Subgradient

2:12 PM

$*$ Subgradient: (Subgradient notes Boyd page 6)

Alternative Proof: Subgradient of primal convex function

Optimal value function of a convex optimization problem: (eq: subgradient of optimal value function) Subgradient Contents:

$$f(x,y) = p^*(x,y) = \left(\begin{array}{ll} \gamma\ f_0(z) & \\ \rceil & \\ \forall_{i \in \{1,\dots,m\}}\ f_i(z) \le x_i & \text{// Here } x_i, y \text{ is perturbation in} \\ Az = y & \text{// resource vector} \end{array} \right)$$

// superscript
//* [illegible]

$$f(x,y) = \inf_z F(x,y,z)$$

$$F(x,y,z) = \begin{cases} f_0(z), & \forall_{i \in \{1,\dots,m\}} f_i(z) \le x_i,\ Az = y \\ +\infty, & \text{else} \end{cases}$$

$$\{D_{(x,y,z)}\}$$

• Goal: Want to find subdifferential of $f(x,y)\big|_{x=\hat{x},\ y=\hat{y}}$

Lets find Dual of (2): Lagrangian First:

$$L_{xy}(z,\lambda,v) = f_0(z) + \sum_{i=1}^{m} \overbrace{\lambda_i}^{\lambda \succeq 0}(f_i(z) - x_i) + v^T(Az - y)$$

$$= f_0(z) + \sum_{i=1}^{m} \lambda_i f_i(z) + v^T Az - \lambda^T x - v^T y$$

$$= \left(f_0(z) + \sum_{i=1}^{m} \lambda_i f_i(z) + v^T Az\right) - \lambda^T x - v^T y$$

dual function will be:

$$g_{xy}(\lambda,v) = -\lambda^T x - v^T y + \underbrace{\inf_z \left(f_0(z) + \sum_{i=1}^{m} \lambda_i f_i(z) + v^T Az\right)}_{g(\lambda,v)}$$

$$= -\lambda^T x - v^T y + g(\lambda,v)$$

$\therefore$ So, the dual problem is:

$$d^*(x,y) = \left(\begin{array}{l} R\ \ g(\lambda,v) - x^T\lambda - y^T v \\ \rceil \\ \lambda \succeq 0 \end{array} \right) \qquad (3)$$

now assume, $p^*(x,y) = d^*(x,y)$ for $x = \hat{x},\ y = \hat{y}$

$$\Rightarrow\ p^*(\hat{x},\hat{y}) = \underbrace{f(\hat{x},\hat{y}) = d^*(\hat{x},\hat{y})}$$

Assume for this, a primal optimal solution is z^*, dual optimal solution is (λ^*, v^*)

Now we will use global perturbation inequality:

$$\left\{ p^*(u,v) = \left(\begin{array}{ll} \gamma\ f_0(x) & \\ \rceil\ \forall_{i \in \{1,\dots,m\}} & f_i(x) \le u_i \\ \forall_{i \in \{1,\dots,p\}} & h_i(x) = v_i \end{array} \right) \right\} \forall_{u,v} \left(p^*(u,v) \ge p^*(0,0) - \lambda^{*T} u - v^{*T} v \right) \leftrightarrow \forall_{u,v}\ p^*(u,v) \ge p^*(\hat{u},\hat{v}) - \lambda^{*T}(u - \hat{u}) - v^{*T}(v - \hat{v})$$

$\{\lambda^*, v^*$ are the optimal dual vector for unperturbed $u=0, v=0$

Alternative Proof: Subgradient of primal convex function

Primal convex problem:

$$p^*(u) = \left(\begin{array}{ll} \gamma_{x \in D}\ f_0(x) & \\ \rceil & \\ \forall_{i \in \{1,\dots,m\}}\ f_i(x) \le u_i & \end{array} \right)$$

At first we will show that p*(u) is convex, then we will find a subgradient for this function

$$L_u(x,\overset{\lambda \succeq 0}{\lambda}) = f_0(x) + \sum_{i=1}^{m} \lambda_i (f_i(x) - u_i)$$

$$= f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) - \underbrace{\sum_{i=1}^{m} \lambda_i u_i}_{\lambda^T u} = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) - \lambda^T u$$

$$g_u(\lambda) = \min_{x \in D} L_u(x,\lambda)$$

the dual problem,

$$d^*(u) = \left(\begin{array}{l} R\ \ g_u(\lambda) \\ \rceil \\ \lambda \succeq 0 \end{array} \right) = \max_{\lambda \succeq 0} g_u(\lambda) = \max_{\lambda \succeq 0} \min_{x \in D} L_u(x,\lambda)$$

$\operatorname{argmax} = \lambda_u^*$ $\therefore d^*(u) = g_u(\lambda_u^*)$ argmax of d*(u)

note that: $L_u(x,\lambda) = f_0(x) + \lambda^T \left(\begin{bmatrix} f_1(x) - u_1 \\ \vdots \\ f_m(x) - u_m \end{bmatrix} \right)$

$$= f_0(x) + \underbrace{\begin{bmatrix} f_1(x) - u_1 \\ \vdots \\ f_m(x) - u_m \end{bmatrix}^T}_{\boxminus^T :\ \boxminus_i = f_i(x) - u_i} \lambda = f_0(x) + \boxminus^T \lambda$$

$$\therefore \max_{\lambda \succeq 0} L_u(x,\lambda) = \max_{\lambda \succeq 0} f_0(x) + \boxminus^T\lambda = f_0(x) + \underbrace{\max_{\lambda \succeq 0} \boxminus^T \lambda}_{h} = f_0(x) + \begin{cases} \infty, & \neg(\forall_i\ \boxminus_i \le 0) \\ 0, & (\forall_i\ \boxminus_i \le 0) \end{cases} = \begin{cases} \infty, & \neg(\forall_i\ \boxminus_i \le 0) \\ f_0(x), & \forall_i\ \boxminus_i \le 0 \end{cases} = \begin{cases} f_0(x), & \text{if } \forall_i\ f_i(x) - u_i \le 0 \\ \infty, & \text{else} \end{cases}$$

{ but, hey note! } $= f_0(x) + \sum_{i=1}^{m} I_{f_i(\boxminus_i) - u_i \le 0}(x)$

$$\# \ \max \sum_{i=1}^{m} \boxminus_i \lambda_i + \sum_{i=1}^{m} I_{\boxminus_i \ge 0}(\lambda) = \max_{\lambda} \sum_{i=1}^{m} \left(\boxminus_i \lambda_i + I_{\boxminus_i \ge 0}\, \lambda_i\right) \quad \text{\# seperable}$$

$$= \sum_{i=1}^{m} \max_{\lambda_i} \left(\boxminus_i \lambda_i + I_{\boxminus_i \ge 0}\, \lambda_i\right) = \sum_{i=1}^{m} \underbrace{\max_{\lambda_i \ge 0} (\boxminus_i \lambda_i)}_{} = \begin{cases} \infty, & (\exists_i\ \boxminus_i > 0) \equiv \neg(\forall_i\ \boxminus_i \le 0) \\ 0, & \forall_i\ \boxminus_i \le 0 \end{cases}$$

• if $\boxminus_i > 0$ then $\lambda_i \to +\infty$ will maximize $\boxminus_i \lambda_i$ with max value $= \infty$

• if $\boxminus_i \le 0$ then $\lambda_i = 0$ will " $\boxminus_i \lambda_i$ " " " $= 0$

Decomposition_using...

