Sub-gradient Methods

Gradient descent methods require ∇f to exist
What if ∇f does not exist at some points?

Recall that when ∇f exists

f is convex on $\mathcal{D} \iff f(y) \geq f(x) + \nabla f(x)^T (y-x), \forall x, y \in \mathcal{D}$

Inequality is strict for strict convexity

__Definition__ For __convex__ f on $\mathbb{R}^n$, g is called a sub-gradient of f at $x \in \mathbb{R}^n$ if

$$f(y) \geq f(x) + g^T (y-x) \quad \forall y \in \mathbb{R}^n$$

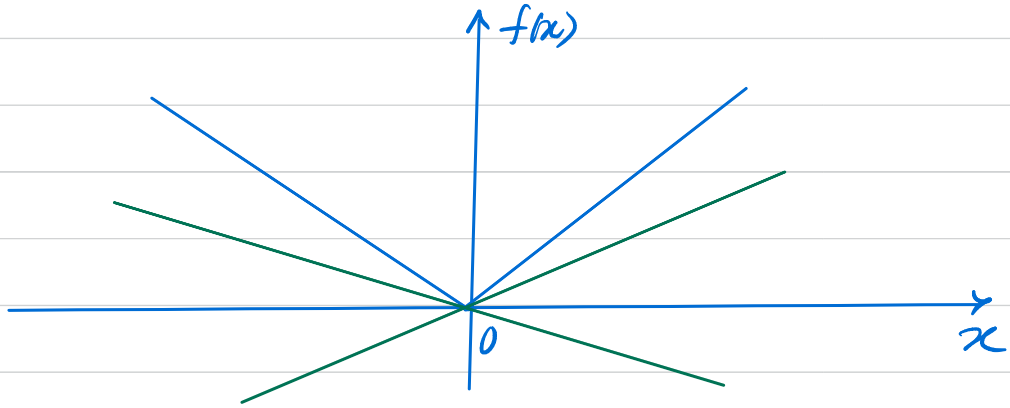
__Properties of sub-gradient (without proof)__

(1) Sub-gradients always exist at any point for convex functions

(2) If ∇f exists at a point x for convex f, Sub-gradient is unique and $= \nabla f(x)$

(3) Same definition for sub-gradient can be applied for non-convex f, but sub-gradient may not exist.

Example $f(x) = |x|, \quad x \in \mathbb{R}$



For $x \neq 0$, ∇f exists and $=$ sub-gradient

For $x = 0$, any $g \in [-1, 1]$ is a sub-gradient

Proof For $y > 0$,

$$f(y) = y \geq \overset{=0}{f(0)} + g\,y, \quad \forall g \in [-1, 1] \quad —(1)$$

For $y < 0$,

$$f(y) = -y \geq f(0) + g\,y, \quad \forall g \in [-1, 1] \quad —(2)$$

Note: Only $g \in [-1, 1]$ can be subgradients at 0. For $g > 1$, (1) fails to hold, and for $g < -1$, (2) fails to hold.

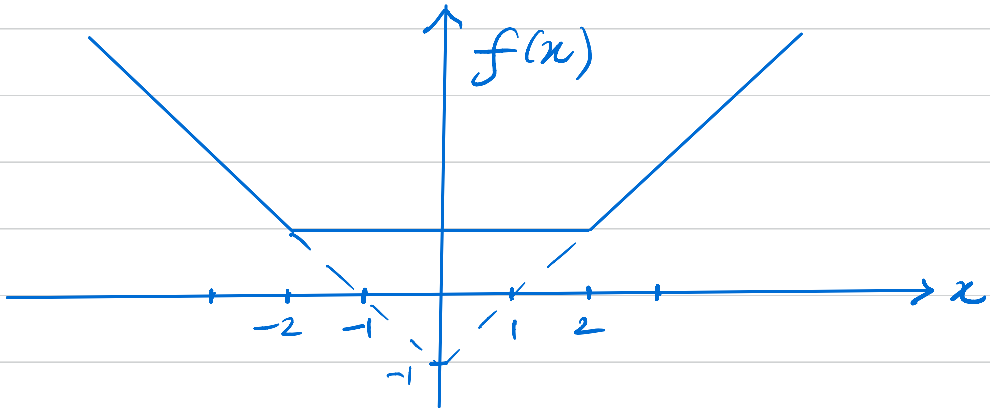
Definition set of all sub-gradients at x is called the sub-differential at x, denoted $\partial f(x)$

For $f(x) = |x|$,

$$\partial f(x) = \begin{cases} -1 & \text{if } x < 0 \\ [-1, 1] & \text{if } x = 0 \\ 1 & \text{if } x > 0 \end{cases}$$

Example $f(x) = \max\{1, |x|-1\}$

Note that f is convex since $f_1(x) = 1$ and $f_2(x) = |x|-1$ are convex and max of convex functions is convex



Guess

$$\partial f(x) = \begin{cases} -1 & \text{if } x < -2 \\ [-1, 0] & \text{if } x = -2 \\ 0 & \text{if } -2 < x < 2 \\ [0, 1] & \text{if } x = 2 \\ 1 & \text{if } x > 2 \end{cases}$$

Need to prove this!

