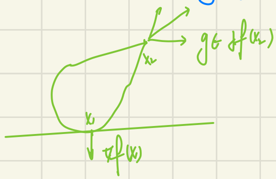
- Subgradients and sublevel sets

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

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

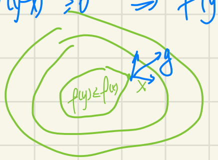


- Quasigradients

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

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

every thing in this half space has a higher function value



quasigradients at x form a cone

(when f is convex, a subgradient is a quasigradient)

eg. linear fractional

$$f(x) = \frac{a^Tx+b}{c^Tx+d} \qquad \mathrm{dom}\, f = \{x \mid c^Tx+d>0\}$$

$g = a - f(x_0)c$ is a quasigradient at x_0

for $c^Tx+d>0$

$$\frac{a^Tx+b}{c^Tx+d} \geq \frac{a^Tx_0+b}{c^Tx_0+d}$$

$$\frac{a^Tx+b}{a^Tx_0+b} \geq \frac{c^Tx+d}{c^Tx_0+d}$$

$$a^Tx+b \geq \frac{c^Tx+d}{c^Tx_0+d}(a^Tx_0+b)$$

$$(a^Tx+b)-(a^Tx_0+b) \geq \left[\frac{c^Tx+d}{c^Tx_0+d}-1\right](a^Tx_0+b)$$

$$a^T(x-x_0) \geq \frac{c^T(x-x_0)}{c^Tx_0+d}(a^Tx_0+b)$$

$$a^T(x-x_0) \geq f(x_0) \cdot c^T(x-x_0)$$

$$[a-f(x_0)\cdot c]^T(x-x_0) \geq 0$$

$g^T(x-x_0) \geq 0 \Rightarrow f(x) \geq f(x_0)$

eg. degree of $a_1 + a_2 t \cdots + a_n t^{n-1}$

$$f(a) = \min\{i \mid a_{i+2} = a_{i+3} \cdots = a_n = 0\}$$

$$g = \text{sign}(a_{k+1}) \cdot e_{k+1} \qquad k = f(a)$$

$$g^T(b-a) = \text{sign}(a_{k+1}) \cdot b_{k+1} - |a_{k+1}| \geq 0$$

$\therefore b_{k+1} \neq 0$

$\therefore f(b) \geq k$

- optimality condition — unconstrained

for a differentiable convex $f(x)$

$$f(x^*) = \inf_x f(x) \iff \nabla f(x^*) = 0$$

$f(y) \geq f(x) + \nabla f(x)^T (y-x)$ (goes from x to y along $-\nabla f(x)$, f decrease if $\nabla f(x) \neq 0$)

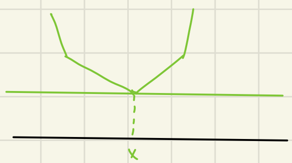
$f(y) \geq f(x)$ when $\nabla f(x) = 0$

for a nondifferentiable convex $f(x)$

$f(x^*) = \inf_x f(x) \iff 0 \in \partial f(x^*)$ (0 is a subgradient)

$f(y) \geq f(x) + g^T(y-x)$ for all y if g is a subgradient

if $g = 0$ $f(y) \geq f(x)$



$0 \in \partial f(x)$

(weak subgradient calculus will not help)

eg. piecewise linear minimization

$f(x) = \max_i (a_i^T x + b_i)$ is a convex

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

$\partial f(x^*) = \text{ConvexHull}\{a_i \mid a_i^T x + b_i = f(x^*)\}$

$\Updownarrow$

there is a λ such that $\underbrace{\lambda \geq 0 \quad \mathbf{1}^T \lambda = 1}_{\text{convex combination}}$ $\quad \sum \lambda_i a_i = 0$ } 0 is in the convex hull

