FDA Short Review

I

① $\delta_{i,j} = \begin{cases} 1, & i=j \\ 0, & \text{otherwise} \end{cases}$

② $\operatorname{range}(A) = \{ Ax : x \in \mathbb{K}^J \}$
$= \operatorname{span}(A_{(:,j)}, i \in J)$

③ $\langle x, y \rangle = y^H x = \sum_{i \in I} x_i \overline{y_i}$

④ Norm $\|\cdot\| : V \to [0, \infty)$
- $\|v\| = 0$ iff $v = 0$
- $\|\lambda v\| = |\lambda| \|v\| \quad \forall v \in V$ and $\lambda \in \mathbb{K}$
- $\|v+w\| \leq \|v\| + \|w\| \quad \forall v, w \in V$

also $\big| \|v\| - \|w\| \big| \leq \|v - w\| \quad \forall v, w \in V$

⑤ $(V, \|\cdot\|)$ Hilbert space if
$\|v\| = \sqrt{\langle v, v \rangle} \quad v \in V$

properties

- $\langle v, v \rangle > 0 \quad \forall v \neq 0$
- $\langle v, w \rangle = \overline{\langle w, v \rangle} \quad \forall v, w \in V$
- $\langle u + \lambda v, w \rangle = \langle u, w \rangle + \lambda \langle v, w \rangle$
- $\langle w, u + \lambda v \rangle = \langle w, u \rangle + \overline{\lambda} \langle w, v \rangle$

$u, v, w \in V$
$\lambda \in \mathbb{K}$

⑥ Cauchy Schwarz
$|\langle v, w \rangle| \leq \|v\| \|w\| \quad v, w \in V$

⑦ **Projection**

- $P_K(v) = \arg\min_{w \in K} \|v - w\|$

- given K is a closed convex set in hilbert space $(V, \langle \cdot, \cdot \rangle)$

$$\mathrm{Re}\left(\langle z - P_K(v), v - P_K(v) \rangle\right) \leq 0$$

$z \in K$

⑧ $$P_W(v) = \sum_{\gamma \in F} \langle v, w_\gamma \rangle w_\gamma$$

if $K = W \subset V$ is close linear subspace of V, then projection onto W.

⑨ **Pythagoras Fourier**

$$\|P_W(v)\|^2 = \sum_{\gamma \in F} |\langle v, w_\gamma \rangle|^2 \quad \forall v \in V$$

⑩ **Trace**

$$\mathrm{tr}(A) = \sum_{i = I} A_{ii}$$

⑪ **Frobenius Norm** : $\|A\|_F = \sqrt{\sum_{i \in I, j \in J} |A_{i,j}|^2} \quad A \in \mathbb{K}^{I \times J}$

12

Hilbert-Schmidt norm :

$$\langle A, B \rangle_F = \sum_{i \in I} \sum_{j \in J} A_{ij} \overline{B_{ij}} = tr(AB^H) = tr(B^H A)$$

$$\|A\|_F^2 = tr(A^H A) = tr(A A^H)$$

II chapter 2

① length of squared projection

$$\|Av\|_2^2 = \sum_{i \in I} |\langle A^{(i)}, v \rangle|^2$$

right singular value

② $$v_k = \underset{\substack{\|v\|_2 = 1 \\ \langle v_1, v \rangle = 0 \dots \langle v_{k-1}, v \rangle = 0}}{\operatorname{argmax}} \|Av\|_2$$

③ $v_k \rightarrow$ subspace spanned by singular vectors $(v_1, v_2, \dots v_k)$ is best fit k dimensional subspace.

④ Schatten-p-norm

$$\|A\|_p = \left(\sum_{k=1}^{r} \sigma_k(A)^p \right)^{1/p}$$

for $p = \infty$ $$\|A\|_\infty = \max_{k=1,\dots r} \sigma_k(A)$$

(5) left singular value

$$u_i = \frac{1}{\sigma_i(A)} A v_i$$

(6) Thm $A = \sum_{i=1}^{r} \sigma_i(A) u_i v_i^H$

Thm :
(7) for $A \in \mathbb{K}^{I \times J}$ 'r' rank matrix
left singular value of A, $u_1, u_2 \dots u_r$ are orthogonal

(8) k- truncated (best k rank approx.)

$$A_k = \sum_{l=1}^{k} \sigma_l u_l v_l^H$$

(9) Rows of A_k are the orthogonal projections of the rows of A onto the subspace V_k spanned by the first k singular vectors of A.

