Aalto University

Nonlinear optimisation, fundamentals (A)

CHEM-E7225 (was E7195), 2022

Francesco Corona (¬_¬)

Chemical and Metallurgical Engineering
School of Chemical Engineering

Overview

Nonlinear optimisation

Overview

An optimisation problem consist of the following three components

- An **objective function** $f(x)$
- The **decision variables** x
- **Constraints** $h(x)$ and $g(x)$

Consider the optimisation (minimisation) problem in standard form,

$$\begin{aligned} \min_{x \in \mathcal{R}^N} \quad & f(x) && \text{(Objective function)} \\ \text{subject to} \quad & g(x) = 0 && \text{(Equality constraints)} \\ & h(x) \geq 0 && \text{(Inequality constraints)} \end{aligned}$$

Overview (cont.)

$$\begin{aligned} \min_{x \in \mathcal{R}^N} \quad & f(x) \\ \text{subject to} \quad & g(x) = 0 \\ & h(x) \geq 0 \end{aligned}$$

All functions are (twice) continuously differentiable functions of a decision variable x

$$f(x) = \underbrace{f(x_1, x_2, \dots, x_N)}_{f : \mathcal{R}^N \to \mathcal{R}}$$

$$g(x) = \underbrace{\begin{bmatrix} g_1(x_1, x_2, \dots, x_N) \\ g_2(x_1, x_2, \dots, x_N) \\ \vdots \\ g_{N_g}(x_1, x_2, \dots, x_N) \end{bmatrix}}_{g : \mathcal{R}^N \to \mathcal{R}^{N_g}}$$

$$h(x) = \underbrace{\begin{bmatrix} h_1(x_1, x_2, \dots, x_N) \\ h_2(x_1, x_2, \dots, x_N) \\ \vdots \\ h_{N_h}(x_1, x_2, \dots, x_N) \end{bmatrix}}_{h : \mathcal{R}^N \to \mathcal{R}^{N_h}}$$

Overview (cont.)

$$\begin{aligned} \min_{x \in \mathcal{R}^N} \quad & f(x) \\ \text{subject to} \quad & g(x) = 0 \\ & h(x) \geq 0 \end{aligned}$$

We define the **feasible set** Ω to be the set of points x that satisfy all the constraints

$$\Omega := \{x \in \mathcal{R}^N : g(x) = 0, h(x) \geq 0\}$$

The feasible set defines the space in which we can search for a solution to the problem

Example

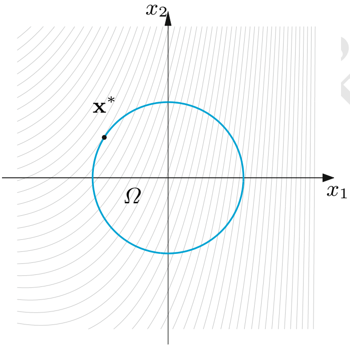
Consider the minimisation of some function $f(x)$ under some equality constraint $g(x)$

Let $f : \mathcal{R}^2 \rightarrow \mathcal{R}$

$$f(x) = \frac{3}{5}x_1^2 + \frac{1}{2}x_1 x_2 - x_2 + 3x_1$$

Let $g : \mathcal{R}^2 \rightarrow \mathcal{R}$

$$g(x) = x_1^2 + x_2^2 - 1$$



$$\min_{x \in \mathcal{R}^2} \quad f(x)$$
$$\text{subject to} \quad g(x) = 0$$

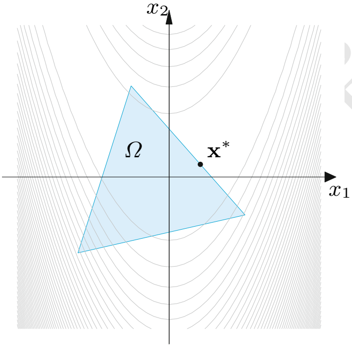
Determine minimiser x^* constrained to set $\Omega \in \mathcal{R}^2$

- In grey, contour lines of the objective $f(x)$
- In cyan, the feasible set $\Omega \in \mathcal{R}^2$

Example

Minimise function $f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$, under inequality constraints $h(x)$

$$\underbrace{\begin{bmatrix} h_1(x) \\ h_2(x) \\ h_3(x) \end{bmatrix} = \begin{bmatrix} -34x_1 - 30x_2 + 19 \\ +10x_1 - 05x_2 + 11 \\ +03x_1 + 22x_2 + 08 \end{bmatrix}}_{h:\mathcal{R}^2 \to \mathcal{R}^3}$$



$$\min_{x \in \mathcal{R}^2} \quad f(x)$$
$$\text{subject to} \quad h(x) \geq 0$$

Determine minimiser x^* constrained to set $\Omega \in \mathcal{R}^2$

- In grey, contour lines of the objective $f(x)$
- In cyan, the feasible set $\Omega \in \mathcal{R}^2$

Overview (cont.)

Example

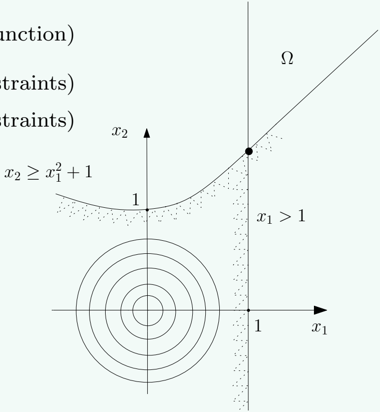
$$\min_{x \in \mathcal{R}^2} \quad x_1^2 + x_2^2 \qquad \text{(Objective function)}$$
$$\text{subject to} \quad x_1 - 1 = 0 \qquad \text{(Equality constraints)}$$
$$x_2 - 1 - x_1^2 \geq 0 \qquad \text{(Inequality constraints)}$$

- $f : \mathcal{R}^2 \rightarrow \mathcal{R}$, with $f \in \mathcal{C}^2\left(\mathcal{R}^2\right)$
- $g : \mathcal{R}^2 \rightarrow \mathcal{R}$, with $g \in \mathcal{C}^2\left(\mathcal{R}^2\right)$
- $h : \mathcal{R}^2 \rightarrow \mathcal{R}$, with $h \in \mathcal{C}^2\left(\mathcal{R}^2\right)$

The feasible set, the set of feasible decisions

$$\Omega = \{x \in \mathcal{R}^2 | h(x) \geq 0, g(x) = 0\}$$

The minimiser x^*, at point •



Overview (cont.)

$$
\begin{aligned}
\min_{w \in \mathcal{R}^N} \quad & f(w) \\
\text{subject to} \quad & g(w) = 0 \\
& h(w) \leq 0
\end{aligned}
$$

We define the **level set** L to be the set of points w such that $f(w) = c$, in which $c \in \mathcal{R}$

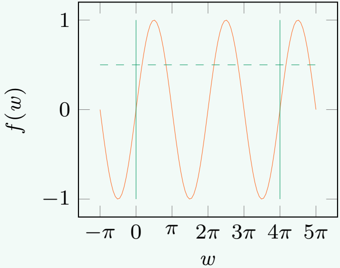
$$\{w \in \mathcal{R}^N : f(w) = c\}$$

We define the **sublevel set** L to be the set of points w such that $f(w) \leq c$, with $c \in \mathcal{R}$

$$\{w \in \mathcal{R}^N : f(w) \leq c\}$$

Overview (cont.)

