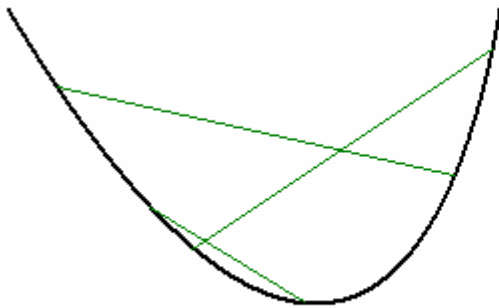


Convexe functies op \mathbb{R} (niet in het boek)

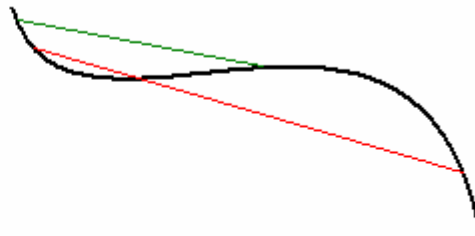
Een functie $f: \mathbb{R} \rightarrow \mathbb{R}$ heet convex, als voor alle $x, y \in \mathbb{R}$ en elke $\lambda \in [0,1]$ geldt dat

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

Voor een functie op \mathbb{R} betekent dit dat als je twee willekeurige punten op de grafiek met een recht lijnstukje verbindt, de grafiek tussen die punten nooit boven dit lijnstukje komt. De grafiek van een convexe functie is “hol” naar boven, “bol” naar beneden, bijvoorbeeld $f(x) = x^2$.



Convex



Niet convex

Een functie $f: \mathbb{R} \rightarrow \mathbb{R}$ heet strikt convex, als voor alle $x, y \in \mathbb{R}$ en elke $\lambda \in]0,1[$ geldt dat

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

Een functie f heet (strikt) concaaf als $-f$ (strikt) convex is. De grafiek van een concave functie is “bol” naar boven, “hol” naar beneden.

Een lineair functie is de enige functie die zowel convex als concaaf is, maar is niet strikt convex of strikt concaaf.

Voorbeeld:

Bewijs met de definitie dat $f(x) = 1/x$ convex is als $x > 0$.

Hiervoor moeten we laten zien dat voor iedere $x, y > 0$ en $\lambda \in [0,1]$ geldt:

$$\frac{1}{\lambda x + (1-\lambda)y} \leq \frac{\lambda}{x} + \frac{1-\lambda}{y} = \frac{\lambda y + (1-\lambda)x}{xy}.$$

Omdat x, y en $\lambda x + (1-\lambda)y$ allemaal positief zijn kunnen we de ongelijkheid hiermee vermenigvuldigen en is dit equivalent met aantonen dat

$$xy \leq \lambda^2 xy + \lambda(1-\lambda)(x^2 + y^2) + (1-\lambda)^2 xy$$

ofwel dat:

$$-\lambda^2(x^2 + y^2 - 2xy) + \lambda(x^2 + y^2 - 2xy) \geq 0.$$

Maar dit is hetzelfde als

$$\lambda(1-\lambda)(x - y)^2 \geq 0$$

en dit is duidelijk waar. Als je een net bewijs wilt moet je dit argument achterstevoren opschrijven! \square

De grafiek van een functie is de verzameling

$$\{(x,y) \in \mathbb{R}^2 \mid y = f(x)\}$$

De epigrafiek van een functie is de verzameling

$$\{(x,y) \in \mathbb{R}^2 \mid y \geq f(x)\}.$$

Dat zijn alle punten die op of boven de grafiek liggen.

Het is eenvoudig in te zien dat een functie convex is precies dan als zijn epigrafiek een convexe verzameling in \mathbb{R}^2 is.

Als f convex is, dan is ook een veelvoud af convex, voor elk getal $a \geq 0$ (en af is concaaf voor elk getal $a \geq 0$).

Als f en g convex zijn, dan ook $f + g$ (maar $f - g$ meestal niet). Een constante functie is convex, dus in het bijzonder blijft een functie convex als je er een constante bij optelt.

Als f en h convex zijn en h bovendien monotoon niet-dalend, dan is $h \circ f$ weer convex. Bewijs:

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

want f is convex en h is niet-dalend

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

want h is convex, dus $h \circ f$ is convex.

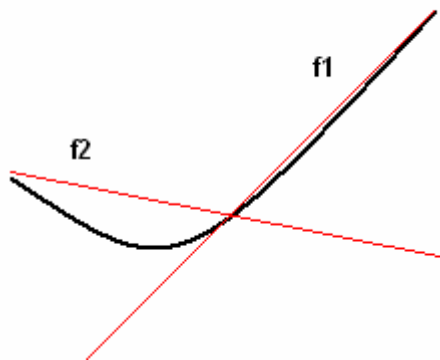
Convexiteit is een behoorlijke beperking voor een functie. Zo kan een convexe functie eigenlijk geen sprongen maken en is zo goed als continu. Alleen op de rand van het domein kunnen er sprongen optreden:

Stelling: Als $f: C \rightarrow \mathbb{R}$ convex is, dan is f continu op het inwendige van C .

Bewijs: Kies x in het inwendige van C . Het interval $[x - \varepsilon, x + \varepsilon]$ is dan helemaal in C bevat voor een zekere $\varepsilon > 0$. Neem voor het gemak aan dat $x = 0$ en dat $f(0) = 0$ (Dat is geen restrictie, door schuiven). Bekijk de functies

$$f_1(x) = \frac{f(\varepsilon)}{\varepsilon}x \quad \text{en} \quad f_2(x) = \frac{f(-\varepsilon)}{-\varepsilon}x$$

De functie f ligt tussen deze twee functies en dat zien we als volgt. Als $x \in [0, \varepsilon]$, dan is $f_2(x) \leq f(x) \leq f_1(x)$.



Dit volgt uit de convexiteitsongelijkheid

$$f(\lambda s + (1-\lambda)t) \leq \lambda f(s) + (1-\lambda)f(t)$$

als je kiest: $s = -\varepsilon$, $t = x$, $\lambda = x/(x+\varepsilon)$, voor de linker ongelijkheid en $s = 0$, $t = \varepsilon$, $1 - \lambda = x/\varepsilon$ voor de rechter.

