Derivatives

First Order

- $f:\mathbb{R}^n\to\mathbb{R},\ f(\boldsymbol{x}),\boldsymbol{x}=[x_1,x_2,\cdots,x_n]^T$

  Gradient: $\nabla f(\boldsymbol{x})=\begin{bmatrix}\frac{\partial f}{\partial x_1}\\ \frac{\partial f}{\partial x_2}\\ \vdots\\ \frac{\partial f}{\partial x_n}\end{bmatrix}$

  Derivative: $Df(\boldsymbol{x})=\left[\frac{\partial f}{\partial x_1},\frac{\partial f}{\partial x_2},\cdots,\frac{\partial f}{\partial x_n}\right]$

  - **Relationship:** $\nabla f(\boldsymbol{x})=Df(\boldsymbol{x})^T$
  - **Conclusions** (use gradient):
    $\frac{\partial(\boldsymbol{x}^T\boldsymbol{a})}{\partial\boldsymbol{x}}=\frac{\partial(\boldsymbol{a}^T\boldsymbol{x})}{\partial\boldsymbol{x}}=\boldsymbol{a}$
    $\frac{\partial(\boldsymbol{x}^T\boldsymbol{x})}{\partial\boldsymbol{x}}=2\boldsymbol{x}$
    $\frac{\partial(\boldsymbol{x}^T\mathbf{A}\boldsymbol{x})}{\partial\boldsymbol{x}}=\mathbf{A}\boldsymbol{x}+\mathbf{A}^T\boldsymbol{x}$
    $\frac{\partial(\boldsymbol{a}^T\boldsymbol{x}\boldsymbol{x}^T\boldsymbol{b})}{\partial\boldsymbol{x}}=\boldsymbol{a}\boldsymbol{b}^T\boldsymbol{x}+\boldsymbol{b}\boldsymbol{a}^T\boldsymbol{x}$
    $\nabla_{\mathbf{A}}\operatorname{tr}(\mathbf{AB})=\mathbf{B}^T$

- $\boldsymbol{f}:\mathbb{R}^n\to\mathbb{R}^m,\ \boldsymbol{f}(\boldsymbol{x}),\boldsymbol{x}=[x_1,x_2,\cdots,x_n]^T$

  Derivative: Jacobian

$$D\boldsymbol{f}(\boldsymbol{x})=\begin{bmatrix}Df_1(\boldsymbol{x})\\ Df_2(\boldsymbol{x})\\ \vdots\\ Df_m(\boldsymbol{x})\end{bmatrix}=\begin{bmatrix}(\nabla f_1(\boldsymbol{x}))^{\mathrm{T}}\\ (\nabla f_2(\boldsymbol{x}))^{\mathrm{T}}\\ \vdots\\ (\nabla f_m(\boldsymbol{x}))^{\mathrm{T}}\end{bmatrix}$$

$$=\begin{bmatrix}\frac{\partial f_1}{\partial x_1}&\frac{\partial f_1}{\partial x_2}&\cdots&\frac{\partial f_1}{\partial x_n}\\ \frac{\partial f_2}{\partial x_1}&\frac{\partial f_2}{\partial x_2}&\cdots&\frac{\partial f_2}{\partial x_n}\\ \vdots&\vdots&\ddots&\vdots\\ \frac{\partial f_m}{\partial x_1}&\frac{\partial f_m}{\partial x_2}&\cdots&\frac{\partial f_m}{\partial x_n}\end{bmatrix}$$

  Conclusions: $D(\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b})=\boldsymbol{A}$

- $\boldsymbol{F}:S^n\to\mathbb{R}$

$$\nabla F(\mathbf{X})=\begin{bmatrix}\frac{\partial F}{\partial X_{11}}&\cdots&\frac{\partial F}{\partial X_{1n}}\\ \vdots&\ddots&\vdots\\ \frac{\partial F}{\partial X_{n1}}&\cdots&\frac{\partial F}{\partial X_{nn}}\end{bmatrix}$$

