

If you want to edit or correct this summary, you can find the source LaTeX code here: <https://github.com/alronz/tum-MA4800-fda-summary>

1 General

- $\log_b(a^c) = c \log_b(a)$
- **Unitary Matrix:** $A^T A = A A^T = I, A^T = A^{-1}, A^+ = A^{-1}, A^+ = A^T$.
- Homogeneity: $\|\lambda v\| = |\lambda| \|v\|$
- Triangle Inequality: $\|v + w\| \leq \|v\| + \|w\|$
- Inverse Triangle Inequality: $|\|v\| - \|w\|| \leq \|v - w\|$
- **Linearity in First Argument:** $\langle u + \lambda v, w \rangle = \langle u, w \rangle + \lambda \langle v, w \rangle$
- **Linearity in Second Argument:** $\langle w, u + \lambda v \rangle = \langle w, u \rangle + \lambda \langle w, v \rangle$
- $\|v\| = \sqrt{\langle v, v \rangle}$
- $|\langle v, w \rangle| \leq \|v\| \|w\|$
- $\|v\|_2 = \sqrt{\sum_{i \in I} |v_i|^2}$
- Orthonormal Set: $\|v_i\| = 1$ and $\langle v_i, v_j \rangle = 0$ for $i \neq j$
- **Cauchy Sequence:** is a sequence $\{v_n\}$ where $\|v_n - v_m\|$ becomes arbitrarily small for sufficiently large n and m .
- C is convex if $tv + (1 - t)w \in C$ for all $v, w \in C$ and $t \in [0, 1]$.
- $P_C(v) = \arg \min_{w \in C} \|v - w\|$
- $\langle z - P_C(v), v - P_C(v) \rangle \leq 0$ for all $z \in C$.
- If W has an orthonormal basis $\{w_\nu\}_{\nu \in F}$:

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

- For $v \in V$:

$$\|P_W(v)\|^2 = \sum_{\nu \in F} |\langle v, w_\nu \rangle|^2.$$

- $P_V(v) = v$ when $W = V$.
- $v = \sum_{\nu \in F} \langle v, w_\nu \rangle w_\nu$.
- $\|v\|^2 = \sum_{\nu \in F} |\langle v, w_\nu \rangle|^2$.
- Linearity: $\text{tr}(\alpha A + \beta B) = \alpha \text{tr}(A) + \beta \text{tr}(B)$
- Cyclic Property: $\text{tr}(AB) = \text{tr}(BA)$
- Sum of Eigenvalues: $\text{tr}(A) = \sum_{i=1}^n \lambda_i$

- **Frobenius Norm:**

$$\|A\|_F = \sqrt{\sum_{i \in I} \sum_{j \in J} |A_{ij}|^2}$$

- $\|A\|_F^2 = \text{tr}(AA^H) = \text{tr}(A^H A)$

- $\|A\| = \|A\|_{X \rightarrow Y} = \sup_{z \neq 0} \frac{\|Az\|_Y}{\|z\|_X}$

- $\|AB\|_F \leq \|A\|_F \|B\|_F$

- $\|AB\| \leq \|A\| \|B\|$

- **PARALLELOGRAM LAW**

$$\|x + y\|_2^2 = 2\|x\|_2^2 + 2\|y\|_2^2 - \|x - y\|_2^2$$

$$\|x - y\|_2^2 = 2\|x\|_2^2 + 2\|y\|_2^2 - \|x + y\|_2^2$$

$$(x - y)^2 = 2x^2 + 2y^2 - (x + y)^2$$

$$\|x + y\|_2^2 = \|x\|_2^2 + \|y\|_2^2 + 2\langle x, y \rangle$$

$$\|x - y\|_2^2 = \|x\|_2^2 + \|y\|_2^2 - 2\langle x, y \rangle$$

- **Taylor Expansions**

- **First-Order:**

$$f(x_0 + t) = f(x_0) + f'(x_0) \cdot t + o(t), \quad \text{as } t \rightarrow 0$$

