MA427 Lecture 8
Convex optimisation: basics

Giacomo Zambelli

Department of Mathematics

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

4 March, 2019

Convex Optimisation

- Today: calculus reminders, convexity, first-order conditions
- Lecture 9: Lagrangian Duality and KKT conditions
- Lecture 10: algorithms for convex optimisation

Optimisation problems

$$\begin{aligned} \min \quad & f_0(x) \\ & f_i(x) \leq 0 \quad i = 1, \dots, m \\ & h_i(x) = 0 \quad i = 1, \dots, k \\ & x \in \mathcal{D} \end{aligned}$$

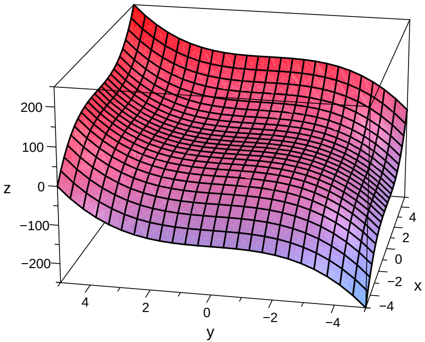
Where $\mathcal{D}$ is the *domain* of the problem:

$$\mathcal{D} = \left(\bigcap_{i=0}^{m} \mathbf{dom}\, f_i \right) \bigcap \left(\bigcap_{i=1}^{k} \mathbf{dom}\, h_i \right).$$

Graph of a function

Graph of $f : \mathbb{R}^n \to \mathbb{R}$:

$$\{(x, f(x)) \in \mathbb{R}^{n+1} \ : \ x \in \textbf{dom}\, f\}$$



Graph of
$f(x_1, x_2) = x_1^3 + x_2^3$

Epigraph: points "above" the graph:

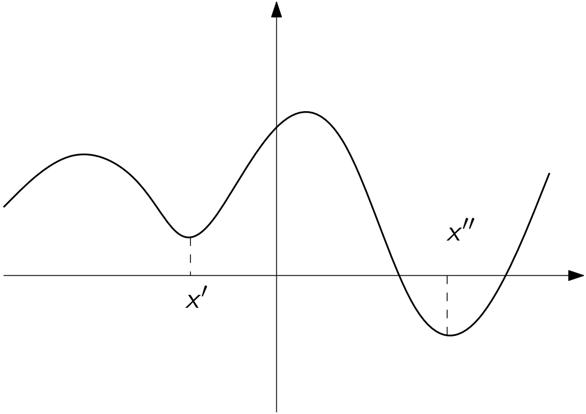
$$\{(x, t) \in \mathbb{R}^{n+1} \ : \ x \in \textbf{dom}\, f, \ f(x) \leq t\}.$$

Global and Local Optima

$f : \mathbb{R}^n \to \mathbb{R}$, $X \subseteq \textbf{dom} f$. Point $x^* \in X$ is a

- global minimum for f in X if $f(x) \geq f(x^*)$ for all $x \in X$.
- local minimum for f in X if there exists $\varepsilon > 0$ such that $f(x) \geq f(x^*)$ for all $x \in X$ such that $\|x - x^*\| \leq \varepsilon$.
- global maximum for f in X if $f(x) \leq f(x^*)$ for all $x \in X$.
- local maximum for f in X if there exists $\varepsilon > 0$ such that $f(x) \leq f(x^*)$ for all $x \in X$ such that $\|x - x^*\| \leq \varepsilon$.

Global and Local Optima



Not all local minima are also global minima.

Gradients

The gradient of $f : \mathbb{R}^n \to \mathbb{R}$ at point $x \in \textbf{dom}\, f$ is the vector

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

$\frac{\partial f(x)}{\partial x_i}$: partial derivative at point x with respect to the variable x_i.

