The University of New South Wales
School of Mathematics and Statistics

MATH3161 & MATH5165 – Optimization, Term 1, 2025[1]

OPTIMIZATION – COURSE SUMMARY

TOPIC 1 – MODEL FORMULATION

Standard Formulation:

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f(x) \\ \text{s.t.} \quad & c_i(x) = 0, \ i = 1, \ldots, m_E, \\ & c_i(x) \le 0, \ i = m_E + 1, \ldots, m. \end{aligned}$$

Conversions:

- Maximize to minimize: $\max f(x) = -\min\big(-f(x)\big)$
- Constraint right-hand-side: $c_i(x) = b_i \ \Leftrightarrow \ c_i(x) - b_i = 0$
- Greater-than-or-equal-to inequalities: $c_i(x) \ge 0 \ \Leftrightarrow \ -c_i(x) \le 0$
- Strict inequality: $c_i(x) < 0 \ \Leftrightarrow \ c_i(x) + \varepsilon \le 0,$ for some $\varepsilon > 0$

TOPIC 2 – MATHEMATICAL BACKGROUND

- **Existence of Global Extrema:** Let Ω be a compact set and f be continuous on Ω. Then the global extrema of f over Ω exist.
- **Gradient:** Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable. The gradient $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$ of f at x is

$$\nabla f(x) = \begin{bmatrix} \dfrac{\partial f(x)}{\partial x_1} \\ \vdots \\ \dfrac{\partial f(x)}{\partial x_n} \end{bmatrix}.$$

[1] Prof Jeya Jeyakumar, School of Mathematics and Statistics, UNSW Sydney.

- **Hessian:** Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable. The Hessian $\nabla^2 f : \mathbb{R}^n \to \mathbb{R}^{n\times n}$ of f at x is

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

- Let $A \in \mathbb{R}^{n\times n}$ be any real square matrix (not necessarily symmetric). Then A is:
  - *positive definite* $\Leftrightarrow x^\top A x > 0$ for all $x \in \mathbb{R}^n$, $x \neq 0$.
  - *positive semi-definite* $\Leftrightarrow x^\top A x \geq 0$ for all $x \in \mathbb{R}^n$.
  - *negative definite* $\Leftrightarrow x^\top A x < 0$ for all $x \in \mathbb{R}^n$, $x \neq 0$.
  - *negative semi-definite* $\Leftrightarrow x^\top A x \leq 0$ for all $x \in \mathbb{R}^n$.
  - *indefinite* $\Leftrightarrow$ there exist $x, y \in \mathbb{R}^n$ such that $x^\top A x < 0$ and $y^\top A y > 0$.

- A real square matrix $A \in \mathbb{R}^{n\times n}$ is positive definite (resp. positive semi-definite, negative definite, negative semi-definite, indefinite) *iff* its symmetric part $\frac{A + A^\top}{2}$ is positive definite (resp. positive semi-definite, negative definite, negative semi-definite, indefinite).

- Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix. Then:
  - A has n real eigenvalues.
  - There exists an orthogonal matrix Q ($Q^\top Q = I$) such that $A = QDQ^\top$ where $D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ and $Q = [v_1 \ \cdots \ v_n]$ with v_i an eigenvector of A corresponding to eigenvalue λ_i.
  - $\det(A) = \prod_{i=1}^n \lambda_i$ and $\operatorname{tr}(A) = \sum_{i=1}^n \lambda_i = \sum_{i=1}^n A_{ii}$.
  - A is positive definite $\Leftrightarrow \lambda_i > 0$ for all $i = 1, \ldots, n$.
  - A is positive semi-definite $\Leftrightarrow \lambda_i \geq 0$ for all $i = 1, \ldots, n$.
  - A is indefinite $\Leftrightarrow$ there exist i, j with $\lambda_i > 0$ and $\lambda_j < 0$.

TOPIC 3 – CONVEXITY OF SETS AND FUNCTIONS

- **Convex set:** A set $\Omega \subseteq \mathbb{R}^n$ is convex $\Leftrightarrow \theta x + (1-\theta)y \in \Omega$ for all $\theta \in [0,1]$ and for all $x, y \in \Omega$.

