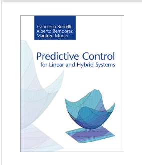


Predictive Control: for Linear and Hybrid Systems
ISBN 9781139061759
I Basics of Optimization
1 Main Concepts

# 1
# Main Concepts

In this chapter, we recall the main concepts and definitions of continuous and discrete optimization. Our intent is to provide only the necessary background for the understanding of the rest of the book. The notions of feasibility, optimality, convexity and active constraints introduced in this chapter will be widely used in this book.

## 1.1 Optimization Problems

An optimization problem is generally formulated as

$$\begin{array}{ll} \inf_z & f(z) \\ \text{subj. to} & z \in S \subseteq Z, \end{array} \tag{1.1}$$

where the vector $z$ collects the decision variables, $Z$ is the optimization problem *domain*, and $S \subseteq Z$ is the set of *feasible* or *admissible* decisions. The function $f$: $Z \to \mathbb{R}$ assigns to each decision $z$ a *cost* $f(z) \in \mathbb{R}$. We will often use the following shorter form of problem (1.1)

$$\inf_{z \in S \subseteq Z} f(z). \tag{1.2}$$

Solving problem (1.2) means to compute the least possible cost $f^*$

$$f^* = \inf_{z \in S} f(z).$$

The number $f^*$ is the *optimal value* of problem (1.2), i.e.,

$$f(z) \geq f(z^*) = f^* \ \forall z \in S, \ \text{with } z^* \in S,$$

or the greatest lower bound of $f(z)$ over the set $S$:

$$f(z) > f^* \ \forall z \in S \ \text{ and } (\forall \varepsilon > 0 \ \exists z \in S : \ f(z) \leq f^* + \varepsilon).$$

If $f^* = -\infty$ we say that the problem is unbounded below. If the set $S$ is empty then the problem is said to be *infeasible* and we set $f^* = +\infty$ by convention. If $S = Z$ the problem is said to be *unconstrained*.

In general, one is also interested in finding an *optimal solution*, that is in finding a decision whose associated cost equals the optimal value, i.e., $z^* \in S$ with $f(z^*) = f^*$. If such $z^*$ exists, then we rewrite problem (1.2) as

$$f^* = \min_{z \in S} f(z) \tag{1.3}$$

and $z^*$ is called an *optimizer*, *global optimizer* or *optimal solution*. *Minimizer* or *global minimizer* are also used to refer to an optimizer of a minimization problem. The set of all optimal solutions is denoted by

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

A problem of determining whether the set of feasible decisions is empty and, if not, to find a point which is feasible, is called a *feasibility problem*.

### 1.1.1 Continuous Problems

In continuous optimization the problem domain $Z$ is a subset of the finite-dimensional Euclidian vector-space $\mathbb{R}^s$ and the subset of admissible vectors is defined through a list of equality and inequality constraints:

$$\begin{array}{lll} \inf_z & f(z) & \\ \text{subj. to} & g_i(z) \leq 0 & \text{for } i = 1, \ldots, m \\ & h_j(z) = 0 & \text{for } j = 1, \ldots, p \\ & z \in Z, & \end{array} \tag{1.4}$$

where $f, g_1, \ldots, g_m, h_1, \ldots, h_p$ are real-valued functions defined over $\mathbb{R}^s$, i.e., $f : \mathbb{R}^s \to \mathbb{R}$, $g_i : \mathbb{R}^s \to \mathbb{R}$, $h_i : \mathbb{R}^s \to \mathbb{R}$. The domain $Z$ is the intersection of the domains of the cost and constraint functions:

$$Z = \{z \in \mathbb{R}^s \ : \ z \in \text{dom } f, \ z \in \text{dom } g_i, \ i = 1, \ldots, m, \ z \in \text{dom } h_j, \ j = 1, \ldots, p\}. \tag{1.5}$$

In the sequel we will consider the constraint $z \in Z$ implicit in the optimization problem and often omit it. Problem (1.4) is unconstrained if $m = p = 0$.

The inequalities $g_i(z) \leq 0$ are called *inequality constraints* and the equations $h_i(z) = 0$ are called *equality constraints*. A point $\bar{z} \in \mathbb{R}^s$ is *feasible* for problem (1.4) if: (i) it belongs to $Z$, (ii) it satisfies all inequality and equality constraints, i.e., $g_i(\bar{z}) \leq 0$, $i = 1, \ldots, m$, $h_j(\bar{z}) = 0$, $i = j, \ldots, p$. The set of feasible vectors is

$$S = \{z \in \mathbb{R}^s : z \in Z, \ g_i(z) \leq 0, \ i = 1, \ldots, m, \ \ h_j(z) = 0, \ j = 1, \ldots, p\}. \tag{1.6}$$

Problem (1.4) is a continuous finite-dimensional optimization problem (since $Z$ is a finite-dimensional Euclidian vector space). We will also refer to (1.4) as a *nonlinear mathematical program* or simply *nonlinear program*. Let $f^*$ be the optimal value of problem (1.4). An optimizer, if it exists, is a feasible vector $z^*$ with $f(z^*) = f^*$.

