Fundamentals of Linear Algebra and Optimization

Introduction to Nonlinear Optimization

Jean Gallier and Jocelyn Quaintance

CIS Department
University of Pennsylvania

jean@cis.upenn.edu

November 18, 2020

The Objective of Optimization Theory

The main goal of *optimization theory* is to construct *algorithms* to find solutions (often approximate) of problems of the form

$$\begin{aligned} &\text{find} \quad u \\ &\text{such that} \quad u \in U \text{ and } J(u) = \inf_{v \in U} J(v), \end{aligned}$$

where U is a given subset of a (real) vector space V (possibly infinite dimensional) and $J\colon \Omega \to \mathbb{R}$ is a function defined on some open subset Ω of V such that $U \subseteq \Omega$.

Optimization Theory Notation

The notation $\inf_{v \in U} J(v)$ denotes the *greatest lower bound* of the set of real numbers $\{J(v) \mid v \in U\}$.

Optimization Theory Notation

The notation $\inf_{v \in U} J(v)$ denotes the *greatest lower bound* of the set of real numbers $\{J(v) \mid v \in U\}$.

The element $\inf_{v \in U} J(v)$ is just $\inf\{J(v) \mid v \in U\}$.

Optimization Theory Notation

The notation $\inf_{v \in U} J(v)$ denotes the *greatest lower bound* of the set of real numbers $\{J(v) \mid v \in U\}$.

The element $\inf_{v \in U} J(v)$ is just $\inf\{J(v) \mid v \in U\}$.

The notation J^* is often used to denote $\inf_{v \in U} J(v)$.

Unbounded Below Optimization Problem

If the function J is not bounded below, which means that for every $r \in \mathbb{R}$, there is some $u \in U$ such that $J(u) < r$, then

$$\inf_{v \in U} J(v) = -\infty,$$

and we say that our minimization problem has *no solution*, or that it is unbounded (below).

Unbounded Below Optimization Problem

If the function J is not bounded below, which means that for every $r \in \mathbb{R}$, there is some $u \in U$ such that $J(u) < r$, then

$$\inf_{v \in U} J(v) = -\infty,$$

and we say that our minimization problem has *no solution*, or that it is unbounded (below).

For example, if $V = \Omega = \mathbb{R}$, $U = \{x \in \mathbb{R} \mid x \leq 0\}$, and $J(x) = x$, then the function $J(x)$ is not bounded below and $\inf_{v \in U} J(v) = -\infty$.

Unsolvable Optimization Problem

The issue is that J^* may not belong to $\{J(u) \mid u \in U\}$, that is, it *may not* be achieved by some element $u \in U$, and solving the above problem consists in finding some $u \in U$ that achieves the value J^* in the sense that $J(u) = J^*$.

Unsolvable Optimization Problem

The issue is that J^* may not belong to $\{J(u) \mid u \in U\}$, that is, it *may not* be achieved by some element $u \in U$, and solving the above problem consists in finding some $u \in U$ that achieves the value J^* in the sense that $J(u) = J^*$.

If no such $u \in U$ exists, again we say that our minimization problem has *no solution*.

Restated Objective of Optimization Theory

The minimization problem

$$\begin{aligned} &\text{find} \quad u \\ &\text{such that} \quad u \in U \text{ and } J(u) = \inf_{v \in U} J(v) \end{aligned}$$

is often presented in the following more informal way:

Restated Objective of Optimization Theory

The minimization problem

$$\begin{aligned} &\text{find} \quad u \\ &\text{such that} \quad u \in U \text{ and } J(u) = \inf_{v \in U} J(v) \end{aligned}$$

is often presented in the following more informal way:

$$\begin{aligned} &\text{minimize} \quad J(v) \\ &\text{subject to} \quad v \in U. \end{aligned} \qquad \textbf{(Problem M)}$$

Minimizer of the Optimization Problem

A vector $u \in U$ such that $J(u) = \inf_{v \in U} J(v)$ is often called a *minimizer* of J over U.

