C1: The Unconstrained Optimization Problem

Niclas Börlin

5DA001 Non-linear Optimization

What's the Problem?

The continuous optimization problem

- A general continuous optimization problem may be written as

$$\min_{x \in \Omega} f(x).$$

- The function $f(x) : \Re^n \to \Re$ is called the objective function and is assumed to be twice continuously differentiable.
- The vector $x \in \Re^n$ contains the variables to be estimated.
- The vector $x \in \Re^n$ is a point in parameter space.
- The feasible set $\Omega \subseteq \Re^n$ is the set of feasible points, i.e. the set of all points that satisfy all constraints.
- For unconstrained problems $\Omega \equiv \Re^n$.
- We do not need to treat maximization problems separately, since

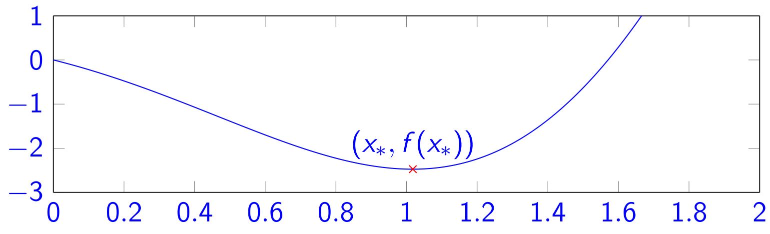
$$\max_x f(x) \equiv -\min_x -f(x).$$

1D example

- Consider the 1D problem with

$$f(x) = -e^x \sin 2x$$

with minimum near $x_* \approx 1.02$.



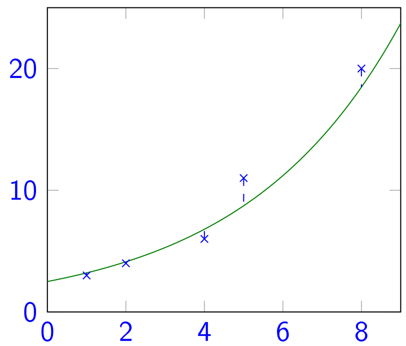
2D example

- Consider the 2D curve fitting problem with

$$f(x) = \sum_{i=1}^{5} (x_1 e^{x_2 t_i} - y_i)^2,$$

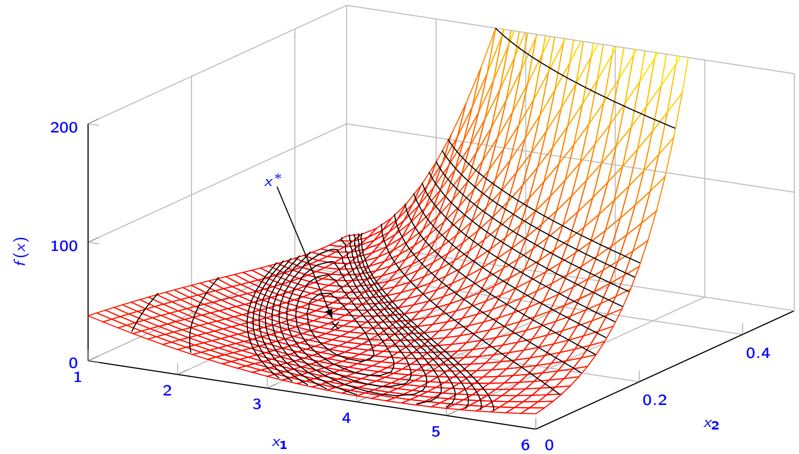
where the given data set is

t_i	:	1	2	4	5	8
y_i	:	3	4	6	11	20



2D example

- The problem has solution $x_* \approx (2.53, 0.21)$.



Optimizers (What are we looking for?)