$$\lim_{t \rightarrow 0} \frac{o(t)}{t} = 0$$

- **Second-Order Taylor**

$$f(x_0 + t) = f(x_0) + f'(x_0) \cdot t + \frac{1}{2} f''(x_0) \cdot t^2 + o(t^2), \quad \text{as } t \rightarrow 0$$

$$\lim_{t \rightarrow 0} \frac{o(t^2)}{t^2} = 0$$

- **Taylor Expansion with Mean Value:**

- **First-Order:**

$$f(x_0 + t) = f(x_0) + f'(z) \cdot t, \quad z \in (x_0, x_0 + t)$$

– **Second-Order**

$$f(x_0 + t) = f(x_0) + f'(x_0) \cdot t + \frac{1}{2} f''(z) \cdot t^2, \quad z \in (x_0, x_0 + t)$$

– **Multidimensional**

$$f(x_0 + ty) = f(x_0) + \nabla f(x_0) \cdot y + \frac{1}{2} t^2 y^T \nabla^2 f(x_0) y + o(t^2), \quad \text{as } t \rightarrow 0$$

- $P_V = VV^T \in \mathbb{R}^{d \times d}$
- $P_V \cdot x = P_V(x)$
- $(f * g)(t) = \int_{-\infty}^{\infty} f(u)g(t-u) du$
- $\phi_{X+Y}(t) = (\phi_X * \phi_Y)(t) = \int_{-\infty}^{\infty} \phi_X(u)\phi_Y(t-u) du$
- Quadratic Function: $f(x) = x^T A x + \langle b, x \rangle + c$:
 - $\nabla f(x) = Ax + A^T x + b$ (For symmetric A : $\nabla f(x) = 2Ax + b$)
 - $\nabla^2 f(x) = A + A^T$ (For symmetric A : $\nabla^2 f(x) = 2A$)
- $\nabla \|x\|_2 = \frac{x}{\|x\|_2}$ for $x \neq 0$.
- $\nabla \|x\|_2^2 = 2x$.
- $B^m = \sum_{k=1}^r \sigma_k^{2m} u_k u_k^T$

2 SVD

- SVD

$$A = \sum_{i=1}^r \sigma_i u_i v_i^T$$

$$A = U \Sigma V^T$$

$$U^T U = I$$

$$U U^T = I$$

$$V^T V = I$$

$$V V^T = I$$

$$\Sigma \Sigma^{-1} = I$$

$$\Sigma^{-1} \Sigma = I$$

- $\sigma_1(A) = \|A v_1\|_2$

- $v_k = \arg \max_{\|x\|_2=1, \langle v_1, v_2, \dots, v_{k-1} \rangle = 0} \|Av\|_2$
- $\|A^{(i)}\|_2^2 = \sum_{k=1}^r |\langle A^{(i)}, v_k \rangle|^2$
- $\sum_{i \in I} \|A^{(i)}\|_2^2 = \sum_{k=1}^r \|A_{\cdot k}\|_2^2 = \sum_{k=1}^r \sigma_k(A)^2$
- $\|A\|_F^2 = \sum_{i \in I} \sum_{j \in J} |A_{ij}|^2 = \sum_{k=1}^r \sigma_k(A)^2$
- $\|A\|_F = \sqrt{\sum_{k=1}^r \sigma_k(A)^2}$
- $\|A\| = \sigma_1(A) = \max_{\|v\|_2=1} \|Av\|_2$
- Frobenius Norm ($p = 2$):

$$\|A\|_F = \left(\sum_{k=1}^r \sigma_k(A)^2 \right)^{1/2}$$

- Spectral Norm ($p = \infty$):

$$\|A\|_\infty = \max_{k=1, \dots, r} \sigma_k(A) = \sigma_1(A)$$

- Nuclear Norm ($p = 1$):

$$\|A\|_* = \|A\|_1 = \sum_{k=1}^r \sigma_k(A)$$

- A matrix A is positive semi-definite if:

$$\langle x, Ax \rangle \geq 0 \quad \forall x \in \mathbb{R}^m.$$

- A matrix A is positive definite if:

$$\langle x, Ax \rangle > 0 \quad \forall x \in \mathbb{R}^m, x \neq 0.$$

- $A^+ = V\Sigma^{-1}U^H$
- $A^+ = \sum_{k=1}^r \sigma_k^{-1} v_k u_k^H$
- if ($m > n$) (long matrix)
 $A^+ = (A^H A)^{-1} A^H$ and $x = A^+ y = (A^H A)^{-1} A^H y$
- if ($m < n$) (fat matrix)
 $A^+ = A^H (A A^H)^{-1}$ and $x = A^+ y = A^H (A A^H)^{-1} y$
- Weyl's Bounds

