Nonexpansive and Contraction Mapping

1:24 PM

Lipschitz constant on a relation means that relation is a function

L {Lipschitz constant on relation F} $\overset{\text{def}}{=}$ $\forall_{xFu}\ \forall_{yFv}\ \|u-v\|_2 \le L\|x-y\|_2$ — $L<1$ হলে F is contraction relation; $L=1$ হলে F is nonexpansive

$\Rightarrow (x=y \Rightarrow \|u-v\|_2 \le L\,0 = 0 \Rightarrow u=v)$

$\leftrightarrow$ same input argument results in a single value, so F is a function

∴ As a result, nonexpansive and contraction mapping they are function too

* Basic properties of Lipschitz function

$\left(F_{\{\text{Lipschitz constant } L\}}, \tilde F_{\{\text{Lipschitz constant } \tilde L\}}\right) \Rightarrow L\tilde L$ {Lipschitz constant of $F\tilde F = F\circ\tilde F$}

so if F and $\tilde F$ are nonexpansive (contraction) $\Leftrightarrow L=1, \tilde L=1$ $(L<1, \tilde L<1)$
then $F\tilde F = F\circ\tilde F$ is also nonexpansive as $L\tilde L=1$ $(L\tilde L<1)$

Lipschitz mapping, (e.g., nonexpansive and contraction mapping) is a function

composition of nonexpansive (contraction) mapping is nonexpansive (contraction) mapping

- $\underbrace{(\ldots)}_{(A)}$ $A \to \alpha F + \tilde\alpha \tilde F$ has Lipschitz constant $|\alpha| L + |\tilde\alpha|\tilde L$, $\alpha \in \mathbb{R}$, $\tilde\alpha \in \mathbb{R}$

// so, $\theta F + (1-\theta)\tilde F$ has Lipschitz constant $|\theta| L + |1-\theta|\tilde L = \theta L + (1-\theta)\tilde L = 1$, so $\theta F + (1-\theta) F$ is nonexpansive (F: nonexpansive, $\tilde F$: nonexpansive, $\theta \in [0,1]$; $L = 1$, $\tilde L = 1$)

So convex combination of two nonexpansive mapping is nonexpansive.

// $\theta \in (0,1)$ হলে and one of F, or $\tilde F$ contraction হলে $\theta L + (1-\theta)\tilde L < 1$ [as we will get 1 only both $L=1, \tilde L = 1$] ($\tilde L \le 1$; <1, <1, $\ll 1$)

$\Downarrow$

$\theta F + (1-\theta) F$ is a contraction mapping then!

So strict convex combination of one expansive mapping and one contraction mapping is contraction mapping

From proximal algorithm we already know that fixed point of operators are important, recall minimizer of f(x) is the fixed point of the proximal operator, i.e., $x^* = \text{prox}_f(x^*)$

- Fixed point set of nonexpansive operator $F_{\{\text{dom } F = \mathbb{R}^n\}}$, $\{x \in \text{dom } F \mid x = F(x)\} = (I-F)^{-1}(0)$ {☐}

So fixed point set of a nonexpansive operator is essentially 0 passed through the resolvent of that operator
Note the scaling factor of the operator is -1

\# $x = F(x)$
$\overset{\text{def}}{=} Ix - F(x) = 0 \overset{\text{def}}{=} (I☐ - F☐)(x) = 0 \overset{\text{def}}{=} x\,(I☐ - F☐)\,0 \overset{\text{def}}{=} (x,0) \in (I☐ - F☐)$
$\overset{\text{def}}{=} (0,x) \in (I☐ - F☐)^{-1} \overset{\text{def}}{=} x \in (I☐ - F☐)^{-1}(0) = (I-F)^{-1}(0)$ #

- Fixed point set of contraction mapping $F_{\{\text{dom } F = \mathbb{R}^n\}}$ is singleton

- Example: 1) Affine function

$F(x) = Ax+b$ has Lipschitz constant $L = \|A\|_2 = \sigma_{\max}(A)$

For any affine function the maximum singular value of the associated matrix is the Lipschitz constant of that function!

$\forall_{x,y,\ x\ne y}\ \|F(x)-F(y)\|_2 = \|Ax+b-Ay-b\|_2 = \|A(x-y)\|_2 \overset{?}{\le} L\|x-y\|_2$

$\Leftrightarrow \forall_{x,y,\ x\ne y}\ \dfrac{\|A(x-y)\|_2}{\|x-y\|_2} \overset{?}{\le} L$

$\Leftrightarrow \underbrace{\max_{(x-y)\ne 0} \dfrac{\|A(x-y)\|_2}{\|x-y\|_2}}\overset{?}{\le} L$

$\max_{z\ne 0} \dfrac{\|Az\|_2}{\|z\|_2} = \sigma_{\max}(A) = \sqrt{\lambda_{\max}(AA^T)}$

So, L can be taken as: $\sigma_{\max}(A)$, if this is <1 $(=1)$ then we can have contraction (nonexpansive) mapping.

2) Differentiable function:

A differentiable function is Lipschitz iif the Jacobian norm of the function is globally bounded by the associated Lipschitz constant.

