Convex Analysis and Gradient Descent Algorithms

Zed

June 30, 2017

Chapter 1

Convex Analysis

The optimization problem (†)

$$\min_{\boldsymbol{x}\in X} f(\boldsymbol{x})$$

is *guaranteed* to have an optimal solution if X is a *convex set* and $f : X \to \mathbb{R}$ is a *convex function.* Such family of problems is called convex optimization problems.

1.1 Basic Concepts

The global optimum of f in $\mathbb{R}^n$ may not fall within the feasible region of problem (†), i.e. the convex set X. Hence it is natural to think about a way to somehow "push back" a point $\boldsymbol{x} = (x_1, ..., x_n)^\top \in \mathbb{R}^n$ into X.

Def.1 **Projection onto Convex Set**: Let $X \subset \mathbb{R}^n$ be a closed convex set. For any $\boldsymbol{y} \in \mathbb{R}^n$, define the closest point to $\boldsymbol{y}$ in X as

$$\mathrm{Proj}_X(\boldsymbol{y}) := \operatorname*{argmin}_{\boldsymbol{x}\in X} \|\boldsymbol{y} - \boldsymbol{x}\|^2$$

The closedness of X is needed to ensure the existence of the projection. Furthermore, it can be shown that the projection onto closed convex set is also unique.

Def.2 **Differentiable Convex Function**: Suppose $f : X \to \mathbb{R}$ is differentiable over X. Then f is convex $\iff$ For any given $\boldsymbol{x}_0 \in \boldsymbol{X}$, and for all $\boldsymbol{x} \in \boldsymbol{X}$,

$$f(\boldsymbol{x}) - f(\boldsymbol{x}_0) \geq \langle \nabla f(\boldsymbol{x}_0), \boldsymbol{x} - \boldsymbol{x}_0 \rangle \tag{1.1}$$

Geometrically, this relation says the tangent plane of f[1] *lies below* the graph of f at any point in X. In another word, the tangent plane underestimates f everywhere in X.
A simple corollary: $\boldsymbol{x}^*$ minimizes differentiable convex function $f \iff \nabla f(\boldsymbol{x}^*)^\top = \boldsymbol{0}$. Because by definition we have $f(\boldsymbol{x}) - f(\boldsymbol{x}^*) \geq 0$ for all $\boldsymbol{x} \in X$.

The gradient can only be calculated for those $\boldsymbol{x}$ where f is differentiable. So we introduce subgradient to extend the notion to non-differentiable points.

Def.3 **Subgradient**: The vector $\boldsymbol{g} = (g_1, ..., g_n) \in \mathbb{R}^n$ is a subgradient of *convex function* f at $\boldsymbol{x}_0 \in X$ if $\forall \boldsymbol{x} \in X$,

$$f(\boldsymbol{x}) - f(\boldsymbol{x}_0) \geq \langle \boldsymbol{g}, \boldsymbol{x} - \boldsymbol{x}_0 \rangle \tag{1.2}$$

[1] Note that the graph of f, i.e. the surface $f(\boldsymbol{x}) - z = 0$ is an $(n+1)$-dimensional object with an extra dimension z. At a given point $\boldsymbol{x}_0 \in X$, the equation of the tangent plane of f is

$$z - f(\boldsymbol{x}_0) = \nabla f(\boldsymbol{x}_0)^\top (\boldsymbol{x} - \boldsymbol{x}_0)$$

And this is denoted by $\boldsymbol{g} \in \partial f(\boldsymbol{x}_0)$.
Geometrically, this relation says that the plane determined by point $\boldsymbol{x}_0$ and normal vector $(g_1, ..., g_n, -1)$ *lies below* the graph of f at any point in X.
Note that subgradient is defined specifically for convex functions, because it is an analogy to definition (1.1), which is only valid for convex functions.
The optimality corollary still holds for subgradient: $\boldsymbol{x}^*$ minimizes $f \iff \boldsymbol{0} \in \partial f(\boldsymbol{x}^*)$; since we have $f(\boldsymbol{x}) - f(\boldsymbol{x}^*) \geq 0$ for all $\boldsymbol{x} \in X$.

Prop.1 (*Existence of Subgradient for Convex f*) Let $X \subset \mathbb{R}^n$ be a convex set, $f : X \to \mathbb{R}$. Then

1. If $\partial f(\boldsymbol{x}) \neq \emptyset$ for all $\boldsymbol{x} \in X \Rightarrow f$ is convex function.
2. If f is convex function $\Rightarrow \partial f(\boldsymbol{x}) \neq \emptyset$ for all $\boldsymbol{x} \in \text{Int}(X)$.

