Optimization and Data Science

Lecture 5: Optimality Conditions

Prof. Dr. Thomas Slawig

Kiel University - CAU Kiel
Dep. of Computer Science

Summer 2020

Contents

1. Optimality Conditions
   - Existence Results
   - Optimality conditions for differentiable functions in one dimension
   - Differentiability in $\mathbb{R}^n$
   - First Order Necessary Condition for Unconstrained Problems
   - First Order Necessary Condition for Constrained Problems

Continuous optimization

- Unconstrained continuous optimization problems:

$$X_{ad} = X = \mathbb{R}^n$$

- Constrained continuous optimization problems:

$$X_{ad} \subset X = \mathbb{R}^n, X_{ad} \neq X.$$

- Integer problems can be **relaxed**:
    - Replace $X = \mathbb{Z}$ by $\hat{X} = \mathbb{R}$
    - Compute solution in $\mathbb{R}$
    - Project onto $\mathbb{Z}$ by rounding to (nearest) integer value.
- Similar for $\mathbb{N} \to \mathbb{R}_{\geq 0}, \{0, 1\} \to [0, 1]$.
- Component-wise for vectors.
- Attention: **Continuous optimization problem** means $x \in \mathbb{R}^n$
- f **continuous** means

$$\lim_{x \to \bar{x}} f(x) = f(\bar{x})$$

Contents

1 Optimality Conditions
- Existence Results
- Optimality conditions for differentiable functions in one dimension
- Differentiability in $\mathbb{R}^n$
- First Order Necessary Condition for Unconstrained Problems
- First Order Necessary Condition for Constrained Problems

Existence result for optimum

Result from mathematical analysis:

Theorem (Weierstrass)

If $f : \mathbb{R}^n \supset X_{ad} \to \mathbb{R}$ is continuous on the closed and bounded set X_{ad}, then there exists at least one global minimizer $x^* \in X_{ad}$ (and also a global maximizer).

- X_{ad} closed: boundary points belong to X_{ad}:

$$(x_k)_{k \in \mathbb{N}} \subset X_{ad}, x_k \to x \Longrightarrow x \in X_{ad}$$

  Examples: $[0, 1]$ closed, $(0, 1), [0, 1)$ not closed, $\mathbb{R}$ closed

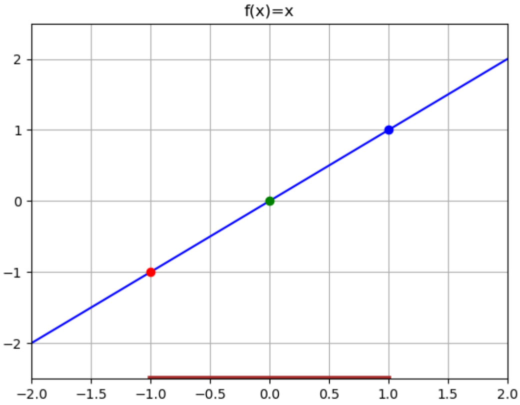
- X_{ad} bounded subset of $\mathbb{R}^n$:

$$\exists M \in \mathbb{R} \ \ \forall x \in X_{ad} : \ \|x\| \leq M.$$

Example for Weierstrass Theorem

The function $f : \mathbb{R} \to \mathbb{R}, f(x) = x$ is continuous. It has ...