- **Convex function:** Let $\Omega \subseteq \mathbb{R}^n$ be a convex set. The function $f : \Omega \to \mathbb{R}$ is convex *iff*
$$f\big(\theta x + (1-\theta)y\big) \leq \theta f(x) + (1-\theta) f(y) \quad \forall\, \theta \in [0,1],\ \forall\, x, y \in \Omega.$$
- **Convex inequality:** Let $c : \mathbb{R}^n \to \mathbb{R}$ be a convex function. Then $\Omega = \{x \in \mathbb{R}^n : c(x) \leq 0\}$ is a convex set.
- Let $\Omega \subset \mathbb{R}^n$ be convex and $f \in C^2(\Omega)$. Then
  - $\nabla^2 f(x)$ is positive semi-definite $\forall x \in \Omega \;\Rightarrow\; f$ is convex on Ω.
  - $\nabla^2 f(x)$ is positive definite $\forall x \in \Omega \;\Rightarrow\; f$ is strictly convex on Ω.
  - Ω is open *and* $\nabla^2 f(x)$ is positive semi-definite $\forall x \in \Omega \Leftrightarrow\; f$ is convex on Ω.
- **Epigraph:** For $f : \Omega \to \mathbb{R}$, $\operatorname{epi} f = \big\{(x, r) \in \mathbb{R}^n \times \mathbb{R} : x \in \Omega,\ f(x) \leq r\big\}$.
- **Epigraph of a convex function:** f is convex on a convex $\Omega \Leftrightarrow$ $\operatorname{epi} f$ is convex in $\mathbb{R}^n \times \mathbb{R}$.

TOPIC 4 – UNCONSTRAINED OPTIMIZATION

- **Problem form:** $\min\limits_{x \in \Omega} f(x)$ with $\Omega = \mathbb{R}^n$.
- **First-order necessary condition:** If x^* is a local minimizer and $f \in C^1(\mathbb{R}^n)$ then $\nabla f(x^*) = 0$.
- **Unconstrained stationary point:** x^* with $\nabla f(x^*) = 0$. Such points may be minimizers, maximizers or saddle points.
- **Second-order necessary conditions:** If $f \in C^2(\mathbb{R}^n)$ then
  - Local minimizer $\Rightarrow \nabla f(x^*) = 0$ *and* $\nabla^2 f(x^*)$ positive semi-definite.
  - Local maximizer $\Rightarrow \nabla f(x^*) = 0$ *and* $\nabla^2 f(x^*)$ negative semi-definite.
- **Second-order sufficient conditions:** If $\nabla f(x^*) = 0$ then
  - $\nabla^2 f(x^*)$ positive definite $\Rightarrow x^*$ is a *strict* local minimizer.
  - $\nabla^2 f(x^*)$ negative definite $\Rightarrow x^*$ is a *strict* local maximizer.
  - $\nabla^2 f(x^*)$ indefinite $\Rightarrow x^*$ is a saddle point.
- **Sufficiency under convexity/concavity:** For $f \in C^2(\mathbb{R}^n)$:
  - f convex $\Rightarrow$ any stationary point is a global minimizer.
  - f *strictly* convex $\Rightarrow$ stationary point is the *unique* global minimizer.
  - f concave $\Rightarrow$ any stationary point is a global maximizer.
  - f *strictly* concave $\Rightarrow$ stationary point is the *unique* global maximizer.

TOPIC 5 – EQUALITY CONSTRAINTS

Standard form:

$$\min_{x\in\mathbb{R}^n}\big\{f(x) : c_i(x) = 0,\ i = 1,\dots,m\big\}. \tag{PE}$$

Lagrangian: For $x \in \mathbb{R}^n$, $\lambda \in \mathbb{R}^m$,

$$L(x,\lambda) := f(x) + \sum_{i=1}^{m} \lambda_i c_i(x).$$

- **Regular point:** A feasible point x is *regular* $\Leftrightarrow$ the gradients $\nabla c_i(x)$, $i = 1,\dots,m$, are linearly independent.
- **First-order necessary optimality conditions:** If x^* is a local minimizer and a regular point of (PE), then $\exists\, \lambda^* \in \mathbb{R}^m$ such that

$$\nabla_x L(x^*,\lambda^*) = 0, \qquad \nabla_\lambda L(x^*,\lambda^*) = 0.$$

- **Constrained stationary point:** Any x^* for which $\exists\, \lambda^*$ satisfying the above equalities.
- **Second-order sufficient conditions:** Let (x^*,λ^*) be a constrained stationary point. Define

$$A(x^*) = \big[\nabla c_1(x^*) \quad \cdots \quad \nabla c_m(x^*)\big], \qquad Z^* \in \mathbb{R}^{n\times(n-t^*)},\ t^* = \operatorname{rank} A(x^*),\ (Z^*)^\top A(x^*) = 0.$$

The *reduced Hessian* of L is $W_Z^* = (Z^*)^\top \nabla^2_{xx} L(x^*,\lambda^*) Z^*$. If W_Z^* is positive definite, then x^* is a strict local minimizer.