$F_{\{:\mathbb{R}^n \to \mathbb{R}^n,\ \mathcal{D}\}}$ {Lipschitz with parameter L} $\Leftrightarrow \forall_x\ \|DF(x)\|_2 \le L$, $\quad (Df(x))_{ij} = \dfrac{\partial f_i(x)}{\partial x_j}$

Proof: $(\leftarrow)$

$\forall_x\ \|DF(x)\|_2 \le L$

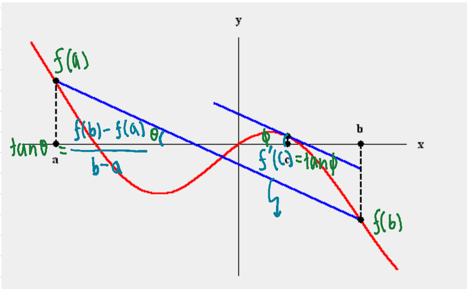
define $g(t) = (F(x)-F(y))^T F(\underbrace{tx + (1-t)y}_{\text{affine combination}})$, $\quad t(x-y)+y$

this is a continuous function in t

so by mean value theorem $\exists_{\xi \in (0,1)}\ g'(\xi) = \dfrac{g(1)-g(0)}{1-0}$

% So for a differentiable and continuous function in some interval, there exists a intermediate point where the slope of the function is equal to the relative increase of the function over the interval

{ mean value theorem: $f(x)_{\{\text{cont}_{[a,b]},\ \mathcal{D}_{(a,b)}\}} \Rightarrow \exists_{c\in(a,b)}\ f'(c) = \dfrac{f(b)-f(a)}{b-a}$



again $g'(t) = \dfrac{d}{dt}(F(x)-F(y))^T F(tx+(1-t)y)$

$= (F(x)-F(y))^T \dfrac{d}{d(t(x-y)-y)}(F(t(x-y)+y))\ \dfrac{d(t(x-y)+y)}{dt}$ // Chain rule $D\,g(f(x)) = D_{f(x)}\,g(f(x))\ D_x f(x)$

$= (F(x)-F(y))^T\ D(F(t(x-y)-y)\ (x-y)$

∴ $g'(\xi) = (F(x)-F(y))^T\ D(F(\xi(x-y)-y)\ (x-y)$

$g(1) = (F(x)-F(y))^T F(x)$
$g(0) = (F(x)-F(y))^T F(y)$
$g(1)-g(0) = (F(x)-F(y))^T (F(x)-F(y))$
$= \|F(x)-F(y)\|_2^2$

$\Leftrightarrow g'(\xi) = \underbrace{(F(x)-F(y))^T}_{a}\ \underbrace{D(F(\xi(x-y)-y)\ (x-y)}_{b} \le |a^Tb| \le \|a\|_2\|b\|_2 = \|F(x)-F(y)\|_2\ \|\underbrace{DF(\xi x+(1-\xi)y)}_{\text{matrix}}\ \underbrace{(x-y)}_{\text{vector}}\|_2$

$\le \|F(x)-F(y)\|_2\ \underbrace{\|DF(\xi(x)+(1-\xi)y)\|_2}_{\le L}\ \|x-y\|_2$

$\le L\ \|F(x)-F(y)\|_2\ \|x-y\|_2$

$\boxed{\text{\# By Cauchy-Schwartz inequality: } |a^Tb| \le \|a\|\,\|b\|}$

// now $\|A\|_p = \max_{u\ne 0} \dfrac{\|Au\|_p}{\|u\|_p}$ (by defn) $\ge \dfrac{\|Au\|_p}{\|u\|_p}$ $\forall_{u\ne 0}$, $p \in \mathbb{Z}_+$ $\quad$ ∴ $\forall_u\ \|Au\|_p \le \|A\|_p\|u\|_p$

যদিও $u=0$ হয় তবে inequality টি trivially true

∴ $\|F(x)-F(y)\|_2^2 \le L\,\|F(x)-F(y)\|_2\ \|x-y\|_2$

$\rightarrow \|F(x)-F(y)\|_2 \le L\|x-y\|_2$ ☺

$(\rightarrow)$

F has a Lipschitz parameter L,

Want to prove $\|DF(x)\|_2 \overset{?}{\le} L$

$\Leftrightarrow \max_{v\ne 0} \dfrac{\|DF(x)v\|_2}{\|v\|_2} \overset{?}{\le} L$

$\Leftrightarrow \forall_{v\ne 0} \dfrac{\|DF(x)v\|_2}{\|v\|_2} \overset{?}{\le} L$

$\Leftrightarrow \forall_v\ \|DF(x)v\|_2 \overset{?}{\le} L\|v\|_2$ [∵ $v=0$ is the equality $0=0$ case]

Now: $\forall_v\ \|DF(x)v\|_2 = \lim_{h\to 0} \dfrac{1}{h}\ \underbrace{\|F(x+hv)-F(x)\|_2}$

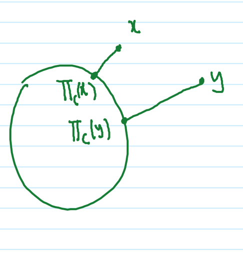
$\le L\|x+hv-x\|_2 = L\|hv\|_2 = L|h|\,\|v\|_2 \le L\|v\|_2$ ☺

[illegible]

* Projections: We will prove, $\Pi_C(☐)$ is a non-expansive operator $\Leftrightarrow \dfrac{\|\Pi_C(x)-\Pi_C(y)\|}{\|x-y\|} \le 1$

{C {[ ], D, ≠ ∅}}

% If there are two lovers in real world and one projection set symbolizing the fantasy world, then the distance between the lovers will be smaller in the fantasy world than that of the real world.



$\Pi_C(x) = \underset{z\in C}{\text{argmin}}\ \frac{1}{2}\|z-x\|_2^2$ // always exists and unique $\quad$ objective strongly convex bole

point in C closest to x

\# Optimality criterion for $\left(\forall_{y\in Y} f_0(y)\right)$ is $\forall_{y\in Y}\ \nabla f_0(y^*)^T(y-y^*) \ge 0$ $\quad$ {f_0} amount moved

Optimality criterion for $\begin{pmatrix} \min & f_0(y) \\ \text{s.t.} & y \in Y \end{pmatrix}$ is $\forall_{y \in Y}\ \underbrace{\nabla f_0(y^*)^T}_{\text{per unit cost change}} \underbrace{(y - y^*)}_{\text{amount moved}} \geq 0$ — total cost change as one moves from y^* to any y is positive if y^* is optimal

This is the first order optimality condition.