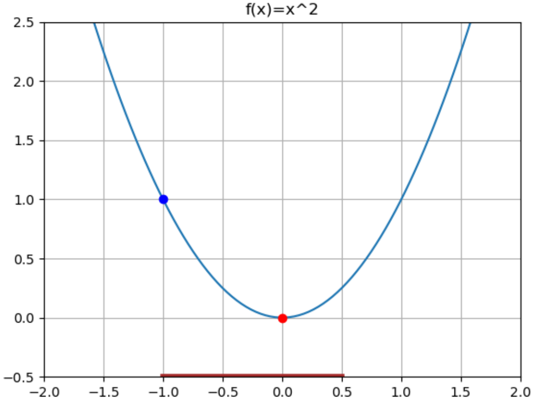
- ... no global minimizer or maximizer on $X_{ad} = X = \mathbb{R}$ (closed, but not bounded)
- ... a global minimizer, but no maximizer on $X_{ad} = \mathbb{R}_{\geq 0}$ (closed, but not bounded)
- ... a global maximizer, but no minimizer on $X_{ad} = \mathbb{R}_{\leq 0}$ (closed, but not bounded)
- ... a global minimizer and maximizer on $X_{ad} = [a, b]$, $a \leq b$ (closed and bounded)
- ... no minimizer or maximizer on $X_{ad} = (a, b)$ (bounded, but not closed)



Example for Weierstrass Theorem

The function $f : \mathbb{R} \to \mathbb{R}, f(x) = x^2$ is continuous.

- ... a global minimizer, but no maximizer on $X_{ad} = X = \mathbb{R}$ (closed, but not bounded)
- ... a global minimizer, but no maximizer on $X_{ad} = \mathbb{R}_{\geq 0}$ (closed, but not bounded)
- ... a global minimizer, but no maximizer on $X_{ad} = \mathbb{R}_{\leq 0}$ (closed, but not bounded)
- ... a global minimizer and maximizer on $X_{ad} = [a, b]$, $a \leq b$ (closed and bounded)
- ... a global minimizer, but no maximizer on $X_{ad} = (a, b)$ (bounded, but not closed)

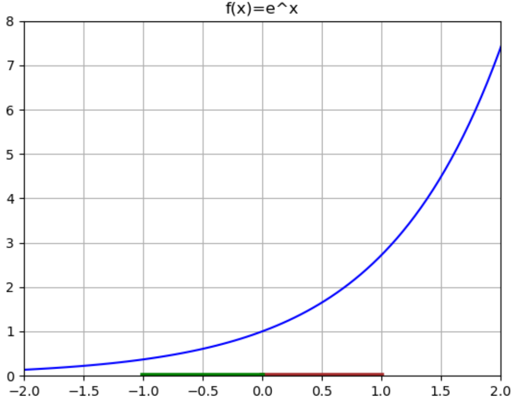


$\rightsquigarrow$ The Weierstrass Theorem provides a **sufficient** condition (X_{ad} closed and bounded).
It is **not a necessary** condition for the existence of a minimizer/maximizer.

Example for Weierstrass Theorem

Consider the function $f : \mathbb{R} \to \mathbb{R}, f(x) = e^x$. Is it continuous? Using the Weierstrass Theorem, what can be said about existence of global minimizers and maximizers ...

- ... on $X_{ad} = (-\infty, 0]$?
- ... on $X_{ad} = (-\infty, 0)$?
- ... on $X_{ad} = (0, \infty)$?
- ... on $X_{ad} = [0, \infty)$?
- ... on $X_{ad} = (0, 1]$ and $X_{ad} = (0, 1)$?
- ... on $X_{ad} = [0, 1]$ and $X_{ad} = [0, 1)$?
- ... on $X_{ad} = (-1, 0]$ and $X_{ad} = (-1, 0)$?
- ... on $X_{ad} = [-1, 0]$ and $X_{ad} = [-1, 0)$?



Radially unbounded functions

The sets $\mathbb{R}$ and $\mathbb{R}^n$ are not bounded $\rightsquigarrow$ theorem not useful for unconstrained problems, but ...

Definition

A function $f : \mathbb{R}^n \to \mathbb{R}$ is called **radially unbounded** if

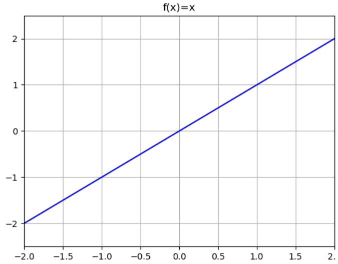
$$\lim_{\|x\| \to \infty} f(x) = +\infty,$$

i.e., for any unbounded sequence, the function value becomes unbounded from above, too.