Example



Consider the optimisation problem

$$\begin{aligned} \min_{w \in \mathcal{R}} \quad & \sin(w) \\ \text{subject to} \quad & w \geq 0 \\ & 4\pi - w \geq 0 \end{aligned}$$

Level set for $c = 0.5$

$$\{w \in \mathcal{R} : f(w) = 0.5\}$$

Sublevel set for $c = 0.5$

$$\{w \in \mathcal{R} : f(w) \leq 0.5\}$$

Overview (cont.)

$$\begin{aligned} \min_{w \in \mathcal{R}^N} \quad & f(w) \\ \text{subject to} \quad & g(w) = 0 \\ & h(w) \geq 0 \end{aligned}$$

A point $w \in \mathcal{R}^N$ is the **global minimiser** of the objective function f, given the constraint functions g and h, if and only if

$$\begin{aligned} & w^* \in \Omega \\ & f(w) \geq f(w^*), \text{ for all } w \in \Omega \end{aligned}$$

- The global minimiser is the point for which the constrained objective is the smallest
- Note that the global minimiser is not necessarily unique

The **global minimum** is the value $f(w^*)$ of the objective at the global minimiser w^*

- The global minimum is unique

Overview (cont.)

$$\begin{aligned} \min_{w \in \mathcal{R}^N} \quad & f(w) \\ \text{subject to} \quad & g(w) = 0 \\ & h(w) \geq 0 \end{aligned}$$

Existence of a global minimiser (Weierstrass)

Let the set $\Omega = \{w \in \mathcal{R}^N \,\big|\, h(w) \geq 0, g(w) = 0\}$ be non-empty, bounded and closed

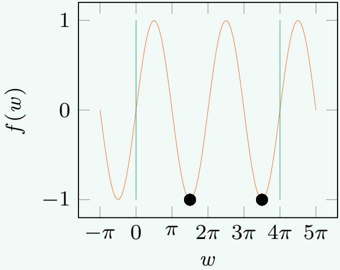
- ⇝ As always, we assume that $f : \Omega \to \mathcal{R}$ is at least $\mathcal{C}^1$
- ⇝ Then, there exists at least one global minimiser

Knowing that there is a global minimiser does not suggest an algorithm to find it

- Importantly, the objective function must be defined over a compact set
- (Weierstrass does not provide guarantees for unconstrained problems)

Overview (cont.)

Example



Consider the optimisation problem

$$\begin{aligned} \min_{w \in \mathcal{R}} \quad & \sin(w) & \\ \text{subject to} \quad & w & \geq 0 \\ & 4\pi - w & \geq 0 \end{aligned}$$

There are two global minimisers

- One global minimum

□

When the global minimiser is unique, then it is called the **strict global minimiser**

$$\begin{aligned} & w^* \in \Omega \\ & f(w) > f(w^*), \text{ for all } w \in \Omega \backslash \{w^*\} \end{aligned}$$

Overview (cont.)

$$\begin{aligned} \min_{w \in \mathcal{R}^N} \quad & f(w) \\ \text{subject to} \quad & g(w) = 0 \\ & h(w) \geq 0 \end{aligned}$$

A point $w \in \mathcal{R}^N$ is the **local minimiser** of the objective function f, given the constraint functions g and h, if and only if

$$w^* \in \Omega$$

and there exists an open ball $\mathcal{N}(w^*)$ about w^* such that

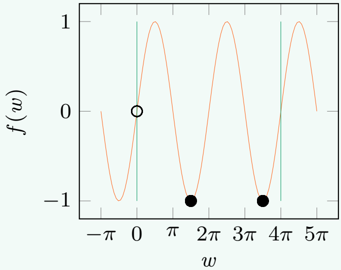
$$f(w) \geq f(w^*) \text{ for all } w \in \mathcal{N}(w) \cap \Omega$$

- The value $f(w^*)$ is the **local minimum**

When the local minimiser is unique in $\mathcal{N}(w^*)$, then it is a **strict local minimiser**

$$f(w) > f(w^*), \text{ for all } w \in \mathcal{N}(w) \cap \Omega \backslash \{w^*\}$$

Example

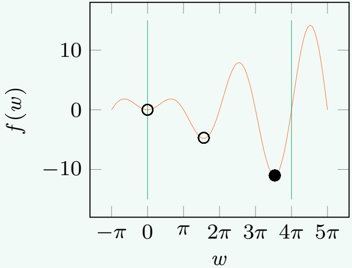


Consider the optimisation problem

$$\begin{aligned} \min_{w \in \mathcal{R}} \quad & \sin(w) \\ \text{subject to} \quad & w && \geq 0 \\ & 4\pi - w && \geq 0 \end{aligned}$$

There are three local minimisers

- Two global minimisers



Consider the optimisation problem

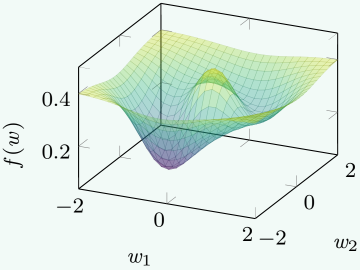
$$\begin{aligned} \min_{w \in \mathcal{R}} \quad & w\sin(w) \\ \text{subject to} \quad & w && \geq 0 \\ & 4\pi - w && \geq 0 \end{aligned}$$

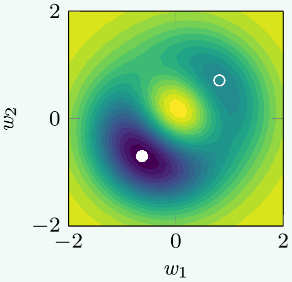
There are three local minimisers

- One global minimiser

Overview (cont.)