In our case the problem is:

$$\begin{pmatrix} \min & \overbrace{\tfrac{1}{2}\|z-x\|_2^2}^{f_0(z)} \\ \text{s.t.} & \\ & z \in C \end{pmatrix}$$

$$\therefore \nabla f_0(z) = (z - x)$$

$$\nabla f_0(z^* = \Pi_C(x)) = (\Pi_C(x) - x)$$

So, the optimality condition is $\forall_{z \in C}\ \underbrace{\nabla f_0(z^*)^T}_{(\Pi_C(x)-x)} \underbrace{(z - z^*)}_{(z-\Pi_C(x))} = (\Pi_C(x) - x)^T (z - \Pi_C(x)) \geq 0$

$z := \Pi_C(y)$

similarly when we are taking projection of y:

$$\forall_{z \in C}\ (\Pi_C(y) - y)^T (z - \Pi_C(y)) \geq 0$$

$z := \Pi_C(x)$

$$(\Pi_C(x) - x)^T (\Pi_C(y) - \Pi_C(x)) = \cancel{(-1)(-1)}^{\,1}\ (x - \Pi_C(x))^T (\Pi_C(x) - \Pi_C(y)) \geq 0$$

$$(\Pi_C(y) - y)^T (\Pi_C(x) - \Pi_C(y)) \geq 0$$

$\oplus$

$$(\Pi_C(y) - y)^T (\Pi_C(x) - \Pi_C(y)) + (x - \Pi_C(x))^T (\Pi_C(x) - \Pi_C(y)) \geq 0$$

$$= ((\Pi_C(y) - y + x - \Pi_C(x))^T (\Pi_C(x) - \Pi_C(y))$$

$$= ((x-y)^T - (\Pi_C(x) - \Pi_C(y))^T)(\Pi_C(x) - \Pi_C(y))$$

$$= (x-y)^T (\Pi_C(x) - \Pi_C(y)) - \|\Pi_C(x) - \Pi_C(y)\|_2^2 \geq 0$$

$$\rightarrow \quad \|\Pi_C(x) - \Pi_C(y)\|_2^2 \leq (\Pi_C(x) - \Pi_C(y))^T (x - y) \leq \|\Pi_C(x) - \Pi_C(y)\|_2 \|x - y\|_2$$

By Cauchy-Schwartz $a^T b \leq \|a\| \|b\|$

$$\therefore \|\Pi_C(x) - \Pi_C(y)\|_2 \leq \underset{L}{1} \|x - y\|_2$$

So $\Pi_C(\square)$ is a nonexpansive operator

Similarly, overprojection operator $R_C = 2\Pi_C - I$ on $C \in \{[\,], D, \neg D\}$ is nonexpansive.

Caution: nonexpansive and contraction mapping are function, but monotone or strongly monotone operator can be nontrivial relations.

So monotone operator always split out vector of same dimension as the input argument.

Intuition of monotone operator: let us start with scalar function:

For that case monotone function is if input argument increases, the function value increases, and vice versa one way of quantifying that is :

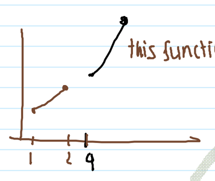
$$(f(x) - f(y))(x - y) \geq 0 \leftrightarrow \begin{pmatrix} (x-y) \geq 0 \rightarrow (f(x) - f(y)) \geq 0 \\ (x-y) \leq 0 \rightarrow f(x) - f(y) \leq 0 \end{pmatrix}$$

Now for imposing that as a definition we want this to be true for all (x,y) pair, so we can give the definition :

$$\forall_x \left(x \in \operatorname{dom} f \leftrightarrow \forall_{y \in \operatorname{dom} f} (f(x) - f(y))(x - y) \geq 0\right)$$

(→) direction টা বুঝলাম, (←) direction এর interpretation কী? Note that (←) এর meaning হলো: $\forall_x \left(\forall_{y \in \operatorname{dom} f} (f(x) - f(y))(x-y) \geq 0 \rightarrow x \in \operatorname{dom} f\right)$

For example



this function is: $f(x) = x^2$, $\operatorname{dom} f = [1,2] \cup [4, \infty)$, so if a function is not maximal monotone then

$$\neg\left(\forall_x \left(\forall_{y \in \operatorname{dom} f} (f(x) - f(y))(x-y) \geq 0 \rightarrow x \in \operatorname{dom} f\right)\right)$$

$$= \exists_x \left(\forall_{y \in \operatorname{dom} f}\ f(x) - f(y))(x-y) \geq 0 \wedge x \notin \operatorname{dom} f\right)$$

now set $x = 3 \notin \operatorname{dom} f$

$\forall_{y \in \operatorname{dom} f = [1,2] \cup [4, \infty]}$

$$(f(x) - f(y))(x - y) = (9 - y^2)(3 - y) = (3+y)(3-y)(3-y) = (3+y)(3-y)^2 \geq 0$$

$$\therefore \exists_{x=3} \left(\forall_{y \in \operatorname{dom} f} (f(x) - f(y))(x-y) \geq 0 \wedge x \notin \operatorname{dom} f\right)$$ $\therefore$ So, f is not maximal monotone

But it can be trivially shown that, $\forall_x \left(x \in \operatorname{dom} f \rightarrow \forall_{y \in \operatorname{dom} f} (f(x) - f(y))(x-y) = (x^2 - y^2)(x-y) = (x+y)(x-y)^2 \geq 0\right)$ i.e., f is monotone

$\therefore$ So, in summary we have a function $f(x)$ which is monotone but not maximal monotone

Now let's extend this for vector valued case, as vectors are partial ordered we cannot just start with

$x \geq y \Rightarrow F(x) \geq F(y)$, however look at the scalar case $\forall_{(x,y)} (f(x) - f(y))(x - y) \geq 0$ and trying to

extend it point wise for a vector case: $\forall_{x,y \in \mathbb{R}^n} \sum_{i=1}^{n} (F(x)_i - F(y)_i)(x_i - y_i) = (F(x) - F(y))^T (x - y) \geq 0$

