CISC 371 Class 1

Introduction to Optimization

Texts: [1] pp.1–10

Main Concepts:
- *Optimum: maximum or minimum*
- *Unconstrained optimization*
- *constrained optimization*

Sample Problem, Spatial Localization: Which point is at the "middle" of a set of given points?

In this course, optimization is the process of selecting a "best" member of a set according to a criterion. We will explore sets that may be vector spaces, vector subspaces, or subsets of vector subspaces. For brevity, we will refer to any member of a set $\mathbb{V}$ as a *point*, which we will write as t or, if $\mathbb{V}$ is a vector space, as $\vec{w}$. In an axiomatic Euclidean geometry, a point is the simplest possible geometrical object and, for this course, a vector satisfies this definition.

The criterion that we will use is a function, which in this course is a mapping from a set of points to the real numbers. The proper usage would be to use a "functional", which is a map from a vector space to the scalar of the vector space; because current usage in optimization is the word "function", we use this term here despite its ambiguity.

A simple example of optimization is a function that maps a real number to a real number, which can be written as

$$f : \mathbb{R} \to \mathbb{R}$$

A simpler, more familiar notation for one such function might be

$$f_1(t) \stackrel{\text{def}}{=} \frac{-t}{t^2+1} \tag{1.1}$$

In this course, we will often refer to such a function as an *objective function*. The term "objective" will be used as an abbreviation for the longer term. A point, from the domain of f, is the *argument* of the function.

An example of the objective function of Equation 1.1 is plotted in Figure 1.1. We see that there are two "best" values. From visual examination of the plot, we can see that the maximum value of f_1 is $+0.5$ and the minimum value of f_1 is -0.5; these occur at the point $t = +1$ and at the point $t = -1$ respectively. In this course we will choose the "best" value of an objective to be its minimum value.

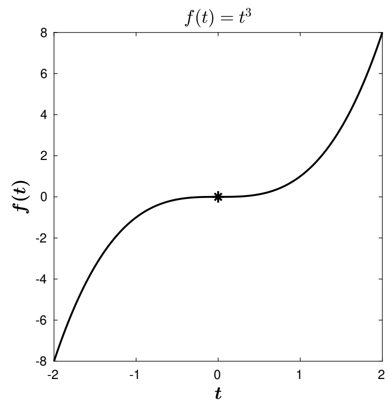


Figure 1.1: An objective function that has a maximum at $t = -1$ and a minimum at $t = +1$.

It is important to distinguish the minimum value of an objective f from the one or more points in the domain of f that map to the minimum value of f. We can summarize these distinctions as:

$$\text{Real Numbers} \quad : \quad \mathbb{R} \tag{1.2}$$

$$\text{Function} \quad : \quad f : \mathbb{R} \to \mathbb{R} \tag{1.3}$$

$$\text{Minimum} \quad : \quad \min_{t \in \mathbb{V}} f(t) \tag{1.4}$$

$$\text{Minimizer} \quad : \quad \operatorname*{argmin}_{t \in \mathbb{V}} f(t) \tag{1.5}$$

We will carefully distinguish the minimum of an objective function, given in Definition 1.4, from the minimizer, given in Definition 1.5. The minimum is always a real number and the minimizer is a *point*. When the set $\mathbb{V}$ contains vectors, this distinction will be easy to maintain; when $\mathbb{V}$ contains real numbers, the distinction may be less clear.

Mostly, nonlinear optimization is applied to vectors in vector spaces. The goal, as for the scalar case, is to minimize an objective function. The difference is that the function is defined for a domain of vectors of a known size.

The domain of the objective function is most often a vector space. If there are n entries of a vector in the vector space, then we will write the vector space as $\mathbb{R}^n$. The objective function maps a vector to a scalar, so $f : \mathbb{R}^n \to \mathbb{R}$. We will usually write the objective function as $f(\vec{w})$ with $\vec{w} \in \mathbb{R}^n$.