Example





$$\min_{w \in \mathcal{R}^2} \quad \frac{2}{5} - \frac{1}{10}\left(5w_1^2 + 5w_2^2 + 3w_1 w_2 - w_1 - 2w_2\right) e^{-\left(w_1^2 + w_2^2\right)}$$

$$w_1 + 2 \geq 0$$

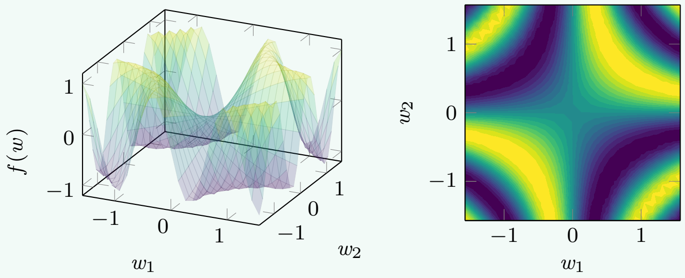
$$w_1 - 2 \geq 0$$

$$w_2 + 2 \geq 0$$

$$w_2 - 2 \geq 0$$

Overview (cont.)

Example



$$
\begin{aligned}
\min_{w \in \mathcal{R}^2} \quad & \sin(\pi w_1 w_2) + 1 \\
& w_1 + 3/2 \geq 0 \\
& w_1 - 3/2 \geq 0 \\
& w_2 + 3/2 \geq 0 \\
& w_2 - 3/2 \geq 0
\end{aligned}
$$

□

Overview (cont.)

$$\begin{aligned} \min_{w \in \mathcal{R}^N} \quad & f(w) \\ \text{subject to} \quad & g(w) = 0 \\ & h(w) \leq 0 \end{aligned}$$

From the given definitions, we understand that to be able to determine the state (global or local) of minimiser w^*, we need to describe the feasibility set in its neighbourhood

$$h(w) = \begin{bmatrix} h_1(w) \\ h_2(w) \\ \vdots \\ h_{N_h}(w) \end{bmatrix}$$

An inequality constraint $h_i(w) \leq 0$ is said to be an **active inequality constraint** at $w^* \in \Omega$ if and only if $h_i(w) = 0$, otherwise it is an **inactive inequality constraint**

- The index set of active inequality constraints is $\mathcal{A}(w^*) \subset \{1, 2, \dots, N_h\}$
- The index set $\mathcal{A}(w^*)$ is denoted as the **active set**
- The cardinality of the active set, $N_{\mathcal{A}} = |\mathcal{A}(w^*)|$

Classification

Nonlinear optimisation

Classification

Nonlinear programs (NLPs, smooth functions)

$$\begin{aligned} \min_{w \in \mathcal{R}^N} \quad & f(w) \\ \text{subject to} \quad & g(w) = 0 \\ & h(w) \geq 0 \end{aligned}$$

Functions f, g, and g are continuously differentiable at least once, often twice or more

The problem data

- $f : \mathcal{R}^N \rightarrow \mathcal{R}$, with $f \in \mathcal{C}^1\left(\mathcal{R}^N\right)$ or more
- $g : \mathcal{R}^N \rightarrow \mathcal{R}^{N_g}$, with $g \in \mathcal{C}^1\left(\mathcal{R}^N\right)$ or more
- $h : \mathcal{R}^N \rightarrow \mathcal{R}^{N_h}$, with $h \in \mathcal{C}^1\left(\mathcal{R}^N\right)$ or more

Differentiability of all problem functions allow to use algorithms based on derivatives

- We consider the nonlinear program as the more general formulation
- No explicit structure to exploit in the general formulation

Classification | Linear programs

Linear programs (LPs, affine functions)

$$
\begin{aligned}
\min_{w \in \mathcal{R}^N} \quad & \underbrace{c^T w}_{f(w) \quad (c_0)} \\
\text{subject to} \quad & \underbrace{Aw - b}_{g(w)} = 0 \\
& \underbrace{Cw - d}_{h(w)} \geq 0
\end{aligned}
$$

Functions f, g, and g are affine, there are efficient solutions (active set/interior point)

The problem data

- $c \in \mathcal{R}^N$ ($c_0 \in \mathcal{R}^N$)
- $A \in \mathcal{R}^{N_g \times N}$ and $b \in \mathcal{R}^{N_g}$
- $C \in \mathcal{R}^{N_h \times N}$ and $d \in \mathcal{R}^{N_h}$

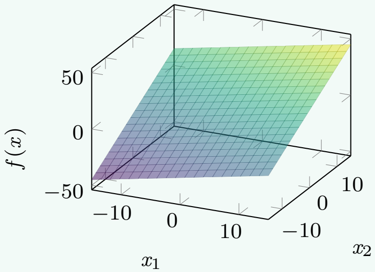
Commonly used software packages for LPs: CPLEX, SOPLEX, lp_solve, lingo, linprog

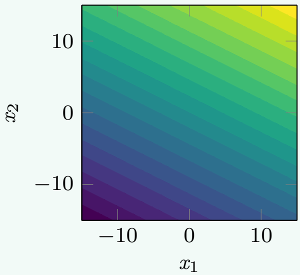
Classification | Linear programs (cont.)

Example

A linear program

$$\begin{aligned} \min_{w \in \mathcal{R}^2} \quad & \begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \\ \text{subject to} \quad & -10 \leq w_1 \leq 10 \\ & -10 \leq w_2 \leq 10 \end{aligned}$$





Classification | Linear programs (cont.)

$$\begin{aligned} \min_{w \in \mathcal{R}^2} \quad & \begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \\ \text{subject to} \quad & -10 \leq w_1 \leq 10 \\ & -10 \leq w_2 \leq 10 \end{aligned}$$

Equivalently, we have

$$\begin{aligned} \min_{w \in \mathcal{R}^2} \quad & \underbrace{w_1 + 2w_2}_{f(w)} \\ \text{subject to} \quad & \underbrace{w_1 + 10}_{h_1(w)} \geq 0 \\ & \underbrace{-w_1 + 10}_{h_2(w)} \geq 0 \\ & \underbrace{w_2 + 10}_{h_3(w)} \geq 0 \\ & \underbrace{-w_2 + 10}_{h_4(w)} \geq 0 \end{aligned}$$

- $f : \mathcal{R}^2 \rightarrow \mathcal{R}$
- $h : \mathcal{R}^2 \rightarrow \mathcal{R}^4$

Classification | Quadratic programs

Quadratic programs (QPs, linear-quadratic objective + affine constraints)

$$\begin{aligned} \min_{w \in \mathcal{R}^N} \quad & \underbrace{c^T w + \frac{1}{2} w^T B w}_{f(w)} \\ \text{subject to} \quad & \underbrace{Aw - b}_{g(w)} = 0 \\ & \underbrace{Cw - d}_{h(w)} \geq 0 \end{aligned}$$

Function f is linear-quadratic and functions g and h are affine

The problem data

- $c \in \mathcal{R}^N$
- ⇝ $B \in \mathcal{R}^{N \times N}$, symmetric
- $A \in \mathcal{R}^{N_g \times N}$ and $b \in \mathcal{R}^{N_g}$
- $C \in \mathcal{R}^{N_h \times N}$ and $d \in \mathcal{R}^{N_h}$

Commonly used packages for QPs: CPLEX, MOSEK, qpOASES, OOQP, quadprog

Classification | Quadratic programs (cont.)