Minimizer of the Optimization Problem

A vector $u \in U$ such that $J(u) = \inf_{v \in U} J(v)$ is often called a *minimizer* of J over U.

Some authors denote the set of minimizers of J over U by $\mathrm{argmin}_{v \in U} J(v)$ and write

$$u \in \mathrm{argmin}_{v \in U} J(v)$$

to express that u is such a minimizer.

Maximization Version of Optimization

If we need to maximize rather than minimize a function, then we try to find some $u \in U$ such that

$$J(u) = \sup_{v \in U} J(v).$$

Maximization Version of Optimization

If we need to maximize rather than minimize a function, then we try to find some $u \in U$ such that

$$J(u) = \sup_{v \in U} J(v).$$

Here $\sup_{v \in U} J(v)$ is the least upper bound of the set $\{J(v) \mid v \in U\}$.

Maximization Version of Optimization

If we need to maximize rather than minimize a function, then we try to find some $u \in U$ such that

$$J(u) = \sup_{v \in U} J(v).$$

Here $\sup_{v \in U} J(v)$ is the least upper bound of the set $\{J(v) \mid v \in U\}$.

Some authors denote the set of *maximizers* of J over U by $\mathrm{argmax}_{v \in U} J(v)$.

Constraints and Functional of the Optimization Problem

In most cases, U is defined as the set of solutions of a finite sets of *constraints*, either equality constraints $\varphi_i(v) = 0$, or inequality constraints $\varphi_i(v) \leq 0$, where the $\varphi_i \colon \Omega \to \mathbb{R}$ are some given functions.

Constraints and Functional of the Optimization Problem

In most cases, U is defined as the set of solutions of a finite sets of *constraints*, either equality constraints $\varphi_i(v) = 0$, or inequality constraints $\varphi_i(v) \leq 0$, where the $\varphi_i \colon \Omega \to \mathbb{R}$ are some given functions.

The function J is often called the *functional* of the optimization problem.

Questions Raised by Optimization Theory

The following questions arise naturally:

Questions Raised by Optimization Theory

The following questions arise naturally:

(1) Results concerning the *existence and uniqueness* of a solution for Problem M.

Questions Raised by Optimization Theory

The following questions arise naturally:

(1) Results concerning the *existence and uniqueness* of a solution for Problem M.

(2) The *characterization* of the possible solutions of Problem M. These are conditions for any element $u \in U$ to be a solution of the problem. Such conditions usually involve the derivative dJ_u of J, and possibly the derivatives of the functions φ_i defining U.

Questions Raised by Optimization Theory

The following questions arise naturally:

(1) Results concerning the *existence and uniqueness* of a solution for Problem M.

(2) The *characterization* of the possible solutions of Problem M. These are conditions for any element $u \in U$ to be a solution of the problem. Such conditions usually involve the derivative dJ_u of J, and possibly the derivatives of the functions φ_i defining U. Some of these conditions become sufficient when the functions φ_i are convex.

Questions Raised by Optimization Theory

(3) The effective construction of *algorithms*, typically iterative algorithms that construct a sequence $(u_k)_{k\geq 1}$ of elements of U whose limit is a solution $u \in U$ of our problem. It is then necessary to understand when and how quickly such sequences converge.

Questions Raised by Optimization Theory

(3) The effective construction of *algorithms*, typically iterative algorithms that construct a sequence $(u_k)_{k\geq 1}$ of elements of U whose limit is a solution $u \in U$ of our problem. It is then necessary to understand when and how quickly such sequences converge.

Gradient descent methods fall under this category. As a general rule, unconstrained problems (for which $U = \Omega = V$) are (much) easier to deal with than constrained problems (where $U \neq V$).

Optimization Problems: Equality Constraints

In a previous module we investigated the problem of determining when a function $J\colon \Omega \to \mathbb{R}$ defined on some open subset Ω of a normed vector space E has a local extremum in a subset U of Ω defined by equational constraints, namely

$$U = \{x \in \Omega \mid \varphi_i(x) = 0,\ \ 1 \leq i \leq m\},$$

where the functions $\varphi_i\colon \Omega \to \mathbb{R}$ are continuous (and usually differentiable).

