Convex and Non-Convex Optimisation

Aayush Bajaj

Version 0.1

2025-05-08

Table of Contents

1. Mathematical Background 3
2. Convexity 4
   - 2.1. Sets 4
   - 2.2. Functions 4
3. Unconstrained Optimisation 5
   - 3.1. Standard Form 5
4. Equality Constraints 5
   - 4.1. Standard Form 5
5. Inequality Constraints 6
6. Line Search Descent 6
7. Newton's Method & Conjugate Gradient Methods 6
8. Penalty Methods 6
9. Optimal Control Theory 6
   - 9.1. Notation 6
   - 9.2. Topic 1 – Model Formulation 7
     - 9.2.1. Standard form 7
       - 9.2.1.1. Typical conversions 7
   - 9.3. Topic 2 – Mathematical Background 7
     - 9.3.1. Gradients and Hessians 7
     - 9.3.2. Definiteness of real matrices 7
   - 9.4. Topic 3 – Convexity of Sets and Functions 7
     - 9.4.1. Sets 7
     - 9.4.2. Functions 7
   - 9.5. Topic 4 – Unconstrained Optimisation 8
     - 9.5.1. First- and second-order tests 8
   - 9.6. Topic 5 – Equality-Constrained Optimisation 8
     - 9.6.1. Problem 8
       - 9.6.1.1. Lagrangian 8
       - 9.6.1.2. Regularity 8
       - 9.6.1.3. First-order (KKT) conditions 8
       - 9.6.1.4. Second-order sufficiency 9
   - 9.7. Topic 6 – Problems with Inequality Constraints 9
       - 9.7.0.1. KKT conditions 9
       - 9.7.0.2. Second-order test 9
     - 9.7.1. Convex programmes 9
     - 9.7.2. Wolfe dual 9
   - 9.8. Topic 7 – Numerical Methods (Unconstrained) 9
     - 9.8.1. General line-search framework 9
       - 9.8.1.1. Convergence rates 10
     - 9.8.2. Steepest Descent 10
     - 9.8.3. Newton's method 10
     - 9.8.4. Conjugate Gradient (non-linear CG) 10
   - 9.9. Topic 8 – Penalty Function Methods 10
   - 9.10. Topic 9 – Optimal Control (Pontryagin Maximum Principle) 10

1. Mathematical Background

Definition 1.1 **Mathspeak**

1. Axiom: A foundational statement accepted without proof. All other results are built ontop.
2. Proposition: A proved statement that is less central than a theorem, but still of interest.
3. Lemma: A helper'' proposition proved to assist in establishing a more important result.
4. Corollary: A statement following from a theorem or proposition, requiring little to no extra proof.
5. Definition: A precise specification of an object, concept or notation.
6. Theorem: A non-trivial mathematical statement proved on the basis of axioms, definitions and earlier results.
7. Remark: An explanatory or clarifying note that is not part of the formal logical chain but gives insight / context.
8. Claim / Conjecture: A statement asserted that requires a proof.

Definition 1.2 **Vector Norm**

A vector norm on $\mathbb{R}^n$ is a function $\|\cdot\|$ from $\mathbb{R}^n$ to $\mathbb{R}$ such that:

a) $\|\mathbf{x}\| \geq 0, \forall\, \mathbf{x} \in \mathbb{R}^n$ and $\|\mathbf{x}\| = 0 \iff \mathbf{x} = \mathbf{0}$
b) $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\| \;\forall\, \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ (Triangle Inequality)
c) $\|\alpha\, \mathbf{x}\| = |\alpha|\; \|\mathbf{x}\| \;\forall \alpha \in \mathbb{R}, \mathbf{x} \in \mathbb{R}^n$

Theorem 1.3 **Cauchy Shwarz-Inequality**

$$|\mathbf{x}^T \mathbf{y}| \leq \|\mathbf{x}\|_2 \;\|\mathbf{y}\|_2 \tag{1}$$

Definition 1.4 **Closed and Bounded Sets**

Definition 1.5 **Functions**

a) Linear:
b) Affine:
c) Quadratic:

Definition 1.6 **Symmetric**

Definition, plus trace and determinant properties

Definition 1.7 **Principal Minors**

2. Convexity

2.1. Sets

Definition 2.1 **Convex Set**

Proposition 2.2 **Intersection of Convex Sets**

Definition 2.3 **Extreme Points**

Definition 2.4 **Convex Combination**

Definition 2.5 **Convex Hull**

Theorem 2.6 **Separating Hyperplane**

Definition 2.7 **Convex Hull**

2.2. Functions

Definition 2.8 **Convex / Concave Functions**

3. Unconstrained Optimisation

3.1. Standard Form

$$\underset{\mathbf{x} \in \Omega}{\text{minimise}}\, f(\mathbf{x}) \tag{2}$$

Theorem 3.1 **First order necessary conditions**

Definition 3.2 **Stationary point**

Definition 3.3 **Saddle point**

Theorem 3.4 **Second order necessary conditions**

Corollary 3.5 **Local maximiser**

$\bar{\mathbf{x}}$ is a local maximiser $\Longrightarrow \nabla f(\bar{\mathbf{x}}) = \mathbf{0}$ and $\nabla^2 f(\bar{\mathbf{x}})$ negative semi-definite.

Note: As the definiteness of the Hessian changes, so does the nature of the maximiser.

Theorem 3.6 **Second order sufficient conditions**

4. Equality Constraints

4.1. Standard Form

$$\begin{aligned} \underset{\mathbf{x} \in \Omega}{\text{minimise}} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{c}_i(\mathbf{x}) \end{aligned} \tag{3}$$

Definition 4.1 **Lagrangian**

Definition 4.2 **Regular Point**

Definition 4.3 **Matrix of Constraint Gradients**

$$A(\mathbf{x}) = [\nabla \mathbf{c}_i(\mathbf{x}) \ \ \dots \ \ \mathbf{c}_m(\mathbf{x})] \tag{4}$$

Definition 4.4 **Jacobian**

$$\begin{aligned} J(\mathbf{x}) &= A(\mathbf{x})^T \\ &= \begin{bmatrix} \nabla \mathbf{c}_i (\mathbf{x})^T \\ \vdots \\ \mathbf{c}_m (\mathbf{x})^T \end{bmatrix} \end{aligned} \tag{5}$$

Proposition 4.5 **First order necessary optimality conditions**

5. Inequality Constraints

Note **Reduced Hessian**

The reduced Hessian W_Z^* is the projection of the Lagrandian's Hessian onto the tangent space of the constraints at the point x^*

6. Line Search Descent

7. Newton's Method & Conjugate Gradient Methods

8. Penalty Methods

9. Optimal Control Theory

9.1. Notation

Unless stated otherwise, $x \in \mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable, ∇f and $\nabla^2 f$ denote the gradient and Hessian respectively, and c_i are the constraint functions of a non-linear programme

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad c_{i(x)} = 0 (i \in E), c_{i(x)} \leq 0 (i \in I). \tag{6}$$

9.2. Topic 1 – Model Formulation

9.2.1. Standard form

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad c_{i(x)} = 0 (i = 1, ..., m_E), c_{i(x)} \leq 0 (i = m_E + 1, ..., m). \tag{7}$$

9.2.1.1. Typical conversions

- **Maximisation.** $\max f(x) = -\min\{-f(x)\}$.
- **Right-hand sides.** $c_{i(x)} = b_i \Longleftrightarrow c_{i(x)} - b_i = 0$.
- **"$\geq$" constraints.** $c_{i(x)} \geq 0 \Longleftrightarrow -c_{i(x)} \leq 0$.
- **Strict inequalities.** $c_{i(x)} < 0 \Longleftrightarrow c_{i(x)} + \varepsilon \leq 0$ for some $\varepsilon > 0$.

