2 Continuous Optimization

Although this course centers on the discrete optimization problems, it is useful first to briefly talk about continuous optimization. Continuous optimization has become a fundamental tool in many different applications. As we will discuss later, continuous optimization methods can also be used for the discrete optimization problems by employing the so-called **relaxation** technique. Hence, it is necessary for us to have a basic understanding of the continuous problems. There is a large class of continuous optimization problems, called **convex** optimization, which can be exactly solved with a reasonable complexity. A very specific type of convex optimizations is **linear** optimization. We will spend a considerable amount of time to understand linear optimization problems, as it will be our basic tool to deal with the discrete optimization problems. I want to remind here that there are a number of other ideas to treat a discrete optimization problem, but we will not have time to cover them in this course.

Since the variable space in a continuous optimization problem is infinite, exhaustive search is impossible. Instead, a number of other algorithms, such as local search are popular in the continuous case. We will briefly talk about them later, but what you need to readily know is that there are well-optimized software packages that implement some state-of-the-art optimization techniques. Some of them are free to use and so you only need to download and learn how to use them. In this course, we will use the CVX optimization toolbox, which is a free MATLAB-based software. I assume that you have a basic knowledge about MATLAB. In the class, we also go through a CVX tutorial, which can be found here:

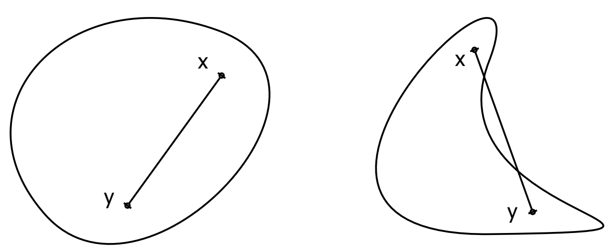
`http://cvxr.com/cvx/doc/quickstart.html`

A more comprehensive introduction to CVX can be found here:

`http://cvxr.com/cvx/doc/CVX.pdf`

2.1 Convex Optimization

To understand a convex optimization, we first need to know what convex functions and convex sets are. A set $\Omega \subseteq \mathbb{R}^n$ is called convex if taking any two points $\mathbf{x}_1, \mathbf{x}_2$, both inside Ω, the line segment connecting $\mathbf{x}_1$ and $\mathbf{x}_2$ is completely inside Ω. A set, which is not convex is called **nonconvex**. Figure 3 depicts a convex and a non convex set in the Euclidean plane $\mathbb{R}^2$.



(a) (b)

Figure 3: a) A convex set. b) A nonconvex set.

Instead of the graphical representation, there is a more convenient way to write the definition of a convex set. Notice that for any two points $\mathbf{x}_1, \mathbf{x}_2$ the line segment connecting $\mathbf{x}_1, \mathbf{x}_2$ is the locus of all points $\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2$, where $\lambda \in [0\ 1]$. So, we can say that a set $\Omega \subseteq \mathbb{R}^n$ is convex if:

$$\forall \mathbf{x}_1, \mathbf{x}_2 \in \Omega, \lambda \in [0\ 1];\ \lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2 \in \Omega \tag{11}$$

One of the main properties of the convex sets is that they are closed under intersection i.e., if I take two convex sets Ω_1 and Ω_2, then $\Omega_1 \cap \Omega_2$ is also convex. Now, let us take the following example:

Example 4. An n-dimensional **half space** S is a subset of $\mathbb{R}^n$, which can be written as

$$S = \{(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n \mid a_1 x_1 + a_2 x_2 + \ldots + a_n x_n + b \leq 0\} \tag{12}$$

where $(a_1, a_2, \ldots, a_n) \in \mathbb{R}^n$ and $b \in \mathbb{R}$ are arbitrary real coefficients. (Try to plot a typical half space in two dimensions $n = 2$). Now, we show that half-spaces are convex. Take two points $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and $\mathbf{y} = (y_1, y_2, \ldots, y_n)$, both in S. Also, take a coefficient $\lambda \in [0\ 1]$. Then, we have that

$$\lambda\mathbf{x} + (1-\lambda)\mathbf{y} = (\lambda x_1 + (1-\lambda)y_1, \lambda x_2 + (1-\lambda)y_2, \ldots, \lambda x_n + (1-\lambda)y_n) \tag{13}$$