What might be an intuition of this? Well it means that in some element of the vector the monotone operator is results in a such scalar monotonic behavior to balance any non-monotonic behavior in other elements. For a 2d case, $(f(x)_1 - f(y)_1)(x_1 - y_1) + (f(x)_2 - f(y)_2)(x_2 - y_2) \geq 0$

So if the first term gets negative for some x,y, then the other terms will counter it by getting strongly positive. Note that this means: $\forall_{(x,u)} \left((x,u) \in F \Rightarrow \forall_{(y,v) \in F} (u - v)^T (x - y) \geq 0\right)$

- this is montone, if we modify the notation as follows:

$$\forall_{(x,u)} \left((x,u) \in F \leftrightarrow \forall_{(y,v) \in F} (u - v)^T (x - y) \geq 0\right)$$ maximal monotone

* Monotone operators: def: monotone operators definitions and related

$F\{\text{relation}, : \mathbb{R}^n\}\{\text{monotone}\} \overset{\text{def}}{\equiv} \forall_{\underbrace{(x,u)}_{x F u} \in F}\ \forall_{\underbrace{(y,v)}_{y F v} \in F} (u - v)^T (x - y) \geq 0$ // e.g. for $F(x) = e^x$ $(e^x - e^y)(x - y) = (+ve)(+ve) \geq 0$

(in the increasing sense)

def: monotone operator

say $x > y$ then $e^x > e^y$

· Maximality:

যদি আর কোন monotone operator $\tilde{F}$ না থাকে, যা কিনা F কে contain করে

$F\{\text{maximal monotone}\} \overset{\text{def}}{\equiv} \neg\left(\exists_{\tilde{F}\{\text{monotone operator}, \neq F\}} F \subset \tilde{F}\right)$; A technical definition $\forall_{(x,u)} \left((x,u) \in F \leftrightarrow \forall_{(y,v) \in F} (v-u)^T (x-y) \geq 0\right)$ def: maximal monotone

quite critical in

because F being a relation is a set of $\{u, F(u)\}$ structure

|| তারমানে → হবে just monotone

|| additionally ← হলে maximally monotone.

· Strong monotonicity:

$F\{\text{strongly monotone}\} \overset{\text{def}}{\equiv} \exists_{m>0} \forall_{(x,u),(y,v) \in F} (u - v)^T (x - y) \geq m\|x - y\|_2^2 \overset{\text{def}}{\equiv} \exists_{m>0} \forall_{x,y \in \operatorname{dom} F} (F(x) - F(y))^T (x - y) \geq m\|x - y\|_2^2$

($u \in F(x)$, $v \in F(y)$)

// so $x \neq y$ হলে হবে $(u - v)^T (x - y) > 0$

$\forall_{x,y \in \operatorname{dom} f} (\|F(x) - F(y)\|_2 \leq L\|x - y\|_2)$ By Cauchy-Schwartz, $(F(x) - F(y))^T (x - y) \leq |(F(x) - F(y))^T (x - y)| \leq \|F(x) - F(y)\|_2 \|x - y\|_2 \leq L\|x - y\|_2^2$

$|a^T b| \leq \|a\| \|b\|$

$F\{\text{strongly monotone, Lipschitz with constant } L\} \overset{\text{def}}{\equiv} \forall_{x,y \in \operatorname{dom} F}\ m\|x - y\|_2^2 \leq (F(x) - F(y))^T (x - y) \leq L\|x - y\|_2^2$

$\Rightarrow \kappa = \frac{L}{m} \geq 1$

condition number of strongly monotone Lipschitz relation F.

* Basic properties of monotone operator:

● Sum and scalar multiple

- $F\{\text{monotone}\}, G\{\text{monotone}\} \Rightarrow F + G$ is monotone [eq: sum of monotone is monotone]

[eq: Sum of relations]

- $(F\{\text{maximal monotone}\}, G\{\text{maximal monotone}\}\{\operatorname{dom} F \cap \operatorname{int} \operatorname{dom} G \neq \emptyset\}) \Rightarrow F + G\{\text{maximal monotone}\}$ # Additivity of maximal monotone operator with fine print

Some interior points of one relation domain has to belong to the other relation domain

- $F\{(\text{maximal}) \text{ monotone}\} \Rightarrow \alpha_{\{>0\}} F\{(\text{maximal}) \text{ monotone}\}$ # Positive scalar multiplicatively of (maximal) monotone operator

- $\begin{pmatrix} F\{\text{strongly monotone with parameter } m\}, \\ G\{\quad " \quad " \quad " \quad " \quad \tilde{m}\} \end{pmatrix}$

$\Rightarrow$

$(F+G)\{\quad " \quad " \quad " \quad " \quad m + \tilde{m}\}$

- $F\begin{bmatrix}\text{strongly monotone} \\ \text{parameter } m\end{bmatrix} \Rightarrow \alpha_{\{>0\}} F\{\text{strongly monotone with parameter } m\}$

● Inverse:

$F\{(\text{maximal}) \text{ monotone}\} \Rightarrow F^{-1}\{(\text{maximal}) \text{ monotone}\}$ (which always exists) [eq: inverse of monotone = monotone]

$F\{\text{strongly monotone with parameter } m\} \Rightarrow F^{-1}\{\text{function with Lipschitz constant}, L = \frac{1}{m}\}$ WOW! # What does that mean? It means that: inverse of a strongly monotone operator (which may be a nontrivial relation) is a function (as any relation with a Lipschitz constant is a function) which is freaking amazing.