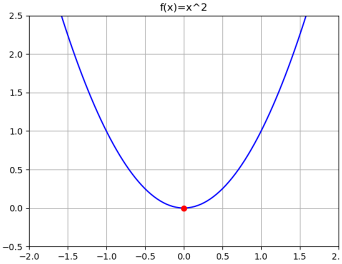
Corollary

If $f : \mathbb{R}^n \to \mathbb{R}$ is continuous and radially unbounded, then it has at least one global minimizer.

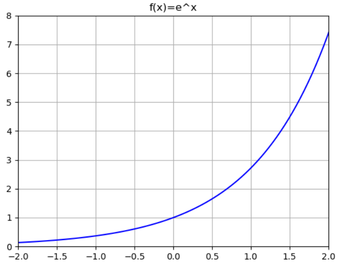
Radially unbounded functions: Examples



not radially unbounded



radially unbounded



not radially unbounded

Minimum of radially unbounded functions

Corollary

If $f : \mathbb{R}^n \to \mathbb{R}$ is continuous and radially unbounded, then it has at least one global minimizer.

Proof.

- Fix any $\hat{x} \in \mathbb{R}^n$. Define $S := \{x \in \mathbb{R}^n : f(x) \leq f(\hat{x})\}$
- S is closed:

  Continuity of f : For any sequence $(x_k)_k \subset S, x_k \to x$, we have $\lim_{k\to\infty} f(x_k) = f(x)$.

  $\Rightarrow f(x) = \lim_{k\to\infty} f(x_k) \leq f(\hat{x}) \Rightarrow x \in S$.
- Moreover, S is bounded:

  Assume S is unbounded: Then there exists a sequence $(x_k)_k$ in S with $\|x_k\| \to \infty$.

  f is radially unbounded: $(f(x_k))_k$ is unbounded, i.e., $f(x_k) > f(\hat{x})$ for k big enough.

  Thus, $x_k \notin S$. Contradiction.

Contents

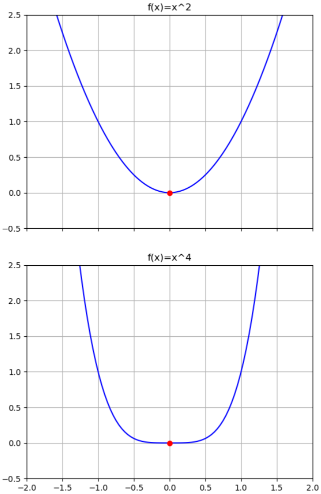
1. Optimality Conditions
   - Existence Results
   - Optimality conditions for differentiable functions in one dimension
   - Differentiability in $\mathbb{R}^n$
   - First Order Necessary Condition for Unconstrained Problems
   - First Order Necessary Condition for Constrained Problems

Optimality conditions for differentiable functions in 1-D, $X_{ad} = X = \mathbb{R}$

- First order necessary condition $f'(x^*) = 0$
  - Example: $f(x) = x^2$ local minimum at $x^* = 0$, there: $f'(x^*) = 2x^* = 0$.
- Second order sufficient condition $f'(x^*) = 0, f''(x^*) > 0$.
  - $f(x) = x^2$: local minimum at $x^* = 0$, $f''(x^*) = 2 > 0$.
  - $f(x) = x^4$: local minimum at $x^* = 0$, $f'(x^*) = 4x^{*3} = 0$. $f''(x^*) = 12x^{*2} = 0$, but $f''(x^*) = 12x^{*2} \not> 0$.
- If $f'(x^*) = 0$ and $f''(x^*) < 0$, then maximum.

$\rightsquigarrow$ Second order necessary condition for minimum: $f'(x^*) = 0$ and $f''(x^*) \geq 0$.