Example

$$\min_{w \in \mathcal{R}^2} \underbrace{\begin{bmatrix} c_1 & c_2 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}^T \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}}_{c_1 w_1 + c_2 w_2 + \frac{1}{2}(b_{11} w_1^2 + (b_{12} + b_{21}) w_1 w_2 + b_{22} w_2^2)}$$

$$\text{subject to} \quad \underbrace{\begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \\ c_{31} & c_{32} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} - \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}}_{q(w)} = 0$$

$$\underbrace{\begin{bmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \\ d_{31} & d_{32} \\ d_{41} & d_{d2} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} - \begin{bmatrix} d_1 \\ d_2 \\ d_3 \\ d_4 \end{bmatrix}}_{h(w)} \geq 0$$

- $f : \mathcal{R}^2 \rightarrow \mathcal{R}$
- $g : \mathcal{R}^2 \rightarrow \mathcal{R}^3$
- $h : \mathcal{R}^2 \rightarrow \mathcal{R}^4$

Classification | Quadratic programs (cont.)

$$\underbrace{c^T w + \frac{1}{2} w^T B w}_{f(w)}$$

If matrix B is positive semi-definite ($z^T B z \geq 0$, for all $z \in \mathcal{R}^N$), then the QP is convex

- If B is positive definite ($z^T B z > 0$, for all $z \in \mathcal{R}^N$), the QP is strictly convex

The positive- and semi-positive definiteness of matrix B is checked from its eigenvalues

Generalised inequality for symmetric matrices

Positive semi-definite matrix, $B \succeq 0$

$$\min \lambda_{\min}(B) \geq 0$$

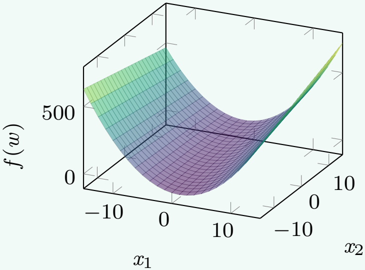
Positive definite matrix, $B \succ 0$

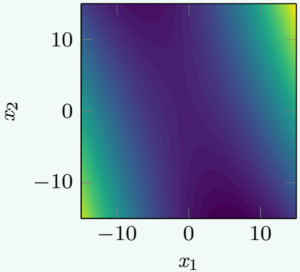
$$\min \lambda_{\min}(B) > 0$$

Example

A convex quadratic program

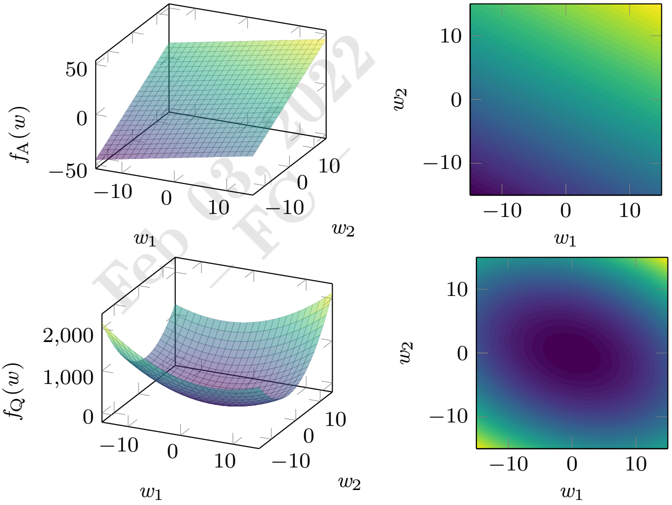
$$\begin{aligned}\min_{w \in \mathcal{R}^2} \quad & \begin{bmatrix}1 & 2\end{bmatrix}\begin{bmatrix}w_1\\ w_2\end{bmatrix} + \frac{1}{2}\begin{bmatrix}w_1\\ w_2\end{bmatrix}^T \begin{bmatrix}5 & 2\\ 2 & 10\end{bmatrix}\begin{bmatrix}w_1\\ w_2\end{bmatrix}\\ \text{subject to} \quad & -10 \leq w_1 \leq 10\\ & -10 \leq w_2 \leq 10\end{aligned}$$





Convex quadratic problems are easy to solve (the local minimum is a global minimum)

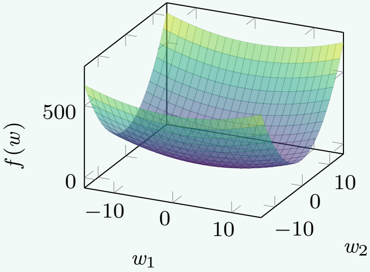
$$f(w) = \underbrace{\begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}}_{f_{\mathrm{A}}(w)} + \underbrace{\frac{1}{2} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}^T \begin{bmatrix} 5 & 2 \\ 2 & 10 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}}_{f_{\mathrm{Q}}(w)}$$

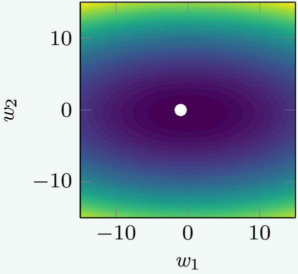


Example

A strictly-convex quadratic program

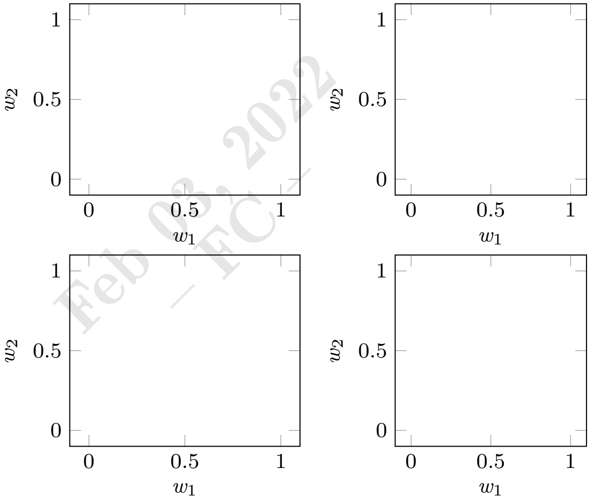
$$\min_{w \in \mathcal{R}^2} \quad \begin{bmatrix} 0 & 2 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}^T \begin{bmatrix} 5 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$
$$\text{subject to} \quad -10 \leq w_1 \leq 10$$
$$-10 \leq w_2 \leq 10$$





