Chapter 3

Nonlinear programming

3.1 Introduction

In this chapter we discuss the basic tools for the solution of optimization problems of the form

$$P_0 \begin{cases} \min_x f(x) \\ g(x) = 0 \\ h(x) \leq 0. \end{cases} \tag{3.1}$$

In the problem P_0 there are both equality and inequality constraints[1]. However, sometimes for simplicity, or because a method has been developed for problems with special structure, we will refer to problems with only equality constraints, *i.e.* to problems of the form

$$P_1 \begin{cases} \min_x f(x) \\ g(x) = 0, \end{cases} \tag{3.2}$$

or to problems with only inequality constraints, *i.e.* to problems of the form

$$P_2 \begin{cases} \min_x f(x) \\ h(x) \leq 0. \end{cases} \tag{3.3}$$

In all the above problems we have $x \in I\!R^n$, $f : I\!R^n \to I\!R$, $g : I\!R^n \to I\!R^m$, and $h : I\!R^n \to I\!R^p$. From a formal point of view it is always possible to transform the equality constraint $g_i(x) = 0$ into a pair of inequality constraints, *i.e.* $g_i(x) \leq 0$ and $-g_i(x) \leq 0$. Hence, problem P_1 can be (equivalently) described by

$$\tilde{P}_1 \begin{cases} \min_x f(x) \\ g(x) \leq 0 \\ -g(x) \leq 0, \end{cases}$$

which is a special case of problem P_2. In the same way, it is possible to transform the inequality constraint $h_i(x) \leq 0$ into the equality constraint $h_i(x) + y_i^2 = 0$, where y_i is an auxiliary variable (also called *slack* variable). Therefore, defining the extended vector $z = [x', y']'$, problem P_2 can be rewritten as

$$\tilde{P}_2 \begin{cases} \min_z f(x) \\ h(x) + Y = 0, \end{cases}$$

with

$$Y = \begin{bmatrix} y_1^2 \\ y_2^2 \\ \vdots \\ y_p^2 \end{bmatrix},$$

which is a special case of problem P_1.

[1] The condition $h(x) \leq 0$ has to be understood element-wise, *i.e.* $h_i(x) \leq 0$ for all i.

Note however, that the transformation of equality constraints into inequality constraints yields an increase in the number of constraints, whereas the transformation of inequality constraints into equality constraints results in an increased number of variables.
Given problem P_0 (or P_1, or P_2), a point x satisfying the constraints is said to be an admissible point, and the set of all admissible points is called the admissible set and it is denoted with $\mathcal{X}$. Note that the problem makes sense only if $\mathcal{X} \neq \emptyset$.
In what follows it is assumed that the functions f, g and h are two times differentiable, however we do not make any special hypothesis on the form of such functions. Note however, that if g and h are linear there are special algorithms, and linear/quadratic programming algorithms are used if f is linear/quadratic and g and h are linear. We do not discuss these special algorithms, and concentrate mainly on algorithms suitable for general problems.

3.2 Definitions and existence conditions

Consider the problem P_0 (or P_1, or P_2). The following definitions are instrumental to provide a necessary condition and a sufficient condition for the existence of local minima.

Definition 6 *An open ball with center $x^\star$ and radius $\theta > 0$ is the set*

$$B(x^\star, \theta) = \{x \in I\!R^n \mid \|x - x^\star\| < \theta\}.$$

Definition 7 *A point $x^\star \in \mathcal{X}$ is a constrained local minimum if there exists $\theta > 0$ such that*

$$f(y) \geq f(x^\star), \tag{3.4}$$

for all $y \in \mathcal{X} \cap B(x^\star, \theta)$.
A point $x^\star \in \mathcal{X}$ is a constrained global minimum if

$$f(y) \geq f(x^\star), \tag{3.5}$$

for all $y \in \mathcal{X}$.
If the inequality (3.4) (or (3.5)) holds with a strict inequality sign for all $y \neq x^\star$ then the minimum is said to be strict.

Definition 8 *The i-th inequality constraints $h_i(x)$ is said to be active at $\tilde{x}$ if $h_i(\tilde{x}) = 0$. The set $I_a(\tilde{x})$ is the set of all indexes i such that $h_i(\tilde{x}) = 0$,* i.e.

$$I_a(\tilde{x}) = \{i \in \{1, 2, \cdots, p\} \mid h_i(\tilde{x}) = 0\}.$$

The vector $h_a(\tilde{x})$ is the subvector of $h(x)$ corresponding to the active constraints, i.e.

$$h_a(\tilde{x}) = \{h_i(\tilde{x}) \mid i \in I_a(\tilde{x}).$$

Definition 9 *A point $\tilde{x}$ is a regular point for the constraints if at $\tilde{x}$ the gradients of the active constraints,* i.e. *the vectors $\nabla g_i(\tilde{x})$, for $i = 1, \cdots, m$ and $\nabla h_i(\tilde{x})$, for $i \in I_a(\tilde{x})$, are linearly independent.*

The definition of regular point is given because, the necessary and the sufficient conditions for optimality, in the case of regular points are relatively simple. To state these conditions, and with reference to problem P_0, consider the Lagrangian function

$$L(x, \lambda, \rho) = f(x) + \lambda' g(x) + \rho' h(x) \tag{3.6}$$

with $\lambda \in I\!R^m$ and $\rho \in I\!R^p$. The vectors λ and ρ are called multipliers.
With the above ingredients and definitions it is now possible to provide a necessary condition and a sufficient condition for local optimality.

Theorem 14 *[First order necessary condition] Consider problem P_0. Suppose $x^\star$ is a local solution of the problem P_0, and $x^\star$ is a regular point for the constraints.*
Then there exist (unique) multipliers $\lambda^\star$ and $\rho^\star$ such that[2]

$$\begin{aligned} &\nabla_x L(x^\star, \lambda^\star, \rho^\star) = 0 \\ &g(x^\star) = 0 \\ &h(x^\star) \le 0 \\ &\rho^\star \ge 0 \\ &(\rho^\star)' h(x^\star) = 0. \end{aligned} \tag{3.7}$$

Conditions (3.7) are known as Kuhn-Tucker conditions.

Definition 10 *Let $x^\star$ be a local solution of problem P_0 and let $\rho^\star$ be the corresponding (optimal) multiplier. At $x^\star$ the condition of strict complementarity holds if $\rho_i^\star > 0$ for all $i \in I_a(x^\star)$.*

Theorem 15 *[Second order sufficient condition] Consider the problem P_0. Assume that there exist $x^\star$, $\lambda^\star$ and $\rho^\star$ satisfying conditions (3.7). Suppose moreover that $\rho^\star$ is such that the condition of strict complementarity holds at $x^\star$. Suppose finally that*

$$s' \nabla^2_{xx} L(x^\star, \lambda^\star, \rho^\star) s > 0 \tag{3.8}$$

for all $s \neq 0$ such that

$$\begin{bmatrix} \dfrac{\partial g(x^\star)}{\partial x} \\ \dfrac{\partial h_a(x^\star)}{\partial x} \end{bmatrix} s = 0.$$

Then $x^\star$ is a strict constrained local minimum of problem P_0.

Remark. Necessary and sufficient conditions for a global minimum can be given under proper convexity hypotheses, *i.e.* if the function f is convex in $\mathcal{X}$, and if $\mathcal{X}$ is a convex set. This is the case, for example if there are no inequality constraints and if the equality constraints are linear. ◇

[2] We denote with $\nabla_x f$ the vector of the partial derivatives of f with respect to x.

Remark. If all points in $\mathcal{X}$ are regular points for the constraints then conditions (3.7) yield a set of points $\mathcal{P}$, *i.e.* the points satisfying conditions (3.7), and among these points there are all constrained local minima (and also the constrained global minimum, if it exists). However, if there are points in $\mathcal{X}$ which are not regular points for the constraints, then the set $\mathcal{P}$ may not contain all constrained local minima. These have to be searched in the set $\mathcal{P}$ and in the set of non-regular points. $\diamond$

Remark. In what follows, we will always tacitly assume that the conditions of regularity and of strict complementarity hold. $\diamond$

3.2.1 A simple proof of Kuhn-Tucker conditions for equality constraints

Consider problem P_1, *i.e.* a minimization problem with only equality constraints, and a point $x^\star$ such that $g(x^\star) = 0$, *i.e.* $x^\star \in \mathcal{X}$. Suppose that[3]

$$\text{rank}\frac{\partial g}{\partial x}(x^\star) = m$$

i.e. $x^\star$ is a regular point for the constraints, and that $x^\star$ is a constrained local minimum. By the implicit function theorem, there exist a neighborhood of $x^\star$, a partition of the vector x, *i.e.*

$$x = \begin{bmatrix} u \\ v \end{bmatrix},$$

with $u \in I\!R^m$ and $v \in I\!R^{n-m}$, and a function ϕ such that the constrains $g(x) = 0$ can be (locally) rewritten as

$$u = \phi(v).$$

As a result (locally)

$$\begin{cases} \min\limits_x f(x) \\ g(x) = 0 \end{cases} \quad \Leftrightarrow \quad \begin{cases} \min\limits_{u,v} f(u,v) \\ u = \phi(v) \end{cases} \quad \Leftrightarrow \quad \min_v f(\phi(v), v),$$

i.e. problem P_1 is (locally) equivalent to a unconstrained minimization problem. Therefore

$$0 = \nabla f(\phi(v^\star), v^\star) = \left(\frac{\partial f}{\partial u}\frac{\partial \phi}{\partial v} + \frac{\partial f}{\partial v}\right)_{x^\star} = \left(-\frac{\partial f}{\partial u}\left(\frac{\partial g}{\partial u}\right)^{-1}\frac{\partial g}{\partial v} + \frac{\partial f}{\partial v}\right)_{x^\star}.$$

Setting

$$\lambda^\star = \left(-\frac{\partial f}{\partial u}\left(\frac{\partial g}{\partial u}\right)^{-1}\right)'_{x^\star}$$

yields

$$\left(\frac{\partial f}{\partial v} + (\lambda^\star)'\frac{\partial g}{\partial v}\right)_{x^\star} = 0 \tag{3.9}$$

[3] Note that m is the number of the equality constraints, and that, to avoid trivial cases, $m < n$.

and

$$\left(\frac{\partial f}{\partial u} + (\lambda^\star)' \frac{\partial g}{\partial u}\right)_{x^\star} = 0. \tag{3.10}$$

Finally, let

$$L = f + \lambda' g,$$

note that equations (3.9) and (3.10) can be rewritten as

$$\nabla_x L(x^\star, \lambda^\star) = 0,$$

and this, together with $g(x^\star) = 0$, is equivalent to equations (3.7).

3.2.2 Quadratic cost function with linear equality constraints

Consider the function

$$f(x) = \frac{1}{2} x' Q x,$$

with $x \in I\!R^n$ and $Q = Q' > 0$, the equality constraints

$$g(x) = Ax - b = 0,$$

with $b \in I\!R^m$ and $m < n$, and the Lagrangian function

$$L(x, \lambda) = \frac{1}{2} x' Q x + \lambda' (Ax - b).$$

A simple application of Theorem 14 yields the necessary conditions of optimality

$$\begin{array}{rcl} \nabla_x L(x^\star, \lambda^\star) & = & Qx^\star + A'\lambda^\star = 0 \\ g(x^\star) & = & Ax^\star - b = 0. \end{array} \tag{3.11}$$

Suppose now that the matrix A is such that $AQ^{-1}A'$ is invertible[4]. As a result, the only solution of equations (3.11) is

$$x^\star = Q^{-1}A'(AQ^{-1}A')^{-1}b \qquad \lambda^\star = -(AQ^{-1}A')^{-1}b.$$

Finally, by Theorem 15, it follows that $x^\star$ is a strict constrained (global) minimum.

3.3 Nonlinear programming methods: introduction

The methods of non-linear programming that have been mostly studied in recent years belong to two categories. The former includes all methods based on the transformation of a constrained problem into one or more unconstrained problems, in particular the so-called (exact or sequential) penalty function methods and (exact or sequential) augmented Lagrangian methods. Sequential methods are based on the solution of a sequence of problems, with the property that the sequence of the solutions of the subproblems converge

[4] This is the case if $\text{rank} A = m$.

to the solution of the original problem. Exact methods are based on the fact that, under suitable assumptions, the optimal solutions of an unconstrained problem coincides with the optimal solution of the original problem.
The latter includes the methods based on the transformation of the original problem into a sequence of constrained quadratic problems.
From the above discussion it is obvious that, to construct algorithms for the solution of non-linear programming problems, it is necessary to use efficient unconstrained optimization routines.
Finally, in any practical implementation, it is also important to quantify the complexity of the algorithms in terms of number and type of operations (inversion of matrices, differentiation, ...), and the speed of convergence. These issues are still largely open, and will not be addressed in these notes.

3.4 Sequential and exact methods

3.4.1 Sequential penalty functions

In this section we study the so-called external sequential penalty functions. This name is based on the fact that the solutions of the resulting unconstrained problems are in general not admissible. There are also internal penalty functions (known as barrier functions) but this can be used only for problems in which the admissible set has a non-empty interior. As a result, such functions cannot be used in the presence of equality constraints.
The basic idea of external sequential penalty functions is very simple. Consider problem P_0, the function

$$q(x) = \begin{cases} 0, & \text{if } x \in \mathcal{X} \\ \\ +\infty, & \text{if } x \notin \mathcal{X} \end{cases} \tag{3.12}$$

and the function

$$F = f + q. \tag{3.13}$$

It is obvious that the unconstrained minimization of F yields a solution of problem P_0. However, because of its discontinuous nature, the minimization of F cannot be performed. Nevertheless, it is possible to construct a sequence of continuously differentiable functions, converging to F, and it is possible to study the convergence of the minima of such a sequence of functions to the solutions of problem P_0.
For, consider a continuously differentiable function p such that

$$p(x) = \begin{cases} 0, & \text{if } x \in \mathcal{X} \\ \\ > 0, & \text{if } x \notin \mathcal{X}, \end{cases} \tag{3.14}$$

and the function

$$F_\epsilon = f + \frac{1}{\epsilon} p,$$

with $\epsilon > 0$. It is obvious that[5]

$$\lim_{\epsilon \to 0} F_\epsilon = F.$$

The function F_ϵ is known as external penalty function. The attribute external is due to the fact that, if $\bar{x}$ is a minimum of F_ϵ in general $p(\bar{x}) \neq 0$, *i.e.* $\bar{x} \notin \mathcal{X}$. The term $\frac{1}{\epsilon}p$ is called penalty term, as it penalizes the violation of the constraints. In general, the function p has the following form

$$p = \sum_{i=1}^{m} (g_i)^2 + \sum_{i=1}^{p} (\max(0, h_i))^2. \tag{3.15}$$

Consider now a strictly decreasing sequence $\{\epsilon_k\}$ such that $\lim_{k\to\infty} \epsilon_k = 0$. The sequential penalty function method consists in solving the sequence of unconstrained problems

$$\min_x F_{\epsilon_k}(x),$$

with $x \in I\!R^n$. The most important convergence results for this methods are summarized in the following statements.

Theorem 16 *Consider the problem P_0. Suppose that for all $\sigma > 0$ the set*[6]

$$\mathcal{X}^\sigma = \{x \in I\!R^n \mid |g_i(x)| \leq \sigma, i = 1, \cdots, m\} \cap \{x \in I\!R^n \mid h_i(x) \leq \sigma, i = 1, \cdots, p\}$$

is compact. Suppose moreover that for all k the function $F_{\epsilon_k}(x)$ has a global minimum x_k. Then the sequence $\{x_k\}$ has (at least) one converging subsequence, and the limit of any converging subsequence is a global minimum for problem P_0.

Theorem 17 *Let $x^\star$ be a strict constrained local minimum for problem P_0. Then there exist a sequence $\{x_k\}$ and an integer $\bar{k} > 0$ such that $\{x_k\}$ converges to $x^\star$ and, for all $k \geq \bar{k}$, x_k is a local minimum of $F_{\epsilon_k}(x)$.*

The construction of the function F_ϵ is apparently very simple, and this is the main advantage of the method. However, the minimization of the function F_ϵ may be difficult, especially for small values of ϵ. In fact, it is possible to show, even via simple examples, that as ϵ tends to zero the Hessian matrix of the function F_ϵ becomes ill conditioned. As a result, any unconstrained minimization algorithm used to minimize F_ϵ has a very slow convergence rate. To alleviate this problem, it is possible to use, in the minimization of $F_{\epsilon_{k+1}}$, as initial point the point x_k. However, this is close to the minimum of $F_{\epsilon_{k+1}}$ only if ϵ_{k+1} is close to ϵ_k, *i.e.* only if the sequence $\{\epsilon_k\}$ converges slowly to zero.
We conclude that, to avoid the ill conditioning of the Hessian matrix of F_ϵ, hence the slow convergence of each unconstrained optimization problem, it is necessary to slow down the convergence of the sequence $\{x_k\}$, *i.e.* slow convergence is an intrinsic property of the method. This fact has motivated the search for alternatives methods, as described in the next sections.

[5] Because of the discontinuity of F, the limit has to be considered with proper *care*.
[6] The set $\mathcal{X}^\sigma$ is sometimes called the relaxed admissible set.

Remark. It is possible to show that the local minima of F_ϵ describe (continuous) trajectories that can be extrapolated. This observation is exploited in some sophisticated methods for the selection of initial estimate for the point x_k. However, even with the addition of this extrapolation procedure, the convergence of the method remains slow. ◇

Remark. Note that, if the function p is defined as in equation (3.15), then the function F_ϵ is not two times differentiable everywhere, *i.e.* it is not differentiable in all points in which an inequality constraints is active. This property restricts the class of minimization algorithms that can be used to minimize F_ϵ. ◇

3.4.2 Sequential augmented Lagrangian functions

Consider problem P_1, *i.e.* an optimization problem with only equality constraints. For such a problem the Lagrangian function is

$$L = f + \lambda' g,$$

and the first order necessary conditions require the existence of a multiplier $\lambda^\star$ such that, for any local solution $x^\star$ of problem P_1 one has

$$\begin{aligned} \nabla_x L(x^\star, \lambda^\star) &= 0 \\ \nabla_\lambda L(x^\star, \lambda^\star) &= g(x^\star) = 0. \end{aligned} \tag{3.16}$$

The first of equations (3.16) is suggestive of the fact that the function $L(x, \lambda^\star)$ has a unconstrained minimum in $x^\star$. This is actually not the case, as $L(x, \lambda^\star)$ is not convex in a neighborhood of $x^\star$. However it is possible to render the function $L(x, \lambda^\star)$ convex with the addition of a penalty term, yielding the new function, known as augmented Lagrangian function[7],

$$L_a(x, \lambda^\star) = L(x, \lambda^\star) + \frac{1}{\epsilon}\|g(x)\|^2, \tag{3.17}$$

which, for ϵ sufficiently small, but such that $1/\epsilon$ is finite, has a unconstrained minimum in $x^\star$. This intuitive discussion can be given a formal justification, as shown in the next statement.

Theorem 18 *Suppose that at $x^\star$ and $\lambda^\star$ the sufficient conditions for a strict constrained local minimum for problem P_1 hold. Then there exists $\bar{\epsilon} > 0$ such that for any $\epsilon \in (0, \bar{\epsilon})$ the point $x^\star$ is a unconstrained local minimum for the function $L_a(x, \lambda^\star)$.*
Vice-versa, if for some $\bar{\epsilon}$ and $\lambda^\star$, at $x^\star$ the sufficient conditions for a unconstrained local minimum for the function $L_a(x, \lambda^\star)$ hold, and $g(x^\star) = 0$, then $x^\star$ is a strict constrained local minimum for problem P_1.

The above theorem highlights the fact that, under the stated assumptions, the function $L_a(x, \lambda^\star)$ is an (external) penalty function, with the property that, to give local minima

[7] To be precise we should write $L_a(x, \lambda^\star, \epsilon)$, however we omit the argument ϵ.

for problem P_1 it is not necessary that $\epsilon \to 0$. Unfortunately, this result is not of practical interest, because it requires the knowledge of $\lambda^\star$. To obtain a useful algorithm, it is possible to make use of the following considerations.
By the implicit function theorem, applied to the first of equation (3.16), we infer that there exist a neighborhood of $\lambda^\star$, a neighborhood of $x^\star$, and a continuously differentiable function $x(\lambda)$ such that (locally)

$$\nabla_x L_a(x(\lambda), \lambda) = 0.$$

Moreover, for any $\epsilon \in (0, \bar{\epsilon})$, as $\nabla^2_{xx} L_a(x^\star, \lambda^\star)$ is positive definite also $\nabla^2_{xx} L_a(x, \lambda)$ is locally positive definite. As a result, $x(\lambda)$ is the only value of x that, for any fixed λ, minimizes the function $L_a(x, \lambda)$. It is therefore reasonable to assume that if λ_k is a good estimate of $\lambda^\star$, then the minimization of $L_a(x, \lambda_k)$ for a sufficiently small value of ϵ, yields a point x_k which is a good approximation of $x^\star$.
On the basis of the above discussion it is possible to construct the following minimization algorithm for problem P_1.

Step 0. Given $x_0 \in I\!R^n$, $\lambda_1 \in I\!R^m$ and $\epsilon_1 > 0$.

Step 1. Set $k = 1$.

Step 2. Find a local minimum x_k of $L_a(x, \lambda_k)$ using any unconstrained minimization algorithm, with starting point x_{k-1}.