Global Optimizers

- Consider the n-dimensional problem

$$\min_{x \in \Omega} f(x).$$

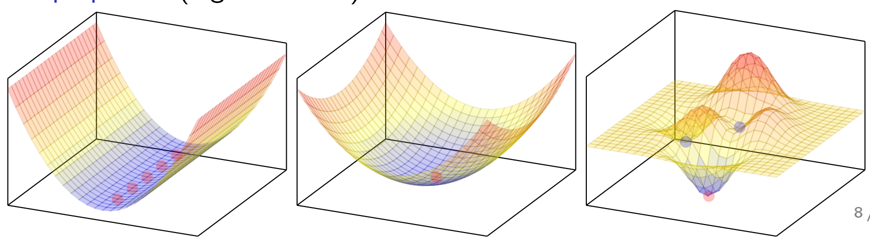
- A point x^* that satisfies

$$f(x^*) \leq f(x) \; \forall x \in \Omega$$

is called a global minimizer to f.
- If

$$f(x^*) < f(x) \; \forall x \neq x^*,$$

the point x^* is called a strict global minimizer and is unique.
- Global minimizer are hard to determine unless f have special properties (e.g. is convex).



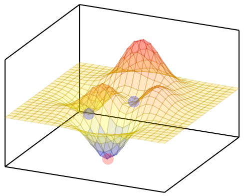
Local optimizers

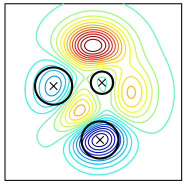
- Often we will have to settle for local minimizers, i.e. x^* such that

$$f(x^*) \leq f(x) \; \forall x \in \mathcal{N},$$

  where $\mathcal{N}$ is a neighbourhood of x^*, i.e. an open set that contains x^*.
- Similarly, a strict local minimizer x^* is defined by

$$f(x^*) < f(x) \; \forall x \in \mathcal{N}, x \neq x^*.$$





How do we determine if a given point is a minimizer?

Wait, first let's talk about Taylor approximations

Taylor approximation

- The Taylor series is a tool for approximating a function f near a specific point x_0.
- Definition: Let x_0 be a specified point and $f : \Re \to \Re$ have n continous derivatives. The n-th order Taylor series approximation is

$$f(x_0 + p) \approx f(x_0) + pf'(x_0) + \frac{p^2}{2} f''(x_0) + \cdots .$$

- For $f : \Re^n \to \Re$, the Taylor approximation becomes

$$f(x_0 + p) = f(x_0) + p^T \nabla f(x_0) + \frac{1}{2} p^T \nabla^2 f(x_0) p + \cdots .$$

- The notation $\nabla f(x_0)$ refers to the gradient of the function f at the point $x = x_0$, i.e. a vector with all first order derivatives of f.
- The notation $\nabla^2 f(x_0)$ refers to the Hessian of f at x_0, i.e. a square matrix with all second order derivatives of f.

Taylor approximation
1D example

- Consider the function

$$\begin{aligned} f(x) &= -e^x \sin 2x, \\ f'(x) &= -e^x \sin 2x - 2e^x \cos 2x = -e^x(\sin 2x + 2\cos 2x), \\ f''(x) &= -e^x(\sin 2x + 2\cos 2x) - e^x(2\cos 2x - 4\sin 2x) \\ &= -e^x(-3\sin 2x + 4\cos 2x). \end{aligned}$$

- For $x_0 = 1.4$,

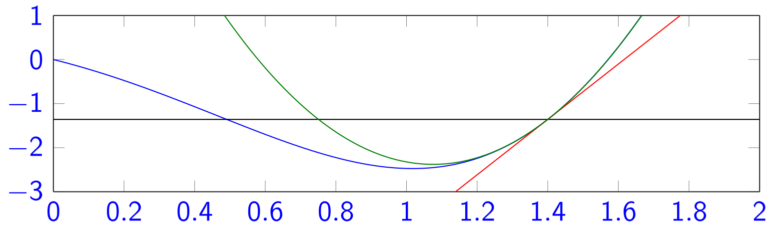
$$\begin{aligned} f(x_0) &= -1.3584, \\ f'(x_0) &= 6.2834, \\ f''(x_0) &= 19.3589. \end{aligned}$$