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

$$|\sigma_k(A + E) - \sigma_k(A)| \leq \|E\|$$

- Mirsky's Bounds

$$\sum_{k=1}^n |\sigma_k(A + E) - \sigma_k(A)|^2 \leq \|E\|_F^2$$

- $\cos \theta(V, W) = \frac{V^T W}{\|V\|_2 \|W\|_2}$

3 Probability

- Random variable X has PDF

$$P(a < X \leq b) = \int_a^b \varphi(t) dt \quad \text{for all } a < b \in \mathbb{R}$$

- **Relationship with Distribution Function:**

$$\varphi(t) = \frac{d}{dt} F(t)$$

where the cumulative distribution function (CDF) $F(x)$ is defined as:

$$F(x) = P(X \leq x) = \int_{-\infty}^x \frac{d}{dt} F(t) dt$$

- Expectation (Mean)

$$E[X] = \int_{\Omega} X(\omega) dP(\omega)$$

$$E[g(X)] = \int_{-\infty}^{\infty} g(t) \varphi(t) dt$$

$$E[X] = \int_{-\infty}^{\infty} t \varphi(t) dt$$

- Moments

$$E[X^p] \quad \text{for } p > 0$$

- Absolute Moments

$$E[|X|^p] \quad \text{for } p > 0$$

- variance

$$Var(X) = E[(X - E[X])^2] = E[X^2] - (E[X])^2$$

- LP Norm:

$$\|X\|_p = (E[|X|^p])^{1/p} \quad \text{for } 1 \leq p < \infty$$

- Inequality:

$$\|X + Y\|_p \leq \|X\|_p + \|Y\|_p$$

- Hölder's Inequality

$$|E[XY]| \leq \|X\|_p \|Y\|_q$$

- Convergence of Random Variables

$$\lim_{n \rightarrow \infty} X_n(\omega) = X(\omega) \quad \text{for all } \omega \in \Omega$$

- Lebesgue's Dominated Convergence Theorem (LDCT)

$$\lim_{n \rightarrow \infty} E[X_n] = E\left[\lim_{n \rightarrow \infty} X_n\right] = E[X]$$

- Fubini's Theorem

$$\int_A \left(\int_B f(x, y) d\mu(y) \right) d\nu(x) = \int_B \left(\int_A f(x, y) d\nu(x) \right) d\mu(y)$$

- Absolute Moments

$$E[|X|^p] = p \int_0^\infty P(|X| \geq t) t^{p-1} dt$$

- $(\mathbb{E}|X + Y|)^p \leq (\mathbb{E}|X|^p) + (\mathbb{E}|Y|^p)$
- Cavalieri's Formula for Expectation

$$E[X] = \int_0^\infty P(X \geq t) dt - \int_0^\infty P(X \leq -t) dt$$

- Markov's Inequality

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

- Generalized Markov Inequality

$$P(|X| \geq t) = P(|X|^p \geq t^p) \leq \frac{E[|X|^p]}{t^p} \quad \text{for all } t > 0$$

- Chebyshev Inequality

$$P(|X - E[X]| \geq t) \leq \frac{\text{Var}(X)}{t^2} \quad \text{for all } t > 0$$

- Laplace Transform (Moment Generating Function)

$$M_X(\theta) = E[e^{\theta X}]$$

- Normal Distribution (Gaussian Distribution)