Differentiable functions

$f : \mathbb{R}^n \to \mathbb{R}$ is differentiable at a point x in the interior of $\mathbf{dom}\, f$ if

$$\lim_{z \in \mathbf{dom}\, f,\, z \to x} \frac{|f(z) - f(x) - \nabla f(x)^\top (z - x)|}{\|z - x\|} = 0.$$

Differentiable functions

$f : \mathbb{R}^n \to \mathbb{R}$ is differentiable at a point x in the interior of $\mathbf{dom}\, f$ if

$$\lim_{z \in \mathbf{dom}\, f,\, z \to x} \frac{|f(z) - f(x) - \nabla f(x)^\top (z - x)|}{\|z - x\|} = 0.$$

It means that, near x, f is well approximated by the affine function

$$h : z \mapsto f(x) + \nabla f(x)^\top (z - x).$$

Differentiable functions

$f : \mathbb{R}^n \to \mathbb{R}$ is differentiable at a point x in the interior of $\mathbf{dom}\, f$ if

$$\lim_{z \in \mathbf{dom}\, f,\, z \to x} \frac{|f(z) - f(x) - \nabla f(x)^\top (z - x)|}{\|z - x\|} = 0.$$

It means that, near x, f is well approximated by the affine function

$$h : z \mapsto f(x) + \nabla f(x)^\top (z - x).$$

f is differentiable if f is continuous, $\mathbf{dom}\, f$ is an open set, and f is differentiable at every point $x \in \mathbf{dom}\, f$.

Differentiable functions

Example. $f(x_1, x_2) = \log(x_1/x_2)$ ($\mathbf{dom}\, f = \{x \in \mathbb{R}^2 : x_1, x_2 > 0\}$) is differentiable:

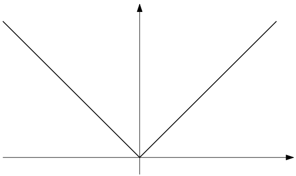
$$\nabla f(x) := \begin{bmatrix} 1/x_1 \\ -1/x_2 \end{bmatrix} \quad \forall x \in \mathbf{dom}\, f$$

Differentiable functions

Example. $f(x_1, x_2) = \log(x_1/x_2)$ ($\mathbf{dom}\, f = \{x \in \mathbb{R}^2 : x_1, x_2 > 0\}$) is differentiable:

$$\nabla f(x) := \begin{bmatrix} 1/x_1 \\ -1/x_2 \end{bmatrix} \quad \forall x \in \mathbf{dom}\, f$$

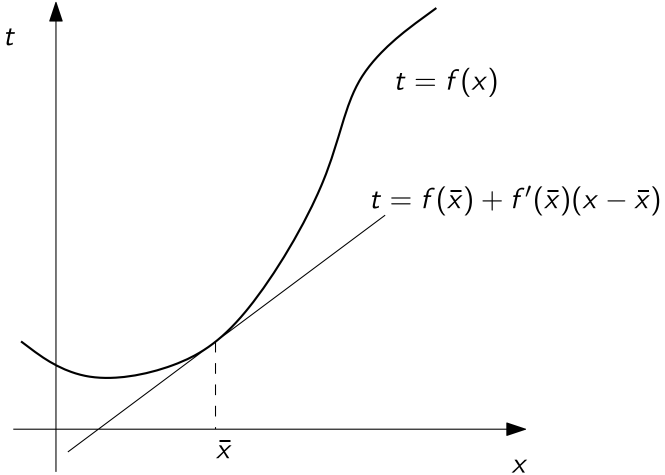
Example. $f : x \mapsto |x|$ is not differentiable, because its derivative does not exist at $x = 0$.



Gradients

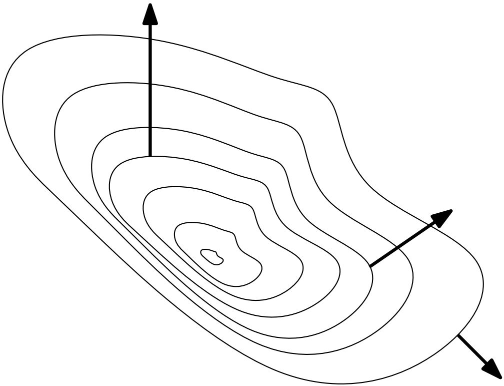
The hyperplane tangent to the graph of f in $(\bar{x}, f(\bar{x}))$ is

$$H = \{(x,t) \in \mathbb{R}^n \times \mathbb{R} \,:\, t = f(\bar{x}) + \nabla f(\bar{x})^\top (x - \bar{x})\}.$$