  Conclusions: $\nabla_{\mathbf{X}}(\log\det\mathbf{X})=\mathbf{X}^{-1}$

Second Order

Hessian (always symmetric)

$f:\mathbb{R}^n\to\mathbb{R},\ f(\boldsymbol{x}),\boldsymbol{x}=[x_1,x_2,\cdots,x_n]^T$

$$\nabla^2 f(\boldsymbol{x})=D[\nabla f(\boldsymbol{x})]=\begin{bmatrix}D\left(\frac{\partial f}{\partial x_1}\right)\\ D\left(\frac{\partial f}{\partial x_2}\right)\\ \vdots\\ D\left(\frac{\partial f}{\partial x_n}\right)\end{bmatrix}$$

$$=\begin{bmatrix}\frac{\partial^2 f}{\partial x_1\partial x_1}&\frac{\partial^2 f}{\partial x_1\partial x_2}&\cdots&\frac{\partial^2 f}{\partial x_1\partial x_n}\\ \frac{\partial^2 f}{\partial x_2\partial x_1}&\frac{\partial^2 f}{\partial x_2\partial x_2}&\cdots&\frac{\partial^2 f}{\partial x_2\partial x_n}\\ \vdots&\vdots&\ddots&\vdots\\ \frac{\partial^2 f}{\partial x_n\partial x_1}&\frac{\partial^2 f}{\partial x_n\partial x_2}&\cdots&\frac{\partial^2 f}{\partial x_n\partial x_n}\end{bmatrix}$$

Taylor Approximation

- $f:\mathbb{R}^n\to\mathbb{R},\ f(\boldsymbol{x}),\boldsymbol{x}=[x_1,x_2,\cdots,x_n]^T$

  $f(\boldsymbol{x})\simeq f(\boldsymbol{x}_0)+\nabla f(\boldsymbol{x}_0)^T(\boldsymbol{x}-\boldsymbol{x}_0)+\frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}_0)^T\nabla^2 f(\boldsymbol{x}_0)(\boldsymbol{x}-\boldsymbol{x}_0)$

- $\boldsymbol{f}:\mathbb{R}^n\to\mathbb{R}^m,\ \boldsymbol{f}(\boldsymbol{x}),\boldsymbol{x}=[x_1,x_2,\cdots,x_n]^T$

  $\boldsymbol{f}(\boldsymbol{x})\simeq\boldsymbol{f}(\boldsymbol{x}_0)+D\boldsymbol{f}(\boldsymbol{x}_0)(\boldsymbol{x}-\boldsymbol{x}_0)$

Norm

Norm is a mapping $\|\cdot\|:\mathbb{R}^n\to\mathbb{R}$.

- Scaling: $\|t\boldsymbol{x}\|=|t|\|\boldsymbol{x}\|$
- Triangle inequality: $\|\boldsymbol{x}+\boldsymbol{y}\|\le\|\boldsymbol{x}\|+\|\boldsymbol{y}\|$
- l_p **norm:** $\|\boldsymbol{x}\|_p=\left(\sum_{i=1}^n|x_i|^p\right)^{1/p}$
  - One-norm: $\|\boldsymbol{x}\|_1=\sum_{i=1}^n|x_i|$
  - Two-norm (Euclidean Norm): $\|\boldsymbol{x}\|=\sqrt{\sum_{i=1}^n x_i^2}$
  - Infinity-norm: $\|\boldsymbol{x}\|_\infty=\max_{i=1}^n|x_i|$.
  - **Dual norm**: $\|\boldsymbol{z}\|_*=\sup_{\boldsymbol{x}}\left\{\boldsymbol{z}^{\mathrm{T}}\boldsymbol{x}\mid\|\boldsymbol{x}\|\le 1\right\}$
- **Cauchy-Schwartz Inequality** $\langle\boldsymbol{x},\boldsymbol{y}\rangle\le\|\boldsymbol{x}\|_2\|\boldsymbol{y}\|_2$