Again, $p^*(u) = \min_{x \in D} f_0(x) + \sum_{i=1}^{m} I_{f_i(\boxminus_i) - u_i \le 0}(x) = \min_{x \in D} \max_{\lambda \succeq 0} L_u(x,\lambda) = \min_{x \in D} \max_{\lambda \succeq 0} f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) - \lambda^T u$

Remember: $\forall_{y \in A}\ f(x,y) : D_x \Rightarrow \left(\sup_{y \in A} f(x,y)\right) = g(x) : D_x$

Now:

$$\forall_{\lambda \succeq 0}\ \underbrace{f_0(x)}_{D_x} + \underbrace{\sum \lambda_i f_i(x)}_{D_x\ \#\lambda_i \ge 0} + \underbrace{(-\lambda^T u)}_{D_u} : D_{(x,u)}$$

$$\Downarrow$$

$$\max_{\lambda \succeq 0} L_u(x,\lambda) = \left(\max_{\lambda \succeq 0} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) - \lambda^T u \right) \right) : D_{(x,u)}$$

again recall:

$$(f : D_{(x,y)}\ ;\ C : D, \neg\phi) \Rightarrow g(x) = \inf_{y \in C} f(x,y) : D_x$$

$$\min_{x \in D,\ D \neq \emptyset} \left(\max_{\lambda \succeq 0} L_u(x,\lambda) \right) : D$$

When strong duality holds,

$$p^*(u) = \min_{x \in D} \max_{\lambda \succeq 0} L_u(x,\lambda) = d^*(u) = \max_{\lambda \succeq 0} \min_{x \in D} L_u(x,\lambda)$$

If strong duality holds $\Rightarrow \min_{x \in D} \max_{\lambda \succeq 0} L_u(x,\lambda) = \max_{\lambda \succeq 0} \min_{x \in D} L_u(x,\lambda)$ # [illegible] min, max can be exchanged at strong # duality

Again, $\sum_{i=1}^{m} \lambda_i f_i(x) = \min_{\substack{\text{st} \\ \forall_i\ v_i \ge f_i(x)}} \sum \lambda_i v_i = \min_{v : \forall_i v_i \ge f_i(x)} \sum_{i=1}^{m} \lambda_i v_i = \min_{v \in \mathbb{R}^m} \lambda^T v + \sum_{i=1}^{m} I_{v_i \ge f_i(x)}(v)$

$$g_u(\lambda) = \min_{x \in D} L_u(x,\lambda) = \min_{x \in D} f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) - \lambda^T u$$

constant w.r.t. v constant w.r.t v

$$= \min\ f_0(x) + \left(\min\ \lambda^T v + \sum I\ (v) \right) - \lambda^T u$$

$p^*(u,v) = \left(\forall_{i\in\{1,\dots,m\}} f_i(x) \le u_i ;\ \forall_{i\in\{1,\dots,p\}} h_i(x) = v_i \right)$ ... u,v

{λ^*, ν^* are the optimal dual vector for unperturbed $u=0, v=0$ problem}

let $x=\hat{x}, y=\hat{y}$ be the unperturbed version of the problem, then global perturbation inequality [illegible]:

$$\forall_{x,y} \quad p^*(x,y) \ge p^*(\hat{x},\hat{y}) - \lambda^{*T}(x-\hat{x}) - \nu^{*T}(y-\hat{y})$$

$$\leftrightarrow \forall_{x,y} \quad f(x,y) \ge f(\hat{x},\hat{y}) + (-\lambda^*, -\nu^*)^T (x-\hat{x}, y-\hat{y})$$

$$\leftrightarrow \forall_{(x,y)} \quad f(x,y) \ge f(\hat{x},\hat{y}) + (-\lambda^*, -\nu^*)^T ((x,y) - (\hat{x},\hat{y}))$$

By definition of subgradient,

$$g\{\text{subgradient of } f:\mathbb{R}^n \to \mathbb{R} \text{ at } x \in \text{dom} f\} \leftrightarrow \forall_{z\in \text{dom} f} \quad f(z) \ge f(x) + g^T(z-x)$$

$$\leftrightarrow (-\lambda^*, -\nu^*) \{\text{subgradient of } f(\cdot,\cdot): \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R} \text{ at } (\hat{x},\hat{y})\}$$

[illegible]: [illegible] convex optimization [illegible] optimal value [illegible] optimization problem [illegible] resource disturbance [illegible] argument, then [illegible] argument [illegible] subgradient [illegible] convex optimization problem (with disturbance vector set at that argument) [illegible] optimal dual vector [illegible] negative.

Page1:

$$g \in \mathbb{R}^n; \{\text{subgradient of } f:\mathbb{R}^n\to\mathbb{R} \text{ at } x\in\text{dom} f\} \stackrel{\text{def}}{=} \forall_{z\in\text{dom} f} \quad f(z) \ge f(x) + g^T(z-x)$$

interpretation:

$g \in \mathbb{R}^n$; {subgradient of $f:\mathbb{R}^n\to\mathbb{R}$ at $x\in\text{dom} f$} $\stackrel{\text{def}}{=}$ the affine function $f(x)+g^T(z-x) = \{z\}$ is a global underestimator of f, no matter what x is picked

$\stackrel{\text{def}}{=}$ $(g,-1)$ supports epi f at $(x, f(x))$ // proof: $\forall_{z\in\text{dom} f} \quad f(z) \ge f(x) + g^T(z-x)$

// $\leftrightarrow \forall_{z\in\text{dom} f} \forall_{t \ge f(z)} \quad t \ge f(z) \ge f(x) + g^T(z-x) \stackrel{\text{def}}{=} \forall_{(z,t)\in \text{epi} f} \quad t \ge f(x) + g^T(z-x)$

$\forall_{(z,t)\in\text{epi} f}$

$\stackrel{\text{def}}{=} \forall_{(z,t)\in\text{epi} f} \quad g^T(z-x) - (t - f(x)) \le 0 \stackrel{\text{def}}{=} \forall_{(z,t)\in\text{epi} f} \quad (g,-1)^T \left(\begin{bmatrix} z \\ t \end{bmatrix} - \begin{bmatrix} x \\ f(x) \end{bmatrix} \right) \le 0 \stackrel{\text{def}}{=} (g,-1)$ supports epi f at $(x, f(x)) \in \text{bd epi} f$ ☺

$\neq 0$

$(g,-1)^T \left(\begin{bmatrix} z-x \\ t-f(x) \end{bmatrix} \right) \le 0 \leftrightarrow (g,-1)^T \left(\begin{bmatrix} z \\ t \end{bmatrix} - \begin{bmatrix} x \\ f(x) \end{bmatrix} \right) \le 0$

Clearly $(x, f(x))$ is a boundary point of epi $f = \{(x,t) \mid x\in\text{dom} f, f(x) \le t\}$

[Subgradient, (g_x,-1) supports epi f at (x,f(x))]

Subgradient Contents:

// Some background:

// Supporting hyperplane theorem:

// $\forall_{C \in \{\mathcal{D}, \neg\emptyset\}} \forall_{x_0 \in \text{bd} C} \exists_{a\neq 0} \underbrace{\left(\forall_{x\in C} \quad a^T(x-x_0) \le 0 \right)}_{\updownarrow}$

(a supports C at x_0)

// epi $f = \{(x,t) \mid x\in\text{dom} f,\ f(x) \le t\}$

Subdifferential: $\partial f(x)$ {set of all subgradients of f at x}

$$\in [\,], \mathcal{D}\} \quad \because \partial f(x) = \bigcap_{z\in\text{dom} f} \{g \mid f(z) \ge f(x) + g^T(z-x)\}$$

Existence of subgradients Subgradient Contents:

Page3:

Existence of subgradients:

Alternative proof: Existence and boundedness of subdifferential

$(f \in \mathcal{D}\},\ x\in \text{int dom} f) \to (\partial f(x) \in \neg\emptyset, \le\})$ # key fact a convex function is locally upper bounded at $x_0 \in \text{int dom} f$

Proof: // Proof strategy: $P \to (Q \wedge R) \stackrel{\text{def}}{=} (P\to Q) \wedge (P \to R)$

// $P \to Q$:

Goal: $\partial f(x) \in \neg\emptyset\}$

Subgradient Contents:

Alternative proof: Existence and boundedness of subdifferential

* Alternative proof: $((f: [\,], \mathcal{D}), x_0 \in \text{int dom} f) \Rightarrow \partial f(x_0): \neg\emptyset, [\,]$