Step 3. Compute a new estimate λ_{k+1} of $\lambda^\star$.

Step 4. Set $\epsilon_{k+1} = \beta \epsilon_k$, with $\beta = 1$ if $\|g(x_{k+1})\| \leq \frac{1}{4}\|g(x_k)\|$ or $\beta < 1$ otherwise.

Step 5. Set $k = k + 1$ and go to **Step 2**.

In **Step 3** it is necessary to construct a new estimate λ_{k+1} of λ_k. This can be done with proper considerations on the function $L_a(x(\lambda), \lambda)$, introduced in the above discussion. However, without providing the formal details, we mention that one of the most used update laws for λ are described by the equations

$$\lambda_{k+1} = \lambda_k + \frac{2}{\epsilon_k} g(x_k), \tag{3.18}$$

or

$$\lambda_{k+1} = \lambda_k - \left[\nabla^2 L_a(x(\lambda_k), \lambda_k)\right]^{-1} g(x_k), \tag{3.19}$$

whenever the indicated inverse exists.
Note that the convergence of the sequence $\{x_k\}$ to $x^\star$ is limited by the convergence of the sequence $\{\lambda_k\}$ to $\lambda^\star$. It is possible to prove that, if the update law (3.18) is used then the algorithm as linear convergence, whereas if (3.19) is used the convergence is superlinear.

Remark. Similar considerations can be done for problem P_2. For, recall that problem P_2 can be recast, increasing the number of variables, as an optimization problem with equality constraints, *i.e.* problem $\tilde{P}_2$. For such an *extended* problem, consider the augmented Lagrangian

$$L_a(x, y, \rho) = f(x) + \sum_{i=1}^{p} \rho_i \left(h_i(x) + y_i^2 \right) + \frac{1}{\epsilon} \sum_{i=1}^{p} \left(h_i(x) + y_i^2 \right)^2,$$

and note that, in principle, it would be possible to make use of the results developed with reference to problem P_1. However, the function L_a can be analytically minimized with respect to the variables y_i. In fact, a simple computation shows that the (global) minimum of L_a as a function of y is attained at

$$y_i(x, \rho) = \sqrt{- \min \left(0, h_i(x) + \frac{\epsilon}{2} \rho_i \right)}.$$

As a result, the augmented Lagrangian function for problem P_2 is given by

$$L_a(x, \rho) = f(x) + \rho' h(x) + \frac{1}{\epsilon} \|h(x)\|^2 - \frac{1}{\epsilon} \sum_{i=1}^{p} \left(\min(0, h_i(x) + \frac{\epsilon}{2} \rho_i) \right)^2.$$

◇

3.4.3 Exact penalty functions

An exact penalty function, for a given constrained optimization problem, is a function of the same variables of the problem with the property that its unconstrained minimization yields a solution of the original problem. The term *exact* as opposed to *sequential* indicates that only one, instead of several, minimization is required.

Consider problem P_1, let $x^\star$ be a local solution and let $\lambda^\star$ be the corresponding multiplier. The basic idea of exact penalty functions methods is to determine the multiplier λ appearing in the augmented Lagrangian function as a function of x, *i.e.* $\lambda = \lambda(x)$, with $\lambda(x^\star) = \lambda^\star$. With the use of this function one has[8]

$$L_a(x, \lambda(x)) = f(x) + \lambda(x)' g(x) + \frac{1}{\epsilon} \|g(x)\|^2.$$

The function $\lambda(x)$ is obtained considering the necessary condition of optimality

$$\nabla_x L_a(x^\star, \lambda^\star) = \nabla f(x^\star) + \frac{\partial g(x^\star)'}{\partial x} \lambda^\star = 0 \tag{3.20}$$

and noting that, if $x^\star$ is a regular point for the constraints then equation (3.20) can be solved for $\lambda^\star$ yielding

$$\lambda^\star = - \left(\frac{\partial g(x^\star)}{\partial x} \frac{\partial g(x^\star)'}{\partial x} \right)^{-1} \frac{\partial g(x^\star)}{\partial x} \nabla f(x^\star).$$

[8] As in previous sections we omit the argument ϵ.

This equality suggests to define the function $\lambda(x)$ as

$$\lambda(x) = -\left(\frac{\partial g(x)}{\partial x}\frac{\partial g(x)'}{\partial x}\right)^{-1}\frac{\partial g(x)}{\partial x}\nabla f(x),$$

and this is defined at all x where the indicated inverse exists, in particular at $x^\star$.
It is possible to show that this selection of $\lambda(x)$ yields and exact penalty function for problem P_1. For, consider the function

$$G(x) = f(x) - g(x)'\left(\frac{\partial g(x)}{\partial x}\frac{\partial g(x)'}{\partial x}\right)^{-1}\frac{\partial g(x)}{\partial x}\nabla f(x) + \frac{1}{\epsilon}\|g(x)\|^2,$$

which is defined and differentiable in the set

$$\tilde{\mathcal{X}} = \{x \in I\!R^n \parallel \text{rank}\frac{\partial g(x)}{\partial x} = m\}, \tag{3.21}$$

and the following statements.

Theorem 19 *Let $\bar{\mathcal{X}}$ be a compact subset of $\tilde{\mathcal{X}}$. Assume that $x^\star$ is the only global minimum of f in $\mathcal{X} \cap \bar{\mathcal{X}}$ and that $x^\star$ is in the interior of $\bar{\mathcal{X}}$. Then there exists $\bar{\epsilon} > 0$ such that, for any $\epsilon \in (0, \bar{\epsilon})$, $x^\star$ is the only global minimum of G in $\bar{\mathcal{X}}$.*

Theorem 20 *Let $\bar{\mathcal{X}}$ be a compact subset of $\tilde{\mathcal{X}}$. Then there exists $\bar{\epsilon} > 0$ such that, for any $\epsilon \in (0, \bar{\epsilon})$, if $x^\star$ is a unconstrained minimum of $G(x)$ and $x^\star \in \bar{\mathcal{X}}$, then $x^\star$ is a constrained local minimum for problem P_1.*

Theorems 19 and 20 show that it is legitimate to search solutions of problem P_1 minimizing the function G for sufficiently small values of ϵ. Note that it is possible to prove stronger results, showing that there is (under certain hypotheses) a one to one correspondence between the minima of problem P_1 and the minima of the function G, provided ϵ is sufficiently small.
For problem P_2 it is possible to proceed as discussed in Section 3.4.2, *i.e.* transforming problem P_2 into problem $\tilde{P}_2$ and then minimizing analytically with respect to the new variables y_i. However, a different and more direct route can be taken. Consider problem P_2 and the necessary conditions

$$\nabla_x L(x^\star, \rho^\star) = \nabla f(x^\star) + \frac{\partial h(x^\star)'}{\partial x}\rho^\star = 0 \tag{3.22}$$

and

$$\rho_i^\star h_i(x^\star) = 0, \tag{3.23}$$

for $i = 1, \cdots, p$. Premultiplying equation (3.22) by $\frac{\partial h(x^\star)}{\partial x}$ and equation (3.23) by $\gamma^2 h_i(x^\star)$, with $\gamma > 0$, and adding, yields

$$\left(\frac{\partial h(x^\star)}{\partial x}\frac{\partial h(x^\star)'}{\partial x} + \gamma^2 H^2(x^\star)\right)\rho^\star + \frac{\partial h(x^\star)}{\partial x}\nabla f(x^\star) = 0,$$

where

$$H(x^\star) = \text{diag}(h_1(x^\star), \cdots, h_p(x^\star)).$$

As a result, a natural candidate for the function $\rho(x)$ is

$$\rho(x) = -\left(\frac{\partial h(x)}{\partial x}\frac{\partial h(x)'}{\partial x} + \gamma^2 H^2(x)\right)^{-1}\frac{\partial h(x)}{\partial x}\nabla f(x),$$

which is defined whenever the indicated inverse exists, in particular in the neighborhood of any regular point. With the use of this function, it is possible to define an exact penalty function for problem P_2 and to establish results similar to those illustrated in Theorems 19 and 20.

The exact penalty functions considered in this section provide, in principle, a theoretically sound way of solving constrained optimization problem. However, in practice, they have two major drawbacks. Firstly, at each step, it is necessary to invert a matrix with dimension equal to the number of constraint. This operation is numerically ill conditioned if the number of constraints is large. Secondly, the exact penalty functions may not be sufficiently regular to allow the use of unconstrained minimization methods with fast speed of convergence, *e.g.* Newton method.

3.4.4 Exact augmented Lagrangian functions

An exact augmented Lagrangian function, for a given constrained optimization problem, is a function, defined on an augmented space with dimension equal to the number of variables plus the number of constraint, with the property that its unconstrained minimization yields a solution of the original problem. Moreover, in the computation of such a function it is not necessary to invert any matrix.

To begin with, consider problem P_1 and recall that, for such a problem, a sequential augmented Lagrangian function has been defined adding to the Lagrangian function a term, namely $\frac{1}{\epsilon}\|g(x)\|^2$, which penalizes the violation of the necessary condition $g(x) = 0$. This term, for ϵ sufficiently small, renders the function L_a convex in a neighborhood of $x^\star$. A *complete* convexification can be achieved adding a further term that penalizes the violation of the necessary condition $\nabla_x L(x, \lambda) = 0$. More precisely, consider the function

$$S(x, \lambda) = f(x) + \lambda' g(x) + \frac{1}{\epsilon}\|g(x)\|^2 + \eta\|\frac{\partial g(x)}{\partial x}\nabla_x L(x, \lambda)\|^2, \tag{3.24}$$

with $\epsilon > 0$ and $\eta > 0$. The function (3.24) is continuously differentiable and it is such that, for ϵ sufficiently small, the solutions of problem P_1 are in a one to one correspondence with the points (x, λ) which are local minima of S, as detailed in the following statements.

Theorem 21 *Let $\bar{\mathcal{X}}$ be a compact set. Suppose $x^\star$ is the unique global minimum of f in the set $\mathcal{X} \cap \bar{\mathcal{X}}$ and $x^\star$ is an interior point of $\bar{\mathcal{X}}$. Let $\lambda^\star$ be the multiplier associated to $x^\star$. Then, for any compact set $\Lambda \subset I\!R^m$ such that $\lambda^\star \in \Lambda$ there exists $\bar{\epsilon}$ such that, for all $\epsilon \in (0, \bar{\epsilon})$, $(x^\star, \lambda^\star)$ is the unique global minimum of S in $\mathcal{X} \times \Gamma$.*

Theorem 22 *Let*[9] $\mathcal{X} \times \Lambda \subset \tilde{\mathcal{X}} \times I\!R^m$ *be a compact set. Then there exists* $\bar{\epsilon} > 0$ *such that, for all* $\epsilon \in (0, \bar{\epsilon})$*, if* $(x^\star, \lambda^\star)$ *is a unconstrained local minimum of* S*, then* $x^\star$ *is a constrained local minimum for problem* P_1 *and* $\lambda^\star$ *is the corresponding multiplier.*

Theorems 21 and 22 justify the use of a unconstrained minimization algorithm, applied to the function S, to find local (or global) solutions of problem P_1.
Problem P_2 can be dealt with using the same considerations done in Section 3.4.2.

3.5 Recursive quadratic programming

Recursive quadratic programming methods have been widely studied in the past years. In this section we provide a preliminary description of such methods. For, consider problem P_1 and suppose that $x^\star$ and $\lambda^\star$ are such that the necessary conditions (3.7) hold. Consider now a series expansion of the function $L(x, \lambda^\star)$ in a neighborhood of $x^\star$, *i.e.*

$$L(x, \lambda^\star) = f(x^\star) + \frac{1}{2}(x - x^\star)' \nabla^2_{xx} L(x^\star, \lambda^\star)(x - x^\star) + ...$$

a series expansion of the constraint, *i.e.*

$$0 = g(x) = g(x^\star) + \frac{\partial g(x^\star)}{\partial x}(x - x^\star) + ...$$

and the problem

$$\widetilde{PQ}_1 \quad \begin{cases} \min\limits_x f(x^\star) + \frac{1}{2}(x - x^\star)' \nabla^2_{xx} L(x^\star, \lambda^\star)(x - x^\star) \\ \frac{\partial g(x^\star)}{\partial x}(x - x^\star) = 0. \end{cases}$$

Note that problem $\widetilde{PQ}_1$ has the solution $x^\star$, and the corresponding multiplier is $\lambda = 0$, which is not equal (in general) to $\lambda^\star$. This phenomenon is called *bias* of the multiplier, and can be avoided by modifying the objective function and considering the new problem

$$PQ_1 \quad \begin{cases} \min\limits_x f(x^\star) + \nabla f(x^\star)'(x - x^\star) + \frac{1}{2}(x - x^\star)' \nabla^2_{xx} L(x^\star, \lambda^\star)(x - x^\star) \\ \frac{\partial g(x^\star)}{\partial x}(x - x^\star) = 0, \end{cases} \tag{3.25}$$

which has solution $x^\star$ with multiplier $\lambda^\star$. This observation suggests to consider the sequence of quadratic programming problems

$$PQ_1^{k+1} \quad \begin{cases} \min\limits_\delta f(x_k) + \nabla f(x_k)'\delta + \frac{1}{2}\delta' \nabla^2_{xx} L(x_k, \lambda_k)\delta \\ g(x_k) + \frac{\partial g(x_k)}{\partial x}\delta = 0, \end{cases} \tag{3.26}$$

[9] The set $\tilde{\mathcal{X}}$ is defined as in equation (3.21).

where $\delta = x - x_k$, and x_k and λ_k are the current estimates of the solution and of the multiplier. The solution of problem PQ_1^{k+1} yields new estimates x_{k+1} and λ_{k+1}. To assess the local convergence of the method, note that the necessary conditions for problem PQ_1^{k+1} yields the system of equations

$$\left[\begin{array}{cc} \nabla^2_{xx}L(x_k,\lambda_k) & \dfrac{\partial g(x_k)'}{\partial x} \\ \dfrac{\partial g(x_k)}{\partial x} & 0 \end{array}\right]\left[\begin{array}{c} \delta \\ \lambda \end{array}\right] = -\left[\begin{array}{c} \nabla f(x_k) \\ g(x_k) \end{array}\right], \tag{3.27}$$

and this system coincides with the system arising from the application of Newton method to the solution of the necessary conditions for problem P_1. As a consequence, the solutions of the problems PQ_1^{k+1} converge to a solution of problem P_1 under the same hypotheses that guarantee the convergence of Newton method.

Theorem 23 *Let $x^\star$ be a strict constrained local minimum for problem P_1, and let $\lambda^\star$ be the corresponding multiplier. Suppose that for $x^\star$ and $\lambda^\star$ the sufficient conditions of Theorem 15 hold. Then there exists an open neighborhood $\Omega \subset I\!R^n \times I\!R^m$ of the point $(x^\star,\lambda^\star)$ such that, if $(x_0,\lambda_0) \in \Omega$, the sequence $\{x_k,\lambda_k\}$ obtained solving the sequence of quadratic programming problems PQ_1^{k+1}, with $k = 0,1,\cdots$, converges to $(x^\star,\lambda^\star)$.*
Moreover, the speed of convergence is superlinear, and, if f and g are three times differentiable, the speed of convergence is quadratic.

Remark. It is convenient to solve the sequence of quadratic programming problems PQ_1^{k+1}, instead of solving the equations (3.27) with Newton method, because, for the former it is possible to exclude converge to maxima or saddle points. $\diamond$

In the case of problem P_2, using considerations similar to the one above, it is easy to obtain the following sequence of quadratic programming problems

$$PQ_2^{k+1} \quad \left\{ \begin{array}{l} \min\limits_{\delta} f(x_k) + \nabla f(x_k)'\delta + \dfrac{1}{2}\delta'\nabla^2_{xx}L(x_k,\lambda_k)\delta \\ \dfrac{\partial h(x_k)}{\partial x}\delta + h(x_k) \leq 0. \end{array} \right. \tag{3.28}$$

This sequence of problems has to be solved iteratively to generate a sequence $\{x_k,\lambda_k\}$ that, under hypotheses similar to those of Theorem 23, converges to a strict constrained local minimum of problem P_2.
The method described are the basis for a large class of iterative methods.
A first disadvantage of the proposed iterative schemes is that it is necessary to compute the second derivatives of the functions of the problem. This computation can be avoided, using the same philosophy of quasi-Newton methods.
A second disadvantage is in the fact that, being based on Newton algorithm, only local convergence can be guaranteed. However, it is possible to combine the method with global convergent methods: these are used to generate a pair $(\tilde{x},\tilde{\lambda})$ sufficiently close to $(x^\star,\lambda^\star)$

and then recursive quadratic programming methods are used to obtain fast convergence to $(x^\star, \lambda^\star)$.
A third disadvantage is in the fact that there is no guarantee that the sequence of admissible sets generated by the algorithm is non-empty at each step.
Finally, it is worth noting that, contrary to most of the existing methods, quadratic programming methods do not rely on the use of a penalty term.

Remark. There are several alternative recursive quadratic programming methods which alleviate the drawbacks of the methods described. These are (in general) based on the use of quadratic approximation of penalty functions. For brevity, we do not discuss these methods. $\diamond$

3.6 Concluding remarks

In this section we briefly summarize advantages and disadvantages of the nonlinear programming methods discussed.
Sequential penalty functions methods are very simple to implement, but suffer from the ill conditioning associated to large penalties (*i.e.* to small values of ϵ). As a result, these methods can be used if approximate solutions are acceptable, or in the determination of initial estimates for more precise, but only locally convergent, methods. Note, in fact, that not only an approximation of the solution $x^\star$ can be obtained, but also an approximation of the corresponding multiplier $\lambda^\star$. For example, for problem P_1, a (approximate) solution $\bar{x}$ is such that

$$\nabla F_{\epsilon_k}(\bar{x}) = \nabla f(\bar{x}) + \frac{2}{\epsilon_k}\frac{\partial g(\bar{x})}{\partial x} g(\bar{x}) = 0,$$

hence, the term $\frac{2}{\epsilon_k} g(\bar{x})$ provides an approximation of $\lambda^\star$.
Sequential augmented Lagrangian functions do not suffer from ill conditioning, and yield faster speed of convergence then that attainable using sequential penalty functions.
The methods based on exact penalty functions do not require the solution of a sequence of problems. However, they require the inversion of a matrix of dimension equal to the number of constraints, hence their applicability is limited to problems with a small number of constraints.
Exact augmented Lagrangian functions can be built without inverting any matrix. However, the minimization has to be performed in an extended space.
Recursive quadratic programming methods require the solution, at each step, of a constrained quadratic programming problem. Their main problem is that there is no guarantee that the admissible set is non-empty at each step.
We conclude that it is not possible to decide which is the best method. Each method has its own advantages and disadvantages. Therefore, the selection of a nonlinear programming method has to be driven by the nature of the problem: and has to take into consideration the number of variables, the regularity of the involved functions, the required precision, the computational cost, ...

3.7 Exercises

Similarly to Section 2.10, this section contains a set of exercises related to the notions, concepts, algorithms and tools discussed in Chapter 3. Since nonlinear programming is a widely studied, and complex, area of optimization, there are several methods and algorithms which are introduced only via exercises and which are not covered in details in the text. It is my hope that the exercises could form the starting point for additional reading.

Exercise 28 Consider the minimization problem

$$\begin{cases} \min\limits_{x_1,x_2} 1 - x_1^2 - x_2^2, \\ x_1 \geq 0, \\ x_2 \geq 0, \\ x_1 + x_2 - 1 \leq 0. \end{cases}$$

a) Show that all points in the admissible set are regular points for the constraints.

b) State the first order necessary conditions of optimality for such a constrained optimization problem.

c) Using the conditions derived in part b), compute candidate optimal solutions.

d) Show that the admissible set is compact. Hence deduce the existence of a global minimizer for the optimization problem. Determine the global minimizer of the problem. Is this minimizer unique?