Op dezelfde manier geldt dat $f_1(x) \leq f(x) \leq f_2(x)$ als $x \in [-\varepsilon, 0]$. De grafiek van f ligt dus eigenlijk ingeklemd tussen de twee rechte lijnen die de grafieken van de functies f_1 en f_2 vormen. Omdat f_1 en f_2 continu zijn met $f_1(0) = f_2(0) = 0$ geldt dat $f(x) \rightarrow f_1(0) = f_2(0) = 0 = f(0)$ als $x \rightarrow 0$, dus f is continu in 0, waarmee we de stelling hebben bewezen. \square

De convexiteit van een functie impliceert niet de continuïteit overal, omdat het op de rand van een gebied mis kan gaan.

Zo is de functie f waarvoor geldt:

$f(0) = f(1) = 1$ en $f(x) = 0$ als $0 < x < 1$

niet continu in 0 en 1, maar wel convex.

Als een differentieerbare functie convex is dan kan de afgeleide niet dalend zijn. Dat betekent dat de tweede afgeleide nooit negatief is. Soms kun je van een functie eenvoudig laten zien dat de tweede afgeleide positief is. Dat is dan een simpele manier om de convexiteit van de functie aan te tonen.

Stelling: $f \in C^2(\mathbb{R}, \mathbb{R})$ is convex dan en slechts dan als $f''(x) \geq 0$ voor alle x .

Bewijs:

“ \Rightarrow ” We moeten aantonen dat de tweede afgeleide van een convexe functie overal niet-negatief is. Omdat f convex is, geldt er (kies $s = x - h$, $t = x + h$, $\lambda = 1/2$) dat :

$$\frac{1}{2} f(x+h) + \frac{1}{2} f(x-h) \geq f\left(\frac{1}{2}(x+h) + \frac{1}{2}(x-h)\right) = f(x).$$

Voor de tweede afgeleide gebruiken we nu het volgende differentiequotiënt, en zien met de vorige ongelijkheid direct dat de tweede afgeleide niet-negatief is:

$$f''(x) = \lim_{h \downarrow 0} \frac{f(x+h) - 2f(x) + f(x-h)}{h^2} \geq 0.$$

(maak een Taylorreeksontwikkeling rond x om te zien dat het differentiequotiënt klopt)

“ \Leftarrow ” Nu laten zien dat een functie met niet-negatieve tweede afgeleide overal convex is.

Integreer de tweede afgeleide te integreren:

$$f'(y) - f'(x) = \int_x^y f''(t) dt.$$

Als $y \geq x$ dan is de integraal niet-negatief omdat de integrand het is, dus $f'(y) \geq f'(x)$. We hebben hiermee aangetoond dat de afgeleide van f niet-dalend is.

We halen nu dezelfde truc uit en integreren f' :

Als $y \geq x$ dan is
$$f(y) - f(x) = \int_x^y f'(t) dt \geq \int_x^y f'(x) dt = f'(x)(y - x).$$

Dit klopt omdat f' niet-dalend is, dus $f'(t) \geq f'(x)$ als $t \geq x$.

Als $y \leq x$ dan is
$$f(y) - f(x) = \int_x^y f'(t) dt = - \int_y^x f'(t) dt \geq \int_y^x -f'(x) dt = f'(x)(y - x).$$

Gevolg: $f(y) - f(x) \geq f'(x)(y - x)$.

De grafiek van f ligt dus nooit onder de raaklijn in het punt x .

Noem $t_\lambda = \lambda x + (1-\lambda)y$, dan is:

$$\begin{aligned} f(y) - f(t_\lambda) &\geq f'(t_\lambda)(y - t_\lambda), \\ f(x) - f(t_\lambda) &\geq f'(t_\lambda)(x - t_\lambda). \end{aligned}$$

Vermenigvuldig de eerste ongelijkheid met $1 - \lambda$, de tweede met λ en tel ze op (deze getallen zijn niet-negatief):

$$\lambda f(x) + (1-\lambda)f(y) - f(t_\lambda) \geq f'(t_\lambda)(\lambda x + (1-\lambda)y - t_\lambda) = 0,$$

want $t_\lambda = \lambda x + (1-\lambda)y$. Hier staat dat

$f(\lambda x + (1-\lambda)y) = f(t_\lambda) \leq \lambda f(x) + (1-\lambda)f(y)$, dus f is convex.

□

Voorbeelden:

Door tweemaal te differentiëren volgt:

$f(x) = ax^2 + bx + c$ is convex als $a \geq 0$,

$f(x) = e^x$ is convex op \mathbb{R} ,

$f(x) = -\log x$ is convex als $x > 0$,

$f(x) = 1/x$ is convex als $x > 0$ (en concaaf als $x < 0$),

$f(x) = x \log x$ is convex als $x > 0$,

$f(x) = x^p$ is convex als $x > 0$ en $p \geq 1$, of $p \leq 0$ (en concaaf als $x > 0$ en $0 \leq p \leq 1$).

Soms kun je met de tweede afgeleide de convexiteit van een functie bewijzen en vervolgens via de definitie van convexiteit ongelijkheden opschrijven die op een andere manier veel moeilijker of niet te bewijzen zijn.

Uit de convexiteit van $-\log x$ volgt bijvoorbeeld dat

$$x^\lambda y^{1-\lambda} \leq \lambda x + (1-\lambda)y$$

Voor $\lambda = 1/2$ staat hier:

$$\sqrt{xy} \leq \frac{x+y}{2}.$$

De linkerkant heet het meetkundig gemiddelde van x en y , de rechterkant is het gewone (rekenkundige) gemiddelde.

Het meetkundig gemiddelde is dus altijd kleiner dan of gelijk aan het rekenkundig gemiddelde.

Je kunt ook nog een harmonisch gemiddelde definiëren als

$$\left[\frac{1}{2} \left(\frac{1}{x} + \frac{1}{y} \right) \right]^{-1} = \frac{2xy}{x+y}$$

Het is eenvoudig in te zien dat het harmonisch gemiddelde altijd kleiner dan of gelijk aan het meetkundig gemiddelde is.

Een eenvoudig gevolg van convexiteit is:

Stelling: Een minimum van een (strikt) convexe functie op een convex gebied is (strikt) absoluut.

Strict convexe functies zijn dus “simpel” in de zin dat ze geen locale minima hebben, er is maar één minimum, dus iteratieve methoden die een lokaal minimum vinden, vinden ook het globale minimum.

Laat zien dat een convexe functie op een niet-convexe verzameling wel locale minima kan hebben.

Een functie $f: \mathbb{R} \rightarrow \mathbb{R}$ heet midpuntconvex, als voor alle $x, y \in \mathbb{R}$ geldt dat

$$f((x+y)/2) \leq (f(x) + f(y))/2.$$

Een functie die convex is, is duidelijk midpointconvex (neem $\lambda = 1/2$)

Geldt het omgekeerde ook?