(10) $$\| A - B \|_F^2 = \sum_{i \in I} \| A^{(i)} - P_V(A^{(i)}) \|_2^2$$

(11) $$\| A - A_k \| = \sigma_{k+1}^2$$

(12) for any matrix B of rank at most k

$$\| A - A_k \| \leq \| A - B \|$$

(13) Principle of power method:

$$A^m = \sum_{k=1}^{r} \sigma_k^m v_k v_k^T$$

(14) spectral decomposition of $B = AA^T$

$$B = \sum_{k=1}^{r} \sigma_k^2 u_k u_k^T$$

$$B^m = \sum_{k=1}^{r} \sigma_k^{2m} u_k u_k^T$$

(15) Lemma: Let $x \in \mathbb{R}^d$ be a unit d-dimensional vector of components $x = (x_1, \dots x_d)$ w.r.t canonical basis and picked at random from the set $\{x : \|x\| \leq 1\}$, the probability that $|x_1| \geq \alpha > 0$ is at least $1 - 2\alpha\sqrt{d-1}$

(16) Theorem: Let $A \in \mathbb{R}^{I \times J}$ and $x \in \mathbb{R}^J$ be a random unit length vectors. Let V be the space spanned by the left singular values of A corresponding to singular values greater than $(1-\varepsilon)\sigma_1$. Let m be $\Omega\left(\frac{\ln(d^\beta/\varepsilon)}{\varepsilon}\right)$, for $\beta \geq 1/2$. Let w^* be the unit vector after m iterations of the power method, namely

$$w^* = \frac{(AA^H)^m x}{\|(AA^H)^m x\|_2}$$

the probability that w^* has a component of at most $O\left(\frac{\varepsilon}{\alpha d^3}\right)$ orthogonal to V is at least

$1 - 2\alpha\sqrt{2d-1}$.

(17) Pseudo inverse : $A^+ = V \Sigma^{-1} U^H$

if num rows ≥ num columns
and columns are linearly independent

$$A^+ = (A^H A)^{-1} A^H$$

(18) Proposition : $A \in \mathbb{K}^{I \times J}$ and $y \in \mathbb{K}^I$. Define $\mathcal{M} \subset \mathbb{K}^J$ to be set of minimizers of the map $x \longmapsto \|Ax - y\|_2$. The convex optimization problem

$$\arg\min_{x \in \mathcal{M}} \|x\|_2$$

has a unique solution $x^* = A^+ y$

Stability of Eigen values

(19) Theorem : (Spectral theorem for Hermetian matrices): if $A \in \mathbb{K}^{I \times I}$ and $A = A^H$, then $\exists$ an orthonormal basis $\{v_1, .. v_n\}$ consisting of eigenvectors of A with real corresponding eigenvalues $\{\lambda_1, ... \lambda_n\}$ such that

$$A = \underbrace{\sum_{k=1}^{n} \lambda_k v_k v_k^H}_{\text{spectral decomposition of } A.}$$

$u_k = \operatorname{sign} \lambda_k v_k$
$\sigma_k = |\lambda_k|$

(20) If $A, E \in \mathbb{K}^{I \times I}$ are two Harmetian matrices then for all $k = 1, \dots n$

$$\lambda_k(A) + \lambda_n(E) \leq \lambda_k(A+E) \leq \lambda_k(A) + \lambda_1(E)$$

this does not depend on magnitude of perturbation

(21) corollary: if $A, E \in \mathbb{K}^{I \times I}$ are two arbitrary matrices, then $\forall\, k = 1, \dots n$

$$|\sigma_k(A+E) - \underset{\text{singular values}}{\sigma_k}(A)| \leq \underset{\text{spectral norm}}{\|E\|}$$

(22) $A, E \in \mathbb{K}^{I \times I}$ are two arbitrary matrices

$$\sqrt{\sum_{k=1}^{n} |\sigma_k(A+E) - \sigma_k(A)|^2} \leq \underset{\text{Frobenius norm.}}{\|E\|_F}$$

<u>Stability of singular norms: Wedin's bound</u>:

(23) <u>Principal angles between subspaces</u>:

$$\cos \theta(v, w) = \frac{w^T v}{\|v\|_2 \|w\|_2}$$

(24) Definition : let $V, W \in \mathbb{K}^{J \times I}$ be orthogonal matrices V and W, n dimensional subspaces of $\mathbb{K}^J$ ($n = \# I \leq \# J = d$).

$$\cos \Theta(V, W) = \Sigma(W^H V)$$

List of singular values of $W^H V$.