Solution 28

a) The admissible set is the shaded area in the figure below. The arrows denote the gradient of the constraints on the boundary of the admissible set. As can be seen, these vectors are always independent, therefore all points are regular points for the constraints.

b) Define the Lagrangian

$$L = 1 - x_1^2 - x_2^2 + \rho_1(-x_1) + \rho_2(-x_2) + \rho_3(x_1 + x_2 - 1).$$

The first order necessary conditions of optimality are

$$\begin{array}{lll} -2x_1 - \rho_1 + \rho_3 = 0, & -2x_2 - \rho_2 + \rho_3 = 0, & \\ -x_1 \leq 0, & -x_2 \leq 0, & x_1 + x_2 - 1 \leq 0, \\ \rho_1 \geq 0, & \rho_2 \geq 0, & \rho_3 \geq 0, \\ -\rho_1 x_1 = 0, & -\rho_2 x_2 = 0, & \rho_3(x_1 + x_2 - 1) = 0. \end{array}$$

c) To compute candidate optimal solutions, note that from the last line of the necessary conditions we have the following possibilities:

$$\begin{array}{ll} \text{P1: } \rho_1 = 0,\ \rho_2 = 0,\ \rho_3 = 0; & \text{P2: } \rho_1 = 0,\ \rho_2 = 0,\ \rho_3 > 0; \\ \text{P3: } \rho_1 = 0,\ \rho_2 > 0,\ \rho_3 = 0; & \text{P4: } \rho_1 = 0,\ \rho_2 > 0,\ \rho_3 > 0; \\ \text{P5: } \rho_1 > 0,\ \rho_2 = 0,\ \rho_3 = 0; & \text{P6: } \rho_1 > 0,\ \rho_2 = 0,\ \rho_3 > 0; \\ \text{P7: } \rho_1 > 0,\ \rho_2 > 0,\ \rho_3 = 0; & \text{P8: } \rho_1 > 0,\ \rho_2 > 0,\ \rho_3 > 0. \end{array}$$

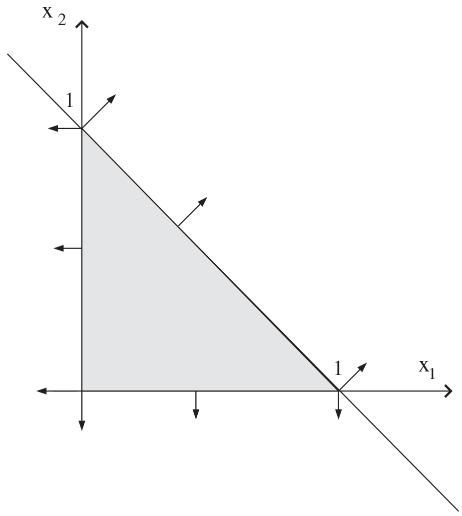
This yields the following candidate points

$$\begin{array}{ll} \text{P1: } x_1 = x_2 = 0 \text{ hence } f = 1; & \text{P2: } x_1 = x_2 = 1/2 \text{ hence } f = 1/2; \\ \text{P3: unfeasible;} & \text{P4: } x_1 = 1,\ x_2 = 0, \text{ hence } f = 0; \\ \text{P5: unfeasible;} & \text{P6: } x_1 = 0,\ x_2 = 1, \text{ hence } f = 0; \\ \text{P7: unfeasible;} & \text{P8: unfeasible.} \end{array}$$

As a result, we have only four candidate points:

$$P_1 = (0,0), \qquad P_2 = (1/2, 1/2), \qquad P_4 = (1,0), \qquad P_6 = (0,1).$$

d) The admissible set is closed (the constraints include the equality sign) and bounded (see the figure), hence compact. By Weierstrass theorem the function f has a global minimum in such a set. The function f attains its global minimum at P_4 and P_6, which are therefore both global minimizers. (This can be also shown noting that the problem is symmetric, i.e. changing x_1 into x_2 and x_2 into x_1 yields the same problem.)



Exercise 29 Consider the optimization problem

$$\begin{cases} \min\limits_{x_1,x_2} -x_1 - x_2 \\ x_1^2 + x_2^2 - 1 = 0 \end{cases}$$

a) Transform this minimization problem into an unconstrained minimization problem using the method of sequential penalty functions.

b) State the necessary conditions of optimality for the unconstrained problem of part a). Hence compute approximate candidate optimal solutions for the unconstrained optimization problem. Discuss the feasibility of these candidate optimal solutions.
(Hint: you may show that optimal points of the unconstrained problem are such that $x_1^\star = x_2^\star$. Moreover, use the fact that the solutions of $1+4x\dfrac{1-2x^2}{\epsilon} = 0$, for ϵ positive and small, are $\frac{\sqrt{2}}{2} + \frac{1}{8}\epsilon$, $-\frac{\sqrt{2}}{2} + \frac{1}{8}\epsilon$, $-\frac{1}{4}\epsilon$.)

c) Consider the stationary points of the sequential penalty function in part b). Consider the limit for $\epsilon \to 0$ of these stationary points and thus determine candidate optimal solutions for the original constrained optimization problem.

Solution 29

a) A sequential penalty function for the constrained problem is

$$F_\epsilon = -x_1 - x_2 + \frac{1}{\epsilon}(x_1^2 + x_2^2 - 1)^2.$$

b) The necessary conditions of optimality for F_ϵ are

$$0 = \nabla F_\epsilon = \begin{bmatrix} -1 + \frac{4x_1}{\epsilon}(x_1^2 + x_2^2 - 1) \\ \\ -1 + \frac{4x_2}{\epsilon}(x_1^2 + x_2^2 - 1) \end{bmatrix}.$$

As a result,

$$\begin{aligned} \frac{1}{x_1} &= \frac{4}{\epsilon}(x_1^2 + x_2^2 - 1) \\ \frac{1}{x_2} &= \frac{4}{\epsilon}(x_1^2 + x_2^2 - 1) \end{aligned}$$

yielding $x_1 = x_2$. Let $x_1 = x_2 = x$. From the first equation we have

$$\frac{1}{x} = \frac{4}{\epsilon}(2x^2 - 1) \Rightarrow 1 + 4x\frac{1 - 2x^2}{\epsilon} = 0.$$

As stated, this equation has approximate solutions

$$\frac{\sqrt{2}}{2} + \frac{1}{8}\epsilon, \qquad -\frac{\sqrt{2}}{2} + \frac{1}{8}\epsilon, \qquad -\frac{1}{4}\epsilon.$$

As a result, F_ϵ has three stationary points:

$$P_1 \approx \left(\frac{\sqrt{2}}{2} + \frac{1}{8}\epsilon, \frac{\sqrt{2}}{2} + \frac{1}{8}\epsilon\right), \qquad P_2 \approx \left(\frac{-\sqrt{2}}{2} + \frac{1}{8}\epsilon, \frac{-\sqrt{2}}{2} + \frac{1}{8}\epsilon\right), \qquad P_3 \approx \left(-\frac{1}{4}\epsilon, -\frac{1}{4}\epsilon\right).$$

Note that none of the above points is feasible, for any $\epsilon > 0$.

c) The stationary points of F_ϵ are such that

$$\lim_{\epsilon\to 0} P_1 = \left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right), \qquad \lim_{\epsilon\to 0} P_2 = \left(-\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}\right), \qquad \lim_{\epsilon\to 0} P_3 = (0, 0).$$

Hence, P_1 and P_2 converge to the admissible set, and P_1 is a (local) solution of the optimization problem considered.

Exercise 30 Consider the optimization problem

$$\begin{cases} \min\limits_{x_1,x_2} x_1^2 - x_2^2, \\ \\ x_1 - x_2^2 = 0. \end{cases}$$

a) Sketch in the (x_1, x_2)-plane the level lines of the function to be minimized and the admissible set. (Hint: plot the level lines corresponding to $x_1^2 - x_2^2 = 0$ and $x_1^2 - x_2^2 = \pm 4$.)

b) Using first order necessary conditions, compute candidate optimal solutions. Use second order sufficient conditions to decide which of the candidate points is a local minimizer or a local maximizer.

c) Compute an exact penalty function for the minimization problem and verify that the candidate optimal solutions determined in part b) are stationary points of the exact penalty function.

Solution 30

a) The level sets and the admissible set are depicted in the figure below.

b) Let

$$L(x, \lambda) = x_1^2 - x_2^2 + \lambda(x_1 - x_2^2).$$

The first order necessary conditions are

$$2x_1 + \lambda = 0, \qquad -2x_2 - 2\lambda x_2 = 0, \qquad x_1 - x_2^2 = 0,$$

and these yield the candidate optimal points

$$P_1 = (0,0), \qquad P_2 = (1/2, \sqrt{2}/2), \qquad P_3 = (1/2, -\sqrt{2}/2),$$

with corresponding multipliers $\lambda_1 = 0, \lambda_2 = -1, \lambda_3 = -1$. The second order sufficient conditions are $s'\nabla^2_{xx}L(x^\star, \lambda^\star)s > 0$ for $s \neq 0$ such that $[1, -2x_2^\star]s = 0$. For P_1 one has $s = [0,1]'$ and $s'(\nabla^2_{xx}L)s < 0$, hence P_1 is a local maximizer. For P_2 one has $s = [\sqrt{2}, 1]'$ and $s'(\nabla^2_{xx}L)s = 4$, and for P_3 one has $s = [\sqrt{2}, -1]'$ and $s'(\nabla^2_{xx})Ls = 4$. Hence, P_2 and P_3 are local minimizers.

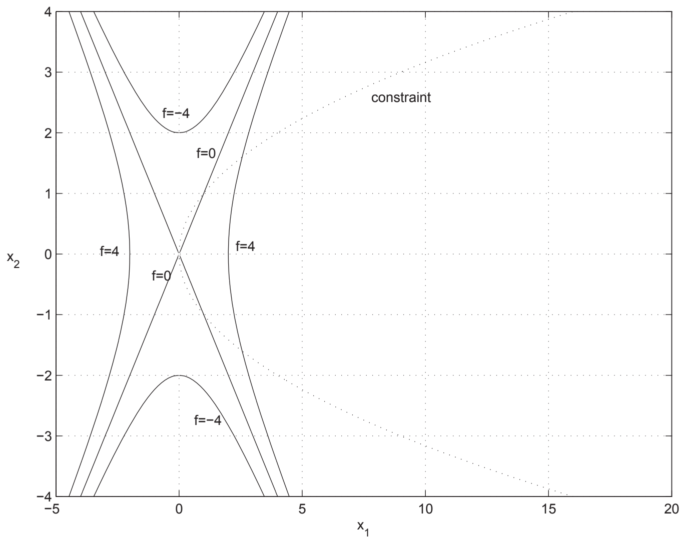
c) An exact penalty function for the considered problem is

$$G(x_1, x_2) = x_1^2 - x_2^2 - \frac{(2x_1 + 4x_2^2)(x_1 - x_2^2)}{1 + 4x_2^2} + \frac{(x_1 - x_2^2)^2}{\epsilon},$$

with $\epsilon > 0$. Its stationary points are the solutions of

$$0 = \nabla G(x_1, x_2) = \left[\begin{array}{c} 2\dfrac{-x_1\epsilon + 4x_1\epsilon x_2^2 - \epsilon x_2^2 + x_1 + 4x_1x_2^2 - x_2^2 - 4x_2^4}{(1+4x_2^2)\epsilon} \\ \\ 2x_2\dfrac{-\epsilon + 8\epsilon x_1^2 - 2x_1\epsilon - 2x_1 - 16x_1x_2^2 - 32x_1x_2^4 + 2x_2^2 + 16x_2^4 + 32x_2^6}{(1+4x_2^2)^2\epsilon} \end{array} \right].$$

By direct substitution we verify that, for any $\epsilon > 0$, the points P_1, P_2 and P_3 are stationary points of G.



Exercise 31 Consider the minimization problem

$$\begin{cases} \min\limits_{x_1,x_2} -x_1x_2, \\ 0 \leq x_1 + x_2 \leq 2, \\ -2 \leq x_1 - x_2 \leq 2, \end{cases}$$

a) Sketch in the (x_1, x_2)-plane the level surfaces of the function to be minimized and the admissible set. Hence show that all points in the admissible set are regular points for the constraints.

b) Using only graphical considerations determine the global solution of the considered problem.

c) State first order necessary conditions of optimality for such a constrained optimization problem. Show that the point determined in part b) satisfies first order necessary conditions of optimality, for some selection of the multiplier ρ.

d) Show that the point determined in part b) satisfies second order sufficient conditions of optimality for such a constrained optimization problem.

Solution 31

a) The admissible set is the shaded area in the figure below. The arrows denote the gradient of the constraints on the boundary of the admissible set. As can be seen, these vectors are always independent, therefore all points are regular points for the constraints. The dashed lines represent level lines of the function f.

b) From the figure it can be seen that the minimum is achieved when the level line of the function f is tangent to the admissible set, *i.e.* at the point $p = (1, 1)$.

c) Consider the Lagrangian

$$L = -x_1x_2 + \rho_1(-x_1 - x_2) + \rho_2(x_1 + x_2 - 2) + \rho_3(-2 - x_1 + x_2) + \rho_4(x_1 - x_2 - 2).$$

The first order sufficient conditions of optimality are

$$0 = \nabla_x L = \begin{bmatrix} -x_2 - \rho_1 + \rho_2 - \rho_3 + \rho_4 \\ -x_1 - \rho_1 + \rho_2 + \rho_3 - \rho_4 \end{bmatrix}$$

$$-x_1 - x_2 \leq 0, \qquad x_1 + x_2 - 2 \leq 0, \qquad -x_1 - x_2 - 2 \leq 0, \qquad x_1 - x_2 - 2 \leq 0,$$

$$\rho_1 \geq 0, \qquad \rho_2 \geq 0, \qquad \rho_3 \geq 0, \qquad \rho_4 \geq 0,$$

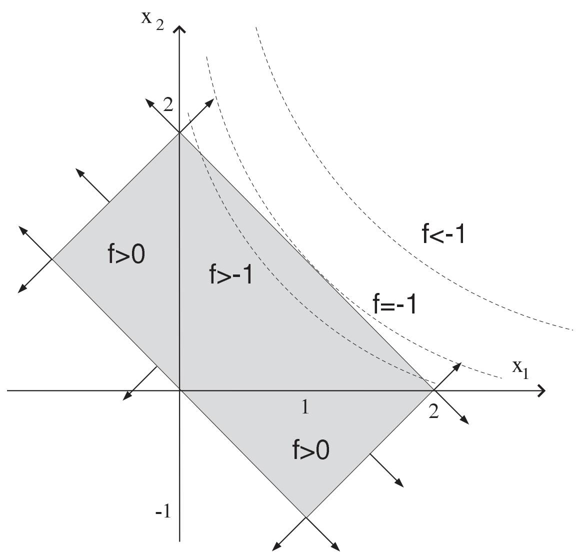
$$\rho_1(-x_1 - x_2) = 0, \qquad \rho_2(x_1 + x_2 - 2) = 0, \qquad \rho_3(-2 - x_1 + x_2) = 0, \qquad \rho_4(x_1 - x_2 - 2) = 0.$$

Setting $(x_1, x_2) = (1, 1)$ and selecting $\rho_1 = 0$, $\rho_3 = 0$, $\rho_4 = 0$ satisfies all the above equations. Hence, the point $(x_1, x_2) = (1, 1)$, together with the given multipliers, satisfies first order necessary conditions of optimality.

d) To check second order sufficient conditions note that for $(x_1, x_2) = (1, 1)$ the only active constraint is $x_1 + x_2 - 2 \leq 0$. Therefore we need to check positivity of $s'\nabla^2_{xx}Ls$ for $s = [s_1\ s_2]'$ such that $[1\ 1]s = 0$. This means $s_1 + s_2 = 0$, hence, solving for s_2, one has

$$s'\nabla^2_{xx}Ls = \begin{bmatrix} s_1 & -s_1 \end{bmatrix} \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} s_1 \\ -s_1 \end{bmatrix} = 2s_1^2 > 0$$

for $s_1 \neq 0$. As a result, the point obtained from graphical considerations in part b) is indeed a local minimizer for the considered problem.



Exercise 32 Consider the optimization problem

$$\begin{cases} \min\limits_{x_1,x_2,x_3} Tx_1^2 + Tx_2^2 + x_3^2, \\ x_1 + x_2 + x_3 - 1 = 0. \end{cases}$$

a) Transform this minimization problem into an unconstrained minimization problem by solving the constraint equation for x_3 and substituting the solution into the objective function.

b) Assume $T > 0$. Consider the unconstrained minimization problem determined in part a). Find (the unique) candidate optimal solution and show that this is indeed a local minimizer.

c) Assume $T > 0$. Exploit the results in parts a) and b) to determine the solution of the constrained optimization problem.

d) Assume $T > 0$. Consider the so-called l_1 penalty function

$$f_p = Tx_1^2 + Tx_2^2 + x_3^2 + \frac{|x_1 + x_2 + x_3 - 1|}{\epsilon},$$

with $\epsilon > 0$ and sufficiently small. Show that the unique stationary point of f_p coincides with the optimal solution determined in part c).
(Hint: recall that that $\frac{d|x|}{dx} = \text{sign}(x)$ and that $\text{sign}(0) \in [-1, 1]$. Moreover, use the fact that the stationary points of f_p do not depend upon the parameter ϵ.)

Solution 32

a) Solving the constrain equation for x_3 yields $x_3 = 1 - x_1 - x_2$. This is replaced in the function to minimize, hence resulting in the unconstrained minimization problem

$$\min_{x_1,x_2} \tilde{f}$$

with

$$\tilde{f} = Tx_1^2 + Tx_2^2 + (1 - x_1 - x_2)^2.$$

b) To determine candidate optimal solution consider the equations

$$0 = \nabla\tilde{f} = \begin{bmatrix} 2Tx_1 + 2x_1 + 2x_2 - 2 \\ 2Tx_2 + 2x_1 + 2x_2 - 2. \end{bmatrix}$$

These have the unique solution

$$x_1^\star = \frac{1}{T+2}, \qquad x_2^\star = \frac{1}{T+2}.$$

Note now that

$$\nabla^2\tilde{f} = \begin{bmatrix} 2T+2 & 2 \\ 2 & 2T+2 \end{bmatrix}$$

and this is positive definite for $T > 0$. Hence, the obtained stationary point is a local minimizer for $\tilde{f}$.

c) To obtain a solution of the original problem it is enough to compute

$$x_3^\star = 1 - x_1^\star - x_2^\star = \frac{T}{T+2}.$$

d) To compute the stationary points of f_p consider the equations

$$0 = \nabla f_p = \begin{bmatrix} 2Tx_1 + \dfrac{\text{sign}(x_1 + x_2 + x_3 - 1)}{\epsilon} \\ 2Tx_2 + \dfrac{\text{sign}(x_1 + x_2 + x_3 - 1)}{\epsilon} \\ 2x_3 + \dfrac{\text{sign}(x_1 + x_2 + x_3 - 1)}{\epsilon} \end{bmatrix}.$$

These can be rewritten as

$$2Tx_1 = 2Tx_2 = 2x_3 = -\frac{\mathrm{sign}(x_1 + x_2 + x_3 - 1)}{\epsilon}$$

yielding $x_1 = x_3/T$ and $x_2 = x_3/T$. Replacing this in the last equation yields

$$2x_3 = -\frac{\mathrm{sign}(x_3/T + x_3/T + x_3 - 1)}{\epsilon}.$$

Note now that the solution of this equation may be independent of ϵ only if $x_3/T + x_3/T + x_3 - 1 = 0$, implying $x_3 = x_3^\star$. Finally, this implies that $x_1 = x_1^\star$ and $x_2 = x_2^\star$, *i.e.* the unique stationary point of f_p coincides with the optimal solution obtained in part c).

Exercise 33 Consider the optimization problem

$$\begin{cases} \min\limits_{x_1, x_2} x_1 x_2, \\ x_1^2 + x_2^2 - 1 \leq 0. \end{cases}$$

a) State first order necessary conditions of optimality for such a constrained optimization problem.

b) Using the conditions derived in part a), compute candidate optimal solutions.

c) This constrained optimization problem can be transformed into an unconstrained optimization problem by defining the so-called barrier function

$$B_\epsilon(x) = x_1 x_2 + \frac{\epsilon}{1 - x_1^2 - x_2^2},$$

with $\epsilon > 0$, and considering the unconstrained minimization of $B_\epsilon(x)$. Determine the stationary points x_ϵ of $B_\epsilon(x)$.
(Hint: show that all stationary points $\bar{x} = (\bar{x}_1, \bar{x}_2)$ are such that $\bar{x}_1 = -\bar{x}_2$, and then note that the solutions of the equation

$$x - \frac{2\epsilon x}{(2x^2 - 1)^2} = 0$$

are $x = 0$ and $x = \pm\frac{\sqrt{2 \pm 2\sqrt{2\epsilon}}}{2}$.)
Discuss the feasibility of the obtained stationary points x_ϵ. Compute $\lim\limits_{\epsilon \to 0} x_\epsilon$ and compare this result with the results obtained in parts a) and b).

d) Discuss the advantages and disadvantages of the proposed barrier function method in comparison with the sequential penalty function method discussed in Section 3.4.1.