Antwoord: Nee

(Walgelijk ingewikkeld) voorbeeld

Voor de reële getallen bestaat er een Hamelbasis $\{x_\alpha \in \mathbb{R} \mid \alpha \in A\}$ (een overaftelbare verz. van reële getallen)

Voor elk reëel getal x zijn er eindig veel rationale getallen ξ_i zodat $x = \sum \xi_i x_{\alpha(i)}$. Deze representatie is uniek (basis).

Definieer nu een functie f op alle elementen van de Hamelbasis: $f_\alpha = f(x_\alpha)$, zodanig dat niet alle getallen $f(x_\alpha)/x_\alpha$ gelijk zijn. Definieer dan f op heel \mathbb{R} :

$$f(x) = f(\sum \xi_i x_{\alpha(i)}) := \sum \xi_i f_{\alpha(i)}$$

Deze functie is bijna lineair, want voldoet aan:

$$f(x+y) = f(x) + f(y) \text{ en}$$

$$f(\lambda x) = \lambda f(x), \text{ maar alleen als } \lambda \in \mathbb{Q}.$$

Hierdoor is f ook midpointconvex.

f is echter in geen enkel punt continu:

Kies de functie f bijvoorbeeld zo dat $f_{\alpha(0)} = 1$ voor één bepaalde $\alpha(0)$ en $f_\alpha = 0$ voor alle andere α .

Kies $x \in \mathbb{R}$, dan is $x = \xi_0 x_{\alpha(0)} + \sum \xi_i x_{\alpha(i)}$ dus $f(x) = \xi_0$.

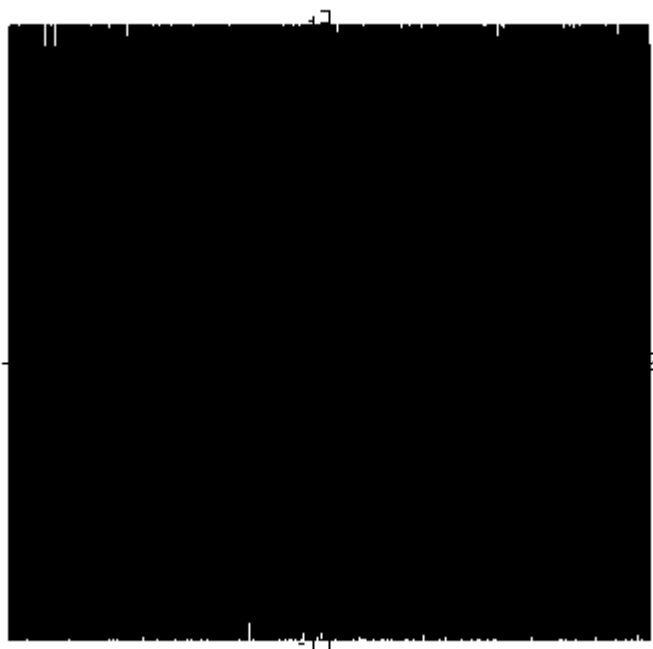
Kies nu een willekeurige $y \in \mathbb{R}$. Kies een $\xi \in \mathbb{Q}$ die minder dan ε van $y - \xi_0$ af zit. Nu ligt $\mathbb{R} \setminus \{x_{\alpha(0)}\mathbb{Q}\}$ dicht in \mathbb{R} , dus je kunt ook een $z \in \mathbb{R} \setminus \{x_{\alpha(0)}\mathbb{Q}\}$ kiezen die minder dan ε van $\xi x_{\alpha(0)}$ af ligt.

Dan ligt $x + \xi x_{\alpha(0)} - z$ minder dan ε van x af, terwijl

$$|f(x + \xi x_{\alpha(0)} - z) - y| = |f(x) + \xi f_{\alpha(0)} - f(z) - y| = |\xi_0 + \xi - y| < \varepsilon.$$

Gevolg: willekeurig dicht bij x liggen punten waarvan de functiewaarde willekeurig dicht bij een willekeurig gekozen y ligt.

De grafiek ziet er dus ongeveer zo uit:



Bij elk punt uit het vlak ligt een punt van de grafiek in de buurt, de grafiek is dus inderdaad helemaal zwart.
 f is dus niet continu en kan dus niet convex zijn.

Convexe functies op \mathbb{R}^n (niet in het boek)

Een functie $f : \mathbb{R}^n \rightarrow \mathbb{R}$ heet convex, als voor alle $x, y \in \mathbb{R}^n$ en elke $\lambda \in [0,1]$ geldt dat

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

Hetzelfde geldt als de functie op een convex deel van \mathbb{R}^n is gedefinieerd. De uitbreidingen naar strikt convex en concaaf zijn volledig analoog aan het ééndimensionale geval.

Voorbeeld: We zullen bewijzen dat de euclidische normfunctie $f(x) = |x|$ convex is op \mathbb{R}^n .

Er geldt dat

$$\begin{aligned} f(\lambda x + (1-\lambda)y)^2 &= \\ |\lambda x + (1-\lambda)y|^2 &= \lambda^2|x|^2 + (1-\lambda)^2|y|^2 + \lambda(1-\lambda)(x,y) \leq \\ \lambda^2|x|^2 + (1-\lambda)^2|y|^2 &+ \lambda(1-\lambda)|x| |y| = (\lambda|x| + (1-\lambda)|y|)^2 = \\ [f(\lambda x) + (1-\lambda)f(y)]^2 \end{aligned}$$

want volgens de ongelijkheid van Schwarz is $(x,y) \leq |x| |y|$. Hieruit volgt dat f convex is. □

Als f convex is, dan is ook af weer convex, voor elke scalar $a \geq 0$.

Als f en g convex zijn, dan ook $f + g$ (maar $f - g$ meestal niet). Een constante functie is convex, dus in het bijzonder blijft een functie convex als je er een constante bij optelt.

Als $f : \mathbb{R}^n \rightarrow \mathbb{R}$ en $h : \mathbb{R} \rightarrow \mathbb{R}$ convex zijn en h bovendien monotoon niet-dalend, dan is $h \circ f$ weer convex.

Er geldt ook weer dat een convexe functie continu is op het inwendige van zijn domein, maar dat er op de rand sprongen kunnen optreden.

Voor de convexiteit van een tweemaal differentieerbare functie is er ook een criterium in termen van zijn tweede afgeleide:

Stelling: $f \in C^2(\mathbb{R}^n, \mathbb{R})$ is convex dan en slechts dan als $\nabla^2 f(x) \geq 0$ voor alle x .