On the other hand, $\mathbf{x}, \mathbf{y} \in S$ implies that

$$\begin{aligned} a_1 x_1 + a_2 x_2 + \ldots + a_n x_n + b \leq 0 \\ a_1 y_1 + a_2 y_2 + \ldots + a_n y_n + b \leq 0 \end{aligned} \tag{14}$$

Multiply the first inequality by λ and the second one by $(1-\lambda)$ and add the two. Notice that this is a valid operation, since λ and $1-\lambda$ are both positive. We get

$$a_1(\lambda x_1 + (1-\lambda)y_1) + a_2(\lambda x_2 + (1-\lambda)y_2) + \ldots + a_n(\lambda x_n + (1-\lambda)y_n) + b \leq 0 \tag{15}$$

From (13), we get that $\lambda\mathbf{x} + (1-\lambda)\mathbf{y} \in S$.

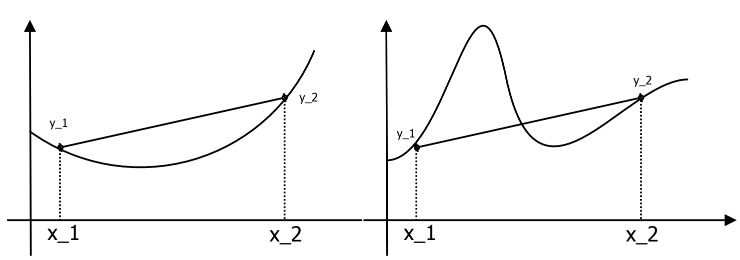
Half spaces are the building blocks of convex sets. It is possible to show that any convex set is an intersection of a (possibly infinite) number of half spaces. An intersection of a finite number of half spaces is called a **convex polyhedral set**.

Now, about convex functions: Take a real function $f : \Omega \to \mathbb{R}$. Notice that this function can be shown by a curve $y = f(\mathbf{x})$ (more generally a surface in high dimensions), like the one in Figure 1. This function is convex if it satisfies the following two conditions:

1. Its domain set Ω is convex.

2. For any two points $(\mathbf{x}_1, y_1 = f(\mathbf{x}_1))$ and $(\mathbf{x}_2, y_2 = f(\mathbf{x}_2))$ on the curve of $f(\mathbf{x})$, the line segment connecting the two points is completely above the curve. Figure 4 shows a convex and a non-convex function. Similar to convex sets, this condition for convex functions can be written as

$$\forall \mathbf{x}_1, \mathbf{x}_2 \in \Omega, \lambda \in [0\ 1];\ f(\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2) \leq \lambda f(\mathbf{x}_1) + (1-\lambda) f(\mathbf{x}_2) \tag{16}$$



(a) (b)

Figure 4: a) A convex function. b) A nonconvex function.

There are two different ways to verify that a function is convex. The first one is simply to use the definition in (16). The second one can only be applied to the twice-differentiable functions. Here, I only talk about the case where the function is univariate $f(x) : \mathbb{R} \to \mathbb{R}$, although the second method can also be used for multivariate functions: A twice-differentiable function $f : \mathbb{R} \to \mathbb{R}$ is convex, if and only if its second derivative $f^{(2)}(x)$ is nonnegative (positive or zero) everywhere. You see that the second method is simpler to check, but I should warn you that this method only applies to second-differentiable functions. There are many interesting functions, which are non-differentiable at least at one point and so the second method is not going to work for them.

Example 5. We want to show that the function $f(x) = x^2$ is convex on $\mathbb{R}$. Notice that this function is twice differentiable and its second derivative is $f^{(2)}(x) = 2 > 0$. Thus, the function is convex. Also, we can use the first method by employing the definition of convex functions. Take two real numbers x_1, x_2 and $\lambda \in [0\ 1]$. Notice that

$$0 \leq \lambda(1-\lambda)(x_1 - x_2)^2 \tag{17}$$

We can write (17) as

$$0 \leq \lambda(1-\lambda)x_1^2 + \lambda(1-\lambda)x_2^2 - 2\lambda(1-\lambda)x_1x_2 \tag{18}$$

