonlinear Programming: Theory and Algorithms.* Wiley, New York, 1993.

[2] J. van Tiel, *Convex Analysis: An Introductory Text.* Wiley, 1984.