Hierbij is $\nabla^2 f(x)$ de tweede afgeleide van f – de Hessematrix – en $\nabla^2 f(x) \geq 0$ betekent dat deze matrix overal niet-negatief definitief is.

Een matrix A heet niet-negatief definitief als voor elke vector x geldt dat $(Ax, x) \geq 0$.

Een matrix A heet positief definitief als voor elke vector $x \neq 0$ geldt dat $(Ax, x) > 0$.

Een matrix is positief (niet-negatief) definitief dan en slechts als al zijn eigenwaarden positief (niet-negatief) zijn.

Voorbeeld: Bekijk de functie $f(x_1, x_2, x_3) = \frac{x_1^2}{x_2} + x_3^2$ voor $x_2 >$

0. De eerste en tweede afgeleiden zijn:

$$\nabla f(x_1, x_2, x_3) = \begin{pmatrix} 2\frac{x_1}{x_2} \\ -\frac{x_1^2}{x_2^2} \\ 2x_3 \end{pmatrix}, \quad \nabla^2 f(x_1, x_2, x_3) = \begin{pmatrix} \frac{2}{x_2} & -\frac{2x_1}{x_2^2} & 0 \\ -\frac{2x_1}{x_2^2} & \frac{2x_1^2}{x_2^3} & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

De eigenwaarden van de Hessematrix zijn 0, 2, $2\frac{x_1^2 + x_2^2}{x_2^3}$. Deze

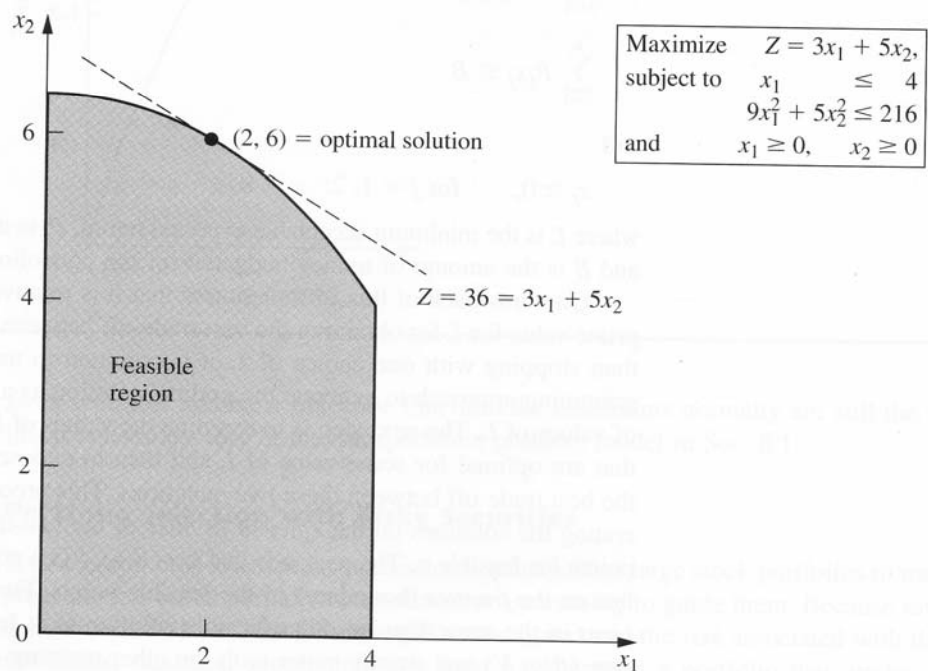
zijn allemaal niet-negatief, dus de matrix is niet-negatief definitief en de functie f is convex. Direct met de definitie van niet-negatief definitief is dit ook te zien:

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}^T \begin{pmatrix} \frac{2}{x_2} & -\frac{2x_1}{x_2^2} & 0 \\ -\frac{2x_1}{x_2^2} & \frac{2x_1^2}{x_2^3} & 0 \\ 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \frac{2}{x_2} \left(y_1 - \frac{x_1 y_2}{x_2} \right)^2 + 2y_3^2 \geq 0.$$

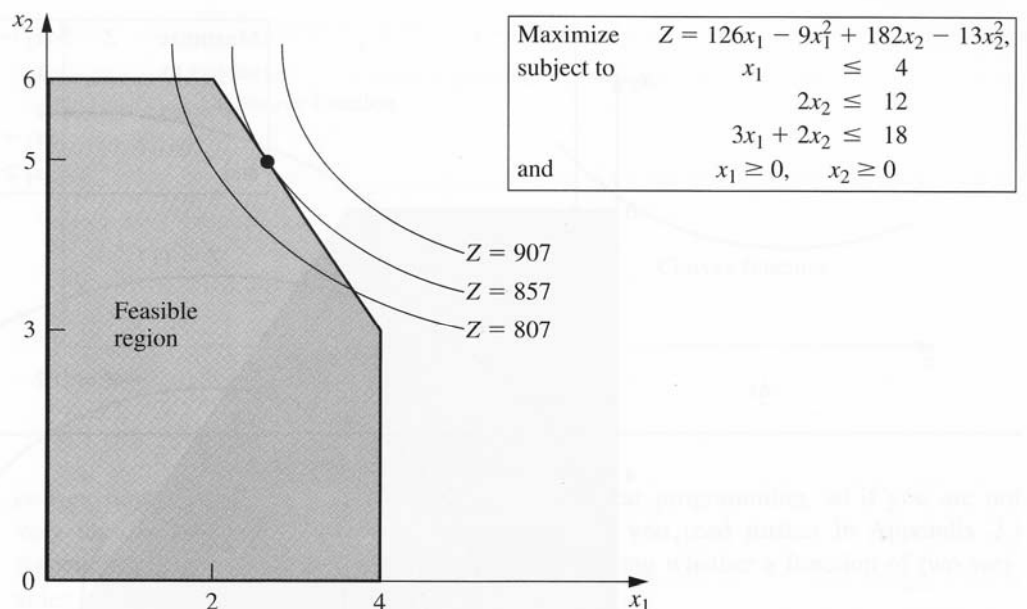
Niet-lineaire Optimalisering (NLP) (Boek H12)

■ **FIGURE 12.5**

The Wyndor Glass Co. example with the nonlinear constraint $9x_1^2 + 5x_2^2 \leq 216$ replacing the original second and third functional constraints.