1.1 Example: Fermat's Problem

A famous example of an objective function of size-2 vectors is *Fermat's Problem*, which can be posed simply. Suppose that there are a number n of vectors in the plane that are called *anchor* points. The problem is to find the "most central" point in the plane – we can write this as the optimal weight vector $\vec{w}^*$ – that has the minimum sum of distances to the anchor points. The distance from any vector $\vec{w}$ to the first anchor $\vec{a}_1$ is

$$\|\vec{w} - \vec{a}_1\|$$

Using a similar distance for each anchor point $\vec{a}_j$, the objective is the sum of these distances:

$$f(\vec{w}) \stackrel{\text{def}}{=} \sum_{j=1}^{n} \|\vec{w} - \vec{a}_j\| \tag{1.6}$$

The minimizer is the vector $\vec{w}^*$ that is the *geometric median*:

$$\vec{w}^* = \underset{\vec{w} \in \mathbb{R}^2}{\text{argmin}} \sum_{j=1}^{n} \|\vec{w} - \vec{a}_j\| \tag{1.7}$$

For $n = 3$, in which the anchor points that are the vertices of the triangle, Evangelista Torricelli published a solution in 1659. There were two cases: when no interior angle equals or exceeds $120°$, and when one interior angle exceeds this value. Solutions of example anchor points are shown in Figure 1.2.

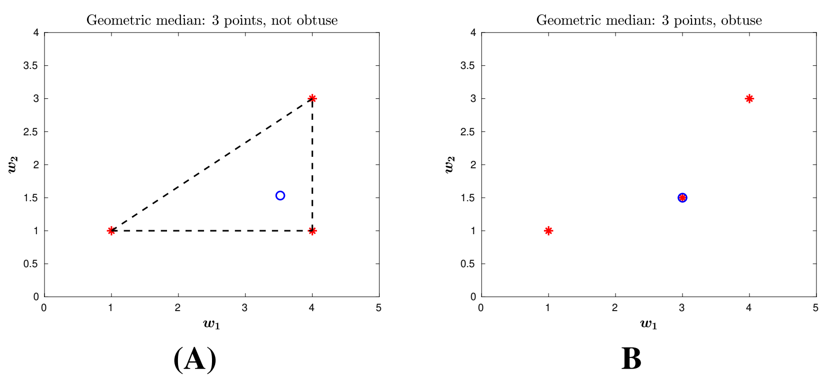
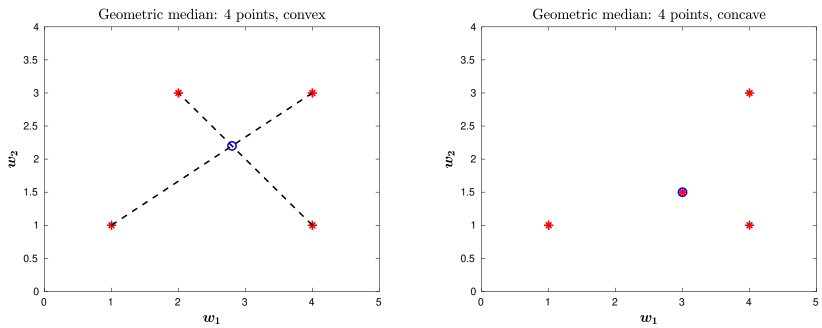


Figure 1.2: Geometric median of 3 anchor points. The anchors are shown as red asterisks and the numerically computed minimizer is shown as a blue circle. (A) No interior angle exceeds $120°$; the minimizer is strictly inside the triangle. (B) One interior angle equals or exceeds $120°$; the minimizer is the corresponding anchor point.

Subsequently, the planar problem was solved geometrically for $n = 4$ points. This, too, has two cases: when one anchor point is inside the triangle described by the other 3 anchor points, or when no anchor point is an interior point. In the first case, the interior point is the minimizer, which is the geometric median. In the second case, the intersection of the lines that join opposing anchor points is the minimizer; this intersection is sometimes named the Radon point. Solutions of example anchor points are shown in Figure 1.3.