Proof:

$\forall_{(x,u) \in F} \forall_{(y,v) \in F}$

$$\underbrace{m\|x - y\|_2^2 \leq (u - v)^T (x - y)}_{\text{definition of strong monotonicity}} \leq \overbrace{|(u - v)^T (x - y)| \leq \|u - v\|_2 \|x - y\|_2}^{\text{Cauchy Schwartz}}$$

if $u = v$ then:

definition of strong monotonicity

if $u=v$ then:

$$\left.\begin{array}{l} \underbrace{m \underbrace{\|x-y\|_2^2}_{\geq 0}}_{\geq 0} \leq 0 \end{array}\right\} \Rightarrow \|x-y\|_2 = 0 \leftrightarrow x=y$$

now as $(x,u)\in F,\ (y,v)\in F \Leftrightarrow (u,x)\in F^{-1},\ (v,y)\in F^{-1}$

now $u=v$ [illegible] $x=y$, so same input [illegible] F^{-1} is a singleton

$\therefore$ $\boxed{F^{-1} \text{ is a function.}}$

$x = F^{-1}(u)$

$y = F^{-1}(v)$

When $x\neq y$, dividing the inequality $m\|x-y\|_2^2 \leq \|u-v\|_2\|x-y\|_2$ by $\|x-y\|_2$

$m\|\underset{F^{-1}(u)}{x}-\underset{F^{-1}(v)}{y}\|_2 \leq \|u-v\|_2$

$\Leftrightarrow\ m\|F^{-1}(u)-F^{-1}(v)\|_2 \leq \|u-v\|_2$

$\therefore\ \|F^{-1}(u)-F^{-1}(v)\|_2 \leq \left(\frac{1}{m}\right)\|u-v\|_2$

$\Leftrightarrow F^{-1}(\cdot)$ is Lipschitz with Lipschitz constant $L=\frac{1}{m}$

(Proved!)

Congruence:

- $\{A\in\mathbb{R}^{s\times t},\ F\{\text{monotone}\}\} \Rightarrow G(\square) = \underbrace{A^T}_{t\times s}F(A\square)\ \{\text{monotone relation}\}$ [eq: congruence of monotone = monotone]

" " maximal " $\Rightarrow$ " " {maximal} //except some pathological //case

- $\{A \in \mathbb{R}^{s\times t}, \text{rank}(A)=t\ (\text{skinny}),\ F\{\text{strongly monotone with parameter } m\}\} \Rightarrow G(\square) = A^TF(A\square)$ {strongly monotone with parameter $m\sigma^2_{min}(A)$}

- $\{A \in \mathbb{R}^{s\times t}, \text{rank}(A)=t\ (\text{skinny}),\ F\{\text{Lipschitz with constant } L\}\} \Rightarrow G(\square) = A^TF(A\square)$ {Lipschitz with constant $m\sigma^2_{max}(A)$}

// [illegible] because they are

- Zero set of monotone operator: // solution of optimization problem = {zeros of associated monotone operator})

$F\{\text{maximal monotone}\} \rightarrow \forall_x\ F(x) = \{z \mid xFz\}\ \{\text{CD}\}$

definition of monotone operator: $F:\{\text{maximal monotone}\} \Leftrightarrow \forall_{(x,u)} \left(xFu \leftrightarrow \forall_{yFv}\ (u-v)^T(x-y)\geq 0 \right)$

Proof: $u_1, u_2 \in F(x) \leftrightarrow (x,u_1)\in F,\ (x,u_2)\in F$

want to check: $\theta u_1 + (1-\theta)u_2 \overset{?}{\in} F(x) \leftrightarrow (x, \theta u_1+(1-\theta)u_2) \overset{?}{\in} F \leftrightarrow \forall_{yFv}\ (\theta u_1+(1-\theta)u_2 - v)^T(x-y) \overset{?}{\geq} 0$ ✓

$\forall_{v\in F(y)}$ ($(y,v)\in F$)

$(\theta u_1 + (1-\theta)u_2 - v)^T(x-y) = (\theta(u_1-v)+(1-\theta)(u_2-v))^T(x-y) = \theta(u_1-v)^T(x-y)+(1-\theta)(u_2-v)^T(x-y) \geq 0$ $\therefore$ F(x) is convex.

$\theta u_1 - \theta v + (1-\theta)u_2 - \underbrace{v+\theta v}_{-(1-\theta)v}$

$= \theta(u_1-v)+(1-\theta)(u_2-v)$

as $(x,u_1)\in F, (x,u_2)\in F, (y,v)\in F$, and F{monotone}

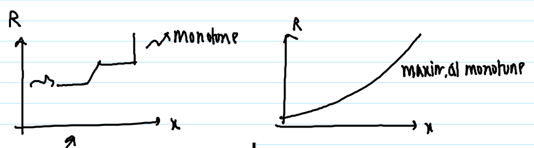
$\Rightarrow (u_1-v)^T(x-y)\geq 0,\ (u_2-v)^T(x-y)\geq 0$ //By defn of monotonicity

$\Rightarrow \theta(u_1-v)^T(x-y)+(1-\theta)(u_2-v)^T(x-y)\geq 0$ #

- $F^{-1}(0) = \{\text{zero set of maximal monotone operator}\}$

$= \{x \mid (x,0)\in F$ [illegible] $xF0\}$: convex

*Examples:



4.2 Examples

Relations on R. We describe this informally. A relation R on $\mathbf{R}$ is monotone if it is a curve in $\mathbf{R}^2$ that is always nondecreasing; it can have horizontal (flat) portions and also vertical (infinite slope) portions. If it is a continuous curve with no end points, then it is maximal monotone.

A relation on $\mathbf{R}$ is strongly monotone with parameter m if it maintains a minimum slope m everywhere; it has Lipschitz constant L if its slope is never more than L.

R{strongly monotone with parameter m} [illegible] minimum slope at any point is $\geq m$

R{monotone with Lipschitz constant m} [illegible] maximum slope at " " is $\leq L$

Continuous functions. A continuous monotone function $F : \mathbf{R}^n \to \mathbf{R}^n$ (with $\mathbf{dom}\, F = \mathbf{R}^n$) is maximal.

Affine functions. An affine function $F(x) = Ax + b$ is maximal monotone if and only if $A + A^T \succeq 0$. It is strongly monotone with parameter $m =$

$\lambda_{\min}(A + A^T)/2$ when $A + A^T \succ 0$.

