

Analyse: van \mathbb{R} naar \mathbb{R}^n hoorcollege

Extrema in \mathbb{R}^n (21)

Gerrit Oomens
G.Oomens@uva.nl

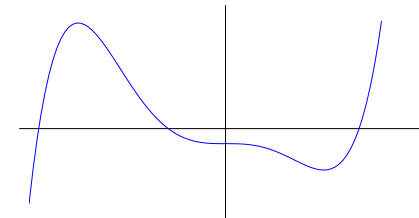
Korteweg-de Vries Instituut voor Wiskunde
Faculteit der Natuurwetenschappen, Wiskunde en Informatica
Universiteit van Amsterdam



Extrema in \mathbb{R}

Bekijk een C^2 -functie $f: E \rightarrow \mathbb{R}$, waar $E \subseteq \mathbb{R}$ open. We zeggen dat f een **lokaal maximum** heeft in a als er een $\delta > 0$ bestaat zodat voor alle $x \in (a - \delta, a + \delta)$ geldt $f(x) \leq f(a)$. Bekend:

- 1 als f een lokaal extremum aanneemt in a , dan $f'(a) = 0$,
- 2 als f een maximum aanneemt in a , dan is $f''(a) \leq 0$,
- 3 als $f'(a) = 0$ en $f''(a) < 0$, dan heeft f een maximum in a .



Extrema in \mathbb{R}^n

We bekijken een functie $f: E \rightarrow \mathbb{R}$, waar $E \subseteq \mathbb{R}^n$.

- f heeft een **lokaal maximum** in $\vec{a} \in E$ als er een $\delta > 0$ is zodat $f(\vec{x}) \leq f(\vec{a})$ voor $\vec{x} \in B(\vec{a}, \delta)$.
- We noemen \vec{a} een **absoluut maximum** als $f(\vec{x}) \leq f(\vec{a})$ voor alle $\vec{x} \in E$. Anders heet \vec{a} **relatief**.
- We noemen \vec{a} een **sterk maximum** als er een $\delta > 0$ is zodat $f(\vec{x}) < f(\vec{a})$ voor $\vec{x} \in B(\vec{a}, \delta)$ en $\vec{x} \neq \vec{a}$. Anders heet \vec{a} **zwak**.
- Het punt \vec{a} kan een **inwendig maximum** ($\vec{a} \in E^\circ$) of een **randmaximum** ($\vec{a} \in \partial E$) zijn.

Extrema en afgeleides

Propositie 13.6

Zij $f: E \rightarrow \mathbb{R}$ en \vec{a} een inwendig extremum van f , dan

- (1) Neem $\vec{u} \in \mathbb{R}^n$. Als de richtingsafgeleide $(D_{\vec{u}}f)(\vec{a})$ bestaat, dan is $(D_{\vec{u}}f)(\vec{a}) = \vec{0}$.
- (2) Als f differentieerbaar is in \vec{a} , dan geldt $f'(\vec{a}) = \vec{0}$.

Een punt met $f'(\vec{a}) = \vec{0}$ heet een **stationair punt**.

Bewijs: neem $\vec{u} \in \mathbb{R}^n$ en definieer $g(t) = f(\vec{a} + t\vec{u})$. Dan is g een functie op \mathbb{R} met een lokaal extremum in 0. Dus

$$(D_{\vec{u}}f)(\vec{a}) := g'(0) = 0.$$

Dit bewijst (1). Merk nu op dat als f differentieerbaar is, dat voor alle \vec{u} geldt

$$f'(\vec{a})\vec{u} = (D_{\vec{u}}f)(\vec{a}) = 0,$$

dus dan volgt $f'(\vec{a}) = \vec{0}$. □

Taylorreeks

Zij $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ een C^2 functie. De Stelling van Taylor geeft