(25) $V, W \in \mathbb{K}^{J \times I}$ be orthogonal matrices spanning V and W, n-dimensional of $\mathbb{K}^J$. ($n \leq \# I \leq \# J = d$)

$$\| P_V - P_W \|_F = \| V V^H - W W^H \|_F = \sqrt{2} \, \| \sin \Theta(V, W) \|_2$$

(26) $W^\perp = \{ v \in \mathbb{K}^J : \langle v, w \rangle = 0 \ \forall \ w \in W \}$

$P_W = 1 - P_{W_\perp}$ $\quad$ $W W^H = I - W_\perp W_\perp^H$

(27) Proposition : let $V, W \in \mathbb{K}^{J \times I}$ be orthogonal matrices spanning V and W, n dim subspace of $\mathbb{K}^J$ ($n = \# I \leq \# J = d$)
$W_\perp$ orthogonal matrix spanning $W_\perp$.

$$\| W_\perp^H V \|_F = \| \sin \Theta(V, W) \|_2$$

$$\| P_{W_\perp} P_V \|_F = \| \sin \Theta(V, W) \|_2$$

(28) $\alpha \geq 0$, $\delta > 0$ s.t. $\sigma_k(\tilde{A}) \geq \alpha + \delta$ $\quad$ $\sigma_{k+1}(A) \leq \alpha$
the for unilaterally invariant norms $*$ (Frobenuis & spectral)

$$\max \{ \| \sin \Theta(\tilde{V}_1, V_1) \|_*, \| \sin \Theta(\tilde{U}_1, U_1) \|_* \} \leq \frac{\max \{ \| R_{11} \|_*, \| R_{21} \|_* \}}{\delta}$$

$$\leq \| E \|_* / \delta$$

Chapter 3 : Probability

① $X(\omega) = \frac{\omega}{\|\omega\|_2} \in S^1$

$X : \Omega \to \mathbb{R}$

$P(\omega \in A) = \frac{\mathrm{Vol}(A)}{\mathrm{Vol}(\Omega)}$

$A = \{ \omega \in \Omega : \theta_1 \leq X(\omega) \leq \theta_2 \}$

$P(A) = \frac{\theta_2 - \theta_1}{2\pi}$

② $(\Omega, \Sigma, \mathbb{P})$ probability space

Σ → σ-algebra on sample space

$\mathbb{P}$ → probability measure

$\mathbb{P}(B) = \int_B dP(\omega) = \int_\Omega I_B(\omega)\, d\mathbb{P}(\omega)$

③ r.v X is a real-valued measurable function on (Ω, Σ)

④ $F(t) = P(X \leq t)$, $t \in \mathbb{R}$

⑤ r.v X has probability density function $\phi : \mathbb{R} \to \mathbb{R}_+$ $P(a < X \leq b) = \int_a^b \phi(t)\, dt$

⑥ $\phi = \frac{d}{dt} F(t)$

⑦ $\mathbb{E} X = \int_\Omega X(\omega)\, d\mathbb{P}(\omega)$

$$\mathbb{E}\, g(X) = \int_{-\infty}^{\infty} g(t)\, \phi(t)\, dt$$

⑧ $(\mathbb{E}|X+Y|^p)^{1/p} \leq (\mathbb{E}|X|^p)^{1/p} + (\mathbb{E}|Y|^p)^{1/p}$

⑨ Hölders inequality : for r.v X & Y on common prob. space and $p, q \geq 1$ with $\frac{1}{p} + \frac{1}{q} = 1$

we have,

$$|\mathbb{E} XY| \leq (\mathbb{E}|X|^p)^{1/p} (\mathbb{E}|Y|^q)^{1/q}$$

⑩ absolute moments of a random variable X can be expressed as :

$$\mathbb{E}|X|^p = p \int_0^\infty \mathbb{P}(|X| \geq t)\, t^{p-1}\, dt, \quad p > 0$$