A feasible point $\bar{z}$ is *locally optimal* for problem (1.4) if there exists an $R > 0$ such that

$$
\begin{array}{lll}
f(\bar{z}) = \inf_z & f(z) & \\
\quad\quad \text{subj. to} & g_i(z) \le 0 & \text{for } i = 1, \ldots, m \\
& h_i(z) = 0 & \text{for } i = 1, \ldots, p \\
& \|z - \bar{z}\| \le R & \\
& z \in Z. &
\end{array}
\tag{1.7}
$$

Roughly speaking, this means that $\bar{z}$ is the minimizer of $f(z)$ in a feasible neighborhood of $\bar{z}$ defined by $||z - \bar{z}|| \le R$. The point $\bar{z}$ is called a *local optimizer* or *local minimizer*.

### Active, Inactive and Redundant Constraints

Consider a feasible point $\bar{z}$. We say that the $i$-th inequality constraint $g_i(z) \le 0$ is *active* at $\bar{z}$ if $g_i(\bar{z}) = 0$. If $g_i(\bar{z}) < 0$ we say that the constraint $g_i(z) \le 0$ is *inactive* at $\bar{z}$. Equality constraints are always active for all feasible points.

We say that a constraint is *redundant* if removing it from the list of constraints does not change the feasible set $S$. This implies that removing a redundant constraint from problem (1.4) does not change its solution.

### Problems in Standard Forms

Optimization problems can be cast in several forms. In this book, we use the form (1.4) where we adopt the convention to minimize the cost function and to have the right-hand side of the inequality and equality constraints equal to zero. Any problem in a different form (e.g., a maximization problem or a problem with "box constraints") can be transformed and arranged into this form. The interested reader is referred to Chapter 4 of [65] for a detailed discussion on transformations of optimization problems into different standard forms.

### Eliminating Equality Constraints

Often in this book we will restrict our attention to problems without equality constraints, i.e., $p = 0$

$$
\begin{array}{lll}
\inf_z & f(z) & \\
\text{subj. to} & g_i(z) \le 0 & \text{for } i = 1, \ldots, m.
\end{array}
\tag{1.8}
$$

The simplest way to remove equality constraints is to replace them with two inequalities for each equality, i.e., $h_i(z) = 0$ is replaced by $h_i(z) \le 0$ and $-h_i(z) \le 0$. Such a method, however, can lead to poor numerical conditioning and may ruin the efficiency and accuracy of a numerical solver.

If one can explicitly parameterize the solution of the equality constraint $h_i(z) = 0$, then the equality constraint can be *eliminated* from the problem. This process can be described in a simple way for linear equality constraints. Assume the equality constraints to be linear, $Az - b = 0$, with $A \in \mathbb{R}^{p\times s}$. If $Az = b$ is inconsistent then the problem is infeasible. The general solution of the equation $Az = b$ can be expressed as $z = Fx + z_0$ where $F$ is a matrix of full rank whose spanned space coincides with the null space of the $A$ matrix, i.e., $\mathcal{R}(F) = \mathcal{N}(A)$, $F \in \mathbb{R}^{s\times k}$, where $k$ is the dimension of the null space of $A$. The variable $x \in \mathbb{R}^k$ is the new optimization variable and the original problem becomes

$$\begin{array}{ll} \inf_x & f(Fx+z_0) \\ \text{subj. to} & g_i(Fx+z_0) \leq 0 \quad \text{for } i=1,\ldots,m. \end{array} \tag{1.9}$$

We want to point out that in some cases the elimination of equality constraints can make the problem harder to analyze and understand and can make a solver less efficient. In large problems it can destroy useful structural properties of the problem such as sparsity. Some advanced numerical solvers perform elimination automatically.

**Problem Description**

The functions $f$, $g_i$ and $h_i$ can be available in analytical form or can be described through an *oracle model* (also called "black box" or "subroutine" model). In an oracle model, $f$, $g_i$ and $h_i$ are not known explicitly but can be evaluated by querying the oracle. Often the oracle consists of subroutines which, called with the argument $z$, return $f(z)$, $g_i(z)$ and $h_i(z)$ and their gradients $\nabla f(z)$, $\nabla g_i(z)$, $\nabla h_i(z)$. In the rest of the book we assume that analytical expressions of the cost and the constraints of the optimization problem are available.

### 1.1.2 Integer and Mixed-Integer Problems

If the decision set $Z$ in the optimization problem (1.2) is finite, then the optimization problem is called *combinatorial* or *discrete*. If $Z \subseteq \{0, 1\}^s$, then the problem is said to be *integer*.

If $Z$ is a subset of the Cartesian product of an integer set and a real Euclidian space, i.e., $Z \subseteq \{[z_c, z_b] : z_c \in \mathbb{R}^{s_c}, z_b \in \{0, 1\}^{s_b}\}$, then the problem is said to be *mixed-integer*. The standard formulation of a *mixed-integer nonlinear program* is

$$\begin{array}{ll} \inf_{[z_c,z_b]} & f(z_c,z_b) \\ \text{subj. to} & g_i(z_c,z_b) \leq 0 \quad \text{for } i=1,\ldots,m \\ & h_j(z_c,z_b) = 0 \quad \text{for } j=1,\ldots,p \\ & z_c \in \mathbb{R}^{s_c},\ \ z_b \in \{0,1\}^{s_b} \end{array} \tag{1.10}$$

where $f$, $g_1$,..., $g_m$, $h_1$,..., $h_p$ are real-valued functions defined over $Z$.

