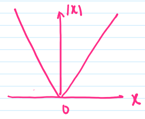
Subgradient Calculus

1:01 PM

Example abs(x), l-1 norm   Subgradient Contents:

g_x subgradient at $x \in \text{relint dom} f$ if $\forall_{y \in \text{dom} f}$ $f(y) \geq f(x) + g_x^T(y-x)$

$\boxed{\text{Subgradient of } f(x_{\in\mathbb{R}}) = |x| \text{ at } x}$



$x>0,\ f(x)=x \Rightarrow \partial f(x_{\{>0\}}) = \{\nabla f(x)\} = \{1\} = \text{sgn}(x)$

$x<0,\ f(x)=-x \Rightarrow \partial f(x_{\{<0\}}) = \{\nabla f(x)\} = \{-1\} = \text{sgn}(x)$

$x=0$, $f(x)$ not differentiable, $\nabla f(x)$ doesnot exist, $\forall_{y \in \mathbb{R}}$ $f(y) = |y| = \max_{-1\leq g \leq 1} gy \geq gy \quad \forall_{g:|g|\leq 1} \rightarrow \forall_{y \in \text{dom} f}$ $\underbrace{f(y) \geq gy}$ $\{-1 \leq g \leq 1\}$

$\Leftrightarrow |g| \leq 1$

$f(y) \geq \underset{0}{\underset{\|}{f(0)}} + g(y-0)$ $\{-1 \leq g \leq 1\}$ → by defn of subgradient, $g \in [-1,1]$ is a subgradient of $f(x)=|x|$ at 0.

∗ We will use the identity: $|z| = \max_{|\rho| \leq 1} \rho z$

e.g. $|-3| = \max_{-1\leq\rho\leq 1} (-3\rho) = \max[-3,3] = 3$

$|3| = \max_{-1\leq\rho\leq 1} (3\rho) = \max[-3,3] = 3$ } so it works!

$\therefore\ \partial f(x_{\{x=0\}}) = [-1,1]$

$$\partial|x| = \begin{cases} \text{sgn}(x), & \text{if } x \neq 0 \\ [-1,1], & \text{if } x=0 \end{cases}$$

subdifferential of abs(x)

$\Leftrightarrow \overbrace{\max_{i\in\{1,\dots,n\}} |g_i|}^{\|g\|_\infty} \leq 1 \Leftrightarrow \|g\|_\infty \leq 1$

∗ Subgradient of ℓ_1 norm function, $f(x) = \|x\|_1$, $x \in \mathbb{R}^n$ at $x=0$

$$f(y) = \|y\|_1 = \sum_{i=1}^n \underbrace{|y_i|}_{\left(\max_{|g_i|\leq 1} g_i y_i\right)} = \sum_{i=1}^n \max_{|g_i|\leq 1} g_i y_i = \underbrace{\left(\max_{|g_1|\leq 1} g_1 y_1\right)}_{\geq g_1 y_1 \{|g_1|\leq 1\}} + \dots + \underbrace{\left(\max_{|g_n|\leq 1} g_n y_n\right)}_{\geq g_n y_n \{|g_n|\leq 1\}} \geq g_1 y_1 + \dots + g_n y_n \ \{\overbrace{\forall_{i\in\{1,\dots,n\}} |g_i| \leq 1}\} = \sum_{i=1}^n g_i y_i \ \{\|g\|_\infty \leq 1\}$$

$\Rightarrow \quad f(y) \geq \sum_{i=1}^n g_i y_i \ \{\|g\|_\infty \leq 1\} = g^T y \ \{\|g\|_\infty \leq 1\}$

$\Leftrightarrow f(y) \geq \underset{\|0\|_1 = 0}{\underset{\|}{f(0)}} + g^T(y-0) \ \{\|g\|_\infty \leq 1\}$

$\Leftrightarrow \ g_{\{\|g\|_\infty \leq 1\}} = \text{subgradient}\left(\|\boxminus\|_1\right)\Big|_{\boxminus=0}$

For general formula of subgradient of l-1 norm at any point, see: [subgradient of l-1 norm]

Subgradient Calculus

Subgradient Calculus:

∗ Chain rule: $q: \mathbb{R}^n \to \mathbb{R}^m$, $h: \mathbb{R}^m \to \mathbb{R}$

$f = h \circ q \Leftrightarrow f(\boxminus) = h(q(\boxminus))$

$\{\nabla\}$

$\partial f(x) = (Dq(x))^T \partial h(q(x))$ # $Dq(x)$ is the jacobian

// origin:

$\nabla f(x) = (Df(x))^T = \left(D(h\boxminus)_{\boxminus=q(x)} D_x q(x)\right)^T = (D_x q(x))^T \left(D(h\boxminus)_{\boxminus=q(x)}\right)^T = \nabla q(x) \left(\nabla h(\boxminus)\right)_{\boxminus=q(x)}$

$Df(x) = D(h(q(x))) = \underbrace{D_{q(x)} h(q(x))}_{D(h(\boxminus))_{\boxminus=q(x)}} D_x q(x)$

Affine transformation: (special case)

