# Optimization - Master Programme Quantitative Finance

Prof. Rüdiger Frey,
ruediger.frey@wu.ac.at

Version from March 31, 2014

# Contents

# Introduction

These notes are the written version of an introductory lecture on optimization that was held in the master QFin at WU Vienna. The notes are based on selected parts of Bertsekas (1999) and we refer to that source for further information. For additional material on linear optimization we refer to Bertsimas & Tsitsiklis (1997) or to **?**.

These notes are of a somewhat preliminary character and they are definitely not "error-free". In fact, I would be grateful to learn about misprints and other mistakes.

## Optimization problems

In its most general for an optimization problem is

$$\text{minimize } f(x) \text{ subject to } x \in \tilde{X} \tag{0.1}$$

Here the *set of admissible points* $\tilde{X}$ is a subset of $\mathbb{R}^n$, and the *cost function* $f$ is a function that maps $\tilde{X}$ to $\mathbb{R}$. Note that maximization problems can be addressed by replacing $f$ with $-f$, as $\sup_{x\in\tilde{X}} f(x) = -\inf_{x\in\tilde{X}}\{-f(x)\}$. Often the set of admissible points is further restricted by explicit inequality constraints; see for instance Chapter 2 below.

**Types of optimization problems.** Optimization problems can be classified according to a number of different criteria:

- Continuous problems. Here $\tilde{X}$ is of 'continuous nature' such as $\tilde{X} = \mathbb{R}^n$ or sets of the form $\tilde{X} = \{x \in \mathbb{R}^n : g(x) \leq 0 \text{ for some } g : \mathbb{R}^n \to \mathbb{R}^n\}$. These problems are usually tackled using calculus or convex analysis.
- Discrete problems. Here $\tilde{X}$ is a (usually large) finite set. A typical example is network optimization, where one considers connections between a large set of nodes and where the optimizer has to decide which connections are open or closed.
- Nonlinear programming. Here $f$ is nonlinear or the constrained set $\tilde{X}$ is specified by nonlinear equations.
- Linear programming. Here $f$ and $g$ are linear, that is (0.1) takes the form

  $$\min c'x \text{ such that } Ax \leq b$$

  for $c, x \in \mathbb{R}^n$, a matrix $A \in \mathbb{R}^{m\times n}$, a vector $b \in \mathbb{R}^m$ and $m \leq n$.

**Optimization problems in finance, economics and statistics.** a) A number of interesting optimization problems stem from the field of *portfolio optimization.* We give two examples.

- Maximization of expected utility. This problem is of the form

  $$\max_{\theta\in\mathbb{R}^n} \mathbb{E}\left(u(V_0 + \theta(S_T - S_0))\right) .$$

  Here $V_0$ is the initial wealth of an investor, $S_0 = (S_0^1, \dots S_0^n$ is the initial asset price, $S_T(\omega) = (S_T^1(\omega), \dots S_T^n(\omega))$ is the terminal asset price and the optimization variable

$\theta$ represents the portfolio strategy. The increasing function $u : \mathbb{R} \to \mathbb{R}$ is the utility function of the investor that is used to model the attitude of the investor towards risk. Typically it is assumed that $u$ is concave which corresponds to the case of a risk-averse investor.

- Markowitz problem. Here one looks for the minimal-variance portfolio under all portfolios with a given mean (see Example 2.6).

b) Calibration problems. Denote by $g_1(\theta), \ldots, g_m(\theta)$ model prices of $m$ financial instruments for a given value of a parameter vector $\theta \in \tilde{X} \subset \mathbb{R}^n$ and by $g_1^*, \ldots, g_m^*$ the prices of these instruments observed on the market. Model calibration leads to the optimization problem

$$\min_{\theta \in \tilde{X}} \frac{1}{2} \sum_{i=1}^{m} (g_i(\theta) - g_i^*)^2.$$

If $g_i$ is linear in $\theta$ we have a standard regression problem; otherwise one speaks of a generalized regression problem.

c) Maximum likelihood methods in statistics.

d) Duality results from convex analysis are crucial in financial mathematics (think of the first fundamental theorem of asset pricing or the superhedging duality).

## Overview

In Chapter 1 we treat unconstrained optimization problems where $\tilde{X} = \mathbb{R}^n$. The focus will be on the characterization of local minima via conditions of first and second order. Moreover we discuss numerical approaches based on these criteria. Chapter 2 introduces the theory of Lagrange multipliers where one uses arguments from calculus in order to derive first and second order characterizations for optima of constrained optimization problems. In Chapter 3 we introduce some key concepts from convex analysis. These concepts are then applied to constrained optimization problems where the objective function $f$ and the constrained set $\tilde{X}$ are convex. A special case are linear optimization problems.

# Chapter 1

# Unconstrained Optimization

Here we consider problems of the form

$$\text{minimize } f(x) \text{ for } \; x \in \tilde{X} = \mathbb{R}^n \tag{1.1}$$

Moreover, we assume that $f$ is once or twice continuously differentiable. Most results hold also in the case where $\tilde{X}$ is an open subset of $\mathbb{R}^n$.

**Notation.** In the sequel we use the following notation:

- Suppose that $f$ is once continuously differentiable ($C^1$). Then $\nabla f(x) = \left(\frac{\partial f}{\partial x_1}(x), \dots, \frac{\partial f}{\partial x_n}(x)\right)'$ . is the *gradient* of $f$ (a row vector);
- Suppose that $f$ is twice continuously differentiable ($C^2$). Then the matrix $Hf$ with $Hf_{ij}(x) = \frac{\partial^2 f}{\partial x_i \partial x_j}(x)$ is the *Hessian matrix* of $f$.
- For a $C^1$ function $g : \mathbb{R}^n \to \mathbb{R}^m$ the *gradient matrix* is given by

  $$\nabla g(x) = (\nabla g_1(x), \dots, \nabla g_m(x)),$$

  where $g(x) = (g_1(x), \dots, g_m(x))'$; $\nabla g$ is the transpose of the (more standard) Jacobian matrix of $g$.

## 1.1 Optimality conditions

**Definition** 1.1. Consider the optimization problem (1.1).

i) $x^*$ is called (unconstrained) local minimum of $f$ if there is some $\delta > 0$ such that $f(x^*) \leq f(x) \; \forall x \in \mathbb{R}^n$ with $\|x - x^*\| < \delta$.

ii) $x^*$ is called global minimum of $f$, if $f(x^*) \leq f(x)$ for all $x \in \mathbb{R}^n$.

iii) $x^*$ is said to be a strict local/global minimum if the inequality $f(x^*) \leq f(x)$ is strict for $x \neq x^*$.

iv) The value of the optimization problem is $f^* := \inf\{f(x) : x \in \mathbb{R}^n\}$

**Remark 1.2.** Local and global maxima are defined analogously.

**Necessary optimality conditions.** The most important tool for deriving necessary conditions for a local minimum of a differentiable function is Taylor expansion. Consider some $x^* \in \mathbb{R}^n$. If $f$ is $C^1$, we get for any $y \in \mathbb{R}^n$

$$f(x^* + y) - f(x^*) = \nabla f(x^*)' y + R(x^*, y) \tag{1.2}$$

where it holds that

$$\lim_{\|y\| \to 0} \frac{R(x^*, y)}{\|y\|} = 0. \tag{1.3}$$

Suppose now that $x^*$ is a local minimum of $f$. Fix some direction $d \in \mathbb{R}^n$ and assume without loss of generality that $\|d\| = 1$. Since $x^*$ is a local minimum of $f$, $f(x^* + hd) - f(x^*) \geq 0$ for all $h > 0$ sufficient small. Dividing by $\|hd\|$ we hence get

$$0 \leq \frac{1}{h}\left(f(x^* + hd) - f(x^*)\right) = \nabla f(x^*)' d + \frac{R(x^*, hd)}{\|hd\|}$$

Using (1.3) we thus get for for $h \to 0$ that

$$\nabla f(x^*)' d \geq 0 \text{ for all } d \in \mathbb{R}^n \text{ with } \|d\| = 1 \tag{1.4}$$

This is possible only for $\nabla f(x^*) = 0$ so that we have shown that at a local minimum the necessary condition $\nabla f(x^*) = 0$ must hold.

If $f$ is $C^2$ the Taylor formula becomes

$$f(x^* + y) - f(x^*) = \nabla f(x^*)' y + \frac{1}{2} y' H f(x^*) y + R_2(x^*, y)$$

where $\lim_{\|y\| \to 0} \frac{R_2(x^*, y)}{\|y\|^2} = 0$. Using the necessary condition $\nabla f(x^*) = 0$, a similar argument as above shows that at a local minimum $x^*$ it must hold that

$$d' H f(x^*) d \geq 0 \text{ for all } d \in \mathbb{R}^n$$

i.e. the Hessian matrix $Hf(x^*)$ should be positive semi-definite.

The following proposition summarizes our discussion.

**Proposition 1.3.** *Let $x^*$ be an unconstrained local minimum of $f : \mathbb{R}^n \to \mathbb{R}$ and assume that $f$ is $C^1$ in an open set $S$ with $x^* \in S$. Then $\nabla f(x^*) = 0$. (First Order Necessary Condition or FONC).*

*If moreover $f$ is in $C^2$ in $S$ one has $Hf(x^*)$ is positive semi-definite (Second Order Necessary Condition or SONC).*

**Definition** 1.4. A point $x^* \in \mathbb{R}^n$ with $\nabla f(x^*) = 0$ is called *stationary point* of $f$.

**Remark 1.5.** The necessary conditions do not guarantee local optimality; consider for instance $f(x) = x^3$ and $x^* = 0$.

**Sufficient conditions for a local minimum.**

**Proposition 1.6.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable in an open set $S \subset \mathbb{R}^n$. Suppose that $x^* \in S$ satisfies the conditions*

$$\nabla f(x^*) = 0, \ Hf(x^*) \text{ strictly positive definite} \tag{1.5}$$

*Then $x^*$ is a strict local minimum. In particular, there exists $\gamma > 0$, $\epsilon > 0$ such that*

$$f(x) \geq f(x^*) + \frac{\gamma}{2}\|x - x^*\|^2 \text{ for all } x \text{ with } \|x - x^*\| < \epsilon.$$

*Proof.* Since $Hf(x^*)$ is positive definite, there exists some $\lambda > 0$ such that $y'Hf(x^*)y > \lambda \|y\|$ for all $y \in \mathbb{R}^n$ ($\lambda$ is the smallest eigenvalue of $Hf$). Using $\nabla f(x^*) = 0$, a second order Taylor expansion of $f$ around $x^*$ gives for $y \in \mathbb{R}^n$ with $\|y\|$ sufficiently smalland

$$\begin{aligned} f(x^* + y) - f(x^*) &= \nabla f(x^*)'y + \frac{1}{2} y' Hf(x^*)y + R_2(x^*, hy) \\ &\geq \frac{1}{2}\lambda \|y\|^2 + R_2(x^*, y) \\ &= \|y\|^2 \left( \frac{1}{2}\lambda + \frac{R_2(x^*, y)}{\|y\|^2} \right) . \end{aligned}$$

Now the term in the bracket converges to $\frac{1}{2}\lambda$ as $\|y\| \to 0$. Hence we may take $\gamma = \frac{\lambda}{2}$ and $\epsilon$ small enough so that $\frac{R_2(x^*,y)}{\|y\|^2} < \frac{1}{4}\lambda$. □

**Remark 1.7.** Of course, $x^*$ need not be a global minimum. (Picture)

**The case of convex functions.** We begin by defining certain fundamental notions related to convexity.

**Definition** 1.8. i) A set $X \subset \mathbb{R}^n$ is convex if $\forall x_1, x_2 \in X, \lambda \in [0, 1]$ the convex combination $\lambda x_1 + (1 - \lambda)x_2$ belongs to $X$.

ii) A function $f : X \subset \mathbb{R}^n \to \mathbb{R}$ ($X$ convex) is called convex if $\forall x_1, x_2 \in X, \lambda \in [0, 1]$

$$f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2); \tag{1.6}$$

f is strict convex if the inequality is strict for $\lambda \in (0, 1)$.

iii) $f : X \subset \mathbb{R}^n \to \mathbb{R}$ is concave $\Leftrightarrow -f$ is (strict) convex $\Leftrightarrow \geq$ holds in (1.6). Strict concavity is defined in the same way.