Taylor approximation
1D example

- For $p = -0.2$,

$$\begin{aligned} f(1.2) &= f(x_0 + p) \\ &\approx f(x_0) + pf'(x_0) + \frac{1}{2}p^2 f''(x_0) \\ &= -1.3584 + (-0.2)\cdot 6.2834 + \frac{1}{2}(-0.2)^2 \cdot 19.3589 \\ &= -1.3584 + -1.2567 + 0.3872 = -2.2279. \end{aligned}$$

- Compare to the exact value $f(1.2) = -2.2426$.



Taylor approximation
2D example

- For the function

$$f(x_1, x_2) = x_1^3 + 5x_1^2 x_2 + 7x_1 x_2^2 + 2x_2^3,$$

the gradient and hessian are

$$\nabla f(x) = \begin{bmatrix} 3x_1^2 + 10x_1x_2 + 7x_2^2 \\ 5x_1^2 + 14x_1x_2 + 6x_2^2 \end{bmatrix},$$

$$\nabla^2 f(x) = \begin{bmatrix} 6x_1 + 10x_2 & 10x_1 + 14x_2 \\ 10x_1 + 14x_2 & 14x_1 + 12x_2 \end{bmatrix}.$$

- Evaluated at the point $x_0 = [-2, 3]^T$ they are

$$f(x_0) = -20, \quad \nabla f(x_0) = \begin{bmatrix} 15 \\ -10 \end{bmatrix}, \quad \nabla^2 f(x_0) = \begin{bmatrix} 18 & 22 \\ 22 & 8 \end{bmatrix}.$$

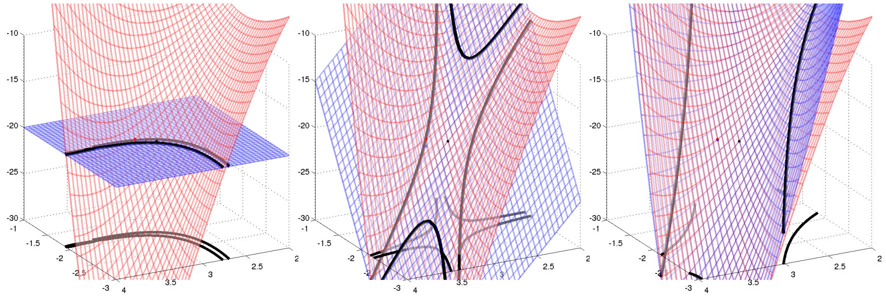
Taylor approximation
2D example

- For $p = [0.1, 0.2]^T$,

$$\begin{aligned} f(-1.9, 3.2) &= f(x_0 + p) \\ &\approx f(x_0) + p^T \nabla f(x_0) + \frac{1}{2}p^T \nabla^2 f(x_0) p \\ &= -20 + \begin{bmatrix} 0.1 \\ 0.2 \end{bmatrix}^T \begin{bmatrix} 15 \\ -10 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} 0.1 \\ 0.2 \end{bmatrix}^T \begin{bmatrix} 18 & 22 \\ 22 & 8 \end{bmatrix} \begin{bmatrix} 0.1 \\ 0.2 \end{bmatrix} \\ &= -20 - 0.5 + 0.69 = -19.81 \end{aligned}$$

- Compare the exact value $f(-1.9, 3.2) = -19.755$.

Approximations



0th order 1st order 2nd order

Ok, that's fine, but how do we determine if a given point is a minimizer?

Almost there, but first we need to understand Definiteness
(almost there, more patience...)