For combinatorial, integer and mixed-integer optimization problems, all definitions introduced in the previous section apply.

## 1.2 Convexity

A set $S \in \mathbb{R}^s$ is *convex* if

$$\lambda z_1 + (1-\lambda) z_2 \in S \text{ for all } z_1 \in S, z_2 \in S \text{ and } \lambda \in [0,1].$$

A function $f : S \to \mathbb{R}$ is convex if $S$ is convex and

$$f(\lambda z_1 + (1-\lambda) z_2) \leq \lambda f(z_1) + (1-\lambda) f(z_2) \quad \text{for all } z_1 \in S, z_2 \in S \text{ and } \lambda \in [0,1].$$

A function $f : S \to \mathbb{R}$ is *strictly convex* if $S$ is convex and

$$f(\lambda z_1 + (1-\lambda)z_2) < \lambda f(z_1) + (1-\lambda) f(z_2)$$
$$\text{for all } z_1 \in S, z_2 \in S \text{ and } \lambda \in (0,1).$$

A twice differentiable function $f : S \to \mathbb{R}$ is *strongly convex* if the Hessian

$$\nabla^2 f(z) \succ 0 \text{ for all } z \in S.$$

A function $f : S \to \mathbb{R}$ is concave if $S$ is convex and $-f$ is convex.

### Operations Preserving Convexity

Various operations preserve convexity of functions and sets. A detailed list can be found in Chapter 3.2 of [65]. A few operations used in this book are mentioned below.

1. The intersection of an arbitrary number of convex sets is a convex set:

$$\text{if } S_1, S_2, \ldots, S_k \text{ are convex, then } S_1 \cap S_2 \cap \ldots \cap S_k \text{ is convex.}$$

   This property extends to the intersection of an infinite number of sets:

$$\text{if } S_n \text{ is convex } \forall n \in \mathbb{N}_+ \text{ then } \bigcap_{n \in \mathbb{N}_+} S_n \text{ is convex.}$$

   The empty set is convex because it satisfies the definition of convexity.
2. The sublevel sets of a convex function $f$ on $S$ are convex:

$$\text{if } f(z) \text{ is convex then } S_\alpha = \{z \in S : f(z) \le \alpha\} \text{ is convex } \forall \alpha \in \mathbb{R}.$$

3. If $f_1,\ldots, f_N$ are convex functions, then $\sum_{i=1}^N \alpha_i f_i$ is a convex function for all $\alpha_i \ge 0$, $i = 1,\ldots, N$.
4. The composition of a convex function $f(z)$ with an affine map $z = Ax + b$ generates a convex function $f(Ax + b)$ of $x$:

$$\text{if } f(z) \text{ is convex then } f(Ax+b) \text{ is convex on } \{x : Ax + b \in \text{dom}(f)\}.$$

5. Suppose $f(x) = h(g(x)) = h(g_1(x),\ldots, g_k(x))$ with $h : \mathbb{R}^k \to R$, $g_i : \mathbb{R}^s \to R$. Then,

   (a) $f$ is convex if $h$ is convex, $h$ is nondecreasing in each argument, and $g_i$ are convex,
   (b) $f$ is convex if $h$ is convex, $h$ is nonincreasing in each argument, and $g_i$ are concave,
   (c) $f$ is concave if $h$ is concave, $h$ is nondecreasing in each argument, and $g_i$ are concave.

6. The pointwise maximum of a set of convex functions is a convex function:

$$f_1(z), \ldots, f_k(z) \text{ convex functions} \Rightarrow f(z)$$
$$= \max\{f_1(z), \ldots, f_k(z)\} \text{ is a convex function.}$$

   This property holds also when the set is infinite.

### Linear and Quadratic Convex Functions

1. A linear function $f(z) = c'z + r$ is both convex and concave.
2. A quadratic function $f(z) = z'Hz + 2q'z + r$ is convex if and only if $H \succeq 0$.
3. A quadratic function $f(z) = z'Hz + 2q'z + r$ is strictly convex if and only if $H \succ 0$. A strictly

convex quadratic function is also strongly convex.

### Convex Optimization Problems

The standard optimization problem (1.4) is said to be *convex* if the cost function $f$ is convex on $Z$ and $S$ is a convex set. A fundamental property of convex optimization problems is that local optimizers are also global optimizers. This is proven next.