Inner Product

- **Vector**: $\langle\boldsymbol{x},\boldsymbol{y}\rangle=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}=\sum_{i=1}^n x_iy_i$
  - Orthogonal $\Leftrightarrow\langle\boldsymbol{x},\boldsymbol{y}\rangle=0$
- **Matrix**: $\langle X,Y\rangle=\operatorname{Tr}\left(X^{\mathrm{T}}Y\right)=\sum_{i=1}^m\sum_{j=1}^n X_{ij}Y_{ij}$
- **Properties:** $f(A,B)=f(B,A)$, $f(A+B,C)=f(A,B)+f(B,C)$, $f(\gamma A,B)=\gamma f(A,B)$, $f(A,A)\ge 0$.
- **Frobenius norm**:
  $\|X\|_F=\operatorname{Tr}\left(X^{\mathrm{T}}X\right)=\sqrt{\langle X,X\rangle}=\sum_{i=1}^m\sum_{j=1}^n X_{ij}^2$

SVD

Such a decomposition is always possible: $A=U\Sigma V^{\mathrm{T}}$

- $A\in\mathbb{R}^{m\times n}$, $U\in\mathbb{R}^{m\times r}$, $V\in\mathbb{R}^{n\times r}$, r is the rank of A.
- $\Sigma=\operatorname{diag}(\sigma_1,\sigma_2,\cdots,\sigma_r)$ where $\sigma_1\ge\sigma_2\ge\cdots\ge\sigma_r$
- U and V and orthogonal: $U^{\mathrm{T}}U=V^{\mathrm{T}}V=I\in\mathbb{R}^{r\times r}$
- Rotate (length i.e. 2-norm unchanged) – scale – rotate.
- $\sigma_{\max}(X)=\max_{\|v\|_2\le 1}\|Xv\|_2$

Spectral Decomposition

For symmetric matrix $A\in S^n$, $A=Q\Lambda Q^{\mathrm{T}}$.

- Q is orthogonal. $\Lambda=\operatorname{diag}(\lambda_1,\lambda_2,\cdots\lambda_n)$, $\lambda_i\in\mathbb{R},\forall i$ (real).
- **PSD**: $A\in S_+^n\Leftrightarrow\boldsymbol{v}^{\mathrm{T}}A\boldsymbol{v}\ge 0\quad\forall\boldsymbol{v}\in\mathbb{R}^n,\boldsymbol{v}\ne 0\Leftrightarrow\lambda_i\ge 0$
- **PD**: $A\in S_{++}^n\Leftrightarrow\boldsymbol{v}^{\mathrm{T}}A\boldsymbol{v}>0\quad\forall\boldsymbol{v}\in\mathbb{R}^n,\boldsymbol{v}\ne 0\Leftrightarrow\lambda_i>0$
- $\boldsymbol{v}_i^{\mathrm{T}}A\boldsymbol{v}_i=\lambda_i$, where $\boldsymbol{v}_i,\lambda_i$ is the i-th eigenvec, eigenval of A.
- $\max_{\|v\|_2\le 1}v^TAv=\lambda_{\max}$, $\quad v^*$ is the eigenvec for $\lambda_{\max}$.
- **Power of symmetric matrix**: $A^k=Q\Lambda^kQ^{\mathrm{T}},k\in\mathbb{Z}$
- **Square Root**: $A^{1/2}=Q\operatorname{diag}\left[\sqrt{\lambda_1}\sqrt{\lambda_2}\cdots\sqrt{\lambda_n}\right]Q^{\mathrm{T}}$

Linear Algebra

Some properties of determinant:

- A matrix $\mathbf{A}$ is non-singular (invertible) iff $\det \mathbf{A} \neq 0$.
- $\det \mathbf{A} = \prod_{n=1}^{N} \lambda_n$, i.e. a matrix is invertible iff $\lambda_n \neq 0, \vee n$.
- $\det \mathbf{AB} = \det \mathbf{A} \det \mathbf{B} = \det \mathbf{BA}$
- $\det \mathbf{A}^{-1} = \frac{1}{\det \mathbf{A}}$
- $\det \mathbf{A}^T = \det \mathbf{A}$
- $\det \alpha \mathbf{A} = \alpha^N \det \mathbf{A}$

Trace

- $\operatorname{trace}(\mathbf{A}) = \sum_{n=1}^{N} \lambda_n$
- $\operatorname{tr}\left(X^T\right) = \operatorname{tr}(X)$
  $\operatorname{tr}(X+Y) = \operatorname{tr}(X) + \operatorname{tr}(Y)$
  $\operatorname{tr}(\lambda X) = \lambda \operatorname{tr}(X)$
- $\langle A, B\rangle = \operatorname{tr}\left(B^T A\right) = \operatorname{tr}\left(\left(B^T A\right)^T\right) = \operatorname{tr}\left(A^T B\right) = \langle B, A\rangle$

Inverse

- $AB = B^{-1}A^{-1}$ if both A and B are invertable.
- $(I+A)^{-1} \approx I - A$ for 'small' A.
- If A is non-singular and symmetric, $A^{-1} = \left(A^{-1}\right)^T$

Quadratic

- $\boldsymbol{x}^{\mathrm{T}} P \boldsymbol{x} = \sum_{i,j=1}^{n} x_i x_j P_{ij}$

Convex Sets

- **Affine set:** if $\forall \boldsymbol{x}_1, \boldsymbol{x}_2 \in C$ then $\theta \boldsymbol{x}_1 + (1-\theta)\boldsymbol{x}_2 \in C, \forall \theta \in \mathbb{R}$
  Affine combination: $\sum_{i=1}^{n} \theta_i \boldsymbol{x}_i$, s.t. $\sum_{\mathrm{i}=1}^{\mathrm{n}} \theta_{\mathrm{i}} = 1, \theta_{\mathrm{i}} \in \mathbb{R}$
  A set $C \subseteq \mathbb{R}^n$ is affine if the <u>line</u> through any two distinct points in C lies in C.
  An affine set contains all affine combinations of points in the set.
- **Convex set:** if $\forall \boldsymbol{x}_1, \boldsymbol{x}_2 \in C$ then $\theta \boldsymbol{x}_1 + (1-\theta)\boldsymbol{x}_2 \in C, \theta \in [0,1]$
  Convex combination: $\sum_{i=1}^{n} \theta_i \boldsymbol{x}_i$, s.t. $\sum_{i=1}^{n} \theta_i = 1$, $\theta_i \geq 0$
  Convex hull of a set C is a set of all convex combinations of points in C.
- **Cones:** if $\forall \boldsymbol{x} \in C$ then $\theta \boldsymbol{x} \in C$, $\forall \theta \geq 0$.
  Conic combination: $\sum_{i=1}^{n} \theta_i \boldsymbol{x}_i, \theta_i \geq 0$
- Proving $\mathscr{P}$ is convex:
  - Pick $\boldsymbol{x}_1 \in \mathscr{P}, \boldsymbol{x}_2 \in \mathscr{P}$
  - Pick any $\theta \in [0,1]$
  - Test $\theta \boldsymbol{x}_1 + (1-\theta)\boldsymbol{x}_2$. Is it in $\mathscr{P}$?