At any point where gradient exists, i.e. for all x, except $x = -2$ and $x = 2$, we have

$$\partial f(x) = \nabla f(x) \text{ – unique}$$

So we need to show $\partial f(x) = \begin{cases} [-1, 0] & \text{if } x = -2 \\ [0, 1] & \text{if } x = 2 \end{cases}$

For $x = 2$, we need to show that for all $y \in \mathbb{R}$,

$$f(y) = \max\{1, |y|-1\} \geq \overset{=1}{f(2)} + g(y-2) \quad \text{--- (3)}$$

iff $g \in [0,1]$.

Case 1: $|y| < 2$, $f(y) = 1$,

RHS of (3) $= 1 + g\underbrace{(y-2)}_{<0} \leq 1 \quad \forall g \in [0,1]$

Thus (3) holds $\forall g \in [0,1]$.

Also, for $g < 0$, $1 + g(y-2) > 1 \Rightarrow$ (3) does not hold

Case 2 : $|y| \geq 2$, $f(y) = |y| - 1$

RHS of (3) $= 1 + g(y-2)$

$\leq 1 + g(|y|-2), \quad \forall g \in [0,1]$

$\leq 1 + |y| - 2 = f(y), \quad \forall g \in [0,1]$.

Thus (3) holds $\forall g \in [0,1]$.

Also, for $g > 1$, and $y > 2$,

$1 + g(y-2) = 1 + g(|y|-2) > 1 + |y| - 2 = f(y)$

$\Rightarrow$ (3) does not hold.

Thus (3) holds $\forall y$ iff $g \in [0,1]$

i.e., $\partial f(2) = [0,1]$

Similarly $\partial f(-2) = [-1, 0]$.

First-order Necessary Conditions for optimality in terms of Subgradients

For convex f,

$$f(x^*) = \min_x f(x) \iff 0 \in \partial f(x^*)$$

Proof x^* is a minimizer

$\iff f(x^*) \le f(y)$ for all $y \in \mathbb{R}^n$

$\iff f(x^*) + 0^T(y-x^*) \le f(y) \quad \forall y \in \mathbb{R}^n$

$\iff 0 \in \partial f(x^*)$

Properties of Subgradients (HW #6)

Let f, f_1, f_2 be convex functions.

(i) scaling: For scalar $a>0$, $\partial(af) = a\,\partial f$, i.e., $\forall x$, g is a subgradient of f at x $\iff$ ag is a subgradient of af at x

(ii) Addition: If g_1 is a subgradient of f_1 at x, and g_2 is a subgradient of f_2 at x, then g_1+g_2 is a subgradient of f_1+f_2 at x

(iii) Let $h(x) = f(Ax+b)$, A square invertible matrix. Then $\partial h(x) = A^T \partial f(Ax+b)$, i.e., $\forall x$, g is a subgradient of f at $Ax+b$ $\iff$ $A^T g$ is a subgradient of h at x.

Sub-gradient Descent for Unconstrained Optimization

Assumptions

(i) f is convex on $\mathbb{R}^n$.

(ii) $f^* = \inf_{x \in \mathbb{R}^n} f(x)$ exists and

there exists an x^* s.t. $f(x^*) = f^*$

(iii) For all $x \in \mathbb{R}^n$ and for all $g \in \partial f(x)$,

$$\|g\| \leq a$$

Subgradient Descent with constant step-size:

$$x_{k+1} = x_k - \alpha g_k \,, \quad g_k \in \partial f(x_k)$$

Analysis

$$\|x_{k+1} - x^*\|^2 = \|x_k - \alpha g_k - x^*\|^2$$

$$= \|x_k - x^*\|^2 + \alpha^2 \|g_k\|^2 - 2\alpha g_k^T (x_k - x^*)$$

$$\leq \|x_k - x^*\|^2 + \alpha^2 a^2 - 2\alpha g_k^T (x_k - x^*)$$

By definition of g_k,

$$f(x_k) + g_k^T (x^* - x_k) \leq f(x^*) = f^*.$$

$$\Rightarrow \|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2 + \alpha^2 a^2 + 2\alpha (f^* - f(x_k))$$

$$\|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2 + \alpha^2 a^2 + 2\alpha\,(f^* - f(x_k))$$

$$\Rightarrow \quad f(x_k) - f^* \leq \frac{\|x_k - x^*\|^2 - \|x_{k+1} - x^*\|^2 + \alpha^2 a^2}{2\alpha}$$

Define $\quad f_N^* = \min\{f(x_0), f(x_1), \ldots, f(x_{N-1})\}$

Note: $\quad f_N^* = \min\{f_{N-1}^*, f(x_{N-1})\}$

$$\sum_{k=0}^{N-1} \left(f(x_k) - f^*\right) \geq \sum_{k=0}^{N-1} \left(f_N^* - f^*\right)$$

$$= N\left(f_N^* - f^*\right)$$

$$\Rightarrow N\left(f_N^* - f^*\right) \leq \sum_{k=0}^{N-1} \frac{\|x_k - x^*\|^2 - \|x_{k+1} - x^*\|^2 + \alpha^2 a^2}{2\alpha}$$

$$= \frac{\|x_0 - x^*\|^2 - \|x_N - x^*\|^2 + N\alpha^2 a^2}{2\alpha}$$

$$\Rightarrow \quad f_N^* \leq f^* + \frac{1}{2\alpha N}\|x_0 - x^*\|^2 + \frac{\alpha a^2}{2}$$

$$\Rightarrow \quad \lim_{N \to \infty} f_N^* \leq f^* + \frac{\alpha a^2}{2}$$

$\Rightarrow$ for α small enough and N large enough f_N^* can be made as close to f^* as desired.