Proof: epi $f = \{(x,t): x\in\text{dom} f,\ t \ge f(x)\}$

bd epi $f = \{(x,t): x\in\text{dom} f,\ t = f(x)\}$

$$g_u(\lambda) = \min_{x\in\mathcal{D}} L_u(x,\lambda) = \min_{x\in\mathcal{D}} f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) - \lambda^T u$$

constant w.r.t. v constant w.r.t. v

$$= \min_{x\in\mathcal{D}} f_0(x) + \left(\min_{v\in\mathbb{R}^m} \lambda^T v + \sum_{i=1}^m \mathbb{1}_{v_i \ge f_i(x)}(v) \right) - \lambda^T u$$

$$= \min_{x\in\mathcal{D}} \left(\min_{v\in\mathbb{R}^m} f_0(x) + \lambda^T v + \sum_{i=1}^m \mathbb{1}_{v_i \ge f_i(x)}(v) - \lambda^T u \right)$$

$$= \min_{x\in\mathbb{R}^n} \min_{v\in\mathbb{R}^m} \left(f_0(x) + \lambda^T v - \lambda^T u + \sum_{i=1}^m \mathbb{1}_{v_i \ge f_i(x)}(v) + \mathbb{1}_{\mathcal{D}}(x) \right)$$

in this format x, v can be exchanged

$$= \min_{v\in\mathbb{R}^m} \left(\min_{x\in\mathbb{R}^n} f_0(x) + \lambda^T(v-u) + \sum_{i=1}^m \mathbb{1}_{v_i \ge f_i(x)}(v) + \mathbb{1}_{\mathcal{D}}(x) \right)$$

const w.r.t. x

$$= \min_{v\in\mathbb{R}^m} \left(\lambda^T(v-u) + \underbrace{\min_{x\in\mathbb{R}^n} f_0(x) + \sum_{i=1}^m \mathbb{1}_{v_i \ge f_i(x)}(v) + \mathbb{1}_{\mathcal{D}}(x)}_{\parallel} \right) = \min_{v\in\mathbb{R}^m} \left(p^*(v) + \lambda^T(v-u) \right)$$

$$\left(\forall_{x\in\mathcal{D}} f_0(x) ;\ \forall_{i\in\{1,\dots,m\}} f_i(x) \le v_i \right) = p^*(v)$$

$$p^*(u) = d^*(u) = g_u(\lambda_u^*) = \min_{v\in\mathbb{R}^m} \left(p^*(v) + \lambda_u^{*T}(v-u) \right)$$

$$\leftrightarrow \forall_{v\in\mathbb{R}^m} \quad p^*(u) \le p^*(v) + \lambda_u^{*T}(v-u)$$

$$\leftrightarrow \forall_{v\in\mathbb{R}^m} \quad p^*(v) \ge p^*(u) + (-\lambda_u^*)^T(v-u)$$

$$\leftrightarrow \quad (-\lambda_u^*) \in \partial p^*(u)$$

$$\min_{x\in\mathcal{D}.\ \mathcal{D}.\neg\emptyset} \left(\max_{\lambda \ge 0} L_u(x,\lambda) \right) : \mathcal{D}_u$$

$\therefore p^*(u): \mathcal{D}_u$

// P → Q:

Goal: $\partial f(x) \in \{\neg\emptyset\}$

$f \in \{D\} \leftrightarrow \text{epi} f = \{(z,t) \mid z \in \text{dom} f,\ t \geq f(z)\} \in \{D\}$ // epigraph of a convex function // is always convex set (this is if and only if)

Thoerem 3.1.2 Nesterov: A function f is convex <=> epi f is convex

Supporting hyperplane theorem [illegible]:

$\forall_{C \in \{D, \neg\emptyset\}} \ \forall_{x_0 \in \text{bd} C} \ \exists_{a \neq 0} \left(\forall_{x \in C} \ a^T(x - x_0) \leq 0 \right)$

$\forall_{(x, f(x)) \in \text{bd} C} \ \exists_{(a,b) \neq 0} \left(\forall_{(z,t) \in \text{epi} f} \ (a,b)^T \left(\begin{bmatrix} z \\ t \end{bmatrix} - \begin{bmatrix} x \\ f(x) \end{bmatrix} \right) \leq 0 \right)$ (Sup HypTh)

[illegible] by defn bd point of epi f will be of form $(x, f(x))$, $x \in \text{dom} f$

$(a,b)^T (z - x,\ t - f(x)) \leq 0$

$\Leftrightarrow a^T(z - x) + b\,(t - f(x)) \leq 0$

true for all $(z,t) \in \text{epi} f$, so this implies $b < 0$ // otherwise, say $b > 0$ or $b = 0$, let consider $b > 0$ first

// $\forall_{(z,t) \in \text{epi} f} \ a^T(z - x) + \underbrace{b}_{>0}(t - f(x)) \leq 0$

// we can construct a $\bar{z}$ such that if $a_i \leq 0$ then $\bar{z}_i < x_i \Leftrightarrow (\bar{z}_i - x_i) < 0 \Rightarrow a_i(\bar{z}_i - x_i) \geq 0$

// if $a_i \geq 0$ then $\bar{z}_i > x_i \Leftrightarrow (\bar{z}_i - x_i) > 0 \Rightarrow a_i(\bar{z}_i - x_i) \geq 0$ } thus for that $\bar{z}$: $a^T(\bar{z} - x) \geq 0$.

// no matter what $f(x), f(\bar{z})$ is as

// by defn $t \geq f(z)$ we can take an $\bar{t}$ such that $\bar{t} - f(x) > 0$ thus $\underbrace{a^T(\bar{z} - x)}_{\geq 0} + \underbrace{b}_{>0} \underbrace{(\bar{t} - f(x))}_{>0} > 0 \Rightarrow$ contradiction as ≤ 0

// $\therefore b \not> 0$

// let consider $b = 0 \to a \neq 0$ as $(a,b) \neq 0$

// lets set $z = x + \varepsilon a$ where $\varepsilon > 0$ then

// $a^T(x + \varepsilon a - x) \leq 0 \Leftrightarrow \varepsilon \|a\|^2 \leq 0 \Leftrightarrow a = 0 \neq 0$ contradiction

So far we have proven:

$\forall_{(x, f(x)) \in \text{bd epi} f} \ \exists_{(a,b) \neq 0} \left(\left(\forall_{(z,t) \in \text{epi} f} \ (a,b)^T \left(\begin{bmatrix} z \\ t \end{bmatrix} - \begin{bmatrix} x \\ f(x) \end{bmatrix} \right) \leq 0 \right) \Rightarrow b < 0 \right)$ [illegible]

$(z,t) \in \text{epi} f \Leftrightarrow z \in \text{dom} f \wedge t \geq f(z)$

[illegible] $(z,t) \in \text{epi} f$ [illegible] (SupHypThm) [illegible] $t = f(z)$ [illegible]: which implies:

$\forall_{(x, f(x)) \in \text{bd} C} \ \exists_{(a,b) \neq 0} \left(\forall_{z \in \text{dom} f} \ \underbrace{(a,b)^T \left(\begin{bmatrix} z \\ f(z) \end{bmatrix} - \begin{bmatrix} x \\ f(x) \end{bmatrix} \right) \leq 0}_{} \right)$

$a^T(z - x) + b(f(z) - f(x)) \leq 0$

$\to \frac{a^T}{b}(z - x) + f(z) - f(x) \geq 0$ [as $b < 0$, (-) flips] existential implication is existential conjunction

$\to f(z) \geq f(x) + \left(-\frac{a}{b}\right)^T (z - x)$ [$\because$ a vector, b scalar]

$\therefore \forall_{x \in \text{int dom} f} \ \exists_{(a,b) \neq 0,\ b < 0} \left(\forall_{z \in \text{dom} f} \ f(z) \geq f(x) + \left(-\frac{a}{b}\right)^T (z - x) \right)$

$\therefore \left(-\frac{a}{b}\right)$ is a subgradient at x

$\to \partial f(x)$ is nonempty

[illegible] $\partial f(x)$ is bounded $\leftrightarrow \sup_{g \in \partial f(x)} \|g\|_\infty \in \mathbb{R}$ ($x \in \text{int dom } f$) // Boyd has not proven it, so may be // some other time.

Subgradient Calculus (Boyd) Subgradient Contents:

Composition

$f(x) = h(f_1(x), \ldots, f_k(x))$

vector inequality

remember this is scalar inequality

Supporting hyperplane theorem: $(C : D, [\,]) \Rightarrow \forall_{x_0 \in \text{bd} C} \ \exists_{a \neq 0} \ \forall_{x \in C} \ a^T(x - x_0) \leq 0$

$(x_0, f(x_0)) \in \text{bd epi} f \Rightarrow \exists_{(d, -\alpha) \neq 0} \ \forall_{(x,t) \in \text{epi} f} \ \begin{bmatrix} d \\ -\alpha \end{bmatrix}^T \left(\begin{bmatrix} x \\ t \end{bmatrix} - \begin{bmatrix} x_0 \\ f(x_0) \end{bmatrix} \right) \leq 0$

$\Leftrightarrow \begin{bmatrix} d \\ -\alpha \end{bmatrix}^T \begin{bmatrix} x - x_0 \\ t - f(x_0) \end{bmatrix} = d^T(x - x_0) - \alpha(t - f(x_0)) \leq 0 \quad \forall_{(x,t) \in \text{epi} f}$