The following lemma gives a characterization of convexity for $C^1$ functions.

**Lemma 1.9.** *Consider an open convex set $X \subset \mathbb{R}^n$. A $C^1$ function $f : X \to \mathbb{R}^n$ is convex on the if and only if it holds for all $x, z \in X$ that*

$$f(z) \geq f(x) + \nabla f(x)'(z - x).$$

*If $f$ is $C^2$ a necessary and sufficient condition for the convexity of $f$ on $X$ is the condition that $Hf(x)$ is positive semi-definite for all $x \in X$.*

For the proof we refer to Appendix B of Bertsekas (1999).

**Proposition 1.10.** *Let $f : X \to \mathbb{R}$ be a convex function on some convex set $X \subset \mathbb{R}^n$. Then the following holds.*

*i) A local minimum of $f$ over $X$ is also a global minimum. If $f$ is strictly convex, there exists at most one global minimum.*

*ii) If $X$ is open, the condition $\nabla f(x^*) = 0$ is a necessary and sufficient condition for $x^* \in X$ to be a global minimum of $f$.*

*Proof.* i) Suppose that $x^*$ is a local minimum and that there is some $\tilde{x} \in X$ with $f(\tilde{x}) < f(x^*)$. Then we have for $0 < \lambda \leq 1$

$$f(x^* + \lambda(\tilde{x} - x^*)) \leq (1-\lambda)f(x^*) + \lambda f(\tilde{x}) < f(x^*). \tag{1.7}$$

For $\lambda \to 0$ we obtain a contradiction to the fact that $x^*$ is a local minimum (the second part of i) is established in the same way).

ii) If $\nabla f(x^*) = 0$ Lemma 1.9 gives $f(z) \geq f(x^*)$ for all $z \in X$ and the result follows. □

**Example 1.11** (Quadratic cost functions)**.** Let $f(x) = \frac{1}{2}x'Qx - b'x$, $x \in \mathbb{R}^n$ for a symmetric $n \times n$ matrix $Q$ and some $b \in \mathbb{R}^n$. Then we have

$$\nabla f(x) = Qx - b \text{ and } Hf(x) = Q.$$

a) Local minima. According to Proposition 1.3 for $x^*$ to be a local minimum we must have

$$\nabla f(x^*) = Qx^* - b = 0,\ Hf(x^*) = Q \text{ positive semi-definite;}$$

hence if $Q$ is not positive semi-definite, $f$ has no local minima.

b) If $Q$ is positive semi-definite, $f$ is convex. In that case we need not distinguish global and local minima, and $f$ has a global minimum if and only if there is some $x^*$ with $Qx^* = b$. c) If $Q$ is positive definite, $Q^{-1}$ exists and the unique global minimum is attained at $x^* = Q^{-1}b$.

### Existence results for a global minimum

**Proposition 1.12** (Weierstrass' Theorem)**.** *Let $X \subset \mathbb{R}^n$ be non-empty and suppose that $f : X \to \mathbb{R}$ is continuous in $X$. Suppose moreover, that one of the following three conditions holds*

*(1) $X$ is compact (closed and bounded).*

*(2) $X$ is closed and $f$ is coercive, that is*

$$\forall (x^k)_{k\in\mathbb{N}} \in X \text{ with } \left\|x^k\right\| \to \infty \text{ one has } \lim_{k\to\infty} f(x^k) = \infty$$

*(3) There is some $\gamma \in \mathbb{R}$ such that the level set $\{x \in X : f(x) \leq \gamma\}$ is non-empty and compact.*

*Then $f$ has at least one global minimum and the set of all global minima is compact.*

**Remark 1.13.** The result holds more generally for functions that are lower semicontinous, where a function $f$ is called lower semicontinuous if for all $x \in X$, all $(x^k)_{k\in\mathbb{N}}$ with $x^k \to x$ it holds that

$$\liminf_{k\to\infty} f(x^k) \geq f(x).$$

**Example 1.14.** Picture where $f$ is not lower semicontinous, and the minimum is not obtained.

**Remark 1.15.** The FONC $\nabla f(x^*) = 0$ can be solved explicitly only in exceptional cases. The main practical use of this condition is in the design of algorithms that converge to a local minimum as in Section 1.2 or in sensitivity analysis.

## 1.2 Numerical Solution via Gradient methods

Next we explain the basic principles behind several numerical methods for finding a stationary point or more generally a local minimum.

**Basic Idea.** Most numerical algorithms for unconstrained minimization problems rely on the idea of *iterative descent*: one starts from an initial guess $x^0$ and generates vectors $x^1, x^2, \dots$ with $f(x_{k+1}) < f(x^k)$, $k = 0, 1, \dots$. These points are typically chosen by a rule of the form

$$x^{k+1} = x^k + \alpha^k d^k, \ \alpha \geq 0, d \in \mathbb{R}^n; \tag{1.8}$$

$\alpha$ is called *step size*, $d$ is the *descent direction.* For differentiable $f$ the choice of $d$ is motivated by following idea: $\nabla f(x^k)$ points to the direction of the strongest increase of $f$; hence if $d$ points to the opposite direction, $f$ should be decreased by moving in direction $d$. 'Pointing to opposite direction' is formalized by the so-called *descent condition*

$$d' \nabla f(x^k) < 0 \tag{1.9}$$

The descent condition does indeed lead to a sequence $x^1, x^2, \dots$ with $f(x_{k+1}) < f(x^k)$, as the following lemma shows.

**Lemma 1.16.** *Suppose that $f$ is $C^1$ and that $\nabla f(x)'d < 0$. Let $x_\alpha = x + \alpha d, \alpha > 0$. Then $f(x_\alpha) \leq f(x)$ for $\alpha$ sufficiently small.*

*Proof.* Taylor expansion gives

$$f(x_\alpha) = f(x) + \alpha \nabla f(x)'d + R(x, \alpha d).$$

Since $\lim_{\alpha \to 0} \frac{R(x,\alpha d)}{\alpha} = 0$, the negative term $\alpha \nabla f(x)'d$ dominates the term $R(x, \alpha d)$ for $\alpha$ sufficiently small so that $f(x_\alpha) < f(x)$. □

### 1.2.1 Choosing the descent direction

A rich set of vectors $d^k$ that satisfy the descent condition $\nabla f(x^k)'d^k < 0$ can be constructed by taking $d^k$ of the form

$$d^k = -D^k \nabla f(x^k) \tag{1.10}$$

for some symmetric, positive definite matrix $D^k$. In fact, as $D^k$ is positive definite,

$$\nabla f(x^k)'d^k = -\nabla f(x^k)' D^k \nabla f(x^k) < 0.$$

Methods where $d^k$ is of the form (1.10) are known as gradient methods.

**a) Steepest descent.** Here $D = I$ and hence $d^k = -\nabla f(x^k)$. This is the simplest choice for $d$, but the convergence to a stationary point can be slow (picture of a "long valley"). The name 'steepest descent method' stems from the observation that among all descent directions $d$ with $\|d\| = 1$ the choice $d = -\nabla f(x^k)/\left\|\nabla f(x^k)\right\|$ minimizes the slope of the map $\alpha \mapsto f(x^k + \alpha d)$ at $\alpha = 0$ (locally steepest descent). In fact, one has

$$\frac{d}{d\alpha} f(x^k + \alpha d)_{|\alpha=0} = \nabla f(x^k)' d\,.$$

Moreover, since $\|d\| = 1$ we get from the Cauchy-Schwarz inequality that

$$\left|\nabla f(x^k)' d\right| \le \left\|\nabla f(x^k)\right\| \|d\| = \left\|\nabla f(x^k)\right\|.$$

Hence we get for $d$ with $\|d\| = 1$ and $\nabla f(x^k)' d < 0$ that

$$\nabla f(x^k)' d = -\left|\nabla f(x^k)' d\right| \ge -\left\|\nabla f(x^k)\right\|;$$

equality holds for for $d = -\nabla f(x^k)/\left\|\nabla f(x^k)\right\|$, so that this choice is in fact the direction of the locally steepest descent.

**b) Newton's method.** Here one takes $D^k = (Hf(x^k))^{-1}$, provided that this matrix is positive definite. The idea underlying this choice is to minimize at each iteration the quadratic approximation of $f$ around $x^k$ given by

$$f^k(x) = f(x^k) + \nabla f(x^k)'(x - x^k) + \frac{1}{2}(x - x^k)' Hf(x^k)(x - x^k).$$

If $Hf(x^k)$ is positive definite, the minimum of $f^k(x)$ is characterized by the FOC

$$\nabla f(x^k) + Hf(x^k)(x - x^k) = 0,$$

which gives $x^{k+1} = x^k - (Hf(x^k))^{-1}\nabla f(x^k)$ and hence the choice $D^k = (Hf(x^k))^{-1}$. If it is applicable, the method usually converges very fast to a local minimum.

**c) Gauss-Newton method.** This method is designed for calibration problems of the form

$$\min_{x \in \mathbb{R}^n} f(x) := \frac{1}{2}\|g(x)\|^2 = \frac{1}{2}\sum_{i=1}^m (g_i(x))^2, \tag{1.11}$$

where $g_1, \ldots, g_m$ are functions from $\mathbb{R}^n$ to $\mathbb{R}$. In the Gauss-Newton method one chooses $D^k = (\nabla g(x^k)\nabla g(x^k)')^{-1}$, provided that the matrix $\nabla g(x^k)\nabla g(x^k)'$ is invertible. This is the case if $\nabla g(x)$ has rank $n$.[1] Since $\nabla f(x^k) = \nabla g(x^k) g(x^k)$, the Gauss-Newton rule gives

$$x^{k+1} = x^k - \alpha^k(\nabla g(x^k)(\nabla g(x^k))')^{-1}\nabla g(x^k) g(x^k). \tag{1.12}$$

It can be shown that this is an approximation to the Newton rule.

[1] Loosely speaking, this is the case if there are more observations than parameters, that is for $m \ge n$.

### 1.2.2 Choosing the stepsize

Recall that in a gradient method one has $x^{k+1} = x^k - \alpha^k D^k \nabla f(x^k)$. In the literature there are a number of rules for the choice of $\alpha$:

- (Limited) minimization rule. Here one takes $\alpha^k$ so that the function $\alpha \mapsto f(x^k + \alpha d^k)$ is minimized over $\alpha \geq 0$ or $\alpha \in [0, s]$, i.e.

  $$f(x^k + \alpha^k d^k) = \min_{\alpha \in [0,s]} f(x^k + \alpha d^k)$$

  (often $\alpha^k$ has to be found numerically).
- constant stepsize,
- many more methods.

### 1.2.3 Convergence Issues: an Overview

We begin with a brief overview of the asymptotic behavior of sequences of the form $x^{k+1} = x^k - \alpha^k D^k \nabla f(x^k)$.