- $X_{ad} \neq X$ (e.g., interval) $\rightsquigarrow$ check boundary points.



Contents

1 Optimality Conditions
- Existence Results
- Optimality conditions for differentiable functions in one dimension
- Differentiability in $\mathbb{R}^n$
- First Order Necessary Condition for Unconstrained Problems
- First Order Necessary Condition for Constrained Problems

Generalization to $\mathbb{R}^n$: First order derivative

- The **gradient** of f at $x \in \mathbb{R}^n$ is defined as

$$\nabla f(x) := \left(\frac{\partial f}{\partial x_k}(x)\right)_{k=1}^{n} \in \mathbb{R}^n$$

with components (the **partial derivatives**):

$$\frac{\partial f}{\partial x_k}(x) := \lim_{h \to 0} \frac{f(x + he_k) - f(x)}{h}, \quad k = 1, \ldots, n$$

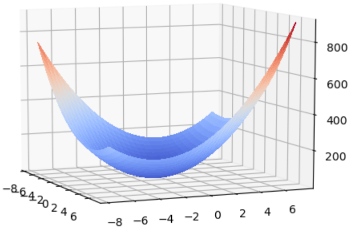
Here $e_k = (0, \ldots, 0, \underset{\substack{\uparrow \\ k}}{1}, 0, \ldots, 1)$ is the k-the unit vector.

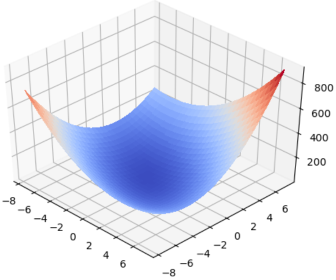
Example: First order derivative

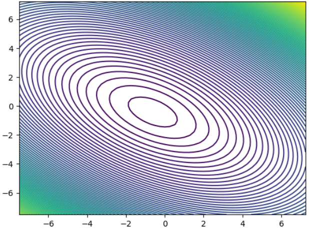
- Function $f : \mathbb{R}^2 \to \mathbb{R}$:

$$f(x) = f(x_1, x_2) = 4x_1^2 + 5x_1x_2 + 6x_2^2 + 7x_1 + 8x_2 + 9.$$

- Compute the partial derivatives and the gradient.







Example: First order derivative

- Function $f : \mathbb{R}^2 \to \mathbb{R}$:

$$f(x) = f(x_1, x_2) = 4x_1^2 + 5x_1x_2 + 6x_2^2 + 7x_1 + 8x_2 + 9.$$

- Partial derivatives:

$$\frac{\partial f}{\partial x_1}(x_1, x_2) = 8x_1 + 5x_2 + 7$$
$$\frac{\partial f}{\partial x_2}(x_1, x_2) = 5x_1 + 12x_2 + 8.$$

- Gradient:

$$\nabla f(x) = \left(\frac{\partial f}{\partial x_k}(x) \right)_{k=1}^{2} = \begin{pmatrix} 8x_1 + 5x_2 + 7 \\ 5x_1 + 12x_2 + 8 \end{pmatrix}$$

Contents

1 Optimality Conditions
- Existence Results
- Optimality conditions for differentiable functions in one dimension
- Differentiability in $\mathbb{R}^n$
- First Order Necessary Condition for Unconstrained Problems
- First Order Necessary Condition for Constrained Problems

First order necessary condition

Theorem

Let $f : \mathbb{R}^n \to \mathbb{R}$ have all partial derivatives in the local minimizer $x^* \in \mathbb{R}^n$. Then $\nabla f(x^*) = 0$.

Proof.