$\Leftrightarrow -\alpha t + d^T x \leq -\alpha f(x_0) + d^T x_0$

Also the direction of $(d, -\alpha)$ matters, so can be normalized: $\left\| \begin{matrix} d \\ -\alpha \end{matrix} \right\|_2 = 1 \Leftrightarrow \|d\|_2^2 + \alpha^2 = 1$

$\boxed{\alpha \geq 0}$

Per contradiction, $\alpha < 0$

now,

if $t \gg f(x_0)$, $(x,t) := (x_0, t) \Rightarrow (x_0, t) \in \text{epi} f$

$\therefore \underbrace{d^T(\underbrace{x - x_0}_{x_0})}_{0} - \underbrace{\alpha}_{<0} \underbrace{(t - f(x_0))}_{>0} \leq 0$, >0 contradiction

* A convex function is locally upper bounded in the interior of its domain:

$\exists_{\varepsilon > 0} \ \exists_{M > 0} \ \underbrace{B_2(x_0, \varepsilon)}_{\text{euclidian ball}} \subseteq \text{dom} f \ \wedge \ \forall_{x \in B_2(x_0, \varepsilon)} \ |f(x) - f(x_0)| \leq M \|x - x_0\|_2$

$\forall_{(x,t) \in \text{epi} f} \ d^T(x - x_0) \leq \alpha \left(t - f(x_0) \right)$

Set (x,t): $x \in B_2(x_0, \varepsilon) \subseteq \text{dom} f$, $t = f(x) \in \text{dom} f$

$d^T(x - x_0) \leq \alpha(f(x) - f(x_0)) \leq \alpha |f(x) - f(x_0)| \leq \alpha M \|x - x_0\|_2$

$x := x_0 + \varepsilon d$ # cool as: $x \in B_2(x_0, \varepsilon) \Leftrightarrow \|x - x_0\|_2 \leq \varepsilon$

$x := x_0 + \varepsilon d \Rightarrow \|x_0 + \varepsilon d - x_0\|_2 = \|\varepsilon d\|_2 = \varepsilon \|d\|_2 \leq \varepsilon$ [$\because \|d\|_2^2 + \alpha^2 = 1 \to \|d\|_2 \leq \sqrt{1 - \alpha^2} \leq 1$]

$d^T(\varepsilon d) \leq \alpha M \|\varepsilon d\|_2 = \alpha M \varepsilon \|d\|_2$

$\Leftrightarrow \cancel{\varepsilon} \|d\|_2^2 \leq M \alpha \cancel{\varepsilon} \|d\|_2$ # $\varepsilon > 0$

$\therefore \|d\|_2^2 \leq M \alpha \|d\|_2$

But $\|d\|_2^2 = 1 - \alpha^2$

$\to 1 - \alpha^2 \leq M \alpha \sqrt{1 - \alpha^2}$

$\Leftrightarrow M \alpha \sqrt{1 - \alpha^2} - (1 - \alpha^2) \geq 0$

$\Leftrightarrow \sqrt{1 - \alpha^2} \left(M\alpha - \sqrt{1 - \alpha^2} \right) \geq 0$

Composition

- $f(x) = h(f_1(x), \dots, f_k(x))$, with h convex nondecreasing, f_i convex
- find $q \in \partial h(f_1(x), \dots, f_k(x))$, $g_i \in \partial f_i(x)$
- then, $g = q_1 g_1 + \cdots + q_k g_k \in \partial f(x)$
- reduces to standard formula for differentiable h, f_i

proof:

$$\begin{aligned} f(y) &= h(f_1(y), \dots, f_k(y)) \\ &\geq h(f_1(x) + g_1^T(y-x), \dots, f_k(x) + g_k^T(y-x)) \\ &\geq h(f_1(x), \dots, f_k(x)) + q^T(g_1^T(y-x), \dots, g_k^T(y-x)) \\ &= f(x) + g^T(y-x) \end{aligned}$$

$f(x) = h(f_1(x), \dots, f_k(x))$ $\quad \{D, \uparrow\}$ $\{D\}$ $\quad f(x) \{\uparrow\}$ $\forall_{x,y \in \mathrm{dom} f}$ $(x \geq y) \leftrightarrow (f(x) \geq f(y))$

remember this is scalar inequality

If $q \in \partial h(f_1(x), \dots, f_k(x)) = \partial f(x)$, $\forall_{i \in \{1,\dots,k\}}$ $g_i \in \partial f_i(x)$,

one subgradient of subdifferential

then $g = \sum_{i=1}^{k} q_i g_i \in \partial f(x)$

$\forall_y$ $f(y) = h(f_1(y), \dots, f_k(y)) = h((f_i(y))_{i=1}^k)$

now: by definition of subgradient, $\forall_y$ $f_i(y) \geq f_i(x) + g_i^T(y-x)$ $\quad \begin{bmatrix} f_1(y) \\ \vdots \\ f_k(y) \end{bmatrix} \geq \begin{bmatrix} f_1(x) + g_1^T(y-x) \\ \vdots \\ f_k(x) + g_k^T(y-x) \end{bmatrix}$

because of the nondecreasing nature of h in its argument $(f_i(y))_{i=1}^k$; i.e. $(z \geq v) \leftrightarrow (h(z) \geq h(v))$

$f(y) = h\left(\begin{bmatrix} f_1(y) \\ \vdots \\ f_k(y) \end{bmatrix}\right) \geq h\left(\begin{bmatrix} f_1(x) + g_1^T(y-x) \\ \vdots \\ f_k(x) + g_k^T(y-x) \end{bmatrix}\right) = h\left(\begin{bmatrix} f_1(x) \\ \vdots \\ f_k(x) \end{bmatrix} + \begin{bmatrix} g_1^T(y-x) \\ \vdots \\ g_k^T(y-x) \end{bmatrix}\right)$

// now subgradient [illegible] definition [illegible]

// [illegible] $\forall_{v \in \mathrm{dom}\,[illegible]}$ $[illegible](v) \geq [illegible](z) + [illegible]^T(v - z) \leftrightarrow \forall_{z+r \in \mathrm{dom}\,[illegible]}$ $[illegible](z+r) \geq [illegible](z) + [illegible]^T [illegible](r)$, with $v = z + r$

// $\because q \in \partial h\left(\begin{bmatrix} f_1(x) \\ \vdots \\ f_k(x) \end{bmatrix}\right) \leftrightarrow \forall_{\begin{bmatrix} f_1(x) \\ \vdots \\ f_k(x) \end{bmatrix} + \begin{bmatrix} g_1^T(y-x) \\ \vdots \\ g_k^T(y-x) \end{bmatrix}}$ $\quad h\left(\begin{bmatrix} f_1(x) \\ \vdots \\ f_k(x) \end{bmatrix} + \begin{bmatrix} g_1^T(y-x) \\ \vdots \\ g_k^T(y-x) \end{bmatrix}\right) \geq h\left(\begin{bmatrix} f_1(x) \\ \vdots \\ f_k(x) \end{bmatrix}\right) + q^T \begin{bmatrix} g_1^T(y-x) \\ \vdots \\ g_k^T(y-x) \end{bmatrix}$

$\geq h\left(\begin{bmatrix} f_1(x) \\ \vdots \\ f_k(x) \end{bmatrix}\right) + q^T \begin{bmatrix} g_1^T(y-x) \\ \vdots \\ g_k^T(y-x) \end{bmatrix}$

$q^T \begin{bmatrix} g_1^T(y-x) \\ \vdots \\ g_k^T(y-x) \end{bmatrix} = [q_1 \dots q_k] \begin{bmatrix} -g_1^T- \\ \vdots \\ -g_k^T- \end{bmatrix} (y-x)$ // everything is cool // because of block multiplication

$= (q_1 g_1^T + \dots + q_k g_k^T)(y-x)$

$= (q_1 g_1 + \dots + q_k g_k)^T (y-x)$

$= g^T(y-x)$ // By definition, $g = q_1 g_1 + \dots + q_k g_k$

$= \underbrace{h\left(\begin{bmatrix} f_1(x) \\ \vdots \\ f_k(x) \end{bmatrix}\right)}_{f(x)} + g^T(y-x) = f(x) + g^T(y-x)$

$\therefore \forall_{y \in \mathrm{dom} f}$ $f(y) \geq f(x) + g^T(y-x) \leftrightarrow g \in \partial f(x)$ ☺ (proved)

$f = \sup_{\alpha \in A} \sup f_\alpha$ $\{A$ is closed and bound, $f_\alpha \{[illegible](x, \alpha)\}\}$

$\Rightarrow \quad \partial f(x) = \mathrm{cl}\ \mathrm{Co} \bigcup \{\partial f_\beta(x) \mid f_\beta(x) = f(x)\}$ usually equality holds