- If the sequence $(x^k)_{k \in \mathbb{N}}$ converges, the limit is usually a stationary point, but in general not more.
- The sequence $(x^k)_{k \in \mathbb{N}}$ need not have a limit point; in fact, $(x^k)_{k \in \mathbb{N}}$ is typically divergent, if f has no local minimum.
- Local minima that are isolated points tend to attract sequences $(x^k)_{k \in \mathbb{N}}$ that start sufficiently close to the local minimum. (This is known as capture theorem, see for instance Bertsekas (1999), Proposition 1.2.5 ).

Finally we quote a precise result concerning the convergence to a stationary point.

**Proposition 1.17.** *Suppose that for all $x \in \mathbb{R}^n$ the eigenvalues of $D^k$ are bounded away from zero and infinity, that is $\exists\, 0 < C_1 \leq C_2 < \infty$ such that*

$$C_1 < \lambda_{\min}(D^k(x)) \leq \lambda_{\max}(D^k(x)) < C_2 < \infty, \ \forall k \in \mathbb{N}, x \in \mathbb{R}^n$$

*Consider the sequence $x^{k+1} = x^k - \alpha^k D^k \nabla f(x^k)$. If $(\alpha^k)_k$ is constructed by the (limited) minimization rule, every limit point of $(x^k)_{k \in \mathbb{N}}$ is a stationary point.*

# Chapter 2

# Lagrange Multipliers via Calculus

## 2.1 Constrained Optimization Problems: Overview

From now on we are interested in constrained optimization problems of the form

$$\min_{x\in X} f(x) \text{ subject to } h_i(x)=0, 1\le i\le m, \quad g_j(x)\le 0, 1\le j\le r. \tag{2.1}$$

Here $X\subset\mathbb{R}^n$ and $f$, $h_i$, $1\le i\le m$, and $g_j$, $1\le j\le r$ are functions from $X$ to $\mathbb{R}$. To simplify the notation we put $h(x)=(h_1(x),\dots,h_m(x))'\in\mathbb{R}^m$ and $g(x)=(g_1(x),\dots,g_r(x))'\in\mathbb{R}^r$, so that (2.1) can be written more succinctly in the form

$$\min_{x\in X} f(x) \text{ subject to } h(x)=0,\ g(x)\le 0. \tag{2.2}$$

In this formulation some of the constraints that define the set of admissible points $\tilde X=\{x\in X\colon h(x)=0,\ g(x)\le 0\}$ have been made explicit. This is often helpful to solve (2.1) for the following reason: we will show in the sequel that under certain conditions of $f$, $h$ and $g$ that there are constants $\lambda^*\in\mathbb{R}^m$, $\mu^*\in\mathbb{R}^r, \mu^*\ge 0$ such that a (local) optimum $x^*$ for problem 2.1 is also a (local) optimum for the unconstrained optimization problem

$$\min_{x\in X} L(x,\lambda^*,\mu^*) := f(x)+\sum_{i=1}^m \lambda_i^* h_i(x)+\sum_{j=1}^r \mu_j^* g_j(x).$$

The function $L$ is called *Lagrange function*, $\lambda^*$ and $\mu^*$ are known as *(Lagrange) multipliers.* Results on the existence of Lagrange multipliers are useful, as they permit the use of concepts and algorithms from unconstrained optimization for the solution of constrained optimization problems; moreover, they frequently lead to interesting statements on the structure of constrained optima.

Basically there are two approaches to establish the existence of Lagrange multipliers.

a) *An approach based on calculus.* Here it is assumed that the functions $f$,$g$ and $h$ are differentiable, and the focus lies on the derivation of necessary conditions for local optima. This theory is the focus of Chapter 2.

b) *Duality theory* for *convex* problems. Here it is assumed that $f$,$g$ and $h$ are convex, the focus is on global optima, and the essential tools come from convex analysis. An introduction to this approach is given in Chapter 3

## 2.2 Lagrange multipliers with equality constraints

In this section we consider the problem

$$\min_{x\in\mathbb{R}^n} f(x) \text{ subject to } h(x) = 0 \tag{2.3}$$

where it is assumed that $f$ and $h = (h_1, \dots, h_m)$ are $C^1$-functions on $\mathbb{R}^n$. Note that all subsequent results hold also for the case where $dom\, f \cap dom\, h$ is an open subset of $\mathbb{R}^n$ that contains a local minimum $x^*$ of 2.3. [1]

**Proposition 2.1** (Existence of Lagrange multipliers)**.** *Let $x^*$ be a local minimum for Problem 2.3, and suppose that the gradients of the constraint functions $\nabla h_1(x^*), \dots, \nabla h_m(x^*)$ are linearly independent. Then there exists a unique vector $\lambda^* \in \mathbb{R}^m$ such that*

$$\nabla f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla h_i(x^*) = 0. \tag{2.4}$$

*If $f$ and $h$ are twice continuously differentiable, one has*

$$y'\Big(Hf(x^*) + \sum_{i=1}^m \lambda_i^* Hh_i(x^*)\Big)y \geq 0 \;\; \textit{for all} \;\; y \in V(x^*). \tag{2.5}$$

*Here $V(x^*)$ is the subspace of* first order feasible variations*, i.e.*

$$V(x^*) = \{y \in \mathbb{R}^n : \nabla h_i(x^*)'y = 0, \;\; \textit{for all } 1 \leq i \leq m\} \tag{2.6}$$

In the sequel a point $x^* \in \mathbb{R}^n$ such that $\nabla h_1(x^*), \dots, \nabla h_m(x^*)$ are linearly independent will be called *regular* (note that this is possible only for $m \leq n$).

**Remark 2.2.** i) $V(x^*)$ is the subspace of variations $\Delta x = (x - x^*)$ for which the constraint $h(x) = 0$ holds 'up to first order'. Condition 2.4 states that the gradient $\nabla f(x^*)$ of the cost function is orthogonal to all these 'locally permissible' variations: for $y \in V(x^*)$ is holds that:

$$\nabla f(x^*)'y = -\sum_{i=1}^m \lambda_i^* \nabla h_i(x^*)'y = 0.$$

This condition is analogous to the condition $\nabla f(x^*) = 0$ in unconstrained optimization.

ii) Using the Lagrange function $L(x, \lambda^*) = f(x) + \sum_{i=1}^m \lambda_i^* h_i(x)$ we may write (2.4) as $\frac{\partial}{\partial x_i} L(x^*, \lambda^*) = 0$, $1 \leq i \leq n$.

**Example 2.3** (geometric interpretation of the condition 2.4)**.** Consider the problem

$$\min_{x\in\mathbb{R}^2} x_1 + x_2 \text{ subject to } x_1^2 + x_2^2 = 2.$$

Obviously, the minimum is attained in $x^* = (-1, -1)$. One has $\nabla f(x^*) = (1, 1)'$; $\nabla h(x^*) = (2x_1^*, 2x_2^*)' = (-2, -2)'$. These vectors are collinear, so that $\nabla f(x^*)$ is orthogonal to $V(x^*)$, as $V(x^*)$ is given by the orthogonal complement of $(-2, -2)$.

[1] $x^*$ is a local minimum of 2.3 if there is some $\epsilon > 0$ such that $f(x^*) \leq f(x^*)$ for all $x \in B_\epsilon(x^*)$ such that $h(x) = 0$.

**Example 2.4** (non-existence of multipliers at an irregular point.)**.** Consider the problem $\min x_1 + x_2$ subject to the equality constraints

$$h_1(x) = (x_1 - 1)^2 + x_2^2 - 1 = 0 \text{ (sphere around the point } (1,0)' \text{ with radius 1)}$$
$$h_2(x) = (x_1 - 2)^2 + x_2^2 - 4 = 0 \text{ (sphere around the point } (2,0)' \text{ with radius 2).}$$

Then $h(x) = 0 \Leftrightarrow x = x^* = (0,0)$, i.e. the constrained set consists of a single point which is then automatically a (local) minimum. One has $\nabla h_1(x^*) = (-2,0)$; $\nabla h_2(x^*) = (-4,0)$. These vectors are collinear, so $x^*$ is not regular; on the other hand $\nabla f(x^*) = (1,1)$ is not a linear combination of $\nabla h_1(x^*)$ and $\nabla h_2(x^*)$, i.e. $x^*$ is a local minimum where no multipliers exist. The example shows that the assumption that $x^*$ is regular is really needed for the existence of multipliers.

*Proof of Proposition 2.1.* We use a penalty function approach, that is we consider an unconstrained problem where there is a large penalty for violating the constraint $h(x) = 0$. As the penalty $k$ gets larger, the local minimum $x^k$ of the unconstrained problem converges to the local minimum $x^*$, and the Lagrange multipliers can be constructed from the FONCs for the unconstrained problem. The idea carries over over to problems with inequality constraints and it can be used for the design of numerical algorithms.

a) *The penalty function.* We let for $k \in \mathbb{N}$, $\alpha > \sigma$.

$$F^k(x) = f(x) + \frac{k}{2}\|h(x)\|^2 + \frac{\alpha}{2}\|x - x^*\|^2, \tag{2.7}$$

where $x^*$ is the local minimum of the original problem. The term $k/2\,\|h(x)\|^2$ is the penalty term; the term $\alpha/2\,\|x - x^*\|^2$ is needed for technical reasons.

Choose now $\epsilon > 0$ sufficiently small so that $f(x) \ge f(x^*)$ for all $x$ with $h(x) = 0$ and $\|x - x^*\| < \epsilon$ and denote by $\bar{S}$ the closed ball

$$\bar{S} = \{x \in \mathbb{R}^n : \|x - x^*\| \le \epsilon\}.$$

Denote by $x^k$ a global minimum of $F^k$ on $\bar{S}$; $x^k$ exists according to the Weierstrass theorem (Proposition 1.12). We now show that $x^k \to x^*$. It holds for $k \in \mathbb{N}$:

$$F^k(x^k) = f(x^k) + \frac{k}{2}\left\|h(x^k)\right\|^2 + \frac{\alpha}{2}\left\|x^k - x^*\right\|^2 \le F^k(x^*) = f(x^*) \tag{2.8}$$

Since $f$ is bounded on $\bar{S}$ we conclude that $\lim_{k\to\infty}\left\|h(x^k)\right\|^2 = 0$. Hence every limit point $\bar{x}$ of $(x^k)_{k\in\mathbb{N}}$ satisfies $h(\bar{x}) = 0$. Moreover, by (2.8) we get $f(x^k) + \frac{\alpha}{2}\left\|x^k - x^*\right\|^2 \le f(x^*)$ and, by passing to the limit,

$$f(\bar{x}) + \frac{\alpha}{2}\|\bar{x} - x^*\|^2 \le f(x^*).$$

On the other hand, since $\bar{x} \in \bar{S}$ and $h(\bar{x}) = 0$, the local optimality of $x^*$ on $\bar{S}$ implies that $f(x^*) \le f(\bar{x})$. Combining these estimates we get

$$f(\bar{x}) \le f(\bar{x}) + \frac{\alpha}{2}\|\bar{x} - x^*\|^2 \le f(x^*) \le f(\bar{x}),$$

which is possible only if $\|\bar{x} - x^*\|^2 = 0$, i.e. for $\bar{x} = x^*$.

b) *Construction of the multipliers.* Since $x^k \to x^*$ we can assume that $x^k \in int(S)$ for $k$ sufficiently large. Since $x^k$ is moreover a local minimum of $F^k$ it satisfies the FONCs

$$0 = \nabla F^k(x^k) = \nabla f(x^k) + k\,\nabla h(x^k)\,h(x^k) + \alpha(x^k - x^*). \tag{2.9}$$

Since $x^*$ is regular, $\nabla h(x^*)$ has rank $m$, and it follows that $\nabla h(x^k)$ has rank $m$ for $k$ sufficiently large. For such $k$ the $m \times m$ matrix $\nabla h(x^k)'\,\nabla h(x^k)$ is invertible, as it is positive definite: one has

$$x'\,\nabla h(x^k)'\,\nabla h(x^k)\,x = \left\|\nabla h(x^k)x\right\|^2 > 0 \text{ for } x \in \mathbb{R}^m, x \neq 0.$$

Pre-multiplying the FONCs with the so-called generalized inverse $(\nabla h'\nabla h)^{-1}\nabla h(x^k)$ we get

$$kh(x^k) = -(\nabla h'\nabla h)^{-1}\nabla h'(x^k)\{\nabla f(x^k) - \alpha(x^k - x^*)\}. \tag{2.10}$$

Taking the limit $k \to \infty$ (and hence $x^k \to x^*$) we get

$$\lim_{k\to\infty} kh(x^k) = -(\nabla h'\nabla h)^{-1}\nabla h(x^*)\;\nabla f(x^*) =: \lambda^*. \tag{2.11}$$

Taking the limit $k \to \infty$ in the FONCs (2.9) we get

$$0 = \nabla f(x^*) + \nabla h(x^k)\lambda^*,$$

so that $\lambda^*$ is the desired Lagrange multiplier. The second claim (the second-order-necessary conditions) are derived from the second-order conditions for unconstrained optimization problems applied to $F^k$; we omit the details.