$\lambda_i = 0$ if $a_i^T x^* + b < f(x^*)$ complemetarity

min. t

s.t. $a_i^T x + b_i \leq t$

$$\mathcal{L}(x, t, \lambda) = t + \lambda^T(Ax + b - t\mathbf{1})$$
$$= (1 - \mathbf{1}^T\lambda)t + (A^T\lambda)^T x + b^T\lambda$$

$g(\lambda) = b^T\lambda$ $(1 - \mathbf{1}^T\lambda = 0 \quad A^T\lambda = 0)$

dual: $\max b^T\lambda$ s.t. $A^T\lambda = 0$ $\mathbf{1}^T\lambda = 1$

- optimality condition — constrained

min. $f_0(x)$

s.t. $f_i(x) \le 0$

assume f_i convex, defined on $\mathbb{R}^n$ (hence subdifferentiable)

Strict feasibility (Slater's condition)

x^* is a primal optimal, λ^* is dual optimal iff

$f_i(x^*) \le 0 \quad \lambda_i^* \ge 0$

$0 \in \partial f_0(x^*) + \sum_i \lambda_i^* \cdot \partial f_i(x^*)$

$\lambda_i^* f_i(x^*) = 0$

$$f_0(x^*) = g(\lambda^*, \nu^*) = \inf_x \left[f_0(x) + \sum_i \lambda_i f_i(x) + \nu^T (Ax - b) \right]$$
$$\le f_0(x^*) + \sum_i \lambda_i f_i(x^*) + \nu^{*T}(Ax^* - b)$$
$$\le f_0(x^*)$$

x^* minimizes the Lagrangian.

- directional derivative

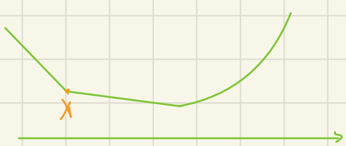
$$f'(x; \delta x) = \lim_{h \searrow 0} \frac{f(x + h\cdot\delta x) - f(x)}{h}$$

can be $+\infty, -\infty$

- f convex, finite near $x \Rightarrow f'(x;\delta x)$ exists
- when f differentiable at x

$\Updownarrow$

for some g $(= \nabla f(x))$ and all δx, $f'(x;\delta x) = g^T \delta x$ ($f'(x;\delta x)$ is a linear function of δx)



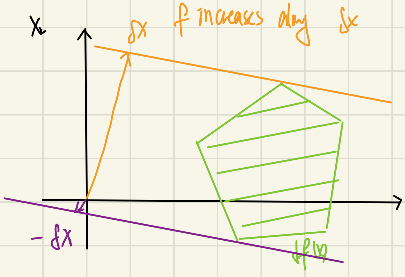
$f'(x;1) = f'_+(x)$ (right-hand slope)

$f'(x;-1) = f'_-(x)$ (left-hand slope)

$\Big\}$ same if f differentiable

- Directional derivative and subdifferential

general formula for convex f: $f'(x;\delta x) = \sup_{g \in \partial f(x)} g^T \delta x$ (support function of subdifferential)

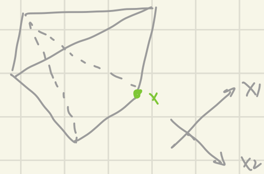


$$f'(x;-\delta x) = \sup_{g \in \partial f(x)} g^T(-\delta x) > 0$$

f increases along $-\delta x$

$\Big\{$ f increase along both δx and $-\delta x$

$\Longrightarrow$



f is non differentiable at x

subgradient > 0 along x_1

subgradient can be pos/neg along x_2

$\underbrace{\hphantom{xxxxxx}}$

$\nabla \rightarrow x_2$

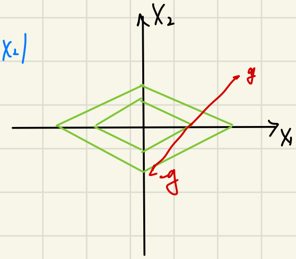
- descent direction

δx is a descent direction for f at x if $f'(x;\delta x) < 0$

(for differentiable f, $\delta x = -\nabla f(x)$ is always a descent direction, except $\nabla f(x) = 0$)