Strictly-convex quadratic programs are the easiest to solve (a unique global minimiser)

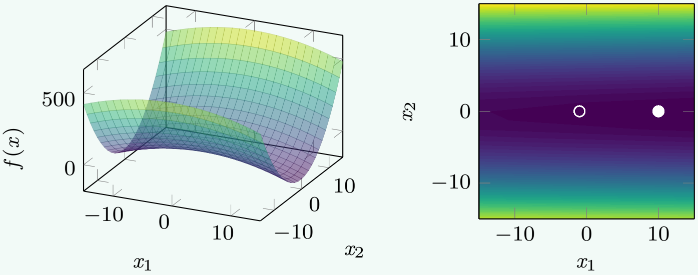
$$f(w) = \begin{bmatrix} 0 & 2 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}^T \begin{bmatrix} 5 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$



Example

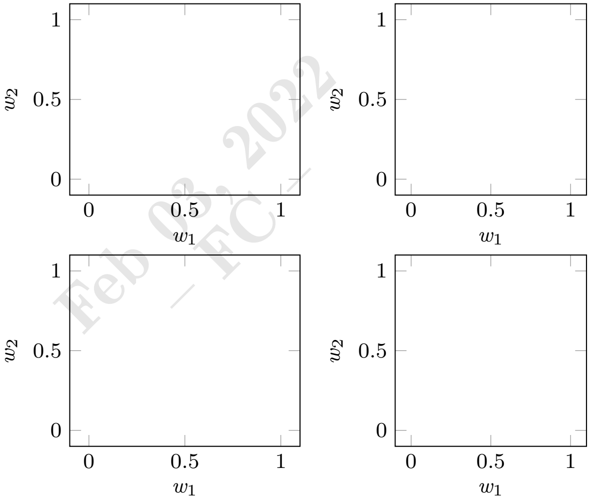
A non-convex quadratic program

$$\begin{aligned} \min_{w \in \mathcal{R}^2} \quad & \begin{bmatrix} 0 & 2 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}^T \begin{bmatrix} 5 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \\ \text{subject to} \quad & -10 \leq w_1 \leq 10 \\ & -10 \leq w_2 \leq 10 \end{aligned}$$



Non-convex quadratic programs can be difficult to solve (for a global minimiser)

$$f(w) = \begin{bmatrix} 0 & 2 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}^T \begin{bmatrix} 5 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$



Classification | Convex programs

Linear and convex quadratic programs are part of an important class of problems

Convex programs

$$\begin{aligned} \min_{x \in \mathcal{R}^N} \quad & f(x) \\ \text{subject to} \quad & g(x) = 0 \\ & h(x) \leq 0 \end{aligned}$$

The feasible set $\Omega = \{x \in \mathcal{R}^N : h(x) \geq 0, g(x) = 0\}$ and function f is also convex

There exists a wide availability of packages that can be used for convex problems

- YAMILP (based on SDP3 and SeDuMi) and CVX (Matlab-based)

Classification | Mixed-integer programs

Mixed-integer nonlinear programs (MINLPs, real and integer decision vars)

$$\begin{aligned} \min_{\substack{w \in \mathcal{R}^N \\ v \in \mathcal{Z}^M}} \quad & f(w, v) \\ \text{subject to} \quad & g(w, v) = 0 \\ & h(w, v) \geq 0 \end{aligned}$$

Mixed-integer nonlinear programs, smooth functions with full or partial relaxations

- Relaxation, by letting variables z to be real vectors

$$\begin{aligned} \min_{\substack{w \in \mathcal{R}^N \\ v \in \mathcal{R}^M}} \quad & f(w, v) \\ \text{subject to} \quad & g(w, v) = 0 \\ & h(w, v) \geq 0 \end{aligned}$$

- Convexification, with branch-and-bound techniques

Convex optimisation

Nonlinear optimisation

Convex optimisation

Linear programs and convex quadratic programs are **convex optimisation** problems

- An important subclass of continuous optimisation problems

⇝ Objective function must be a convex function

⇝ The feasible set must be a convex set

For this class of problems, any local minimiser is a global minimiser (given w/o proof)

Convex optimisation | Convex sets

Convex sets

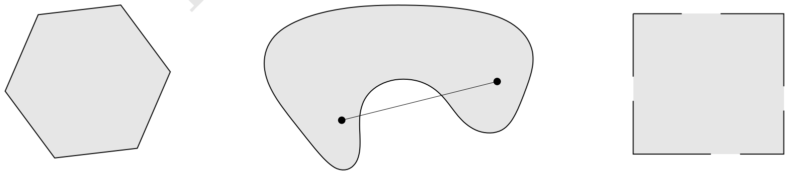
Consider set $\Omega \subset \mathcal{R}^N$

Set Ω is convex if and only if, for all pairs $(w, w') \in \Omega$ and scalars $\lambda \in [0, 1]$, we have

$$w + \lambda(w' - w) \in \Omega$$

- $w + \lambda(w' - w)$ are points on the line segment bounded by w and w'
- When $\lambda = 0$ we obtain point w, when $\lambda = 1$ we obtain w'

Equivalently, we say that 'all connecting segments lie in the set'



Convex optimisation | Convex functions

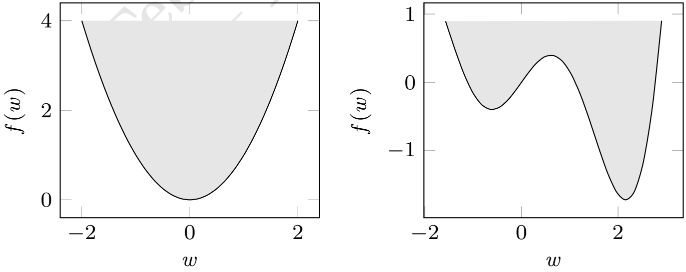
Convex functions

Consider some function $f : \Omega \rightarrow \mathcal{R}$

Function f is convex if and only if, set Ω is convex set and for all the pairs $(w, w') \in \Omega$ and scalars $\lambda \in [0, 1]$, we have