Pointwise supremum

if $f = \sup_{\alpha \in \mathcal{A}} f_\alpha$,

$$\mathbf{cl}\,\mathbf{Co} \bigcup \{\partial f_\beta(x) \mid f_\beta(x) = f(x)\} \subseteq \partial f(x)$$

(usually get equality, but requires some technical conditions to hold, *e.g.*, $\mathcal{A}$ compact, f_α cts in x and α)

roughly speaking, $\partial f(x)$ is closure of convex hull of union of subdifferentials of active functions

Weak rule for pointwise supremum

$f = \sup f$ // note that A can be any set

$y^T A(x) y = y^T Q \Lambda Q\ y$

$M\alpha\sqrt{1-\alpha^2} - (1-\alpha^2) \geq 0$

$\leftrightarrow \underbrace{\sqrt{1-\alpha^2}}_{\geq 0}\left(M\alpha - \sqrt{1-\alpha^2}\right) \geq 0$ # so $\sqrt{1-\alpha^2}$ [illegible] other part positive [illegible],

$\rightarrow M\alpha - \sqrt{1-\alpha^2} \geq 0$ # note this is not same # as division by $\sqrt{1-\alpha^2}$

$\leftrightarrow M\alpha \geq \sqrt{1-\alpha^2}$

$\leftrightarrow M^2\alpha^2 \geq 1 - \alpha^2 \leftrightarrow (1+M^2)\alpha^2 \geq 1 \leftrightarrow \alpha \geq \dfrac{1}{\sqrt{1+M^2}} > 0$ ($M > 0$)

$\forall_{(x,t) \in \mathrm{epi} f}$ $d^T(x - x_0) - \alpha(t - f(x_0)) \leq 0$

$(x,t) := (x, f(x))$:

$\forall_{x \in \mathrm{dom} f}$ $d^T(x-x_0) - \alpha(f(x) - f(x_0)) \leq 0$

$\leftrightarrow \left(\dfrac{d^T}{\alpha}\right)(x - x_0) - f(x) + f(x_0) \leq 0$

$\leftrightarrow f(x) \geq f(x_0) + \left(\dfrac{d^T}{\alpha}\right)(x - x_0)$

$\therefore \left(\dfrac{d^T}{\alpha}\right) \in \partial f(x_0)$ ($\in \mathrm{int\,dom} f$.) $\quad \therefore (\partial f(x_0) : x_0 \in \mathrm{int\,dom} f) = \neg\emptyset$

[illegible] $\partial f(x_0)$ bounded $\leftrightarrow \exists_{M>0} \forall_{g \in \partial f(x_0)}$ $\|g\|_2 \leq M$

$g \in \partial f(x_0)$, $g \neq 0$ $\leftrightarrow \forall_{x \in \mathrm{dom} f}$ $f(x) \geq f(x_0) + g^T(x - x_0)$

as $x_0 \in \mathrm{int\,dom} f$ so $\exists_{\varepsilon>0} \exists_{M>0} \forall_{x \in B_2(x_0,\varepsilon) \subseteq \mathrm{dom} f}$ $|f(x) - f(x_0)| \leq M\|x - x_0\|$ # locally upper bounded principle of D in int dom f

clearly $x := x_0 + \varepsilon \dfrac{g}{\|g\|_2}$ [illegible] $\|x - x_0\|_2 = \left\|\varepsilon \dfrac{g}{\|g\|_2}\right\| = \dfrac{\varepsilon}{\|g\|_2}\|g\|_2 = \varepsilon \rightarrow x \in B_2(x_0, \varepsilon)$

set: $x := x_0 + \varepsilon \dfrac{g}{\|g\|_2}$

$f(x) \geq f(x_0) + \underbrace{g^T\left(\varepsilon \dfrac{g}{\|g\|_2}\right)}_{\varepsilon \frac{\|g\|_2^2}{\|g\|_2} = \varepsilon\|g\|_2}$

$\leftrightarrow \varepsilon\|g\|_2 \leq f(x) - f(x_0) \leq |f(x) - f(x_0)| \leq M\|x - x_0\|$

$= M\varepsilon$

#$\varepsilon > 0$

$\rightarrow \cancel{\varepsilon}\|g\|_2 \leq M\cancel{\varepsilon}$

Weak rule for pointwise supremum

$$f = \sup_{\alpha \in \mathcal{A}} f_\alpha$$

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

example

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

where $A(x) = A_0 + x_1 A_1 + \cdots + x_n A_n$, $A_i \in \mathbf{S}^k$

- f is pointwise supremum of $g_y(x) = y^T A(x) y$ over $\|y\|_2 = 1$
- g_y is affine in x, with $\nabla g_y(x) = (y^T A_1 y, \ldots, y^T A_n y)$
- hence, $\partial f(x) \supseteq \mathbf{Co}\{\nabla g_y \mid A(x)y = \lambda_{\max}(A(x))y,\ \|y\|_2 = 1\}$ (in fact equality holds here)

to find **one** subgradient at x, can choose **any** unit eigenvector y associated with $\lambda_{\max}(A(x))$; then

$$(y^T A_1 y, \ldots, y^T A_n y) \in \partial f(x)$$

$f = \sup_{\alpha \in A} f_\alpha$ // note that A can be any set

algorithm for finding $g \in \partial f(x)$:

```
for β in A
    if ( f_β(x) == f(x) )
        find g in ∂f_β(x)
        return g   break
    end
end
```

(alternative calculation নিচে)

• **Example:**

$A(x_0) = A_0 + x_1 A_1 + \ldots + x_n A_n \in S^k$

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

$= \sup_{\alpha \in A = \{\alpha : \|\alpha\|_2 = 1\}} \{\underbrace{\alpha^T A(x) \alpha}_{f_\alpha(x)}\}$ // this is not a bilinear // program.

$= \sup_{\alpha \in A} f_\alpha(x)$

say, at x, $f(x) = f_\alpha(x) = \alpha^T A(x) \alpha$ which is continuous in α, x

then $g \in \partial f_\alpha(x) = \partial(\alpha^T (A_0 + \sum_{i=1}^n x_i A_i) \alpha)$ continuous in x

$\therefore g = \nabla_x f_\alpha(x) = \nabla_x \left(\mathrm{tr}(\underbrace{\alpha\alpha^T}_{z} (A_0 + \sum_{i=1}^n x_i A_i)) \right)$

$\mathrm{tr}(zA_0 + \sum_{i=1}^n x_i (zA_i))$

$= \mathrm{tr}(zA_0) + \mathrm{tr}(\sum_{i=1}^n x_i (zA_i))$ {$\because$ tr is a linear operator $\therefore \mathrm{tr}(\sum x_i A_i B_i) = \sum x_i \mathrm{tr}(A_i B_i)$}

$= \mathrm{tr}(zA_0) + \sum_{i=1}^n x_i \mathrm{tr}(zA_i) = \mathrm{tr}(zA_0) + x_1 \mathrm{tr}(zA_1) + \ldots + x_n \mathrm{tr}(zA_n)$

$\therefore g = \nabla_x f_\alpha(x) = \left(\frac{\partial}{\partial x_1} f_\alpha(x), \ldots, \frac{\partial}{\partial x_n} f_\alpha(x)\right) = (\mathrm{tr}(zA_1), \ldots, \mathrm{tr}(zA_n))$

$= (\mathrm{tr}(\alpha\alpha^T A_i))_{i=1}^n = (\alpha^T A_i \alpha)_{i=1}^n$

$= (\alpha^T A_1 \alpha, \alpha^T A_2 \alpha, \ldots, \alpha^T A_n \alpha) \in \partial f(x)$ ☺

$y^T A(x) y = y^T Q \Lambda Q\ y$

$= (Q^T y) \Lambda (Q^T y)$

$= \sum_{i=1}^k \lambda_i (Q^T y)_i^2$ $[\lambda_1 \geq \ldots \geq \lambda_k]$

$\leq \lambda_1 \underbrace{\sum_{i=1}^k (Q^T y)_i^2}_{\|Q^T y\|_2^2 = \|y\|_2^2}$ $[\because \forall_i\ \lambda_1 \geq \lambda_i$

$\rightarrow \forall_i\ \lambda_1 (Q^T y)_i^2 \geq \lambda_i (Q^T y)_i^2$

$\therefore \sum_{i=1}^k \lambda_1 (Q^T y)_i^2 \geq \sum_{i=1}^k \lambda_i (Q^T y)_i^2$