$$\begin{aligned} f(\vec{a} + \vec{h}) &= f(\vec{a}) + \sum_{j_1=1}^2 \frac{D_{j_1} f(\vec{a})}{1!} h_{j_1} + \sum_{j_1, j_2=1}^2 \frac{D_{j_1 j_2} f(\vec{a})}{2!} h_{j_1} h_{j_2} + o(\|\vec{h}\|^2) \\ &= f(\vec{a}) + D_1 f(\vec{a}) h_1 + D_2 f(\vec{a}) h_2 \\ &\quad + \frac{1}{2} [D_{11} f(\vec{a}) h_1^2 + 2D_{12} f(\vec{a}) h_1 h_2 + D_{22} f(\vec{a}) h_2^2] + o(\|\vec{h}\|^2) \\ &= f(\vec{a}) + [D_1 f(\vec{a}) \quad D_2 f(\vec{a})] \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} \\ &\quad + \frac{1}{2} [h_1 \quad h_2] \begin{bmatrix} D_{11} f(\vec{a}) & D_{12} f(\vec{a}) \\ D_{12} f(\vec{a}) & D_{22} f(\vec{a}) \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} + o(\|\vec{h}\|^2) \\ &= f(\vec{a}) + f'(\vec{a}) \vec{h} + \frac{1}{2} \vec{h}^T H_f(\vec{a}) \vec{h} + o(\|\vec{h}\|^2), \end{aligned}$$

waarbij

$$H_f(\vec{a}) = \begin{bmatrix} D_{11} f(\vec{a}) & D_{12} f(\vec{a}) \\ D_{12} f(\vec{a}) & D_{22} f(\vec{a}) \end{bmatrix}$$

de **Hesse-matrix** of **Hessiaan** van f wordt genoemd.

Taylorreeks: meer algemeen

Zij $f: \mathbb{R}^n \rightarrow \mathbb{R}$ een C^2 functie. Er geldt

$$\begin{aligned} f(\vec{a} + \vec{h}) &= f(\vec{a}) + \sum_{j_1=1}^n \frac{D_{j_1} f(\vec{a})}{1!} h_{j_1} + \sum_{j_1, j_2=1}^n \frac{D_{j_1 j_2} f(\vec{a})}{2!} h_{j_1} h_{j_2} + o(\|\vec{h}\|^2) \\ &= f(\vec{a}) + f'(\vec{a}) \vec{h} + \frac{1}{2} \vec{h}^T H_f(\vec{a}) \vec{h} + o(\|\vec{h}\|^2), \end{aligned}$$

waarbij

$$H_f(\vec{a}) = \begin{pmatrix} D_{11} f(\vec{a}) & \cdots & D_{1n} f(\vec{a}) \\ \vdots & & \vdots \\ D_{n1} f(\vec{a}) & \cdots & D_{nn} f(\vec{a}) \end{pmatrix}$$

de **Hesse-matrix** of **Hessiaan** van f wordt genoemd. In het bijzonder geldt in een stationair punt

$$f(\vec{a} + \vec{h}) = f(\vec{a}) + \frac{1}{2} \vec{h}^T H_f(\vec{a}) \vec{h} + o(\|\vec{h}\|^2).$$

Tweede afgeleide en extrema

Propositie 13.9

Zij $f: E \rightarrow \mathbb{R}^n$ een C^2 functie en $\vec{a} \in E^\circ$. Als f een lokaal minimum aanneemt in \vec{a} , dan geldt $\vec{h}^T H_f(\vec{a}) \vec{h} \geq 0$ voor alle $\vec{h} \in \mathbb{R}^n$.