Definiteness

- A square matrix A is positive semi-definite if

$$x^T A x \geq 0, \forall x.$$

- A square matrix A is positive definite if

$$x^T A x > 0, \forall x \neq 0.$$

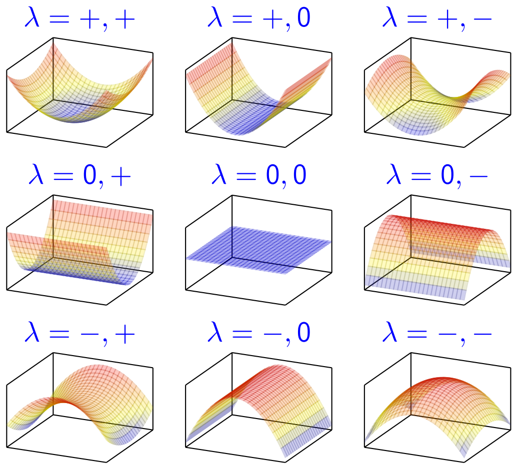
- A square matrix A which is neither positive semi-definite nor negative semi-definite is indefinite.
- A positive definite matrix has only positive eigenvalues since for all eigenpairs (x, λ) of A

$$x^T A x > 0 \Rightarrow x^T \underbrace{Ax}_{\lambda x} = x^T \lambda x = \lambda x^T x > 0 \Rightarrow \lambda > 0.$$

- Positive semi-definite matrices have $\lambda \geq 0$.

Definiteness and curvature

- The definiteness and sign of the eigenvalues correspond to the curvatures of the quadratic expression $x^T Ax$.



Fine, whatever turns you on. Can we now please determine if a given point is a minimizer?

Yes.

It's called the First and Second order conditions on a minimizer. (Wake up, this is it!)

The First-Order *Necessary* Conditions

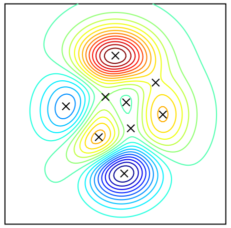
- Assume x^* is a local minimizer to f. Study f around x^*:

$$f(x^* + p) = f(x^*) + \nabla f(x^*)^T p + \frac{1}{2} p^T \nabla^2 f(\xi) p$$

- If x^* is a local minimizer this implies that

$$\nabla f(x^*) = 0.$$

- This condition is called the $\boxed{\text{first-order necessary condition}}$ for a minimizer.
- A point that satisfies $\nabla f(x^*) = 0$ is called a stationary point to f.

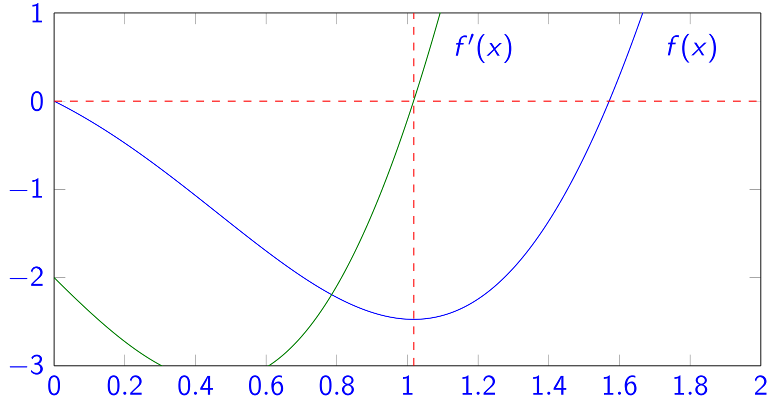


The First-Order Necessary Conditions

1D example

- In 1-D, the first-order necessary condition simplifies to

$$f'(x) = 0.$$



The Second-Order *Necessary* Conditions

- Study f around a stationary point x^*. Assume x^* is a local minimizer:

$$f(x) = f(x^* + p) = f(x^*) + \underbrace{\nabla f(x^*)^T p}_{=0} + \frac{1}{2} p^T \nabla^2 f(\xi) p.$$