$f(x) = h(Ax+b)$

$\therefore\ \nabla f(x) = \nabla(h(Ax+b))$

$Df(x) = D(h(Ax+b)) = D_{Ax+b} h(Ax+b) \cdot D_x(Ax+b)$ $\left\{ D_x(Ax+b) = D_x Ax = \left(\frac{\partial}{\partial x_j} \sum_k a_{ik} x_k\right)_{ij} = \left(\sum_k a_{ik} \frac{\partial x_k}{\partial x_j}\right)_{ij} = \left(\sum_k a_{ik} \{k=j\}\right)_{ij} = (a_{ij})_{ij} = A \right\}$

$= (Dh(\boxminus))_{\boxminus=Ax+b} A$

So for differentiable case: $\nabla f(x) = \left(Dh(\boxminus)_{\boxminus=Ax+b} A\right)^T = A^T \underbrace{\left(Dh(\boxminus)_{\boxminus=Ax+b}\right)^T}_{(\nabla h(\boxminus))_{\boxminus=Ax+b}} = A^T \nabla h(\boxminus)_{\boxminus=Ax+b}$

Extending to subgradient:

$\partial f(x) = A^T (\partial h(\boxminus))_{\boxminus=Ax+b \in \text{int dom} h}$

∗ Example:

$f(x) = |a^T x - b| = (\text{abs} \circ a^T\boxminus - b)(x) = \text{abs}(a^T x - b)$

// pretend differentiable $\nabla f(x) = ?$

// $Df(x) = D_{a^Tx+b} \text{abs}(a^T x - b)\ D_x(a^T x - b) = (D_\square |\square|)_{\square = a^T x + b} (D_x(a^T x - b))$

$\therefore\ \partial f(x) = Df(x)^T = \left((D_\square|\square|)_{\square=a^Tx\dots} (D_x(a^Tx-b))\right)^T$

// $Df(x) = D_{a^Tx+b}\, \mathrm{abs}(a^Tx-b)\; D_x(a^Tx-b) = \left(D_\square |\square|\right)_{\square = a^Tx+b} \left(D_x(a^Tx-b)\right)$

$$\therefore\; \partial f(x) = Df(x)^T = \left(\left(D_\square |\square|\right)_{\square = a^Tx-b} \left(D_x(a^Tx-b)\right)\right)^T$$

$$= \left(D_x(a^Tx-b)\right)^T \left(D_\square |\square|\right)^T_{\square = a^Tx-b}$$

$$= \underbrace{\nabla_x (a^Tx-b)}_{a} \left(\partial_\square |\square|\right)_{\square = a^Tx-b}$$

not differentiable, hence subgradient

$$\partial_\square |\square| = \begin{cases} \mathrm{sgn}\,\square, & \square \neq 0 \\ [-1,1], & \square = 0 \end{cases} = \mathrm{sgn}\,\square \,\{\square \neq 0\} + [-1,1]\,\{\square = 0\}$$

$$= a\left(\mathrm{sgn}(a^Tx-b)\,\{a^Tx-b \neq 0\} + [-1,1]\,\{a^Tx-b = 0\}\right)$$

* Sum or linear combination

$$f(x) = \overset{\{\geq 0\}}{\alpha}\, \underbrace{h(x)}_{\{\square\}} + \overset{\{\geq 0\}}{\beta}\, \underbrace{q(x)}_{\{\square\}}$$

$\forall\, x \in \mathrm{relint\ dom}\, h \cap \mathrm{relint\ dom}\, q$

$$\partial f(x) = \alpha\, \partial h(x) + \beta\, \partial q(x)$$

$$\Leftrightarrow \partial(\alpha h(x) + \beta q(x)) = \alpha\, \partial h(x) + \beta\, \partial q(x)$$

// proper parsing all elements of $\partial h(x)$ are multiplied by α } → all the elements of these resultant

// 〃 〃 〃 $\partial q(x)$ 〃 〃 〃 β } sets are added together.

Example: $f(x) = \sum_{i=1}^n |a_i^Tx+b|$

$$\partial f(x) = \partial\left(\sum_{i=1}^n |a_i^Tx+b|\right) = \sum_{i=1}^n \partial |a_i^Tx+b| = \sum_{i=1}^n \left(a_i\left(\mathrm{sgn}(a_i^Tx-b)\,\{a_i^Tx-b \neq 0\} + [-1,1]\,\{a_i^Tx-b = 0\}\right)\right)$$

now. $\partial f(x) = \partial \|x\|_1 = \partial \sum_{i=1}^n |x_i| = \partial \sum_{i=1}^n |e_i^Tx+0| = \sum_{i=1}^n e_i\; \mathrm{sgn}(\underbrace{e_i^Tx}_{x_i}+0)\,\{\underbrace{e_i^Tx}_{x_i}+0 \neq 0\} + [-1,1]\,\{\underbrace{e_i^Tx}_{x_i}+0 = 0\}$

$$= \sum_{i=1}^n e_i\, \mathrm{sgn}(x_i)\,\{x_i \neq 0\} + [-1,1]\,\{x_i = 0\}$$

[subgradient of l-1 norm]