Optimization Problems: Equality Constraints

In a previous module we investigated the problem of determining when a function $J\colon \Omega \to \mathbb{R}$ defined on some open subset Ω of a normed vector space E has a local extremum in a subset U of Ω defined by equational constraints, namely

$$U = \{x \in \Omega \mid \varphi_i(x) = 0,\ \ 1 \leq i \leq m\},$$

where the functions $\varphi_i\colon \Omega \to \mathbb{R}$ are continuous (and usually differentiable).

We gave a *necessary condition* in terms of the *Lagrange multipliers*.

Optimization Problems: Inequality Constraints

Our goal is to find a necessary criterion for a function $J \colon \Omega \to \mathbb{R}$ to have a minimum on a subset U defined by *inequality* constraints $\varphi_i(x) \leq 0$, where the functions φ_i are convex.

Optimization Problems: Inequality Constraints

Our goal is to find a necessary criterion for a function $J \colon \Omega \to \mathbb{R}$ to have a minimum on a subset U defined by *inequality* constraints $\varphi_i(x) \leq 0$, where the functions φ_i are convex.

There is a *necessary condition* for a function J to have a minimum on a subset U defined by qualified inequality constraints in terms of the *Karush–Kuhn–Tucker conditions* (for short KKT conditions), which involve *nonnegative Lagrange multipliers*.

Optimization Problems: Inequality Constraints

Our goal is to find a necessary criterion for a function $J\colon \Omega \to \mathbb{R}$ to have a minimum on a subset U defined by *inequality* constraints $\varphi_i(x) \leq 0$, where the functions φ_i are convex.

There is a *necessary condition* for a function J to have a minimum on a subset U defined by qualified inequality constraints in terms of the *Karush–Kuhn–Tucker conditions* (for short KKT conditions), which involve *nonnegative Lagrange multipliers*.

The proof relies on a version of the Farkas–Minkowski lemma.

Optimization Problems: Inequality Constraints

In general, the KKT conditions are useless unless the constraints are convex. In this case, there is a manageable notion of qualified constraint given by *Slater's conditions*.

Optimization Problems: Inequality Constraints

In general, the KKT conditions are useless unless the constraints are convex. In this case, there is a manageable notion of qualified constraint given by *Slater's conditions*.

Furthermore, if J is also convex and if the KKT conditions hold, then J has a global minimum.

Equality Constraints as Inequalities

From now on we assume that U is defined by a set of inequalities, that is

$$U = \{x \in \Omega \mid \varphi_i(x) \leq 0,\ \ 1 \leq i \leq m\},$$

where the functions $\varphi_i\colon \Omega \to \mathbb{R}$ are continuous (and usually differentiable).

Equality Constraints as Inequalities

From now on we assume that U is defined by a set of inequalities, that is

$$U = \{x \in \Omega \mid \varphi_i(x) \leq 0,\ 1 \leq i \leq m\},$$

where the functions $\varphi_i \colon \Omega \to \mathbb{R}$ are continuous (and usually differentiable). An equality constraint $\varphi_i(x) = 0$ is treated as the conjunction of the two inequalities $\varphi_i(x) \leq 0$ and $-\varphi_i(x) \leq 0$.

Equality Constraints as Inequalities

From now on we assume that U is defined by a set of inequalities, that is

$$U = \{x \in \Omega \mid \varphi_i(x) \leq 0, \ \ 1 \leq i \leq m\},$$

where the functions $\varphi_i \colon \Omega \to \mathbb{R}$ are continuous (and usually differentiable). An equality constraint $\varphi_i(x) = 0$ is treated as the conjunction of the two inequalities $\varphi_i(x) \leq 0$ and $-\varphi_i(x) \leq 0$.

Later on we will see that when the functions φ_i are convex, since $-\varphi_i$ is not necessarily convex, it is desirable to treat equality constraints separately, but for the time being we won't.

Role of Convexity in Optimization

Since the astute reader will notice the word *convex* has appeared numerous times throughout this lesson, we need to first define the notion of a convex function.