**Theorem 1.1** *Consider a convex optimization problem and let $\bar{z}$ be a local optimizer. Then, $\bar{z}$ is a global optimizer.*

*Proof:* By hypothesis $\bar{z}$ is feasible and there exists $R$ such that

$$f(\bar{z}) = \min\{f(z) \ : \ g_i(z) \leq 0 \ i = 1, \ldots, m, \ h_j(z) = 0, \ j = 1, \ldots, p \ \|z - \bar{z}\| \leq R\}. \tag{1.11}$$

Now suppose that $\bar{z}$ is not globally optimal. Then, there exist a feasible $y$ such that $f(y) < f(\bar{z})$, which implies that $\|y - \bar{z}\| > R$. Now consider the point $z$ given by

$$z = (1 - \theta)\bar{z} + \theta y, \quad \theta = \frac{R}{2\|y - \bar{z}\|}.$$

Then $\|z - \bar{z}\| = R/2 < R$ and by convexity of the feasible set $z$ is feasible. By convexity of the cost function $f$

$$f(z) \leq (1 - \theta)f(\bar{z}) + \theta f(y) < f(\bar{z}),$$

which contradicts (1.11). ■

Theorem 1.1 does not make any statement about the existence of a solution to problem (1.4). It merely states that all local minima of problem (1.4) are also global minima. For this reason, convexity plays a central role in the solution of continuous optimization problems. It suffices to compute a local minimum to problem (1.4) to determine its global minimum. Convexity also plays a major role in most nonconvex optimization problems which are solved by iterating between the solutions of convex subproblems.

It is difficult to determine whether the feasible set $S$ of the optimization problem (1.4) is convex or not except in special cases. For example, if the functions $g_1(z),\ldots, g_m(z)$ are convex and all the $h_i(z)$ (if any) are affine in $z$, then the feasible set $S$ in (1.6) is an intersection of convex sets and is therefore convex. Moreover there are nonconvex problems which can be transformed into convex problems through a change of variables and manipulations of cost and constraints. The discussion of this topic goes beyond the scope of this overview on optimization. The interested reader is referred to [65].

---

**Remark 1.1** With the exception of trivial cases, integer and mixed-integer optimization problems are always nonconvex problems because {0, 1} is not a convex set.

---

## 1.3 Optimality Conditions

In general, an analytical solution to problem (1.4), restated below, does not exist.

$$
\begin{array}{lll}
\inf_z & f(z) & \\
\text{subj. to} & g_i(z) \le 0 & \text{for } i = 1, \dots, m \\
& h_j(z) = 0 & \text{for } j = 1, \dots, p \\
& z \in Z. &
\end{array}
\tag{1.12}
$$

Solutions are usually computed by iterative algorithms which start from an initial guess $z_0$ and at step $k$ generate a point $z_k$ such that the sequence $\{f(z_k)\}_{k=0,1,2,\ldots}$ converges to $f^*$ as $k$ increases. These algorithms iteratively use and/or solve conditions for optimality, i.e., analytical conditions that a point $z$ must satisfy in order to be an optimizer. For instance, for *convex, unconstrained* optimization problems with a *smooth* cost function the most commonly used optimality criterion requires the gradient to vanish at the optimizer, i.e., $z$ is an optimizer if and only if $\nabla f(z) = 0$. In this chapter we summarize necessary and sufficient optimality conditions for unconstrained and constrained optimization problems.

### 1.3.1 Optimality Conditions for Unconstrained Problems

The proofs of the theorems presented next can be found in Chapter 4 and Section 8.6.1 of [27].

#### Necessary Conditions

**Theorem 1.2** *Suppose that* $f : \mathbb{R}^s \to \mathbb{R}$ *is differentiable at* $\bar{z}$. *If there exists a vector* $d$ *such that* $\nabla f(\bar{z})'d < 0$, *then there exists a* $\delta > 0$ *such that* $f(\bar{z} + \lambda d) < f(\bar{z})$ *for all* $\lambda \in (0, \delta)$.

The vector $d$ in the theorem above is called a *descent direction*. At a given point $\bar{z}$ a descent direction $d$ satisfies the condition $\nabla f(\bar{z})'d < 0$. Theorem 1.2 states that if a descent direction exists at a point $\bar{z}$, then it is possible to move from $\bar{z}$ towards a new point $\tilde{z}$ whose associated cost $f(\tilde{z})$ is lower than $f(\bar{z})$. The direction of *steepest descent* $d_s$ at a given point $\bar{z}$ is defined as the normalized direction where $\nabla f(\bar{z})'d_s < 0$ is minimized. The direction $d_s$ of steepest descent is $d_s = -\frac{\nabla f(\bar{z})}{\|\nabla f(\bar{z})\|}$.

Two corollaries of Theorem 1.2 are stated next.

**Corollary 1.1** *Suppose that* $f : \mathbb{R}^s \to \mathbb{R}$ *is differentiable at* $\bar{z}$. *If* $\bar{z}$ *is a local minimizer, then* $\nabla f(\bar{z}) = 0$.

**Corollary 1.2** *Suppose that* $f : \mathbb{R}^s \to \mathbb{R}$ *is twice differentiable at* $\bar{z}$. *If* $\bar{z}$ *is a local minimizer, then* $\nabla f(\bar{z}) = 0$ *and the Hessian* $\nabla^2 f(\bar{z})$ *is positive semidefinite.*

#### Sufficient Condition

**Theorem 1.3** *Suppose that* $f : \mathbb{R}^s \to \mathbb{R}$ *is twice differentiable at* $\bar{z}$. *If* $\nabla f(\bar{z}) = 0$ *and the Hessian of* $f(z)$ *at* $\bar{z}$ *is positive definite, then* $\bar{z}$ *is a local minimizer.*

#### Necessary and Sufficient Condition

**Theorem 1.4** *Suppose that* $f : \mathbb{R}^s \to \mathbb{R}$ *is differentiable at* $\bar{z}$. *If* $f$ *is convex, then* $\bar{z}$ *is a global minimizer if and only if* $\nabla f(\bar{z}) = 0$.

