Convex Functions

1 Definitions

1.1 Jensen's Inequality

Definition 1. A function $f : \mathbb{R}^n \to \mathbb{R}$ is **convex** iff the line segment between any two points on its graph lie at or above the graph. This is formalized by Jensen's Inequality:

$$f(\theta \boldsymbol{x} + (1-\theta)\boldsymbol{y}) \leq \theta f(\boldsymbol{x}) + (1-\theta) f(\boldsymbol{y}) \ \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n, \theta \in [0,1]$$

It follows from this definition that f is convex iff its **restriction** to any line $\{\boldsymbol{x} + t\boldsymbol{v} : t \in \mathbb{R}\}$ is convex. So, all the 2D "slices" of the function must be convex.

Definition 2. A function $f : \mathbb{R}^n \to \mathbb{R}$ is **strictly convex** iff Jensen's Inequality is strictly satisfied everywhere $\boldsymbol{x} \neq \boldsymbol{y}$.

Definition 3. A function $f : \mathbb{R}^n \to \mathbb{R}$ is **concave** iff $-f$ is convex. Note that if f is both convex and concave, it must be affine.

1.2 First-Order Definition

Definition 4. A once-differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ is convex iff the first-order approximation at any point underestimates the value of the function everywhere. Formally,

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + (\nabla f(\boldsymbol{x}))^\intercal (\boldsymbol{y} - \boldsymbol{x}) \ \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$$

This definition allows us to conclude global information from local information because if $\nabla f(\boldsymbol{x}) = 0$ then $\boldsymbol{x}$ is a global minimizer of f.

1.3 Second-Order Definition

Definition 5. A twice-differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ is convex iff its Hessian matrix is positive semi-definite for all $\boldsymbol{x}$: $\nabla^2 f(\boldsymbol{x}) \succeq 0$. In the one-dimensional case, this simply means that the second derivative of f is non-negative everywhere.

1.4 Epigraph Definition

Definition 6. The **epigraph** of a function is the set of all points above the graph of the function. Formally, given a function $f : \mathbb{R}^n \to \mathbb{R}$, $epi(f) = \{(\boldsymbol{x}, t) : t \geq f(\boldsymbol{x})\}$. f is convex iff $epi(f)$ is a convex set; this useful fact allows us to translate a problem from convex functions to convex sets, which are generally easier to work with.

1.5 Generalized Convexity

In general, Jensen's Inequality applies for any convex combination of points in the domain of f. That is, given $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ in the domain of f and $\theta_1, \ldots, \theta_k \geq 0$ such that $\sum_i \theta_i = 1$,

$$f\Big(\sum_i \theta_i \boldsymbol{x}_i\Big) \leq \sum_i \theta_i f(\boldsymbol{x}_i)$$

If we consider infinitely many points, then we can translate Jensen's Inequality to the language of probability. Given a probability measure D with $\boldsymbol{x} \sim D$,

$$f\big(\mathbb{E}[\boldsymbol{x}]\big) \leq \mathbb{E}[f(\boldsymbol{x})]$$

This implies that adding noise to $\boldsymbol{x}$ can only increase $\mathbb{E}[f(\boldsymbol{x})]$.

2 Properties

2.1 Minima

Definition 7. A point $\boldsymbol{x}$ is a **local minimum** of a function f if there exists an open ball B containing $\boldsymbol{x}$ such that $f(\boldsymbol{y}) \geq f(\boldsymbol{x})\ \forall y \in B$.

If f is convex, then $\boldsymbol{x}$ is a local minimum of f iff $\boldsymbol{x}$ is a global minimum of f. This fact underlies much of the tractability of convex optimization, and is one reason why convex functions are so convenient to work with.

2.2 Sublevel Set

Definition 8. The α-sublevel set of a function f is $\{\boldsymbol{x} \in domain(f) : f(\boldsymbol{x}) \leq \alpha\}$.

Every sublevel set of a convex function is a convex set; however, not every function with convex sublevel sets is convex (these functions are called **quasiconvex**).

2.3 Other Properties

1. Real-valued convex functions are continuous on the interior of their domain.
2. If $f : D \to \mathbb{R}$ then we can transform it into $f : \mathbb{R}^n \to \mathbb{R}$ by setting $f(\boldsymbol{x}) = \infty$ where $\boldsymbol{x} \notin D$.

3 Convexity-Preserving Operations

To prove a function is convex, we can either prove it directly via Jensen's Inequality or express the function as a combination of other convex functions under certain convexity-preserving operations.

3.1 Combinations

Non-negative weighted combinations of convex functions are convex; this extends to infinite combinations in the form of integrals or expectations. Thus, $\min \mathbb{E}[f]$ for a convex f is a convex optimization problem – that is, stochastic convex optimization is equivalent to convex optimization.

Since non-negative weighted combinations are also conic combinations, this fact implies that convex functions form a convex cone in the vector space of functions from $\mathbb{R}^n \to \mathbb{R}$. Furthermore, the set of all convex functions is the intersection of an infinite set of homogenous linear inequalities:

$$f(\theta \boldsymbol{x} + (1-\theta)\boldsymbol{y}) - \theta f(\boldsymbol{x}) - (1-\theta) f(\boldsymbol{y}) \leq 0$$

3.2 Compositions

The composition of a convex function with an affine function is convex. That is, $g(\boldsymbol{x}) = f(\mathbf{A}\boldsymbol{x} + \boldsymbol{b})$ is convex when $f : \mathbb{R}^n \to \mathbb{R}$ is convex with $\mathbf{A} \in \mathbb{R}^{n \times m}, \boldsymbol{b} \in \mathbb{R}^n, g : \mathbb{R}^m \to \mathbb{R}$.

3.3 Maximization

If $f_1, \ldots, f_n$ are convex functions, then $\max_i f_i(\boldsymbol{x})$ is convex. This implies that when a convex cost function is uncertain, minimizing the worst-case scenario is a convex optimization problem. Furthermore, if $f(\boldsymbol{x}, \boldsymbol{y})$ is convex, and C is a convex nonempty set, then $g(\boldsymbol{x}) = \inf_{y \in C} f(\boldsymbol{x}, \boldsymbol{y})$ is a convex function.