⑪ For r.v X, the expectation satisfies

$$\mathbb{E} X = \int_0^\infty \mathbb{P}(X \geq t)\, dt - \int_0^\infty \mathbb{P}(X \leq -t)\, dt$$

⑫ Let X be a r.v, then

$$\mathbb{P}(|X| \geq t) \leq \frac{\mathbb{E}|X|}{t}$$

(13) $\mathbb{P}(|X| \geq t) = \mathbb{P}(|X|^p \geq t^p) \leq t^{-p}\, \mathbb{E}|X|^p$

$p=2$ is called Chebyschev's inequality

(14) Laplace transform / moment generating f^n

$P(X \geq t) = \mathbb{P}(\exp(\theta X) \geq \exp(\theta t)) \leq \exp(-\theta t)\, \mathbb{E}\exp(\theta X)$

(15) gaussian random variable X

$$\varphi(t) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right)$$

(16) Joint probability distribution

$F(t_1, t_2, \dots t_n) = \mathbb{P}(X_1 \leq t_1, \dots X_n \leq t_n),\ t_1, \dots t_n \in \mathbb{R}$

(17) $\mathbb{P}(X \in D) = \int_D \phi(t_1, \dots t_n)\, dt_1 \dots dt_n$

(18) in (16) if r.v are independent

$P(X_1 \leq t_1, \dots X_n \leq t_n) = \prod_{i=1}^{n} \mathbb{P}(X_i \leq t_i)$

$\mathbb{E}\left[\prod_{i=1}^{n} X_i\right] = \prod_{i=1}^{n} \mathbb{E}[X_i]$

$\phi(t_1, \dots t_n) = \phi_1(t_1) \times \dots \times \phi(t_n)$

(19) $\phi_{X+Y}(t) = (\phi_X * \phi_Y)(t) = \int_{-\infty}^{\infty} \phi_X(u)\, \phi_Y(t-u)\, du$

(20) Let $f: \mathbb{R}^n \to \mathbb{R}$ be a convex function, and let $X \in \mathbb{R}^n$ be a random vector. Then,

$$f(\mathbb{E}X) \leq \mathbb{E} f(X)$$

(22) Cramér's Theorem and Hoeffding's inequality

cumulant generat-ing fn $C_X(\theta) = \ln \mathbb{E} \exp(\theta X)$

→ Cramér's theorem

(23) Thm: Let $X_1, .. X_M$ be a sequence of independent (real valued) random variables, with cumulant generating function C_{X_l}, $l \in [M]$, then, for $t > 0$

$$\mathbb{P}\left(\sum_{l=1}^{M} X_l \geq t\right) \leq \exp\left(\inf_{\theta>0} \left\{-\theta t + \sum_{l=1}^{M} C_{X_l}\right\}\right)$$

(24) Hoeffding's inequality

Thm Let $X_1, .. X_M$ be a sequence of independent r.v such that $\mathbb{E} X_l = 0$ and $|X_l| \leq B_l$ almost surely $l \in [M]$ then

$$\mathbb{P}\left(\sum_{l=1}^{M} X_l \geq t\right) \leq \exp\left(-\frac{t^2}{2\sum_{l=1}^{M} B_l^2}\right)$$

and consequently

$$\mathbb{P}\left(\left|\sum_{l=1}^{M} X_l\right| \geq t\right) \leq 2\exp\left(-\frac{t^2}{2\sum_{l=1}^{M} B_l^2}\right)$$

25

Bernstien's Theorem :

Thm 3.9 Let $x_1, \ldots x_M$ be independent mean zero random variables such that, all integers $n \geq 2$

$$\mathbb{E}|x_\ell|^n \leq \frac{n!}{2} R^{n-2} \sigma_\ell^2 \quad , \forall \ell \in [M]$$

for some constant $R > 0$ and $\sigma_\ell > 0$, $\ell \in [M]$. Then, for all $t > 0$

$$\mathbb{P}\left(\left| \sum_{\ell=1}^{M} x_\ell \right| \geq t \right) \leq 2\exp\left(- \frac{t^2}{2(\sigma^2 + Rt)} \right)$$

where $\sigma^2 = \sum_{\ell=1}^{M} \sigma_\ell^2$

JL Theorem :

