Let \( f : \mathbb{R}^n \to \mathbb{R} \) be continuous (not essential, but simplify the def's).

We say \( y \) is a \textit{regular subgradient} of \( f \) at \( x \) (\( y \in \partial f(x) \)) if

\[
\liminf_{\|z^{(n)}\| \to 0} \frac{f(x + z^{(n)}) - f(x) - y^T z^{(n)}}{\|z^{(n)}\|} \geq 0.
\]

We sometimes write this as

\[
f(x + z) \geq f(x) + y^T z + o(\|z\|)
\]

Compare this with the convex case, in which \( y \in \partial f(x) \) requires

\[
f(x + z) \geq f(x) + y^T z \text{ must hold for all small } z.
\]

Now, a similar inequality must hold \textit{for all small } \( z \).

\( f : \mathbb{R} \to \mathbb{R} \)

\[
\begin{align*}
\hat{\partial}f(x_1) &= (-\infty, \infty) \\
\hat{\partial}f(x_2) &= \mathbb{R} \\
\hat{\partial}f(x_3) &= [-1, 1.5] \\
\hat{\partial}f(x_4) &= \emptyset
\end{align*}
\]

\( \hat{\partial}f(x) \) is called the \textit{regular subdifferential} of \( f \) at \( x \).
By def., \( \hat{\partial} f(x) \) is closed and convex but not necessarily compact or nonempty.

In case \( n \geq 1 \), \[ \begin{bmatrix} y \\ -1 \end{bmatrix} \] is normal to a hyperplane in \( \mathbb{R}^{n+1} \) passing through \((x, f(x))\) and lying underneath the graph of \( f \) locally, TO FIRST ORDER.

We say \( y \) is a (GENERAL) SUBGRADIENT of \( f \) at \( x \) (\( y \in \hat{\partial} f(x) \))

\[ \exists \{ x^{(n)} \}, \{ y^{(n)} \} \text{ with } \]
\[ x^{(n)} \to x, \quad y^{(n)} \to y, \quad y^{(n)} \in \hat{\partial} f(x^{(n)}). \]

Clearly \( \hat{\partial} f(x) \subseteq \partial f(x) \) (take \( x^{(n)} \equiv x \), \( y^{(n)} \equiv y \in \hat{\partial} f(x) \)).

In our example, \( \hat{\partial} f(x) \neq \partial f(x) \) ?

Answer: only \( x_4 \), with \( \partial f(x_4) = [1.5, -1] \) NOT A CONVEX SET.

We say \( y \) is a HORIZON SUBGRADIENT of \( f \) at \( x \) (\( y \in \partial^{\infty} f(x) \))

\[ \exists \{ x^{(n)} \}, \{ y^{(n)} \} \in \mathbb{R}^n, \{ t_n \} \in \mathbb{R}_+ \text{ with } \]
\[ x^{(n)} \to x, \quad t_n y^{(n)} \to y, \quad t_n \to 0 \quad (\text{ie.} \ t_n \downarrow 0) \]
\[ y^{(n)} \in \partial f(x^{(n)}). \]

If \( \hat{\partial} f(x) \neq \emptyset \), then \( 0 \in \partial^{\infty} f(x) \) (take \( x^{(n)} \equiv x \), \( y^{(n)} \in \hat{\partial} f(x) \), \( t_n \equiv 0 \)).

In our example, \( \partial^{\infty} f(x_1) \neq \emptyset \)?

Answer: \( x_1, x_2 \) : \( \partial^{\infty} f(x_1) = (-\infty, 0), \partial^{\infty} f(x_2) = \mathbb{R} \).
We call $\partial f(x)$ the subdifferential of $f$ at $x$ and $\partial^\infty f(x)$ the horizon subdifferential of $f$ at $x$.

If $f$ is convex or $f$ is $C^1$ (continuously differentiable) at $x$, then
\[
\partial f(x) = \hat{\partial} f(x), \quad \partial^\infty f(x) = \{0\}.
\]

Note: If $f(x) = -|x|$, then $\hat{\partial} f(0) = \emptyset$, $\partial f(0) = \{-1, 1\}$
and $\partial^\infty f(0) = \{0\}$.

**Simplest Nontrivial Example**

(see also, "Nonsmooth Analysis of Eigenvalues")

Let $\varphi_k(x) = k^{th}$ largest element of $\{x_1, \ldots, x_n\}$.

Let $\varphi_k(x) = \chi_{\{x\}}$ in BV notation.

Clearly $\varphi_k$ is convex iff $k = 1$.

Thus $\hat{\partial} \varphi_k(x) = \begin{cases} \text{Conv} \{e_i : x_i = \varphi_k(x)\} & \text{if } k = 1 \\
\{e_{k+1}, \ldots, e_n\} & \text{if } k > 1 \text{ and } \varphi_{k-1}(x) > \varphi_k(x) \\
\emptyset & \text{otherwise}
\end{cases}$

where $e_i = \begin{bmatrix} 0 \\
1 \end{bmatrix}$.