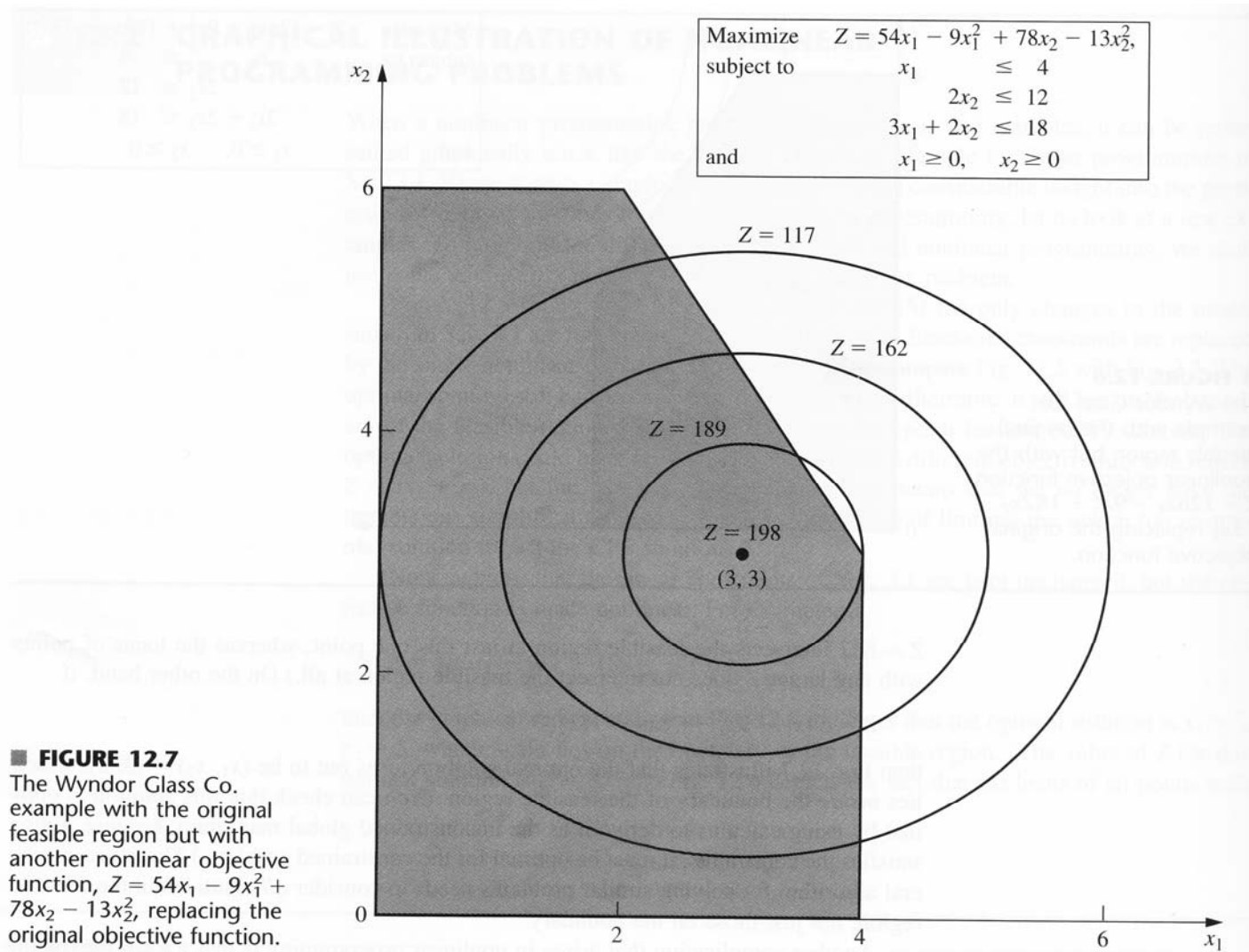
Lineaire doelfunctie, niet-lineaire constraints



■ **FIGURE 12.6**

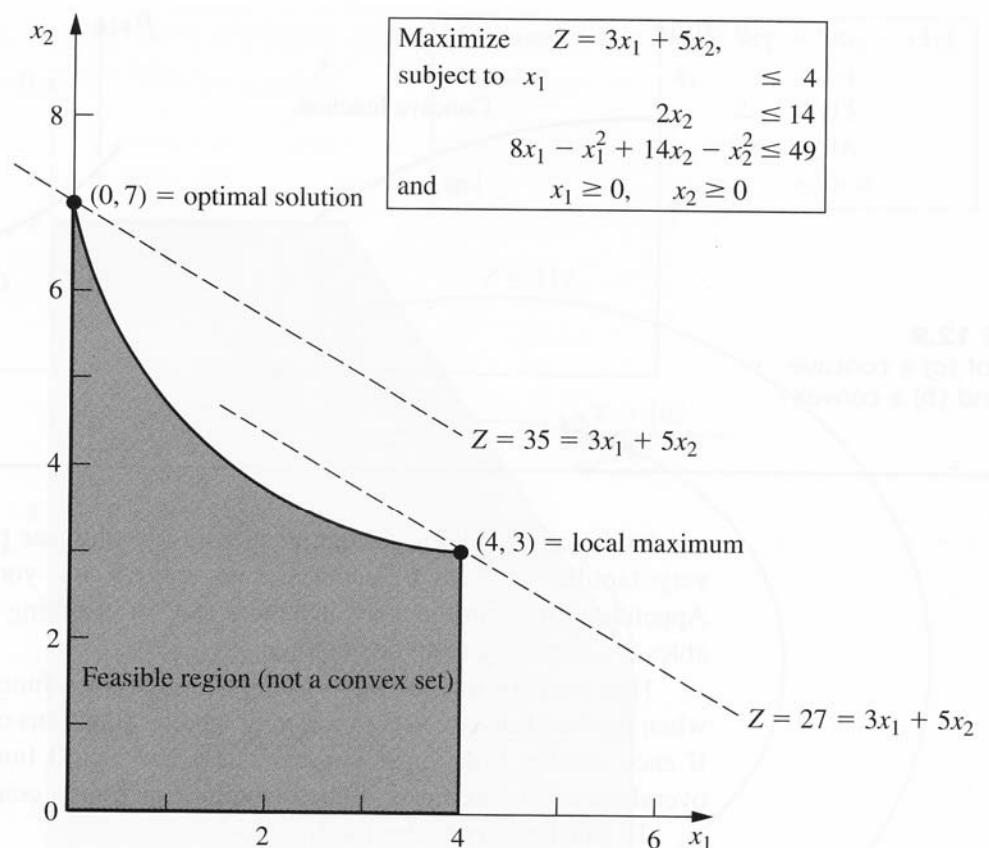
The Wyndor Glass Co. example with the original feasible region but with the nonlinear objective function $Z = 126x_1 - 9x_1^2 + 182x_2 - 13x_2^2$ replacing the original objective function.