Solution 33

a) The Lagrangian of the problem is

$$L = x_1 x_2 + \rho(x_1^2 + x_2^2 - 1).$$

The first order necessary conditions of optimality are

$$0 = \nabla_x L = \begin{bmatrix} x_2 + 2\rho x_1 \\ x_1 + 2\rho x_2 \end{bmatrix},$$

$$x_1^2 + x_2^2 - 1 \leq 0, \qquad \rho \geq 0, \qquad (x_1^2 + x_2^2 - 1)\rho = 0.$$

b) From the first two equations we have that if $\rho \neq 1/2$ then $x_1 = x_2 = 0$. If $\rho = 1/2$ then $x_1 + x_2 = 0$ and from the last equation $x_1^2 + x_2^2 - 1 = 0$. As a result $x_1 = \pm\frac{1}{\sqrt{2}}$ and $x_2 = \mp\frac{1}{\sqrt{2}}$. In conclusion we have three candidate solutions

$$P_1 = (0, 0), \qquad P_2 = \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right), \qquad P_3 = \left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right).$$

c) To determine the stationary points of B_ϵ, consider the equation

$$0 = \nabla_x B_\epsilon = \begin{bmatrix} x_2 + 2\epsilon \dfrac{x_1}{(1-x_1^2-x_2^2)^2} \\ x_1 + 2\epsilon \dfrac{x_2}{(1-x_1^2-x_2^2)^2} \end{bmatrix}.$$

From these we obtain

$$\begin{aligned} \tfrac{x_2}{x_1} &= -2\epsilon \frac{1}{(1-x_1^2-x_2^2)^2}, \\ \tfrac{x_1}{x_2} &= -2\epsilon \frac{1}{(1-x_1^2-x_2^2)^2}, \end{aligned}$$

hence $x_1/x_2 < 0$ and $x_2/x_1 = x_1/x_2$. As a result $x_1 = -x_2$. Replacing in the second equation yields

$$x_1 = 2\epsilon \frac{x_1}{(1-2x_1^2)^2}$$

hence we obtain five candidate solutions, namely

$$P_a = (0,0), \qquad P_b = \left(\frac{\sqrt{2+2\sqrt{2\epsilon}}}{2}, -\frac{\sqrt{2+2\sqrt{2\epsilon}}}{2} \right), \qquad P_c = \left(-\frac{\sqrt{2+2\sqrt{2\epsilon}}}{2}, \frac{\sqrt{2+2\sqrt{2\epsilon}}}{2} \right),$$

$$P_d = \left(\frac{\sqrt{2-2\sqrt{2\epsilon}}}{2}, -\frac{\sqrt{2-2\sqrt{2\epsilon}}}{2} \right), \qquad P_e = \left(-\frac{\sqrt{2-2\sqrt{2\epsilon}}}{2}, \frac{\sqrt{2-2\sqrt{2\epsilon}}}{2} \right).$$

Note that P_a, P_d and P_e are feasible, whereas P_b and P_c are not feasible. Finally $P_a = P_1$,

$$\lim_{\epsilon\to 0} P_b = \lim_{\epsilon\to 0} P_d = P_2$$

and

$$\lim_{\epsilon\to 0} P_c = \lim_{\epsilon\to 0} P_e = P_3.$$

d) The proposed method is preferable to the sequential penalty function method because it provides feasible solutions also for $\epsilon > 0$. However, the function B_ϵ is not defined on all $I\!R^2$, hence it may be difficult to perform a numerical minimization.

Exercise 34 Let $Q \in I\!R^{n\times n}$ with $Q = Q' > 0$, $x \in I\!R^n$, $A \in I\!R^{m\times n}$ $b \in I\!R^m$ and $y \in I\!R^m$. Consider the minimization problem

$$P : \begin{cases} \min\limits_x \dfrac{1}{2} x'Qx, \\ Ax - b \le 0, \end{cases}$$

and the so-called dual problem

$$D : \begin{cases} \min\limits_y \dfrac{1}{2} y'AQ^{-1}A'y + b'y, \\ -y \le 0. \end{cases}$$

a) Write first order necessary conditions of optimality for the problem P. (Denote the multiplier with ρ.)

b) Write first order necessary conditions of optimality for the problem D. (Denote the multiplier with σ.)

c) Let $y_\star$ and $\sigma_\star$ be such that the optimality conditions in part b) hold. Show that

$$x_\star = -Q^{-1}A'y_\star, \qquad \rho_\star = y_\star,$$

are such that the optimality conditions in part a) hold.

d) Consider the minimization problem

$$P_1 : \begin{cases} \min_x \frac{1}{2}x'x, \\ x_1 + 1 \leq 0, \end{cases}$$

with $x \in I\!R^n$ and $x = [x_1, x_2, \cdots, x_n]'$. Exploiting the results above solve this problem. (Hint: write the dual D_1 of problem P_1, solve problem D_1, and then obtain a solution to problem P_1 exploiting the results in part c).)

Solution 34

a) Let $L_P = \frac{1}{2}x'Qx + \rho'(Ax - b)$ be the Lagrangian for problem P. The first order necessary conditions of optimality for problem P are

$$Qx_\star + A'\rho_\star = 0, \qquad Ax_\star - b \leq 0, \qquad \rho_\star \geq 0, \qquad \rho_\star'(Ax_\star - b) = 0.$$

b) Let $L_D = \frac{1}{2}y'AQ^{-1}A'y + b'y + \sigma'(-y)$ be the Lagrangian for problem D. The first order necessary conditions of optimality for problem D are

$$AQ^{-1}A'y_\star + b - \sigma_\star = 0, \qquad -y_\star \leq 0, \qquad \sigma_\star \geq 0, \qquad \sigma_\star'(-y_\star) = 0.$$

c) Replacing $x_\star = -Q^{-1}A'y_\star$ and $\rho_\star = y_\star$ in the equations in part a) yields

$$\begin{array}{rcl} Q(-Q^{-1}A'y_\star) + A'y_\star & = & 0, \\ A(-Q^{-1}A'y_\star) - b & \leq & 0, \\ y_\star & \geq & 0, \\ y_\star'(A(-Q^{-1}A'y_\star) - b) & = & 0. \end{array}$$

The first of the above equations holds trivially. For the second one note that

$$A(-Q^{-1}A'y_\star) - b = -\sigma_\star \leq 0,$$

by the third of the equations in b). The third equation holds by the second of the equations in b). The fourth equation holds exploiting the first and the fourth of the equations in b), hence we conclude the claim.

d) Problem P_1 is of the form of problem P with $Q = I$, $A = [1, 0, \cdots, 0]$ and $b = -1$. Hence, the dual D_1 is

$$D_1 : \begin{cases} \min_y \frac{1}{2}y^2 - y, \\ -y \leq 0, \end{cases}$$

with $y \in I\!R$. The problem D_1 has the solution $y_\star = 1$ and $\sigma_\star = 0$. Hence, the solution to problem P_1 is

$$x_\star = -[1, 0, \cdots, 0]', \qquad \rho_\star = 1.$$

Exercise 35 Consider the optimization problem

$$\begin{cases} \min_{x_1,\, x_2} x_1^2 + x_2, \\ x_1^2 + (x_2 - 1)^2 = 4. \end{cases}$$

a) Sketch in the (x_1, x_2)-plane the level lines of the function to be minimized and the admissible set. Hence show that all points in the admissible set are regular points for the constraints.

b) Using only graphical considerations determine the solution of the considered problem.

c) Show that the considered problem can be solved by eliminating the variable x_1 and obtaining the optimization problem

$$\begin{cases} \min_{x_2} 4 - (x_2 - 1)^2 + x_2 \\ -1 \leq x_2 \leq 3. \end{cases}$$

d) Solve the optimization problem in part c) and hence obtain a solution for the considered optimization problem.

e) Suppose that one wants to solve the considered optimization problem using recursive quadratic programming methods. Write the quadratic programming problem associated with the considered optimization problem.

Solution 35

a) The level lines and the admissible set are depicted in the figure below. Note that the constraint is always active, and that the gradient of the constraint is never zero, hence all points are regular points.

b) The solution of the problem is obtained when the level set of f is tangent to the admissible set in its lower point. As a result, the optimal point is $(x_1, x_2) = (0, -1)$.

c) We can solve the constraint yielding

$$x_1^2 = 4 - (x_2 - 1)^2.$$

Replacing in f we obtain the function to minimize

$$\tilde{f} = 4 - (x_2 - 1)^2 + x_2.$$

Note that x_2 is not *free*. In fact, from the constraint

$$(x_2 - 1)^2 = 4 - x_1^2 \leq 4$$

we obtain

$$-1 \leq x_2 \leq 3.$$

This shows that eliminating the variable x_1 yields the constrained scalar problem given in part b).

d) Note that a solution to the minimization problem in part c) is obtained at a stationary point of $\tilde{f}$ or at the boundary of the admissible set. The function $\tilde{f}$ has a stationary point (a local maximizer) for $x_2 = 3/2$. Note now that

$$\tilde{f}(-1) = -1, \qquad \tilde{f}(3/2) = 21/4, \qquad \tilde{f}(3) = 3.$$

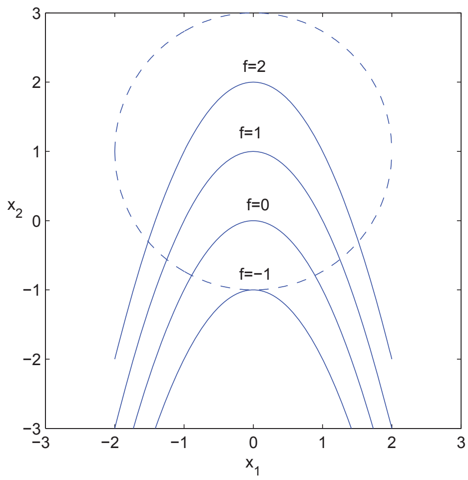
Therefore the function $\tilde{f}$ attains its minimum for $x_2 = -1$. Replacing this into the constraint yields $x_1 = 0$ and this coincides with the optimal solution obtained in part b).

e) Consider the optimization problem $\min_x f(x)$ subject to $g(x) = 0$. Using recursive quadratic programming methods to solve this problem one obtains the quadratic programming problem

$$PQ_1^{k+1} \begin{cases} \min_{\delta} f(x_k) + \nabla f(x_k)'\delta + \frac{1}{2}\delta' \nabla^2_{xx} L(x_k, \lambda_k)\delta, \\ \frac{\partial g(x_k)}{\partial x}\delta = 0, \end{cases}$$

where $L = f + \lambda' g$, $\delta = x - x_k$, and x_k and λ_k are the current estimates of the solution and of the multiplier. For the specific example one has to replace the functions f and g in the above expression yielding

$$PQ_1^{k+1} \begin{cases} \min_{\delta_1, \delta_2} x_{1,k}^2 + x_{2,k} + 2x_{1,k}\delta_1 + \delta_2 + \delta_1^2, \\ 2x_{1,k}\delta_1 + 2(x_{2,k} - 1)\delta_2 = 0. \end{cases}$$



Exercise 36 Consider the optimization problem

$$\begin{cases} \displaystyle\max_{x_1,\ x_2,\ x_3} \ x_1^\alpha \ x_2^\alpha \ x_3^\alpha \\ x_1 + x_2 + x_3 - 1 = 0 \\ x_1 \geq 0 \\ x_2 \geq 0 \\ x_3 \geq 0, \end{cases}$$

with $\alpha > 1$.

a) State first order necessary conditions of optimality for this constrained optimization problem.

b) Using the conditions derived in part a), compute candidate optimal solutions. Show that there is only one candidate solution such that $x_1 \neq 0$, $x_2 \neq 0$ and $x_3 \neq 0$.

c) Consider the candidate optimal solution with $x_1 \neq 0$, $x_2 \neq 0$ and $x_3 \neq 0$ determined in part b). Show using second order sufficient conditions that such a candidate optimal point is a local maximizer.

Solution 36

a) Let (note the change of sign in the objective function to transform the problem into a minimization problem)

$$L = -x_1^\alpha x_2^\alpha x_3^\alpha + \lambda(x_1 + x_2 + x_3 - 1) + \mu_1(-x_1) + \mu_2(-x_2) + \mu_3(-x_3).$$

The first order necessary conditions of optimality for the problem are

$$-\alpha x_1^{\alpha-1} x_2^\alpha x_3^\alpha + \lambda - \mu_1 = 0, \qquad -\alpha x_1^\alpha x_2^{\alpha-1} x_3^\alpha + \lambda - \mu_2 = 0, \qquad -\alpha x_1^\alpha x_1^\alpha x_3^{\alpha-1} + \lambda - \mu_3 = 0,$$

$$x_1 + x_2 + x_3 - 1 = 0, \qquad -x_1 \leq 0, \qquad -x_2 \leq 0, \qquad -x_3 \leq 0,$$

$$\mu_1 \geq 0, \qquad \mu_2 \geq 0, \qquad \mu_3 \geq 0, \qquad x_1\mu_1 = 0, \qquad x_2\mu_2 = 0, \qquad x_3\mu_3 = 0.$$

b) Consider the condition $x_1\mu_1 = 0$. This implies $\mu_1 = 0$ or $x_1 = 0$. If $x_1 = 0$ then, since $\alpha > 1$,

$$\lambda = \mu_1 = \mu_2 = \mu_3 = \kappa \geq 0,$$

for some constant κ. If $\kappa > 0$ then $x_2 = 0$ and $x_3 = 0$ which is not feasible. If $\kappa = 0$ then any x_2 and x_3 such that

$$x_2 + x_3 - 1 = 0, \qquad x_2 \geq 0, \qquad x_3 \geq 0,$$

satisfy necessary conditions of optimality. We obtain similar conclusions from the conditions $x_2\mu_2 = 0$ and $x_3\mu_3 = 0$. To obtain other candidate solutions we have to consider the case $x_1 \neq 0$, $x_2 \neq 0$ and $x_3 \neq 0$. In this case, $\mu_1 = \mu_2 = \mu_3 = 0$, and

$$\alpha x_1^{\alpha-1} x_2^\alpha x_3^\alpha - \lambda = 0,$$
$$\alpha x_1^\alpha x_2^{\alpha-1} x_3^\alpha - \lambda = 0,$$
$$\alpha x_1^\alpha x_1^\alpha x_3^{\alpha-1} - \lambda = 0.$$

The above equations imply $x_1 = x_2 = x_3$ which, together with the constraint $x_1 + x_2 + x_3 - 1 = 0$, yields the candidate optimal solution $(x_1, x_2, x_3) = (1/3, 1/3, 1/3)$. In summary, all candidate optimal solutions are

$$x_1 = x_2 = x_3 = 1/3$$
$$x_1 = 0,\ x_2 + x_3 = 1,\ x_2 \geq 0,\ x_3 \geq 0,$$
$$x_2 = 0,\ x_1 + x_3 = 1,\ x_1 \geq 0,\ x_3 \geq 0,$$
$$x_3 = 0,\ x_1 + x_2 = 1,\ x_1 \geq 0,\ x_2 \geq 0.$$

c) At the point $(x_1, x_2, x_3) = (1/3, 1/3, 1/3)$ the only active constraint is the equality constraint. Hence, the second order sufficient conditions of optimality are $s'\nabla^2 L s > 0$, for all $s \neq 0$ such that $[1,\ 1,\ 1]s = 0$. Note now that at the considered point

$$\nabla^2 L = -\left(\frac{1}{3}\right)^{3\alpha-2} \begin{bmatrix} \alpha(\alpha-1) & \alpha^2 & \alpha^2 \\ \alpha^2 & \alpha(\alpha-1) & \alpha^2 \\ \alpha^2 & \alpha^2 & \alpha(\alpha-1) \end{bmatrix}$$

and the admissible s can be parameterized as

$$s_a = [\beta, 0, -\beta]' \qquad s_b = [\gamma, -\gamma, 0]'.$$

As a result

$$s_a'\nabla^2 L s_a = 2\left(\frac{1}{3}\right)^{3\alpha-2} \beta^2\alpha > 0, \qquad s_b'\nabla^2 L s_b = 2\left(\frac{1}{3}\right)^{3\alpha-2} \gamma^2\alpha > 0,$$

which show that the considered point is a local minimizer (hence a maximizer for the original problem).

Exercise 37 Consider the problem of approximating a matrix $Q \in I\!R^{n\times n}$ with a matrix of the form $A = \rho I$, with I the identity matrix of dimension $n \times n$ and $\rho \geq 0$.
As a measure of the distance between the two matrices we could use either the square of the Frobenius norm or the infinity norm. The Frobenius norm of a matrix $L \in I\!R^{n\times n}$ is defined as

$$\|L\|_F = \sqrt{\sum_{i=1}^n \sum_{j=1}^n L_{ij}^2},$$

where the L_{ij}'s denote the entry of the matrix L. The infinity norm of a matrix $L \in I\!R^{n\times n}$ is defined as

$$\|L\|_\infty = \max_i \sum_{j=1}^n |L_{ij}|.$$

The optimal approximation problem is thus the problem of determining the nonnegative constant ρ which minimizes

$$\|Q - \rho I\|_F^2$$

or

$$\|Q - \rho I\|_\infty.$$

a) Show that the considered optimal approximation problems can be written as constrained minimization problems with one inequality constraint.

b) Consider the Frobenius norm. Solve the problem derived in part a). Show that if $\text{trace}(Q) > 0$ then the optimal ρ is positive, and if $\text{trace}(Q) \leq 0$ then the optimal ρ is zero.

c) Consider the infinity norm and assume that $n = 2$, hence

$$Q = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix},$$

that $0 < Q_{11} < Q_{22}$ and that $|Q_{12}| = |Q_{21}|$.

i) Sketch the graph of the function to be minimized.

ii) Argue that the optimal solution $\rho_\star$ is such that

$$0 < Q_{11} < \rho_\star < Q_{22}.$$

iii) Compute the optimal solution $\rho_\star$.

Solution 37

a) The optimal approximation problems can be written as

$$P_f : \begin{cases} \min_\rho \|Q - \rho I\|_F^2, \\ \rho \geq 0, \end{cases} \qquad \text{or} \qquad P_\infty : \begin{cases} \min_\rho \|Q - \rho I\|_\infty, \\ \rho \geq 0. \end{cases}$$

b) Note that

$$\begin{aligned} \|Q - \rho I\|_F^2 &= (Q_{11} - \rho)^2 + Q_{12}^2 + \cdots + Q_{1n}^2 + \\ &\quad Q_{21}^2 + (Q_{22} - \rho)^2 + Q_{23}^2 + \cdots + Q_{2n}^2 + \cdots + \\ &\quad Q_{n1}^2 + \cdots + Q_{2,n-1}^2 + (Q_{nn} - \rho)^2 \end{aligned}$$

hence

$$\|Q - \rho I\|_F^2 = n\rho^2 - 2\rho \overbrace{(Q_{11} + Q_{22} + \cdots + Q_{nn})}^{\text{trace}(Q)} + \text{constant terms}.$$

If $\text{trace}(Q) > 0$ the function $\|Q - \rho I\|_F^2$, which is convex, has a global minimizer for $\rho = \dfrac{\text{trace}(Q)}{n}$. If $\text{trace}(Q) \leq 0$ the function $\|Q - \rho I\|_F^2$ is monotonically increasing for $\rho \geq 0$, hence it achieves its minimum, in the set $\rho \geq 0$, for $\rho = 0$.

c) The optimal approximation problem is now

$$\tilde{P}_\infty : \begin{cases} \min_\rho \Big(\max \big(|Q_{11} - \rho| + |Q_{12}|, |Q_{21}| + |Q_{22} - \rho| \big) \Big), \\ \rho \geq 0. \end{cases}$$

A sketch of the function to be minimized is in the figure below. From this it is clear that $0 < Q_{11} < \rho_\star < Q_{22}$. Note that $\rho_\star$ is such that

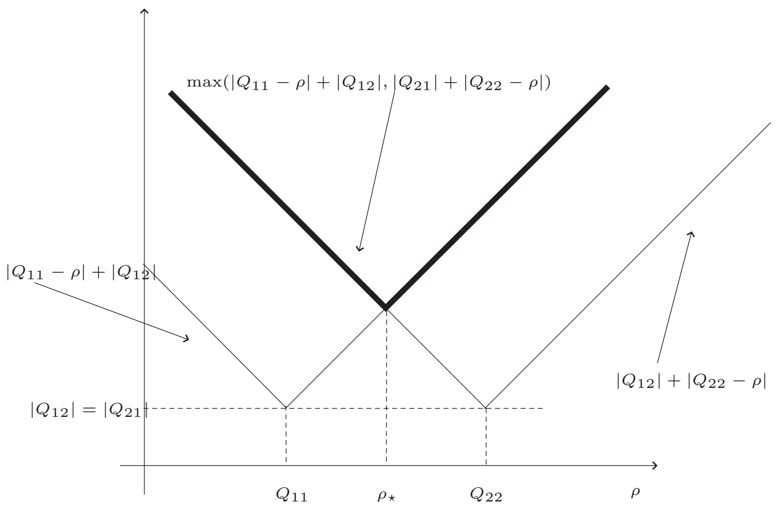
$$|Q_{11} - \rho_\star| + |Q_{12}| = |Q_{21}| + |Q_{22} - \rho_\star|.$$

However, because $0 < Q_{11} < \rho_\star < Q_{22}$ this can be rewritten as

$$\rho_\star - |Q_{11}| + |Q_{12}| = |Q_{21}| + |Q_{22}| - \rho_\star.$$

As a result (recall that $Q_{11} > 0$, $Q_{22} > 0$ and $|Q_{12}| = |Q_{21}|$)

$$\rho_\star = \frac{Q_{11} + Q_{22}}{2}.$$



Exercise 38 Consider the optimization problem

$$\begin{cases} \min\limits_{x_1,x_2} x_1^2 + x_2, \\ x_1^2 + (x_2-1)^2 \geq 1, \\ x_1^2 + (x_2-2)^2 \leq 4. \end{cases}$$

a) Sketch in the (x_1, x_2)-plane the admissible set and show that there is a point which is not a regular point for the constraints.
b) State first order necessary conditions of optimality for such a constrained optimization problem.
c) Find candidate optimal solutions for the considered problem.
d) Prove that the non-regular point for the constraints is the global minimizer for the considered problem.