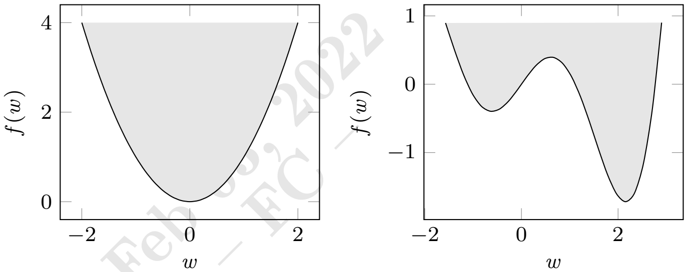
$$f\left(w + \lambda(w - w')\right) \leq f\left(w\right) + \lambda(f\left(w'\right) - f\left(w\right))$$

- $f\left(w\right) + \lambda(f\left(w'\right) - f\left(w\right))$ are points on the segment bounded by $f\left(w\right)$ and $f\left(w'\right)$
- $f\left(w + \lambda(w - w')\right)$ are function values at points in the segment $w + \lambda(w - w')$



Convex optimisation | Convex functions (cont.)

Equivalently, we say that 'all secants are above the graph of f'



Similarly, we can say that 'the epigraph of f is a convex set'

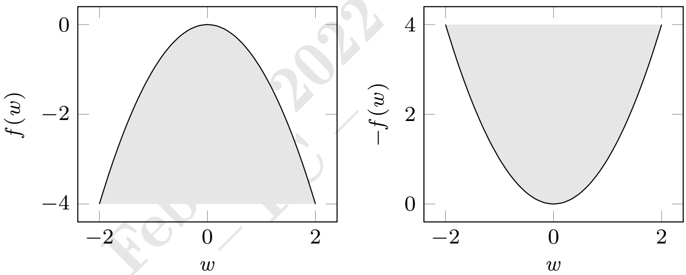
$$\text{epi}(f) = \{(w, s) \in \mathcal{R}^N \times \mathcal{R} : x \in \Omega, s \geq f(w)\}$$

This theorem combines convexity of sets and functions

Convex optimisation | Convex functions (cont.)

Concave functions

A function $f : \Omega \to \mathcal{R}$ is a concave function if function $-f$ is convex



The domain of definition Ω of the function $(-f)$ must be a convex set

The Hessian matrix of a concave function is negative semi-definite

$$\nabla^2 f(w) \preceq 0$$

Convex optimisation | Properties

Convex programs

$$\begin{aligned} \min_{x \in \mathcal{R}^N} \quad & f(x) \\ \text{subject to} \quad & g(x) = 0 \\ & h(x) \leq 0 \end{aligned}$$

The feasible set $\Omega = \{x \in \mathcal{R}^N : h(x) \geq 0, g(x) = 0\}$ and function f is also convex

---

For convex programs local optimality implies global optimality

- That is, every local minimiser is also a global minimiser
- Global optimality is retrieved from local information

Consider a local minimiser w^*, we have

$$f\left(w'\right) \geq f\left(w^*\right), \quad \text{for all } w' \in \Omega$$

Convex optimisation | Properties (cont.)

$$f\left(w'\right) \geq f\left(w^*\right), \quad \text{for all } w' \in \Omega$$

If w^* is a local minimiser, then for all $\overline{w} \in \mathcal{N}(w^*) \cap \Omega$ we have that $f\left(\overline{w}\right) \geq f\left(w^*\right)$

- By convexity of Ω, the segment

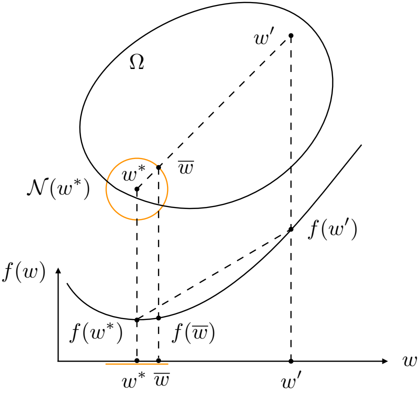
$$w^* + \lambda(w' - w^*) \in \Omega$$

- Point $\overline{w}$ is in the segment, thus

$$\begin{aligned} f\left(w^*\right) &\leq f\left(\overline{w}\right) \\ &\leq f\left(w^* + \lambda(w' - w^*)\right) \end{aligned}$$

- By convexity of f, we have

$$\begin{aligned} f\left(w^*\right) &\leq f\left(\overline{w}\right) \\ &\leq f\left(w^* + \lambda(w' - w^*)\right) \\ &\leq f\left(w^*\right) + \lambda(f\left(w'\right) - f\left(w^*\right)) \end{aligned}$$



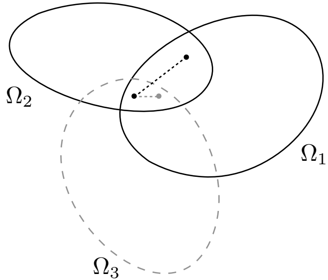
Subtract $f\left(w^*\right)$ from both sides, divide by $\lambda \neq 0$ ($\overline{w}$ is not w^*), and then rearrange

Convex optimisation | Convex sets and functions

Convexity-preserving operations for sets

- **Intersections**

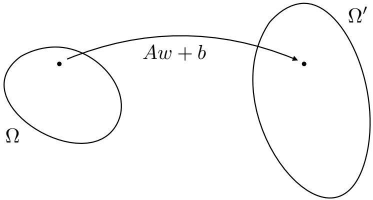
  The intersection of (finitely or infinitely many) convex sets is also a convex set



- **Affine images**

Affine transformations $\Omega' = A\Omega + b$ of a convex set Ω are also convex sets

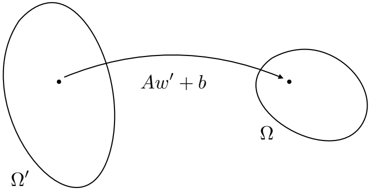
$$\Omega' = \{w' \in \mathcal{R}^M : \exists w \in \Omega : w' = Aw + b, A \in \mathcal{R}^{M \times N}, b \in \mathcal{R}^M\}$$



- **Affine pre-images**

If set Ω is convex, then there exists a convex set Ω' such that $\Omega = A\Omega' + b$

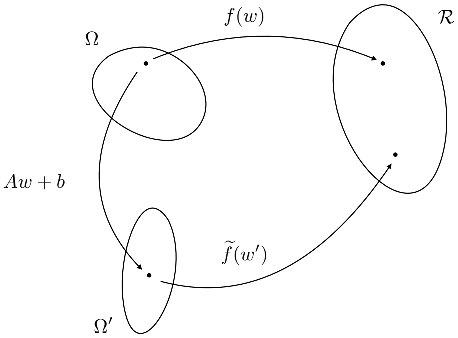
$$\Omega' = \{w' \in \mathcal{R}^M : w = Aw' + b, A \in \mathcal{R}^{N \times M}, b \in \mathcal{R}^N\}$$



Convex optimisation | Convex sets and functions (cont.)

