

Sub-gradient Methods

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

Recall that when ∇f exists

$$f \text{ is convex on } S \Leftrightarrow f(y) \geq f(x) + \nabla f(x)^T(y-x), \forall x, y \in S$$

Inequality is strict for strict convexity

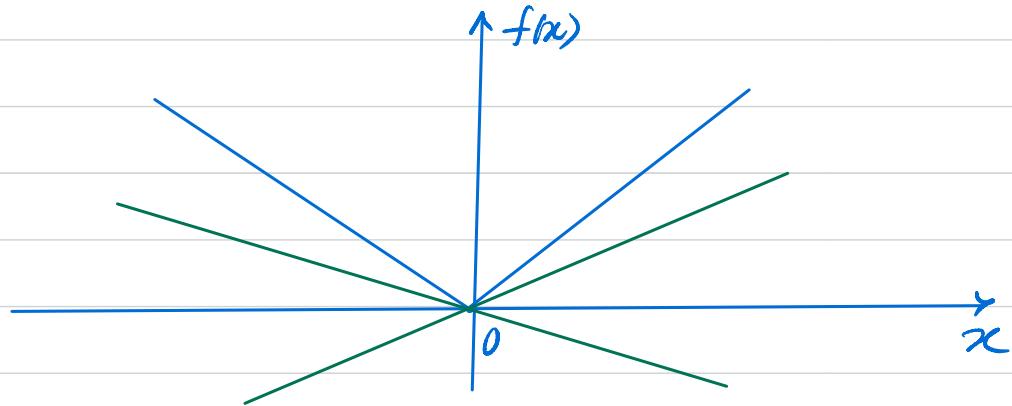
Definition For convex f on \mathbb{R}^n , g is called a sub-gradient of f at $x \in \mathbb{R}^n$ if

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

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 at all points.

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



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

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

Proof

For $y > 0$,

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

For $y < 0$,

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

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.

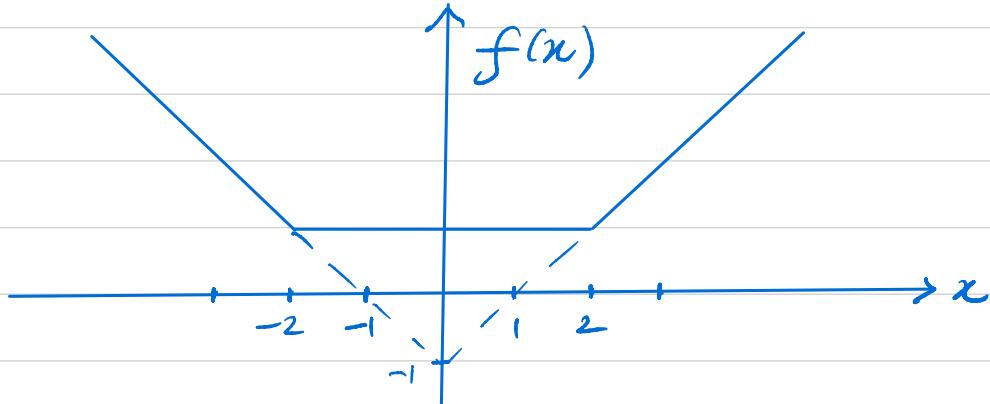
Definition Set of all sub-gradients of f at x is called the sub-differential at x , denoted $\partial f(x)$

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 $f(x) = \max\{f_1(x), f_2(x)\}$

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 $y \in \mathbb{R}$,

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

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

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

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

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

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

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

$$\begin{aligned}\text{RHS of (3)} &= 1 + g(y-2) \\ &\leq 1 + g(|y|-2), \quad \forall g \in [0, 1] \\ &\leq 1 + |y| - 2 = f(y), \quad \forall g \in [0, 1].\end{aligned}$$

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 $g \in [0, 1]$

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

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

First-order necessary/sufficient conditions for optimality
in terms of subgradients For convex f ,

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

Proof x^* is a minimizer

$$\Leftrightarrow f(x^*) \leq f(y) \text{ for all } y \in \mathbb{R}^n$$

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

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

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 $\Leftrightarrow ag$ is a subgradient
of f at x 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 $Ax+b$ $\Leftrightarrow A^T g$ 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 $g \in \partial f(x)$,
 $\|g\| \leq \alpha$

Subgradient Descent with constant step-size:

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

Analysis

$$\begin{aligned} \|x_{k+1} - x^*\|^2 &= \|x_k - \alpha g_k - x^*\|^2 \\ &= \|x_k - x^*\|^2 + \alpha^2 \|g_k\|^2 - 2\alpha g_k^T (x_k - x^*) \\ &\leq \|x_k - x^*\|^2 + \alpha^2 \alpha^2 - 2\alpha g_k^T (x_k - x^*) \end{aligned}$$

By definition of g_k ,

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

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

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

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

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

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

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

$$\Rightarrow N(f_N^* - f^*) \leq \sum_{k=0}^{N-1} \frac{\|x_k - x^*\|^2 - \|x_{k+1} - x^*\|^2 + \alpha^2 a^2}{2\alpha} \\ = \frac{\|x_0 - x^*\|^2 - \|x_N - x^*\|^2 + N\alpha^2 a^2}{2\alpha}$$

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

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

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