Solution 38

a) The admissible set is the set outside a circle of radius one and centered at $(0, 1)$ and inside a circle of radius two and centered at $(0, 2)$, which is the shaded region in the figure below. The point $(0, 0)$ is not a regular point for the constraints because at this point both constraints are active and their gradients, namely

$$\begin{bmatrix} 2x_1 \\ 2(x_2-1) \end{bmatrix}, \qquad \begin{bmatrix} 2x_1 \\ 2(x_2-2) \end{bmatrix},$$

evaluated at the point, are linearly dependent.

b) To write necessary conditions of optimality rewrite first the constraints as

$$1 - x_1^2 - (x_2-1)^2 \leq 0 \qquad x_1^2 + (x_2-2)^2 - 4 \leq 0$$

and define the Lagrangian function

$$L(x_1, x_2, \mu_1, \mu_2) = x_1^2 + x_2 + \mu_1(1 - x_1^2 - (x_2-1)^2) + \mu_2(x_1^2 + (x_2-2)^2 - 4).$$

The necessary conditions of optimality are

$$\begin{aligned} \frac{dL}{dx_1} &= 2x_1 - 2\mu_1 x_1 + 2\mu_2 x_1 = 0, & \frac{dL}{dx_2} &= 1 - 2\mu_1(x_2-1) + 2\mu_2(x_2-2) = 0, \\ 1 - x_1^2 - (x_2-1)^2 &\leq 0, & x_1^2 + (x_2-2)^2 - 4 &\leq 0, \\ \mu_1 &\geq 0, & \mu_2 &\geq 0, \\ \mu_1(1 - x_1^2 - (x_2-1)^2) &= 0, & \mu_2(x_1^2 + (x_2-2)^2 - 4) &= 0. \end{aligned}$$

c) To find candidate optimal solutions we exploit the complementarity conditions, hence we have four possibilities.

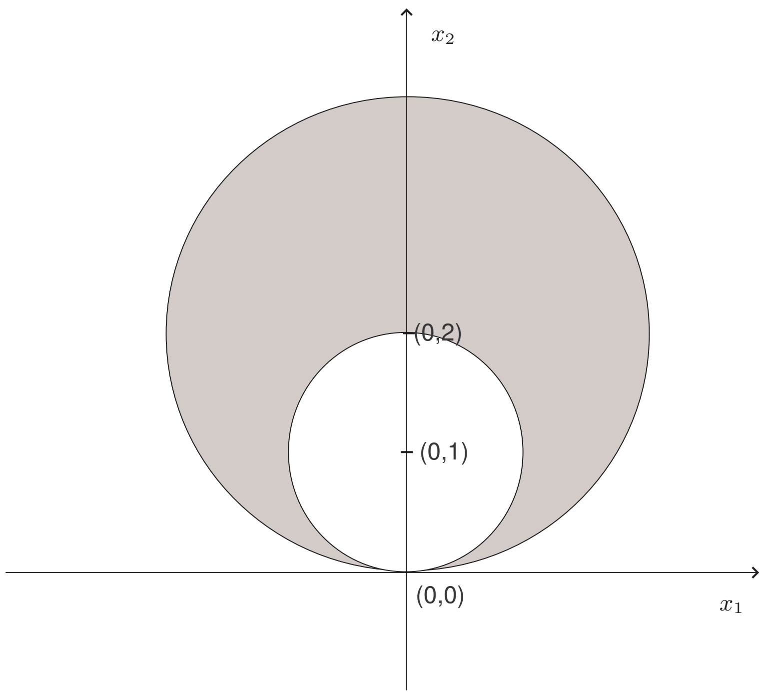
- $\mu_1 = 0$ and $\mu_2 = 0$.
  This selection yields $0 = \frac{dL}{dx_2} = 1$, hence no candidate optimal solution.
- $\mu_1 = 0$ and $x_1^2 + (x_2 - 2)^2 - 4 = 0$.
  This selection yields, from $0 = \frac{dL}{dx_1}$, either $x_1 = 0$ or $\mu_2 = -1$. The first option yields $x_2 = 0$ or $x_2 = 4$, whereas the second option violates the positivity of μ_2. Moreover, the selection $x_1 = 0$ and $x_2 = 4$ yields, from $0 = \frac{dL}{dx_2}$, $\mu_2 < 0$, hence it is not a candidate solution.
- $1 - x_1^2 - (x_2 - 1)^2 = 0$ and $\mu_2 = 0$.
  This selection yields, from $0 = \frac{dL}{dx_1}$, $x_1 = 0$ or $\mu_1 = 1$. The first option yields $x_2 = 0$ or $x_2 = 2$. The second option yields, from $0 = \frac{dL}{dx_2}$, $x_2 = 3/2$, hence, from $1 - x_1^2 - (x_2 - 1)^2 = 0$, $x_1 = \pm\frac{\sqrt{3}}{2}$.
- $1 - x_1^2 - (x_2 - 1)^2 = 0$ and $x_1^2 + (x_2 - 2)^2 - 4 = 0$.
  The only point consistent with these conditions is $(0, 0)$.

In summary the candidate solutions obtained so far are as follows.

- $(0, 0)$.
- $(0, 2)$.
- $\left(\pm\frac{\sqrt{3}}{2}, \frac{3}{2}\right)$.

Hence there are four candidate optimal solutions.

d) The nonregular point $(0, 0)$ is such that $x_1^2 + x_2 = 0$. Note now that the function $x_1^2 + x_2$ is always nonnegative in the admissible set and it is zero, in the admissible set, if and only if $x_1 = x_2 = 0$. Hence the nonregular point is a global minimizer for the considered problem. Note that it is not possible to associate, in a unique way, a pair of optimal multipliers to this optimal point.



Exercise 39 Consider the optimization problem

$$\begin{cases} \min\limits_{x_1,\ x_2} x_1^2 + x_2^2, \\ -x_1 \le 0, \\ x_2 - x_1 - 1 = 0. \end{cases}$$

a) Sketch in the (x_1, x_2)-plane the level surfaces of the function to be minimized and the admissible set. Hence show that all points in the admissible set are regular points for the constraints.

b) Using only graphical considerations determine the solution of the considered problem.

c) This constrained optimization problem can be transformed into an unconstrained optimization problem by defining the so-called mixed penalty-barrier function

$$F_\epsilon(x_1, x_2) = x_1^2 + x_2^2 + \frac{1}{\epsilon}(x_2 - x_1 - 1)^2 + \frac{\epsilon}{x_1},$$

with $\epsilon > 0$ and considering the unconstrained minimization of $F_\epsilon(x_1, x_2)$. Determine the stationary points of $F_\epsilon(x_1, x_2)$.
(Hint: solve $\nabla_{x_2} F_\epsilon(x_1, x_2) = 0$ for x_2, and replace the obtained solution in the equation $\nabla_{x_1} F_\epsilon(x_1, x_2) = 0$. Solve this last equation assuming that $x_1 = \alpha\epsilon^{1/2}$, for some $\alpha > 0$ to be determined, and neglecting all terms ϵ^k, for $k \ge 1/2$.)

d) Show that the stationary point of $F_\epsilon(x_1, x_2)$ computed in part c) tends, as ϵ tends to zero, to the optimal solution determined in part b).

Solution 39

a) The admissible set, and the level surfaces of the function to be minimized are as in the figure below. There are two constraints active at the point $(0, 1)$ and their gradients, at this point, are independent. At any other admissible point there is only one active constraint, the equality constraint, and its gradient is always nonzero (it is a constant vector). Thus all points are regular points for the constraints.

b) The optimal solution is obtained considering the smallest circle centered at the origin intersecting the admissible set. Hence, the optimal solution is the point $(0, 1)$.

c) The stationary points of the mixed penalty-barrier function are the solutions of

$$0 = \nabla F_\epsilon = \begin{bmatrix} 2x_1 - \frac{2}{\epsilon}(x_2 - x_1 - 1) - \frac{\epsilon}{x_1^2} \\ 2x_2 + \frac{2}{\epsilon}(x_2 - x_1 - 1) \end{bmatrix}.$$

Solving the second equation yields

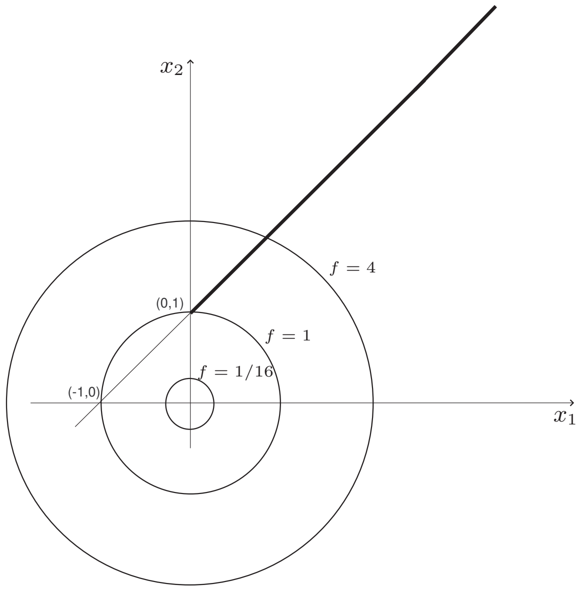
$$x_2 = \frac{x_1 + 1}{\epsilon + 1},$$

and replacing this in the first equation yields

$$0 = \frac{x_1^3(2\epsilon + 4) + 2x_1^2 - \epsilon(1 + \epsilon)}{(\epsilon + 1)x_1^2}.$$

Setting $x_1 = \alpha\sqrt{\epsilon}$ and neglecting all ϵ^k terms, with $k \ge 1/2$, yields $0 = (2\alpha^2 - 1)$, hence (recall that $\alpha > 0$) $x_1 = \sqrt{\epsilon/2}$, and $x_2 = \frac{\sqrt{\epsilon/2}+1}{\epsilon+1}$.

d) As $\epsilon \to 0$, the stationary point of the mixed penalty-barrier function tends to $(0, 1)$, which is the optimal solution of the considered problem.



Exercise 40 Consider the optimization problem

$$\begin{cases} \min\limits_{x_1,\ x_2} x_1 x_2, \\ \frac{1}{2}x_1^2 + 2x_2^2 = 1. \end{cases}$$

a) State first order necessary conditions of optimality for such a constrained optimization problem.

b) Using the conditions in part a) determine candidate optimal solutions for the considered problem.

c) Transform the minimization problem into an unconstrained minimization problem using the method of the exact augmented Lagrangian functions and write explicitly the exact augmented Lagrangian function for the considered problem.

d) Show that the candidate optimal solutions determined in part b) are stationary points of the exact augmented Lagrangian function.

e) Find the global minimizer for the considered problem. Is the global minimizer unique?

Solution 40

a) Define the Lagrangian

$$L(x_1, x_2, \lambda) = x_1 x_2 + \lambda(\frac{1}{2}x_1^2 + 2x_2^2 - 1).$$

The first order necessary conditions of optimality are

$$0 = \frac{dL}{dx_1} = x_2 + \lambda x_1, \qquad 0 = \frac{dL}{dx_2} = x_1 + 4\lambda x_2, \qquad \frac{1}{2}x_1^2 + 2x_2^2 - 1 = 0.$$

b) The conditions $\frac{dL}{dx_1} = \frac{dL}{dx_2} = 0$ can be rewritten as

$$\begin{bmatrix} \lambda & 1 \\ 1 & 4\lambda \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0.$$

If $4\lambda^2 - 1 \neq 0$ the above equation implies $x_1 = x_2 = 0$, which is not an admissible point. If $4\lambda^2 - 1 = 0$, or $\lambda = \pm\frac{1}{2}$. then $x_2 = \mp\frac{1}{2}x_1$, and replacing in the constrains yields the candidate solutions with the corresponding multipliers, namely

$$(x_1, x_2, \lambda) = \left(1, -\frac{1}{2}, \frac{1}{2}\right), \qquad (x_1, x_2, \lambda) = \left(-1, \frac{1}{2}, \frac{1}{2}\right),$$

$$(x_1, x_2, \lambda) = \left(1, \frac{1}{2}, -\frac{1}{2}\right), \qquad (x_1, x_2, \lambda) = \left(-1, -\frac{1}{2}, -\frac{1}{2}\right).$$

c) The exact augmented Lagrangian function for a constraint optimization problem with equality constraints is

$$S(x, \lambda) = f(x) + \lambda' g(x) + \frac{1}{\epsilon}\|g(x)\|^2 + \eta \|\frac{\partial g(x)}{\partial x} \nabla_x L(x, \lambda)\|^2,$$

with $\epsilon > 0$ and $\eta > 0$. For the considered problem, we have

$$S(x_1, x_2, \lambda) = x_1 x_2 + \lambda(\frac{1}{2}x_1^2 + 2x_2^2 - 1) + \frac{1}{\epsilon}(\frac{1}{2}x_1^2 + 2x_2^2 - 1)^2 + \eta \left(\begin{bmatrix} x_1 & 4x_2 \end{bmatrix} \begin{bmatrix} x_2 + \lambda x_1 \\ x_1 + 4\lambda x_2 \end{bmatrix} \right)^2.$$

d) The stationary points of the function $S(x_1, x_2, \lambda)$ are the solutions of the equations

$$0 = \frac{dS}{dx_1} = x_2 + \lambda x_1 + \frac{2x_1}{\epsilon}(\frac{1}{2}x_1^2 + 2x_2^2 - 1) + 2\eta(5x_1x_2 + \lambda x_1^2 + 16\lambda x_2^2)(5x_2 + 2\lambda x_1),$$

$$0 = \frac{dS}{dx_2} = x_1 + 4\lambda x_2 + \frac{8x_2}{\epsilon}(\frac{1}{2}x_1^2 + 2x_2^2 - 1) + 2\eta(5x_1x_2 + \lambda x_1^2 + 16\lambda x_2^2)(5x_1 + 32\lambda x_2),$$

$$0 = \frac{dS}{d\lambda} = \frac{1}{2}x_1^2 + 2x_2^2 - 1 + 2\eta(5x_1x_2 + \lambda x_1^2 + 16\lambda x_2^2)(x_1^2 + 16x_2^2).$$

Replacing the candidate points obtained in part b) shows that indeed they are stationary points for the augmented Lagrangian function. (Note that this is true for any ϵ and η.)

e) To find the global minimum we evaluate the function to be minimized at the candidate optimal solutions:

$$(x_1x_2)_{x_1=1, x_2=-1/2} = -\frac{1}{2}, \qquad (x_1x_2)_{x_1=-1, x_2=1/2} = -\frac{1}{2},$$

$$(x_1x_2)_{x_1=1, x_2=1/2} = \frac{1}{2}, \qquad (x_1x_2)_{x_1=-1, x_2=-1/2} = \frac{1}{2}.$$

Hence, the points $(1, -1/2)$ and $(-1, 1/2)$ are both global minimizers. (Note that the points $(1, 1/2)$ and $(-1, -1/2)$ are both global maximizers.)

Exercise 41 Consider the optimization problems

$$P_{min} \left\{ \begin{array}{l} \min\limits_{x_1,\ x_2} |x_1| + |x_2|, \\ \\ x_1^2 + x_2^2 = 1, \end{array} \right.$$

and

$$P_{max} \left\{ \begin{array}{l} \max\limits_{x_1,\ x_2} |x_1| + |x_2|, \\ \\ x_1^2 + x_2^2 = 1. \end{array} \right.$$

a) Sketch in the (x_1, x_2)-plane the admissible set and the level lines of the function $|x_1| + |x_2|$.

b) Using only graphical considerations determine the solutions of the considered problems.

c) State first order necessary conditions of optimality for these constrained optimization problems. Show that the optimal solutions determined in part b) satisfy the necessary conditions of optimality. (Hint: use the fact that $\text{sign}(0) = 0$.)

d) Write a penalty function F_ϵ for problem P_{max}. Show that, for $\epsilon > 0$ and sufficiently small, the stationary points of F_ϵ approach the optimal solutions determined in part b). (Do not compute explicitly the stationary points of F_ϵ.)
(Hint: for ϵ sufficiently small, the stationary points of F_ϵ are such that $x_1 \neq 0$ and $x_2 \neq 0$.)

Solution 41

a) The admissible set is the circle of radius one and with center at $(0, 0)$. The level lines of the function $|x_1| + |x_2|$ are squares with their vertices on the x_1- and x_2- axes, as indicated in the figure below.

b) The solution to problem P_{min} is obtained considering the smallest square level set intersecting the admissible set. Hence there are four optimal solutions, namely the points $(0, \pm 1)$ and $(\pm 1, 0)$.
The solution to problem P_{max} is obtained considering the largest square level set intersecting the admissible set. Hence there are four optimal solutions, namely the points $\left(\pm\frac{\sqrt{2}}{2}, \pm\frac{\sqrt{2}}{2}\right)$.

c) Define the Lagrangian

$$L(x_1, x_2, \lambda) = \pm(|x_1| + |x_2|) + \lambda(x_1^2 + x_2^2 - 1),$$

where the $+$ sign has to be used for P_{min} and the $-$ sign has to be used for P_{max}. The first order necessary conditions of optimality are

$$0 = \frac{dL}{dx_1} = \text{sign}(x_1) + 2\lambda x_1, \qquad 0 = \frac{dL}{dx_2} = \text{sign}(x_2) + 2\lambda x_2, \qquad x_1^2 + x_2^2 - 1 = 0,$$

and a direct substitution shows that the solutions determined in part b) satisfy the necessary conditions of optimality.

d) A penalty function for problem P_{max} is

$$F_\epsilon(x_1, x_2) = -(|x_1| + |x_2|) + \frac{1}{\epsilon}(x_1^2 + x_2^2 - 1)^2.$$

The stationary points of F_ϵ are the solutions of the equations

$$0 = -\text{sign}(x_1) + \frac{4}{\epsilon}x_1(x_1^2 + x_2^2 - 1), \qquad 0 = -\text{sign}(x_2) + \frac{4}{\epsilon}x_2(x_1^2 + x_2^2 - 1).$$

If we assume that the stationary points of F_ϵ, for ϵ sufficiently small, are away from $x_1 = 0$ and from $x_2 = 0$, then the stationary points are such that

$$\frac{\text{sign}(x_1)}{x_1} = \frac{\text{sign}(x_2)}{x_2},$$

which implies $x_2 = \pm x_1$. Replacing this in the first of the equations above yields

$$0 = -\text{sign}(x_1) + \frac{4}{\epsilon}x_1(2x_1^2 - 1),$$

or equivalently

$$\frac{\epsilon}{4}\text{sign}(x_1) = x_1(2x_1^2 - 1).$$

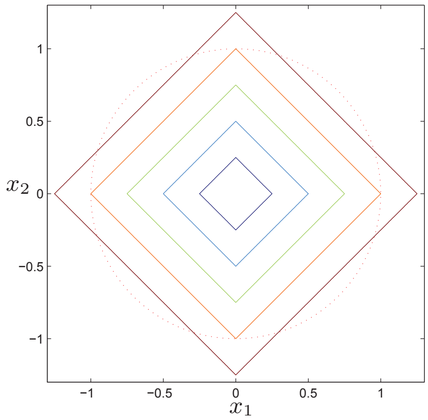
For ϵ sufficiently small the solutions of this equation are of the form

$$x_1 = \pm\frac{\sqrt{2}}{2} + o(\epsilon).$$

As a result, the stationary points of F_ϵ are of the form

$$\left(\pm\left(\frac{\sqrt{2}}{2} + o(\epsilon)\right), \pm\left(\frac{\sqrt{2}}{2} + o(\epsilon)\right)\right),$$

i.e. they are close to the optimal solutions of the problem P_{max} for ϵ sufficiently small.



Exercise 42 Consider the optimization problem

$$\begin{cases} \min\limits_{x_1,\ x_2} x_1^3 - x_1^2 x_2 + 2x_2^2, \\ x_1 \geq 0, \\ x_2 \geq 0. \end{cases}$$

a) State first order necessary conditions of optimality for this constrained optimization problem.

b) Using the conditions derived in part a) compute candidate optimal solutions. Show that there is one candidate solution on the boundary of the admissible set and one in the interior of the admissible set.

c) Using second order sufficient conditions of optimality show that the candidate solution inside the admissible set is not a local minimizer.

d) Show that the candidate optimal solution on the boundary of the admissible set is a local minimizer. (Hint: show that the function to be minimized is zero at the candidate optimal solution, and it is strictly positive in all admissible points in a neighborhood of the candidate optimal solution).

e) Show that the function to be minimized is not bounded from below in the admissible set. Hence, argue that the problem does not have a global solution.
(Hint: consider the function to be minimized along the line $x_2 = 2x_1$, and study its behaviour for $x_1 > 0$ and large.)

Solution 42

a) Define the Lagrangian

$$L(x_1, x_2, \rho_1, \rho_2) = x_1^3 - x_1^2 x_2 + 2x_2^2 + \rho_1(-x_1) + \rho_2(-x_2).$$

The first order necessary conditions of optimality are

$$0 = \frac{dL}{dx_1} = 3x_1^2 - 2x_1 x_2 - \rho_1, \qquad 0 = \frac{dL}{dx_2} = -x_1^2 + 4x_2 - \rho_2,$$

$$-x_1 \leq 0, \qquad -x_2 \leq 0, \qquad \rho_1 > 0, \qquad \rho_2 > 0,$$

$$-x_1\rho_1 = 0, \qquad -x_2\rho_2 = 0.$$

b) Using the complementarity conditions, *i.e.* the last two conditions, we have four possibilities.

- $\rho_1 = 0$ and $\rho_2 = 0$. This yields the candidate optimal solutions $(x_1, x_2) = (0, 0)$ and $(x_1, x_2) = (6, 9)$.

- $\rho_1 = 0$ and $x_2 = 0$. This yields the candidate optimal solution $(x_1, x_2) = (0, 0)$.
- $x_1 = 0$ and $\rho_2 = 0$. This yields the candidate optimal solution $(x_1, x_2) = (0, 0)$.
- $x_1 = 0$ and $x_2 = 0$.