4.1 Theorem: For any $0 < \varepsilon < 1$ and any integer n, let k be a positive integer such that

$$k \geq 2\beta\left(\frac{\varepsilon^2}{2} - \frac{\varepsilon^3}{3}\right)^{-1} \ln(n),$$

for some $\beta \geq 2$, The for any set P of n points in $\mathbb{R}^d$, there is a map $f: \mathbb{R}^d \to \mathbb{R}^k$ s.t $\forall\, v, w \in P$

$$(1-\varepsilon)\|v-w\|_2^2 \leq \|f(v)-f(w)\|_2^2 \leq (1+\varepsilon)\|v-w\|_2^2$$

Furthermore, the map can be generated at random with probability $1-n^{2-\beta}-n^{1-\beta}$.

Functions fullfilling (4.1) are called JL embeddings

Lemma: Let $k < d$, then

(a) if $\alpha < 1$, then

$$\mathbb{P}\left(L \leq \alpha\frac{k}{d}\right) \leq \alpha^{k/2}\left(1+\frac{(1-\alpha)k}{d-k}\right)^{\frac{d-k}{2}} \leq \exp\left(\frac{k}{2}(1-\alpha+\ln\alpha)\right)$$

(b) if $\alpha > 1$ then

$$\mathbb{P}\left(L \geq \alpha\frac{k}{d}\right) \leq \alpha^{k/2}\left(1+\frac{(1-\alpha)k}{d-k}\right)^{\frac{(d-k)}{2}} \leq \exp\left(\frac{k}{2}(1-\alpha+\ln\alpha)\right)$$

4.3 Theorem: For any $0 < \varepsilon < 1/2$ and any integer n, let K be a positive integer and such that

$$K \geq \beta \varepsilon^{-2} \ln n.$$

Then for any set P of n points in $\mathbb{R}^d$, there is a map $f : \mathbb{R}^d \to \mathbb{R}^K$ s.t $\forall\ v, w \in P$.

$$(1-\varepsilon)\|v-w\|_2^2 \leq \|f(v) - f(w)\|_2^2 \leq (1+\varepsilon)\|v-w\|_2^2$$

Furthermore, the map can be generated at random with probability $1 - (n^{2-\beta(1-\varepsilon)} - n^{1-\beta(1-\varepsilon)})$

Lemma 4.4

$$P\left(\left| \left\| \frac{1}{\sqrt{K}} Ax \right\|_2^2 - \|x\|_2^2 \right| > \varepsilon \|x\|_2^2 \right) \leq 2e^{-(\varepsilon^2 - \varepsilon^3)K/4}$$

or equivalently,

$$P\left((1-\varepsilon)\|x\|_2^2 \leq \left\| \frac{1}{\sqrt{K}} Ax \right\|_2^2 \leq (1-\varepsilon)\|x\|_2^2 \right) \geq 1 - 2e^{-(\varepsilon^2 - \varepsilon^3)\frac{K}{4}}$$

Convex Analysis :

① convex sets : subset $K \subseteq \mathbb{R}^N$ is called a convex if for all $x, z \in K$

$$tx + (1-t)z \in K \quad \forall\ t \in [0,1]$$

② convex hull : Let $T \in \mathbb{R}^N$ be a set, it's convex hull $conv(T)$ is the smallest convex set containing T.

③ Cone : $K \subset \mathbb{R}^N$ is called a cone if $\forall\ x \in K$ and $\forall\ t \geq 0$ also tx is contained in K.

④ convex functions :

Defn : An extended valued function $F : \mathbb{R}^N \to (-\infty, \infty]$ is called convex if $\forall\ x, z \in \mathbb{R}^N$ and $t \in [0,1]$

$$F(tx + (1-t)z) \leq tF(x) + (1-t)F(z)$$

⑤ strongly convex function :

F is called strongly convex with parameter $\gamma > 0$ if $\forall\ x, z \in \mathbb{R}^N$ and $t \in [0,1]$

$$F(tx + (1-t)z) \leq tF(x) + (1-t)F(z) - \frac{\gamma}{2}(1-t)\|x - z\|_2^2$$

⑥ epigraph : A function is convex iff it's epigraph

$$epi(F) = \{(x, r) : r \geq F(x)\} \subset \mathbb{R}^N \times \mathbb{R}$$

is a convex set.