$F(x) = Ax+b = \{\text{maximally monotone}\} \Leftrightarrow A+A^T\succeq 0$

$F(x) = Ax+b = \{\text{strongly monotone with parameter } m=\lambda_{min}\left(\frac{A+A^T}{2}\right)\} \Leftarrow A+A^T\succ 0$

Subdifferential Mapping: • If $f\{\text{CCP}, \mathbb{R}^n\rightarrow\mathbb{R}\cup\{\infty\}\} \Rightarrow \partial f\{\text{maximal monotone}\}$ (C: closed, C: convex, P: proper) def: CCP

Proof: [illegible] monotone part [illegible]: # This proof just needs convexity, closed-ness and properness is not needed

assume, $u\in \underset{\partial f(x)}{F(x)}$, $v\in F(y) = \partial f(y)$

$(x,u)\in F$ $(y,v)\in F$

$\Leftrightarrow \forall_{\tilde{x}\in \text{dom} f}\ f(\tilde{x}) \geq f(x)+u^T(\tilde{x}-x)$: $\tilde{x}=y \Rightarrow f(y)\geq f(x)+u^T(y-x) \leftrightarrow f(y)-f(x)\geq u^T(y-x)$

$\forall_{\tilde{y}\in \text{dom} f}\ f(\tilde{y}) \geq f(y)+v^T(\tilde{y}-y)$: $\tilde{y}=x \Rightarrow f(x)\geq f(y)+v^T(x-y) \leftrightarrow f(x)-f(y)\geq v^T(x-y)$

$0 \geq \underbrace{u^T(y-x)}_{-u^T(x-y)}+v^T(x-y)$

$\Leftrightarrow (u^T-v^T)(x-y) = (u-v)^T(x-y)\geq 0$ $\therefore$ ∂f is a monotone operator.

def: monotone operator

note that in the proof convexity was not required

So, for any function f, ∂f is a monotone operator

However it turns out that (Ernest told me) maximality is indeed needed for proposing efficient algorithm. for example, for nonconvex function too subdifferential operator is monotone but not maximally monotone. So if f is a nonconvex function, then $\forall_{(x,u)}\left(\partial f(x)\ni u \rightarrow \forall_{v:\partial f(y)\ni v}\ (u-v)^T(x-y)\geq 0\right)$

but the converse is not true, $\forall_{(x,u)}\left(\neg\left(\forall_{v:\partial f(y)\ni v}\ (u-v)^T(x-y)\geq 0 \rightarrow \partial f(x)\ni u\right)\right)$

Normal Cone Operator: normal cone operator

$\{C\in\mathbb{R}^n, [\,], D\}$

$N_C\{\text{normal cone operator}\} = \begin{cases} \phi, & x\notin C \\ \{g \mid \forall_{z\in C}\ g^T(z-x)\leq 0\}, & x\in C \end{cases}$

$x \rightarrow N_C(\square)$: $x\notin C \rightarrow \phi$; $x\in C \rightarrow \{g \mid \forall_{z\in C}\ g^T(z-x)\leq 0\} = \left(N_C(x) = \{\} \,//\, \left(z\in C \,//\, \text{find } g: g^T(z-x)\leq 0 \,//\text{push!}(N_C(x), g)\right)_{\forall z}\right)$

the angle between $\angle g, (z-x)$ is greater than 90°

$\forall_{x \in \text{int} C} \quad N_C(x) = \{0\}$

$\forall_{x \in \partial C} \quad N_C(x) = \text{nontrivial}$ # The intuition behind this is that remember at the boundary of any function, the subdifferential might not exist. One strange function in this regard is the function of a $I_C(X)$ convex set

Why $N_C(x)$ matters? Because this is the subdifferential mapping of the convex indicator function $I_C(\boxminus) = \begin{cases} 0, & \boxminus \in C \\ \infty, & \boxminus \notin C \end{cases}$, i.e., $N_C = \partial I_C$.

Makes sense because remember, for any convex function the subdifferential always exists in the int(dom f). At the boundary the subdifferential may not exist, i.e, $\forall_{x \in bd(C)} \; \partial I_C(x)$ might not exist, as a result $N_C(x \in \partial C) = \partial I_C(x \in \partial C) = \text{nontrivial}$

But remember for at all the interior points, $N_C(x) = \partial I_C(x) = \{0\}$

[eq: subdiffernetial is monotone mapping]

proof: $I_C(x) = \begin{cases} 0, & x \in C \\ \infty, & x \notin C \end{cases}$, want to determine, $\partial I_C(x) = N_C(x)$

kenona shokol y er jonno def ti khate hobe bole, min[0,\infty]=0 bibechona korlei cholbe, equivalently kebol y \in C bibechona korlei cholbe, as in the lhs, we are considering only y \in C er corresponding I_C(.) value.

By definition of subgradient, $g_x \in \partial I_C(x) \Leftrightarrow \forall_{y \in \text{dom} I_C} \; \underbrace{I_C(y)}_{[0,\infty]} \geq I_C(x) + g_x^T(y-x) \Leftrightarrow \forall_{y \in \text{dom} I_C} \; \min[0,\infty] = 0 \geq I_C(x) + g_x^T(y-x) \Leftrightarrow \forall_{y \in C} \; 0 \geq I_C(x) + g_x^T(y-x)$

- case 1: $x \in C$, then $\forall_{y \in C} \; 0 \geq \underbrace{I_C(x)}_{=0} + g_x^T(y-x)$

  $\Leftrightarrow \forall_{y \in C} \; g_x^T(y-x) \leq 0$

  $\therefore x \in C \Rightarrow \partial I_C(x) = \{g_x \mid \forall_{y \in C} \; g_x^T(y-x) \leq 0\}$

- case 2: $x \notin C$, $\Rightarrow$ $0 \geq \infty + \underbrace{g_x^T(y-x)}_{\text{finite}} \Rightarrow$ no finite g_x can exist $\therefore \partial I_C(x) = \phi$

$\therefore \partial I_C(x) = N_C(x) = \begin{cases} \phi, & x \notin C \\ \{g_x \mid \forall_{y \in C} \; g_x^T(y-x) \leq 0\} = \begin{cases} 0, & x \in \text{int}(C) \\ \text{nontrivial}, & x \in bd(C) \end{cases} \end{cases}$

Not the most rigorous proof ...