□

**Example 2.5** (Arithmetic-geometric-mean inequality)**.** We want to show that for $x_1, \ldots, x_n \geq 0$ the following inequality holds.

$$(x_1 \cdot x_2 \cdot \ldots \cdot x_n)^{\frac{1}{n}} \leq \frac{1}{n}(x_1 + \ldots + x_n). \tag{2.12}$$

The lhs is the geometric mean of $x_1, \ldots, x_n$; the rhs is the arithmetic mean of $x_1, \ldots, x_n$, so that the inequality is known as *arithmetic-geometric-mean inequality.*

With the substitution $y_i := \ln x_i$ (2.12) translates to the inequality

$$\exp\left(\frac{1}{n}\sum_{i=1}^n y_i\right) \leq \frac{1}{n}\sum_{i=1}^n e^{y_i} \text{ for all } y_1, \ldots, y_n \in \mathbb{R}^n. \tag{2.13}$$

In order to prove this inequality we minimize for arbitrary $s \in \mathbb{R}$ the function $f(y) = e^{y_1} + \ldots + e^{y_n}$ over all $y \in \mathbb{R}^n$ with $\sum_{i=1}^n y_i = s$ and show that the minimum is no less than $ne^{s/n}$. We thus have to consider for $s > 0$ fixed the following optimization problem with equality constraints

$$\min_{y\in\mathbb{R}^n} e^{y_1} + \ldots + e^{y_n} \text{ subject to } \sum_{i=1}^n y_i = s \tag{2.14}$$

The Lagrange function is $L(y, \lambda) = \sum_{i=1}^m e^{y_i} + \lambda\,(s - \sum_{i=1}^m y_i)$. The first order necessary conditions are

$$0 = e_i^y - \lambda, \quad 1 \leq i \leq m$$

It follows that $y_i = \ln \lambda$ for all $i$, in particular the $y_i$ are all equal. Using the constraint $\sum_{i=1}^n y_i = s$ we thus get that the only point that satisfies the FONCs is $y_i^* = s/n$, $1 \le i \le n$. It is easily seen that $y^*$ is in fact a global minimum for (2.14). Since $s$ was arbitrary this gives the inequality (2.13) and hence the claim.

**Example 2.6** (Markowitz portfolio optimization)**.** Consider a one-period financial market model with $n$ risky assets with price $(S_{t,i})$, $1 \le i \le n$, $t = 0, T$. The return of asset $i$ over $[0, T]$ is

$$R_i = \frac{S_{T,i} - S_{0,i}}{S_{0,i}} \approx \ln S_{T,i} - \ln S_{0,i} \tag{2.15}$$

Consider now an investor with initial wealth $V_0$. Denote by $\theta_i$ the number of units of asset $i$ in his portfolio and by $\pi_i = \theta_i S_{0,i}/V_0$ the weight of asset $i$. Note that by definition $\sum_{i=1}^n \pi_i = 1$. The change in the portfolio value is

$$V_T - V_0 = \sum_{i=1}^n \theta_i (S_{T,i} - S_{0,i}) = V_0 \sum_{i=1}^n \pi_i R_i\,; \tag{2.16}$$

the return on the portfolio is thus

$$(V_T - V_0)/V_0 = \sum_{i=1}^n \pi_i R_i = \pi' R.$$

Assume now that the investor evaluates portfolios only according to mean and variance of the return, preferring higher mean to lower mean and lower variance to higher variance (all else equal). This can be justified for instance by assuming that the rvs $(R_1, \dots, R_m)$ are multivariate normal, since mean and variance fully characterize the return distribution in that case. In order to determine optimal portfolios we first solve the following problem.

(2.17) Minimize the portfolio variance over all portfolios with a given mean $m$

Denote by $\mu = (\mu_1, \dots, \mu_n)'$ the mean vector of $R = (R_1, \dots, R_m)$ and by $Q$ with $q_{ij} = \mathrm{cov}(R_i, R_j)$ the covariance matrix. Since the portfolio $X^\pi$ that corresponds to given portfolio weights $(\pi_1, \dots, \pi_n)$ has expected return $\mathbb{E}(X^\pi) = \pi' \mu$ and variance

$$\mathrm{var}(X^\pi) = \mathrm{var}(\sum_{i=1}^n \pi_i R_i) = \pi' Q \pi,$$

in mathematical terms the problem (2.17) can be written as

$$\min_{\pi \in \mathbb{R}^n} \pi' Q \pi \;\text{ subject to } \pi' \mathbf{1} = 1, \pi' \mu = m, \tag{2.18}$$

where $\mathbf{1} = (1, \dots, 1)' \in \mathbb{R}^n$. In order to solve this problem we make the following assumptions:

A1) The matrix $Q$ is positive definite, i.e. the assets $S_1, \dots, S_n$ do not generate/contain the riskless asset.

A2) The vectors $\mathbf{1}$ and $\mu$ are linearly independent, that is not all risky assets have the same expected return.

Assumption A1) will be relaxed below, when we discuss the optimal investment with a traded riskless asset. We want to use Lagrange multiplier theory to study how the solution of (2.18) varies as we vary the target return $m$. Denote by $\lambda_1$ and $\lambda_2$ the multipliers corresponding to the equality constraints in (2.18). The Lagrange function is

$$L(\pi; \lambda_1, \lambda_2) = \pi' Q \pi + \lambda_1(\mathbf{1}'\pi - 1) + \lambda_2(\mu'\pi - m). \tag{2.19}$$

A2) implies in particular that every point $\pi$ is regular. Hence at a local optimum $\pi^*$ there will be multipliers $\lambda_1^*$, $\lambda_2^*$ such that $\nabla_\pi L(\pi; \lambda_1^*, \lambda_2^*) = 0$. This gives by differentiating (2.19) that $0 = 2Q\pi + \lambda_1^*\mathbf{1} + \lambda_2^*\mu$. Since $Q$ is invertible by assumption, we get

$$\pi^* = -\frac{1}{2}\lambda_1^* Q^{-1}\mathbf{1} - \frac{1}{2}\lambda_2^* Q^{-1}\mu. \tag{2.20}$$

In order to determine $\lambda_1^*$ and $\lambda_2^*$ we use the equations $\mathbf{1}'\pi^* = 1, \mu'\pi^* = m$; substituting (2.20) gives the following linear $2 \times 2$ equation system for $\lambda_1^*$, $\lambda_2^*$.

$$\begin{aligned} 1 = \mathbf{1}'\pi^* &= -\frac{1}{2}\lambda_1^*\mathbf{1}'Q^{-1}\mathbf{1} - \frac{1}{2}\lambda_2^*\mathbf{1}'Q^{-1}\mu, \\ m = \mu'\pi^* &= -\frac{1}{2}\lambda_1^*\mu'Q^{-1}\mathbf{1} - \frac{1}{2}\lambda_2^*\mu'Q^{-1}\mu. \end{aligned} \tag{2.21}$$

Since $Q^{-1}$ is positive definite (as $Q$ is positive definite) the matrix of the coefficients in (2.21) has full rank. Hence we get the representation

$$\lambda_1^* = a_1 + b_1 m \ \text{ and } \lambda_2^* = a_2 + b_2 m, \tag{2.22}$$

where the matrix $\begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix}$ is the inverse of the coefficient matrix in the linear system (2.21). Substituting back into (2.20) gives

$$\pi^* = m\left(-\frac{1}{2}b_1 Q^{-1}\mathbf{1} - \frac{1}{2}b_2 Q^{-1}\mu\right) - \frac{1}{2}a_1 Q^{-1}\mathbf{1} - \frac{1}{2}a_2 Q^{-1}\mu =: mv + w,$$

for $v, w \in \mathbb{R}^n \backslash \{0\}$, independent of $m$. Hence the variance of the optimal (variance minimizing) portfolio is

$$\operatorname{var}(X^{\pi^*}) = (mv + w)'Q(mv + w) =: \sigma^2(m),$$

i.e. $\sigma^2(m)$ is the minimal variance of a portfolio with expected return $m$. We may write

$$\sigma^2(m) = m^2 \underbrace{v'Qv}_{=:a} + 2m \underbrace{v'Qw}_{=:b} + \underbrace{w'Qw}_{=:c},$$

so that $\sigma^2(m) = am^2 + bm + c$ with $a > 0$ (a parabola in $m$). It is more common to plot the relation between standard deviation and $m$, and we get

$$\sigma(m) = \sqrt{am^2 + bm + c}$$

In the following graph (only on the blackboard) we plot $\sigma^2(m)$ and $\sigma(m)$ as a function of $m$; the return $m^*$ and the value $\sigma^*$ correspond to the risky portfolio with the smallest variance. It can be shown that $\sigma(m)$ is convex with minimum at $m^*$. Of course, for an investor only the portfolios with mean $m \geq m^*$ (the right branch of the parabola) are of interest.

Finally we consider portfolios $\tilde{\pi} = (\pi_0, \pi_1, \ldots, \pi_n) \in \mathbb{R}^{n+1}$, consisting of a risk free asset with return $r$ and variance zero (asset 0) and the $n$ risky assets considered previously. Now $\tilde{\pi}$ has to satisfy the constraint $\sum_{i=0}^{n} \pi_i = 1$. Put $\pi = (1-\pi_0)^{-1}(0, \pi_1, \ldots, \pi_n)'$, so that $\tilde{\pi}$ consists of $\pi_0$ units of the risk-free asset and $(1-\pi_0)$ units of the portfolio $\pi$. One obviously has

$$\mathbb{E}\left(X^{\tilde{\pi}}\right) = \pi_0 r + \sum_{j=1}^{n} \pi_j R_j = \pi_0 r + (1-\pi_0)\mathbb{E}\left(X^{\pi}\right)$$

$$\operatorname{var}(X^{\tilde{\pi}}) = \operatorname{var}\left((1-\pi_0)\sum_{j=1}^{n} \frac{\pi_j}{(1-\pi_0)} R_j\right) = \left(1-\pi_0)^2 \operatorname{var}(X^{\pi}\right)$$

and the standard deviation satisfies

$$\sigma(X^{\tilde{\pi}}) = |1-\pi_0|\,\sigma(X^{\pi}).$$

If we plot the set of all attainable $\mu, \sigma$) pairs we get the following picture (only on the blackboard). It is clear from the picture that optimal portfolios are on the red line (the efficient frontier) that 'touches' the upper boundary of the $(m, \sigma)$ pairs. Therefore every investor will choose a portfolio that is a combination of the risk-free asset and $\tilde{\pi}^*$. This result is known as two-funds theorem; it is a central part of the derivation of the CAPM.

