Numerical Optimisation: Background

f.rullan@cs.uc

Department of Computer Sc
Centre for Medical Image Com
Centre for Inverse Problems
University College London

Lecture 1

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{subject to} \quad \begin{array}{ll} c_i(x) = 0, & i \in \mathcal{E} \\ c_i(x) \geq 0, & i \in \mathcal{I} \end{array}$$

-
-

  $i \in \mathcal{I}$ inequality constraints.
- $x \in \mathbb{R}^n$: optimisation variable

Optimal solution $x^\star$ has the smallest value of f among all x which satisfy the constraints.

Geodesics are the shortest surface paths between two points.

Figure: `https://academo.org/demos/geodesics/`

A very short and incomplete early history

Source `http://www.mitrikitti.fi/opthist.html`

- **Antiquity: geometrical optimisation problems**
  300 BC Euclid considers the minimal distance between a point a line, and proves that a square has the greatest area among the rectangles with given total length of edges

-
  smallest variation of volume w.r.t. barrel p [1]
  Early version of the secretary problem (opt problem) when he started to look for a new wi
  1636 P. Fermat shows that at the extreme po of a function vanishes. In 1657 Fermat shows that light travels between two points in minimal time.

[1] http://www.maa.org/press/periodicals/convergence/kepler-the-volume-of-a-wine-barrel-solving-the-problem-of-maxima-wine-barrel-design

Source http://www.mitrikitti.fi/opthist.html

- **Calculus of Variations**

1696 Johann and Jacob Bernoulli study Br problem, calculus of variations is born

1740 L. Euler's publication begins the resea theory of calculus of variations

Source http://www.mitrikitti.fi/opthist.html

- **Least squares**
  1806 A. M. Legendre presents the least square method, which also J.C.F. Gauss claims to have invented. Legendre made contributions in the field of CoV, too

-

  1939 L.V. Kantorovich presents LP-mod
  for solving it. In 1975 Kantorovich and T.C. K
  the Nobel memorial price in economics for t
  on LP-problem
  1947 G. Dantzig, who works for US air-force, presents the Simplex method for solving LP-problems, von Neumann establishes the theory of duality for LP-problems

Example: transportation problem

- 2 factories, F_i
- 12 retail outlets, R_j
- each factory F_i can produce up to a_i tones of a certain
- the
- the cost of shipping of one tone of the compoun to R_j is c_{ij}

Goal: what is the optimal amount to ship from each factory to each outlet which satisfies demand at minimal cost.

$$\min \sum_{ij} c_{ij} x_{ij}$$

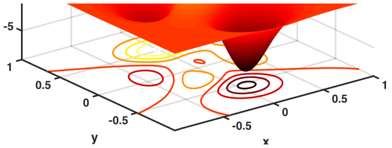
$$\sum_{i=1} x_{ij} \geq b_j,\ j = 1 \ldots 12$$

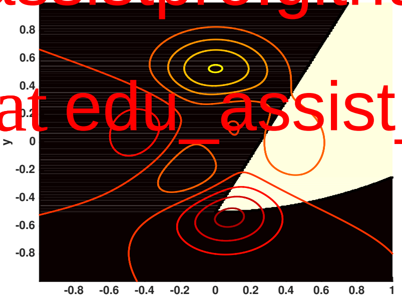
$$x_{ij} \geq 0,\ i$$

Linear programming problem because the objective function and all constraints are linear.

$$\min f(x, y)$$

$$\text{subject to} \quad -y + 2x - \frac{1}{2} \geq 0$$





A set $\mathbb{S} \subset \mathbb{R}^n$ is **convex** if for any two points $x, y \in \mathbb{S}$ the line segment connecting them lies entirely in $\mathbb{S}$

$$\alpha x + (1 - \alpha) y \in \mathbb{S}, \quad \forall \alpha \in [0, 1].$$

Examples:

