Convex and Non-Convex Optimisation

Aayush Bajaj

Version 0.1

2025-05-09

Table of Contents

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:

1. $\|\mathbf{x}\| \geq 0, \forall\, \mathbf{x} \in \mathbb{R}^n$ and $\|\mathbf{x}\| = 0 \Longleftrightarrow \mathbf{x} = \mathbf{0}$
2. $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\| \;\forall\, \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ (Triangle Inequality)
3. $\|\,\alpha\,\mathbf{x}\| = |\alpha|\; \|\mathbf{x}\| \;\forall \alpha \in \mathbb{R}, \mathbf{x} \in \mathbb{R}^n$

Definition 1.3 **Continuous Derivatives**

The notation

$$f \in C^k(\mathbb{R}^n) \tag{1}$$

means that the function $f : \mathbb{R}^n \to \mathbb{R}$ possesses continuous derivatives up to order k on $\mathbb{R}^n$.

Example

1. $f \in C^1(\mathbb{R}^n)$ implies each $\frac{\partial f}{\partial x_i}$ exists, and $\nabla f(x)$ is continuous on $\mathbb{R}^n$
2. $f \in C^2(\mathbb{R}^n)$ implies each $\frac{\partial f^2}{\partial x_i y_i}$ exists, and $\nabla^2 f(x)$ forms a continuous Hessian matrix.

Theorem 1.4 **Cauchy Shwarz-Inequality**

$$|\mathbf{x}^T \mathbf{y}| \leq \|\mathbf{x}\|_2 \, \|\mathbf{y}\|_2 \tag{2}$$

Definition 1.5 **Closed and Bounded Sets**

A set $\Omega \subset \mathbb{R}^n$ is *closed* if it contains all the limits of convergent sequences of points in Ω.

A set $\Omega \subset \mathbb{R}^n$ is *bounded* if $\exists K \in \mathbb{R}^+$ for which $\Omega \subset B[0, K]$, where $B[0, K] = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\| \leq K\}$ is the ball with centre 0.

Definition 1.6 **Standard Vector Function Forms**

If $f_0 \in \mathbb{R}, \mathbf{g} \in \mathbb{R}^n, G \in R^{\{n \times n\}}$:
1. Linear: $f(\mathbf{x}) = \mathbf{g}^T \mathbf{x}$
2. Affine: $f(\mathbf{x}) = \mathbf{g}^T \mathbf{x} + f_0$
3. Quadratic: $f(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T G \mathbf{x} + \mathbf{g}^T \mathbf{x} + f_0$

Definition 1.7 **Symmetric**

Let $A \in \mathbb{R}^{\{n \times n\}}$ be a symmetric matrix. Then:

1. A has n real eigenvalues.
    1. There exists an orthogonal matrix Q ($Q^\top Q = I$) such that $A = QDQ^\top$ where $D = \operatorname{diag}(\lambda_1, ..., \lambda_n)$ and $Q = [v_1 \dots v_n]$ with v_i an eigenvector of A corresponding to eigenvalue λ_i.
    2. $\det(A) = \prod_{i=1}^n \lambda_i$ and $\operatorname{tr}(A) = \sum_{i=1}^n \lambda_i = \sum_{i=1}^n A_{ii}$.
    3. A is positive definite $\Longleftrightarrow \lambda_i > 0$ for all $i = 1, ..., n$.
    4. A is positive semi-definite $\Longleftrightarrow \lambda_i \geq 0$ for all $i = 1, ..., n$.
    5. A is indefinite $\Longleftrightarrow$ there exist i, j with $\lambda_i > 0$ and $\lambda_j < 0$.

Definition 1.8 **Leading Principal Minors / Sylvester's Criterion**

A symmetric matrix A is **positive definite** if and only if all leading principal minors of A are positive. The ith principal minor δ_i of A is the determinant of the leading $i \times i$ submatrix of A.

If $\delta_i, i = 1, 2, ..., n$ has the sign of $(-1)^i, i = 1, 2, ..., n$, that is, the values of δ_i are alternatively negative and positive, then A is **negative definite**.