Lineaire constraints, niet-lineaire doelfunctie



■ **FIGURE 12.7**
 The Wyndor Glass Co.
 example with the original
 feasible region but with
 another nonlinear objective
 function, $Z = 54x_1 - 9x_1^2 +$
 $78x_2 - 13x_2^2$, replacing the
 original objective function.

Lineaire constraints, niet-lineaire doelfunctie



■ **FIGURE 12.10**
 The Wyndor Glass Co.
 example with $2x_2 \leq 14$ and a
 nonlinear constraint,
 $8x_1 - x_1^2 + 14x_2 - x_2^2 \leq 49$,
 replacing the original second
 and third functional
 constraints.

Typen NLP problemen (Boek H 12.3)

- **NLP zonder nevenvoorwaarden.**

Er is alleen een niet-lineaire functie gegeven, zonder constraints, waarvoor een optimum moet worden gevonden.

Noodzakelijke voorwaarde is: $Df(x) = 0$. Dit zijn niet-lineaire vergelijkingen, meestal niet analytisch op te lossen.

- **NLP met lineaire voorwaarden.**

De doelfunctie is lineair, de constraints zijn lineair.

- **Kwadratische programmering.**

De doelfunctie is kwadratisch, de constraints zijn lineair

- **Convexe programmering.**

De doelfunctie is convex of concaaf, de constraints zijn van de vorm $g_i(x) \leq b_i$, met g_i convex (het toegelaten gebied is dan convex). Een optimum is dan absoluut

- **Separabele programmering.**

Dit is een convex programmeringsprobleem waarin de doelfunctie en de constraint functies separabel zijn, een som van functies van de afzonderlijke coördinaten.

- **Niet-convexe programmering**

Alle andere niet-lineaire programmeringsproblemen. Er kunnen veel locale optima optreden.

- Fractional Programming

Doelfunctie is van de vorm $f(x) = f_1(x)/f_2(x)$

Deze problemen kunnen soms getransformeerd worden

Voorbeeld:

$$\text{Max} \quad f(x) = \frac{c^T x + c_0}{d^T x + d_0}$$

$$\text{z.d.d.} \quad Ax \leq b$$

$$\text{en} \quad x \geq 0$$

$$\text{Noem} \quad y = \frac{x}{d^T x + d_0}, \quad t = \frac{1}{d^T x + d_0}$$

Dan wordt het probleem:

$$\text{Max} \quad Z = c^T y + c_0 t$$

$$\text{z.d.d.} \quad Ay - bt \leq 0$$

$$d^T y + d_0 t = 1$$

$$\text{en} \quad y \geq 0, t \geq 0$$

Dit is een LP probleem en kan met simplex worden opgelost.

- Complementariteitsproblemen

Vind een toegelaten oplossing van

$$w = F(z), w \geq 0, z \geq 0$$

$$w^T z = 0 \quad (\text{complementariteitsconditie})$$

Er moet dan gelden: $w_i = 0$ of $z_i = 0$

Geen doelfunctie.

Eéndimensionale zoekmethoden (Boek H 12.4)

De bisectiemethode.

Als f concaaf en differentieerbaar is met maximum in x^* dan is $f'(x) > 0$ als $x < x^*$, $f'(x) = 0$ als $x = x^*$, $f'(x) < 0$ als $x > x^*$.

Kies nauwkeurigheid $\varepsilon > 0$.

Vind ondergrens $x^- < x^*$ en bovengrens $x^+ > x^*$

1. Bereken nieuw punt $x' = (x^- + x^+)/2$

2. Als $f'(x') \geq 0$, dan $x^- := x'$

3. Als $f'(x') \leq 0$, dan $x^+ := x'$

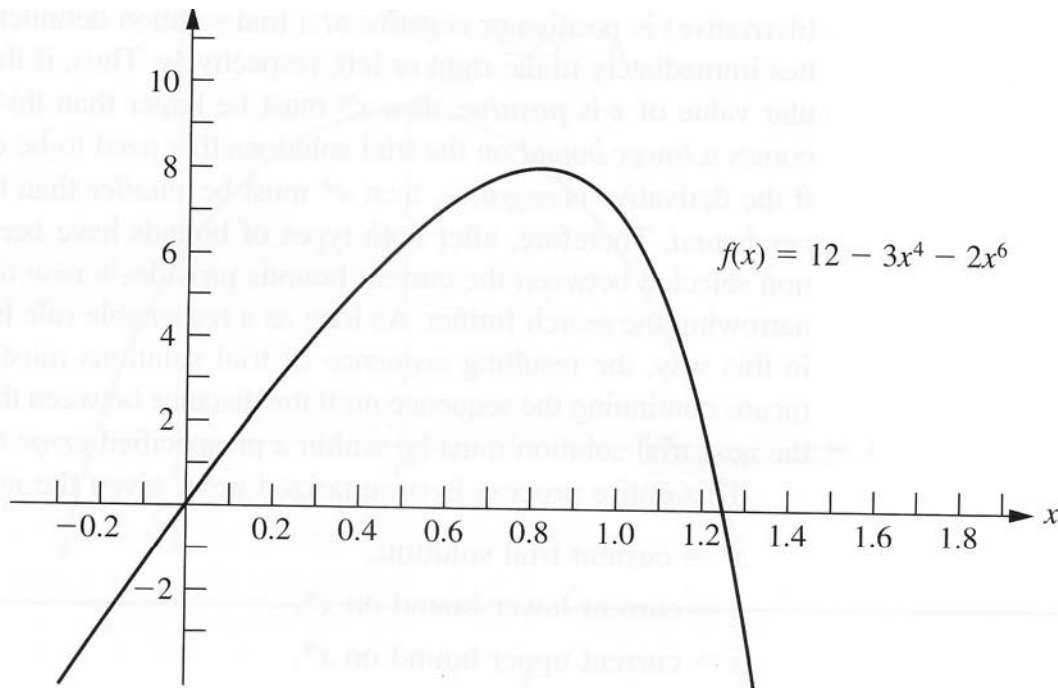
4. Als $x^+ - x^- \leq 2\varepsilon$ dan stop, anders ga naar 1

Invariant: $f'(x^-) \leq 0$ en $f'(x^+) \geq 0$,

terwijl $x^+ - x^-$ elke iteratie in lengte halveert.

