Chapter 1

Introduction

1.1 Optimization problems

A generic optimization problem

$$\text{minimize } f(x) \quad \text{subject to} \quad g(x) = 0 \quad \text{and} \quad h(x) < 0 \tag{1.1}$$

is specified by a triple of functions (f, g, h)

- $f : \mathbb{R}^n \to \mathbb{R}$ is called *cost function* (or objective function in case of maximization problems),
- $g : \mathbb{R}^n \to \mathbb{R}^m$ defines *equality constraints*, and
- $h : \mathbb{R}^n \to \mathbb{R}^p$ defines *inequality constraints*.

The vector $x \in \mathbb{R}^n$ over which the optimization is carried out is called a *decision variable*. Let the domain of f be $\mathscr{D} \subseteq \mathbb{R}^n$. The set

$$\mathscr{F} := \{\, x \in \mathscr{D} \mid g(x) = 0 \text{ and } h(x) < 0 \,\}$$

is called a *feasibility set* of the optimization problem (1.1). A vector of decision variables $x \in \mathbb{R}^n$ is called *feasible* if $x \in \mathscr{F}$. If $\mathscr{F}$ is an empty set, then the problem (1.1) is called *infeasible*. Otherwise, the problem is called *feasible*. A solution of the optimization problem (1.1) must be a feasible vector, so that feasibility is a necessary condition for existence of solution. Note however that feasibility is not a sufficient condition for existence of solution. (Consider, for example, the problem: minimize $1/x$ subject to $x > 0$.)

The vector x^* is called a *globally optimal solution* of the optimization problem (1.1) if

$$x \in \mathscr{F} \quad \implies \quad f(x) \geq f(x^*).$$

The vector x^* is called a *locally optimal solution* of the optimization problem (1.1) if there is a neighborhood $\mathscr{E}$ containing x^*, such that

$$x \in \mathscr{F} \cap \mathscr{E} \quad \implies \quad f(x) \geq f(x^*).$$

("local" means that the candidate solutions are chosen from the neighborhood $\mathscr{E}$ instead of from the whole $\mathbb{R}^n$.) Two optimization problems are called *equivalent* if there is a bijection between the (locally) optimal solutions of the two problems. In some cases the solution of the problem involves a manipulation of the functions (f, g, h), aiming to derive an equivalent problem (f', g', h'), in which f', g', and h' are simpler and make the solution obvious. For example, transforming the given problem to a standard problem the solution of which is known. Typical manipulations are elimination of the constraints and decision variables.

1.2 Numerical methods

An important aspect of optimization theory is to find methods and algorithms that exploit the properties of particular classes of problems in order to find effectively and efficiently locally or globally optimal solutions. In general, the wider the class of problems a method can solve is, the less efficient and less reliable this method is. Therefore, although in theory the most general methods are the most widely applicable, in practice one is interested to find the most particular method for the problem at hand.

In a few simple cases, the problems allow an analytical solution, i.e., necessary and sufficient conditions for existence of a solution are given and if a solution exists all solutions are given by an explicit *analytic expression*. This type of solution requires hand computation or a symbolic computational environment, such as Maple. A prototypical example of a problem with analytical solution is the problem of computing a least-squares approximate solution to an overdetermined system of equations.

Other problems allow "partial" analytic solution in the sense that they can be reduced to some well understood standard problems, e.g., eigenvalue and singular value matrix decompositions. In this case, there are still necessary and sufficient conditions for existence of solution and procedures for (numerically) finding all optimal solutions, if one exists. For all practical purposes the partial analytic solution is as good as an analytic solution, so we will not make a distinction between them. A prototypical example of a problem with analytical solution is computing a low-rank approximation of a given matrix in the induced 2-norm or Frobenius norm.

A key classification for problems that have no analytic solution is in convex vs nonconvex problems. The important property of convex optimization problems is that they have a *unique* solution. Convex optimization theory gives necessary and sufficient conditions for optimality and leads to effective solution methods. In addition, there is readily available software for solving certain classes of convex optimization problems. In contrast, the theory for nonconvex problems is local and the methods are at best guaranteed to find only one locally optimal solution.

1.3 Applications

Optimization methods have applications throughout science and engineering. In this course we will show examples from machine learning (PCA, ICA, and SVM), control (linear quadratic optimal control and minimum time control), approximation (data fitting and system identification), and estimation (Kalman faltering and maximum likelihood estimation).