Gradients

- Gradient at point $\bar{x}$ is orthogonal to the contour of f at point $\bar{x}$ (i.e. $\{x : f(x) = f(\bar{x})\}$), and pointing in the direction of the steepest ascent.
- The direction of maximum ascent in H is the direction of the gradient $\nabla f(\bar{x})$, and the slope of H in the direction of $\nabla f(\bar{x})$ is $\|\nabla f(x)\|$.

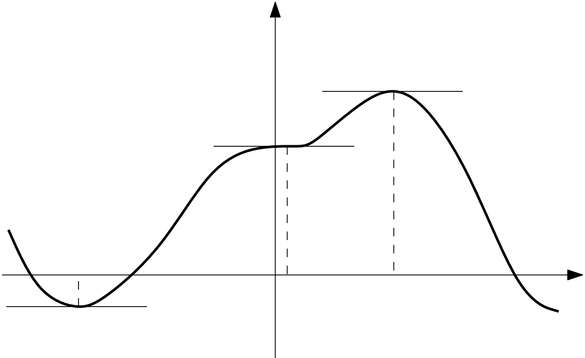


First-order necessary conditions

Theorem

Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable, and let x^* be a point in $\mathbf{dom}\, f$*. If x^* is a local maximum or a local minimum for f, then $\nabla f(x^*) = 0$.*

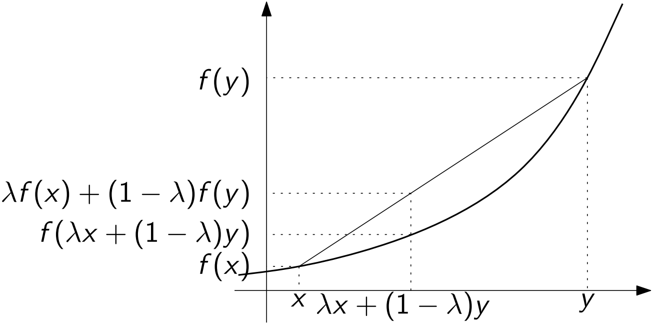
Conditions are necessary, but not sufficient:



Convex functions

Definition. $f : \mathbb{R}^n \to \mathbb{R}$ is *convex* if $\mathbf{dom}\, f$ is convex and, for every $x, y \in \mathbf{dom}\, f$, and $\lambda \in [0, 1]$,

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y).$$



A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if and only if the epigraph of f is a convex set.

Affine functions are convex $f(x) = p^\top x + r$.

Univariate convex functions

Theorem

Let $f : \mathbb{R} \to \mathbb{R}$ be twice differentiable on $\mathbf{dom}\, f$*. Then f is convex if and only if $f''(x) \geq 0$ for all $x \in$* $\mathbf{dom}\, f$*.*

Examples:

- $f(x) = -\log x$, where $\mathbf{dom}\, f = \{x \in \mathbb{R} : x > 0\}$.
- $f(x) = e^x$, where $\mathbf{dom}\, f = \mathbb{R}$.
- $f(x) = 1/x$, where $\mathbf{dom}\, f = \{x \in \mathbb{R} : x > 0\}$.
- $f(x) = x \log x$, where $\mathbf{dom}\, f = \{x \in \mathbb{R} : x > 0\}$.

