CHAPTER 1

Introduction

In mathematical optimization we seek to either minimize or maximize a function over a set of alternatives. The function is called the *objective function*, and we allow it to be transfinite in the sense that at each point its value is either a real number or it is one of the to infinite values $\pm\infty$. The set of alternatives is called the *constraint region*. Since every maximization problem can be restated as a minimization problem by simply replacing the objective f_0 by its negative $-f_0$ (and visa versa), we choose to focus only on minimization problems. We denote such problems using the notation

$$\begin{array}{ll} & \underset{x\in X}{\text{minimize}} \;\; f_0(x) \\ (1) & \\ & \text{subject to} \;\; x \in \Omega, \end{array}$$

where $f_0 : X \to \mathbb{R} \cup \{\pm\infty\}$ is the objective function, X is the space over which the optimization occurs, and $\Omega \subset X$ is the constraint region. This is a very general description of an optimization problem and as one might imagine there is a taxonomy of optimization problems depending on the underlying structural features that the problem possesses, e.g., properties of the space X, is it the integers, the real numbers, the complex numbers, matrices, or an infinite dimensional space of functions, properties of the function f_0, is it discrete, continuous, or differentiable, the geometry of the set Ω, how Ω is represented, properties of the underlying applications and how they fit into a broader context, methods of solution or approximate solution, ... For our purposes, we assume that Ω is a subset of $\mathbb{R}^n$ and that $f_0 : \mathbb{R}^n \to \mathbb{R} \cup \{\pm\infty\}$. This severely restricts the kind of optimization problems that we study, however, it is sufficiently broad to include a wide variety of applied problems of great practical importance and interest. For example, this framework includes *linear programming* (LP).

Linear Programming:
In the case of LP, the objective function is linear, that is, there exists $c \in \mathbb{R}^n$ such that

$$f_0(x) = c^T x = \sum_{j=1}^{n} c_j x_j,$$

and the constraint region is representable as the set of solution to a finite system of linear equation and inequalities,

$$(2) \qquad \Omega = \left\{ x \in \mathbb{R}^n \;\middle|\; \sum_{i=1}^{n} a_{ij} x_j \le b_j,\; i = 1, \dots, s,\; \sum_{i=1}^{n} a_{ij} x_j = b_j,\; i = s+1, \dots, m \right\},$$

where $A := [a_{ij}] \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$.

However, in this course we are primarily concerned with nonlinear problems, that is, problems that cannot be encoded using finitely many linear function alone. A natural generalization of the LP framework to the nonlinear setting is to simply replace each of the linear functions with a nonlinear function. This leads to the general *nonlinear programming* (NLP) problem which is the problem of central concern in these notes.

Nonlinear Programming:
In nonlinear programming we are given nonlinear functions $f_i : \mathbb{R}^n \to \mathbb{R},\; i = 1, 2, \dots, m$, where f_0 is the objective function in (1) *and the functions $f_i,\; i = 1, 2, \dots, m$ are called the constraint functions which are used to define the constrain region in* (1) *by setting*

$$(3) \qquad \Omega = \{x \in \mathbb{R}^n \,|\, f_i(x) \le 0,\; i = 1, \dots, s,\; f_i(x) = 0,\; i = s+1, \dots, m\} \;.$$

If $\Omega = \mathbb{R}^n$, then we say that the problem (1) is an *unconstrained optimization problem*; otherwise, it called a constrained problem. We begin or study with unconstrained problems. They are simpler to handle since we are only concerned with minimizing the objective function and we need not concern ourselves with the constraint region.

However, since we allow the objective to take infinite values, we shall see that every explicitly constrained problem can be restated as an ostensibly unconstrained problem.

In the following section, we begin our study of unconstrained optimization which is arguably the most widely studied and used class of unconstrained unconstrained nonlinear optimization problems. This is the class of *linear least squares* problems. The theory an techniques we develop for this class of problems provides a template for how we address and exploit structure in a wide variety of other problem classes.

Linear Least Squares:
A linear least squares problem is one of the form

$$\underset{x\in\mathbb{R}^n}{\text{minimize}}\ \tfrac{1}{2}\,\|Ax-b\|_2^2\,, \tag{4}$$

where

$$A\in\mathbb{R}^{m\times n},\ b\in\mathbb{R}^m, \quad \text{and} \quad \|y\|_2^2 := y_1^2+y_2^2+\cdots+y_m^2\ .$$

Problems of this type arise in a diverse range of application, some of which are discussed in later chapters. Whole books have been written about this problem, and various instances of this problem remain a very active area of research. This problem formulation is usually credited to Legendre and Gauss who made careful studies of the method around 1800. But others had applied the basic approach in a ad hoc manner in the previous 50 years to observational data and, in particular, to studying the motion of the planets.

The second class most important class of unconstrained nonlinear optimization problems is the minimization of *quadratic functions*. As we will see, the linear least squares problem is a member of this class of problems. It is an important for a wide variety of reasons, not the least of which is the relationship to the second-order Taylor approximations for functions mapping $\mathbb{R}^n$ into $\mathbb{R}$.

Quadratic Functions:
A function $f:\mathbb{R}^n\to\mathbb{R}$ is said to be quadratic if there exists $\alpha\in\mathbb{R}$, $g\in\mathbb{R}^n$ and $H\in\mathbb{R}^{n\times n}$ such that

$$f(x)=\alpha+g^Tx+\tfrac{1}{2}x^THx\ .$$

The first thing to notice about such functions is that we may as well assume that the matrix H is symmetric since

$$x^THx=\tfrac{1}{2}(x^THx+x^THx)=\tfrac{1}{2}((x^THx)^T+x^THx)=\tfrac{1}{2}(x^TH^Tx+x^THx)=x^T(\tfrac{1}{2}(H^T+H))x,$$

that is, we may as well replace the matrix H by its symmetric part $\frac{1}{2}(H^T+H)$.

Having quadratic functions in hand, one arrives at an important nonlinear generalization of linear programming where we simply replace the LP linear objective with a quadratic function.

Quadratic Programming:
In quadratic programming we minimize a quadratic objective function subject convex polyhedral constraints of the form (2).

The linear least squares problem and the optimization of quadratic functions are the themes for our initial forays into optimization. The theory and methods we develop for these problems as well as certain variations on these problems form the basis for our extensions to other problem classes. For this reason, we study these problems with great care. Notice that although these problems are nonlinear, their component pieces come from linear algebra, that is matrices and vectors. Obviously, these components play a key role in understanding the structure and behavior of these problems. For this reason, our first task is to review and develop the essential elements from linear algebra that provide the basis for our investigation into these problems.