–

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

– MGF: $M_X(\theta) = \exp\left(\mu\theta + \frac{1}{2}\sigma^2\theta^2\right)$

– μ : Mean

– σ^2 : Variance

– Mean: $E[X] = \mu$

– Variance: $\text{Var}(X) = \sigma^2$

$$X \sim N(\mu, \sigma^2)$$

– **Expectation:**

$$\mathbb{E}[X] = \frac{1}{N} \sum_{i=1}^N X_i$$

– **Variance:**

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

Alternatively,

$$\text{Var}(X) = \frac{1}{N} \sum_{i=1}^N X_i^2 - \left(\frac{1}{N} \sum_{i=1}^N X_i\right)^2$$

– **Standard Deviation:**

$$\sigma_X = \sqrt{\text{Var}(X)}$$

- **for normal distribution PDF:**

– $\phi(t) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right)$

– MGF: $M_X(\theta) = \exp\left(\frac{1}{2}\theta^2\right)$

- $P(X_1 \leq t_1, \dots, X_n \leq t_n) = \prod_{\ell=1}^n P(X_\ell \leq t_\ell)$

- $E[\prod_{\ell=1}^n X_\ell] = \prod_{\ell=1}^n E[X_\ell]$

- join PDF

$$\varphi(t_1, \dots, t_n) = \varphi_1(t_1) \cdot \varphi_2(t_2) \cdots \varphi_n(t_n)$$

- PDF of $X+Y$

$$\varphi_{X+Y}(t) = (\varphi_X * \varphi_Y)(t) = \int_{-\infty}^{\infty} \varphi_X(u) \varphi_Y(t-u) du$$

- Fubini's Theorem for Expectations

$$E[|f_1(X)|] = E[|f_2(Y)|] = E[|f(X,Y)|]$$

- Multivariate Normal Distribution (Gaussian Vector)

$$X = Ag + \mu$$

$$\Sigma = AA^T = E[(X - \mu)(X - \mu)^T]$$

$$\psi(x) = \frac{1}{(2\pi)^{n/2} \sqrt{\det(\Sigma)}} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right)$$

- Jensen's Inequality

– **For Convex Functions:**

$$f(E[X]) \leq E[f(X)]$$

– **Convex Function:** $f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)$ for all $x, y \in \mathbb{R}^n$ and $\lambda \in [0, 1]$.

– **For Concave Functions:**

$$E[f(X)] \leq f(E[X])$$

– **Concave Function:** $f(\lambda x + (1 - \lambda)y) \geq \lambda f(x) + (1 - \lambda)f(y)$ for all $x, y \in \mathbb{R}^n$ and $\lambda \in [0, 1]$.

- Moment Generating Function (MGF)

$$M_X(\theta) = E[\exp(\theta X)]$$

- Cumulant Generating Function (CGF)

$$K_X(\theta) = \ln(E[\exp(\theta X)])$$

- Cramer's Theorem

$$P\left(\sum_{\ell=1}^M X_{\ell} \geq t\right) \leq \exp\left(\inf_{\theta > 0} \left\{-\theta t + \sum_{\ell=1}^M C_X(\theta)\right\}\right)$$

- Hoeffding's Inequality

- One-Sided Bound:

$$\mathbb{P}\left(\sum_{\ell=1}^M X_{\ell} \geq t\right) \leq \exp\left(-\frac{t^2}{2\sum_{\ell=1}^M B_{\ell}^2}\right)$$

- Two-Sided Bound:

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

- For a single symmetric Bernoulli random variable, Hoeffding's inequality states:

$$\mathbb{P}(|X| > t) \leq 2\exp\left(-\frac{2t^2}{(b-a)^2}\right)$$

- Bernstein's Inequality

- For all $t > 0$:

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

- Where $\sigma^2 = \sum_{t=1}^M \sigma_t^2$.