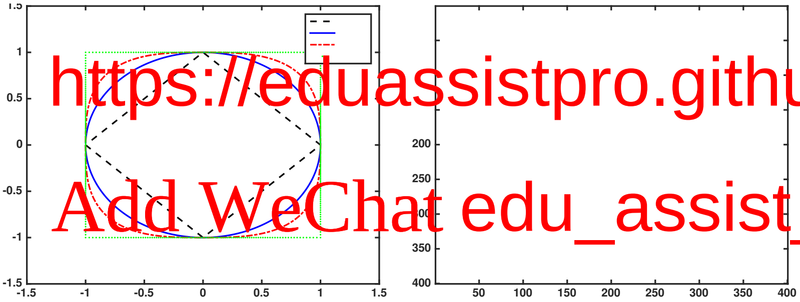


Figure: (a) unit ball $\{x \in \mathbb{R}^n : \|x\|_p \leq 1\}, \; p \geq 1$; (b) polyheadron $\{x \in \mathbb{R}^n : Ax = b, Cx \leq d\}$

A function f is **convex** if

- its domain $\mathbb{S}$ is a convex set,
- for any two points $x, y \in \mathbb{S}$,

$$f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y), \quad \forall \alpha \in [0, 1].$$

A func

1).

A function f is **concave** if $-f$ is convex.

Examples:

- linear function $f(x) = c^{\mathrm{T}} x + \alpha$, where $c \in \mathbb{R}^n$, $\alpha \in \mathbb{R}$
- convex quadratic function $f(x) = x^{\mathrm{T}} H x$, where $H \in \mathbb{R}^{n \times n}$ symmetric positive (semi)definite

- **convex** vs non-convex
- **smooth vs non-smooth**
-
-
-

- **local** vs global
- stochastic vs deterministic
- discrete vs **continuous**

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a smooth function for which we can evaluate f and its derivatives at any given point $x \in \Omega \subseteq \mathbb{R}^n$.

Unconstraint optimisation problem

$$\min \quad f(x). \tag{1}$$

A poin

$$f(x^\star) \leq f(x), \ \forall x$$

A point $x^\star$ is a **local minimiser** if

$$\exists \mathcal{N}(x^\star) : f(x^\star) \leq f(x), \ \forall x \in \mathcal{N}(x^\star),$$

$\mathcal{N}(y)$ is a neighbourhood of y (an open set which contains y).

A point $x^\star$ is a **strict (or strong) local minimiser** if

$$(x^\star) : f(x^\star) < f(x), \quad x \qquad (x^\star),\ x \neq x^\star.$$

Exa

- $f(x) = (x-2)^4 :\ x^\star = 2$ is a strict local global one)

- $f(x) = \cos(x) :\ x^\star = \pi + 2k\pi, k$ minimisers (but not strict global on

A point $x^\star$ is an **isolated local minimiser** if

$$\exists \mathcal{N}(x^\star) : x^\star \text{ is the only local minimiser in } \mathcal{N}(x^\star).$$

Som

has a strict local minimiser at $x^\star = 0$ but t
minimisers (albeit $f(x_j) \geq x_j^4 > 0 = f($
$x_j \to 0,\ j \to \infty$.

However, all isolated local minimiser are strict.

Unconstraint minimisation

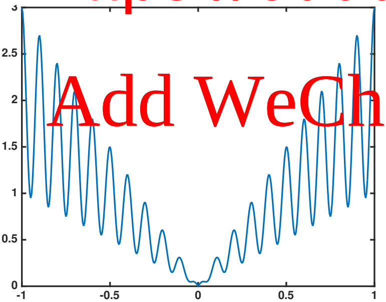
Difficulties with global minimisation:

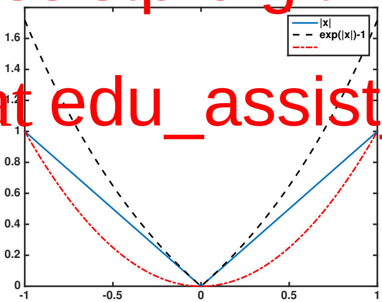
$$f(x) = (\cos(20\pi x) + 2)|x|$$

has a unique global minimiser $x^\star = 0$, but the algorithms usually get trapped into one of the many local minima.

For co

mini





Taylor theorem

Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable. Then for $p \in \mathbb{R}^n$ we have

$$f(x + p) = f(x) + \nabla f(x + tp)^{\mathrm{T}} p$$

for some $t \quad (0, 1)$.

If mor

0

and

$$f(x + p) = f(x) + \nabla f(x)^{\mathrm{T}} p + \frac{}{2} p \;\; \nabla \;\; f(x + tp) p,$$

fore some $t \in (0, 1)$.

Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable in an open neighbourhood of a **local minimiser** $x^\star$, then $\nabla f(x^\star) = 0$.

Proof: [by contradiction]
Suppose that $\nabla f(x^\star) \neq 0$ and define $p = -\nabla f(x^\star)$. Note that $p^{\mathrm{T}}\nabla$ $\star$ $\star$ 2 nuous near

By Taylor's theorem, for any $\bar{t} \in (0, T$

$$f(x^\star + \bar{t}p) = f(x^\star) + \bar{t}p^{\mathrm{T}}\nabla f($$

Hence $f(x^\star + \bar{t}p) < f(x^\star)$ for all $\bar{t} \in (0, T]$, and we have found a direction leading away from $x^\star$ along which f decreases which is in contradiction with $x^\star$ being a local minimiser.

We ca

By Th
a stati

If $x^\star$ is a **local minimiser** of f and $\nabla^2 f$ exists and is continuous in an open neighbourhood of $x^\star$, then $\nabla f(x^\star) = 0$ and $\nabla^2 f(x^\star)$ is positive semidefinite.

Proof: [by contradiction]

By Th ... $^\star) = 0$.

Assu

vect

conti

$p^{\mathrm{T}} \nabla^2 f(x^\star + tp) p < 0$ for all $t \in [0, T$

By Taylor theorem, we have for any $\bar{t}$ ... $, \bar{T})$

$$f(x^\star + \bar{t}p) = f(x^\star) + \bar{t}p^{\mathrm{T}} \nabla f(x^\star) + \frac{1}{2}\bar{t}^2 p^{\mathrm{T}} f(x^\star + tp) p < f(x^\star).$$

We have found a decrease direction for f away from $x^\star$ which contradicts $x^\star$ being a local minimiser.

Let $f : \mathbb{R}^n \to \mathbb{R}$ with $\nabla^2 f$ continuous in an open neighbourhood of $x^\star$. If $\nabla f(x^\star) = 0$ and $\nabla^2 f(x^\star)$ is positive definite, then $x^\star$ is a **strict local minimiser** of f.

Proof:

Because the Hessian $\quad {}^2 f$ is continuous and positive definite at $x^\star$, we ca

defin

$2 < r\}$.

For a $\quad / \quad$) and

$$f(x^\star + p) = f(x^\star) + p^{\mathrm{T}} \nabla f(x^\star \quad - \tag{2}$$

$$= f(x^\star) + \frac{1}{2} p^{\mathrm{T}} \nabla^2 f \tag{3}$$

for some $t \in (0, 1)$.

Furthermore, $x^\star + tp \in B_2(x^\star, r)$ thus $p^{\mathrm{T}} \nabla^2 f(x^\star + tp)p > 0$ and therefore $f(x^\star + p) > f(x^\star)$.

2nd order sufficient condition guarantees a stronger statement than the necessary conditions (strict local minimiser).

A stric

$$f(x) = x \,, \quad f\;(x) = 4x \,, \quad f\;\;(x) = 12x$$

$x^\star = 0$ is a strict local minimiser while
the necessary but not the sufficient conditions.

If $f : \mathbb{R}^n \to \mathbb{R}$ is convex, any local minimiser $x^\star$ is also a global minimiser of f. If, f is also differentiable, then any stationary point $x^\star$ is a global minimiser.

Proof:

Supp $\star$

$\exists z \in$ $x^\star$ and

z i.e.

$$\mathcal{L}(x^\star, z) = \{x : x = \lambda z + (1 -$$

by convexity of f we have

$$f(x) \leq \lambda f(z) + (1 - \lambda) f(x^\star) < f(x^\star).$$

For any neighbourhood $\mathcal{N}(x^\star) \cap \mathcal{L}(x^\star, z) \neq \emptyset$, hence $\exists x \in \mathcal{N}(x^\star) : f(x) < f(x^\star)$ and $x^\star$ is not a local mininiser.

Proof: cont.

For the second part, we suppose that $x^\star$ is not global minimiser.

For all z chosen as before by convexity of f it follows

$$\lim_{\lambda \to 0}$$

$$\leq \lim_{\lambda \to 0} \frac{\lambda f(z)}{}$$

$$= f(z) - f(x$$

Hence $\nabla f(x^\star) \neq 0$ and $x^\star$ is not a stationary point.