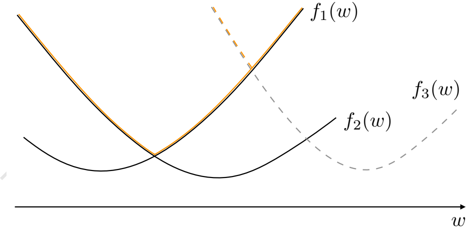
Convexity-preserving operations for functions

- The (point-wise) sum of two (or more) convex functions is also a convex function
- Positively weighted sums of two (or more) convex functions is a convex function
- Affine transformations $Aw + b$ of the independent variable $w \in \Omega$ of a convex function $f : \Omega \to \mathcal{R}$ lead to convex functions $\widetilde{f} : \Omega' \to \mathcal{R}$ from the set $\Omega' = \{w' \in \mathcal{R}^M | w' = Aw + b, w \in \Omega, A \in \mathcal{R}^{M \times N}, b \in \mathcal{R}^M\}$ such that $\widetilde{f}(w) = f(Aw + b)$



Convex optimisation | Convex sets and functions (cont.)

- The supremum $f(w) = \sup_{1,\dots,N_h} f_{n_h}(w)$ over a set of convex functions $\{f_{n_h}\}_{n_h=1}^{N_h}$ is a convex function, because its epigraph is the intersection of convex epigraphs



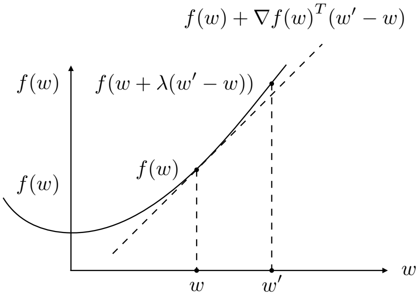
Convex optimisation | Convex sets and functions (cont.)

Convexity of $\mathcal{C}^1$ functions

Let $\Omega \in \mathcal{R}^N$ be a convex set and let $f : \Omega \rightarrow \mathcal{R}$ be a continuously differentiable function

Function $f \in \mathcal{C}^1\left(\mathcal{R}^N\right)$ is convex if and only if for all pairs of points $(w, w') \in \Omega$,

$$f\left(w'\right) \geq \underbrace{f\left(w\right) + \nabla f\left(w\right)^T\left(w' - w\right)}_{\text{Taylor's expansion at } w}$$



- Equivalently, was can say that 'all tangent lines lies below the graph of f'
- (Remember that by convexity 'all secant lines lies above the graph')

This theorem provides a possibility to check for convexity, by testing all pairs (w, w')

Convex optimisation | Convex sets and functions (cont.)

$$f\left(w'\right) \geq \underbrace{f\left(w\right) + \nabla f\left(w\right)^T\left(w' - w\right)}_{\text{Taylor's expansion at } w}$$

Suppose that f is a convex function over the convex set Ω

Because of the convexity of function f, we can write

$$f\left(w + \lambda(w' - w)\right) \leq f\left(w\right) + \lambda(f\left(w'\right) - f\left(w\right))$$

Rearranging, we get,

$$f\left(w + \lambda(w' - w)\right) - f\left(w\right) \leq \lambda(f\left(w'\right) - f\left(w\right))$$

Using the definition of (directional) derivative, we have

$$\begin{aligned}\nabla f\left(w\right)^T(w - w') &= \lim_{\lambda \to 0} \frac{f\left(w + \lambda(w - w')\right) - f\left(w\right)}{\lambda} \\ &\leq f\left(w'\right) - f\left(w\right)\end{aligned}$$

Convex optimisation | Convex sets and functions (cont.)

Convexity of $\mathcal{C}^2$ functions

Let $\Omega \in \mathcal{R}^N$ be a convex set and let $f : \Omega \rightarrow \mathcal{R}$ be twice continuously differentiable

Function $f \in \mathcal{C}^2\left(\mathcal{R}^N\right)$ is convex if, for any point $w \in \Omega$, we have

$$\nabla^2 f\left(w\right) \succeq 0$$

- The Hessian matrix must positive semi-definite

$$\min \lambda_{\min}(\nabla^2 f\left(w\right)) \geq 0$$

This theorem provides a possibility to check for convexity, by testing single pairs w

Convex optimisation | Convex sets and functions (cont.)

$$\nabla^2 f(w) \succeq 0$$

We consider the second-order Taylor's expansion of function f along $\lambda(w - w')$

$$f\left(w + \lambda(w' - w)\right) = f(w) + \lambda \nabla f(w)^T (w' - w) + \frac{1}{2}\lambda^2 (w' - w)^T \nabla^2 f(w)(w' - w) + \mathcal{O}(\lambda^2 (w' - w)^2)$$

Because of the convexity of function f, we have $f(w') \geq f(w) + \nabla f(w)^T (w' - w)$

$$f(w') - f(w) - \nabla f(w)^T (w' - w) \geq 0$$

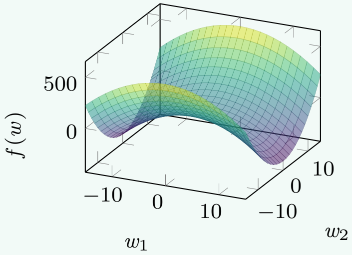
Thus,

$$f\left(w + \lambda(w - w')\right) - f(w) - \lambda \nabla f(w)^T (w - w') = \frac{1}{2}\lambda^2 (w - w')^T \underbrace{\nabla^2 f(w)}_{\succeq 0}(w - w') + \mathcal{O}(\lambda^2 (w - w')^2)$$

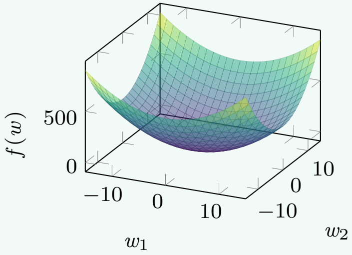
$$\geq 0$$

Convex optimisation | Convex sets and functions (cont.)