- Johnson-Lindenstrauss Lemma

$$k \geq \beta\epsilon^{-2} \log(2n)$$

$$(1 - \epsilon)\|v - w\|_2^2 \leq \|f(v) - f(w)\|_2^2 \leq (1 + \epsilon)\|v - w\|_2^2 \quad \text{for all } v, w \in P$$

$$|\langle f(\mathbf{v}), f(\mathbf{w}) \rangle| \leq |\langle \mathbf{v}, \mathbf{w} \rangle|(1 + \epsilon)$$

4 Optimization

- To find global minimized using gradient decent, following conditions needed:

- **Convexity:** f is convex if $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $\lambda \in [0, 1]$:

$$f(\lambda \mathbf{x} + (1 - \lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1 - \lambda)f(\mathbf{y}).$$

- **Lipschitz Continuity of Gradient:** There exists $L > 0$ such that $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$:

$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\| \leq L\|\mathbf{x} - \mathbf{y}\|.$$

- **Step Size:** The step size α_k satisfies:

$$0 < \alpha < \frac{2}{L}$$

- **Convex Set:** A set $K \subseteq \mathbb{R}^N$ is convex if for all $x, z \in K$ and $t \in [0, 1]$: $t \cdot x + (1 - t) \cdot z \in K$
- **Convex Hull:** For $T \subseteq \mathbb{R}^N$, the convex hull $\text{conv}(T)$ is the smallest convex set containing T .
- **Cones:** A set $K \subseteq \mathbb{R}^N$ is a cone if for all $x \in K$ and all $t \geq 0$, $t \cdot x \in K$.
- **Convex Cone:** A set K is a convex cone if it is a cone and convex. For all $x, z \in K$ and $s, t \geq 0$, $s \cdot x + t \cdot z \in K$.
- **Dual Cones:** For a cone $K \subseteq \mathbb{R}^N$, the dual cone K^* is:

$$K^* = \{z \in \mathbb{R}^N : \langle x, z \rangle \geq 0 \text{ for all } x \in K\}$$

- Bidual Cone: If K is a closed cone, then $(K^*)^* = K$.
- Polar Cones: For a cone $K \subseteq \mathbb{R}^N$, the polar cone K° is:

$$K^\circ = \{z \in \mathbb{R}^N : \langle x, z \rangle \leq 0 \text{ for all } x \in K\}$$

- $K^\circ = -K^*$, where K^* is the dual cone.
- Geometrical Hahn-Banach Theorem

For convex sets $K_1, K_2 \subseteq \mathbb{R}^N$ with empty interior intersection, there exists a vector $w \in \mathbb{R}^N$ and a scalar λ such that:

$$K_1 \subseteq \{x \in \mathbb{R}^N : \langle x, w \rangle \leq \lambda\}$$

$$K_2 \subseteq \{x \in \mathbb{R}^N : \langle x, w \rangle \geq \lambda\}$$

- Defined as $\text{dom}(F) = \{x \in \mathbb{R}^N : F(x) \neq \infty\}$.
- A function is proper if $\text{dom}(F) \neq \emptyset$.
- Convex Functions

$F : \mathbb{R}^N \rightarrow (-\infty, \infty)$ is convex if for all $x, z \in \mathbb{R}^N$ and $t \in [0, 1]$:

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

- Strictly Convex Functions F is strictly convex if for all $x \neq z$ and $t \in (0, 1)$:

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

- Strongly Convex Functions F is strongly convex with parameter $\gamma > 0$ if for all $x, z \in \mathbb{R}^N$ and $t \in [0, 1]$:

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

- The epigraph of F is $\text{epi}(F) = \{(x, r) : r \geq F(x)\}$.

- F is convex if and only if $\text{epi}(F)$ is a convex set.
- For $(x_1, r_1), (x_2, r_2) \in \text{epi}(F)$ and $t \in [0, 1]$:

$$(tx_1 + (1 - t)x_2, tr_1 + (1 - t)r_2) \in \text{epi}(F)$$

- A differentiable function $F : \mathbb{R}^N \rightarrow \mathbb{R}$ is convex if and only if:

$$F(x) \geq F(y) + \nabla F(y)^T(x - y)$$

- F is strongly convex with parameter $\gamma > 0$ if and only if:

$$F(x) \geq F(y) + \nabla F(y)^T(x - y) + \frac{\gamma}{2}\|x - y\|^2$$

- A twice differentiable function F is convex if and only if its Hessian $\nabla^2 F(x)$ is positive semi-definite:

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

- to check convexity:

- Convexity can be checked using the gradient. If the tangent line (or hyperplane) at any point y lies below the function, F is convex. Condition from above.
- **Strong Convexity:** Strong convexity includes an additional quadratic term that provides a lower bound on the curvature of F .
- **Hessian Condition:** For twice differentiable functions, convexity can be checked by ensuring the Hessian matrix is positive semi-definite at every point.