$\leftrightarrow \lambda_1 \sum_{i=1}^k (Q^T y)_i^2 \geq \sum_{i=1}^k \lambda_i (Q^T y)_i^2]$

$\rightarrow y^T A(x) y \leq \lambda_1 \|y\|_2^2$

$\rightarrow \max_{y \neq 0} y^T A(x) y \leq \lambda_1 \|y\|_2^2$

clearly if $A(x) y^* = \lambda_1 y^*$, i.e., the eigenvector corresponding to λ_1 will result in the maximum value, then $y^{*T} A(x) y^* = y^{*T}(\lambda_1 y^*) = \lambda_1 \|y^*\|_2^2$

$\rightarrow \max_{y \neq 0} y^T A(x) y = \lambda_1 \|y\|_2^2$ // equality [illegible], as we know y,

// λ_{max} এর corresponding eigenvector will be attaining the upper bound

we can keep our attention to normalized version of the eigenvector, i.e., $\|y^*\|_2 = 1$

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

Alternative and shorter calculation:

suppose $\sup_{\|y\|_2=1} y^T A(x) y$ is achieved at $y = \alpha$

then at x $f(x) = \sup_{\|y\|_2=1} y^T A(x) y = \alpha^T A(x) \alpha = f_\alpha(x)$

$g \overset{?}{\in} \partial f_\alpha(x)$

note: $f_\alpha(x) = y^T A(x) y = y^T (A_0 + \sum_{i=1}^n x_i A_i) y$

$= y^T A_0 y + \sum_{i=1}^n x_i (y^T A_i y)$ # this affine in x, so differentiable # in x

$= \underbrace{y^T A_0 y}_{d} + \underbrace{[y^T A_1 y \ \ldots \ y^T A_n y]}_{b^T} \underbrace{\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}}_{x} = d + b^T x$

$\therefore \nabla f_\alpha(x) = g_x = b = (y^T A_1 y, \ldots, y^T A_n y) \in \partial f(x)$ ☺

$\#\varepsilon > 0$

$\rightarrow \cancel{\varepsilon}\|g\|_2 \leq M \cancel{\varepsilon}$

$\rightarrow \|g\|_2 \leq M$

$\therefore \forall_{g: \in \partial f(x_0), \neq 0} \|g\|_2 \leq M$ $\# M > 0$

Obviously if $g: \in \partial f(x_0), = 0$ $\|g\|_2 = 0 < M$

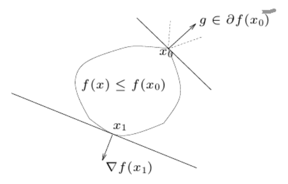
$\} \forall_{g \in \partial f(x_0)} \|g\|_2 \leq M \leftrightarrow \partial f(x_0)$ bounded.

☺

Subgradients and sublevel sets

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

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



- f differentiable at x_0: $\nabla f(x_0)$ is normal to the sublevel set $\{x \mid f(x) \leq f(x_0)\}$
- f nondifferentiable at x_0: subgradient defines a supporting hyperplane to sublevel set through x_0

$\nabla f(x_0) \perp \{x \mid f(x) \leq f(x_0)\}$

$\leftrightarrow \forall_{x \in \mathrm{dom} f} \quad \nabla f(x_0)^T x = 0$

Proof:

Given	Goal
$f \{D, \delta^2\}$	$\nabla f(x_0)^T x = 0$
$\forall_x$	
$x \in \{\bar{x} \in \mathrm{dom} f \mid f(\bar{x}) \leq f(\bar{x}_0)\}$	

$g \in \partial f(x) \leftrightarrow \forall_{y \in \mathrm{dom} f} \quad f(y) \geq f(x) + g^T(y-x)$

$\leftrightarrow g^T(y-x) \leq f(y) - f(x)$

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

So, when we are doing a minimization on $f(x)$: the set $\{y \mid g^T(y-x) \geq 0\}$ দ্বারা বর্ণিত যেই halfspace ওটাতে search করার কোন দরকার নাই.

$f(x) \leq f(x_0)$

$\because f\{D, \delta^2\} \leftrightarrow \forall_{x \in \mathrm{dom} f} \quad \underbrace{f(x) \geq f(x_0) + \nabla f(x_0)^T(x-x_0)}$

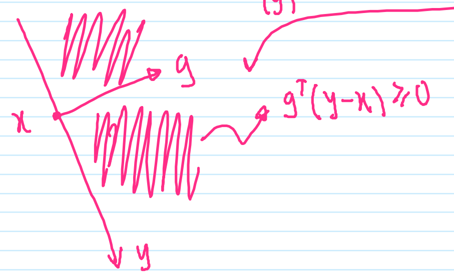
$f(x_0) + \nabla f(x_0)^T(x-x_0) \leq f(x) \leq f(x_0)$

$\rightarrow f(x_0) + \nabla f(x_0)^T(x-x_0) \leq f(x_0)$

$\rightarrow \nabla f(x_0)^T(x-x_0) \leq 0$

কারণ এই conditionটি থাকলে $g^T(y-x) \geq 0$ দ্বারা বর্ণিত hyperspace এর যে কোনো point (y) এর function value $f(y)$, $f(x)$ এর চেয়ে বড় হবে, i.e. for a minimization problem এ hyperspace এ search করার দরকার হবে না.

where x is the current point of the subgradient algorithm



* Page 21:

$f(x^*) = \inf_x f(x) \leftrightarrow 0 \in \partial f(x^*)$

$f(x^*) = \inf_x f(x) \leftrightarrow 0 \in \partial f(x^*)$

Proof:

$f(x^*) = \inf_x f(x)$

$\leftrightarrow \forall_{x \in \mathrm{dom} f} \quad f(x^*) \le f(x)$

$\leftrightarrow \forall_{x \in \mathrm{dom} f} \quad f(x) \ge f(x^*) + 0^T(x - x^*)$

$\leftrightarrow \quad 0 \in \partial f(x^*) \quad$ (proved)

Piecewise linear minimization   Subgradient Contents:

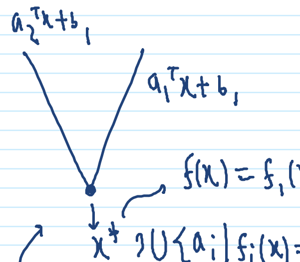
*Page23:

$f(x) = \max_{i \in \{1,\dots,m\}} (a_i^T x + b_i)$: Our goal is to find the optimality condition for the solution in terms of subgradient and show that it is infact equivalent to KKT condition

$x^* = \forall_{x \in \mathrm{dom} f} f(x) \leftrightarrow 0 \in \partial f(x^*)$

// Fact: $f = \max_{i=1,\dots,m} f_i(x) \rightarrow \partial f(x) = \mathbf{co} \bigcup \{\partial f_i(x) \mid f_i(x) = f(x)\}$

// $f(\cdot) = \max_{i=1,\dots,m} (a_i^T(\cdot) + b_i) \rightarrow \partial f(x) = \mathbf{co} \bigcup \{\partial f_i(x) \mid a_i^T x + b_i = f(x)\}$



$f(x) = f_1(x) = a_1^T x + b_1,\ f(x) = f_2(x) = a_2^T x + b_2$

$\bigcup \{a_i \mid f_i(x) = f(x)\} = \bigcup \{a_1 \mid f(x) = a_1^T x + b_1\} \cup \{a_2 \mid f(x) = a_2^T x + b_2\}$

$= \{a_1\} \cup \{a_2\}$

$0 \in \partial f(x^*) \leftrightarrow 0 \in \mathbf{co} \bigcup \{\partial f_i(x^*) \mid \underbrace{f_i(x^*) = a_i^T x^* + b_i}_{\text{continuous in } x,\ \partial f_i(x) = \{\nabla_x f_i(x) = \nabla_x (a_i^T x + b_i) = a_i\}}\} = \mathbf{co} \bigcup \{a_i \mid f_i(x^*) = a_i^T x^* + b_i\}$ //union U [illegible] !