When the optimization is constrained and the cost function is not sufficiently smooth, the conditions for optimality become more complicated. The intent of this chapter is to give an overview of some important optimality criteria for constrained nonlinear optimization. The optimality conditions derived here will be the main building blocks for the theory developed later in this book.

## 1.4 Lagrange Duality Theory

Consider the nonlinear program (1.12). Let $f^*$ be the optimal value. Denote by $Z$ the domain of cost and constraints (1.5). Any feasible point $\bar{z}$ provides an upper bound to the optimal value $f(\bar{z}) \geq f^*$. Next, we will show how to generate a lower bound on $f^*$.

Starting from the standard nonlinear program (1.12) we construct another problem with different variables and constraints. The original problem (1.12) will be called the primal problem while the new one will be called the *dual* problem. First, we augment the objective function with a weighted sum of the constraints. In this way the *Lagrange dual function* (or Lagrangian) $L$ is obtained

$$\begin{aligned} L(z,u,v) = {} & f(z) + u_1 g_1(z) + \cdots + u_m g_m(z) + \\ & + v_1 h_1(z) + \cdots + v_p h_p(z), \end{aligned} \tag{1.13}$$

where the scalars $u_1,\ldots, u_m, v_1,\ldots, v_p$ are real variables called *dual variables* or *Lagrange multipliers*. We can write Equation (1.13) in the compact form

$$L(z,u,v) = f(z) + u'g(z) + v'h(z), \tag{1.14}$$

where $u = [u_1,\ldots, u_m]'$, $v = [v_1,\ldots, v_p]'$ and $L : \mathbb{R}^s \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$. The components $u_i$ and $v_i$ are called dual variables. Note that the $i$-th dual variable $u_i$ is associated with the $i$-th inequality constraint of problem (1.12), the $i$-th dual variable $v_i$ is associated with the $i$-th equality constraint of problem (1.12).

Let $z$ be a feasible point: for arbitrary vectors $u \geq 0$ and $v$ we trivially obtain a lower bound on $f(z)$

$$L(z,u,v) \leq f(z). \tag{1.15}$$

We minimize both sides of Equation (1.15)

$$\inf_{z\in Z,\ g(z)\leq 0,\ h(z)=0} L(z,u,v) \leq \inf_{z\in Z,\ g(z)\leq 0,\ h(z)=0} f(z) \tag{1.16}$$

in order to reconstruct the original problem on the right-hand side of the expression. Since for arbitrary $u \geq 0$ and $v$

$$\inf_{z\in Z} L(z,u,v) \leq \inf_{z\in Z,\ g(z)\leq 0,\ h(z)=0} L(z,u,v), \tag{1.17}$$

we obtain

$$\inf_{z\in Z} L(z,u,v) \leq \inf_{z\in Z,\ g(z)\leq 0,\ h(z)=0} f(z). \tag{1.18}$$

Equation (1.18) implies that for arbitrary $u \geq 0$ and $v$ the solution to

$$\inf_{z \in Z} L(z, u, v) \tag{1.19}$$

provides us with a lower bound to the original problem. The "best" lower bound is obtained by maximizing problem (1.19) over the dual variables

$$\sup_{(u,v),\ u \geq 0} \inf_{z \in Z} L(z, u, v) \leq \inf_{z \in Z,\ g(z) \leq 0,\ h(z)=0} f(z).$$

Define the dual cost $d(u, v)$ as follows

$$d(u, v) = \inf_{z \in Z} L(z, u, v) \in [-\infty, +\infty]. \tag{1.20}$$

Then the Lagrange dual problem is defined as

$$\sup_{(u,v),\ u \geq 0} d(u, v), \tag{1.21}$$

and its optimal solution, if it exists, is denoted by $(u^*, v^*)$. The dual cost $d(u, v)$ is the optimal value of an *unconstrained optimization problem*. Problem (1.20) is called the *Lagrange dual subproblem*. Only points $(u, v)$ with $d(u, v) > -\infty$ are interesting for the Lagrange dual problem. A point $(u, v)$ will be called *dual feasible* if $u \geq 0$ and $d(u, v) > -\infty$. $d(u, v)$ is *always a concave function* since it is the pointwise infimum of a family of affine functions of $(u, v)$. This implies that the *dual problem is a convex optimization problem (max of a concave function over a convex set) even if the original problem is not convex.* Therefore, it is easier in principle to solve the dual problem than the primal (which is in general nonconvex). However, in general, the solution to the dual problem is only a lower bound of the primal problem:

$$\sup_{(u,v),\ u \geq 0} d(u, v) \leq \inf_{z \in Z,\ g(z) \leq 0,\ h(z)=0} f(z).$$

Such a property is called *weak duality*. In a simpler form, let $f^*$ and $d^*$ be the primal and dual optimal value, respectively,

$$f^* = \inf_{z \in Z,\ g(z) \leq 0,\ h(z)=0} f(z), \tag{1.22a}$$

$$d^* = \sup_{(u,v),\ u \geq 0} d(u, v), \tag{1.22b}$$

then, we always have

$$f^* \geq d^* \tag{1.23}$$

and the difference $f^* - d^*$ is called the *optimal duality gap*. The weak duality inequality (1.23) holds also when $d^*$ and $f^*$ are infinite. For example, if the primal problem is unbounded below, so that $f^* = -\infty$, we must have $d^* = -\infty$, i.e., the Lagrange dual problem is infeasible. Conversely, if the dual problem is unbounded above, so that $d^* = +\infty$, we must have $f^* = +\infty$, i.e., the primal problem is infeasible.