- Let F, G be convex functions on \mathbb{R}^N . Then, for $\alpha, \beta \geq 0$ the function $\alpha F + \beta G$ is convex.
- Let $F : \mathbb{R} \rightarrow \mathbb{R}$ be convex and nondecreasing, and $G : \mathbb{R}^N \rightarrow \mathbb{R}$ be convex. Then $H(x) = F(G(x))$ is convex.

- Lower Semicontinuity:

$$\liminf_{j \rightarrow \infty} F(x_j) \geq F(x)$$

- **Global Minimum:** A point $x \in \mathbb{R}^N$ is a global minimum of F if:

$$F(x) \leq F(y) \quad \text{for all } y \in \mathbb{R}^N$$

- **Local Minimum:** A point $x \in \mathbb{R}^N$ is a local minimum of F if there exists $\epsilon > 0$ such that:

$$F(x) \leq F(y) \quad \text{for all } y \text{ satisfying } \|x - y\|_2 \leq \epsilon$$

- For a convex function F , any local minimum is also a global minimum.
- The set of minima of a convex function F is convex.
- If F is strictly convex, the minimum is unique.
- For extreme point x , $x = ty + (1 - t)z$ with $t \in (0, 1)$ implies $x = y = z$.
- Convex Conjugate (F^* is always a convex function, even if F is not.)

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

- Fenchel-Young Inequality: For all $x, y \in \mathbb{R}^N$:

$$\langle x, y \rangle \leq F(x) + F^*(y)$$

- Biconjugate:

- F^{**} is the largest lower semicontinuous convex function satisfying $F^{**}(x) \leq F(x)$.
- If F is convex and lower semicontinuous, then $F = F^{**}$.
- Scaling Argument: For $\tau \neq 0$:

$$(F_\tau)^*(y) = F^*\left(\frac{y}{\tau}\right)$$

- Scaling Function: For $\tau > 0$:

$$(\tau F)^*(y) = \tau F^*\left(\frac{y}{\tau}\right)$$

- Translation: For $z \in \mathbb{R}^N$:

$$(F_z)^*(y) = \langle z, y \rangle + F^*(y)$$

- **Subdifferential and Subgradients:** For a convex function $F : \mathbb{R}^N \rightarrow \mathbb{R}$ at $x \in \mathbb{R}^N$,

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

v is a subgradient at x if it satisfies:

$$F(z) - F(x) \geq \langle v, z - x \rangle$$

The subdifferential $\partial F(x)$ is always non-empty for a convex function.

If F is differentiable at x ,

$$\partial F(x) = \{\nabla F(x)\}$$

- A vector x is a minimum of a convex function F if and only if:

$$0 \in \partial F(x)$$

- For a convex lower semicontinuous function F :

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

- Proximal Mapping:

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

The function $x \mapsto F(x) + \frac{1}{2} \|x - z\|_2^2$ is strictly convex, ensuring a unique minimizer.

For a convex function $F : \mathbb{R}^N \rightarrow (-\infty, \infty]$,

$$x = P_F(z) \text{ if and only if } z \in x + \partial F(x)$$

$$P_F = (I + \partial F)^{-1}$$

- Moreau's Identity: For a lower semicontinuous convex function $F : \mathbb{R}^N \rightarrow (-\infty, \infty]$ and all $z \in \mathbb{R}^N$,

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

- Nonexpansiveness of Proximal Mappings:

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

- Lagrange Function:

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

- $x \in \mathbb{R}^N$: Decision variables.
- $\xi \in \mathbb{R}^m$: Lagrange multipliers for equality constraints. **It can take any real values, positive or negative**
- $\nu \in \mathbb{R}^M$: Lagrange multipliers for inequality constraints ($\nu_l \geq 0$).