- Saddle Subdifferential:

$f\{: \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R} \cup \{\pm\infty\}, \text{CCP}\}$ assume $f(\boxminus, \boxminus)$ is convex in $\boxminus$, concave $\boxminus$

$F(x,y) = \{\text{saddle subdifferential relation}\}$

$= \begin{bmatrix} \partial_x f(x,y) \\ \partial_y \underbrace{(-f(x,y))}_{-f(x,y)\text{: convex in } f} \end{bmatrix} \{\forall_{x,y} \; \partial_x f(x,y) \neq \phi, \partial_y f(x,y) \neq \phi\}$

*Why this is called saddle subdifferential?

Because $(x,y) F 0 \Leftrightarrow (x,y) \in F^{-1}(0) \Leftrightarrow 0 \in F(x,y) = \begin{bmatrix} \partial_x f(x,y) \\ \partial_y(-f(x,y)) \end{bmatrix} \Leftrightarrow$

$0 \in \partial_x f(x,y) \Leftrightarrow x \in \underset{\tilde{x} \in \text{dom}}{\text{argmin}} f(\tilde{x},y) \Leftrightarrow \forall_{(\tilde{x},y) \in \text{dom} f} \; f(x,y) \leq f(\tilde{x},y)$

$0 \in \partial_y \underbrace{(-f(x,y))}_{\text{convex}} \Rightarrow y \in \underset{\tilde{y} \in \text{dom}}{\text{argmin}} (-f(x,\tilde{y})) \Leftrightarrow \forall_{(x,\tilde{y}) \in \text{dom} f} \; \underbrace{-f(x,\tilde{y}) \geq -f(x,y)}_{\Updownarrow} \quad f(x,\tilde{y}) \leq f(x,y)$

$\Rightarrow \forall_{(\tilde{x},y) \in \text{dom} f, \; (x,\tilde{y}) \in \text{dom} f} \; f(x,\tilde{y}) \leq f(x,y) \leq f(\tilde{x},y)$

$\therefore$ 0 set of F is the saddle points of f.

Saddle subdifferential relation for CCP function $f(x,y)$ is maximal

*KKT operator: # Manuscript version

$\begin{pmatrix} \underset{x}{\text{minimize}} \; f_0(x) \sim D_x \\ \text{subject to} \\ \forall_{i \in \{1,\ldots,m\}} \; f_i(x) \leq 0 \; (D_x) \\ \forall_{i \in \{1,\ldots,p\}} \; \underbrace{h_i(x) = 0}_{\text{affine}} \end{pmatrix}$

$L(x,\lambda,\nu) = \begin{cases} f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x), & \text{if } \lambda \succeq 0 \\ -\infty, & 0 \end{cases}$

$T(x,\lambda,\nu) = \{\text{KKT operator}\} = \begin{bmatrix} \partial_x L(x,\lambda) \\ -F(x) + N_{\{\lambda \succeq 0\}} \\ -H(x) \end{bmatrix}$, so $(x^*, \lambda^*, \nu^*) T 0 \Leftrightarrow$

$\partial_x L(x,\lambda) \ni 0$ // vanishing gradient of Lagrangian

$-F(x) + \partial I_{\{\lambda \succeq 0\}} = [-f_i(x) + \partial I_{\{\lambda_i \succeq 0\}}]_{i=1}^{m} = 0$ (?) explanation needed // primal inequality feasibility + // dual variable feasibility

$-H(x) = \begin{bmatrix} -h_1(x) \\ \vdots \\ -h_p(x) \end{bmatrix} = 0$

$N_{\{\lambda \succeq 0\}} = \partial I_{\{\lambda \succeq 0\}}$ // ensures that // dual feasibility condition // holds

$H(x) = \begin{bmatrix} h_1(x) \\ \vdots \\ h_p(x) \end{bmatrix}$

$F(x) = \begin{bmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{bmatrix}$

$T(x,\lambda,\nu)$ {special case of saddle subdifferential} = {monotone}

- $T(x^*, \lambda^*, \nu^*) \ni T(x^*, \lambda^*, \nu^*) \Leftrightarrow (x^*, \lambda^*, \nu^*)$ solves the optimization problem

*KKT operator (Slide version)

minimize $f(x)$

subject to $Ax = b$

$L(x,y) = f(x) + y^T(Ax - b)$

KKT operator: $F(x,y) = \begin{bmatrix} \partial_x L(x,y) \\ -\partial_y L(x,y) \end{bmatrix}$

\# gives L·H·S of vanishing gradient at Lagrangian condition (one of the KKT condition)

\# gives L·H·S of primal feasibility (another KKT condition)

Note that this is an equality constrained optimization problem, so it will have only these two KKT conditions with RHS zero and the LHS I given by the rows of the KKT operator. So if the optimal (x*,y*) is inputted through the KKT operator then we will have zero.

$F(\boxminus, \boxminus) = \begin{bmatrix} \partial_{\boxminus} L(\boxminus, \boxminus) \\ -\partial_{\boxminus} L(\boxminus, \boxminus) \end{bmatrix}$

so if $F(x^*, y^*) = 0 \Leftrightarrow (x^*, y^*)$: optimal primal dual pair

$\Leftrightarrow ((x^*, y^*), 0) \in F$

$\Leftrightarrow (0, (x^*, y^*)) \in F^{-1}$

$\Leftrightarrow (x^*, y^*) \in F^{-1}(0)$ So optimal primal dual pair will belong to the output set of the inverse KKT operator with 0 fed into it!