## 2.3 Necessary conditions for local optima with inequality constraints

Next we consider optimization problems with equality *and* inequality constraints, which are of the form

$$\min f(x) \ \text{ subject to } h(x) = 0, g(x) \le 0, \tag{2.23}$$

for $C^1$-functions $f$, $h_1, \ldots, h_m$ and $g_1, \ldots, g_r$. We want to derive conditions for the existence of multipliers at a local minimum of (2.23) using our result for the case of pure equality constraints. For this we use:

**Definition** 2.7. The constraint $g_j(x) \le 0$ is called *active* or *binding* at a feasible point $x$ if $g_j(x) = 0$; the set $A(x) = \{1 \le j \le r : g_j(x) = 0\}$ is the set of active constraints at the point $x$.

Now if $x^*$ is a local minimum for (2.23), it is also a local minimum for the following problem with equality constraints:

$$\min f(x) \ \text{ subject to } \ h(x) = 0, g_j(x) = 0 \ \text{ for all } \ j \in A(x^*). \tag{2.24}$$

If $x^*$ is regular for (2.24), by Proposition 2.1 there exist multipliers $\lambda_1^*, \ldots, \lambda_m^*$, $\mu_j^*$, for $j \in A(x^*)$ such that

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_j^* \nabla h_j(x^*) + \sum_{j \in A(x^*)} \mu_j^* g_j(x^*) = 0.$$

If we put in addition $\mu_j = 0$ for $j \in \{1, \dots, r\} \setminus A(x^*)$ we get the existence of multipliers $\lambda^* \in \mathbb{R}^n$, $\mu^* \in \mathbb{R}^r$ such that

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_j^* \nabla h_j(x^*) + \sum_{j=1}^{r} \mu_j^* g_j(x^*) = 0.$$

This can be be viewed as an analogue to Proposition 2.1.

In the case with inequality constraints there are additional restrictions on the sign of the multipliers of the inequality constraints: it has to hold that $\mu_j^* \geq 0$ for all $j \in A(x)$. This is illustrated in the following example.

**Example 2.8** (sign restrictions for inequality constrains)**.** Consider the problem

$$\min x_1 + x_2 \text{ subject to } x_1^2 + x_2^2 \leq 2.$$

One has $x^* = (-1, -1)'$, $\nabla f(x^*) = (1, 1)'$, $\nabla g(x^*) = (-2, -2)'$ and hence $\mu^* = \frac{1}{2} > 0$. The fact that $\mu^* > 0$ implies that $\nabla f(x^*)$ and $\nabla g(x^*)$ point to opposite directions of the hyperplane $\nabla g(x^*) = 0$. Such a condition should hold, because if $x^*$ is locally optimal, if one moves in the feasible direction $-\nabla g(x^*)$ the value of the objective function $f$ should become larger and not smaller.

Summarizing we get with $L(x, \lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i h_i(x) + \sum_{j=1}^{r} \mu_j g_j(x)$:

**Proposition 2.9** (Karush-Kuhn-Tucker condition, KKT-conditions)**.** *Let $x^*$ be a local minimum of* (2.23)*, and suppose that $x^*$ is regular for* (2.24)*. Then there exist unique Lagrange multiplier vectors $\lambda^* = (\lambda_1^*, \dots, \lambda_m^*)$, $\mu^* = (\mu_1^*, \dots, \mu_r^*)$ such that*

$$\nabla_x L(x^*, \lambda^*, \mu^*) = 0, \quad \mu_j^* \geq 0, \ 1 \leq j \leq r, \ \mu_j^* = 0 \textit{ for } j \notin A(x^*). \tag{2.25}$$

The result follows from the preceding discussion except for the fact that $\mu_j^* \geq 0$ for $j \in A(x^*)$. This statement can be established by a variant of the penalty function approach, see Bertsekas (1999), page 317.

Note that the restrictions on $\mu^*$ can be written in the form

$$\mu^* \geq 0, \ \sum_{j=1}^{r} \mu_j^* g_j(x^*) = 0, \tag{2.26}$$

Equation (2.26) is known as *complementary slackness condition.*

**Remark 2.10.** In the case with inequality constraints the assumption that $x^*$ is regular is more restrictive than in the case with only equality constraints. In fact for many interesting problems $r$ is much larger than $n$ and it can happen that more than $n$ constraints are active at $x^*$. For this reason weaker forms of the KKT-conditions have been developed; these conditions are known as so-called *Fritz-John conditions.* We refer to Bertsekas (1999) for details.

The following example illustrates the use of the KKT conditions for the analytical solution of constrained optimization problems.

**Example 2.11.** Consider the problem

$$\min(x_1^2 + x_2^2 + x_3^2) \text{ subject to } x_1 + x_2 + x_3 \leq -3.$$

For a local minimum the FOCs (2.25) give

$$2x_1^* + \mu^* = 0,\ 2x_2^* + \mu^* = 0,\ 2x_3^* + \mu^* = 0. \tag{2.27}$$

Now there are two possibilities

a) The constraint is not binding. Then $\mu^* = 0$ and hence $x^* = (0,0,0)'$ which contradicts the inequality constraint.

b) The constraint is binding, $x_1^* + x_2^* + x_3^* = -3$. Then we get by summing over (2.27) that $\mu^* = 2$ and hence $x_i^* = -1, i = 1,2,3$. It is easily checked that $x^*$ is in fact the unique global minimum of the problem.

# Chapter 3

# Duality and Convex Programming

In this chapter we we study constrained optimization problems using some tools from convex analysis. We begin by introducing certain basic notions of convex analysis, in particular separation theorems; in Section 3.2 we consider the so-called dual problem associated with the Lagrange function; duality results (existence results for multipliers) are given in Section 3.3.

## 3.1 Convex sets and separating hyperplanes

Recall that a set $C \subset \mathbb{R}^n$ is convex if for all $x_1, x_2 \in C$, and all $\lambda \in [0,1]$ it holds that $\lambda x_1 + (1-\lambda)x_2 \in C$. Convex sets arise frequently in optimization.

Some special types of convex sets.

- A set $C \subset \mathbb{R}^n$ is a called finitely generated convex set or a *polytope* if there are $x_1, \ldots, x_m$ in $\mathbb{R}^n$ such that every $x \in C$ has a representation $x = \sum_{i=1}^n \gamma_i x_i$ for $\gamma_i \geq 0$, $\sum_i \gamma_i = 1$.
- A set $C \subset \mathbb{R}^n$ is called a *convex cone*, if $C$ is convex and if it holds that for $x \in C$ $\lambda x \in C$ for all $\lambda \geq 0$; equivalently, $C$ is a convex cone if $x_1, \ldots, x_m \in C \Rightarrow \sum_{i=1}^n \lambda_i x_i \in C$ for all $\lambda_i \in \mathbb{R}, \lambda_i \geq 0$.
- $C \subset \mathbb{R}^n$ is called a *polyhedron*, if it is of the form $C = \{x \in \mathbb{R}^n : Ax \leq b$ for some $A \in \mathbb{R}^{m \times n}, b \in R^m\}$.

In the next lemma we collect some simple properties of convex sets.

**Lemma 3.1.** *1) If $C \subset \mathbb{R}^n$ is convex, the closure $\bar{C}$ is convex as-well.*

*2) Define for $C_1, C_2 \subset \mathbb{R}^n$ the sum $C_1 + C_2 = \{z \in \mathbb{R}^n, z = x_1 + x_2$ for some $x_1 \in C_1, x_2 \in C_2\}$. Suppose that $C_1$ and $C_2$ are convex. Then the sum $C_1 + C_2$ is convex.*

*3) Suppose that the sets $C_1, C_2, \ldots \subset \mathbb{R}^n$ are convex. Then the intersection $\bigcap_{i=1,2,\ldots} C_i$ is convex.*

**Theorem 3.2** (Projection on convex sets.)**.** *Let $C \subset \mathbb{R}^n$ be a closed convex set, and $\|\cdot\|$ be the Euclidean norm. Then the following holds* *a) For every $x \in \mathbb{R}^n$ there is a unique $x^* \in C$ that minimizes $\|z - x\|$ over all $z \in C$, that is $\|x^* - x\| = \min_{z \in C} \|z - x\|$. The point $x^*$ is known as projection of $X$ on $C$.*

*b) Given $x \in \mathbb{R}^n$, $z \in C$ is equal to $x^*$ if and only if $(y - z)'(x - z) \leq 0$ for all $y \in C$.*

*c) The mapping $\mathbb{R}^n \to C$, $x \mapsto x^*$ is a contraction, that is $\|x^* - y^*\| \leq \|x - y\| \, \forall x, y \in R^n$.*

*Proof.* (see also Bertsekas, page 705). **a)** Fix $x \in \mathbb{R}^n$ and consider $w \in C$. Minimizing $\|x - z\|$ over $C$ is equivalent to minimizing $\|x - z\|$ over $\{z \in C : \|x - z\| \leq \|x - w\|\}$. This is a compact set, and the mapping $z \mapsto \|x - z\|^2$ is smooth. Hence existence of a minimizer follows from the Weierstrass Theorem (Proposition 1.12). For uniqueness note that if $x_1^*, x_2^*$ are two different solutions of the projection problem we have by the strict convexity of $x \mapsto \|x\|^2$ that for $\alpha \in (0, 1)$

$$\begin{aligned} \|x - (\alpha x_1^* + (1 - \alpha) x_2^*)\|^2 &= \|\alpha(x - x_1^*) + (1 - \alpha)(x - x_2^*)\|^2 \\ &< \alpha \|x - x_1^*\|^2 + (1 - \alpha) \|x - x_2^*\|^2 = \|x - x_1^*\|^2 , \end{aligned}$$

which contradicts the optimality of $x_1^*$.

**b)** For $y$ and $z$ in $C$ one has, using that $\|y - x\|^2 = (y - x)'(y - x)$,

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

Suppose now that for $z \in C$ one has $(x - z)'(y - z) \leq 0$ for all $y \in C$. Then we have that for all $y \in C$ $-2(y - z)'(x - z) \geq 0$ and hence $\|y - x\|^2 \geq \|z - x\|^2$, which implies that $z = x^*$. Conversely, we want to show that for $z = x^*$ the inequality

$$(y - x^*)'(x - x^*) \leq 0 \text{ for all } y \in C$$

holds. Suppose to the contrary that there is some $\tilde{y} \in C$ with $(\tilde{y} - x^*)'(x - x^*) > 0$. Define for $\alpha > 0$ $y_\alpha = \alpha \tilde{y} + (1 - \alpha) x^*$ and note that $y_\alpha \in C$ as $C$ is convex. It follows that

$$\frac{\partial}{\partial \alpha} \left( \|x - y_\alpha\|^2 \right) |_{\alpha = 0} = -2(x - x^*)'(\tilde{y} - x^*) < 0.$$

Hence for $\alpha$ small $\|x - y_\alpha\|^2 < \|x - x^*\|^2$, contradicting the optimality of $x^*$.