and further modified to

$$\lambda^2 x_1^2 + (1-\lambda)^2 x_2^2 + 2\lambda(1-\lambda)x_1x_2 \leq \lambda x_1^2 + (1-\lambda)x_2^2 \tag{19}$$

which leads to

$$\lambda^2 x_1^2 + (1-\lambda)^2 x_2^2 + 2\lambda(1-\lambda)x_1x_2 \leq \lambda x_1^2 + (1-\lambda)x_2^2 \tag{20}$$

and

$$(\lambda x_1 + (1-\lambda)x_2)^2 \leq \lambda x_1^2 + (1-\lambda)x_2^2 \tag{21}$$

This is exactly the definition of a convex function in 16.

Example 6. Another example of a convex function is an **affine** function. An affine function is any function $f(\mathbf{x}) = f(x_1, x_2, \ldots, x_n) : \mathbb{R}^n \to \mathbb{R}$ of the following form

$$f(x_1, x_2, \ldots, x_n) = a_1x_1 + a_2x_2 + \ldots + a_nx_n + b \tag{22}$$

where $(a_1, a_2, \ldots, a_n) \in \mathbb{R}^n$ and $b \in \mathbb{R}$ are arbitrary real coefficients. It is left as an exercise (not mandatory!) to show that an affine function is convex. Similar to half planes, affine functions are building blocks of all convex functions. It is more difficult to explain this topic and I skip it in favor of simplicity.

Now, I can explain what a convex optimization is. A minimization problem is called convex if

1. The feasible region is a convex set.
2. The cost function is a convex function.

In particular, an optimization in the form of (2) is convex if:

1. The variable space Ψ is $\mathbb{R}^n$ for some n and the function f is convex on $\mathbb{R}^n$.
2. The functions $g_i(\mathbf{x})$ are convex on $\mathbb{R}^n$.
3. The functions $h_j(\mathbf{x})$ are affine.

Let us take an example:

Example 7. Consider the optimization

$$\begin{aligned}
\min_{(x_1,x_2,x_3,x_4)\in\mathbb{R}^4} \; & x_1 + x_2 + x_3 + x_4 \\
\text{subject to} & \\
x_1^2 + x_2^2 &\le 1 \\
x_4 &\ge 0 \\
2x_1 + x_2 - x_3 + 3x_4 &= 1 \\
x_1 + x_2 + 2x_3 - x_4 &= 3
\end{aligned} \tag{23}$$

This is a convex optimization as the variable space is $\mathbb{R}^4$, the cost function $f(\mathbf{x}) = x_1 + x_2 + x_3 + x_4$ is convex (affine in fact). The inequality constrains are associated with convex functions $g_1(x_1, x_2, x_3, x_4) = x_1^2 + x_2^2 - 1$ and $g_2(x_1, x_2, x_3, x_4) = -x_4$, and the equality constraints are affine; $h_1 = 2x_1 + x_2 - x_3 + 3x_4 - 1$ and $h_2 = x_1 + x_2 + 2x_3 - x_4 = 3$.

Note that we can eliminate the inequality constraints by writing them as

$$\begin{aligned}
2x_1 + x_2 &= x_3 - 3x_4 + 1 \\
x_1 + x_2 &= -2x_3 + x_4 + 3
\end{aligned} \tag{24}$$

which can be solved for x_1 and x_2 to obtain

$$\begin{aligned}
x_1 &= 3x_3 - 4x_4 - 2 \\
x_2 &= -5x_3 + 5x_4 + 5
\end{aligned} \tag{25}$$

We can now replace these relations in the optimization in (23) to eliminate x_1, x_2 as well as the equality constraints

$$\begin{aligned}
\min_{(x_3,x_4)\in\mathbb{R}^2} \; & -x_3 + 2x_4 + 3 \\
\text{subject to} & \\
(3x_3 - 4x_4 - 2)^2 + (-5x_3 + 5x_4 + 5)^2 &\le 1 \\
x_4 &\ge 0
\end{aligned} \tag{26}$$

After solving (26), we can calculate x_1 and x_2 from (25).

As seen, the affine equality constraints can always be eliminated by calculating a number of variables in terms of the others. Hence, we may not consider equality constraints in a convex optimization.