Note that *PSD* only applies if you check **all** principal minors, of which there are $2^n - 1$, as opposed to checking n submatrices here.

2. Convexity

Definition 2.1 **Convex**

A set $\Omega \subseteq \mathbb{R}^n$ is convex $\iff \theta ve(x) + (1-\theta)\,\mathbf{y} \in \Omega$ for all $\theta \in [0,1]$ and for all $\mathbf{x}, \mathbf{y} \in \Omega$. **Note:** there is no such thing as a *concave* set.

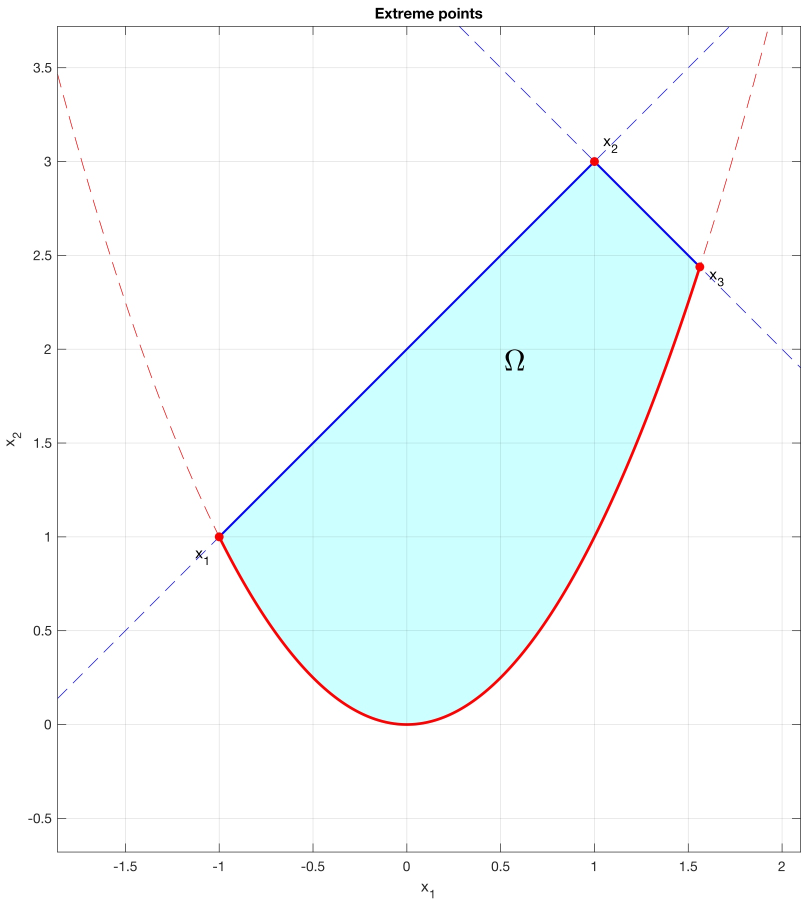
Proposition 2.2 **Intersection of Convex Sets**

Let $\Omega_1, \ldots, \Omega_n \subseteq \mathbb{R}^n$ be convex, then their intersections $\Omega = \Omega_1 \cap \ldots \cap \Omega_n$ is convex.

Definition 2.3 **Extreme Points**

Let $\Omega \subseteq \mathbb{R}^n$ be a convex set $\bar{\mathbf{x}} \in \Omega$ is an extreme point of $\Omega \iff \mathbf{x}, \mathbf{y} \in \Omega, \theta \in [0,1]$ and $\bar{\mathbf{x}} = \theta\,\mathbf{x} + (1-\theta)y \implies \mathbf{x}, \mathbf{y} \in \mathbb{R}$, or $x = y$.

In other words, a point is in an extreme point if it cannot be on a line segment in Ω.