TOPIC 6 – INEQUALITY CONSTRAINTS

Standard form (with equalities and inequalities):

$$\min_{x\in\mathbb{R}^n}\big\{f(x) : c_i(x) = 0,\ i \in E,\ c_i(x) \le 0,\ i \in I\big\}. \tag{NLP}$$

- **Active set at x^*:** $A(x^*) = \{i \in E \cup I : c_i(x^*) = 0\}$.
- **Regular point:** Feasible x^* such that $\{\nabla c_i(x^*) : i \in A(x^*)\}$ are linearly independent.
- **Constrained stationary point:** Feasible x^* for which $\exists\, \lambda_i^*$ $(i \in A(x^*))$ with $\nabla f(x^*) + \sum_{i\in A(x^*)} \lambda_i^* \nabla c_i(x^*) = 0$.
- **Karush–Kuhn–Tucker (KKT) necessary conditions:** If x^* is a local minimizer and a regular point, then $\exists\, \lambda_i^*$ $(i \in A(x^*))$ such that

$$\nabla f(x^*) + \sum_{i\in A(x^*)} \lambda_i^* \nabla c_i(x^*) = 0,$$

with $c_i(x^*) = 0$ $(i \in E)$, $c_i(x^*) \le 0$ $(i \in I)$, $\lambda_i^* \ge 0$ $(i \in I)$, and $\lambda_i^* = 0$ for $i \notin A(x^*)$.

- **Second-order sufficient conditions for strict local minimum:** Let $t^* = |A(x^*)|$, $A^* = [\nabla c_i(x^*) \mid i \in A(x^*)]$. If $t^* < n$ and A^* has full rank, let $Z^* \in \mathbb{R}^{n\times(n-t^*)}$ with $(Z^*)^\top A^* = 0$. Define

$$W^* = \nabla^2 f(x^*) + \sum_{i\in A(x^*)} \lambda_i^* \nabla^2 c_i(x^*), \qquad W_Z^* = (Z^*)^\top W^* Z^*.$$

  If $\lambda_i^* > 0\ \forall i \in I \cap A(x^*)$ and W_Z^* is positive definite, then x^* is a strict local minimizer.

- **Convex problem:** (NLP) is a *convex optimization* problem if f is convex on the feasible set, c_i is affine $\forall i \in E$, and c_i is convex $\forall i \in I$.

- **KKT sufficient conditions for global minimum:** If (NLP) is convex and x^* satisfies the KKT conditions with $\lambda_i^* \geq 0$ for all $i \in I \cap A(x^*)$, then x^* is a global minimizer.

- **Dual problem (Wolfe dual):**

$$\max_{y\in\mathbb{R}^n,\,\lambda\in\mathbb{R}^m} \left\{ f(y) + \sum_{i=1}^m \lambda_i c_i(y) : \nabla f(y) + \sum_{i=1}^m \lambda_i \nabla c_i(y) = 0,\ \lambda_i \geq 0\ (i \in I) \right\}. \tag{D}$$

TOPIC 7 – NUMERICAL METHODS

- **Rates of convergence of iterative methods.** If $x_k \to x^*$ and $\dfrac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|^\alpha} \to \beta$ as $k \to \infty$, the method has *α-th order* convergence. Key cases: $\alpha = 1$ (*linear*), $\alpha = 1$ with $\beta = 0$ (*superlinear*), $\alpha = 2$ (*quadratic*).

- **Line search methods.** Given $s^{(k)}$ at $x^{(k)}$, set $x^{(k+1)} = x^{(k)} + \alpha^{(k)} s^{(k)}$ where $\alpha^{(k)}$ minimizes or approximately minimizes $\ell_k(\alpha) = f(x^{(k)} + \alpha s^{(k)})$.

- **Descent direction:** $(g^{(k)})^\top s^{(k)} < 0$.

- **Exact line search condition:** $\ell_k'(\alpha) = g(x^{(k)} + \alpha s^{(k)})^\top s^{(k)} = 0$.

- If $s^{(k)}$ is a descent direction, a line search yields $\alpha^{(k)} > 0$ with $f^{(k+1)} < f^{(k)}$.

- *Global convergence*: convergence to a stationary point from any $x^{(1)}$.

- *Quadratic termination*: method finds minimizer of a strictly convex quadratic in finite known iterations.

Steepest Descent Method

- Search direction: $s^{(k)} = -g^{(k)}$.
- Descent direction: Yes.
- Global convergence: Yes.
- Quadratic termination: No.
- Rate: Linear with exact line searches. If f is strictly convex quadratic, then for each k,
$$\|x^{(k+1)} - x^*\| \leq \left(\frac{\kappa - 1}{\kappa + 1}\right)^k \|x^{(1)} - x^*\|,$$
where κ is the condition number of $\nabla^2 f$.