1.2 Optimality Condition

Prop.2. f is a convex function and X convex set. $\boldsymbol{x}^*$ is an optimal solution of problem (†) $\iff$ there exists $\boldsymbol{g}^* \in \partial f(\boldsymbol{x}^*)$ such that $\forall \boldsymbol{x} \in X$,

$$\langle \boldsymbol{g}^*, \boldsymbol{x} - \boldsymbol{x}^* \rangle \geq 0 \tag{1.3}$$

Proof. Define the (reverse) indicator function of X

$$\tilde{1}_X(\boldsymbol{x}) = \begin{cases} 0, & \boldsymbol{x} \in X \\ \infty, & \boldsymbol{x} \notin X \end{cases} \tag{1.4}$$

Clearly, $\tilde{1}_X(\boldsymbol{x})$ is a convex function. And the problem (†) can be rewritten as unconstrained minimization, with the value at points beyond X being ∞; i.e. $\min f(\boldsymbol{x}) + \tilde{1}_X(\boldsymbol{x})$. By Corollary, $\boldsymbol{x}^*$ is optimal point $\iff \boldsymbol{0} \in \partial(f(\boldsymbol{x}^*) + \tilde{1}_X(\boldsymbol{x}^*)) \iff \exists \boldsymbol{g}^* \in \partial f(\boldsymbol{x}^*), \exists \boldsymbol{w}^* \in \partial \tilde{1}_X(\boldsymbol{x}^*)$, such that $\boldsymbol{g}^* + \boldsymbol{w}^* = \boldsymbol{0}$ (‡).

We want to study the property of the subgradient $\boldsymbol{w}$ of the artificially defined indicator. At a given point $\boldsymbol{x}_0 \in X$ and $\forall \boldsymbol{x} \in \mathbb{R}^n$, by definition

$$\tilde{1}_X(\boldsymbol{x}) - 0 \geq \langle \boldsymbol{w}, \boldsymbol{x} - \boldsymbol{x}_0 \rangle$$

I.e. $0 \geq \langle \boldsymbol{w}, \boldsymbol{x} - \boldsymbol{x}_0 \rangle$ if $\boldsymbol{x} \in X$; $\infty \geq \langle \boldsymbol{w}, \boldsymbol{x} - \boldsymbol{x}_0 \rangle$ otherwise. $\Rightarrow$

$$\partial 1_X(\boldsymbol{x}_0) = \{\boldsymbol{w} : \langle \boldsymbol{w}, \boldsymbol{x} - \boldsymbol{x}_0 \rangle \leq 0, \forall \boldsymbol{x} \in X\}$$

We have: $\boldsymbol{w}^* \in \partial 1_X(\boldsymbol{x}^*) \iff \langle \boldsymbol{w}^*, \boldsymbol{x} - \boldsymbol{x}^* \rangle \leq 0$ for all $\boldsymbol{x} \in X$.
$\boldsymbol{g}^* = \boldsymbol{0} - \boldsymbol{w}^* \iff \langle \boldsymbol{g}^*, \boldsymbol{x} - \boldsymbol{x}^* \rangle = 0 - \langle \boldsymbol{w}^*, \boldsymbol{x} - \boldsymbol{x}^* \rangle \geq 0$ for all $\boldsymbol{x} \in X$. (‡) becomes: $\boldsymbol{x}^*$ is optimal point $\iff \exists \boldsymbol{g}^* \in \partial f(\boldsymbol{x}^*)$, such that $\langle \boldsymbol{g}^*, \boldsymbol{x} - \boldsymbol{x}^* \rangle \geq 0$ for all $\boldsymbol{x} \in X$. □

1.3 Lagrange Duality

We add constraints to (†), and obtain a problem of the type:

$$\begin{aligned} & \min_{\boldsymbol{x} \in X} f(\boldsymbol{x}) \\ \text{s.t.} \quad & g_i(\boldsymbol{x}) \leq 0; \quad i = 1, ..., m; \\ & h_j(\boldsymbol{x}) = 0; \quad j = 1, ..., p \end{aligned} \tag{1.5}$$

where $X \subset \mathbb{R}^n$ is closed convex set. $f, g_i : X \to \mathbb{R}$ are convex functions. $h_j : X \to \mathbb{R}$, of the form $h_j(\boldsymbol{x}) = \boldsymbol{\theta}_j^\top \boldsymbol{x} + b_j$ are affine functions.