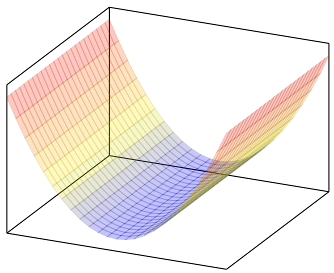
- For x close to x^*, $\nabla^2 f(\xi)$ will be close to $\nabla^2 f(x^*)$.
- If $\nabla^2 f(x^*)$ is *not* positive semi-definite, there exists a v such that $v^T \nabla^2 f(x^*) v < 0$, and there is a p close to v such that $p^T \nabla^2 f(\xi) p < 0$ meaning that

$$f(x) = f(x^* + p) = f(x^*) + \underbrace{\frac{1}{2} p^T \nabla^2 f(\xi) p}_{<0} < f(x^*),$$

which is a contradiction, since x^* was assumed to be a minimizer.

The Second-Order Necessary Conditions

- Thus, $\nabla^2 f(x^*)$ must be positive semi-definite in order for the stationary point x^* to be a minimizer.
- This condition is called the $\boxed{\text{second-order necessary condition}}$.



The Second-Order *Sufficient* Conditions

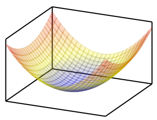
- Study f around x^* when $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*)$ positive definite.
- Then

$$f(x) = f(x^* + p) = f(x^*) + \underbrace{\nabla f(x^*)^T p}_{=0} + \frac{1}{2} p^T \nabla^2 f(\xi) p.$$

- For x close to x^*, $\nabla^2 f(\xi)$ will also be positive definite and

$$f(x) = f(x^* + p) = f(x^*) + \underbrace{\frac{1}{2} p^T \nabla^2 f(\xi) p}_{>0\ \forall p \neq 0} > f(x^*) \quad \forall p$$

- Thus, x^* is a strict minimizer of f.
- This is called the $\boxed{\text{second-order sufficient condition}}$ on a minimizer.



Second-Order Conditions

1D example

- In 1-D, the necessary second-order condition simplifies to

$$f''(x_*) \geq 0,$$

whereas the sufficient condition simplifies to

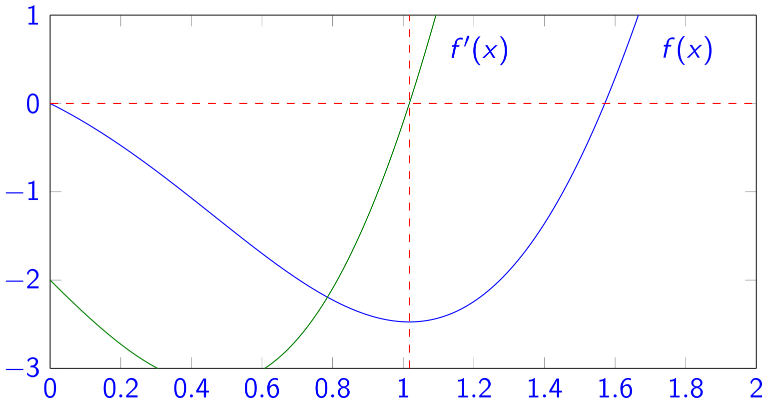
$$f''(x_*) > 0.$$

Second-Order Conditions

1D example

- In this case

$$f''(1.02) \approx 12.$$



Discussion time

- Discuss the first- and second-order conditions (necessary and sufficient).
- What conditions can be used to determine if
  - a point x_* is a minimizer (inclusion criteria),
  - a point x_* is not a minimizer (exclusion criteria)?
- Are there cases when the first and second order conditions are not sufficient to determine if a point x_* is a minimizer or not?
- Suggest a 1D function where the first and second order conditions cannot be used to determine if $x = 0$ is a minimizer (and $x = 0$ is indeed a minimizer).
- Suggest a 1D function where the first and second order conditions cannot be used to determine if $x = 0$ is a minimizer (and $x = 0$ is not a minimizer).