**Remark 1.2** We have shown that the dual problem is a convex optimization problem even if the original problem is nonconvex. As stated in this section, for nonconvex optimization problems it is "easier" to solve the dual problem than the primal problem. However the evaluation of $d(\bar{u}, \bar{v})$ at a point $(\bar{u}, \bar{v})$ requires the solution of the nonconvex unconstrained optimization problem (1.20), which, in general, is "not easy."

### 1.4.1 Strong Duality and Constraint Qualifications

If $d^* = f^*$, then the duality gap is zero and we say that **strong duality** holds:

$$\sup_{(u,v),\ u\geq 0} d(u,v) = \inf_{z\in Z,\ g(z)\leq 0,\ h(z)=0} f(z). \tag{1.24}$$

This means that the best lower bound obtained by solving the dual problem coincides with the optimal cost of the primal problem. In general, strong duality does not hold, even for convex primal problems. Constraint qualifications are conditions on the constraint functions which imply strong duality for convex problems. A detailed discussion on constraint qualifications can be found in Chapter 5 of [27].

A well-known simple constraint qualification is "Slater's condition":

**Definition 1.1 (Slater's condition)** *Consider problem (1.12). There exists $\hat{z} \in \mathbb{R}^s$ which belongs to the relative interior of the problem domain Z, which is feasible $(g(\hat{z}) \leq 0,\ \ h(\hat{z}) = 0)$ and for which $g_j(\hat{z}) < 0$ for all j for which $g_j$ is not an affine function.*

**Remark 1.3** Note that Slater's condition reduces to feasibility when all inequality constraints are linear and $Z = \mathbb{R}^n$.

**Theorem 1.5 (Slater's theorem)** *Consider the primal problem (1.22a) and its dual problem (1.22b). If the primal problem is convex, Slater's condition holds and $f^*$ is bounded then $d^* = f^*$.*

### 1.4.2 Certificate of Optimality

Consider the (primal) optimization problem (1.12) and its dual (1.21). Any feasible point $z$ gives us information about an upper bound on the cost, i.e., $f^* \leq f(z)$. If we can find a dual feasible point $(u, v)$ then we can establish a lower bound on the optimal value of the primal problem: $d(u, v) \leq f^*$. In summary, without knowing the exact value of $f^*$ we can give a bound on how suboptimal a given feasible point is. In fact, if $z$ is primal feasible and $(u, v)$ is dual feasible then $d(u, v) \leq f^* \leq f(z)$. Therefore $z$ is $\varepsilon$-suboptimal, with $\varepsilon$ equal to the primal-dual gap, i.e., $\varepsilon = f(z) - d(u, v)$.

The optimal value of the primal (and dual) problems will lie in the same interval

$$f^* \in [d(u,v), f(z)] \text{ and } d^* \in [d(u,v), f(z)].$$

For this reason $(u, v)$ is also called a *certificate* that proves the (sub)optimality of $z$. Optimization algorithms make extensive use of such criteria. Primal-dual algorithms iteratively solve primal and dual problems and generate a sequence of primal and dual feasible points $z_k$, $(u_k, v_k)$, $k \geq 0$

until a certain $\varepsilon$ is reached. The condition

$$f(z_k) - d(u_k, v_k) < \varepsilon,$$

for terminating the algorithm guarantees that when the algorithm terminates, $z_k$ is $\varepsilon$-suboptimal. If strong duality holds the condition can be met for arbitrarily small tolerances $\varepsilon$.

## 1.5 Complementary Slackness

Consider the (primal) optimization problem (1.12) and its dual (1.21). Assume that strong duality holds. Suppose that $z^*$ and $(u^*, v^*)$ are primal and dual feasible with zero duality gap (hence, they are primal and dual optimal):

$$f(z^*) = d(u^*, v^*).$$

By definition of the dual problem, we have

$$f(z^*) = \inf_{z \in Z} \left( f(z) + u^{*\prime} g(z) + v^{*\prime} h(z) \right).$$

Therefore

$$f(z^*) \leq f(z^*) + u^{*\prime} g(z^*) + v^{*\prime} h(z^*), \tag{1.25}$$

and since $h(z^*) = 0$, $u^* \geq 0$ and $g(z^*) \leq 0$ we have

$$f(z^*) \leq f(z^*) + u^{*\prime} g(z^*) \leq f(z^*). \tag{1.26}$$

From the last equation we can conclude that $u^{*\prime} g(z^*) = \sum_{i=1}^{m} u_i^* g_i(z^*) = 0$ since $u_i^* \geq 0$ and $g_i(z^*) \leq 0$, we have

$$u_i^* g_i(z^*) = 0, \; i = 1, \ldots, m. \tag{1.27}$$

Conditions (1.27) are called *complementary slackness* conditions. Complementary slackness conditions can be interpreted as follows. If the $i$-th inequality constraint of the primal problem is inactive at the optimum ($g_i(z^*) < 0$), then the $i$-th dual optimizer has to be zero $(u_i^* = 0)$. Vice versa, if the $i$-th dual optimizer is different from zero $(u_i^* > 0)$, then the $i$-th constraint is active at the optimum ($g_i(z^*) = 0$).