Dit convergeert naar x^* .

Voorbeeld: $f(x) = 12x - 3x^4 - 2x^6$
 $f'(x) = 12(1 - x^3 - x^5)$
 $f''(x) = -12(3x^2 + 5x^4) < 0$, dus f is concaaf

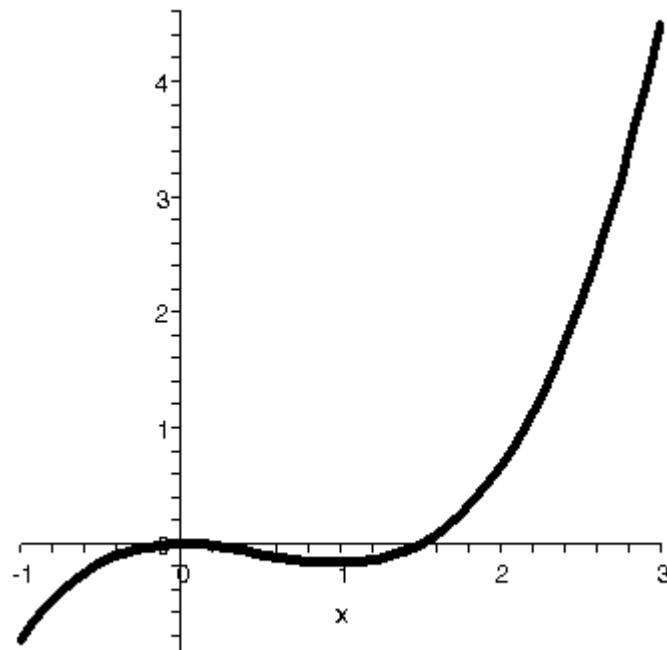


■ **TABLE 12.1** Application of the bisection method to the example

Iteration	$\frac{df(x)}{dx}$	\underline{x}	\bar{x}	New x'	$f(x')$
0		0	2	1	7.0000
1	-12	0	1	0.5	5.7812
2	+10.12	0.5	1	0.75	7.6948
3	+4.09	0.75	1	0.875	7.8439
4	-2.19	0.75	0.875	0.8125	7.8672
5	+1.31	0.8125	0.875	0.84375	7.8829
6	-0.34	0.8125	0.84375	0.828125	7.8815
7	+0.51	0.828125	0.84375	0.8359375	7.8839
Stop					

Wat als f niet concaaf?

Voorbeeld: $f(x) = x^3/3 - x^2/2$, $x^- = -1$, $x^+ = 3$



$$f'(x) = x^2 - x$$

Neem $x^- = -1$, $x^+ = 3$, dan is

$$x' = 1, \quad f'(1) = 0$$

$$x^- := x' = 1$$

$$x^+ := x' = 1$$

Het algoritme vindt een “maximum” in $x = 1$, maar dat is een minimum.

$x^- = -1$, $x^+ = 2$ convergeert naar lokaal maximum $x = 0$

$x^- = -1$, $x^+ = 4$ convergeert naar absoluut maximum $x = 4$

Methode van Newton

$$f(x_{j+1}) = f(x_j) + f'(x_j)(x_{j+1} - x_j) + f''(x_j) (x_{j+1} - x_j)^2/2 + O(|x_{j+1} - x_j|^3)$$

$$f'(x_{j+1}) = f'(x_j) + f''(x_j)(x_{j+1} - x_j) + O(|x_{j+1} - x_j|^2)$$

Probeer een stationair punt te vinden: $f'(x_{j+1}) = 0$, dus

$$f'(x_j) + f''(x_j)(x_{j+1} - x_j) = 0$$

$$x_{j+1} := x_j - f'(x_j)/f''(x_j)$$

Voorbeeld: $f(x) = 12x - 3x^4 - 2x^6$
 $f'(x) = 12(1 - x^3 - x^5)$
 $f''(x) = -12(3x^2 + 5x^4) < 0$, dus f is concaaf

■ TABLE 12.2 Application of Newton's method to the example

Iteration i	x_i	$f(x_i)$	$f'(x_i)$	$f''(x_i)$	x_{i+1}
1	1	7	-12	-96	0.875
2	0.875	7.8439	-2.1940	-62.733	0.84003
3	0.84003	7.8838	-0.1325	-55.279	0.83763
4	0.83763	7.8839	-0.0006	-54.790	0.83762

Orde van convergentie

Hoe hard convergeert een rij?

Lineaire convergentie: De fout wordt na elke iteratie ongeveer met een vaste factor kleiner:

$$|\alpha_{k+1} - \alpha| \approx \beta |\alpha_k - \alpha|$$

voor een zekere constante $\beta \leq 1$.

Kwadratische convergentie:

$$|\alpha_{k+1} - \alpha| \approx \beta |\alpha_k - \alpha|^2$$

voor een zekere constante $\beta \leq 1$.

Een rij (α_k) convergeert met orde p naar een limiet α als

$$p = \sup \left\{ q \mid \limsup_{k \rightarrow \infty} \frac{|\alpha_{k+1} - \alpha|}{|\alpha_k - \alpha|^q} \right\}.$$

De convergentiefactor bij deze orde is

$$\beta = \limsup_{k \rightarrow \infty} \frac{|\alpha_{k+1} - \alpha|}{|\alpha_k - \alpha|^p}.$$

Als $p = 1$ heet de convergentie lineair.

Als ook $\beta = 0$, dan heet de convergentie superlineair (de fout wordt telkens verkleind met een factor die zelf steeds kleiner wordt).

Als daarentegen $\beta = 1$, dan heet de convergentie sublineair (de fout wordt telkens verkleind met een factor die steeds dichterbij 1 komt, de convergentie gaat dus steeds trager).

Voorbeeld: De meetkundige rij $\alpha_k = a^k$ voor $0 < a < 1$.

Deze rij convergeert naar $\alpha = 0$, dus:

$$\frac{|\alpha_{k+1} - \alpha|}{|\alpha_k - \alpha|^p} = \frac{a^{k+1}}{a^{kp}} = a^{k(1-p)+1} = a(a^{1-p})^k.$$

Als $p > 1$, dan is $a^{1-p} > 1$, dus de rij is niet begrensd als $k \rightarrow \infty$.

Als $p = 1$, dan is $a^{1-p} = 1$, dus de rij is constant a .

Als $p < 1$, dan is $a^{1-p} < 1$, dus de rij nadert naar 0.

De orde van convergentie is dus $p = 1$, en de convergentiefactor is $\beta = a$, dus lineaire convergentie.