Definition 2.4 **Convex Combination**

The convex combination of $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(m)} \in \mathbb{R}^m$ is

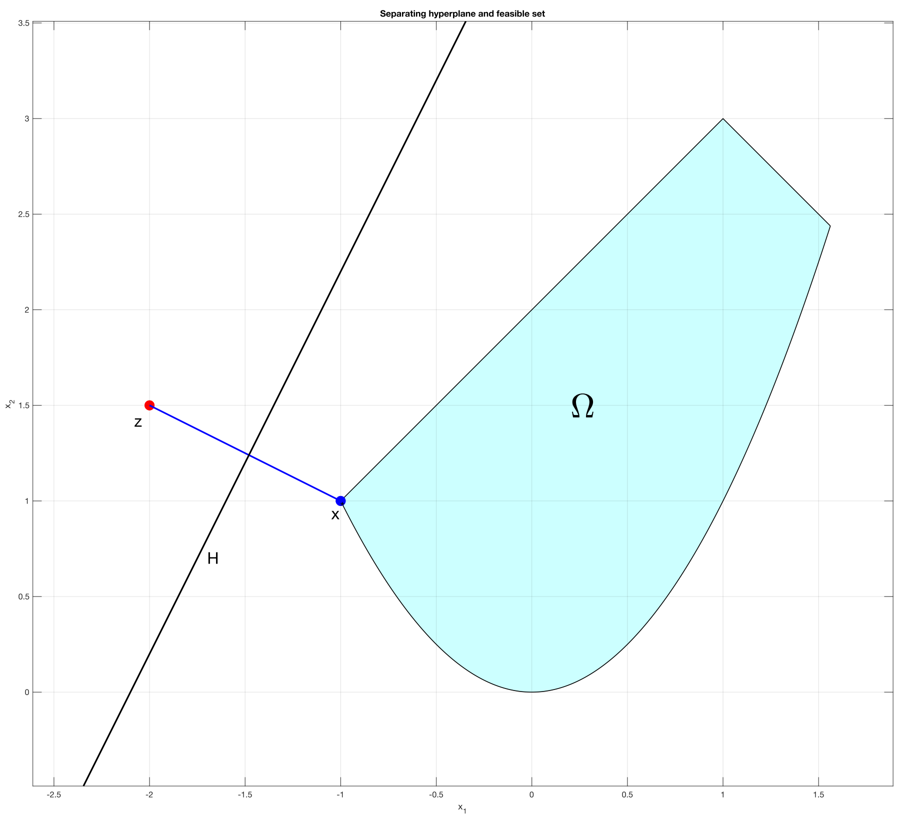
$$\mathbf{x} = \sum_{i=1}^{m} \alpha_i \, \mathbf{x}^{(i)}, \text{ where } \sum_{i=0}^{m} \alpha_i = 1 \text{ and } \alpha_i \geq 0, i = 1, \ldots, m \quad (3)$$

Definition 2.5 **Convex Hull**

The convex hull $\operatorname{conv}(\Omega)$ of a set Ω is the set of all convex combinations of points in Ω.

Theorem 2.6 **Separating Hyperplane**

Let $\Omega \subseteq \mathbb{R}^n$ be a non-empty closed convex set and let $z \notin \Omega$. There exists a hyperplane $H = \{\mathbf{u} \in \mathbb{R}^n : \mathbf{a}^T \mathbf{u} = \beta\}$ such that $\mathbf{a}^T \mathbf{z} < \beta$ and $\mathbf{a}^T \mathbf{x} \geq \beta$ for all $x \in \Omega$.



Definition 2.7 **Convex / Concave Functions**

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

- **convex** if $f(\theta\,\mathbf{x} + (1-\theta)\,\mathbf{y}) \leq \theta f(\mathbf{x}) + (1-\theta) f(\mathbf{y})$;
- **strictly convex** if strict inequality holds whenever $\mathbf{x} \neq \mathbf{y}$;
- **concave** if $-f$ is convex.

3. Unconstrained Optimisation

Definition 3.1 **Standard Form**

$$\underset{\mathbf{x}\in\Omega}{\text{minimise}}\, f(\mathbf{x}) \tag{4}$$

Remark: $\max f(\mathbf{x}) = -\min\{-f(\mathbf{x})\}$

Definition 3.2 **Hessian**

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

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix} \tag{5}$$

Theorem 3.3 **First order necessary conditions**

If x^* is a local minimizer and $f \in C^1(\mathbb{R}^n)$ then $\nabla f(x^*) = 0$.

