1 Lecture 2: Basic concepts in convex optimization

Recall that the training algorithms in machine learning boil down to the following constrained optimization problem:

$$\min_{x \in A} f(x), \quad A \subset \mathbb{R}^d. \tag{1.1}$$

In order to be able to solve the above optimization problem, we need to know the properties of f.

Remark 1.1. In unconstrained optimization: if f is convex then (1.1) attains a global minimum.

The goal of this lecture is to minimize a continuous and convex function over a convex subset of Euclidean space. Therefore, let $A \subset \mathbb{R}^d$ be a bounded convex and compact set in the Euclidean space. In the sequel, we will give some basic definitions which are used in convex optimization. The first step is define the convex sets.

Definition 1.2 A set A is convex if for any $x, y \in A$ all the points on the line segment connecting x and y also belong to A, i.e.,

$$\forall t \in [0,1] \quad ty + (1-t)x \in A. \tag{1.2}$$

Observe that the line segment is described parametrically. The line segment between x and y, which corresponds to $t = 0$ and $t = 1$ is contained to A. Now we give the definition for the convex functions.

Definition 1.3 A function $f : A \to \mathbb{R}$ is convex if for any $x, y \in A$ we have

$$\forall t \in [0,1], \quad f(ty + (1-t)x) \leq (1-t)f(x) + tf(y). \tag{1.3}$$

Intuitively, the line joining $(x, f(x))$ and $(y, f(y))$ lies above the function graph. This inequality is known as Jensen's inequality. Next, we give the definition of a convex function with a necessary and sufficient condition.

Definition 1.4 If f is differentiable, that is, its gradient exists for any $x \in A$ then it is convex if-f

$$\forall x, y \in A \quad f(y) \geq f(x) + \nabla f(x)^\top (x - y). \tag{1.4}$$

Intuitively, this means that a convex function if the gradient bounds from below the function.

1.1 Projections into convex sets

In the constrained optimization, we shall make use of a projection operation onto a convex set, which is defined as the closest point in terms of Euclidean distance inside the convex set to a given point. Formally,

$$\Pi_A(y) := argmin_{x \in A} \|x - y\|. \tag{1.5}$$

Remark 1.5. **Computational complexity of projections**. How do we know if x belongs to A or no? If we can devise a procedure that tells if x belongs or not to a set, then projections are computed in polynomial time. In some cases near linear time.

1.2 Optimality conditions

Optimality conditions are the conditions when a function attains a local optimum or a saddle point. These conditions are called KKT (Kusher-Kuhn-Tucker). Generally there are a lot of points in which a function is minimised. We define this set as $\{argmin_{x \in A} f(x)\}$.

A minimum of a convex and differentiable function on $\mathbb{R}$ is a point in which is derivative is equal to zero. Formally,

$$\nabla f(x) = 0 \Leftrightarrow x \in \{argmin_{x \in A} f(x)\}. \tag{1.6}$$

Now the optimality condition in $\mathbb{R}^d$ is very similar. It is given by the following theorem and it is called KKT condition.

Theorem 1.6 *Let $A \subset \mathbb{R}^d$ be a convex set and $x^* \in \{argmin_{x \in A} f(x)\}$. Then for any $y \in A$ we have:*

$$\nabla f(x^*)(y - x^*) \geq 0$$

.

Intuitively, if the inner product is positive one could improve the objective by moving towards in the direction of the projected negative gradient.