Basic Newton's Method

- Search direction: solve $G^{(k)}\delta^{(k)} = -g^{(k)}$ where $G^{(k)}$ is the Hessian.
- Descent direction: Yes, if $G^{(k)}$ positive definite.
- Global convergence: No (Hessian may be singular).
- Quadratic termination: Yes (one iteration for strictly convex quadratics).
- Rate: Quadratic if G^* positive definite.
- Usage: When Hessian can be evaluated and is positive definite.

Conjugate Gradient Method

- Search direction: $s^{(k)} = -g^{(k)} + \beta^{(k)} s^{(k-1)}$.
- Descent direction: Yes.
- Quadratic termination: Yes with exact line searches.
- Usage: Large n; stores only vectors, avoids solving linear systems.

TOPIC 8 – PENALTY FUNCTION METHODS

Standard formulation: Same as (NLP).
Penalty problem:

$$\min_{x\in\mathbb{R}^n}\big(f(x)+\mu P(x)\big), \tag{P_μ}$$

with

$$P(x)=\sum_{i=1}^{m_E}[c_i(x)]^2+\sum_{i=m_E+1}^{m}[c_i(x)]_+^2, \qquad [c_i(x)]_+=\max\{c_i(x),0\}.$$

- If $c:\mathbb{R}^n\to\mathbb{R}$ is convex, then $\max\{c(x),0\}^2$ is convex.
- $\dfrac{\partial}{\partial x_j}\big(\max\{c(x),0\}^2\big)=2\,\max\{c(x),0\}\,\dfrac{\partial c}{\partial x_j}$.
- **Convergence Theorem:** For each $\mu>0$ let x_μ minimize (P_μ) and set $\theta(\mu)=f(x_\mu)+\mu P(x_\mu)$. Suppose $\{x_\mu\}$ lies in a closed bounded set. Then

  $$\min_x\{f(x):c_i(x)=0,\ i\in E,\ c_i(x)\le 0,\ i\in I\}=\lim_{\mu\to\infty}\theta(\mu).$$

  Moreover, any cluster point x^* of $\{x_\mu\}$ solves the original problem, and $\mu P(x_\mu)\to 0$ as $\mu\to\infty$.

TOPIC 9 – OPTIMAL CONTROL

A typical *autonomous* optimal control problem with fixed target is

$$(C)\quad \min_{u(t)\in U}\int_{t_0}^{t_1} f_0\big(x(t),u(t)\big)\,dt \quad \text{subject to} \quad x'(t)=f\big(x(t),u(t)\big),\ x(t_0)=x_0,\ x(t_1)=x_1.$$

Hamiltonian:

$$H(x,\hat z,u)=\hat z^\top\dot{\hat x}=\sum_{i=0}^{n} z_i(t)f_i\big(x(t),u(t)\big),$$

where $\hat z(t)=[z_0(t),\dots,z_n(t)]^\top$, $\hat x(t)=[x_0(t),\dots,x_n(t)]^\top$, $x_0(t)=f_0(x(t),u(t))$, $x_0(t_0)=0$.
Co-state equations: $\dot{\hat z}=-\dfrac{\partial H}{\partial \hat x}$.

Pontryagin Maximum Principle (autonomous, fixed targets). Suppose (x^*,u^*) is optimal for (C). Then

- $z_0=-1$ (*normal* case), so

  $$H\big(x,\hat z,u\big)=-f_0\big(x,u\big)+\sum_{i=1}^{n} z_i(t)f_i\big(x,u\big).$$

- Co-state equations admit a solution $\hat{z}^*$.
- u^* maximizes $H(x^*, \hat{z}^*, u)$ over $u \in U$.
- x^* satisfies state equation with $x^*(t_0) = x_0$, $x^*(t_1) = x_1$.
- The Hamiltonian is constant along the optimal path and equals 0 if t_1 is free:

$$H(x^*, \hat{z}^*, u^*) = \begin{cases} \text{constant}, & t_1 \text{ fixed}, \\ 0, & t_1 \text{ free}. \end{cases}$$

Partially free targets: If target is intersection of k surfaces $g_i(x_1) = 0$, $i = 1, \ldots, k$, then transversality condition is $z_1 = \sum_{i=1}^{k} c_i \nabla g_i(x_1)$ for some constants c_i.
Completely free target: If $x(t_1)$ is unrestricted, then $z(t_1) = 0$.

Non-autonomous problems. For state $x' = f(x, u, t)$ and cost $J = \int_{t_0}^{t_1} f_0(x, u, t)\, dt$, introduce extra state x_{n+1} with $x'_{n+1} = 1$, $x_{n+1}(t_0) = t_0$, $x_{n+1}(t_1) = t_1$.