(A) **B**

Figure 1.3: Geometric median of 4 anchor points. The anchors are shown as red asterisks and the numerically computed minimizer is shown as a blue circle. (A) No anchor point is an interior point; the minimizer is the intersection of the lines that join opposing anchor points. (B) one anchor point is inside the triangle described by the other 3 anchor points; the minimizer is this anchor point.

Remarkably, for $n \geq 5$ planar anchor points, there is no general solution from either geometry or algebra. The only way to estimate a minimizer point is to numerically approximate the minimum of the objective in Equation 1.7. For $n = 5$ example anchor points, a numerical solution is illustrated in Figure 1.4.

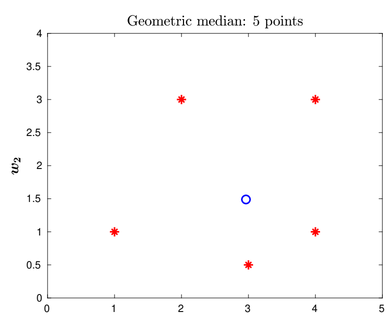


Figure 1.4: Geometric median of 5 anchor points, computed by numerical approximation of Fermat's objective function. The anchors are shown as red asterisks and the numerically computed minimizer is shown as a blue circle.

In this course, we will examine 3 kinds of optimization problems:

1. Unconstrained minimization of an objective with a scalar argument
2. Unconstrained minimization of an objective with a vector argument
3. Constrained minimization of an objective with a vector argument

The first kind of problem is relatively easy to solve. The solutions that we will examine will depend on how much information is provided about a potential minimizer.

The second kind of problem is much more difficult. We will explore this kind of problem by relying on a key observation: this can be done by searching in a vector space for the *direction* from the current approximation to the next approximation, and the *distance* by which to "carry" the current approximation. This framework requires much less differential calculus than the historically earlier framework of posing an optimization purely as a calculus problem.

The third kind of problem – when it can be solved at all – is often reduced to the second kind of problem by introducing additional variables. We will explore constrained optimization in the context of machine learning, for which linear algebra can be a useful tool.

1.2 Our Approach: Mainly Linear Algebra

Most work that was reported in the 20^{th} century, and the early part of the 21^{st} century, used a great deal of calculus and some statistics. These mathematical tools are important for deriving and proving results, particularly the convergence of optimization algorithms.

In this undergraduate course, we will concentrate on the implementation and use of optimization. When optimization is reduced to code, we will discover that this is partly nonlinear function computations, and is mainly linear algebra. There are reasons for believing that this should be so – some results in differential geometry suggest that computations can be predominantly computed in the tangent space of a manifold – and in practice this is what we will implement.

Because the implementations rely heavily on linear algebra, we will persistently look for algebraic explanations and simplifications. This approach will not be especially apparent when working with a function of a scalar argument. When we turn to functions with vector arguments, both unconstrained and constrained, we will see an increasing reliance on linear algebra to structure our solutions.

1.3 Definitions: Minimum, Minimizer, Interior Point, Open Set

In this course, we will need to distinguish between the minimum of a function and the associated minimizer. A minimizer is a value of the argument such that the function is at a minimum. These definitions should be familiar from prerequisite material and are presented for reference.

For these definitions, we will use a point $\vec{w}$ that is a member of a vector space $\mathbb{R}^n$. The scalar case, which is $t \in \mathbb{R}$, will be interpreted as a 1-D vector case. The set $\mathbb{R}_{++}$ is the set of positive real numbers, that is, the set of all real numbers that are greater than zero.

A *strict global minimizer* is an argument of a function f such that the function takes on the lowest possible value.

<u>Definition:</u> strict global minimizer of $f(\vec{w})$

> For any $\vec{w} \in \mathbb{R}^n$, and any $f : \mathbb{R}^n \to \mathbb{R}$, a vector $\vec{t}^* \in \mathbb{R}^n$ is a *strict global minimizer* of f is defined as
> $$(\vec{w} \neq \vec{t}^*) \to (f(\vec{t}^*) < f(\vec{w})) \tag{1.8}$$

A *strict local minimizer* is an argument of a function f such that, for all arguments in an open set around the strict local minimizer, the function takes on the lowest possible value.