In summary there are two candidate optimal solutions: the point $(0, 0)$, on the boundary of the admissible set, and the point $(6, 9)$ in the interior of the admissible set.

c) The second order sufficient condition of optimality for the candidate point in the interior of the admissible set is $\nabla^2 L(6, 9) > 0$. Note that

$$\nabla^2 L(6,9) = 2 \begin{bmatrix} 9 & -6 \\ -6 & 2 \end{bmatrix},$$

and that $\det \nabla^2 L(6, 9) < 0$, which implies that $\nabla^2 L(6, 9)$ is not positive definite. Hence the candidate optimal point in the interior of the admissible set is not a local minimizer (tt is a saddle point).

d) To show that the point $(0, 0)$ is a local minimizer note that the function f to be minimized is such that $f(0, 0) = 0$, $f(x_1, 0) > 0$ for $x_1 > 0$, and $f(0, x_2) > 0$ for $x_2 > 0$. Consider now straight lines described by $x_2 = \alpha x_1$, with $\alpha > 0$. Then

$$f(x_1, \alpha x_1) = \alpha^2 \left(\frac{1-\alpha}{\alpha^2} x_1^3 + 2x_1^2 \right),$$

which is positive for all $\alpha > 0$ and all $x_1 > 0$ and sufficiently small. Since the function f is zero at the candidate optimal solution $(0, 0)$ and strictly positive in all admissible point in a neighborhood of this point, then the point is a local minimizer.

e) The function f along the line $x_2 = 2x_1$ is given by $f(x_1, 2x_1) = -x_1^3 + 4x_1^2$ and this function is not bounded from below, *i.e.* $\lim_{x_1 \to \infty} f(x_1, 2x_1) = -\infty$. This implies that the considered optimization problem does not have a global solution.

Exercise 43 Consider the optimization problem

$$\begin{cases} \max\limits_{x_1,\ x_2,\ x_3} (x_1 x_2 + x_2 x_3 + x_1 x_3), \\ x_1 + x_2 + x_3 = 3. \end{cases}$$

a) State first order necessary conditions of optimality for this constrained optimization problem and show that there exists only one candidate optimal solution.

b) Using second order sufficient conditions of optimality show that the candidate solution is a local maximizer.

c) Consider the use of an exact penalty function for the solution of the problem.

  i) Write an exact penalty function G for the problem.

  ii) Show that the function is well-defined for every (x_1, x_2, x_3).

  iii) Show that the exact penalty function has only one stationary point and this coincides with the optimal solution of the problem determined in part b).

Solution 43

a) Define the Lagrangian (note the change in sign due to the transformation of the maximization problem into a minimization problem)

$$L(x_1, x_2, x_3, \lambda) = -(x_1 x_2 + x_2 x_3 + x_1 x_3) + \lambda(x_1 + x_2 + x_3 - 3).$$

The first order necessary conditions of optimality are

$$0 = \frac{dL}{dx_1} = -x_2 - x_3 + \lambda, \qquad 0 = \frac{dL}{dx_2} = -x_1 - x_3 + \lambda,$$

$$0 = \frac{dL}{dx_3} = -x_2 - x_1 + \lambda, \qquad\qquad 0 = x_1 + x_2 + x_3 - 3.$$

This is system a linear equations with the unique solution $(x_1, x_2, x_3, \lambda) = (1, 1, 1, 2)$. Hence the problem has only one candidate optimal solution.

b) Note that

$$\nabla^2 L = \begin{bmatrix} 0 & -1 & -1 \\ -1 & 0 & -1 \\ -1 & -1 & 0 \end{bmatrix}$$

and

$$\frac{\partial g}{\partial x} = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}.$$

The candidate optimal solution is a minimizer if $s'\nabla^2 Ls > 0$ for all $s \neq 0$ such that $s'\frac{\partial g}{\partial x} = 0$. The set of such s's can be described by linear combinations of the vectors

$$s_1' = \begin{bmatrix} 1 & -1 & 0 \end{bmatrix} \qquad\qquad s_2' = \begin{bmatrix} 1 & 0 & -1 \end{bmatrix}.$$

Note that

$$[s_1, s_2]'\nabla^2 L[s_1, s_2] = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} > 0,$$

hence the candidate optimal solution is a local minimizer.

c) The exact penalty function for a constraint optimization problem with equality constraints is

$$G(x) = f(x) - g'(x)\left(\frac{\partial g}{\partial x}\frac{\partial g'}{\partial x}\right)^{-1}\frac{\partial g}{\partial x}\nabla f + \frac{1}{\epsilon}\|g(x)\|^2,$$

with $\epsilon > 0$.

i) For the considered problem we have

$$G(x_1, x_2, x_3) = -(x_1x_2 + x_2x_3 + x_1x_3) + \frac{2}{3}(x_1 + x_2 + x_3 - 3)(x_1 + x_2 + x3) + \frac{1}{\epsilon}(x_1 + x_2 + x_3 - 3)^2.$$

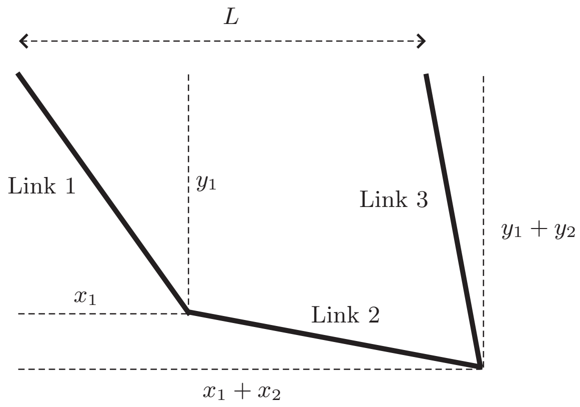
ii) The function is well-defined for all (x_1, x_2, x_3) since $\frac{\partial g}{\partial x}\frac{\partial g'}{\partial x}$ is a full rank matrix (it is a nonzero constant).

iii) The stationary points of the function $G(x_1, x_2, x_3)$ are the solutions of the equations

$$0 = \nabla G = \begin{bmatrix} \frac{1}{3}(4x_1 + x_2 + x_3) - 2 + \frac{2}{\epsilon}(x_1 + x_2 + x_3 - 3) \\ \frac{1}{3}(x_1 + 4x_2 + x_3) - 2 + \frac{2}{\epsilon}(x_1 + x_2 + x_3 - 3) \\ \frac{1}{3}(x_1 + x_2 + 4x_3) - 2 + \frac{2}{\epsilon}(x_1 + x_2 + x_3 - 3) \end{bmatrix}.$$

These equations have a unique solution $(x_1, x_2, x_3) = (1, 1, 1)$ which does not depend upon ϵ and coincides with the optimal solution determined in part b).

Exercise 44 A chain with three links, each of length one, hangs between two points at the same height, a distance $L > 1$ apart (see the figure below). To find the form in which the chain hangs we minimize the potential energy.
Let (x_i, y_i) be the displacement of the right end of the ith link , from the right end of the $(i-1)$th link.



The potential energy is therefore

$$V(y_1, y_2, y_3) = \frac{1}{2}y_1 + (y_1 + \frac{1}{2}y_2) + (y_1 + y_2 + \frac{1}{2}y_3).$$

a) The condition that the hanging points are a distance L apart can be translated in the constraint

$$x_1 + x_2 + x_3 = L.$$

Express this constraint in terms of the variables y_i.
(Hint: use Pythagoras' Theorem!)

b) Show that the condition that the height of the hanging points is the same can be expressed with the constraint

$$y_1 + y_2 + y_3 = 0.$$

c) Consider the problem of minimizing the potential energy $V(y_1, y_2, y_3)$ subject to the constraints determined in parts b) and c).

i) Write necessary conditions of optimality for the considered optimization problem.

ii) Using physical considerations it may be noted that candidate optimal solutions should be such that the chain has a _/ shape or a /‾\\ shape. Show that these two shapes yield values for y_1, y_2, y_3 and for the Lagrangian multipliers such that the necessary conditions of optimality are met.
(Hint: note that for both shapes $y_2 = 0$.)

iii) By evaluating the potential energy at the candidate optimal solutions determined in part c.ii) determine the shape that minimizes the potential energy.

Solution 44

a) From the figure above we obtain, for $i = 1, 2, 3$, $x_i^2 + y_i^2 = 1$, hence $x_i = \sqrt{1 - y_i^2}$, yielding the constrain

$$\sqrt{1 - y_1^2} + \sqrt{1 - y_2^2} + \sqrt{1 - y_3^2} = L.$$

b) Since the height of the left hanging point is at zero, and the y-coordinate of the last link is $y_1+y_2+y_3$ then the condition that both hanging points are at the same height is given by $y_1 + y_2 + y_3 = 0$.

c) The optimization problem to solve is thus

$$\begin{cases} \min\limits_{y_1, y_1, y_3} \frac{5}{2}y_1 + \frac{3}{2}y_2 + \frac{1}{2}y_3, \\ \\ \sqrt{1 - y_1^2} + \sqrt{1 - y_2^2} + \sqrt{1 - y_3^2} - L = 0, \\ \\ y_1 + y_2 + y_3 = 0. \end{cases}$$

i) Define the Lagrangian

$$L(y_1, y_2, y_3, \lambda_1, \lambda_2) = \frac{5}{2}y_1 + \frac{3}{2}y_2 + \frac{1}{2}y_3 + \lambda_1(\sqrt{1-y_1^2} + \sqrt{1-y_2^2} + \sqrt{1-y_3^2} - L) + \lambda_2(y_1 + y_2 + y_3).$$

The first order necessary conditions of optimality are

$$0 = \frac{\partial L}{\partial y_1} = \frac{5}{2} - \lambda_1 \frac{y_1}{\sqrt{1-y_1^2}} + \lambda_2, \qquad 0 = \frac{\partial L}{\partial y_2} = \frac{3}{2} - \lambda_1 \frac{y_2}{\sqrt{1-y_2^2}} + \lambda_2,$$

$$0 = \frac{\partial L}{\partial y_3} = \frac{1}{2} - \lambda_1 \frac{y_3}{\sqrt{1-y_3^2}} + \lambda_2,$$

$$\sqrt{1-y_1^2} + \sqrt{1-y_2^2} + \sqrt{1-y_3^2} - L = 0, \qquad y_1 + y_2 + y_3 = 0.$$

ii) The indicated shapes are such that $y_2 = 0$ and$y_1 = -y_3$. Replacing these conditions in the necessary conditions of optimality yields

$$0 = \frac{\partial L}{\partial y_1} = \frac{5}{2} - \lambda_1 \frac{y_1}{\sqrt{1-y_1^2}} + \lambda_2, \qquad 0 = \frac{\partial L}{\partial y_2} = \frac{3}{2} - \lambda_2,$$

$$0 = \frac{\partial L}{\partial y_3} = \frac{1}{2} - \lambda_1 \frac{y_1}{\sqrt{1-y_1^2}} + \lambda_2,$$

$$2\sqrt{1-y_1^2} + 1 - L = 0, \qquad 0 = 0.$$

These equations have the two solutions

$$y_1 = \pm\frac{1}{2}\sqrt{3 + 2L - L^2}, \qquad \lambda_1 = \pm\frac{L-1}{\sqrt{3+2L-L^2}}, \qquad \lambda_2 = \frac{3}{2}.$$

The one with positive y_1 corresponds to the /‾\ shape, the one with negative y_1 corresponds to the _/ shape. Note that all square roots are well-defined since $L > 1$.

iii) The potential energy for the above shapes is $V(y_1, y_2, y_3) = 2y_1$. Hence the candidate optimal solution with negative y_1 yields a local minimizer.

Exercise 45 The economy class luggage policy of an airline on a transatlantic flight reads:

Each passenger is allowed one piece of luggage. The three linear dimensions, when added together, must not exceed 150 *cm.*

The problem of maximizing the volume of the luggage can be posed and solved with the following steps.

a) Let $x_1 > 0$, $x_2 > 0$ and $x_3 > 0$ be the three linear dimensions (in cm) of a piece of luggage. Write the considered optimization problem as a minimization problem subject to one inequality constraint. (Do not include the constraints $x_1 > 0$, $x_2 > 0$ and $x_3 > 0$ in the formulation of the problem.)

b) State first order necessary conditions of optimality for this constrained optimization problem.

c) Using the conditions derived in part b) compute candidate optimal solutions.

d) Using second order sufficient conditions of optimality determine which of the candidate optimal solutions determined in part c) is a local maximizer.

e) Which is the geometric shape of the 'optimal luggage'?

Solution 45

a) The considered optimization problem can be written as

$$\begin{cases} \max\limits_{x_1,x_2,x_3} \ x_1x_2x_3, \\ x_1 + x_2 + x_3 \leq 150. \end{cases}$$

b) Define the Lagrangian (note the change in sign of the objective function)

$$L(x_1, x_2, x_3, \rho) = -x_1x_2x_3 + \rho(x_1 + x_2 + x_3 - 150).$$

The first order necessary conditions of optimality are

$$0 = \frac{\partial L}{\partial x_1} = -x_2x_3 + \rho, \qquad 0 = \frac{\partial L}{\partial x_2} = -x_1x_3 + \rho, \qquad 0 = \frac{\partial L}{\partial x_3} = -x_2x_1 + \rho,$$

$$\rho \geq 0, \qquad x_1 + x_2 + x_3 - 150 \leq 0, \qquad \rho(x_1 + x_2 + x_3 - 150) = 0.$$

c) Using the complementarity condition, *i.e.* the last condition, we have two cases.
Case 1: $\rho = 0$. This implies $x_1x_2 = x_2x_3 = x_1x_3 = 0$, yielding the sets of candidate solutions

$$x_1 = x_2 = \rho = 0, x_3 \leq 150, \qquad x_1 = x_3 = \rho = 0, x_2 \leq 150, \qquad x_2 = x_3 = \rho = 0, x_1 \leq 150.$$

Case 2: $x_1 + x_2 + x_3 = 150$. This yields the candidate solutions

$$x_1 = x_2 = \rho = 0, x_3 = 150, \qquad x_1 = x_3 = \rho = 0, x_2 = 150, \qquad x_2 = x_3 = \rho = 0, x_1 = 150,$$

and

$$x_1 = x_2 = x_3 = 50, \rho = 50^2.$$

d) Note that

$$\nabla^2 L(x_1, x_2, x_3) = -\begin{bmatrix} 0 & x_3 & x_2 \\ x_3 & 0 & x_1 \\ x_2 & x_1 & 0 \end{bmatrix}.$$

All candidate solutions obtained in Case 1, for which no constrain is active, are such that $\nabla^2 L$ has a positive, a negative and a zero eigenvalue. As a result, all solutions obtained in Case 1, are saddle points. Consider now the candidate solutions obtained in Case 2, and such that $\rho = 0$. For such solutions the condition of strict complementarity does not hold, hence it is not possible to use second order sufficient conditions to classify these points. Finally, consider the candidate optimal solution

$$x_1 = x_2 = x_3 = 50, \qquad \rho = 50^2.$$

The second order sufficient condition require $s'\nabla^2 L(50, 50, 50)s > 0$ for all non-zero s such that

$$\begin{bmatrix} 1 & 1 & 1 \end{bmatrix} s = 0.$$

Such s can be parameterized as

$$s = \begin{bmatrix} s_1 & s_2 & -s_1 - s_2 \end{bmatrix},$$

yielding

$$s'\nabla^2 L(50, 50, 50)s = 100(s_1^2 + s_2^2 + s_1s_2),$$

which is positive for all non-zero s_1 and s_2. As a result, this candidate optimal solution is a local minimizer. (It is a local maximizer for the original problem).

e) The optimal luggage is a cube!

Exercise 46 Consider the optimization problem

$$\begin{cases} \min\limits_{x_1, x_2} \; x_1x_2^2, \\ x_1^2 + x_2^2 \leq 2. \end{cases}$$

a) State first order necessary conditions of optimality for this constrained optimization problem.

b) Using the conditions derived in part a) compute candidate optimal solutions.

c) Evaluating the objective function at the candidate optimal solutions determined in part b) derive the solution of the considered optimization problem.

d) The considered constrained optimization problem can be solved minimizing the so-called logarithmic penalty function given by
$$P_l(x_1, x_2) = x_1 x_2^2 - \epsilon \log(2 - x_1^2 - x_2^2),$$
with $\epsilon > 0$.

i) State first order necessary condition of optimality for P_l.

ii) Show that the stationary points of P_l are such that
$$x_2^2 = 2x_1^2.$$

iii) Using the results in part d.ii) show that the stationary points of P_l are such that
$$x_1(3x_1^3 - 2x_1 - \epsilon) = 0.$$
Hence argue that, as ϵ approaches zero the stationary points of P_l approach candidate optimal solutions for the considered problem.

Solution 46

a) Define the Lagrangian
$$L(x_1, x_2, \rho) = x_1 x_2^2 + \rho(x_1^2 + x_2^2 - 2).$$
The first order necessary conditions of optimality are
$$0 = \frac{\partial L}{\partial x_1} = x_2^2 + 2\rho x_1, \qquad 0 = \frac{\partial L}{\partial x_2} = 2x_1 x_2 + 2\rho x_2,$$
$$x_1^2 + x_2^2 - 2 \leq 0, \qquad \rho > 0, \qquad \rho(x_1^2 + x_2^2 - 2) = 0.$$

b) Using the complementarity conditions, *i.e.* the last condition, we have two possibilities.

- $\rho = 0$. This yields the candidate optimal solutions
$$P_1: \ (x_1, x_2) = (\alpha, 0)$$
with $|\alpha| \leq \sqrt{2}$. Note that at $(x_1, x_2) = (\pm\sqrt{2}, 0)$ the condition of strict complementarity does not hold.
- $x_1^2 + x_2^2 - 2 = 0$. This yields the candidate optimal solutions
$$P_2: \ (x_1, x_2) = (\pm\sqrt{2}, 0),$$
with $\rho \geq 0$, and
$$P_3: \ (x_1, x_2) = (-\frac{\sqrt{6}}{3}, \pm\frac{\sqrt{12}}{3}),$$
with $\rho = \frac{\sqrt{6}}{3}$.

In summary there are infinitely many candidate optimal solutions, some of which such that second order sufficient conditions cannot be used.

c) The values of the objective function at candidate optimal points are
$$f(P_1) = 0, \qquad f(P_2) = 0, \qquad f(P_3) = -\frac{4}{9}\sqrt{6}.$$
Hence P_3 is the solution of the considered problem.

d) i) The first order necessary condition of optimality for P_l are
$$0 = \frac{\partial P_l}{\partial x_1} = x_2^2 + 2\epsilon \frac{x_1}{2 - x_1^2 - x_2^2}, \qquad 0 = \frac{\partial P_l}{\partial x_2} = 2x_1 x_2 + 2\epsilon \frac{x_2}{2 - x_1^2 - x_2^2}.$$

ii) The equations defining the stationary points of P_l yield, for nonzero x_1 and x_2,

$$-2\epsilon\frac{1}{2-x_1^2-x_2^2} = \frac{x_2^2}{x_1} = 2x_1,$$

hence stationary points are such that

$$x_2^2 = 2x_1^2.$$

If $x_1 = 0$ then the necessary conditions yield $x_2 = 0$, and similarly for $x_2 = 0$. Hence, the above relation holds for any x_1 and x_2.

iii) Replacing the above relation in the equation

$$0 = \frac{\partial P_l}{\partial x_1}$$

yields

$$2x_1\frac{3x_1^3 - 2x_1 - \epsilon}{3x_1^2 - 2} = 0$$

from which we infer that, as $\epsilon \to 0$, $x_1 \to 0$ or $x_1 \to \pm\frac{\sqrt{6}}{3}$. As a result, as ϵ goes to zero the stationary points of P_l *approach* the candidate optimal solutions of the problem.

Exercise 47 Consider the optimization problem

$$\begin{cases} \max\limits_{x_1,x_2,x_3} \; x_1 + 2x_2 + x_3, \\ x_1^2 + x_2^2 + x_3^2 \leq 1. \end{cases}$$

a) State first order necessary condition of optimality for this constrained optimization problem.

b) Using the conditions derived in part a) compute candidate optimal solutions.

c) Using second order sufficient conditions of optimality determine the solution of the optimization problem.

d) Consider the change of variables

$$x_1 = r\cos\theta\sin\phi, \qquad x_2 = r\sin\theta\sin\phi, \qquad x_3 = r\cos\phi,$$

with $r \geq 0$, $\theta \in [0, 2\pi)$, and $\phi = [0, 2\pi)$.

i) Rewrite the considered optimization problem in the new variables and show that the resulting problem can be written in the form

$$\begin{cases} \max\limits_{r,\theta,\phi} \; r\Psi(\theta,\phi), \\ r \leq 1, \\ \theta \in [0, 2\pi), \\ \phi \in [0, 2\pi), \end{cases}$$

Determine the function $\Psi(\theta, \phi)$.

ii) Argue that the problem is equivalent to the unconstrained optimization problem

$$\max_{\theta,\phi} \Psi(\theta, \phi).$$

iii) Find candidate solutions of the unconstrained optimization problem in part d.ii), and show that one of the candidate solutions coincides with the optimal solution determined in part c).