<u>**Examples:**</u>

- **Hyperplane:** $\left\{\boldsymbol{x} \mid \boldsymbol{a}^{\mathrm{T}} \boldsymbol{x} = \boldsymbol{b}\right\} = \left\{\boldsymbol{x} \mid \boldsymbol{a}^{\mathrm{T}}\left(\boldsymbol{x} - \boldsymbol{x}_0\right) = 0\right\}$
  $\boldsymbol{a}$ is a normal vector, $\boldsymbol{b}$ determines the offset from origin.
- **Half space:** $\left\{\boldsymbol{x} \mid \boldsymbol{a}^{\mathrm{T}} \boldsymbol{x} \leq \boldsymbol{b}\right\} = \left\{\boldsymbol{x} \mid \boldsymbol{a}^{\mathrm{T}}\left(\boldsymbol{x} - \boldsymbol{x}_0\right) \leq 0\right\}$.
- **Polyhedra:** $\{\boldsymbol{x} \mid \mathbf{A}\boldsymbol{x} \leq \boldsymbol{b}, \mathbf{C}\boldsymbol{x} = \boldsymbol{d}\}$
- **Norm ball:** $\mathcal{B} = \{\boldsymbol{x} \|\boldsymbol{x}\| \leq 1\}$, is a convex set for all norms.
- **Euclidean Ball:** $\mathcal{B}\left(\boldsymbol{x}_c, r\right) = \left\{\boldsymbol{x} \mid \left(\boldsymbol{x} - \boldsymbol{x}_c\right)^{\mathrm{T}}\left(\boldsymbol{x} - \boldsymbol{x}_c\right) \leq r^2\right\}$
- **Ellipse:** $\mathcal{E}\left(\boldsymbol{x}_c, \mathbf{P}\right) = \left\{\boldsymbol{x} \mid \left(\boldsymbol{x} - \boldsymbol{x}_c\right)^{\mathrm{T}} \mathbf{P}^{-1}\left(\boldsymbol{x} - \boldsymbol{x}_c\right) \leq 1\right\}, \mathbf{P} \in S_{++}^n$
  - Euclidean ball is an ellipse with $\mathbf{P} = \mathbf{I}r^2$
  - Geometry.
  - Use cols of Q.
  - $x = Q\tilde{x} + x_c$
  - Volume is proportional to $\sqrt{\det P} = \sqrt{\prod_{i=1}^{n} \lambda_i}$
- S_+^n is a convex cone.
- **Generalized inequalities:** $\boldsymbol{x} \leq_K \boldsymbol{y} \leftrightarrow \boldsymbol{y} - \boldsymbol{x} \in K$

<u>**Operation that preserves convexity:**</u>

- **Intersection**: If S_α is (affine,convex, conic) then $\cap_\alpha S_\alpha$ is (affine,convex, conic) (perhaps infinitely many).
- **Affine map**: $\boldsymbol{f}: \mathbb{R}^n \rightarrow \mathbb{R}^m, \boldsymbol{f}(\boldsymbol{x}) = \mathbf{A}\boldsymbol{x} + \boldsymbol{b}$. If S is convex, then
  – $\boldsymbol{f}(S) = \{\boldsymbol{f}(\boldsymbol{x}) \mid \boldsymbol{x} \in S\}$ is convex, i.e., image of a convex set under affine map is convex.
  – $\boldsymbol{f}^{-1}(S) = \{\boldsymbol{x} \mid \boldsymbol{f}(\boldsymbol{x}) \in S\}$ is convex, i.e., pre-image of ..

<u>**Properties of convex sets:**</u>

- **Separating hyperplanes:** If $S, T \subseteq \mathbb{R}^n$ are convex and disjoint i.e. $S \cap T = \emptyset$, then $\exists \boldsymbol{a} \in \mathbb{R}^n, \boldsymbol{a} \neq 0$ and $\boldsymbol{b} \in \mathbb{R}$ s.t. $\boldsymbol{a}^{\mathrm{T}} \boldsymbol{x} \geq b, \forall \boldsymbol{x} \in S$ and $\boldsymbol{a}^{\mathrm{T}} \boldsymbol{x} \leq b, \forall \boldsymbol{x} \in T$.
- **Supporting hyperplane:** If S is convex then $\forall x_0 \in \partial S$, $\exists$ $\boldsymbol{a} \neq 0 \in \mathbb{R}^n$ s.t. $\boldsymbol{a}^{\mathrm{T}} \boldsymbol{x} \leq \boldsymbol{a}^{\mathrm{T}} x_0, \forall \boldsymbol{x} \in S$

