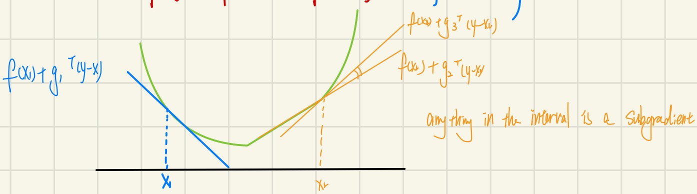
- **Subgradients**

for differentiable f: $f(y) \geq f(x) + \nabla f(x)^T(y-x)$

(first-order taylor is a global underestimator

g is a subgradient of f (not necessarily convex) if

$$f(y) \geq f(x) + \nabla f(x)^T(y-x) \quad \text{for all } y$$

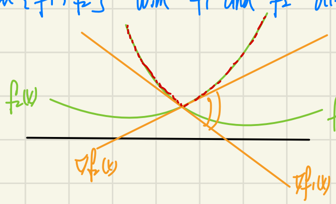


g is a subgradient of f at x iff $(g,-1)$ supports epi f at $(x, f(x))$

g is a subgradient iff $f(x) + g^T(y-x)$ is a global affine underestimator of f

if f is cvx and differentiable, $\nabla f(x)$ is the only subgradient

e.g. $f = \max\{f_1, f_2\}$ with f_1 and f_2 differentiable



when $f_1(x) > f_2(x)$: unique subgradient $g = \nabla f_1(x)$

when $f_2(x) > f_1(x)$: unique subgradient $g = \nabla f_2(x)$

when $f_1(x) = f_2(x)$: subgradients form a line segment $[\nabla f_1(x), \nabla f_2(x)]$

- **Subdifferential**

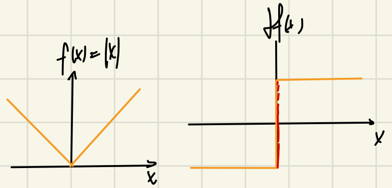
$\partial f(x)$ set of all subgradients of f is called the subdifferential of f at x,

in general, $\partial f(x)$ is a closed convex set (can be empty e.g. f not cvx)

if f is convex:

$\partial f(x)$ is nonempty for $x \in$ relint dom f

$\partial f(x) = \{\nabla f(x)\}$ iff f is differentiable at x



] the jump is the discontinuing of the jump

· Subgradient calculus

weak subgradient calculus: formulas to find one subgradient $g \in \partial f(x)$

strongly subgradient calculus: formulas to find the whole subdifferential

· basic rules

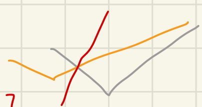
$\partial f(x) = \{\nabla f(x)\}$ if f is differentiable at x

· scaling: $\partial(\alpha f) = \alpha \cdot \partial f$ if $\alpha > 0$

· addition: $\partial(f_1 + f_2) = \partial f_1 + \partial f_2$ (addition of sets)

· Affine transformation $g(x) = f(Ax+b)$ $\partial g(x) = A^T \partial f(Ax+b)$

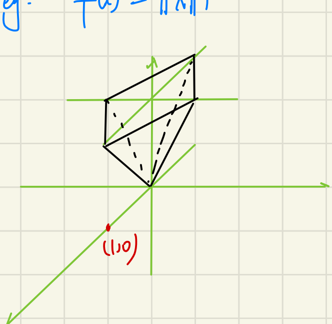
· finite pointwise maximum: $f = \max_{i=1,\dots,m} f_i$

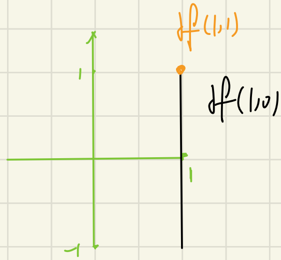


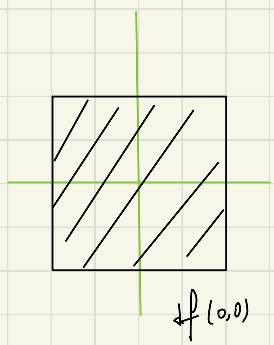
$\partial f = \text{Convex Hull}\left[\, \bigcup \{ \partial f_i(x) \mid f_i(x) = f(x) \} \right]$

convex hull of union of subdifferentials of "active functions"