Operations that preserve convexity

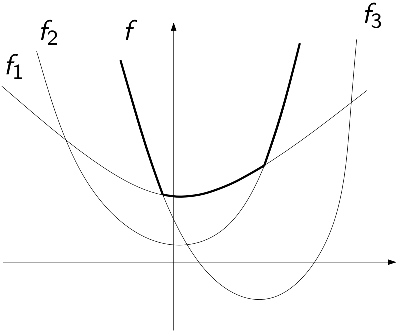
- Nonnegative linear combination. If $f_1, \ldots, f_m : \mathbb{R}^n \to \mathbb{R}$ are convex and $\lambda_1, \ldots, \lambda_m \geq 0$, then $f = \lambda_1 f_1 + \cdots + \lambda_m f_m$ is convex,

Operations that preserve convexity

- Nonnegative linear combination. If $f_1, \ldots, f_m : \mathbb{R}^n \to \mathbb{R}$ are convex and $\lambda_1, \ldots, \lambda_m \geq 0$, then $f = \lambda_1 f_1 + \cdots + \lambda_m f_m$ is convex,
- Point-wise supremum. If $f_\alpha : \mathbb{R}^n \to \mathbb{R}$ $(\alpha \in \mathcal{A})$ is a family of convex functions, then the function f defined by

$$f(x) = \sup_{\alpha \in \mathcal{A}} f_\alpha(x)$$

is convex.



Local vs Global minima and convexity

Theorem

Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex, and let $X \subseteq \mathbf{dom}\, f$ be a convex set. Then every local minimum for f in X is a global minimum for f in X.

First order characterisation of convexity

Theorem

Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable. Then f is convex if and only if, for all $x, y \in \mathbf{dom}\, f$,

$$f(x) \geq f(y) + \nabla f(y)^\top (x - y).$$

First order characterisation of convexity

Theorem

Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable. Then f is convex if and only if, for all $x, y \in \mathbf{dom}\, f$,

$$f(x) \geq f(y) + \nabla f(y)^\top (x - y).$$

Recall: the hyperplane tangent to the graph of f at point $(y, f(y))$ is

$$H = \{(x, t) \in \mathbb{R}^n \times \mathbb{R} \,:\, t = f(y) + \nabla f(y)^\top (x - y)\},$$

therefore the theorem states that a function is convex if and only if, for every $y \in \mathbf{dom}\, f$, the graph of f lies above the hyperplane tangent to the graph of f at point $(y, f(y))$.

First order optimality conditions and convexity

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable convex function.

Theorem (Unconstrained convex minimisation)

A point $x^* \in \mathbf{dom}\, f$ is a global minimum of f if and only if $\nabla f(x^*) = 0$.

First order optimality conditions and convexity

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable convex function.

Theorem (Unconstrained convex minimisation)

A point $x^* \in \mathbf{dom}\, f$ is a global minimum of f if and only if $\nabla f(x^*) = 0$.

Theorem (Constrained convex minimisation)

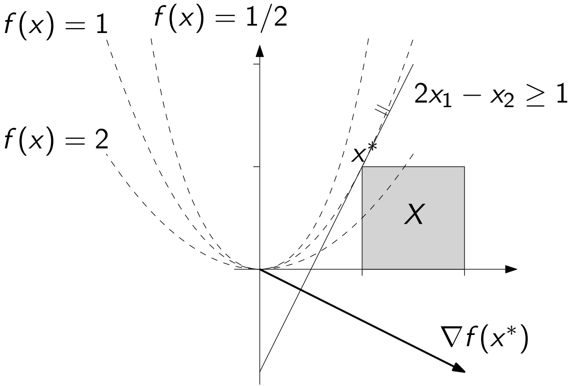
Let $X \subseteq \mathbf{dom}\, f$ be a convex set. A point $x^* \in X$ is a global minimum of f over X if and only if

$$\nabla f(x^*)^\top (x - x^*) \geq 0 \text{ for all } x \in X.$$

Example

$$f(x) = \frac{x_1^2}{x_2}, \ \mathbf{dom}\, f = \{x \,:\, x_2 > 0\}, \ X = \left\{ x \,:\, \begin{array}{l} 1 \leq x_1 \leq 2 \\ 0 < x_2 \leq 1 \end{array} \right\}$$

Show that $x^* = (1,1)^\top$ minimizes f in X.



Second order characterisation of convexity

Taylor expansion of univariate functions

Theorem

Assume that $f : \mathbb{R} \to \mathbb{R}$ *is twice differentiable on* $\mathbf{dom}\, f$*. Then for every* $x, y \in \mathbf{dom}\, f$*, we can write*

$$f(x) = f(y) + f'(y)(x - y) + \frac{1}{2} f''(\bar{x})(x - y)^2$$

for some $\bar{x} \in [x, y]$*.*

What is the notion of second derivative for multivariate functions?