Solution 47

a) Define the Lagrangian (note the sign change due to the transformation of the maximization problem into a minimization problem)

$$L(x_1, x_2, x_3, \rho) = -x_1 - 2x_2 - x_3 + \rho(x_1^2 + x_2^2 + x_3^2 - 1).$$

The first order necessary conditions of optimality are

$$0 = \frac{\partial L}{\partial x_1} = -1 + 2\rho x_1, \qquad 0 = \frac{\partial L}{\partial x_2} = -1 + 2\rho x_2, \qquad 0 = \frac{\partial L}{\partial x_2} = -1 + 2\rho x_3,$$

$$\rho \geq 0, \qquad x_1^2 + x_2^2 + x_3^2 - 1 \leq 0, \qquad \rho(x_1^2 + x_2^2 + x_3^2 - 1) = 0.$$

b) Using the complementarity conditions, *i.e.* the last condition, we have two possibilities.

- $\rho = 0$. This does not yield any candidate optimal solution.
- $x_1^2 + x_2^2 + x_3^2 - 1 = 0$, $\rho > 0$. This yields the candidate optimal solution

$$P: \ (x_1, x_2, x_3) = \frac{1}{2\rho}(1, 2, 1)$$

with $\rho \geq 0$ such that $\frac{3}{2\rho^2} = 1$.

In summary there is only one candidate optimal solution given by

$$P: \ (x_1, x_2, x3) = \frac{\sqrt{6}}{6}(1, 2, 1).$$

c) The Hessian of the Lagrangian is $\nabla^2 L = 2\rho I$, with I the identity matrix. Hence, the Hessian is positive definite at the candidate optimal solution which is therefore a (local) minimizer for the problem (note that we have changed the sign of the objective function to transform the maximization problem into a minimization one).

d) i) Applying the change of variable to the objective function yields the transformed objective function

$$r(\cos\theta \sin\phi + 2\sin\theta \sin\phi + \cos\phi),$$

whereas the constraint is transformed into $r^2 \leq 1$, which is equivalent to $r \leq 1$ since $r \geq 0$. As a result, the function Ψ is given by

$$\Psi(\theta, \phi) = (\cos\theta \sin\phi + 2\sin\theta \sin\phi + \cos\phi).$$

ii) The objective function in the transformed variables is separable, *i.e.* it is the product of two functions of different variables, namely r and Ψ. As a result, the maximization is achieved maximizing Ψ and r. The latter is maximized for $r = 1$. The former has to be maximized for $\theta \in [0, 2\pi)$ and $\phi \in [0, 2\pi)$. However, since Ψ is periodic in θ and ϕ it can be maximized disregarding the constraints.

iii) The stationary points of Ψ are the solutions of

$$\sin\phi(2\cos\theta - \sin\theta) = 0 \qquad \cos\theta\cos\phi + 2\sin\theta\cos\phi + \sin\phi = 0.$$

The first equation yields

- $\phi = 0$ or $\phi = \pi$, which replaced in the second equation yield $\theta = -\arctan 2$;
- $\theta = \arctan 2$, yielding $\phi = \arctan\sqrt{5}$.

In the original coordinates the first candidate solutions yield $(x_1, x_2, x_3) = (0, 0, \pm 1)$, whereas the second candidate solution give the optimal solution determined in part c).

Exercise 48 The methods of optimization can be used to solve simple geometric problems. Consider the following list of problems. For each of them, formulate the problem as an optimisation problem, defining the decision variables, the cost to be optimised, and the admissible set, and provide explicit solutions.

a) Show that of all rectangles with a fixed positive area the one with the smallest perimeter is a square.

b) Show that of all rectangles with a fixed positive perimeter the one with the largest area is a square.

c) Find the rectangle of largest area that has its base on the x-axis and its other two vertices above the x-axis and on the parabola $y = 8 - x^2$.

d) A piece of wire 10 meters long is cut into two pieces. One piece is bent to form a square, the other piece is bent to form an equilateral triangle. How should the wire be cut so that the total area of the square and of the triangle is a maximum or a minimum?

Solution 48

a) Let x and y be the height and length of the rectangle, then we want to minimize $P = 2x + 2y$ subject to $xy = A$, $x > 0$ and $y > 0$. Using the constraint on the area we have

$$y = \frac{A}{x},$$

hence we need to minimize

$$P = 2x + 2\frac{A}{x}$$

subject to $x > 0$. Ignoring the positivity constraint on x, the stationary points of P are the solutions of

$$0 = \frac{dP}{dx} = 2 - 2\frac{A}{x^2},$$

i.e. $x = \pm\sqrt{A}$. The only feasible solution is $x = \sqrt{A}$, which is a global minimizer, since P is convex for $x > 0$, yielding $y = \sqrt{A}$, hence the rectangle with minimum perimeter is a square with perimeter $P = 4\sqrt{A}$.

b) Let x and y be the height and length of the rectangle, then we want to maximize $A = xy$ subject to $P = 2x + 2y$, $x > 0$ and $y > 0$. Using the constraint on the perimeter we have

$$A = x\frac{P - 2x}{2} = \frac{1}{2}Px - x^2,$$

subject to $x > 0$. Ignoring the positivity constraint on x, the stationary points of A are the solutions of $0 = \frac{dA}{dx} = \frac{1}{2}P - 2x$, *i.e.* $x = \frac{1}{4}P$. This solution is positive, hence feasible, and it is a global maximizer since the function is concave, hence the rectangle with maximum area is a square with area $A = \frac{P^2}{16}$.

c) The area of the rectangle is (note that to have two vertices on the given parabola, the two vertices on the x-axis should be symmetric with respect to the y-axis)

$$A = 2xy = 2x(8 - x^2) = 16x - 2x^3,$$

with $0 \le x \le \sqrt{8}$. The function A is continuous in the interval $[0, \sqrt{8}]$, hence its global maximum is either a stationary point or an extreme of the interval. Stationary points are the solutions of $0 = \frac{dA}{dx} = 16 - 6x^2$, *i.e.* $x = \pm\sqrt{8/3}$. Note now that

$$x = 0 \Rightarrow A = 0, \qquad x = \sqrt{8/3} \Rightarrow A = \frac{64}{3}\frac{\sqrt{2}}{\sqrt{3}}, \qquad x = \sqrt{8} \Rightarrow A = 0,$$

hence the optimal solution is $x = \sqrt{8/3}$.

d) Let x be the length of the wire used for the square, and $10 - x$ the length used for the triangle. Each side of the square is $x/4$, and its area is $x^2/16$. Each side of the triangle is $(10 - x)/3$, and its area is $\frac{\sqrt{3}}{4}\frac{(10-x)^2}{3^2}$. The total area enclosed is

$$A = \frac{1}{16}x^2 + \frac{\sqrt{3}}{36}(10 - x)^2,$$

with $x \in [0, 10]$. Extrema, *i.e.* minimizers and maximizers are either stationary points in $[0, 10]$ or the extremes of the intervals. The function A has only a stationary point (since it is a quadratic function), namely $x = \frac{80\sqrt{3}}{18+8\sqrt{3}} \approx 4.35$. Note now that

$$x = 0 \Rightarrow A = \frac{100\sqrt{3}}{36} \approx 4.81, \qquad x \approx 4.35 \Rightarrow A \approx 2.72, \qquad x = 10 \Rightarrow A = 6.25.$$

Hence, to have the minimum area, use 4.35m of wire for the square and the rest for the triangle, to have maximum area use all the wire for the square!

Exercise 49 Consider the optimisation problem

$$\begin{cases} \min\limits_{x_1,x_2} \; -x_1 + x_2, \\ 0 \le x_1 \le 1, \\ x_2 \ge x_1^2. \end{cases}$$

a) State first order necessary conditions of optimality for this constrained optimisation problem.

b) Using the conditions in part a) determine a candidate optimal solution $x^\star$ for the considered optimisation problem.

c) Transform the optimization problem into an optimization problem with equality constrains by adding an auxiliary variable and disregarding, for simplicity, the constraints $0 \le x_1 \le 1$.

State first order necessary conditions of optimality for this transformed problem. Determine a candidate optimal solution and show that it coincides with the solution determined in part b).

Solution 49

a) Define the Lagrangian

$$L(x_1, x_2, \rho_1, \rho_2, \rho_3) = -x_1 + x_2 + \rho_1(-x_1) + \rho_2(x_1 - 1) + \rho_3(x_1^2 - x_2).$$

The first order necessary conditions of optimality are

$$0 = \frac{\partial L}{\partial x_1} = -1 - \rho_1 + \rho_2 + 2\rho_3 x_1, \qquad 0 = \frac{\partial L}{\partial x_2} = 1 - \rho_3,$$

$$-x_1 \le 0, \qquad x_1 - 1 \le 0, \qquad x_1^2 - x_2 \le 0, \qquad \rho_1 \ge 0, \qquad \rho_2 \ge 0, \qquad \rho_3 \ge 0,$$

$$\rho_1 x_1 = 0, \qquad \rho_2(x_1 - 1) = 0, \qquad \rho_3(x_1^2 - x_2) = 0.$$

b) To begin with note that ρ_3 has to be equal to one, hence $x_1^2 = x_2$. Consider now the following four possibilities.

- $\rho_1 = 0$, $\rho_2 = 0$. This yields $x_1 = 1/2$ and $x_2 = 1/4$.
- $\rho_1 = 0$, $\rho_2 > 0$. This yields $x_1 = 1$, $x_2 = 1$, $\rho_2 = -1$, which is not admissible.
- $\rho_1 > 0$, $\rho_2 = 0$. This yields $x_1 = 0$, $x_2 = 0$, $\rho_1 = -1$, which is not admissible.
- $\rho_1 > 0$, $\rho_2 > 0$. This yields $x_1 = 0$ and $x_1 = 1$, which is meaningless.

In summary the only candidate solution is $x_1 = 1/2$, $x_2 = 1/4$, $\rho_1 = 0$, $\rho_2 = 0$, $\rho_3 = 1$.

c) The inequality constraint $x_1^2 - x_2 \le 0$ can be rewritten as $x_1^2 - x_2 + y^2 = 0$, where y is an auxiliary variable. The problem is thus transformed into the problem

$$\begin{cases} \min\limits_{x_1,x_2,y} \; -x_1 + x_2, \\ x_1^2 - x_2 + y^2 = 0. \end{cases}$$

The Lagrangian for this problem is $L = -x_1 + x_2 + \lambda(x_1^2 - x_2 + y^2)$, and the necessary conditions of optimality are $0 = \frac{dL}{dx_1} = -1 + 2\lambda x_1$, $\quad 0 = \frac{dL}{dx_2} = 1 - \lambda$, $\quad 0 = \frac{dL}{dy} = 2\lambda y$, $\quad x_1^2 - x_2 + y^2 = 0. The only candidate solution is \lambda = 1$, $y = 0$, $x_1 = 1/2$, $x_2 = 1/4$, which coincides with the one determined in part b).

Exercise 50 Consider the optimisation problem

$$\begin{cases} \min\limits_{y_1,y_2} \ y_1 y_2, \\ y_1^2 + y_2^2 \leq 1. \end{cases}$$

a) Sketch in the (y_1, y_2)-plane the admissible set and the level lines of the the function $y_1 y_2$. Hence, using only graphical considerations determine the optimal solutions of the considered problem.

b) State first order necessary conditions of optimality for this constrained optimisation problem.

c) Using the conditions derived in part b) compute candidate optimal solutions. Show that the optimal solutions derived graphically in part a) satisfy the necessary conditions of optimality.

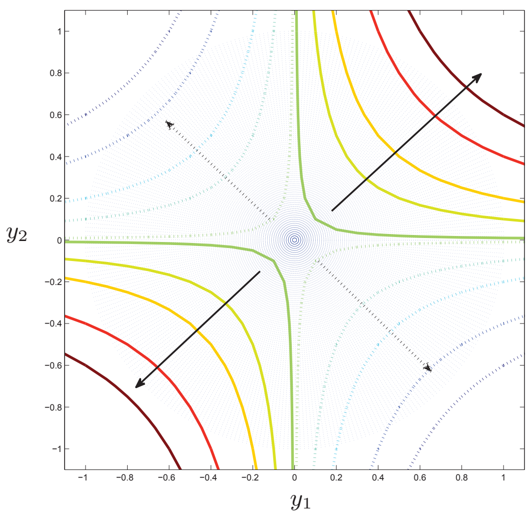
d) The considered problem can be transformed into a linear programming problem using the change of variable $x_1 = (y_1 - y_2)^2, x_2 = (y_1 + y_2)^2$.

   i) Write the equations describing the transformed problem.
   (Hint: note that the transformed problem has three inequality constraints.)

   ii) Sketch in the (x_1, x_2)-plane the admissible set and the level lines of the cost function. Hence determine the optimal solution of the transformed problem.

   iii) Show how the optimal solution of the transformed problem can be used to determine the optimal solutions of the original problem.

Solution 50



a) The admissible set is the shaded area in the figure above. The level lines are the solid (positive values of f) and dotted (negative values of f) lines. The value of the function f increases in the direction of the solid arrows, and decreases in the direction of the dotted arrows. The solution of the problem is obtained for negative values of f at a point in which the level lines are tangent to the circle $y_1^2 + y_2^2 = 1$. At such points $y_1 = -y_2$, hence the (global) minimizers are the point

$$P_1 = \left(\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2} \right), \qquad P_2 = \left(\frac{-\sqrt{2}}{2}, \frac{\sqrt{2}}{2} \right).$$

The value of the function at the optimal points is $f(P_1) = f(P_2) = -\frac{1}{2}$.

b) The Lagrangian of the problem is

$$L(y_1, y_2, \rho) = y_1 y_2 + \rho(y_1^2 + y_2^2 - 1).$$

The first order necessary conditions of optimality are

$$0 = \frac{\partial L}{\partial y_1} = y_2 + 2\rho y_1, \qquad 0 = \frac{\partial L}{\partial y_2} = y_1 + 2\rho y_2,$$

$$y_1^2 + y_2^2 - 1 \leq 0, \qquad \rho \geq 0, \qquad \rho(y_1^2 + y_2^2 - 1) = 0.$$

c) Consider the two cases.

- $\rho = 0$. In this case $y_1 = y_2 = 0$.
- $\rho > 0$. Consider the equations
$$0 = \frac{\partial L}{\partial y_1} = \frac{\partial L}{\partial y_2}.$$
For any $\rho > 0$, $y_1 = y_2 = 0$ is a solution. In addition, if $\rho = 1/2$ there are infinitely many solutions of the form $(y_1, y_2) = (\alpha, -\alpha)$, where α is any real number. Note now that $\rho > 0$ implies, by the complementarity condition, $y_1^2+y_2^2-1 = 0$. Hence, $2\alpha^2 = 1$, yielding $\alpha = \pm\frac{\sqrt{2}}{2}$.

In summary the candidate optimal solutions are

- $(y_1, y_2) = (0, 0)$, with $\rho \geq 0$.
- P_1 and P_2 with $\rho = 1/2$.

d) i) Note that x_1 and x_2 are non-negative by definition and that $x_1 = y_1^2 - 2y_1y_2 + y_2^2$, and $x_2 = y_1^2 + 2y_1y_2 + y_2^2$. Hence

$$\frac{x_2 - x_1}{4} = y_1 y_2, \qquad \frac{x_1 + x_2}{2} = y_1^2 + y_2^2.$$

As a result, in the variables x_1 and x_2 the problem is

$$\begin{cases} \min\limits_{x_1, x_2} \dfrac{x_2 - x_1}{4}, \\ x_1 \geq 0, \\ x_2 \geq 0, \\ \dfrac{x_1 + x_2}{2} \leq 1. \end{cases}$$

ii) The admissible set is the shaded area in the figure below. The level lines are the solid (positive values of f) and dotted (negative values of f) lines. The value of the function f increases in the direction of the solid arrow. The optimal solution is the point

$$P = (2, 0).$$

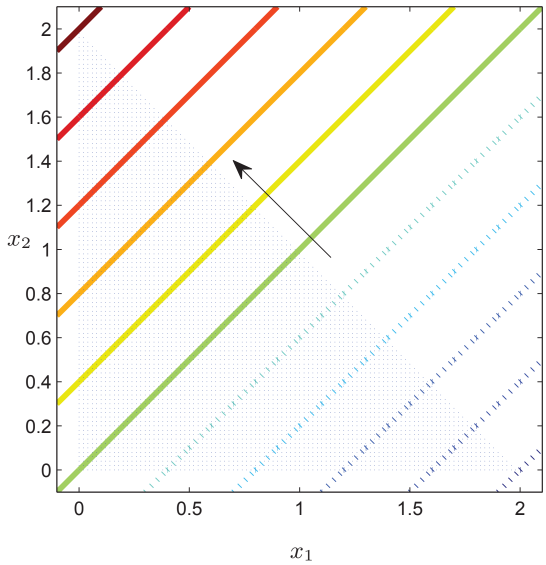
iii) The point P in the (x_1, x_2)-variables is transformed into points in the (y_1, y_2)-variables solving the equations

$$(y_1 - y_2)^2 = 2 \qquad (y_1 + y_2)^2 = 0.$$

These equations have the solutions

$$(y_1, y_2) = \left(\pm\frac{\sqrt{2}}{2}, \mp\frac{\sqrt{2}}{2}\right),$$

which coincide with the points P_1 and P_2 determined in part c).



2
1.8
1.6
1.4
1.2
x_2
1
0.8
0.6
0.4
0.2
0
0
0.5
1
1.5
2
x_1

Chapter 4

Global optimization

4.1 Introduction

Given a function $f : I\!R^n \to I\!R$, global optimization methods aim at finding the global minimum of f, *i.e.* a point $x^\star$ such that

$$f(x^\star) \le f(x)$$

for all $x \in I\!R^n$. Among these methods it is possible to distinguish between deterministic methods and probabilistic methods. In the following sections we provide a very brief introductions to global minimization methods. It is worth noting that this is an active area of research.

4.2 Deterministic methods

4.2.1 Methods for Lipschitz functions

Consider a function $f : I\!R^n \to I\!R$ and suppose it is Lipschitz with constant $L > 0$, *i.e.*

$$|f(x_1) - f(x_2)| \le L\|x_1 - x_2\|, \tag{4.1}$$

for all $x_1 \in I\!R^n$ and $x_2 \in I\!R^n$. Note that equation (4.1) implies that

$$f(x) \ge f(x_0) - L\|x - x_0\| \tag{4.2}$$

and that

$$f(x) \le f(x_0) + L\|x - x_0\|, \tag{4.3}$$

for all $x \in I\!R^n$ and $x_0 \in I\!R^n$, see Figure 4.1 for a geometrical interpretation.

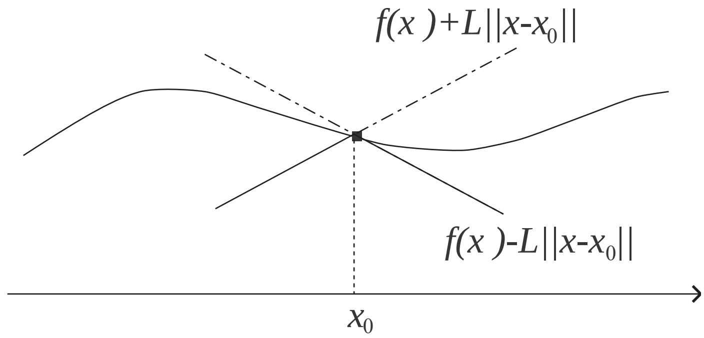


Figure 4.1: Geometrical interpretation of the Lipschitz conditions (4.2) and (4.3).

Methods for Lipschitz functions are suitable to find a global solution of the problem

$$\min_x f(x),$$

with

$$x \in I_n = \{x \in I\!R^n \mid A_i \le x_i \le B_i\},$$

and $A_i < B_i$ given, under the assumptions that the set I_n contains a global minimizer of f, the function f is Lipschitz in I_n and the Lipschitz constant L of f in I_n is known. Under these assumptions it is possible to construct a very simple global minimization algorithm, known as Schubert-Mladineo algorithm, as follows.

Step 0. Given $x_0 \in I_n$ and $\tilde{L} > L$.

Step 1. Set $k = 0$.

Step 2. Let

$$F_k(x) = \max_{j=0,\cdots,k} \{f(x_j) - \tilde{L}\|x - x_j\|\}$$

and compute x_{k+1} such that

$$F_k(x_{k+1}) = \min_{x \in I_n} F_k(x).$$

Step 4. Set $k = k + 1$ and go to **Step 2**.

Remark. The functions F_k in **Step 2** of the algorithm have a very special form. This can be exploited to construct special algorithms solving the problem

$$\min_{x \in I_n} F_k(x)$$

in a finite number of iterations. ◇

For Schubert-Mladineo algorithm it is possible to prove the following statement.

Theorem 24 *Let $f^\star$ be the minimum value of f in I_n, let $x^\star$ be such that $f(x^\star) = f^\star$ and let $F_k^\star$ be the minima of the functions F_k in I_n. Let*

$$\Phi = \{x \in I_n \mid f(x) = f^\star\}$$

and let $\{x_k\}$ be the sequence generated by the algorithm. Then

- $\lim_{k \to \infty} f(x_k) = f^\star$;
- *the sequence $\{F_k^\star\}$ is non-decreasing and* $\lim_{k \to \infty} F_k^\star = f^\star$;
- $\lim_{k \to \infty} \inf_{x \in \Phi} \|x - x_k\| = 0$;
- $f(x_k) \geq f^\star \geq F_{k-1}(x_k)$.