Bewijs: neem $\vec{h} \in \mathbb{R}^n$ en bekijk

$$g(t) := f(\vec{a} + t\vec{h}).$$

Als f een lokaal minimum aanneemt, dan neemt g een lokaal minimum aan in 0. Dus $g''(0) \geq 0$. Met wat rekenwerk (Lemma 12.5) zien we dat

$$g''(0) = \vec{h}^T H_f(\vec{a}) \vec{h}.$$

Tweede afgeleide en extrema

Stelling 13.10

Zij $f: E \rightarrow \mathbb{R}$ een C^2 -functie en $\vec{a} \in E^\circ$ met $f'(\vec{a}) = \vec{0}$. Stel dat voor alle $\vec{h} \neq 0$ geldt $\vec{h}^T H_f(\vec{a}) \vec{h} > 0$. Dan neemt f een sterk lokaal minimum aan in \vec{a} .

- Bekijk $\Lambda: \mathbb{R}^n \rightarrow \mathbb{R}$ gegeven door $\Lambda(\vec{h}) = \vec{h}^T H_f(\vec{a}) \vec{h}$.
- Dit is continu en neemt dus een minimum aan op de compacte verzameling $\{\vec{h} \in \mathbb{R}^n : \|\vec{h}\| = 1\}$.
- Dus voor zekere \vec{h}_0 geldt $\mu := \inf_{\|\vec{h}\|=1} \Lambda(\vec{h}) = \Lambda(\vec{h}_0) > 0$.
- Voor $\vec{h} \neq 0$ willekeurig geldt nu

$$\Lambda(\vec{h}) = \|\vec{h}\|^2 \frac{\vec{h}^T H_f(\vec{a}) \vec{h}}{\|\vec{h}\|^2} = \|\vec{h}\|^2 \left[\frac{\vec{h}}{\|\vec{h}\|} \right]^T H_f(\vec{a}) \left[\frac{\vec{h}}{\|\vec{h}\|} \right] \geq \mu \|\vec{h}\|^2.$$

- Er bestaat $\delta > 0$ zodat als $\|\vec{h}\| < \delta$, dan $|o(\|\vec{h}\|^2)| < \frac{\mu}{4} \|\vec{h}\|^2$.
- Er volgt

$$f(\vec{a} + \vec{h}) - f(\vec{a}) = \frac{1}{2} \vec{h}^T H_f(\vec{a}) \vec{h} + o(\|\vec{h}\|^2) \geq \frac{\mu}{2} \|\vec{h}\|^2 - \frac{\mu}{4} \|\vec{h}\|^2 > 0.$$

Symmetrische en positief definitie matrices

Bekijk een symmetrische matrix A .

- We noemen A **positief semidefinit** als $\vec{x}^T A \vec{x} \geq 0$ voor alle \vec{x} .
- Als geldt $\vec{x}^T A \vec{x} > 0$ voor $\vec{x} \neq 0$, dan heet A **positief definit**.

Analoog definiëren we **negatief (semi)definit**. Uit de lineaire algebra weten we dat we kunnen schrijven

$$A = O^T D O = \left[\begin{array}{c|c|c} \vec{e}_1 & \cdots & \vec{e}_n \end{array} \right] \left[\begin{array}{ccc} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{array} \right] \left[\begin{array}{c} \vec{e}_1^T \\ \vdots \\ \vec{e}_n^T \end{array} \right],$$

waarbij de \vec{e}_i de orthonormale eigenvectoren van A zijn en de λ_i de eigenwaarden.

Neem $\vec{x} \in \mathbb{R}^n$ en schrijf $O\vec{x} = [\tilde{x}_1 \ \cdots \ \tilde{x}_n]^T$. Dan

$$\vec{x}^T A \vec{x} = (O\vec{x})^T D O\vec{x} = \langle O\vec{x}, D O\vec{x} \rangle = \sum_{j=1}^n \lambda_j \tilde{x}_j^2.$$

We zien: A positief definit \Leftrightarrow alle eigenwaarden zijn positief.

Complicaties

Indien de Hessiaan in een stationair punt \vec{a} te veel eigenwaarden 0 heeft, geeft deze geen informatie. We moeten dan op een andere manier het gedrag van f rond \vec{a} onderzoeken, bijvoorbeeld door naar de **niveaукrommen** $f(\vec{x}) = c$ te kijken.