Hessian

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

- *Symmetric:* $\frac{\partial^2 f(x)}{\partial x_i \partial x_j} = \frac{\partial^2 f(x)}{\partial x_j \partial x_i}$.
- *Twice differentiable functions*: **dom** f is an open set and the Hessian of f exists at every point in **dom** f.

Taylor expansion of multivariate functions

Theorem

Assume that $f : \mathbb{R}^n \to \mathbb{R}$ *is twice differentiable on* $\textbf{dom}\, f$. *Then for every* $x, y \in \textbf{dom}\, f$, *we can write*

$$f(x) = f(y) + \nabla f(y)^T (x - y) + \frac{1}{2}(x - y)^\top \nabla^2 f(\bar{x})(x - y)$$

for some $\bar{x} \in [x, y] = \{\lambda x + (1 - \lambda) y \,:\, 0 \leq \lambda \leq 1\}$

(Semi)definite matrices

$A \in \mathbb{R}^{n \times n}$: *symmetric* matrix: $a_{ij} = a_{ji}$ for every i, j.

- A is *positive definite* if, for all $x \in \mathbb{R}^n \setminus \{0\}$, $x^\top Ax > 0$.
- A is *positive semidefinite* if, for all $x \in \mathbb{R}^n$, $x^\top Ax \geq 0$.
- A is *negative definite* if, for all $x \in \mathbb{R}^n \setminus \{0\}$, $x^\top Ax < 0$.
- A is *negative semidefinite* if, for all $x \in \mathbb{R}^n$, $x^\top Ax \leq 0$.

(Semi)definite matrices

$A \in \mathbb{R}^{n \times n}$: *symmetric* matrix: $a_{ij} = a_{ji}$ for every i, j.

- A is *positive definite* if, for all $x \in \mathbb{R}^n \setminus \{0\}$, $x^\top A x > 0$.
- A is *positive semidefinite* if, for all $x \in \mathbb{R}^n$, $x^\top A x \geq 0$.
- A is *negative definite* if, for all $x \in \mathbb{R}^n \setminus \{0\}$, $x^\top A x < 0$.
- A is *negative semidefinite* if, for all $x \in \mathbb{R}^n$, $x^\top A x \leq 0$.

Theorem

Let $A \in \mathbb{R}^{n \times n}$ be a symmetric matrix. Then, A is positive definite if and only if all its eigenvalues are positive. A is positive semidefinite if and only if all its eigenvalues are nonnegative.

Second order characterisation of local extrema

Theorem

Let $f : \mathbb{R}^n \to \mathbb{R}$ is twice differentiable on $\mathbf{dom}\, f$. If $x^* \in \mathbf{dom}\, f$ is a local minimum, then $\nabla^2 f(x^*)$ is positive semidefinite. If $x^* \in \mathbf{dom}\, f$ is a local maximum, then $\nabla^2 f(x^*)$ is negative semidefinite.

Second order characterisation of convexity

Theorem

Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice differentiable.

(i) *If $\nabla^2 f(z)$ is positive semidefinite for every $z \in \textbf{dom}\, S$, then f is convex.*

(ii) *Assume that f is convex, and $\nabla^2 f$ is continuous on $\textbf{dom}\, f$. Then, $\nabla^2 f(z)$ is positive semidefinite for every $z \in \textbf{dom}\, f$.*

Example: $f(x) = \dfrac{x_1^2}{x_2}$, $\textbf{dom}\, f = \{x : x_2 > 0\}$.

Quadratic functions

- *Quadratic function* $f : \mathbb{R}^n \to \mathbb{R}$: degree two polynomial function. E.g. $f(x_1, x_2) = -x_1^2 + 3x_1x_2 + 2x_2^2 - 5x_1 + 6x_2 + 3$.
- Standard form:

$$f(x) = x^\top Qx + p^\top x + r,$$

Example: $Q = \begin{pmatrix} -1 & 1.5 \\ 1.5 & 3 \end{pmatrix}$, $p = (-5, 6)$, $r = 3$.