9.3. Topic 2 – Mathematical Background

9.3.1. Gradients and Hessians

For $f \in C^2(\mathbb{R}^n)$

$$\nabla f(x) = \begin{pmatrix} \partial \frac{f}{\partial} x_1 \\ \vdots \\ \partial \frac{f}{\partial} x_n \end{pmatrix}, \quad \nabla^2 f(x) = \left[\partial^2 \frac{f}{\partial} x_i \partial x_j\right]_{i,j=1}^n. \tag{8}$$

9.3.2. Definiteness of real matrices

A (not necessarily symmetric) $A \in \mathbb{R}^{n \times n}$ is

| | | |—|—|—| | positive definite | $\Longleftrightarrow x^\top Ax > 0 \forall x \neq 0$,| | positive semi-def. | $\Longleftrightarrow x^\top Ax \geq 0 \forall x$,| | negative definite | $\Longleftrightarrow x^\top Ax < 0 \forall x \neq 0$,| | negative semi-def. | $\Longleftrightarrow x^\top Ax \leq 0 \forall x$,| | indefinite | $\Longleftrightarrow \exists x, z : x^\top Ax < 0, z^\top Az > 0$.|

For a **symmetric** matrix the signs of the eigenvalues $\lambda_1, ..., \lambda_n$ fully determine definiteness; e.g. $A \succ 0 \Longleftrightarrow \lambda_i > 0 \forall i$. A convenient test for $A \succ 0$ is that all leading principal minors are positive (Sylvester's criterion).

9.4. Topic 3 – Convexity of Sets and Functions

9.4.1. Sets

$\Omega \subset \mathbb{R}^n$ is **convex** if $\theta x + (1 - \theta) y \in \Omega$ for every $x, y \in \Omega$ and $\theta \in [0, 1]$.

9.4.2. Functions

A function $f : \Omega \to \mathbb{R}$ (with Ω convex) is

- **convex** if $f(\theta x + (1 - \theta) y) \leq \theta f(x) + (1 - \theta) f(y)$;

- **strictly convex** if strict inequality holds whenever $x \neq y$;
- **concave** if $-f$ is convex.

Useful characterisations:

$$f \quad \text{convex} \iff (\forall x, y \in \Omega) f(y) \geq f(x) + \nabla f(x)^\top (y - x); \tag{9}$$

$$f \quad \text{convex on open} \quad \Omega \iff \nabla^2 f(x) \succeq 0 \forall x \in \Omega; \tag{10}$$

$$\text{epigraph} \quad \text{epi } f = \{(x, r) : x \in \Omega, f(x) \leq r\} \quad \text{is convex.} \tag{11}$$

9.5. Topic 4 – Unconstrained Optimisation

9.5.1. First- and second-order tests

For $f \in C^1$:

$$x^* \quad \text{local min} \Rightarrow \nabla f(x^*) = 0. \tag{12}$$

For $f \in C^2$:

$$x^* \quad \text{local min} \Rightarrow \nabla f(x^*) = 0, \nabla^2 f(x^*) \succeq 0. \tag{13}$$

Moreover, if $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*) \succ 0$ then x^* is a **strict** local minimiser; $\prec 0$ gives a maximiser; an indefinite Hessian implies a saddle.

For **convex** (resp. concave) f, **any** stationary point is a global minimum (resp. maximum).

9.6. Topic 5 – Equality-Constrained Optimisation

9.6.1. Problem

$$\min f(x) \quad \text{s.t.} \quad c_{i(x)} = 0 (i = 1, ..., m). \tag{14}$$

9.6.1.1. Lagrangian

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

9.6.1.2. Regularity

A feasible x is **regular** if $\left\{\nabla c_{i(x)}\right\}_{i=1}^m$ are linearly independent.

9.6.1.3. First-order (KKT) conditions

If x^* is a local minimiser and regular, then

$$\nabla_x L(x^*, \lambda^*) = 0, \quad c_{i(x^*)} = 0. \tag{15}$$

Any point satisfying these with some multipliers is a **constrained stationary point**.

9.6.1.4. Second-order sufficiency

Let Z^* whose columns form a basis for ker $A^\top$ with $A = \left[\nabla c_1(x^*) \ldots \nabla c_{m(x^*)}\right]$. Define $W^* = \nabla^2 f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla^2 c_{i(x^*)}$. If $(Z^*)^\top W^* Z^* \succ 0$ then x^* is a strict local minimum.

9.7. Topic 6 – Problems with Inequality Constraints

Given (NLP)

$$\min f(x) \quad \text{s.t.} \quad c_{i(x)} = 0 (i \in E), c_{i(x)} \leq 0 (i \in I), \tag{16}$$

let the **active set** $A(x) = \left\{i \in E \cup I : c_{i(x)} = 0\right\}$.