E.g., $x = \begin{bmatrix} 1 \\
3 \\
2 \end{bmatrix} \in \mathbb{R}^5$

$\hat{\partial} \varphi_1(x) = \begin{bmatrix} 0 & 0 \\
0 & 1 - \tau \end{bmatrix}$, $\tau \in [0, 1]$

$\hat{\partial} \varphi_2(x) = \emptyset$ etc.
Let $J = \{ i : x_i = \phi_k(x) \}$

If $k = 1$, then $\phi_k$ is convex so

$$\hat{\delta}(x) = \delta(x) = \{ y : \phi_k(x + z) = \phi_k(x) + y^T z \quad \forall z \in \mathbb{R}^n \}$$

$$= \operatorname{conv} \{ e^i : i \in J \}.$$

If $k > 1$ and $\phi_{k-1}(x) > \phi_k(x)$ then (sufficiently) close to $x$,

$\phi_k$ is equivalent to

$$w \mapsto \max_{i \in J} (w_i)$$

This is convex with subdifferential $\operatorname{conv} \{ e^i : i \in J \}$

so this set is $\hat{\delta}(\phi_k(x))$.

On the other hand, if $\phi_{k-1}(x) = \phi_k(x)$, where $|J| \geq 2$, suppose $y \in \hat{\delta}(\phi_k(x))$ so

$$\phi_k(x + z) \geq \phi_k(x) + y^T z + o(z) \quad \text{(near sequence z \to 0)}$$

for any index $i \in J$, all suff small $\delta > 0$, where

$$\phi_k(x + \delta e^i) = \phi_k(x)$$

since the perturbation $\delta e^i$ changes only one of the two or more entries equal to $\phi_k(x)$. So, $y_i \leq 0$. Also

$$\phi_k(x - \delta \sum_{i \in J} e^i) = \phi_k(x) - \delta$$

since the perturbation changes all the entries equal to $\phi_k(x)$ so $y^T z = -\delta \sum y_i \leq -\delta$, i.e. $\sum y_i \geq 1$: contradiction.

$\therefore \hat{\delta}(\phi_k(x)) = \emptyset$.
Since \( \tilde{\psi}_n(x) \) is bounded, clearly \( \tilde{\psi}_n(x) \rightarrow 0^+ \) for, by def.

Then \( \partial \psi_n(x) = \{ y : y \in \text{conv} \{ \epsilon_i : x_i = \psi_n(x) \} \}

and \( \# \{ y_i \text{ that are } > 0 \} \leq \# \{ x_i \text{ that are } \geq \psi_n(x) \} - k + 1 \) \( \ast \). 

E.g. \( x = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 3 \\ 3 \\ 2 \\ 2 \\ 2 \\ 2 \\ 2 \end{bmatrix} \)

\[ \partial \psi_1(x) = \left\{ \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} : z \in [0,1] \} , \text{a.e.} (x) = 2(1+1) = 4 
\]

(Which we already knew)

\[ \partial \psi_2(x) = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \right\} , \text{a.e.} (x) = 2-2+1 = 1 \]

NOT CONVEX.

\[ \text{Pf.} \text{ Cong. see Lewis, NSAE.} \]

Def. the \{ recession core \} of a nonempty, closed, convex set \( S \), denoted \( S^\infty \), is

\[ \{ r : w + r e S \forall e R_+ \} \]

where \( w \) is any given element of \( S \).

E.g. \( S \)

\[ (\text{doesn't matter whether } 0 \in S) \]

\[ (\text{always } 0 \in S^\infty) \]

E.g. \( \{ [-1, \infty) \}^\infty = \{ [1, \infty) \}^\infty = R_+ \)
REGULARITY

$f \in \{\text{Clarks\ subdifferentially}\} \iff \text{REGULAR at } x$.

1. $\hat{\partial} f(x) = \hat{\partial} f(x) \neq \emptyset$ (not standard, assume for simplicity) and
2. $\bar{\partial} f(x) = (\hat{\partial} f(x))^\circ = \text{horizon cone of } \hat{\partial} f(x).

In our example, in what $x$ is $f$ not regular at $x$?

Answer: only $x_4$.

Fact: $f$ is convex $\implies f$ is regular at all $x$.

Thus regularity generalizes smoothness AND convexity.

e.g. $f(x) = \begin{cases} x^2 \sin \frac{1}{x} & x \neq 0 \\ 0 & x = 0 \end{cases}$

(Differentiable is not enough, e.g. $f(x) = \begin{cases} x^2 \sin \frac{1}{x} & x \neq 0 \\ 0 & x = 0 \end{cases}$)

This regularity generalizes smoothness AND convexity.

e.g. $\ell_k$ ($k^{th}$ largest element) is regular at $x$ if

$k = 1$ or $k > 1$ and $\ell_{k-1}(x) > \ell_k(x)$ (corollary of previous theorem p. SGN3).
Why do we care about regularity?