**c)** For this part we refer to the book of Bertsekas. □

**Definition 3.3** (Hyperplanes). i) A hyperplane with normal vector $a \in \mathbb{R}^n \backslash \{0\}$ at the level $b \in \mathbb{R}$ is the set $H = H(a, b) = \{x \in R^n : a'x = b\}$.

ii) A hyperplane with normal vector $a \in R^n$ through the point $x_0 \in \mathbb{R}^n$ is the set $\{x \in \mathbb{R}^n : a'x = a'x_0\}$.

iii) Consider the hyperplane $H(a, b) = \{x : a'x = b\}$. The two sets $\{x \in \mathbb{R}^n : a'x \geq b\}$ and $\{x \in \mathbb{R}^n : a'x \leq b\}$ are the positive respectively the negative half space associated with $H(a, b)$.

Fix $x_0 \in H$. Then $H$ consists of all points of the form $\{x = x_0 + z : a'z = 0\}$, i.e. $H$ is the displaced orthogonal complement of the vector $a$.

**Proposition 3.4** (Supporting Hyperplane Theorem)**.** *Consider some convex set $C \subset \mathbb{R}^n$ and suppose that $\bar{x}$ is not an interior point of $C$. Then there exists a hyperplane $H$ through $\bar{x}$ such that $C$ is in the positive half-space of $H$, that is there is $a \in \mathbb{R}^n \setminus \{0\}$ with $a'x \geq a'\bar{x}$ for all $x \in C$.*

*Proof.* (Idea of the proof) Since $\bar{x}$ is not an interior point of $C$ and since $C$ is convex, there exists a sequence $(x^k)_{k\in\mathbb{N}}$ such that $x^k \notin \bar{C}$ and such that $x^k \to x$ (see Bertsekas, Prop. B8). Denote by $(x^k)^* \in \bar{C}$ the projection of these points on the convex set $\bar{C}$, see Theorem 3.2. Then it is easily seen from point b) of the theorem that $\bar{C}$ lies within the positive halfspace of the hyperplane through $x^k$ with normal vector $a^k = \frac{(x^k)^* - x_k}{\|(x^k)^* - x_k\|}$. Since the sequence $a_k$ belongs to the compact set $S_1 = \{z \in \mathbb{R}^n \colon \|z\| = 1\}$, it contains a convergent subsequence with limit vector $a^\infty \in S_1$. By continuity, $C$ lies in the positive halfspace of the hyperplane through $\bar{x}$ with normal vector $a^\infty$. □

**Proposition 3.5** (Separating Hyperplane Theorem)**.** *Let $C_1$ and $C_2$ be two nonempty convex subsets of $\mathbb{R}^n$ such that $C_1 \cap C_2 = \emptyset$. Then there exists a hyperplane that separates the two sets, i.e. there is a vector $a \in \mathbb{R}^n \setminus \{0\}$ with $a'x_1 \leq a'x_2$ for all $x_1 \in C_1, x_2 \in C_2$.*

*Proof.* Consider the convex set $C = \{x \in \mathbb{R}^n : x = x_2 - x_1 \text{ for some } x_1 \in C_1, x_2 \in C_2\}$. Then $0 \notin C$, as $C_1$ and $C_2$ are disjoint. By the supporting hyperplane theorem there is some $a \neq 0 \in \mathbb{R}^n$ with $a'x \geq a'0 = 0$ for all $x \in C$. Since $x = x_2 - x_1$ it follows that $a'x_2 \geq a'x_1 \; \forall \; x_1 \in C_1, x_2 \in C_2$ as claimed. □

**Proposition 3.6** (Strict Separation)**.** *If $C_1$ and $C_2$ are two nonempty and disjoint convex sets such that $C_1$ is closed and $C_2$ is compact there exists a hyperplane that strictly separates these sets, that is there is some $a \in \mathbb{R}^n, a \neq 0$ and some $b \in \mathbb{R}$ such that*

$$a'x_1 < b < a'x_2 \; \forall x_1 \in C_1, x_2 \in C_2.$$

For the proof we refer to Bertsekas, Prop. B14.

## 3.2 The Dual Problem for constrained optimization

Now we return to the analysis of constrained optimization problems. Consider the problem

$$\min_{x\in X} f(x) \;\text{ subject to }\; h(x) = 0, g(x) \leq 0, \tag{3.1}$$

where $f$, $h = (h_1, \dots, h_m)$ and $g = (g_1, \dots, g_r)$ are defined on the set $X \subset \mathbb{R}^n$. In the sequel we are particularly concerned with the case where $f$, $g$ and $h$ are convex or even linear. A point $x \in X$ with $h(x) = 0$ and $g(x) \leq 0$ is sometimes called *feasible*. The *optimal value* of problem (3.1) is given by

$$f^* = \inf\{f(x) : x \in X, h(x) = 0, g(x) \leq 0\}.$$

The next definition is motivated by the Lagrange multiplier theory of Chapter 3.

**Definition** 3.7. a) The *Lagrange function* associated with problem (3.1) is

$$L(x, \lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i h_i(x) + \sum_{j=1}^{r} \mu_j g_j(x),\; x \in X,\; \lambda \in \mathbb{R}^m,\; \mu \in \mathbb{R}^r. \tag{3.2}$$

b) The *dual function* associated with problem (3.1) is

$$q(\lambda, \mu) = \inf_{x \in X} L(x, \lambda, \mu); \tag{3.3}$$

the domain of $q$ is $\operatorname{dom} q = \{\lambda \in \mathbb{R}^m, \mu \in \mathbb{R}^r, q(\lambda, \mu) > -\infty\}$.

Since $q(\lambda, \mu)$ is the infimum of linear functions, i.e.

$$q(\lambda, \mu) = \inf_{x \in X} f(x) + \lambda' h(x) + \mu' g(x),$$

it holds has that $q(\lambda, \mu)$ is concave in $(\lambda, \mu)$ and $\operatorname{dom} q$ is convex.

The next lemma gives a first link between the dual function and the problem (3.1).

**Lemma 3.8.** *For $\lambda \in \mathbb{R}^m$, $\mu \in \mathbb{R}^r$ such that $\mu \geq 0$ it holds that $q(\lambda, \mu) \leq f^*$.*

*Proof.* Consider some $x \in X$ with $h(x) = 0$, $g(x) \leq 0$. We get, as $\mu \geq 0$, that

$$f(x) + \lambda' h(x) + \mu' g(x) \leq f(x).$$

Hence we get by taking the infimum over $x$,

$$\begin{aligned} q(\lambda, \mu) = \inf_{x \in X} L(x, \lambda, \mu) &\leq \inf\{L(x, \lambda, \mu) : x \text{ feasible in } (3.1)\} \\ &\leq \inf\{f(x) : x \text{ feasible in } (3.1)\}, \end{aligned}$$

and the last infimum is obviously equal to $f^*$. □

### 3.2.1 The dual problem

According to lemma 3.8, $q(\lambda, \mu)$ is a lower boundary to $f^*$ if $\mu \geq 0$. It is natural to look for the largest lower bound to $f^*$. This leads to the so-called *dual problem*

$$\max_{(\lambda,\mu) \in \operatorname{dom} q} q(\lambda, \mu) \text{ subject to } \mu \geq 0. \tag{3.4}$$

Note that the dual problem is a convex optimization problem as $q$ is concave and $\operatorname{dom} q$ is convex. Lemma 3.8 immediately gives

**Corollary 3.9** (weal duality)**.** *One has the inequality*

$$q^* := \sup\{q(\lambda, \mu) : (\lambda, \mu) \in \operatorname{dom} q, \mu \geq 0\} \leq f^*.$$

**Example 3.10** (Dual problem for linear programming)**.** Consider a linear optimization problem of the form

$$\min c'x \text{ subject to } Ax \leq b \tag{3.5}$$

for $c \in \mathbb{R}^n$, $b \in \mathbb{R}^r$, $A \in \mathbb{R}^{r \times n}$. The Lagrangian is

$$L(x, \mu) = c'x + \mu'(Ax - b) = (c + A'\mu)'x - \mu' b.$$

The dual function is

$$q(\mu) = \inf_{x \in \mathbb{R}^n} (c + A'\mu)'x - \mu' b = \begin{cases} -\mu' b, & if\, A'\mu + c = 0 \\ -\infty, & else \end{cases}$$

as $\min_{x \in \mathbb{R}^n} a'x = -\infty$ for any $a \neq 0$. The dual problem is therefore

$$\max_{\mu \in \mathbb{R}^r} -\mu' b \text{ subject to } A'\mu = -c, \mu \geq 0. \tag{3.6}$$

### 3.2.2 Multipliers and optimality conditions.

One says that there is a *duality gap* for the optimization problem problem (3.1) if $q^* < f^*$; if, on the other hand, $q^* = f^*$ there is no duality gap for the problem ( recall that$q^* \le f^*$ by Corollary 3.9). In the sequel we are interested in conditions that ensure that there is no duality gap.

**Definition** 3.11. A pair $\lambda^* \in R^m$, $\mu^* \in R^r$ with $\mu^* \ge 0$ is called a multiplier for problem (3.1) if it holds that

$$f^* = \inf_{x \in X} L(x, \lambda^*, \mu^*).$$

Part of the interest in the existence of multipliers stems from the following observation: if $(\lambda^*, \mu^*)$ is a multiplier for the problem (3.1), the optimal value $f^*$ for problem (3.1) can be obtained as solution of the unconstrained problem $\min_{x \in X} L(x, \lambda^*, \mu^*)$. The following lemma shows that multipliers are closely related to the solutions of the dual problem.

**Lemma 3.12** (Multipliers and duality gap). *a) If there is a multiplier pair* $(\lambda^*, \mu^*)$ *for problem* (3.1) *then there is no duality gap and* $(\lambda^*, \mu^*)$ *are solutions of the dual problem* (3.4)*.*

*b) Suppose that* $(\lambda^*, \mu^*)$ *is a solution of the dual problem. If* $q^* = f^*$ *(no duality gap), then* $(\lambda^*, \mu^*)$ *is a multiplier pair.*

Note in particular that optimization problems with a duality gap do not admit multipliers in the sense of Definition **??**.

*Proof.* Statement a) follows from the definitions in a straightforward way: if $(\lambda^*, \mu^*)$ is a multiplier pair we have

$$q(\lambda^*, \mu^*) = \inf_{x \in X} L(x, \lambda^*, \mu^*) = f^*.$$

Since $f^* \ge q^*$ by weak duality, we get $q(\lambda^*, \mu^*) \ge q^*$, and $(\lambda^*, \mu^*)$ is optimal in the dual problem.

For b) we argue as follows. If $(\lambda^*, \mu^*)$ is optimal in the dual problem and if $f^* = q^*$, we have $q(\lambda^*, \mu^*) = f^*$, so that $(\lambda^*, \mu^*)$ is a multiplier. □

**Example 3.13** (Existence of Multipliers)**.** Consider the linear problem

$$\min_{x \in X} f(x) = x_1 - x_2 \text{ subject to } g(x) = x_1 + x_2 - 1 \le 0,$$
$$X = \{x_1, x_2 : x_i \ge 0\}.$$

In geometric terms the set of feasible points is just the unit simplex, and it is easily seen that the optimal point is $x^* = (0, 1)$ and that $f^* = -1$.

Next we consider the dual problem. We get for $\mu \in \mathbb{R}$

$$L(x, \mu) = x_1 - x_2 + \mu(x_1 + x_2 - 1), \tag{3.7}$$

and hence for $\mu \ge 0$

$$q(\mu) = \inf_{x_1, x_2 \ge 0} L(x, \mu) = \inf_{x_1, x_2 \ge 0} x_1(1 + \mu) + x_2(\mu - 1) - \mu = \begin{cases} -\infty & , \mu \le 1 \\ -\mu & , \mu \ge 1. \end{cases}$$

Hence the solution of the dual problem is $q^* = \sup_{\mu \geq 1}\{-\mu\} = -1$, and the dual optimum is attained at $\mu^* = 1$. It follows that $f^* = q^*$ (no duality gap), and that $\mu^* = 1$ is a multiplier. It is well-known that for linear optimization problems multipliers always exist, if the primal problem has a solution.