for nondifferentiable (convex) functions, $\delta x = -g$ with $g \in \partial f(x)$ may not be a descent direction

eg. $f(x) = |x_1| + 2|x_2|$



goes up along both g and $-g$

- subgradients and distance to sublevel sets

the negative subgradient is a descent direction for the distance to the optimal point

if f is convex, $f(z) < f(x)$, $g \in \partial f(x)$, then for small $t > 0$

$$\| x - tg - z \|_2 < \|x - z\|_2$$

thus $-g$ is a descent direction for $\|x-z\|_2$, for any z with $f(z) < f(x)$ (eg. x^*)

$$\begin{aligned} \|x - tg - z\|_2^2 &= (x - tg - z)^T (x - tg - z) \\ &= (x-z)^T(x-z) - 2tg^T(x-z) + (tg)^T(tg) \\ &\le \|x-z\|_2^2 + \underbrace{t^2\|g\|_2^2}_{\text{scale } t^2} \underbrace{- 2t(f(x) - f(z))}_{\text{scale } t} \end{aligned}$$

$f(z) \ge f(x) + g^T(z-x)$

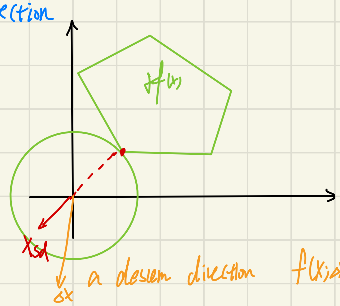
- descent direction and optimality

for f convex, finite near x,

either $\begin{cases} ① \ 0 \in \partial f(x) \\ ② \text{ there is a descent direction} \end{cases}$

x is optimal iff there is no descent direction for f at x

$$\delta x_{sd} = -\underset{z \in \partial f(x)}{\operatorname{argmin}} \|z\|$$



a descent direction $f'(x;\delta x) = \sup_{g \in \partial f(x)} g^T \delta x < 0$

- Subgradient methods

$$x^{(k+1)} = x^{(k)} - \alpha_k \cdot g^{(k)}$$

$g^{(k)}$ is any subgradient of f at x

not a descent method. need to keep track of the best point so far: $f_{best}^{(k)} = \min_i f(x^{(i)})$

- Step-size rules

Step sizes are fixed ahead of time

1. Costant step size: $\alpha_k = \alpha$ (costant)
2. Costant step length: $\alpha_k = \gamma / \|g^{(k)}\|_2$ (so $\|x^{(k+1)} - x^{(k)}\|_2 = \gamma$)
3. square summable but not summable:
   step sizes satisfy: $\sum_{k=1}^{\infty} a_k^2 < \infty$ $\sum_{k=1}^{\infty} a_k = \infty$
4. nonsummable diminishing:
   step sizes satisfy: $\lim_{k\to\infty} a_k = 0$ $\sum_{k=1}^{\infty} a_k = \infty$

- Assumptions

$f^* = \inf_x f(x) > -\infty$ with $f(x^*) = f^*$

$\|g\|_2 \le G$ for all $g \in \partial f$ (equivalent to Lipschitz condition on f)

$R \ge \|x^{(1)} - x^*\|_2$ (can take "=" here)

- Convergence results

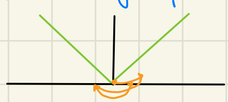
define $\bar{f} = \lim_{k\to\infty} f_{best}^{(k)}$

constant step size: $\bar{f} - f^* \le G^2 \cdot \alpha/2$

converges to $G^2 \cdot \alpha/2$ suboptimal

converges to f^* if f differential and α small enough

eg. $\min |x|$, returning $\partial f(x) = 0.75$ at $x=0$, which is a valid subgradient



end up oscillating

constant step length: $\bar{f} - f^* \le G\gamma/2$ converges to $G\gamma/2$ suboptimal

diminishing step size: $\bar{f} = f^*$ converges

eg. $\alpha^{(k)} = 1/k$