7. F is convex if and only if $\forall\, x, y \in \mathbb{R}^N$

$$F(x) \geq F(y) + \langle \nabla F(y), x-y \rangle$$

where gradient is defined as $\nabla F y = (\partial_{y_1} F(y) \dots \partial_{y_n} F(y))^T$

8. F is strongly convex with parameter $\gamma > 0$ if and only if $x, y \in \mathbb{R}^N$

$$F(x) \geq F(y) + \langle \nabla F(y), x-y \rangle + \frac{\gamma}{2} \|x-y\|_2^2$$

9. assume that F is twice differentiable. then F is convex if and only if

$$\nabla^2 F(x) \succcurlyeq 0$$

$\forall\, x \in \mathbb{R}^N$, where $\nabla^2 F$ is the Hessian of F

10. if F & G are convex function on $\mathbb{R}^N$. Then for $\alpha, \beta \geq 0$, the function $\alpha F + \beta G$ is also convex.

11. Let $F : \mathbb{R}^N \to \mathbb{R}$ be convex and nondecreasing and $G : \mathbb{R}^N \to \mathbb{R}$ be convex, then $H(x) = F(G(x))$ is convex.

12. every norm $\|\cdot\|$ on $\mathbb{R}^N$ is convex by triangle inequality and homogenity

13. l_p-norms are strictly convex for $1 < p < \infty$ and not strictly convex for $p = 1, \infty$.

14. $A \in \mathbb{R}^{N \times N}$ +ve semi definite, $F(x) = x^T A x$ is convex

(15) $f_K(x) = \begin{cases} 0, & x \in K \\ \infty, & \text{else} \end{cases}$ {for convex set K}

characteristic fn

(16) $F: \mathbb{R}^N \to \mathbb{R}$ be a convex function. Then F is continuous on $\mathbb{R}^N$

(17) $F: \mathbb{R}^N \to (-\infty, \infty]$ is called lower semicontinuous if for all $x \in \mathbb{R}^N$ and every sequence $(x_j)_{j \in \mathbb{N}} \subset \mathbb{R}^N$ converging to x it holds

$$\lim_{j \to \infty} F(x_j) \geq F(x)$$

- nontrivial example is $f_K(x)$
- function is lower semi continuous if the epigraph is closed.

(18) $F: \mathbb{R}^N \to (-\infty, \infty]$ be convex,
(a) local minimum of F is a global minimum
(b) the set of minima of F is convex
(c) if F is strictly convex the minima is unique

(19) Let $K \subset \mathbb{R}^N$ be a compact convex set, and $F: K \to \mathbb{R}$ be a convex function. Then F attains its maximum at an extreme point of K.

(20) convex conjugate:
Defn: $F: \mathbb{R}^N \to (-\infty, \infty]$. Then it's convex conjugate (or Fenchel dual) function $F^*: \mathbb{R}^N \to (-\infty, \infty]$ is defined by

$$F^*(y) = \sup_{x \in \mathbb{R}^N} \{ \langle x, y \rangle - F(x) \}$$

(21) Fenchel inequality ***

$$\langle x, y\rangle \leq F(x) + F^*(y) \quad \forall\, x, y \in \mathbb{R}^N$$

(22) Let $F: \mathbb{R}^N \to (-\infty, \infty]$

- convex conjugate F^* is lower semi-continuous
- For $\tau \neq$ $F_\tau(x) = F(\tau x)$. Then, $(F_\tau)^*(y) = F^*\left(\frac{y}{\tau}\right)$
- $\tau > 0$, $(\tau F)^*(y) = \tau F^*\left(\frac{y}{\tau}\right)$
- $F^z = F(x-z)$. Then, $(F^{(z)})^*(y) = \langle z, y\rangle + F^*(y)$

(24) subdifferential

Defn: The subdifferential of a convex function $F: \mathbb{R}^N \to \mathbb{R}$ at a point $x \in \mathbb{R}^N$ is defined by

$$\partial F(x) = \{ v \in \mathbb{R}^N : F(z) - F(x) \geq \langle v, z-x\rangle \ \forall\, z \in \mathbb{R}^N \}$$

The element of $\partial F(x)$ are called subgradient of F at x

convex functions always have a subdifferential

(24.1) A vector x is minimum if and only if $0 \in \partial F(x)$

(25) Theorem: Let $F: \mathbb{R}^N \to (-\infty, \infty]$ be a convex function and $x, y \in \mathbb{R}^N$, the following conditions are equivalent