\# Multiplier to residual mapping is very important as it has a connection with ADMM

* Multiplier to residual mapping. def: multiplier to residual mapping

$\begin{pmatrix} \text{minimize } f(x) \{\text{CCP strongly}\} \\ \text{subject to } Ax = b \end{pmatrix}$

$L(x,y) = f(x) + y^T(Ax - b)$ {more technically: $F(\square) = b - A \underset{x}{\text{argmin}} L(x, \boxminus) \therefore F(y) = (b - A \underset{x}{\text{argmin}} L(x, \boxminus))(y) = b - A \underset{x}{\text{argmin}} L(x,y)$

define $F(y) = b - A x^*(y)$ {$x^*(y) = \underset{x}{\text{argmin}} L(x,y)$, because $f(x)$: strongly convex, $y^T(Ax-b) = (A^T y)^T x - b^T y$: affine in $x \Rightarrow f(x) + y^T(Ax-b)$: strongly convex, so will have unique minimizer

\# Note: convex + strongly convex = convex

This F is called multiplier to residual mapping, because it takes the lagrange multiplier and outputs the residue that associated with the sub-optimality of x.

$\therefore x^*(y)$: $\underbrace{\partial_x L(x,y)}_{\partial_x [f(x) + y^T Ax - y^T b] = \partial_x f(x) + (A^T y)} \ni 0 \Leftrightarrow \partial_x f(x) + A^T y \ni 0 \Leftrightarrow \exists_{\eta_x \in \partial_x f(x)} \; \eta_x + A^T y = 0$

$\Leftrightarrow \exists_{\eta_x \in \partial_x f(x)} \; \eta_x = -A^T y$

$\Leftrightarrow -A^T y \in \partial_x f(x) \Leftrightarrow (x, -A^T y) \in \partial_x f$

$\Leftrightarrow (-A^T y, x) \in (\partial_x f)^{-1}$

$\Leftrightarrow x^*(y)$: $(-A^T y, x^*(y)) \in (\partial_{x^*(y)} f)^{-1}$

$\Updownarrow$

$x^*(y) = (\partial_{x^*(y)} f)^{-1}(-A^T y)$ # but $x^*(y)$ is unique as it is the minimizer of a strongly convex function.

{Alternative definition to multiplier to residual mapping operator}

$\boxed{F(y) = b - A(\partial_{x^*(y)} f)^{-1}(-A^T y)}$ this is a monotone operator, so multiplier to residual mapping is a monotone operator

{proof: $(\partial_{x^*(y)} f)$ {monotone} [eq: subdiffernetial is monotone mapping]

$\Rightarrow (\partial_{x^*(y)} f)^{-1}$ {monotone} [eq: inverse of monotone = monotone]

$\Rightarrow \forall_{M} \; M^T (\partial_{x^*(y)} f)^{-1} (M \boxminus)$ {monotone} [eq: congruence of monotone = monotone]

\# F: monotone $\rightarrow A^T F(A \square)$: monotone

$\downarrow \{M = -A^T$

$-A(\partial_{x^*(y)} f)^{-1}(-A^T \boxminus)$ {monotone}

$\Rightarrow \{b - A(\partial_{x^*(y)} f)^{-1}(-A^T \boxminus)\}$ {monotone} } addition of a constant vector to each element of a relation set will not change monotonicity of a relation ; equivalent logic.

$b \mathbb{I} \boxminus$ is a monotone operator

$-A(\partial_{x^*(y)} f)^{-1}(-A^T \boxminus)$ is a monotone operator

$\Rightarrow$ $\oplus$ $(b\mathbb{I}\boxminus + (-A(\partial_{x^*(y)} f^{-1})(-A^T\boxminus))$ is a monotone operator

[eq: sum of monotone is monotone]

Now we show:

$*$ $F(y) = b - A(\partial f)^{-1}(-A^T y)$

$= \partial_y [\, b^T y + f^*(-A^T y) \,]$

Proof:

$$\partial_y [\, b^T y + f^*(-A^T y) \,] = b + \underbrace{\partial_y [\, f^*(-A^T y) \,]}_{-A\, [\partial_y f^*(y)]_{y := -A^T y}}$$

/* chain rule: $\partial_x f(Ax) = A^T [\partial_x f(x)]_{x = Ax}$ */

now $\partial_y f^*(y) = \partial_y [\, \sup_x - f(x) + y^T x \,]$

$= \partial_y [\, -f(x^*) + y^T x^* \,] = x^* : x^* = \operatorname{argmax}_x -f(x) + y^T x$

$x^* = \operatorname{argmax}_x (-f(x) + y^T x)$

$= -f(x^*) + y^T x^* : x^* = \operatorname{argmax}_x -f(x) + y^T x$

$= \operatorname{argmin}_x f(x) - y^T x$

/* recall, if $f = \sup_{\alpha \in A} (f_\alpha)$, $f = \sup_{\alpha \in A} f_\alpha = f_\beta$ $\Rightarrow$ find $\eta \in \partial f_\beta \Rightarrow \eta \in \partial f$ */

/* $x^* = \operatorname{argmin}_x f(x) - y^T x$

$\leftrightarrow$ $[\partial_x (f(x) - y^T x)]_{x = x^*} \ni 0$

$\leftrightarrow$ $\partial f(x^*) - y \ni 0 \leftrightarrow \partial f(x^*) \ni y$

$\leftrightarrow$ $(x^*, y) \in \partial f \leftrightarrow (y, x^*) \in \underbrace{(\partial f)^{-1}}_{\text{this is a function}} \leftrightarrow x^* = (\partial f^{-1})(y)$ */

this is a function (inverse of subdifferential is a function)

$= (\partial f^{-1})(y)$

$\therefore$ $\partial_y f^*(y) = (\partial f^{-1})(y)$

$\therefore$ $\partial_y (b^T y + f^*(-A^T y)) = b - \partial_y (f^*(-A^T y)) = b - [(\partial f^{-1})(y)]_{y := -A^T y}$

$= b - A(\partial f^{-1})(-A^T y) = b - A\, [\partial f^{-1}(y)]_{y = -A^T y} = b - A\, (\partial f^{-1})(-A^T y)$ $\blacksquare$