· eg. $f(x) = \|x\|_1$



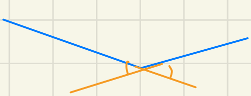


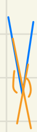


the subdifferential of a norm $\|\cdot\|$ at the original

is the unit ball of the dual norm

the size of the subdifferential is associated with the non-differentiability





kind of like the dual cone (duality)

· Weak rule for pointwise supremum

$$f = \sup_{a \in A} f_a$$

find any β for which $f_\beta(x) = f(x)$ (assuming supremum is achieved)
for any $g \in \partial f_\beta(x)$, $g \in \partial f(x)$

eg. maximum eigenvalue of a symmetric matrix under an affine function

$$f(x) = \lambda_{max}(A(x)) = \sup_{\|y\|_2 = 1} y^T A(x) y$$

$$A(x) = A_0 + x_1 A_1 + \cdots + x_n A_n$$

$g_y(x) = y^T A(x) y$ is a pointwise supremum over $\|y\| = 1$
$g_y(x)$ is affine in x with gradient $\nabla g_y(x) = [y^T A_1 y \ \cdots \ y^T A_n y]$

find the maximum eigenvalue of $A(x)$ and any associated eigenvector $\tilde{y}$
then $\nabla g_y(x) = [\tilde{y}^T A_1 \tilde{y} \ \cdots \ \tilde{y}_n^T A_n \tilde{y}_n]$ is a subgradient

· Expectation

$f(x) = E f(x, u)$ is convex in x if $f(x, u)$ is cvx in x for each u
(Expectation preserves convexity)

for each u, choose any $g_u \in \partial f(x, u)$, ($u \mapsto g_u$ is a function)
$g = E g_u \in \partial f(x)$ (expectation of subgradients)

Can be written analytically if u is gaussian, exponential ...

approximate $f(x)$ and a $g \in \partial f(x)$
generate iid samples $u_1, u_2 \cdots u_k$
$f(x) \approx 1/k \cdot \Sigma f(x, u_i)$ $g_i \in \partial f(x, u_i)$
$g = 1/k \cdot \Sigma g_i(x, u_i)$ is a approximated subgradient

- Minimization

define $g(y)$ as the optimal value of

$$\min_x \; f_0(x)$$
$$\text{s.t.} \quad f_i(x) \le y_i \qquad f_i \text{ convex}$$

with λ^* an optimal dual variable

$$g(x) \ge g(y) - \sum_i \lambda_i^*(x_i - y_i) \qquad (\text{perturbation analysis})$$

$-\lambda^*$ is a subgradient of g at y

- Composition

$f(x) = h[f_1(x) \cdots f_k(x)]$ with h cvx and nondecreasing, f_i convex

find $q \in \partial h(f_1(x) \cdots f_k(x))$ $\quad g_i \in \partial f_i(x)$

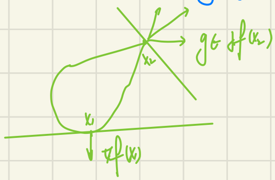
$g = \sum_i q_i \cdot g_i \in \partial f(x)$

$$\begin{aligned} f(y) &= h[f_1(y) \cdots f_k(y)] \\ &\ge h[f_1(x) + g_1^T(y-x), \cdots f_k(x) + g_k^T(y-x)] \\ &\ge h[f_1(x) \cdots f_k(x)] + q^T[g_1^T(y-x) \cdots g_k^T(y-x)] \\ &= f(x) + (\sum q_i \cdot g_i)^T(y-x) \end{aligned}$$

- Subgradients and sublevel sets

g is a subgradient at x means $f(y) \ge f(x) + g^T(y-x)$

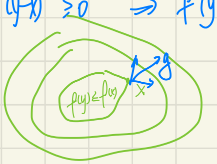
hence $f(y) \le f(x) \Rightarrow g^T(y-x) \le 0$



- Quasigradients

$g \ne 0$ is a quasigradient of f at x if

$$g^T(y-x) \ge 0 \Rightarrow f(y) \ge f(x)$$



quasigradients at x form a cone