9.7.0.1. KKT conditions

At a regular local minimiser x^* there exist multipliers λ^* such that

$$\nabla f(x^*) + \sum_{i \in A(x^*)} \lambda_i^* \nabla c_{i(x^*)} = 0, \quad \lambda_i^* \geq 0 (i \in I \cap A(x^*)). \tag{17}$$

9.7.0.2. Second-order test

With Z^*, W^* defined as before and $t^* = |A(x^*)| < n$: if $\lambda_i^* > 0 \forall i \in I \cap A(x^*)$ and $(Z^*)^\top W^* Z^* \succ 0$ then x^* is a strict local minimum.

9.7.1. Convex programmes

If f is convex, c_i affine $(i \in E)$, and c_i convex $(i \in I)$, then **any** point satisfying the KKT conditions with $\lambda_i \geq 0 (i \in I)$ is a **global** minimiser.

9.7.2. Wolfe dual

For $m = |E \cup I|$

$$\max_{y, \lambda} f(y) + \sum_{i=1}^m \lambda_i c_{i(y)} \quad \text{s.t.} \quad \nabla f(y) + \sum_{i=1}^m \lambda_i \nabla c_{i(y)} = 0, \lambda_i \geq 0 (i \in I). \tag{18}$$

Strong duality holds in the convex case.

9.8. Topic 7 – Numerical Methods (Unconstrained)

9.8.1. General line-search framework

Given descent direction s^k at x^k, choose $\alpha^k > 0$ (exact or inexact) and set $x^{k+1} = x^k + \alpha^k s^k$.

9.8.1.1. Convergence rates

If $x^k \to x^*$ and $\frac{\| x^{k+1} - x^* \|}{\| x^k - x^* \|^\alpha} \to \beta$ then the method is **α-th-order**: $\alpha = 1$ linear, $\alpha = 1, \beta = 0$ super-linear, $\alpha = 2$ quadratic.

9.8.2. Steepest Descent

$$s^k = -\nabla f(x^k). \tag{19}$$

Globally convergent, linear rate in the quadratic case; no quadratic termination.

9.8.3. Newton's method

$$\nabla^2 f(x^k) \delta^k = -\nabla f(x^k), \quad s^k = \delta^k. \tag{20}$$

Quadratic convergence near a non-singular minimiser; single-step termination for strictly convex quadratics; may fail globally if $\nabla^2 f$ is singular or indefinite.

9.8.4. Conjugate Gradient (non-linear CG)

$$s^k = -g^k + \beta^k s^{k-1}, \quad g^k = \nabla f(x^k). \tag{21}$$

Descent directions, quadratic termination (exact line search), especially attractive for large-scale problems because only vector operations are required.

9.9. Topic 8 – Penalty Function Methods

For (P) with mixed constraints define

$$P(x) = \sum_{i \in E} c_{i(x)}^2 + \sum_{i \in I} \left[c_{i(x)}\right]_+^2, \quad [x]_+ = \max\{x, 0\}. \tag{22}$$

The penalty subproblem

$$\min_{x \in \mathbb{R}^n} f(x) + \mu P(x) \quad (:= P_\mu) \tag{23}$$

is unconstrained. Under mild boundedness assumptions, every sequence of minimisers x_μ with $\mu \to \infty$ has accumulation points that solve the original constrained problem, and $\mu P(x_\mu) \to 0$.

9.10. Topic 9 – Optimal Control (Pontryagin Maximum Principle)

For an autonomous system with fixed end-points

$$\min_{u(.)} \int_{t_0}^{t_1} f_0(x(t), u(t))\,\mathrm{d}t, \quad \dot{x} = f(x(t), u(t)), x(t_0) = x_0, x(t_1) = x_1, \tag{24}$$

define the Hamiltonian $H(x, \hat{z}, u) = \hat{z}^\top (f_0(x, u), f(x, u))$. There exists a non-trivial adjoint $\hat{z}(t)$ with $\dot{\hat{z}} = -\partial \frac{H}{\partial} x$ such that the optimal control $u^*(t)$ maximises $H(x^*(t), \hat{z}(t), u)$ for all $u \in U$. For fixed end-time the Hamiltonian is constant along the optimal trajectory; it vanishes when the terminal time is free. If only some components of $x(t_1)$ are fixed, a transversality condition relates adjoint values to the gradients of the terminal constraints.