Convex Functions

- If $f: \mathbb{R}^n \rightarrow \mathbb{R}$ is defined on a convex domain (i.e. $\operatorname{dom} f \subseteq \mathbb{R}^n$ is a convex set), then f is <u>**convex**</u> if $\forall \boldsymbol{x}, \boldsymbol{y} \in \operatorname{dom} f, \forall \theta \in [0,1], f(\theta \boldsymbol{x} + (1-\theta)\boldsymbol{y}) \leq \theta f(\boldsymbol{x}) + (1-\theta) f(\boldsymbol{y})$
- The <u>**epigraph**</u> of a function $f: \mathbb{R}^n \rightarrow \mathbb{R}$ is the **set** $\operatorname{epi} f = \left\{(\boldsymbol{x}, t) \in \mathbb{R}^{n+1} \mid \boldsymbol{x} \in \operatorname{dom} f, t \geq f(\boldsymbol{x})\right\}$, where $\boldsymbol{x} \in \mathbb{R}^n, t \in \mathbb{R}$.
  – f is a convex function $\Leftrightarrow$ $\operatorname{epi} f$ is a convex set.
- The <u>**sublevel set**</u> of a function $f: \mathbb{R}^n \rightarrow \mathbb{R}$ is $C(\alpha) = \{\boldsymbol{x} \in \operatorname{dom} f \mid f(\boldsymbol{x}) \leq \alpha\}$
  – f is a convex function $\Rightarrow C(\alpha)$ is a convex set.
  – **Quasiconvex**: all sublevel sets are convex.
  – **Quasiconcave**: all superlevel sets are convex.

Ways to show a function is convex
<u>Use 1st order (differentiable)/2nd order (twice) conditions.</u>

- f convex $\Leftrightarrow$ $\operatorname{dom} f$ is a convex set and $\forall x, x_0 \in \operatorname{dom} f$, $f(x) \geq f\left(x_0\right) + \nabla f\left(x_0\right)^{\mathrm{T}}\left(x - x_0\right)$
- f convex $\Leftrightarrow$ $\operatorname{dom} f$ is convex, $\forall x \in \operatorname{dom} f, \nabla^2 f(x) \geq 0$

<u>Reduce to scalar scenario.</u>

- f is convex $\Leftrightarrow f(x_0 + tv)$ is convex in t.
- To prove $f(x)$ is convex, choose a starting point $x_0 \in \mathbb{R}^n$ and a direction $v \in \mathbb{R}^n$, and prove $g(t) = f(x_0 + tv)$ is convex in $t \in \mathbb{R}$.

Use properties of operations that preserve convexity.

- **Nonnegative weighted sums**. If $f_1, \ldots, f_m$ are convex, the nonnegative weighted sum of them $f = w_1 f_1 + \cdots + w_m f_m$ is convex.
- **Composition with an affine mapping.** Suppose $f : \mathbf{R}^n \to \mathbf{R}, A \in \mathbf{R}^{n \times m}$, and $b \in \mathbf{R}^n$. Define $g : \mathbf{R}^m \to \mathbf{R}$ by $g(x) = f(Ax + b)$ with $\operatorname{dom} g = \{x \mid Ax + b \in \operatorname{dom} f\}$. Then if f is convex, so is g.
- **Pointwise maximum.** If $f_1, \ldots, f_m$ are convex, their pointwise maximum $f(x) = \max\{f_1(x), \ldots, f_m(x)\}$ is convex.
  - Sum of r largest components of $x \in \mathbb{R}^n$ is a convex function. $f(x) = \sum_{i=1}^r x_{[i]}$. Further, if $w_1 \geq w_2 \geq \cdots \geq w_r \geq 0$, then $\sum_{i=1}^r w_i x_{[i]}$ is convex.
- **Composition.** $f(x) = h(g(x))$
  - $g : \mathbb{R}^n \to \mathbb{R}, h : \mathbb{R} \to \mathbb{R}, f$ is convex if