say, at x^*, $f(x^*) = f_1(x^*) = \dots = f_{\bar m}(x^*)$ //we can always enumerate the functions // active at x^* by $1, \dots, \bar m$

then $\bigcup \{a_i \mid f_i(x^*) = a_i^T x^* + b_i\}$

$= \{a_1\} \cup \dots \cup \{a_{\bar m}\} = \{a_1, \dots, a_{\bar m}\}$

$\therefore \mathbf{co}[\{a_1, \dots, a_{\bar m}\}] = \{\sum_{i=1}^{\bar m} \bar\lambda_i a_i \mid \mathbf{1}^T \bar\lambda = 1, \bar\lambda \ge 0\}$

so: $0 \in \mathbf{co}\{\partial f_i(x^*) \mid f_i(x^*) = a_i^T x + b_i\} \leftrightarrow \exists_{\bar\lambda \ge 0, \mathbf{1}^T \bar\lambda = 1}\ 0 = \sum_{i=1}^{\bar m} \bar\lambda_i a_i \leftrightarrow \boxed{\exists_{\lambda \ge 0, \mathbf{1}^T \lambda = 1}\ 0 = \sum_{i=1}^{m} \lambda_i a_i}$ (1)

by 0 padding, $\lambda = (\bar\lambda, \bar 0)$, $\bar 0 = (0_{\bar m + 1}, \dots, 0_m)$

not possible as $f(x) = \max_{j=1,\dots,m} f_j(x) \not< f_1(x)$

$\forall_{i \in \{\bar m + 1, \dots, m\}} \left[f_i(x) \ne f(x) \leftrightarrow (f_i(x) > f(x) \lor f_i(x) < f(x)) \rightarrow f_i(x) < f(x) \right)$

$\boxed{\therefore a_i^T x + b_i < f(x) = \max_{i \in \{1,\dots,m\}} f_i(x) \Rightarrow \lambda_i = 0}$ (2)

and

Consider the epigraph form:

$\left(\begin{array}{l} \forall\ t \\ \rceil \\ \left(\begin{array}{l} t \ge \max_{i \in \{1,\dots,m\}} (a_i^T x + b_i) \\ \leftrightarrow (\forall_{i \in \{1,\dots,m\}}\ t \ge a_i^T x + b_i) \\ \leftrightarrow t\bar{\mathbf{1}} \ge \underbrace{\begin{bmatrix} -a_1^T- \\ \vdots \\ -a_m^T- \end{bmatrix}}_{A} x + \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} = Ax + b \leftrightarrow [A\ -\bar{\mathbf{1}}] \begin{bmatrix} x \\ t \end{bmatrix} + b \le 0 \end{array} \right) \end{array} \right)$

$= \left(\begin{array}{l} \forall\ (\bar 0, 1)^T (x, t) \\ \rceil \\ [A\ -\bar{\mathbf{1}}] \begin{bmatrix} x \\ t \end{bmatrix} + b \le 0 \end{array} \right)$

$L((x,t), \underset{\ge 0}{\lambda}) = (\bar 0, 1)^T (x, t) + \lambda^T \left([A\ -\bar{\mathbf{1}}] \begin{bmatrix} x \\ t \end{bmatrix} + b \right)$

$= (\bar 0, 1)^T (x, t) + \lambda^T [A\ -\bar{\mathbf{1}}] (x, t) + \lambda^T b$

$= \left((\bar 0, 1) + [A\ -\bar{\mathbf{1}}]^T \lambda \right)^T (x, t) + \lambda^T b$

$g(\lambda) = \inf_{(x,t)} L((x,t), \lambda) = \begin{cases} \lambda^T b, & \text{if } \begin{bmatrix} \bar 0 \\ 1 \end{bmatrix} + \begin{bmatrix} A^T \\ -\bar{\mathbf{1}}^T \end{bmatrix} \lambda = 0,\ \lambda \ge 0 \\ -\infty & \text{else} \end{cases}$

dual problem

so (1) and (2) are implications of $0 \in \partial f(\tilde{x})$

same as dual feasibility + complementary conditions

dual problem

$$\text{maximize } b^T\lambda$$

subject to $\left(\begin{bmatrix} \bar{0} \\ 1 \end{bmatrix} + \begin{bmatrix} A^T\lambda \\ \underbrace{-\bar{1}^T\lambda}_{\text{scalar}} \end{bmatrix} = 0 \right) \leftrightarrow \left(\begin{array}{l} A^T\lambda = 0 \leftrightarrow \sum_{i=1}^m \underbrace{(A^T)_i}_{\text{row}_i(A) = a_i} \lambda_i = \sum_{i=1}^m a_i \lambda_i = 0 \\ \bar{1}^T\lambda = 1 \end{array} \right)$

vector; scalar

$\lambda \succeq 0$

$$= \left(\begin{array}{l} \max_{\lambda} \; b^T\lambda \\ \text{s.t.} \; \boxed{\begin{array}{l} \sum_{i=1}^m a_i\lambda_i = 0 \\ \bar{1}^T\lambda = 1 \\ \lambda \succeq 0 \end{array}} \end{array} \right)$$

so optimality condition for subgradient is equivalent to dual feasibility

* Complementary slackness:

each component of innerproduct

$$\lambda^T \left(\begin{bmatrix} A & -\bar{1} \end{bmatrix} \begin{bmatrix} x \\ t \end{bmatrix} + b \right) \text{ will be } 0$$

$$= \lambda^T (Ax - t\bar{1} + b)$$

$$= \sum_{i=1}^m \lambda_i (a_i^T x + b_i - t)$$

$$\forall_{i \in \{1,\dots,m\}} \quad \lambda_i (a_i^T x + b - t) = 0$$

$$\Rightarrow \boxed{\underbrace{t = \max_{i \in \{1,\dots,m\}} f(x)}_{\substack{\text{because at optimal} \\ \text{solution, } t = \max_{i \in \{1,\dots,m\}} f(x)}} > (a_i^T x + b) \Rightarrow \lambda_i = 0}$$

Optimality Conditions - Constrained   Subgradient Contents:

$$L(x,\lambda,\nu) = f(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \nu_i h_i(x)$$

Page 25: Consider a differentiable optimization problem: $\left(\nabla_0 f(x) \;\; \forall_{i \in \{1,\dots,m\}} f_i(x) \le 0, \; \forall_{i \in \{1,\dots,p\}} h_i(x) = 0 \right)$ then the KKT conditions are:

Primal Feasibility:

$\forall_{i \in \{1,\dots,m\}} \quad f_i(x) \le 0$

$\forall_{i \in \{1,\dots,p\}} \quad h_i(x) = 0$

Dual feasibility:

$\lambda \succeq 0$

Vanishing gradient at Lagrangian:

$\nabla f(x) + \sum^{m} \dots \sum^{p} \dots = 0$

extension to nondifferentiable case

Primal Feasibility:

$\forall_{i \in \{1,\dots,m\}} \quad f_i(x) \le 0$

$\forall_{i \in \{1,\dots,p\}} \quad h_i(x) = 0$

Dual feasibility:

$\lambda \succeq 0$

Vanishing subgradient at Lagrangian:

$\sum^{m} \dots \sum^{p}$

case

Vanishing gradient at Lagrangian:

$$\nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x) + \sum_{i=1}^{p} \nu_i \nabla h_i(x) = 0$$

Complementary slackness: (Each term in the λ sum in the Lagrangian, $\sum_{i=1}^{p} \lambda_i f_i(x)$ is zero

$$\forall_{i \in \{1,\dots,m\}} \quad \lambda_i f_i(x) = 0$$

Vanishing subgradient at Lagrangian:

$$\partial f(x) + \sum_{i=1}^{m} \lambda_i \partial f_i(x) + \sum_{i=1}^{p} \nu_i \; \partial h_i(x) \ni 0$$

Complementary slackness: (Each term in the λ sum in the Lagrangian, $\sum_{i=1}^{p} \lambda_i f_i(x)$ is zero

$$\forall_{i \in \{1,\dots,m\}} \quad \lambda_i f_i(x) = 0$$

KKT condition for nondifferentiable problems

Directional derivative [Subgradient Contents:]

Directional Derivative: // this definition has issues, for a nondifferentiable function

$$f'(x;\delta x) \triangleq \lim_{h \downarrow 0} \frac{f(x+h\delta x) - f(x)}{h} \in [-\infty, \infty]$$

at the point of nondifferentiability
the limit of directional derivative does not exist.

