Lecture 5

Convex Optimization (I)

5.1 Lecture Objectives

- Understand what are convex functions.
- Recognize the implications of convex optimization problems in terms of global optimization.
- Recognize the implications of non-convex optimization in terms of the (general) inability to guarantee global optimization.
- Recognize what operations preserve convexity.

5.2 Local vs Global Convergence

The main reason to classify our optimization problems as convex vs non-convex is the fact that most algorithms used to solve such optimization problems are iterative descent algorithms that start out from some initial guess $\mathbf{x_0}$ and descend on the cost function $f(\mathbf{x})$ until a locally optimal solution is found.

In this context, the "topography" of $f(\mathbf{x})$ becomes very important: if it contains multiple mountains and valleys, it will generally be very challenging to ensure that our iterative algorithms converge to globally optimal solutions. The following section shows examples of functions that have different topography, intuitively leading to different challenges for iterative algorithms.

A central idea to ensure that our optimization problems have a favorable topography is the concept of convexity. We will study this concept over the next two lectures.

5.3 Convex Functions

5.3.1 Convexity

A function $f(\mathbf{x}) : \mathbb{R}^N \to \mathbb{R}$ is defined as convex if, for any two locations $\mathbf{x_1}$ and $\mathbf{x_2}$, and a scalar $t \in [0, 1]$:

$$f(t\mathbf{x_1} + (1 - t)\mathbf{x_2}) \leq tf(\mathbf{x_1}) + (1 - t)f(\mathbf{x_2}) \tag{5.1}$$

or in other words, the function values lie below the straight line that joins any two points as depicted in figure 5.1. The implications of convexity for local vs global optimization are demonstrated pictorially in figure 5.2.

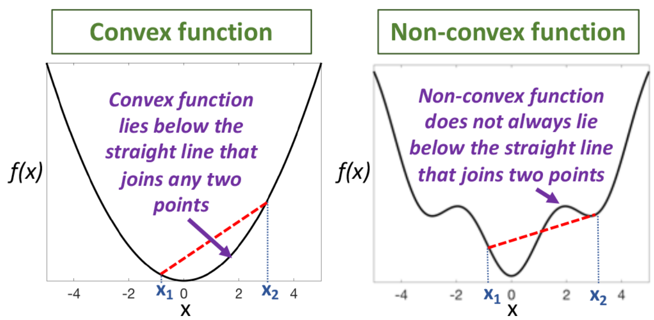

Figure 5.1: Pictorial depiction of the concept of convexity.

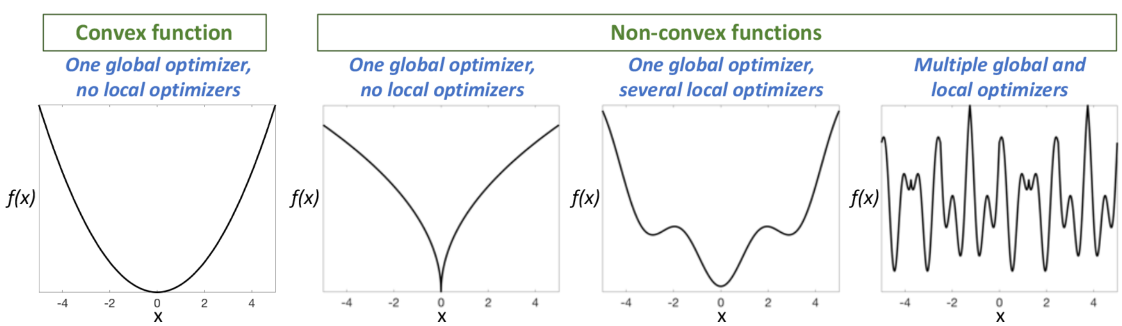

Figure 5.2: Several examples of convex and non-convex functions in 1D, with varying numbers of global and local optimizers. Note that a convex function cannot have local optimizers that are not also globally optimal, and it cannot have multiple separate 'valleys'. Non-convex functions may have all sorts of local and global optimizers, and multiple valleys that complicate the solution of optimization problems using iterative descent-based algorithms.

5.3.2 Strict convexity

Convex functions for which the "less than or equal to" ($\leq$) symbol in Equation 5.2 is actually a "less than" ($<$) are strictly convex:

$$f(t\mathbf{x_1} + (1-t)\mathbf{x_2}) < tf(\mathbf{x_1}) + (1-t)f(\mathbf{x_2}) \tag{5.2}$$

for $t \in (0, 1)$ (note that the interval for t now does not include 0 or 1, since equality is trivially guaranteed at these extreme points). From the definition, strictly convex functions are always convex, but the converse is not necessarily true.

5.4 Operations that Preserve Convexity

5.4.1 Convexity is preserved by several key operations

A powerful component of the concept of convexity is that it is preserved under a set of important operations, including:

- *Non-negative weighted sum*, ie: linear combination with non-negative weights. If we have several convex functions f_1, f_2, ..., f_K, and corresponding non-negative weights $w_1 \geq 0$, $w_2 \geq 0$, ..., $w_K \geq 0$, then the function $f(\mathbf{x}) = w_1 f_1(\mathbf{x}) + w_2 f_2(\mathbf{x}) + \cdots + w_K f_K(\mathbf{x})$ is also convex.
- *Pointwise maximum.* If we have several convex functions f_1, f_2, ..., f_K, then the function $f(\mathbf{x}) = \max_k [f_k(\mathbf{x})]$ (where at each $\mathbf{x}$ we simply calculate the maximum of $f_1(\mathbf{x})$, $f_2(\mathbf{x})$, ..., $f_K(\mathbf{x})$), is also convex.
- *Composition with an affine mapping.* If $f(\mathbf{x})$ is convex, then $f(\mathbf{Ax} + \mathbf{b})$ is also convex, for any $M \times N$ matrix $\mathbf{A}$, and $M \times 1$ vector $\mathbf{b}$.
- *Restriction to a line.* If $f(\mathbf{x})$ is convex, and $\mathbf{x_1}$ and $\mathbf{x_2}$ are two points in $\mathbb{R}^N$, then $g(\alpha) = f(\mathbf{x_1} + \alpha\mathbf{x_2})$, for $\alpha \in \mathbb{R}$, is also convex.

5.4.2 Why all the fuss about convexity?

Question: Consider the first two examples in figure 5.2 above. The first one is a quadratic function ($f(x) = x^2$), and the second one is a square root function ($f(x) = \sqrt{|x|}$). Both of these functions have a single minimizer, at the bottom of a single valley. In this sense, they both seem similarly appropriate for iterative descent-based algorithms. However, $f(x) = x^2$ is convex, whereas $f(x) = \sqrt{|x|}$ is not. For each of these functions, how would it behave if we add up two shifted versions of itself, eg: $g_1(x) = x^2 + (x-2)^2$, and $g_2(x) = \sqrt{|x|} + \sqrt{|x-2|}$? Does g_1 still have a single valley with a single minimizer? How about g_2?