Definition 3.4 **(Unconstrained) Stationary point**

x^* is an unconstrained stationary point $\iff \nabla f(x^*) = 0$

Example

local min, local max, saddle point.

Definition 3.5 **Saddle point**

A stationary point $\mathbf{x}^* \in \mathbb{R}^n$ is a saddle point of f if for any $\delta > 0$ there exist $\mathbf{x}, \mathbf{y}$ with $\|\mathbf{x} - \mathbf{x}^*\| < \delta$, $\|\mathbf{y} - \mathbf{x}^*\| < \delta$ such that:

$$f(\mathbf{x}) < f(\mathbf{x}^*) \quad \text{and} \quad f(\mathbf{y}) > f(\mathbf{x}^*) \tag{6}$$

Proposition 3.6 **Second order necessary conditions**

If $f \in C^2(\mathbb{R}^n)$ then
1. Local minimiser $\implies \nabla f(\mathbf{x}^*) = 0$ *and* $\nabla^2 f(\mathbf{x}^*)$ positive semi-definite.
2. Local maximiser $\implies \nabla f(\mathbf{x}^*) = 0$ *and* $\nabla^2 f(\mathbf{x}^*)$ negative semi-definite.

Corollary 3.7 **Local maximiser**

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

Theorem 3.8 **Second order sufficient conditions**

If $\nabla f(\mathbf{x}^*) = 0$ then
1. $\nabla^2 f(\mathbf{x}^*)$ positive definite $\implies \mathbf{x}^*$ is a *strict* local minimiser.
2. $\nabla^2 f(\mathbf{x}^*)$ negative definite $\implies \mathbf{x}^*$ is a *strict* local maximiser.
3. $\nabla^2 f(\mathbf{x}^*)$ indefinite $\implies \mathbf{x}^*$ is a saddle point.

1. $\nabla^2 f(\mathbf{x}^*)$ positive semi-definite $\implies \mathbf{x}^*$ is *either* a local minimiser or a saddle point!
2. $\nabla^2 f(\mathbf{x}^*)$ negative semi-definite $\implies \mathbf{x}^*$ is *either* a local maximiser or a saddle point! Be careful with these.

Corollary 3.9 **Global Optimums**

From the sufficiency of stationarity as above, and under the convexity / concavity of $f \in C^2(\mathbb{R}^n)$:
1. f convex $\implies$ any stationary point is a global minimiser.
2. f *strictly* convex $\implies$ stationary point is the *unique* global minimiser.
3. f concave $\implies$ any stationary point is a global maximiser.
4. f *strictly* concave $\implies$ stationary point is the *unique* global maximiser.

4. Equality Constraints

Definition 4.1 **Standard Form**

$$\begin{aligned} \underset{\mathbf{x}\in\Omega}{\text{minimise}} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{c}_i(\mathbf{x}) = 0 \end{aligned} \qquad \text{(EP)}$$

EP = equality problem

Definition 4.2 **Lagrangian**

For $\mathbf{x} \in \mathbb{R}^n, \boldsymbol{\lambda} \in \mathbb{R}^m$,

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) := f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i c_i(\mathbf{x}) \qquad (8)$$

Note

λ_i are termed Lagrange Multipliers

Definition 4.3 **Regular Point**

A feasible point $\bar{\mathbf{x}}$ is *regular* $\Longleftrightarrow$ the gradients $\nabla c_i(\bar{\mathbf{x}})$, $i = 1, ..., m$, are linearly independent.

feasible means the constraint is satisfied at $\bar{\mathbf{x}}$

Definition 4.4 **Matrix of Constraint Gradients**

$$A(\mathbf{x}) = [\nabla\, \mathbf{c}_i(\mathbf{x}) \ \ \dots \ \ \mathbf{c}_m(\mathbf{x})] \qquad (9)$$