Schubert-Mladineo algorithm can be given, if $x \in I_1 \subset I\!R$, a simple geometrical interpretation, as shown in Figure 4.2.
The main advantage of Schubert-Mladineo algorithm is that it does not require the computation of derivatives, hence it is also applicable to functions which are not everywhere

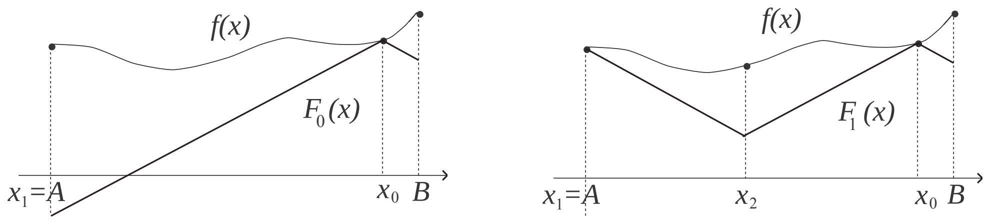


Figure 4.2: Geometrical interpretation of Schubert-Mladineo algorithm.

differentiable. Moreover, unlike other global minimization algorithms, it is possible to prove the convergence of the sequence $\{x_k\}$ to the global minimizer. Finally, it is possible to define a simple *stopping* condition. For, note that if $\{x_k\}$ and $\{F_k^\star\}$ are the sequences generated by the algorithm, then

$$f(x_k) \geq f^\star \geq F_k^\star$$

and

$$f(x_k) \geq f^\star \geq f(x_k) + r_k,$$

where $r_k = F_k^\star - f(x_k)$ and $\lim_{k\to\infty} r_k = 0$. As a result, if $|r_k| < \epsilon$, for some $\epsilon > 0$, the point x_k gives a good approximation of the minimizer of f.
The main disadvantage of the algorithm is in the assumption that the set I_n contains a global minimizer of f in $I\!R^n$. Moreover, it may be difficult to compute the Lipschitz constant L.

4.2.2 Methods of the trajectories

The basic idea of the global optimization methods known as methods of the trajectories is to construct trajectories which go through all local minimizers. Once all local minimizers are determined, the global minimizer can be easily isolated. These methods have been originally proposed in the 70's, but only recently, because of increased computer power and of a reformulation using tools from differential geometry, they have proved to be effective.
The simplest and first method of the trajectories is the so-called Branin method. Consider the function f and assume ∇f is continuous. Fix x_0 and consider the differential equations

$$\frac{d}{dt}\nabla f(x(t)) = \pm\nabla f(x(t)) \qquad x(0) = x_0. \tag{4.4}$$

The solutions $x(t)$ of such differential equations are such that

$$\nabla f(x(t)) = \nabla f(x_0)e^{\pm t},$$

i.e. $\nabla f(x(t))$ is parallel to $\nabla f(x_0)$ for all t. Using these facts it is possible to describe Branin algorithm.

Step 0. Given x_0.

Step 1. Compute the solution $x(t)$ of the differential equation

$$\frac{d}{dt}\nabla f(x(t)) = -\nabla f(x(t))$$

with $x(0) = x_0$.

Step 2. The point $x^\star = \lim_{t\to\infty} x(t)$ is a stationary point of f, in fact $\lim_{t\to\infty} \nabla f(x(t)) = 0$.

Step 3. Consider a perturbation of the point $x^\star$, *i.e.* the point $\tilde{x} = x^\star + \epsilon$ and compute the solution $x(t)$ of the differential equation

$$\frac{d}{dt}\nabla f(x(t)) = \nabla f(x(t)).$$

Along this trajectory the gradient $\nabla f(x(t))$ increases, hence the trajectory escapes from the *region of attraction* of x_0.

Step 4. Fix $\bar{t} > 0$ and assume that $x(\bar{t})$ is sufficiently away from x_0. Set $x_0 = x(\bar{t})$ and go to **Step 1**.

Note that, if the perturbation ϵ and the time $\bar{t}$ are properly selected, at each iteration the algorithm generates a new stationary point of the function f.

Remark. If $\nabla^2 f$ is continuous then the differential equations (4.4) can be written as

$$\dot{x}(t) = \pm\left[\nabla^2 f(x(t))\right]^{-1}\nabla f(x(t)).$$

Therefore Branin method is a continuous equivalent of Newton method. Note however that, as $\nabla^2 f(x(t))$ may become singular, the above equation may be meaningless. In such a case it is possible to modify Branin method using ideas borrowed from quasi-Newton algorithms. ◇

Branin method is very simple to implement. However, it has several disadvantages.

- It is not possible to prove convergence to the global minimizer.
- Even if the method yields the global minimizer, it is not possible to know how many iterations are needed to reach such a global minimizer, *i.e.* there is no stopping criterion.
- The trajectories $x(t)$ are attracted by all stationary points of f, *i.e.* both minimizers and maximizers.
- There is not a systematic way to select ϵ and $\bar{t}$.

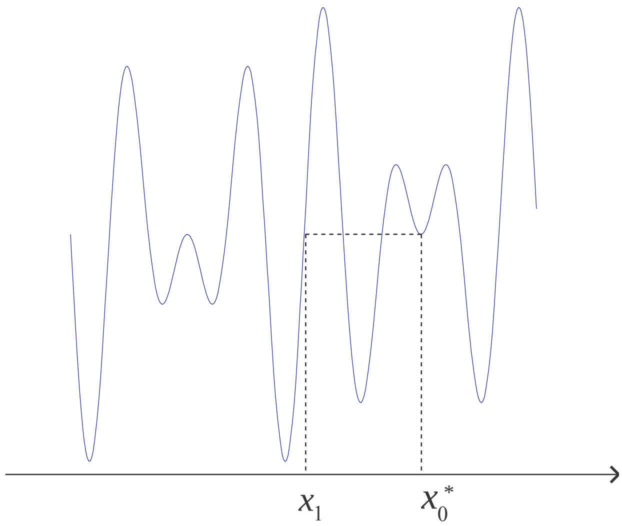


Figure 4.3: Interpretation of the tunneling phase.

4.2.3 Tunneling methods

Tunneling methods have been proposed to find, in an efficient way, the global minimizer of a function with several (possibly thousands) of local minimizers.
Tunneling algorithms are composed of a sequence of cycles, each having two phases. The first phase is the minimization phase, *i.e.* a local minimizer is computed. The second phase is the tunneling phase, *i.e.* a new starting point for the minimization phase is computed.

Minimization phase

Given a point x_0, a local minimization, using any unconstrained optimization algorithm, is performed. This minimization yields a local minimizer $x_0^\star$.

Tunneling phase

A point $x_1 \neq x_0^\star$ such that

$$f(x_1) = f(x_0^\star)$$

is determined. See Figure 4.3 for a geometrical interpretation.

In theory, tunneling methods generate a sequence $\{x_k^\star\}$ such that

$$f(x_{k+1}^\star) \leq f(x_k^\star)$$

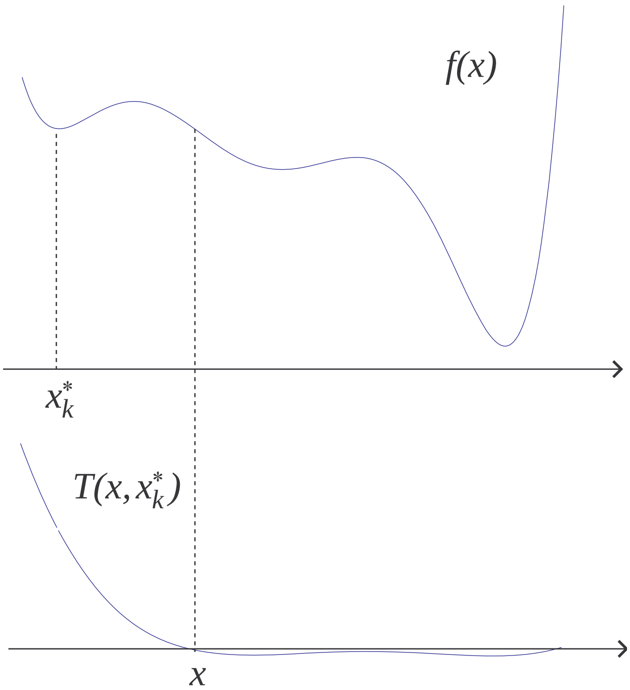


Figure 4.4: The functions $f(x)$ and $T(x, x_k^\star)$.

and the sequence $\{x_k^\star\}$ converges to the global minimizer without *passing* through all local minimizers. This is the most important advantage of tunneling methods. The main disadvantage is the difficulty in performing the tunneling phase. In general, given a point $x_k^\star$ a point x such that $f(x) = f(x_k^\star)$ is constructed searching for a zero of the function (see Figure 4.4)

$$T(x, x_k^\star) = \frac{f(x) - f(x_k^\star)}{\|x - x_k^\star\|^{2\lambda}},$$

where the parameter $\lambda > 0$ has to be selected such that $T(x_k^\star, x_k^\star) > 0$.
Finally, it is worth noting that tunneling methods do not have a stopping criterion, *i.e.* the algorithm attempts to perform the tunneling phase even if the point $x_k^\star$ is a global minimizer.

4.3 Probabilistic methods

4.3.1 Methods using random directions

In this class of algorithms at each iteration a randomly selected direction, having unity norm, is selected. The theoretical justification of such an algorithm rests on Gaviano theorem. This states that the sequence $\{x_k\}$ generated using the iteration

$$x_{k+1} = x_k + \alpha_k d_k,$$

where d_k is randomly selected on a unity norm sphere and α_k is such that

$$f(x_k + \alpha_k d_k) = \min_{\alpha} f(x_k + \alpha d_k),$$

is such that for any $\epsilon > 0$ the probability that

$$f(x_k) - f^\star < \epsilon,$$

where $f^\star$ is a global minimum of f, tends to one as $k \to \infty$.

4.3.2 Multistart methods

Multistart methods are based on the fact that for given sets D and A, with measures $m(D)$ and $m(A)$, and such that

$$1 \geq \frac{m(A)}{m(D)} = \alpha \geq 0,$$

the probability that, selecting N random points in D, one of these points is in A is

$$P(A, N) = 1 - (1 - \alpha)^N.$$

As a result

$$\lim_{N \to \infty} P(A, N) = 1.$$

Therefore, if A is a neighborhood of a global minimizer of f in D, we conclude that, selecting a sufficiently large number of random points in D, one of these will (almost surely) be close to the global minimizer. Using these considerations it is possible to construct a whole class of algorithms, with similar properties, as detailed hereafter.

Step 0. Set $f^\star = \infty$.

Step 1. Select a random point $x_0 \in I\!R^n$.

Step 2. If $f(x_0) > f^\star$ go to **Step 1**.

Step 3. Perform a local minimization starting from x_0 and yielding a point $x_0^\star$. Set $f^\star = f(x_0^\star)$.

Step 4. Check if $x_0^\star$ satisfies a stopping criterion. If not, go to **Step 1**.

4.3.3 Stopping criteria

The main disadvantage of probabilistic algorithms is the lack of a theoretically sound stopping criterion. The most promising and used stopping criterion is based on the construction of a *probabilistic approximation* $\tilde{P}(w)$ of the function

$$P(w) = \frac{m(\{x \in D \mid f(x) \leq w\})}{m(D)}.$$

Once the function $\tilde{P}(w)$ is known, a point $x^\star$ is regarded as a good approximation of the global minimizer of f if

$$\tilde{P}(f(x^\star)) \leq \epsilon \ll 1.$$

4.4 Exercises

Similarly to Sections 2.10 and 3.7, this section contains some exercises to illustrate how global minimization methods can be used.

Exercise 51 Consider the discrete time system

$$x_{k+1} = a x_k$$

with $x_k \in I\!R$, and output $y_k = x_k$. Consider also the auxiliary discrete time system

$$\xi_{k+1} = \alpha \xi_k$$

with $\xi_k \in I\!R$, output $\eta_k = \xi_k$, and such that $\xi_0 = x_0 \neq 0$.
Consider now the problem of determining the constant α such that the cost

$$J(\alpha) = \frac{1}{2}\left(e_1^2 + e_2^2 + \cdots e_N^2\right)$$

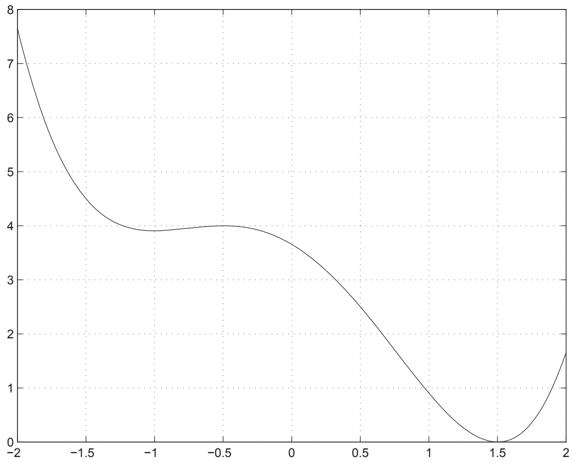
is minimized, where $e_i = y_i - \eta_i$ and $N \geq 1$.

a) Pose the above problem as an unconstrained optimization problem in the decision variable α, parametrized by a and x_0.

b) Assume $N = 1$. Show that $J(a) = 0$ and $J(\alpha) > 0$ for all $\alpha \neq a$. Hence show that the function $J(\alpha)$ has a unique local minimizer which is also a global minimizer.

c) Suppose $N = 2$. Compute the stationary points of $J(\alpha)$. Note that the number of stationary points is a function of the value of a. Hence, determine the local minimizers and the local maximizers of the function $J(\alpha)$.

d) For $N = 2$ and $a = 3/2$, the function $J(\alpha)$ is as shown in the figure below. Let $L = 12$ be the Lipschitz constant of $J(\alpha)$ for $\alpha \in [-2, 2]$. Apply four steps of the Schubert-Mladineo algorithm for the minimization of the function $J(\alpha)$ assuming that a global minimizer is in the set $I_1 = \{\alpha \in I\!R \mid -2 \leq \alpha \leq 2\}$ and that the starting point of the algorithm is selected to be $\alpha = 2$.



Solution 51

a) The problem can be formulated as

$$\min_{\alpha \in I\!R} J(\alpha) = \min_{\alpha \in I\!R} \frac{1}{2}(a - \alpha)^2 x_0^2 + (a^2 - \alpha^2)^2 x_0^2 + \cdots + (a^N - \alpha^N)^2 x_0^2,$$

i.e. as an unconstrained optimization problem in the decision variable α and parameterized by a and x_0.

b) For $N = 1$ one has $J(\alpha) = \frac{1}{2}(a-\alpha)^2 x_0^2$. Hence, $J(a) = 0$ and $J(\alpha) > 0$ for all $\alpha \neq a$. This shows (using the very definition of global minimum) that $\alpha = a$ is a global minimum.

c) If $N = 2$ one has $J(\alpha) = \frac{1}{2}\left((a-\alpha)^2 + (a^2-\alpha^2)^2\right) x_0^2$. Hence $\frac{dJ(\alpha)}{d\alpha} = -(a-\alpha)(2\alpha^2 + 2\alpha a + 1)x_0^2$. Therefore, the stationary points are

$$P_1 = a, \qquad P_2 = -\frac{a}{2} + \frac{\sqrt{a^2-2}}{2}, \qquad P_3 = -\frac{a}{2} - \frac{\sqrt{a^2-2}}{2}.$$

We conclude that, if $|a| < \sqrt{2}$ there is only one stationary point, whereas if $|a| \geq \sqrt{2}$ there are three stationary points. Computing second derivatives we have that P_1 is always a local minimizer, and, for $|a| \geq \sqrt{2}$, P_2 is a local maximizer and P_3 is a local minimizer.

d) A sketch of the application of the Schubert-Mladineo algorithm is shown at the end of the chapter. Note that x_4 is very close to the global minimizer.

Exercise 52 Consider the function

$$f = x_1^4 - x_1 x_2 + x_2^4$$

and the problem of finding its global minimizer.

a) Write the formulae for the so-called Branin system, that is the system

$$\dot{x} = -[\nabla^2 f]^{-1} \nabla f,$$

for the considered function f.

b) Compute the equilibria of the Branin system determined in part a). Show that these equilibria coincide with the stationary points of the function f. Show that f is radially unbounded. Hence determine the global minimizer of f.

c) Consider the linearization of the Branin system, computed in part a), around its equilibrium at $x = 0$. Show that this linearized system has two eigenvalues equal to -1, hence deduce that the point $x = 0$ is locally attractive.

d) Write now the formulae for the modified Branin system

$$\dot{x} = -\det(\nabla^2 f)[\nabla^2 f]^{-1} \nabla f,$$

for the function f above. Consider the linearization of the modified Branin system at $x = 0$ and show that this equilibrium point is unstable.

e) Give reasons for the modified Branin method being preferable to the Branin method when determining a global minimizer for the considered function f.

Solution 52

a) The Branin system is

$$\dot{x} = \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \frac{1}{144x_1^2x_2^2 - 1} \begin{bmatrix} -48x_2^2x_1^3 - 8x_2^3 - x_1 \\ 8x_1^3 + x_2 - 48x_1^2x_2^3 \end{bmatrix}.$$

b) The equilibria of the Branin system are $P_1 = (0,0)$, $P_2 = (1/2, 1/2)$ and $P_3 = (-1/2, -1/2)$. Note now that these are also such that $\nabla f(P_i) = 0$, for $i = 1,2,3$. Hence, the equilibria of Branin system coincide with the stationary points of f. Note now that $\lim_{\|x\|\to\infty} f(x) = +\infty$, hence f is radially unbounded. Moreover $f(P_1) = 0$ and $f(P_2) = f(P_3) = -1/8$. Hence the global minimum of f is $-1/8$ and there are two global minimizers, P_2 and P_3.

c) The linearization of the Branin system around the point P_1 is described by

$$\dot{x} = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} x.$$

The linearized system has two eigenvalues equal to -1 and this shows that the point P_1 is locally attractive.

d) The modified Branin system is

$$\dot{x} = \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} -48x_2^2x_1^3 - 8x_2^3 - x_1 \\ 8x_1^3 + x_2 - 48x_1^2x_2^3 \end{bmatrix}.$$

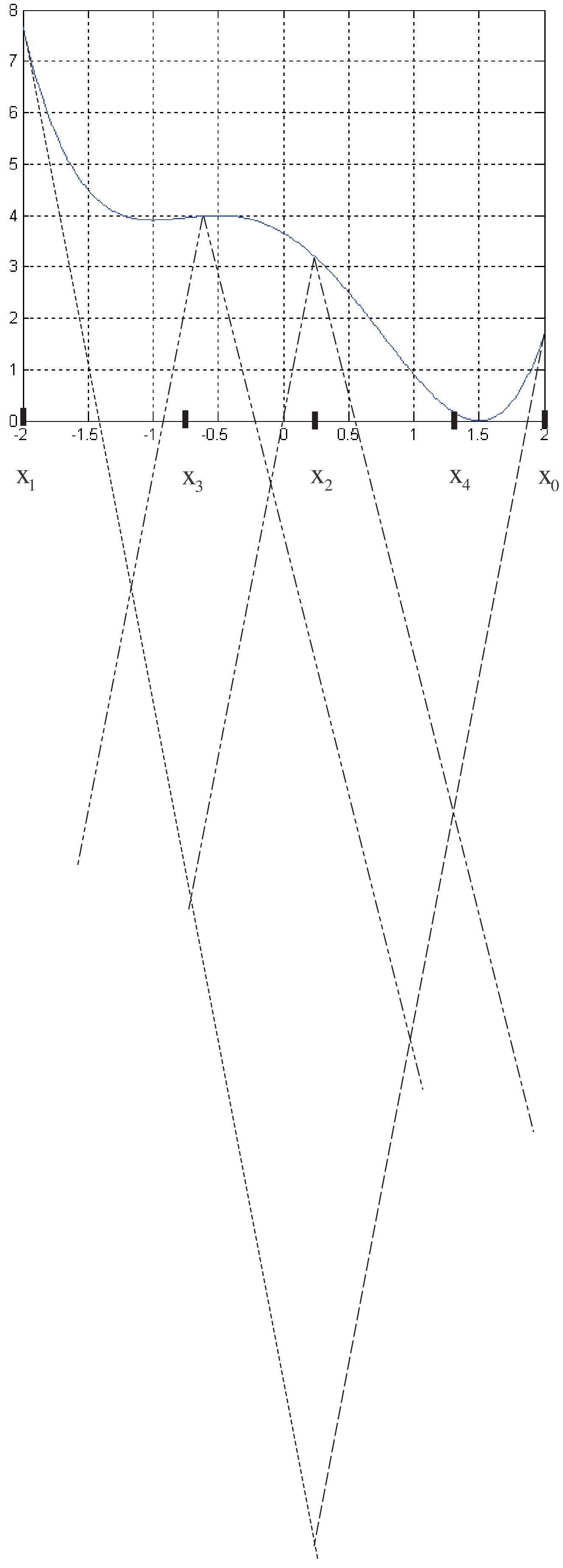
Its linearization around P_1 is described by

$$\dot{x} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} x,$$

and this shows that the point P_1 is an unstable equilibrium of the modified Branin system.

e) The modified Branin system has the following advantages:

- the differential equations are defined for all x;
- the point P_1, which is a local maximizer, is unstable therefore almost all trajectories of the system are not be *attracted* by P_1.


8
7
6
5
4
3
2
1
0
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
x_1
x_3
x_2
x_4
x_0