Relation (1.27) implies that the inequality in (1.25) holds as equality

$$f(z^*) + \sum_i u_i^* g_i(z^*) + \sum_j v_j^* h_j(z^*) = \min_{z \in Z} \left( f(z) + \sum_i u_i^* g_i(z) + \sum_j v_j^* h_j(z) \right). \tag{1.28}$$

Therefore, complementary slackness conditions implies that $z^*$ is a minimizer of $L(z, u^*, v^*)$.

## 1.6 Karush-Kuhn-Tucker Conditions

Consider the (primal) optimization problem (1.12) and its dual (1.21). Assume that strong duality holds. Assume that the cost functions and constraint functions $f$, $g_i$, $h_i$ are differentiable. Let $z^*$ and $(u^*, v^*)$ be primal and dual optimal points, respectively. Complementary slackness conditions implies that $z^*$ minimizes $L(z, u^*, v^*)$ under no constraints (Equation (1.28)). Since $f$, $g_i$, $h_i$ are differentiable, the gradient of $L(z, u^*, v^*)$ must be zero at $z^*$

$$\nabla f(z^*) + \sum_i u_i^* \nabla g_i(z^*) + \sum_j v_j^* \nabla h_j(z^*) = 0.$$

In summary, the primal and dual optimal pair $z^*$, $(u^*, v^*)$ of an optimization problem with differentiable cost and constraints and zero duality gap, have to satisfy the following conditions:

$$\nabla f(z^*) + \sum_{i=1}^{m} u_i^* \nabla g_i(z^*) + \sum_{j=1}^{p} v_j^* \nabla h_i(z^*) = 0, \tag{1.29a}$$

$$u_i^* g_i(z^*) = 0, \quad i = 1, \dots, m \tag{1.29b}$$

$$u_i^* \geq 0, \quad i = 1, \dots, m \tag{1.29bc}$$

$$g_i(z^*) \leq 0, \quad i = 1, \dots, m \tag{1.29d}$$

$$h_j(z^*) = 0, \quad j = 1, \dots, p \tag{1.29e}$$

where Equations (1.29d)–(1.29e) are the primal feasibility conditions, Equation (1.29c) is the dual feasibility condition and Equation (1.29b) are the complementary slackness conditions.

Conditions (1.29a)–(1.29e) are called the *Karush-Kuhn-Tucker* (KKT) conditions. We have shown that the KKT conditions are necessary conditions for any primal-dual optimal pair if strong duality holds and the cost and constraints are differentiable, i.e., any primal and dual optimal points $z^*$, $(u^*, v^*)$ must satisfy the KKT conditions (1.29). If the primal problem is also convex then the KKT conditions are sufficient, i.e., a primal dual pair $z^*$, $(u^*, v^*)$ which satisfies conditions (1.29a)–(1.29e) is a primal dual optimal pair with zero duality gap.

There are several theorems which characterize primal and dual optimal points $z^*$ and $(u^*, v^*)$ by using KKT conditions. They mainly differ on the type of constraint qualification chosen for characterizing strong duality. Next we report just two examples.

If a convex optimization problem with differentiable objective and constraint functions satisfies Slater's condition, then the KKT conditions provide necessary and sufficient conditions for optimality.

**Theorem 1.6** *[27, p. 244] Consider problem (1.12) and let $Z$ be a nonempty set of $\mathbb{R}^s$. Suppose that problem (1.12) is convex and that cost and constraints $f$, $g_i$ and $h_i$ are differentiable at a feasible $z^*$. If problem (1.12) satisfies Slater's condition then $z^*$ is optimal if and only if there are $(u^*, v^*)$ that, together with $z^*$, satisfy the KKT conditions (1.29).*

If a convex optimization problem with differentiable objective and constraint functions has a linearly independent set of active constraints gradients, then the KKT conditions provide necessary and sufficient conditions for optimality.

**Theorem 1.7 (Section 4.3.7 in [27])** *Consider problem (1.12) and let $Z$ be a nonempty open set of $\mathbb{R}^s$. Let $z^*$ be a feasible solution and $A = \{i: g_i(z^*) = 0\}$ be the set of active constraints at $z^*$. Suppose cost and constraints $f$, $g_i$ are differentiable at $z^*$ for all $i$ and that $h_j$ are continuously* *differentiable at $z^*$ for all $j$. Further, suppose that $\nabla g_i(z^*)$ for $i \in A$ and $\nabla h_j(z^*)$ for $j = 1,\dots, p$, are linearly independent. If $z^*$, $(u^*, v^*)$ are primal and dual optimal points, then they satisfy the KKT conditions (1.29). In addition, if problem (1.12) is convex, then $z^*$ is optimal if and only if there are $(u^*, v^*)$ that, together with $z^*$, satisfy the KKT conditions (1.29).*

The KKT conditions play an important role in optimization. In a few special cases, it is possible to solve the KKT conditions (and therefore, the optimization problem) analytically. Many algorithms for convex optimization are conceived as, or can be interpreted as, methods for solving the KKT conditions as Boyd and Vandenberghe observe in [65].