(a) $y \in \partial F(x)$
(b) $F(x) + F^*(y) = \langle x, y\rangle$

if additionally F is lower semicontinuous then (a) and (b) are equivalent to

(c) $x \in \partial F^*(y)$

(26) Proximal Mapping

let $F : \mathbb{R}^N \to (-\infty, \infty]$ be a convex function

$$x \longmapsto F(x) + \frac{1}{2} \|x - z\|_2^2$$

is strictly convex due to the strict convexity of $x \longmapsto \|x\|_2^2$, the mapping

$$P_F(z) = \arg\min \left\{ F(x) + \frac{1}{2} \|x - z\|_2^2 : x \in \mathbb{R}^N \right\}$$

is called the proximal mapping associated with F.

in special case, $F = \chi_K$ is the charachteristic fun-
-ction of a convex set K, then $P_K = P_{\chi_K}$ is
is orthogonal projection onto K.

$$P_K(z) = \arg\min_{x \in K} \|x - z\|_2$$

Let $F : \mathbb{R}^N \to (-\infty, \infty]$ be a convex function. Then $x = P_F(z)$ if and only if $z \in x + \partial F(x)$

(27) Moreau's Identity

: let $F : \mathbb{R}^N \to (-\infty, \infty]$ be a lower subcontinuous convex functions. Then for all $z \in \mathbb{R}^N$

$$P_F(z) + P_{F^*}(z) = z$$

using Moreau's identity $P_{\tau F}(z) + P_{(\tau F)^*}(z) = z$

$$P_{\tau F}(z) + \tau P_{\tau^{-1} F^*}\left(\frac{z}{\tau}\right) = z$$

(28) for convex function $F: \mathbb{R}^N \to (-\infty, \infty]$ the proximal mapping P_F is a contraction

$$\|P_F(z) - P_F(z')\|_2 \leq \|z - z'\|_2 \quad \forall z, z' \in \mathbb{R}^N$$

(29) <u>convex optimization</u>

$$\min_{x \in \mathbb{R}^N} \quad \text{s.t} \quad Ax = y$$
$$F_j(x) \leq b_j, \quad j \in [M] = \{1, \dots M\}$$

(30) <u>Lagrange function of an optimization problem</u>:

$$L(x, \xi, \nu) = F_0(x) + \xi^*(Ax - y) + \sum_{l=1}^{M} \nu_l (F_l(x) - b_l)$$

an optim problem without inequality constr.

$$L(x, \xi) = F_0(x) + \xi^*(Ax - y)$$

$\xi, \nu \to$ Lagrange multipliers

(31) <u>Lagrange dual function</u>:

$$H(\xi, \nu) = \inf_{x \in \mathbb{R}^N} L(x, \xi, \nu) \qquad \xi \in \mathbb{R}^m, \nu \in \mathbb{R}^M, \ \nu \succeq 0$$

if $x \mapsto L(x, \xi, \nu)$ is unbounded from below, then we set $H(\xi, \nu) = -\infty$

Dual function is always point wise infimum of a family of affine functions, even if original problem is not convex.

Dual function is bounded by optimal value of primal $F_0(x^{\#})$

$$L(x, \xi, \nu) = F_0(x) + \xi^*(Ax - y) + \sum_{l=1}^{M} \nu_l (F_l(x) - b_l) \leq F_0(x)$$

tight lower bound, we can also use.

$$\max H(\xi, \nu) \quad s.t \quad \nu \succeq 0$$

$\#$ $H(\xi^{\#}, \nu^{\#}) \leq F(x^{\#})$ (weak duality)

$\#$ $H(\xi^{\#}, \nu^{\#}) = F(x^{\#})$ (strong duality)

(32) Slater's constraint qualification : assume that $F_0, F_1, \ldots F_M$ are convex function with $\text{dom}(F_0) = \mathbb{R}^N$.
If $\exists$ $x \in \mathbb{R}^N$ s.t $Ax = y$ and $F_l(x) < b_l$ $\forall$ $l \in [M]$, then strong duality holds for the optimization holds for the optimization problem.

(33) Primal dual gap :

$$E(x, \xi, \nu) = F(x) - H(\xi, \nu)$$

for strong duality $E(x^{\#}, \xi^{\#}, \nu^{\#})$