Introduction

1 Examples of optimization problems

Parametric regression

Suppose we are given experimental measurements of some physical process such as the height y_i at time t_i of a ball thrown from the top of Pisa tower. Elementary physics suggests us that

$$h(t) = x_1 + x_2 t + x_3 t^2,$$

where x_1 is the initial height of the ball, x_2 is its initial velocity, and x_3 should be the free fall acceleration. We will call $h(t; \mathbf{x})$ a *model* with model parameters $\mathbf{x} = (x_1, x_2, x_3)$. In order to fit the model to the data, we can try finding the parameters that best describe the data. To do so, we need to define an *error criterion* such as the sum of squared differences

$$f(\mathbf{x}) = \sum_{i=1}^{m} (h(t_i; \mathbf{x}) - y_i)^2 = \sum_{i=1}^{m} (x_1 + x_2 t_i + x_3 t_i^2 - y_i)^2$$

and look for such $\mathbf{x}$ that brings $f(\mathbf{x})$ to a *minimum.*

Such problems are called *unconstrained continuous minimization* or, more generally, *optimization* problems. The term "continuous" refers to the fact that the *optimization variable* $\mathbf{x}$ is continuous (in our example, $\mathbf{x} \in \mathbb{R}^3$), while the term "unconstrained" means that no constraints are imposed on $\mathbf{x}$. The function $f(\mathbf{x})$ is called the *objective function* (or, simply, the *objective*) of the optimization problem, which is frequently denoted as

$$\min_{\mathbf{x}} f(\mathbf{x}).$$

The minimum value attained by f is called the (global) *minimum*, while the value of the optimization variable at which the minimum is obtained is called the *minimizer* or the *solution* of the problem. The latter is usually denoted as

$$\mathbf{x}^* = \arg\min_{\mathbf{x}} f(\mathbf{x}).$$

Oftentimes, the two notions are confused, and the term "minimum" is (ab)used to denote both.

Going back to our example, we may complicate the physics a bit and introduce, for example, another term to account for the air friction. In general, we can write a model

$$h(t;\mathbf{x}) = \sum_{k=1}^{n} x_k \varphi_k(t),$$

where $\varphi_k(t)$ are some functions. Observe that despite the fact that these functions are typically nonlinear (in t), the model h is *linear* in $\mathbf{x}$ (for this reason, regression problems of this kind are called "linear regression"), and the objective $f(\mathbf{x})$ is *quadratic* in $\mathbf{x}$. The latter fact allows to give a simple closed-form expression for the minimizer (as well as the minimum) of the optimization problem.

Exercise 1. *Show a closed-form solution for the linear parametric regression discussed above. Suggestion: use matrix notation.*

Linear regression problems were among the first practically used optimization problems, mainly by the nineteenth century statisticians, long before the advent of computers and computational methods.

Nonlinear regression

In many practical problems, the dependence of the model on its parameters is *nonlinear*. For example, consider $h(\mathbf{t};\mathbf{A},\mathbf{b}) = \varphi(\mathbf{A}\mathbf{t}+\mathbf{b})$ where φ is a nonlinear function, and the matrix $\mathbf{A}$ and the vector $\mathbf{b}$ are the model parameters. We can still try solving an optimization problem minimizing the quadratic objective function we had before. However, the difference between the optimization problem involving this nonlinear model and its linear counterpart is vast: while in the latter case we had a closed-form expression for the solution, in the former case no known algorithm can find the global minimizer in reasonable time.

It is important to realize that general optimization problems are *unsolvable.* This often gives rise to the dilemma: is it better to use a bad model that we can solve, or a good model that we cannot (or, more precisely, are not guaranteed to solve but still can try to solve). While the ultimate answer to this dilemma probably belongs to the domain of philosophy, my personal opinion favors the second choice. Engineering practice is full of examples of "bad but tractable" models (band-limited signals in Shannon's sampling theory, linear models in classification and regression problems, Gaussian sources in information theory, etc.), with lots of useful applications. Yet, attempts to break the limitations of these models by trying to solve "better but unsolvable" models sometimes lead to breakthroughs (for example, compressed sensing, and artificial neural networks).

In this course, we will see what can be done with general optimization problems and what cannot. We will also encounter a class of solvable non-linear models called *convex optimization problems.*

Optimal resource assignment

As another example of an optimization problem, consider a particular problem of resource assignment. Let there be M power plants, each producing p_m units of power, $m = 1, \dots, M$. Let there be N customers, each demanding q_n units of power, $n = 1, \dots, N$. The geographic locations of the customers and the plants result in certain costs: the cost of transferring a unit of electrical power from plant m to customer n is c_{mn}. Our goal is to solve the optimal assignment problem: find the amount of electricity x_{mn} each plant has to transfer to each customer to minimize the total cost, while answering the customers' demands and not exceeding the plants' capacities. Mathematically, the problem can be formulated as

$$\min_{\mathbf{X} \in \mathbb{R}^{m \times n}} \sum_{m,n} c_{mn} x_{mn} \quad \text{subject to} \quad \begin{cases} \sum_n x_{mn} \leq p_m, \; m = 1, \dots, M \\ \sum_m x_{mn} \leq q_n, \; n = 1, \dots, N. \end{cases}$$

Such optimization problems are called *constrained* because the optimization variable $\mathbf{X}$ is restricted to a sub-set of $\mathbb{R}^{m \times n}$; the particular types of constraints are called *inequality constraints*. Note that both the objective and the constraints are linear in $\mathbf{X}$. Optimization problems of this kind are called *linear programs*.