The following example [27] shows a convex problem where the KKT conditions are not fulfilled at the optimum. In particular, both the constraint qualifications of Theorem 1.7 and Slater's condition in Theorem 1.6 are violated.

**Example 1.1** [27, p. 196] Consider the convex optimization problem

$$\begin{array}{ll} \min & z_1 \\ \text{subj. to} & (z_1-1)^2+(z_2-1)^2 \le 1 \\ & (z_1-1)^2+(z_2+1)^2 \le 1. \end{array} \tag{1.30}$$

From the graphical interpretation in Figure 1.1 it is immediate that the feasible set is a single point $\bar{z} = [1, 0]'$. The optimization problem does not satisfy Slater's conditions and moreover $\bar{z}$ does not satisfy the constraint qualifications in Theorem 1.7. At the optimum $\bar{z}$ Equation (1.29a) cannot be satisfied for any pair of nonnegative real numbers $u_1$ and $u_2$.

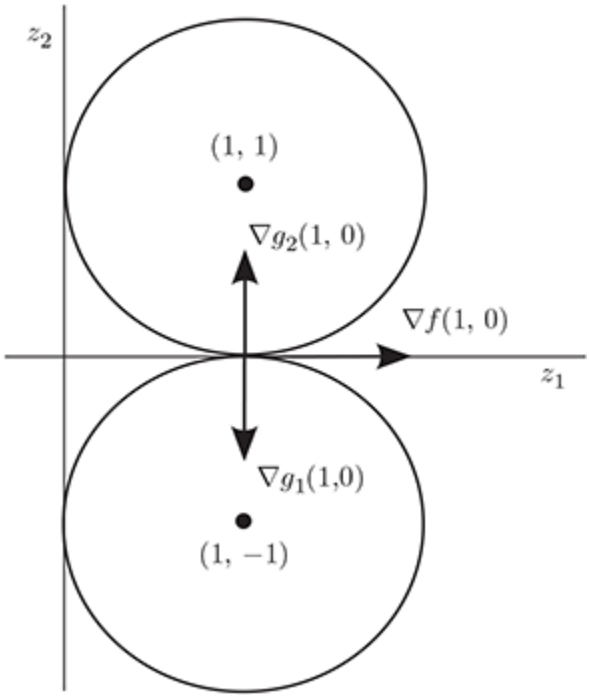


**Figure 1.1** Example 1.1. Constraints, feasible set and gradients. The feasible set is the point (1, 0) that satisfies neither Slater's condition nor the constraint qualification condition in Theorem 1.7.

## 1.6.1 Geometric Interpretation of KKT Conditions

A geometric interpretation of the KKT conditions is depicted in Figure 1.2 for an optimization problem in two dimensions with inequality constraints and no equality constraints.

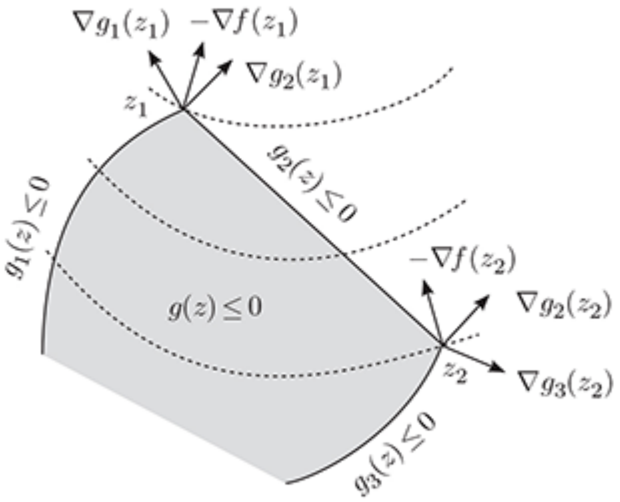


**Figure 1.2** Geometric interpretation of KKT conditions [27].

Equation (1.29a) and equation (1.29c) can be rewritten as

$$-\nabla f(z) = \sum_{i \in A} u_i \nabla g_i(z), \quad u_i \geq 0, \tag{1.31}$$

where $A = \{1, 2\}$ at $z_1$ and $A = \{2, 3\}$ at $z_2$. This means that the negative gradient of the cost at the optimum $-\nabla f(z^*)$ (which represents the direction of steepest descent) has to belong to the cone spanned by the gradients of the active constraints $\nabla g_i$ (since inactive constraints have the corresponding Lagrange multipliers equal to zero). In Figure 1.2, condition (1.31) is not satisfied at $z_2$. In fact, one can move within the set of feasible points $g(z) \leq 0$ and decrease $f$, which implies that $z_2$ is not optimal. At point $z_1$, on the other hand, the cost $f$ can only decrease if some constraint is violated. Any movement in a feasible direction increases the cost. Conditions (1.31) are fulfilled and hence $z_1$ is optimal.