• Lagrange Dual Function:

$$H(\xi, \nu) = \inf_{x \in \mathbb{R}^N} L(x, \xi, \nu),$$

- **Concavity:** The dual function $H(\xi, \nu)$ is concave.
- **Dual Feasible:** (ξ, ν) is dual feasible if $\xi \in \mathbb{R}^m$ and $\nu \geq 0$.
- **Dual Optimal:** (ξ^*, ν^*) maximizes $H(\xi, \nu)$.
- **Primal-Dual Optimal:** (x^*, ξ^*, ν^*) where x^* is optimal for the primal problem and (ξ^*, ν^*) are optimal for the dual problem.
- Weak Duality: $H(\xi^*, \nu^*) \leq F_0(x^*)$
- Strong Duality: $H(\xi^*, \nu^*) = F_0(x^*)$

• Slater's Constraint Qualification Theorem:

- **Assumption:** F_0, F_1, \dots, F_M are convex functions with $\text{dom}(F_0) = \mathbb{R}^N$.
- **Condition:** There exists $x \in \mathbb{R}^N$ such that:

$$Ax = y \quad F_i(x) < b_i, \quad \forall i \in \{1, \dots, M\}$$

- **Conclusion:** If the above conditions hold, then strong duality holds for the optimization problem.

• Saddle-Point Interpretation:

- **Optimal Value:**

$$F_0(x^*) = \inf_{x \in \mathbb{R}^N} \sup_{\xi \in \mathbb{R}^m} L(x, \xi).$$

- **Supremum:**

$$\sup_{\xi \in \mathbb{R}^m} \inf_{x \in \mathbb{R}^N} L(x, \xi) \leq \inf_{x \in \mathbb{R}^N} \sup_{\xi \in \mathbb{R}^m} L(x, \xi)$$

- **Strong Duality:**

$$\sup_{\xi \in \mathbb{R}^m} \inf_{x \in \mathbb{R}^N} L(x, \xi) = \inf_{x \in \mathbb{R}^N} \sup_{\xi \in \mathbb{R}^m} L(x, \xi)$$

- For a primal-dual optimal pair (x^*, ξ^*) :

$$L(x^*, \xi) \leq L(x^*, \xi^*) \leq L(x, \xi^*)$$

- **Lipschitz Continuity:**

- Gradient $\nabla f(x)$ is Lipschitz continuous with constant L :

$$\|\nabla f(x) - \nabla f(y)\|_2 \leq L\|x - y\|_2$$

- If f is twice differentiable:

$$\lambda_{\max}(\nabla^2 f(x)) \leq L$$

- A convex function f satisfies:

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

For all $x, y \in \mathbb{R}^d$.

- If $\nabla f(x)$ is Lipschitz continuous with constant L , then:

$$f(y) \leq f(x) + \nabla f(x)^T(y - x) + \frac{L}{2}\|y - x\|_2^2$$

- Convergence of Gradient Descent:

$$\alpha = \alpha^{(n)} \leq \frac{1}{L}$$

- **Update Rule:**

$$x^{(n+1)} = x^{(n)} - \alpha^{(n)} \nabla f(x^{(n)})$$

- **Strong Convexity:**

- **Gradient Condition:**

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

- **Gradient Difference Condition:**

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \gamma\|x - y\|_2^2$$

- **Hessian Condition:**

$$w^T \nabla^2 f(x) w \geq \gamma\|w\|_2^2 \quad \forall x, w \in \mathbb{R}^n$$

* Alternatively:

$$\nabla^2 f(x) \succeq \gamma I$$

- **Jensen's Inequality for Strongly Convex Functions:**

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y) - \frac{\gamma}{2}\lambda(1 - \lambda)\|x - y\|_2^2$$

- **Gradient Inequality:** $f(x + v) \geq f(x) + \nabla f(x) \cdot v \quad \forall x, v$

- Use the spectral norm (largest singular value) to estimate the Lipschitz constant.

$$\|\nabla^2 f(x)\| = \text{Largest Singular Value}$$