Definition 4.5 **Jacobian**

$$\begin{aligned} J(\mathbf{x}) &= A(\mathbf{x})^T \\ &= \begin{bmatrix} \nabla\, \mathbf{c}_1\,(\mathbf{x})^T \\ \vdots \\ \mathbf{c}_m\,(\mathbf{x})^T \end{bmatrix} \end{aligned} \qquad (10)$$

Proposition 4.6 **First order necessary optimality conditions**

If $\mathbf{x}^*$ is a local minimiser and a regular point of (EP) 4.1, then $\exists\, \boldsymbol{\lambda}^* \in \mathbb{R}^m$ such that

$$\nabla_x \mathcal{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*) = 0, \qquad \nabla_\lambda \mathcal{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*) = \mathbf{0} \tag{11}$$

Corollary 4.7 **Constrained Stationary Point**

Any $\mathbf{x}^*$ for which $\exists\, \boldsymbol{\lambda}^*$ satisfying the first order conditions 11.

Proposition 4.8 **Second order sufficient conditions**

Let $\mathbf{x}^*$ be a constrained stationary point of (EP) 4.1 so there exist Lagrange multipliers $\boldsymbol{\lambda}^{\mathbf{x}}$ such that

$$\begin{aligned} \nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*) = \nabla f(\mathbf{x}^*) + A(\mathbf{x}^*)\, \boldsymbol{\lambda}^* = \mathbf{0} \\ \nabla_{\boldsymbol{\lambda}} \mathcal{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*) = \mathbf{c}(\mathbf{x}^*) = \mathbf{0} \end{aligned} \tag{12}$$

If W_Z^* is positive definite $\Longrightarrow$ $\mathbf{x}^*$ is a strict local minimiser.

Here,

$$A(\mathbf{x}^*) = \begin{bmatrix} \nabla c_1(\mathbf{x}^*) & \dots & c_m(\mathbf{x}^*) \\ \end{bmatrix} \tag{13}$$

$$W_Z^* := (Z^*)^\top \nabla_{\mathbf{x}}^2 \mathcal{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*) Z^* \tag{14}$$

$$Z^* \in \mathbb{R}^{n \times (n - t^*)} \qquad , t^* = \operatorname{rank}(A(\mathbf{x}^*)) \tag{15}$$

$$(Z^*)^\top A(\mathbf{x}^*) = \mathbf{0} \tag{16}$$

Remark

where W_Z^* is the reduced Hessian of the Lagrangian, and that in turn can be thought of as the projection of the Lagrangian's Hessian onto the tangent space of the constraints at the point $\mathbf{x}^*$

5. Inequality Constraints

Definition 5.1 **Standard Form**

$$\begin{aligned} \underset{\mathbf{x} \in \Omega}{\text{minimise}} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{c}_i(\mathbf{x}) = 0, \quad & i = 1, ..., m_E \\ & \mathbf{c}_i(\mathbf{x}) \leq 0, \quad & i = m_E + 1, ..., m \end{aligned} \tag{NLP}$$

NLP = non-linear problem

Definition 5.2 **Convex Problem**

The problem (NLP) 5.1 is a standard form convex optimisation problem if the objective function f is convex on the feasible set, $\mathbf{c}_i$ is affine for each $i \in \epsilon$, and $\mathbf{c}_i$ is convex for each $i \in \mathcal{I}$.

Definition 5.3 **Active Set**

The set of active constraints at a feasible point $\mathbf{x}$ is

$$\mathcal{A}(\mathbf{x}) = \{i \in 1, ..., m : \mathbf{c}_i(\mathbf{x}) = 0\} \tag{18}$$

Note that this concept only applies to inequality constraints.

Definition 5.4 **Regular Point**

Feasible $\mathbf{x}^*$ such that $\left\{\nabla c_{i(\mathbf{x}^*)} : i \in \mathcal{A}(\mathbf{x}^*)\right\}$ are linearly independent.