- x^* local minimizer $\Rightarrow f(x^* + he_k) - f(x^*) \geq 0$ for $h > 0$ small.
- Division by $h > 0 \Rightarrow \dfrac{f(x^* + he_k) - f(x^*)}{h} \geq 0$.
- f partially differentiable $\Rightarrow \dfrac{\partial f}{\partial x_k}(x) = \lim\limits_{h \to 0} \dfrac{f(x^* + he_k) - f(x^*)}{h} \geq 0$.
- Analogously: $f(x^* + he_k) - f(x^*) \geq 0$ for $h < 0$ small, division by $h < 0$ changes inequality $\Rightarrow \dfrac{\partial f}{\partial x_k}(x^*) = \lim\limits_{h \to 0} \dfrac{f(x^* + he_k) - f(x^*)}{h} \leq 0$.
- $\dfrac{\partial f}{\partial x_k}(x^*) \geq 0$ and $\dfrac{\partial f}{\partial x_k}(x^*) \leq 0$ gives $\dfrac{\partial f}{\partial x_k}(x^*) = 0$.This holds for all k.

Example: First order necessary optimality condition

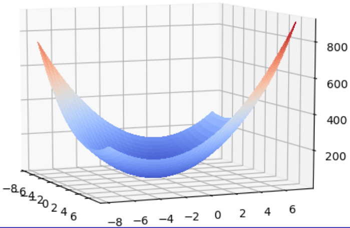
- Function $f : \mathbb{R}^2 \to \mathbb{R}$:

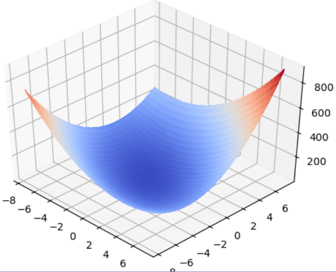
$$f(x) = f(x_1, x_2) = 4x_1^2 + 5x_1x_2 + 6x_2^2 + 7x_1 + 8x_2 + 9.$$

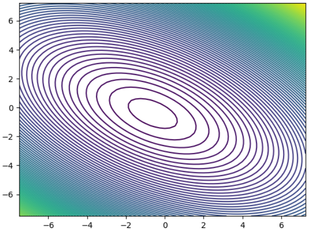
- Gradient:

$$\nabla f(x) = \begin{pmatrix} 8x_1 + 5x_2 + 7 \\ 5x_1 + 12x_2 + 8 \end{pmatrix} = 0$$

- Two linear equations: there exists a (unique) solution.





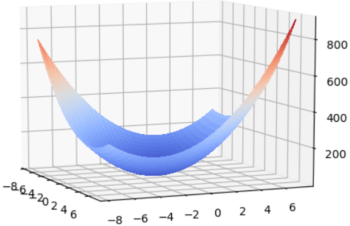


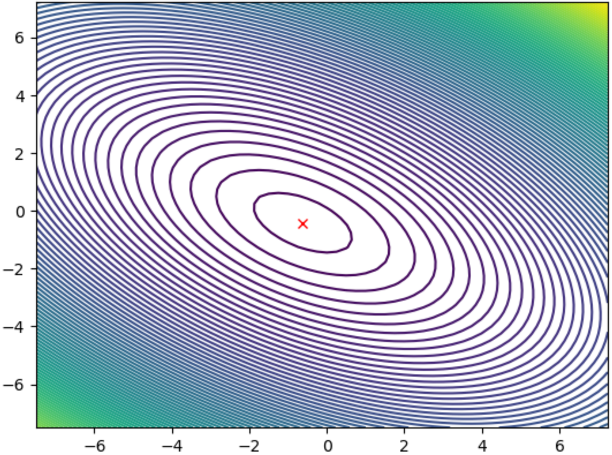
Example: First order necessary optimality condition

- Gradient: Two linear equations: there exists a (unique) solution.

$$\nabla f(x) = \begin{pmatrix} 8x_1 + 5x_2 + 7 \\ 5x_1 + 12x_2 + 8 \end{pmatrix} = 0$$

- It is a minimum, since the function has no maximum:





Contents

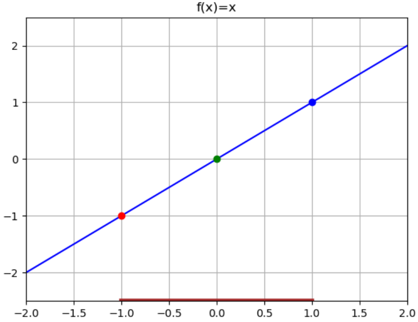
1 Optimality Conditions
- Existence Results
- Optimality conditions for differentiable functions in one dimension
- Differentiability in $\mathbb{R}^n$
- First Order Necessary Condition for Unconstrained Problems
- First Order Necessary Condition for Constrained Problems

Admissible directions

- If X_{ad} is bounded, a minimizer can be on the boundary.
- At the boundary, we cannot compute the limit in the partial derivatives

$$\frac{\partial f}{\partial x_k}(x) := \lim_{h \to 0} \frac{f(x + he_k) - f(x)}{h},$$

- It can only be considered in special directions:



Definition

For $x \in X_{ad} \subset \mathbb{R}^n$ we call $d \in \mathbb{R}^n$ an **admissible direction** (in x), if there exists $h_0 > 0$ with

$$x + hd \in X_{ad} \quad \text{for all } h \in (0, h_0).$$

Directional derivatives

Definition

Let $f : X_{ad} \to \mathbb{R}, x \in X_{ad} \subset \mathbb{R}^n$ and $d \in \mathbb{R}^n$ be an admissible direction (in x). If the limit

$$Df(x;d) := \lim_{h \downarrow 0} \frac{f(x+hd) - f(x)}{h} \tag{1}$$

exists, f is called **differentiable in direction** d, and (1) is called **directional derivative**.

- The directional derivative is a generalization of the partial derivative.
- It is useful at boundary points ...
- and if f is not differentiable in some point x,

$$f : \mathbb{R} \to \mathbb{R}, \quad f(x) = |x|,$$

- If f has continuous partial derivatives in x, then

$$Df(x;d) = \nabla f(x)^\top d \ \text{ for all } d \in \mathbb{R}^n.$$

Generalization of first order necessary condition

Theorem

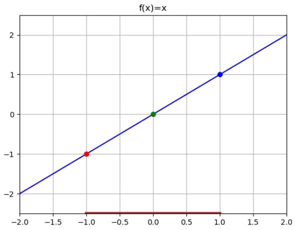
Let $x^* \in X_{ad}$ be a local minimizer of $f : X_{ad} \to \mathbb{R}$. If f is differentiable at x^* in direction $d \in \mathbb{R}^n$, then $Df(x^*; d) \geq 0$.

Proof.

- $Df(x^*; d) = \lim_{h \downarrow 0} \dfrac{f(x^* + hd) - f(x^*)}{h}$ exists.

$\Rightarrow$ d is admissible direction $\Rightarrow x^* + hd \in X_{ad}$ for small $h > 0$.

- x^* local minimizer: $\Rightarrow f(x^* + hd) - f(x^*) \geq 0$ for small $h > 0$.
- $\Rightarrow \dfrac{f(x^* + hd) - f(x^*)}{h} \geq 0$ for small $h > 0$.
- Let $h \downarrow 0$: $Df(x^*; d) = \lim_{h \downarrow 0} \dfrac{f(x^* + hd) - f(x^*)}{h} \geq 0$.



$\square$

What is important?

- The Weierstrass Theorem gives the existence of global optima in a closed bounded set for continuous functions ...
- ... but it is not directly applicable for unbounded sets, i.e., unconstrained problems.
- Radial boundedness may help to use the theorem also for unconstrained problems.
- Differentiability of the cost function provides a first order necessary condition for minima.
- The concept of differentiability can be extended to n dimensions.
- For unconstrained problems, the necessary condition is that the gradient vanishes.
- For constrained problems, we use the directional derivative and admissible directions.
- At a minimizer, the directional derivative is greater or equal to zero in all admissible directions.