Example



$$f(w) = \frac{1}{2}\begin{bmatrix} w_1 \\ w_2 \end{bmatrix}^T \begin{bmatrix} 5 & 0 \\ 0 & -3 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$
$$\rightsquigarrow \nabla^2 f(w) \prec 0$$



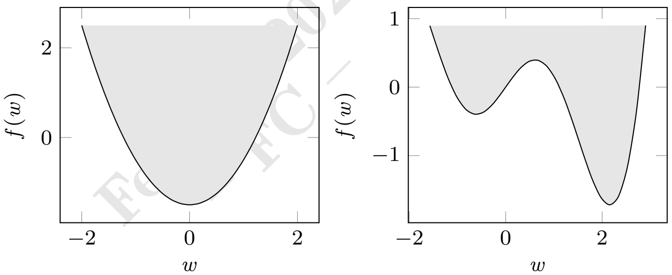
$$f(w) = \frac{1}{2}\begin{bmatrix} w_1 \\ w_2 \end{bmatrix}^T \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$
$$\rightsquigarrow \nabla^2 f(w) \succ 0$$

Convex optimisation | Convex sets and functions (cont.)

Convexity of level-sets

Consider the level set $\{w \in \Omega : f(w) \leq c, c \in \mathcal{R}\}$ of any convex function $f : \Omega \rightarrow \mathcal{R}$

- The level-set is a convex set, for any constant c



The theorem suggests that convex sets can be created from functions with inequalities

Convex optimisation | Convex sets and functions (cont.)

Example

Consider a collection of convex functions $\{f_{n_h} : \mathcal{R}^N \rightarrow \mathcal{R}\}_{n_h=1}^{N_h}$

Consider the intersection of their sub-level sets

$$\Omega = \{w \in \mathcal{R}^N : \{f_{n_h}(w) \leq 0\}_{n_h=1}^{N_h}\}$$

Set Ω is a convex set

Level sets Ω_{n_h} of convex functions are convex sets

- Their intersection is also a convex set

$$\Omega = \bigcap_{n_h=1}^{N_h} \Omega_{n_h}$$

□

Convex optimisation | Formulation

Consider the general form of a nonlinear optimisation problem

$$\begin{aligned} \min_{w \in \mathcal{R}^N} \quad & f(w) \\ \text{subject to} \quad & g(w) = 0 \\ & h(w) \geq 0 \end{aligned}$$

We defined the feasible set Ω to be the set of points w that satisfy all the constraints

$$\Omega = \left\{ w \in \mathcal{R}^N \middle| g(w) = 0, h(w) \geq 0 \right\}$$

In order to have a feasible set Ω that is convex, the equality constraints must be affine functions and the (positive defined) inequality constraints must be concave functions

---

If f is convex and the above holds, then the problem is convex (a sufficient condition)

$$\begin{aligned} \min_{w \in \mathcal{R}^N} \quad & f(w) && \text{(Objective function, convex)} \\ \text{subject to} \quad & \underbrace{Aw - b}_{g(w)} = 0 && \text{(Equality constraints, affine)} \\ & \widetilde{h}(w) \leq 0 && \text{(Inequality constraints, convex)} \end{aligned}$$

Convex optimisation | Formulation (cont.)

$$\begin{aligned} \min_{w \in \mathcal{R}^N} \quad & f(w) && \text{(Objective function, convex)} \\ \text{subject to} \quad & \underbrace{Aw - b}_{g(w)} = 0 && \text{(Equality constraints, affine)} \\ & \widetilde{h}(w) \leq 0 && \text{(Inequality constraints, convex)} \end{aligned}$$

The inequality constraint functions $\widetilde{h}_1, \widetilde{h}_2, \dots, \widetilde{h}_{N_h}$ must be convex functions

- We know that their intersection is a convex set

The equality constraint function $g_1, g_2, \dots, g_{N_h}$ must be affine functions

- They are affine pre-images to a convex set, point 0

The intersection of a convex set with a convex set is a convex set

⇝ The feasible set Ω is convex

Convex optimisation | Optimality

First-order optimality conditions for convex problems (constrained)

Consider the convex problem with set $\Omega = \{w \in \mathcal{R}^N : g(w) = 0, h(w) \leq 0\}$

$$\begin{aligned} \min_{w \in \mathcal{R}^N} \quad & f(w) && \text{(Objective function, convex and differentiable)} \\ \text{subject to} \quad & Aw + b = 0 && \text{(Equality constraints, affine)} \\ & h(w) \leq 0 && \text{(Inequality constraints, convex)} \end{aligned}$$

For convex optimisation problems, a local minimiser is also a global minimiser

Points $w* \in \Omega$ is a global minimiser if and only if, for all $w \in \Omega$

$$\nabla f(w^*)^T (w - w^*) \geq 0$$

Convex optimisation | Optimality (cont.)

$$\nabla f\left(w^*\right)^T(w - w^*) \geq 0$$

If the condition holds, by the convexity characterisation of $\mathcal{C}^1$ functions we have

$$\begin{aligned} f\left(w'\right) &\geq f\left(w\right) + \nabla f\left(w^*\right)^T(w' - w^*) \quad (\text{for all } w' \in \Omega) \\ &\geq f\left(w^*\right) \end{aligned}$$

We can also assume the existence of $w' \in \Omega$ such that $\nabla f\left(w^*\right)(w' - w^*) < 0$

Then, by a first-order Taylor's expansion

$$f\left(w^* + \lambda(w' - w^*)\right) \approx f\left(w^*\right) + \lambda \underbrace{\nabla f\left(w^*\right)^T(w' - w^*)}_{<0}$$

For some small λ, this yields

$$f\left(w^* + \lambda(w' - w^*)\right) < f\left(w^*\right)$$

Convex optimisation | Optimality (cont.)

First-order optimality conditions for convex problems (unconstrained)

Consider the convex optimisation problem with feasibility set $\Omega = \mathcal{R}^N$

$$\min_{w \in \mathcal{R}^N} \quad f(w) \quad \text{(Convex and differentiable)}$$

A point $w* \in \Omega$ is a global minimiser if and only if the following holds

$$\nabla f(w^*)^T = 0$$

Convex optimisation | Optimality (cont.)

Example

Consider the strictly convex quadratic problem

$$\min_{w \in \mathcal{R}^N} \quad \Big(c^T w + \frac{1}{2} w^T \underbrace{B}_{>0} w \Big)$$

For the gradient vector evaluated at the minimiser, we have

$$\nabla f(w^*) = c + Bw = 0$$

By solving the system of linear equations, we get

$$w^* = -B^{-1} c$$

By substitution, we get the optimal function value

$$f(w^*) = -\frac{1}{2} c^T B^{-1} c$$

□