**THM (Chain Rule - simplest version)**

Let $A \in \mathbb{R}^{n \times m}$, $b \in \mathbb{R}^m$

Let $f: \mathbb{R}^n \to \mathbb{R}$, and define $h : \mathbb{R}^m \to \mathbb{R}$ by

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

Suppose

1. $f$ is regular at $A\bar{x} + b$ for some $\bar{x} \in \mathbb{R}^n$
2. $A^Ty = 0$ and $y \in \partial f(A\bar{x} + b) \Rightarrow y = 0$

i.e. $N(A^T) \cap \partial f(A\bar{x} + b) = \{0\}$

Then $h$ is regular at $\bar{x}$, with

$$\partial h(\bar{x}) = A^T \partial f(A\bar{x} + b)$$

$$\partial^2 h(\bar{x}) = A^T \partial^2 f(A\bar{x} + b)$$

**Ref:** Rockafellar & Wets, Springer 1998, Chapter 10.

Extremely useful property.
Def: \( f \) is Lipschitz (with constant \( L \)) on a set \( S \) if
\[ \| f(x) - f(y) \| \leq L \| x - y \| \quad \forall x, y \in S. \]

If \( f \) is Lipschitz on a set \( S \in \mathbb{R}^n \), we say \( f \) is locally Lipschitz at \( x \).

In our example, at which \( x \) is \( f \) not locally Lipschitz?

Answer: \( x_1, x_2 \).

Fact: If \( f \) is locally Lipschitz at \( x \), then \( \partial^c f(x) = \{ 0 \} \).

Suppose \( f \) is locally Lipschitz at \( x \). We say that \( y \) is a Clarke subgradient of \( f \) at \( x \) (written \( y \in \partial^c f(x) \)) if \( y \) is a convex combination of subgradients of \( f \) at \( x \), i.e.,
\[ \partial^c f(x) = \text{conv} (\partial f(x)). \]

E.g. \( f(x) = \| x \|_1 \) \( \partial^c f(x) = \text{conv} (\{-1, +1\}) = [-1, 1] \)

Fact \( \partial^c f(x) = \text{conv} \, G(x) \)

where \( G(x) = \{ g : \exists \epsilon(x) \to x, \text{ } f \text{ is differentiable at } x \}
with \( \nabla f(x) \to g \} \).

Note: If \( f \) is locally Lipschitz and regular at \( x \), then \( \partial^c f(x) = \partial f(x) = \hat{\partial} f(x) \) and \( \partial^c f(x) = \{ 0 \} \).

i.e. all 3 kinds of subgradients (Clarke, "general", regular) are the SAME.
Optimality Conditions

0 \in \hat{\nabla} f(x) \implies \text{Nec. condition on } x \text{ minimize } f, \\
\text{locally.}

Sufficient conditions are more complicated. Here is a very strong one.

Def \( \hat{x} \) : a sharp local minimizer of \( f \) if \( \exists \, \varepsilon > 0 \)

\( (*) \quad f(\hat{x} + z) - f(\hat{x}) \geq \varepsilon \| z \| \quad \forall z \text{ with } \| z \| \text{ suff. small,} \)

\[ f(\hat{x} + z) - f(\hat{x}) - \varepsilon w^T z \geq \frac{\varepsilon \| z \|}{\| z \|} \geq 0 \]

\( \forall z, \text{ open neighborhoods } \{ z \} \rightarrow 0 \)

so taking \( \liminf \), \( \forall w \in \hat{\nabla} f(\hat{x}) \), \( \forall z \in B \subset \hat{\nabla} f(\hat{x}) \)

where \( B = \{ w : \| w \| \leq \varepsilon \} \) (unit ball).

* 0 \in \text{int } \hat{\nabla} f(\hat{x}) \).

\((\Rightarrow) \quad \text{ Suppose } 0 \in \text{int } \hat{\nabla} f(\hat{x}), \text{ so } \exists \varepsilon > 0 \text{ with } \varepsilon B \subset \hat{\nabla} f(\hat{x}), \text{ or} \\
\liminf_{z \to 0} \frac{f(\hat{x} + z) - f(\hat{x}) - \varepsilon w^T z}{\| z \|} = 0 \quad \forall w \in B, \text{ inel. } w \text{ with } \| w \| = 1. \)
In particular, this is true for the sequence $x^{(n)} = \delta_n w$
$\delta_n \in \mathbb{R}, \delta_n \downarrow 0, \|w\| = 1.$

So,

$$\liminf_{\delta_n \downarrow 0} \frac{f(x + \delta_n w) - f(x)}{\delta_n} \geq \sigma$$

Let $\varepsilon < \sigma$. Then, for sufficiently small $\delta_n$

$$f(x + \delta_n w) - f(x) \geq \varepsilon \cdot \delta_n$$

Since this is true for any $w$ with $\|w\| = 1$, we have $(\ast)$. 