**Proposition 3.14** (Optimality conditions)**.** *Consider $x^* \in X$, $\lambda^* \in \mathbb{R}^m$, $\mu^* \in \mathbb{R}^r$. Then the following statements are equivalent:*

*a) $x^*$ is optimal in the primal problem* (3.1) *and $(\lambda^*, \mu^*)$ is a multiplier (this implies $\mu^* \geq 0$ by definition).*

*b) The following four optimality conditions hold*

  *i) $x^* \in X$, $g(x^*) \leq 0$, $h(x^*) = 0$ (primal feasibility),*

  *ii) $\mu^* \geq 0$ (dual feasibility),*

  *iii) $x^* \in \operatorname{argmin}_{x \in X} L(x, \lambda^*, \mu^*)$ (Lagrangian optimality),*

  *iv) $\sum_{j=1}^r \mu_j^* g_j(x^*) = 0$ (complementary slackness).*

*Proof.* $a \Rightarrow b$. Primal and dual feasibility ((i) and (ii)) hold by definition. Since $x^*$ is optimal we have, as $\mu^* \geq 0$,

$$f^* = f(x^*) \geq f(x^*) + (\mu^*)' g(x^*) = L(x^*, \lambda^*, \mu^*) \geq \inf_x L(x, \lambda^*, \mu^*) = f^*,$$

where the last equality holds as $(\lambda^*, \mu^*)$ is a pair of multipliers. Hence equality holds everywhere in the above chain of equations, which gives Lagrangian optimality and complementary slackness.

$b \Rightarrow a$. Statement $b$) implies the following chain of inequalities:

$$f(x^*) \overset{(iv)}{=} L(x^*, \lambda^*, \mu^*) \overset{(iii)}{=} q(\lambda^*, \mu^*) \leq q^* \leq f^*, \tag{3.8}$$

where the last inequality follows from weak duality. Since $x^*$ feasible in the primal problem the inequality $f(x^*) \leq f^*$ implies that $x^*$ is optimal in the primal problem. Moreover, we must have equality throughout in (3.8). In particular we get $q(\lambda^*, \mu^*) = f^*$, which shows that $(\lambda^*, \mu^*)$ are multipliers. □

The proposition suggests the following strategy for finding solutions to constraint optimization problems.

1. Find for given $\lambda \in \mathbb{R}^n$, $\mu \in \mathbb{R}^r$, $\mu \geq 0$ a vector $\hat{x} = \hat{x}(\lambda, \mu)$ with $\hat{x} \in \operatorname{argmin}_{x \in X} L(x, \lambda, \mu)$.

2. Determine $\lambda^*$, $\mu^* \geq 0$ so that $\hat{x}(\lambda^*, \mu^*)$ is primally feasible and so that $(\mu^*)' g(\hat{x}(\lambda^*, \mu^*)) = 0$.

If these steps can be carried out successfully, $x^* = \hat{x}(\lambda^*, \mu^*)$ is optimal in the primal problem and $(\lambda^*, \mu^*)$ is a multiplier.

Note however, that this strategy will fail if the primal problem (3.1) does not admit multipliers, e.g. because there is a duality gap. Hence it is important to have conditions that guarantee the existence of multipliers (see Section 3.3).

**Example 3.15** (Portfolio optimization)**.** Consider an arbitrage-free and complete security market model on some finite state space $\Omega = \{\omega_1, \dots, \omega_n\}$. Denote the historical measure by $p_i = P(\{\omega_i\})$, $1 \le i \le n$ (we assume $p_i > 0$ for all $i$), the risk-free interest rate by $r \ge 0$ and the equivalent martingale measure by $q_i = Q(\{\omega_i\})$, $1 \le i \le n$. Recall that the price of any contingent claim $X = (x_1, \dots, x_n)$ with $x_i = X(\omega_i)$ is

$$\Pi_X = \frac{1}{1+r} \sum_i q_i x_i.$$

Consider a utility function $u$, i.e. some $C^2$-function $u : \mathbb{R}^+ \to \mathbb{R}$ with $u' > 0$, $u'' < 0$, where the concavity of $u$ is used to model risk aversion. Popular examples are $u(x) = \ln x$ or $u(x) = \frac{1}{\gamma} x^\gamma$ for $\gamma < 1$, $\gamma \neq 0$. Consider an investor with initial wealth $V_0 > 0$ who wants to maximize the expected utility $\mathbb{E}\left(u(X)\right) = \sum_i p_i u(x_i)$ of his terminal wealth $X = (x_1, \dots, x_n)$ with $x_i = X(\omega_i)$. Since the agent can create only wealth profiles $X$ whose current value is no larger than $V_0$, we are led to the problem

$$\max_{x_1,\dots,x_n \ge 0} \sum_{i=1}^n p_i u(x_i) \text{ subject to } \sum_{i=1}^n q_i x_i \le (1+r)V_0$$

With $v(x) := -u(x)$ we obtain the following equivalent minimization problem

$$\min_{x_1,\dots,x_n \ge 0} \sum_{i=1}^n p_i v(x_i) \text{ subject to } \sum_i q_i x_i \le (1+r)V_0.$$

In the sequel we assume that

$$\lim_{x \to 0} u'(x) = \infty, \lim_{x \to \infty} u'(x) = 0.; \tag{3.9}$$

these conditions are known as Inada conditions.

In order to find an optimal terminal wealth we now carry out the strategy explained above.

Step 1. The Lagrange function is for $\mu \ge 0$ given by

$$L(x, \mu) = \sum_{i=1}^n \{p_i v(x_i) + \mu q_i x_i\} - \mu(1+r)V_0.$$

Minimization of $x \mapsto L(x, \mu)$ gives the FOCs

$$p_i v'(x_i) + \mu q_i = 0, \ 1 \le i \le n. \tag{3.10}$$

Denote by $I$ the inverse function of $v'$. $I$ exists, as $v'$ is increasing because of the convexity of $v = -u$. Moreover, because of the Inada conditions we obtain that

$$\text{dom } I = \text{range } v' = (-\infty, 0); \quad \text{range } I = \text{dom } v' = (0, \infty).$$

Hence the FOC (3.10) gives

$$\hat{x}_i = \hat{x}_i(\mu) = I\left(-\mu \frac{q_i}{p_i}\right) > 0, \quad 1 \le i \le n.$$

Note that the point $\hat{x}(\mu) = (\hat{x}_1(\mu), \dots, \hat{x}_n(\mu))$ is in fact a global minimum of $x \mapsto L(x, \mu)$ as $L$ is convex in $x$.

Step 2. By Proposition 3.14 $\hat{x}(\mu^*)$ is optimal if we can find a multiplier $\mu^* > 0$ such that

$$\sum_{i=1}^{n} q_i \hat{x}_i(\mu^*) = (1+r)V_0. \tag{3.11}$$

Such a multiplier exists as $I$ is increasing with

$$\lim_{z\to-\infty} I(z) = 0 \lim_{z\to 0} I(z) = \infty.$$

Hence if $\mu$ increases from 0 to $\infty$, $\hat{x}_i(\mu)$ decreases from $\infty$ to 0, and there is a unique solution$\mu^*$ of (3.11). The optimal wealth is than $x^* = \hat{x}(\mu^*)$; the optimal trading strategy can be found by means of a replication argument.

For a specific example take $u(x) = \ln x$. In that case $v'(x) = -\frac{1}{x}$ and $I(y) = -\frac{1}{y}$. Hence we get $\hat{x}_i(mu) = \frac{1}{\mu}\frac{p_i}{q_i}$. To determine $\mu^*$ we use the condition

$$\sum_{i=1}^{n} q_i \hat{x}_i(\mu^*) = \sum_{i=1}^{n} q_i \frac{1}{\mu^*}\frac{p_i}{q_i} = (1+r)V_0.$$

This gives $\mu^* = ((1+r)V_0)^{-1}$ and hence the optimal wealth

$$x_i^* = \frac{1}{\mu^*}\frac{p_i}{q_i} = (1+r)V_0\frac{p_i}{q_i}, 1 \le i \le n.$$

The next result is a reformulation of the optimality conditions from Proposition 3.14.

**Proposition 3.16.** *(Saddle Point Theorem) Given $x^* \in X \subset \mathbb{R}^n$, $\lambda^* \in \mathbb{R}^m$, $\mu \in \mathbb{R}^n \ge 0$. Then the following are equivalent:*

*a) $x^*$ is optimal in* (3.1) *and $(\lambda^*, \mu^*)$ is a multiplier.*

*b) $(x^*, (\lambda^*, \mu^*))$ is a saddle point of the Lagrange-function, that is*

$$L(x^*, (\lambda, \mu)) \le L(x^*, (\lambda^*, \mu^*)) \le L(x, (\lambda^*, \mu^*)) \tag{3.12}$$

*for all $x \in X$, $\lambda \in \mathbb{R}^m$, $\mu \ge 0 \in \mathbb{R}^r$.*

*Proof.* $a \Rightarrow b$. Lagrangian optimality implies the right inequality in (3.12). Since $g(x^*) \le 0$ we moreover have $L(x^*, (\lambda, \mu)) \le f(x^*) = L(x^*, (\lambda^*, \mu^*))$, which gives the left inequality. $b \Rightarrow a$. It holds that

$$\sup_{\lambda,\mu\ge 0} L(x^*, \lambda, \mu) = \sup_{\lambda,\mu\ge 0} f(x^*) + \lambda' h(x^*) + \mu' g(x^*) = \begin{cases} f(x^*), & \text{if } h(x^*) = 0 \text{ and } g(x^*) \le 0 \\ \infty, & \text{else} \end{cases}$$

Now the left inequality in (3.12) implies that

$$\sup_{\lambda,\mu} L(x^*, (\lambda, \mu)) \le L(x^*, (\lambda^*, \mu^*)) < \infty,$$

so that we must have $h(x^*) = 0$, $g(x^*) \le 0$, and hence primal feasibility of $x^*$. The primal feasiblity of $x^*$ implies that $L(x^*, (\lambda^*, \mu^*)) = f(x^*)$ and hence complementary slackness. Lagrangian optimality finally follows from the right side of (3.12). □

## 3.3 Duality results

In section we discuss existence results for multipliers, so-called duality results.

### 3.3.1 Visualization of multipliers.

For simplicity we consider only problems with inequality constraints, i.e. we consider the problem

$$\min_{x \in X} f(x) \ \text{ subject to } \ g(x) \leq 0.$$

We define the set $S \subset \mathbb{R}^{n+1}$ by $S = \{(g(x), f(x)) : x \in X\}$, so that

$$f^* = \inf\{\omega \in \mathbb{R} : \exists z \in \mathbb{R}^r, z \leq 0, \text{ with } (z, \omega) \in S\}.$$

Recall that a hyperplane $H$ through some $(\bar{z}, \bar{\omega}) \in \mathbb{R}^{r+1}$ with normal vector $(\mu, \mu_0) \in \mathbb{R}^{r+1}$ is

$$H = \{(z, w) \in \mathbb{R}^r \times \mathbb{R} : \mu' z + \mu_0 w = \mu' \bar{z} + \mu_0 \bar{w}\},$$

and recall the notion of the associated half-space $H^+$.