<u>Definition:</u> strict local minimizer of $f(\vec{w})$

> For any $\vec{w} \in \mathbb{R}^n$, and any $f : \mathbb{R}^n \to \mathbb{R}$, a vector $\vec{t}^* \in \mathbb{R}^n$ is a *strict local minimizer* of f is defined as
> $$\exists r \in \mathbb{R}_{++} \big((\|\vec{w} - \vec{t}^*\| < r) \wedge (\vec{w} \neq \vec{t}^*)\big) \to (f(\vec{t}^*) < f(\vec{w})) \tag{1.9}$$

We will sometimes need to relax the requirements of strict minimization. The definitions of general minimization are closely related to the definitions of strict minimization.

<u>Definition:</u> global minimizer of $f(\vec{w})$

> For any $\vec{w} \in \mathbb{R}^n$, and any $f : \mathbb{R}^n \to \mathbb{R}$, a vector $\vec{t}^* \in \mathbb{R}^n$ is a *global minimizer* of f is defined as
> $$f(\vec{t}^*) \leq f(\vec{w}) \tag{1.10}$$

Definition: local minimizer of $f(\vec{w})$

For any $\vec{w} \in \mathbb{R}^n$, and any $f : \mathbb{R}^n \to \mathbb{R}$, a vector $\vec{t}^* \in \mathbb{R}^n$ is a *local minimizer* of f is defined as

$$\exists r \in \mathbb{R}_{++}(\|\vec{w} - \vec{t}^*\| < r) \to (f(\vec{t}^*) \leq f(\vec{w})) \tag{1.11}$$

Observation: In the definitions of minimizers, the function f can be discontinuous, e.g., can be defined over the domain of integers.

We will occasionally need to use a few basic concepts from point-set topology. The main concepts are interior points, boundary points, and open sets. Because our points are always members of vector spaces, our definitions are relatively straightforward.

An *interior point* of a set $\mathbb{V} \subseteq \mathbb{R}^n$ is a point that has a "ball" of points around it that are all in $\mathbb{V}$.
Definition: interior point of $\mathbb{V}$

For any $\mathbb{V} \subseteq \mathbb{R}^n$ and any $\vec{u} \in \mathbb{V}$, the vector $\vec{u}$ is an *interior point* of $\mathbb{V}$ is defined as

$$\exists r \in \mathbb{R}_{++}, \forall \vec{w} \in \mathbb{R}^n \left((\|\vec{w} - \vec{u}\| < r) \to (\vec{w} \in \mathbb{V}) \right) \tag{1.12}$$

A *boundary point* is a point in a set that is not an interior point.
Definition: boundary point of $\mathbb{V}$

For any $\mathbb{V} \subseteq \mathbb{R}^n$ and any $\vec{u} \in \mathbb{V}$, the vector $\vec{u}$ is a *boundary point* of $\mathbb{V}$ is defined as

$$\neg \left(\exists r \in \mathbb{R}_{++}, \forall \vec{w} \in \mathbb{R}^n \left((\|\vec{w} - \vec{u}\| < r) \to (\vec{w} \in \mathbb{V}) \right) \right) \tag{1.13}$$

A set $\mathbb{V} \subseteq \mathbb{R}^n$ is an *open set* if $\mathbb{V}$ contains only interior points.
Definition: open set

A set $\mathbb{V} \subseteq \mathbb{R}^n$ is an *open set* of $\mathbb{V}$ is defined as

$$\forall \vec{u} \in \mathbb{V}, \exists r \in \mathbb{R}_{++}, \forall \vec{w} \in \mathbb{R}^n \left((\|\vec{w} - \vec{u}\| < r) \to (\vec{w} \in \mathbb{V}) \right) \tag{1.14}$$

References

[1] Beck A: Introduction to Nonlinear Optimization: Theory, Algorithms, and Applications with MATLAB. Siam Press, 2014