Proposition 5.5 **Constrained Stationary Point**

Feasible $\mathbf{x}^*$ for which $\exists \boldsymbol{\lambda}_{\mathbf{i}}^*$ for $i \in \mathcal{A}(\mathbf{x}^*)$) with

$$\nabla f(\mathbf{x}^*) + \sum_{(i \in \mathcal{A}(\mathbf{x}^*)\}} \lambda_i^* \nabla c_{i(\mathbf{x}^*)} = \mathbf{0} \tag{19}$$

Theorem 5.6 **Karush Kuhn Tucker (KKT) necessary optimality conditions**

If $\mathbf{x}^*$ is a local minimiser and a regular point, then $\exists \lambda_i^*$ $(i \in \mathcal{A}(\mathbf{x}^*))$ such that

$$\nabla f(\mathbf{x}^*) + \sum_{(i \in \mathcal{A}(\mathbf{x}^*)\}} \lambda_i^* \nabla c_{i(\mathbf{x}^*)} = \mathbf{0}, \tag{20}$$

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

Note

KKT generalises Lagrange Multipliers 8

Theorem 5.7 **Second-order sufficient conditions for strict local minimum**

Let $t^* = \mid \mathcal{A}(\mathbf{x}^*) \mid$, $\mathcal{A}^* = \left[\nabla c_{i(\mathbf{x}^*)} \mid i \in \mathcal{A}(\mathbf{x}^*)\right]$. If $t^* < n$ and $\mathcal{A}^*$ has full rank, let $Z^* \in \mathbb{R}^{n \times (n - t^*)}$ with $(Z^*)^\top \mathcal{A}^* = 0$. Define $W^* = \nabla^2 f(\mathbf{x}^*) + \sum_{(i \in \mathcal{A}(\mathbf{x}^*)\}} \lambda_i^* \nabla^2 c_{i(\mathbf{x}^*)}$

$$W_Z^* = (Z^*)^\top W^* Z^* \tag{21}$$

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

Theorem 5.8 **KKT sufficient conditions for global minimum**

If (NLP) 5.1 is convex and $\mathbf{x}^*$ satisfies the KKT 5.6 conditions with $\lambda_i^* \ge 0$ for all $i \in \mathcal{I} \cap \mathcal{A}(\mathbf{x}^*)$, then $\mathbf{x}^*$ is a global minimiser.

Theorem 5.9 **Wolfe Dual Problem**

$$
\begin{gathered}
\max_{y \in \mathbb{R}^n \; \lambda \in \mathbb{R}^m} f(y) + \sum_{i=1}^{m} \lambda_i c_{i(\mathbf{y})} \\
\text{s.t. } \nabla f(\mathbf{y}) + \sum_{i=1}^{m} \lambda_i \nabla c_{i(\mathbf{y})} = 0 \\
\lambda_i > 0 (i \in \mathcal{I})
\end{gathered}
\tag{22}
$$

strong duality, weak duality?

6. Numerical Methods (unconstrained)

Definition 6.1 **Rates of convergence of iterative methods**

Algorithm 6.2 **Line Search Algorithms**

Algorithm 6.3 **Steepest Descent Method**

Algorithm 6.4 **Newton's Method**

Algorithm 6.5 **Conjugate Gradient Method**

7. Penalty Methods

Definition 7.1 **Penalty function**

Remark

1. $c : \mathbb{R}^n \to \mathbb{R}$ is a convex function $\Longrightarrow \max\{\mathbf{c}(\mathbf{x}), 0\}^2$ is a convex function
2. $\frac{\partial}{\partial x_i}[\max\{\mathbf{c}(\mathbf{x}), 0\}]^2 = 2\max\{\mathbf{c}(\mathbf{x}), 0\}\frac{\partial}{\partial x_i}$

Theorem 7.2 **Convergence Theorem**

8. Optimal Control Theory

Definition 8.1 **Standard Form**

Definition 8.2 **Hamiltonian**

break this up so it is understandable

Definition 8.3 **Co-state Equations**

Theorem 8.4 **Pontryagin Maximum Principle**

partially free target

non-autonmous problem