Quadratic functions

- *Quadratic function* $f : \mathbb{R}^n \to \mathbb{R}$: degree two polynomial function. E.g. $f(x_1, x_2) = -x_1^2 + 3x_1x_2 + 2x_2^2 - 5x_1 + 6x_2 + 3$.
- Standard form:

$$f(x) = x^\top Qx + p^\top x + r,$$

Example: $Q = \begin{pmatrix} -1 & 1.5 \\ 1.5 & 3 \end{pmatrix}$, $p = (-5, 6)$, $r = 3$.

Theorem
Let $Q \in \mathbb{R}^{n\times n}$ be a symmetric matrix, $p \in \mathbb{R}^n$, and $r \in \mathbb{R}$. The quadratic function $f(x) = x^\top Qx + p^\top x + r$ is convex if and only if Q is positive semidefinite.

Concave functions

- A function f is *concave* if $-f$ is convex.
- The previous theorems concerning minima of convex functions are true if we replace "convex" with "concave" and "minimum" with "maximum".
  - Local maximum = Global Maximum.
  - Local maximum iff gradient is zero.

Convex Optimisation Problems

$$\begin{array}{ll} \min & f_0(x) \\ & f_i(x) \leq 0 \quad i = 1, \ldots, m \\ & h_i(x) = 0 \quad i = 1, \ldots, k \end{array}$$

- Domain of problem: $\mathcal{D} = \left(\bigcap_{i=0}^{m} \textbf{dom}\, f_i \right) \bigcap \left(\bigcap_{i=1}^{k} \textbf{dom}\, h_i \right)$.
- Feasible region: set X of all points in $\mathcal{D}$ satisfying the constraints.

Convex Optimisation Problems

$$\begin{aligned} \min \quad & f_0(x) \\ & f_i(x) \leq 0 \quad i = 1, \ldots, m \\ & h_i(x) = 0 \quad i = 1, \ldots, k \end{aligned}$$

- Domain of problem: $\mathcal{D} = \left(\bigcap_{i=0}^{m} \textbf{dom}\, f_i \right) \bigcap \left(\bigcap_{i=1}^{k} \textbf{dom}\, h_i \right)$.
- Feasible region: set X of all points in $\mathcal{D}$ satisfying the constraints.

Convex problem if:

- $f_0, \ldots, f_m$ are convex,
- $h_1, \ldots, h_k$ are affine (i.e. there exist $a_i \in \mathbb{R}^n$, $b_i \in \mathbb{R}$ such that $h_i(x) = a_i^\top x - b_i$)

Important facts

1. If f_i convex, then $\{x \in \mathbb{R}^n : f_i(x) \leq 0\}$ is a convex set.
2. The feasible region X is convex, because it is the intersection of convex sets.
3. Every local optimum for f_0 in X is also a global optimum.

Example: Portfolio optimisation

The Markowitz portfolio optimisation problem is convex

$$\begin{gathered}\min\ x^\top \Sigma x \\ \text{s.t.} \\ \bar{p}^\top x \geq r_{\min} \\ \textstyle\sum_{i=1}^n x_i = B \\ x \geq 0\end{gathered}$$

The objective function $x^\top \Sigma x$ is a quadratic convex function (because Σ is positive semidefinite).
Indeed $x^\top \Sigma x = \mathrm{Var}(p) = \mathbb{E}[(p^\top x - \bar{p}^\top x)^2]$.

Concave maximisation

- A maximisation problem of the form

$$\begin{aligned} \max \quad & f_0(x) \\ & f_i(x) \leq 0 \quad i = 1, \ldots, m \\ & h_i(x) = 0 \quad i = 1, \ldots, k \end{aligned}$$

  is a convex optimisation problem if f_0 is concave, while $f_1, \ldots, f_m$ are convex and $h_1, \ldots, h_k$ affine,
- Indeed, the equivalent problem obtained replacing "$\max f_0(x)$" with "$\min -f_0(x)$" is a convex optimisation problem.