Als we bekijken $f: E \rightarrow \mathbb{R}$ waar E niet open is, zullen we ook naar randextrema moeten zoeken. Hiervoor kunnen we de rand parametriseren als een kromme $(x(t), y(t))$ en zoeken naar extrema van

$$g(t) = f(x(t), y(t)).$$

Vervolgens moeten we onderzoeken of deze ook extrema van f zijn.

Vinden van extrema (in \mathbb{R}^2)

Om extrema te vinden van een differentieerbare functie, zoeken we een punt met $f'(\vec{a}) = 0$. We willen dan onderzoeken wat voor punt \vec{a} is. Wanneer f een C^2 functie is, bekijken we de **Hessiaan**

$$H_f(\vec{a}) = \begin{bmatrix} (D_{11}f)(\vec{a}) & (D_{12}f)(\vec{a}) \\ (D_{21}f)(\vec{a}) & (D_{22}f)(\vec{a}) \end{bmatrix}.$$

- Als de eigenwaarden van $H_f(\vec{a})$ kleiner dan 0 zijn ($H_f(\vec{a})$ is **negatief definit**), dan neemt f een sterk lokaal maximum aan.
- Als de eigenwaarden van $H_f(\vec{a})$ groter dan 0 zijn ($H_f(\vec{a})$ is **positief definit**), dan neemt f een sterk lokaal minimum aan.
- Als $H_f(\vec{a})$ zowel positieve als negatieve eigenwaarden heeft (**indefinit**), dan heeft f geen extremum in \vec{a} (**zadelpunt**).

Truc: het teken van de eigenwaarden λ_1, λ_2 is makkelijk te bepalen door $\text{tr } H_f(\vec{a}) = \lambda_1 + \lambda_2$ en $\det H_f(\vec{a}) = \lambda_1 \lambda_2$.

Voorbeeld

Bekijk $f(x, y) = x^2 + y^2 + x$ op $E = \{(x, y) : x^2 + y^2 \leq 1\}$.

- We bepalen de partiële afgeleiden:

$$\begin{aligned} D_1 f(x, y) = 2x + 1 = 0 & \Rightarrow x = -\frac{1}{2} \\ D_2 f(x, y) = 2y = 0 & \Rightarrow y = 0 \end{aligned}$$

- We zien $f'(x, y) = 0$ in $(-\frac{1}{2}, 0)$. Er geldt

$$H_f(x, y) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix},$$

dus $H_f(-\frac{1}{2}, 0)$ is positief definit: dit punt is een inwendig sterk minimum.

Voorbeeld

Bekijk $f(x, y) = x^2 + y^2 + x$ op $E = \{(x, y) : x^2 + y^2 \leq 1\}$. We hebben gezien dat f een inwendig sterk minimum heeft in $(-\frac{1}{2}, 0)$.

- We parametriseren de rand met $(\cos t, \sin t)$ waar $t \in [0, 2\pi)$:

$$g(t) := f(\cos t, \sin t) = \cos^2 t + \sin^2 t + \cos t = 1 + \cos t.$$

- Dan is $g'(t) = -\sin t$. We vinden 2 mogelijke randextrema:

$$t = 0 \quad \Rightarrow \quad (x, y) = (1, 0)$$

$$t = \pi \quad \Rightarrow \quad (x, y) = (-1, 0)$$

- Merk op: 0 is een maximum van g . De functie f neemt op E ergens een maximum aan, maar niet in een inwendig punt. Dus moet $(1, 0)$ een absoluut (sterk) randmaximum van f zijn.
- Het punt π is een minimum van g . Het zou dus kunnen dat $(-1, 0)$ een randminimum van f is met $f(-1, 0) = 0$. Echter

$$f(x, 0) = x^2 + x = x(x + 1) < 0 \quad \text{als } x > -1.$$

We concluderen dat $(-1, 0)$ geen extremum van f is.