* $f \notin \{D$: finite near $x\} \Rightarrow \exists \; f'(x;\delta x)$

* $f \in \{\underbrace{\mathcal{D}_x^1}_{\text{differentiable}}\} \Leftrightarrow \exists_{g = \nabla f(x)} \; \forall_{\delta x} \quad f'(x;\delta x) = g^T \delta x$ // relation between $\overbrace{f'(x;\delta x)}^{\text{directional derivative}}$ and g → subgradient

We can extend this definition for nondifferentiable function:

$$\{f \notin \mathcal{D}\} \quad f'(x;\delta x) = \sup_{g \in \partial f(x)} g^T \delta x$$

* Why directional derivatives are important:

Say:

$$f'(x;\delta x) \triangleq \lim_{h \downarrow 0} \frac{f(x+h\delta x) - f(x)}{h} < 0$$

$$\rightarrow \forall_{h: h \downarrow 0, h \in \mathbb{R}_{++}} \left(\frac{f(x+h\delta x) - f(x)}{h} < 0 \right)$$

$$\rightarrow \forall_{h: h \downarrow 0, h \in \mathbb{R}_{++}} \left(f(x + \underbrace{h}_{\text{scalar}} \underbrace{\delta x}_{\text{search direction}}) < f(x) \right)$$

: equivalent to saying δx is a cost reducing direction!

$\therefore$ negative directional directive provides a cost reducing direction!

alternative importance: if

$$f'(x, \underbrace{\delta x}_{(y-x)}) = \sup_{g \in \partial f(x)} g^T \delta x \leq 0$$

$$\rightarrow \forall_{g \in \partial f(x)} \; g^T \delta x \leq 0 \quad ; \quad \delta x = y - x$$

0 → $\sup g^T \delta x$

$f(y) \geq f(x) + g^T \delta x \rightarrow f(y) - f(x) \geq g^T \delta x$: does not lead to anything good

Negative subgradient is a descent direction for distance to optimal point:

(say $g \in \partial f(x)$ and let $z^* = \operatorname{argmin}_{y \in \operatorname{dom} f} f(y)$) $\Rightarrow \|x^+ - z^*\|_2 < \|x - z^*\|$

$x^+ = x + \underline{t}(-g)$

যদি negative subgradient কে search direction হিসেবে select তাহলে resultant new point $x^+ = x + t(-g)$ will become strictly closer to the optimal point.

Proof:

$$\|x^+ - z^*\|_2^2 = \|x - tg - z^*\|_2^2 = (x - tg - z^*)^T (x - tg - z^*)$$

$$= \left((x - z^*)^T - tg^T\right)\left((x - z^*) - tg\right) = \|x - z^*\|_2^2 + t^2 \|g\|_2^2 + 2t(-g^T(x - z^*))$$

$$\leq \|x - z^*\|_2^2 + \underbrace{t^2\|g\|_2^2}_{\text{small}^2 \text{ positive number}} + \underbrace{2t\,(f(z^*) - f(x))}_{\text{small negative number}}$$

$$< \|x - z^*\|_2^2 \quad \text{[By suitable choice of } t\text{]}$$

$$\therefore \|x^+ - z\|_2 < \|x - z\|_2$$

$$g \in \partial f(x) \leftrightarrow \forall_{\bar{y} \in \mathrm{dom} f} \; f(\bar{y}) \geq f(x) + g^T(\bar{y} - x)$$

$$\therefore \bar{y} = z^* \Rightarrow \quad f(z^*) \geq f(x) + g^T(z^* - x) \;\wedge\; f(x) \geq f(z^*)$$

$$\Rightarrow g^T(z^* - x) \leq f(z^*) - f(x) \leq 0$$

$$\Rightarrow -g^T(x - z^*) \leq \underbrace{f(z^*) - f(x)}_{\leq 0}$$

* Sugradient of the dual function [Subgradient Contents:]

Subgradients of the dual function:

[Calafiore 286]

$$p^* = \begin{pmatrix} \gamma_x & f_0(x) \\ \text{s.t.} & \forall_{i \in \{1,\ldots,m\}} \; f_i(x) \leq 0 \end{pmatrix} = \begin{pmatrix} \gamma & f_0(x) + I_D(x) \\ \text{s.t.} & \forall_{i \in \{1,\ldots,m\}} \; f_i(x) \leq 0 \end{pmatrix}$$

$$L(x, \lambda) = f_0(x) + I_D(x) + \sum_{i=1}^m \lambda_i f_i(x)$$

$$g(\lambda) = \inf_x \left[f_0(x) + I_D(x) + \sum_{i=1}^m \lambda_i f_i(x) \right]$$

$$= \inf_{x \in D} \left[f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) \right] = f_0(x_\lambda) + \sum_{i=1}^m \lambda_i f_i(x_\lambda)$$

$$\# \quad x^*_\lambda = \underset{x \in D}{\mathrm{argmin}} \left[f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) \right]$$

Now:

$$g(\bar{\lambda}) = \gamma_{x \in D} \left[f_0(x) + \sum_{i=1}^m \bar{\lambda}_i f_i(x) \right] \quad \text{\# let's } x^*_{\bar{\lambda}} \text{ be the argmin]}$$

$$= f_0(x^*_{\bar{\lambda}}) + \sum_{i=1}^m \bar{\lambda}_i f_i(x^*_{\bar{\lambda}})$$

$$\leq f_0(x^*_\lambda) + \sum_{i=1}^m \bar{\lambda}_i f_i(x^*_\lambda)$$

[[illegible] x^*_λ is argmin of this, but not necessarily for $\{\bar{\lambda}\}$]

$$= \underbrace{f_0(x^*_\lambda) + \sum_{i=1}^m \lambda_i f_i(x^*_\lambda)}_{g(\lambda)} + \sum_{i=1}^m \bar{\lambda}_i f_i(x^*_\lambda) - \sum_{i=1}^m \lambda_i f_i(x^*_\lambda)$$

$$= g(\lambda) + \sum_{i=1}^m (\bar{\lambda}_i - \lambda_i)\, f_i(x^*_\lambda)$$

$$\sum_{i=1}^m f_i(x^*_\lambda)(\bar{\lambda}_i - \lambda_i) = \underbrace{(f_1(x^*_\lambda), f_2(x^*_\lambda), \ldots, f_m(x^*_\lambda))}_{F(x_\lambda)}^T \left(\underbrace{(\bar{\lambda}_1, \ldots, \bar{\lambda}_m)}_{\bar{\lambda}} - \underbrace{(\lambda_1, \ldots, \lambda_m)}_{\lambda} \right) = F(x^*_\lambda)^T(\bar{\lambda} - \lambda)$$

$$= g(\lambda) + F(x^*_\lambda)^T(\bar{\lambda} - \lambda)$$

$$\therefore \forall \; g(\bar{\lambda}) \leq g(\lambda) + F(x^*_\lambda)^T(\bar{\lambda} - \lambda)$$

$= g(\lambda) + F(x_\lambda^*)^T(\bar{\lambda}-\lambda)$ $\qquad F(x_\lambda)$ $\qquad \bar{\lambda}$ $\qquad \lambda$

$\therefore \forall_{\bar{\lambda}} \; g(\bar{\lambda}) \leq g(\lambda) + F(x_\lambda^*)^T(\bar{\lambda}-\lambda)$

now remember $g(\lambda)$ is concave in λ by structure, so for definition of subgradient we need convex $-g(\lambda)$

$\forall_{\bar{\lambda}} \; (-g(\bar{\lambda})) \leq (-g(\lambda)) + (-F(x_\lambda^*)^T)(\bar{\lambda}-\lambda)$

so, $(-F(x_\lambda^*)) = (-f_i(x_\lambda^*))_{i=1}^m$ is subgradient of the negative dwell function $-\underbrace{g(\lambda)}$ at λ

$x_\lambda^* = \underset{x \in D}{\text{argmin}} \left(f_0(x) + \sum \lambda_i f_i(x)\right)$ $\qquad g(\lambda) = f_0(x_\lambda) + \sum_{i=1}^m \lambda_i f_i(x)$ # $g(\lambda)$ evaluate করার জন্য x_λ এমনিতেই বের করতে হবে. while we minimize the Lagrangian $L(x,\lambda)$

so there is no additional cost,

$\therefore (-f_i(x_\lambda^*))_{i=1}^m \in \partial(-\underbrace{g(\lambda)})$ [eq: subgradient for negative dual function] ret: [Projected Subgradient for dual problem]

dual function of the primal problem : $\begin{pmatrix} \forall \; f_0(x) \\ \neg \\ \forall_{i \in \{1,\dots,m\}} \; f_i(x) \leq 0 \\ x \in D \end{pmatrix} = \begin{pmatrix} \forall \; f_0(x) + I_D(x) \\ \neg \\ \forall_{i \in \{1,\dots,m\}} \; f_i(x) \leq 0 \end{pmatrix}$

$x_\lambda^* = \underset{x}{\text{argmin}} \; \underbrace{L(x,\lambda)}_{= f_0(x) + \sum_{i=1}^m \lambda_i f_i(x)}$