**Lemma 3.17.** *(Visualization of Lagrange multipliers)*

*a) The hyperplane with normal $(\mu, 1)$ through some vector $(g(x), f(x)) \in S$ intercepts the vertical axis $\{(0, w) : w \in \mathbb{R}\}$ at the level $L(x, \mu)$.*

*b) Among all hyperplanes with normal $(\mu, 1)$ that contain the set $S$ in their positive half-space the highest attained level of interception is $\inf_{x \in X} L(x, \mu) = q(\mu)$.*

*c) $\mu^*$ is a multiplier if $\mu^* \geq 0$ and if the highest attained level of interception of all hyperplanes with normal $(\mu, 1)$ containing $S$ in their positive half space is $f^*$.*

*Proof.* a) The hyperplane is the set

$$H = \{(z, w) : \mu' z + w = \mu' g(x) + f(x)\}.$$

A point $(0, w)$ is in $H$ if and only if $w = \mu' g(x) + f(x) = L(x, \mu)$.

b) Denote by $(0, w)$ the level of interception of some hyperplane with normal $(\mu, 1)$. By definition, $S \in H^+$ if and only if for all $x \in X$.

$$\mu' g(x) + f(x) \geq \mu' 0 + w = w,$$

that is if $w \leq L(x, \mu)$. Hence the highest possible point of interception is $\inf_{x \in X} L(x, \mu) = q(\mu)$.

c) This follow immediately from the definiton of a multiplier.

□

**Example 3.18.** a) Consider the problem

$$\min_{x\in\mathbb{R}^2} \frac{1}{2}(x_1^2 + x_2^2) \text{ subject to } g(x) = x_1 - 1 \le 0.$$

Apparent by, the optimum is attainend at $x^* = (0,0)$ and $f^* = 0$. In order to draw the set $S$ we use its lower boundary is given by

$$\begin{aligned}\inf\{\omega : \exists x \in \mathbb{R}^2 \text{ with } g(x) = z, f(x) = w\}\\ = \inf\{\omega : \exists x_2 \in \mathbb{R} \text{ with } \frac{1}{2}(z+1)^2 + \frac{1}{2}x_2^2 \le w\} = \frac{1}{2}(z+1)^2\}.\end{aligned}$$

Hence we get the following picture: (picture only on blackboard)

b) Consider the problem

$$\min_{x\in\mathbb{R}} x \text{ subject to } g(x) = x^2 \le 0.$$

Since $x = 0$ is the only feasible point we have $x^* = 0$ and $f^* = 0$. The dual function is given by

$$q(\mu) = \inf_{x\in\mathbb{R}} x + \mu x^2 = -\frac{1}{4\mu}, \ \mu \ge 0.$$

Hence we get $q^* = \sup_{\mu\ge 0} q(\mu) = 0$, so that there is no duality gap. However, there is no multiplier as $q^*$ is not attained.

Now we turn to the geometric interpretation. The set $S$ is given by $S = \{(z,w) : z \ge 0, w = \pm\sqrt{z}\}$. From a geometric viewpoint, the non-existence of multipliers is due to the fact that the supporting hyperplane $H$ for the set $S$ at the 'optimal point' $(0,0)$ is vertical, so that the normal vector of $H$ is collinear to $(1,0)$.

### 3.3.2 Duality results for problems with inequality constraints.

We consider the problem $\min_{x\in X} f(x)$ subject to $g(x) \le 0$. We make the following assumptions on the data of the problem.

**Assumption 3.19** (Convexity and inner point)**.** *The admissible domain $\{x \in X : g_j(x) \le 0$ for $j = 1,\dots,r\}$ is nonempty and the optimal value $f^*$ is finite. Moreover, $f$ and $g_1,\dots,g_r$ are convex functions on the convex set $X$. Finally, the so-called Slater condition holds, that is there is some $\bar{x} \in X$ with $g_j(\bar{x}) < 0$ for all $j$.*

**Theorem 3.20** (Duality theorem for inequality constraints.)**.** *Under assumption* (3.19) *there is no duality gap, and there is at least one multiplier $\mu^*$.*

Note that the Slater condition is not satisfied in example (3.18) b), but all other assumptions of the theorem are. Hence the example shows that the Slater condition or some other constraint qualification is really needed to ensure the existence of multipliers.

*Proof.* We consider the set $A = S + \mathbb{R}_+^{n+1}$, that is

$$A = \{(z,w) \in \mathbb{R}^{n+1} : \exists x \in X \text{ with } g_j(x) \le z_j, 1 \le j \le r \text{ and } f(x) \le w\}$$

Convexity of $f, g_1, \ldots, g_r$ and of $X$ imply that $A$ is a convex subset of $\mathbb{R}^{r+1}$. Moreover, $(0, f^*)$ is not an inner point of $A$ (otherwise $(0, f^* - \epsilon)$ would be belong to $A$ for $\epsilon > 0$ sufficiently small, which contradicts the definition of $f^*$ as optimal value). The supporting hyperplane theorem implies the existence of $(\mu, \beta) \neq (0,0) \in \mathbb{R}^{r+1}$ such that

$$\beta f^* = \beta f^* + \mu' 0 \leq \beta w + \mu' z \text{ for all } (z, w) \in \bar{A} \tag{3.13}$$

Since $(0, f^* + 1) \in A$ we get from (3.13) that $\beta \geq 0$, and since $(e_j, f^*) \in \bar{A}$ for $j = 1, \ldots, r$ we get that $\mu_j \geq 0$ for all $j$. Next we show that the Slater condition implies that $\beta$ is strictly bigger than zero. In fact, if $\beta = 0$, (3.13) implies that $0 \leq \mu' z$ for all $(z, w) \in A$. Since $(g_1(\bar{x}), \ldots, g_r(\bar{x}), f(\bar{x})) \in A$, we would get

$$0 \leq \mu' g(\bar{x}) = \sum_{j=1}^{r} \mu_j g_j(\bar{x}).$$

Since $g_j(\bar{x}) < 0$ for all $j$, we thus must have $\mu_j = 0$ for all j. Hence we have $(\mu, \beta) = (0, 0)$ which contradicts our previous assertion $(\mu, \beta) \neq (0, 0)$.

Hence we have $\beta > 0$, and by division through $\beta$ we may assume that $\beta = 1$. Then (3.13) gives, as $(g(x), f(x)) \in A$ for $x \in X$,

$$f^* \leq f(x) + \mu' g(x) \text{ for all } x \in X.$$

Taking the infimum over $x \in X$ we get, as $\mu \geq 0$,

$$f^* \leq \inf_{x \in X} \{f(x) + \mu' g(x)\} = q(\mu) \leq q^*,$$

so that there is no duality gap. Since $q(\mu) \leq f^*$ by weak duality, we have that $\mu$ is a solution of the dual problem and hence a multiplier. □

**Example 3.21** (Cost-minimizing production)**.** Assume that at least $A$ units of some good (say, electricity) need to be produced by $n$ different independent production units (say, power plants). Denote by $x_i$ the quantity produced by unit $i$, and by $f_i(x_i)$ the cost function of that unit. Assume moreover, that one has capacity constraints of the form $\alpha_i \leq x_i \leq \beta_i$ for the output of unit $i$. Then the problem of finding a cost-minimizing production plan $(x_1, \ldots, x_n)$ can be written in the form

$$\min_{\alpha_i \leq x_i \leq \beta_i} \sum_{i=1}^{n} f_i(x) \text{ subject to } \sum_{i=1}^{n} x_i \geq A,$$

In order to proceed we assume that $f_1, \ldots, f_n$ are strictly convex and $C^1$ on $(a_i, \beta_i)$, and that $\sum_i \beta_i > A$. Hence Assumption 3.19 is satisfied for our problem and we can safely apply the solution methodology from Proposition 3.14, since we know that multipliers exist by Theorem 3.20.

Step 1. We consider the Lagrange function

$$L(x, \mu) = \sum_{i=1}^{n} f_i(x_i) + \mu(A - \sum_{i=1}^{n} x_i) = \sum_{i=1}^{n} (f_i(x_i) - \mu x_i) + \mu A$$

where $x \in X = \{x \in \mathbb{R}^n : \alpha_i \leq x_i \leq \beta_i \text{ for all } i\}$, that is we include the capacity constraints in the set $X$. Because of due the special form of $L(x, \mu)$, Lagrangian minimization amounts to solving the $n$ one-dimensional problems $\min_{\alpha_i \le x_i \le \beta_i}(f(x_i) - \mu x_i)$. For given $\mu \ge 0$ the optimum is attained at

$$x_i(\mu) = \begin{cases} \alpha_i, & \text{if } f'(\alpha_i) \ge \mu_i \\ I_i(\mu), & \text{if } f'(\alpha_i) < \mu < f'(\beta_i) \\ \beta_i, & \text{if } f'(\beta_i) \le \mu \end{cases}$$

here $I_i(\cdot)$ is the inverse of the strictly increasing function $f_i'$.

Step 2. Next we determine $\mu^* \ge 0$ such that the complementary slackness condition $\mu^*(A - \sum_i x_i) = 0$ is satisfied. For this we use that the mapping $\mu \mapsto x_i^*(\mu)$ is continuous and increasing. We distinguish two cases.

a) $\sum_{i=1}^n x_i^*(0) \ge A$. In that case we take $\mu^* = 0$ and $x^* = (x_1^*(0), \dots, x_n^*(0))$.

b) $\sum_{i=1}^n x_i^*(0) < A$. As $\lim_{\mu \to \infty} x_i^*(\mu) = \beta_i$ for all $i$, we can find some $\mu^* > 0$ such that $\sum_{i=1}^n x_i^*(\mu^*) = A$ (using the assumption that $\sum_i \beta_i > A$). In that case $(x_1^*(\mu^*), \dots, x_n^*(\mu^*))$ solves the problem.

**Remark 3.22.** The expression $g_i(x_i, \mu) = \mu x_i - f_i(x_i)$ can be viewed as profit of production unit $i$ given that the good can be sold at a price $\mu$ on the market. Hence the problem $\min_{\alpha_i \le x_i \le \beta_i} -g(x_i)$ can be viewed as problem of choosing a profit-maximizing output amount for unit $i$. This implies that the problem of finding an optimal production plan could be decentralized if a central unit announces a shadow price $\mu$ and lets the subunits respond by choosing their optimal units $x_i(\mu)$. In the second step the shadow price $\mu$ has to be adapted so that the constraint $\mu^*(A - \sum_{i=1}^n x_i(\mu^*)) = 0$ is satisfied. The existence of an shadow price $\mu^*$ is ensured by Theorem 3.20.

### 3.3.3 Duality results for convex cost function and linear constraints

In the following we give results for the existence of multipliers for problems with linear constraints of the form.

$$\min_{x \in X} f(x) \text{ subject to } a_j' x \le b_j, 1 \le j \le n, e_i' x = \delta_i, 1 \le i \le m. \tag{3.14}$$

**Assumption 3.23.** *$f$ is convex on $\mathbb{R}^m$, and $X$ is a polyhedral set, that is a set generated by linear inequalities.*

**Proposition 3.24.** *Suppose that assumption 3.23 holds and that the optimal value $f^*$ of problem* (3.14) *is finite. Then strong duality holds and there is at least one multiplier pair $(\lambda^*, \mu^*)$.*

**Corollary 3.25.** *(Duality for linear programming) Suppose that $f$ is linear, that is of the form $f(x) = c'x$. If the optimal value is finite, the primal and the dual problem have optimal solutions and there is no duality gap.*

# References

Bertsekas, D. (1999), *Nonlinear Programming*, 2nd edn, Athena Scientific, Belmont, Massachusetts.

Bertsimas, D. & Tsitsiklis, J. (1997), *Introduction to Linear Optimization*, Athena Scientific, Belmont, Massachusetts.