    g convex, h convex and non-decreasing

    g concave, h convex and non-increasing
  - $g : \mathbb{R}^n \to \mathbb{R}^k, h : \mathbb{R}^k \to \mathbb{R}$, f is convex if

    g_i is convex for all $i \in [k]$, h convex and non-decreasing in each argument

    $h(g(x)) = h(g_1(x), g_2(x), ..., g_k(x)), g_k : \mathbb{R}^n \to \mathbb{R}$.

Optimality Conditions

For unconstrained problems:

- x^* is a local minimum of $f \Rightarrow \nabla f(x^*) = 0$ and $\nabla^2 f(x^*) \geq 0$.
- x^* is a local minimum of $f \Leftarrow \nabla f(x^*) = 0$ and $\nabla^2 f(x^*) > 0$.

For constrained problems:

If f_0 is differentiable, then

x^* is optimal $\Leftrightarrow \forall y \in C, \nabla f_0(x^*)^T (y - x^*) \geq 0$

Lagrange Method

- **Primal problem**:

$$\begin{array}{ll} \underset{x}{\text{minimize}} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p \end{array}$$

  - Domain $\mathcal{D} = \bigcap_{i=0}^m \operatorname{dom} f_i \cap \bigcap_{i=1}^p \operatorname{dom} h_i$
  - Optimum: x^*, optimal value: p^*
- **Lagrangian**:

  $L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \nu_i h_i(x)$,
- **Dual function**: $g(\lambda, \nu) = \min_{x \in \mathcal{D}} L(x, \lambda, \nu)$
- **Dual optimization problem**:

$$\begin{array}{ll} \underset{\lambda, \nu}{\text{maximize}} & g(\lambda, \nu) \\ \text{subject to} & \lambda \succeq 0 \end{array}$$

  - Optimum: λ^*, ν^*, optimal value: d^*

Properties

- $g(\lambda, \nu)$ is concave in (λ, ν)
- $g(\lambda, \nu) \leq g(\lambda^*, \nu^*) \leq f_0(x^*) \leq f_0(x)$
- Sufficient condition for strong duality: problem convex & Slater's condition ($\exists x, \text{s.t.} f_i(x) < 0, \forall i \in [m], Ax = b$) holds.

KKT Conditions

- **Primal feasibility**:

  $f_i(x^\star) \leq 0, \forall i \in [m], \quad h_i(x^\star) = 0, \forall i \in [p]$
- **Dual feasibility**:

  $\lambda_i^\star \geq 0, \quad \forall i \in [m]$
- **First order condition**:

  $\nabla f_0(x^\star) + \sum_{i=1}^m \lambda_i^\star \nabla f_i(x^\star) + \sum_{i=1}^p \nu_i^\star \nabla h_i(x^\star) = 0$
- **Complementary slackness**:

  $\lambda_i^\star f_i(x^\star) = 0, \quad \forall i \in [m]$

Theorems

- If strong duality holds, (x^*, λ^*, μ^*) are primal & dual optimal, then (x^*, λ^*, μ^*) satisfies KKT.
- For **convex** opt. problem, if $(\tilde{x}, \tilde{\lambda}, \tilde{\mu})$ satisfies KKT, then strong duality holds and $(\tilde{x}, \tilde{\lambda}, \tilde{\mu})$ are primal & dual optimal.
- If problem is differentiable, convex, strong duality holds, (x, λ, μ) satisfies KKT $\Longleftrightarrow (x, \lambda, \mu)$ are primal & dual optimal.