Voorbeeld: De rij $\alpha_k = a^{(2^k)}$ voor $0 < a < 1$.

Deze rij convergeert ook naar $\alpha = 0$, maar sneller dan de meetkundige. We laten zien dat deze rij kwadratisch convergeert door $p = 2$ te nemen:

$$\frac{|\alpha_{k+1} - \alpha|}{|\alpha_k - \alpha|^2} = \frac{a^{(2^{k+1})}}{(a^{(2^k)})^2} = \frac{(a^{(2^k)})^2}{(a^{(2^k)})^2} = 1.$$

De orde van convergentie is dus $p = 2$, en de convergentiefactor is $\beta = 1$, dus kwadratische convergentie.

Voorbeeld: De rij $\alpha_k = 1/k$ convergeert naar $\alpha = 0$. Nu geldt:

$$\frac{|\alpha_{k+1} - \alpha|}{|\alpha_k - \alpha|} = \frac{k}{k+1}.$$

De limiet van deze getallen is 1, de orde van convergentie is dus $p = 1$, en de convergentiefactor is $\beta = 1$, dus we hebben te maken met (trage) sublineaire convergentie.

Voorbeeld: Bekijk de rij $\alpha_k = k^{-k}$. Nu is:

$$\frac{|\alpha_{k+1} - \alpha|}{|\alpha_k - \alpha|^p} = \frac{k^{pk}}{(k+1)^{k+1}} = \frac{k^{(p-1)k}}{(k+1)\left(1 + \frac{1}{k}\right)^k} = \frac{(k^k)^{p-1}}{(k+1)\left(1 + \frac{1}{k}\right)^k}.$$

Er geldt dat $\left(1 + \frac{1}{k}\right)^k \rightarrow e$.

Als $p > 1$, dan is de bovenstaande rij niet begrensd als $k \rightarrow \infty$.

Als $p = 1$, nadert de bovenstaande rij naar 0.

Als $p < 1$, nadert de bovenstaande rij naar 0.

De orde van convergentie is dus $p = 1$, en de convergentiefactor is $\beta = 0$, dus superlineaire convergentie.

Voorbeeld:

Bisectiemethode: lineaire convergentie met factor $1/2$

Newtonmethode: kwadratische convergentie met factor $f'''(x^*)/(2f''(x^*))$

$$\frac{x_{k+1} - x^*}{|x_k - x^*|^2} = \frac{1}{|x_k - x^*|^2} \left(x_k - \frac{f'(x_k)}{f''(x_k)} - x^* \right) =$$

$$\frac{1}{|x_k - x^*|^2} \left(x_k - x^* - \frac{(x_k - x^*)f''(x^*) + (x_k - x^*)^2 f'''(x^*) + O((x_k - x^*)^3)}{f''(x^*) + (x_k - x^*)f'''(x^*) + O((x_k - x^*)^2)} \right) =$$

$$\frac{f'''(x^*)/2 + O((x_k - x^*))}{f''(x^*) + O((x_k - x^*))} \approx \frac{f'''(x^*)}{2f''(x^*)}$$

Uniform zoeken (Niet in het boek)

Voor de bisectiemethode is in elke iteratie de evaluatie van een afgeleide nodig.

Voor de Newtonmethode is in elke iteratie de evaluatie van een afgeleide en een tweede afgeleide nodig.

Wat als er geen afgeleiden beschikbaar zijn, bijvoorbeeld omdat functie-evaluaties uit computersimulaties komen.

Uniform zoeken: Verdeel in het interval $[a,b]$ uniform n nieuwe punten:

$$x_k = a + k \frac{b-a}{n+1}, \quad k = 1, \dots, n.$$

Reken de functiewaarde uit in elk punt en zoek de kleinste. Als die wordt aangenomen in x_k vervang je het interval $[a, b]$ door $[x_{k-1}, x_{k+1}]$. Het nieuwe interval heeft als lengte

$$\frac{2}{n+1}(b-a).$$

n functie-evaluaties leveren dus een reductie van $2/(n+1)$. Per functie-evaluatie reduceert de lengte dus (worst case) met een factor

$$\left(\frac{2}{n+1} \right)^{\frac{1}{n}}.$$

Deze factor is minimaal is als $n = 3$. De reductiefactor per evaluatie is dan

$$2^{-1/3} = 0.793700\dots$$

Convergentie is dus lineair met factor 0.793700...

Gulden snede zoekmethode (Niet in het boek)

De gulden snede zoekmethode (*golden section search*) is efficiënter dan uniform zoeken.

Het idee is dat je in het interval $[a,b]$ een extra punt c hebt waarvan de functiewaarde bekend is, en dat $f(a) \geq f(c) \leq f(b)$. Vervolgens wordt er een nieuw punt d gekozen. We nemen aan dat dit tussen a en c ligt, maar tussen c en b gaat het analoog.

Als $f(d) \geq f(c)$, dan is $f(d) \geq f(c) \leq f(b)$.

Als $f(d) \leq f(c)$, dan is $f(a) \geq f(d) \leq f(c)$.

Het interval $[a,b]$ met tussenpunt c kan worden vervangen door $[d,b]$ met tussenpunt c , of door $[a,c]$ met tussenpunt d , terwijl telkens de waarde in het tussenpunt het kleinst is.

Kies nu de posities van c en d zo dat de lengtes van de nieuwe intervallen gelijk zijn, dan is de reductie telkens gelijk.

Kies d op verhouding λ en c op verhouding $1-\lambda$ op het interval $[a,b]$, ofwel:

$$d = a + \lambda(b-a), \quad c = a + (1-\lambda)(b-a).$$

De slimme truc van de methode bestaat er nu in dat je λ zo kiest dat je na reductie het middelste punt weer kunt gebruiken.

Het punt op verhouding λ kan op verhouding $(1-\lambda)$ liggen in het nieuwe interval. Omdat het nieuwe interval een factor $(1-\lambda)$ kleiner is geworden is de nieuwe factor van het tussenpunt nu $\lambda/(1-\lambda)$. Als deze verhouding gelijk is aan $1-\lambda$, dan is het punt weer te gebruiken.

Dit is het geval als $\lambda^2 - \lambda - 1 = 0$, dus als

$$\lambda = \frac{3 - \sqrt{5}}{2} = 0,381966..., \quad 1 - \lambda = \frac{\sqrt{5} - 1}{2} = 0,618033...$$

Dit is de gulden snede verhouding. Met één functie-evaluatie wordt het interval dus verkleind met een factor $1 - \lambda = 0,618033$, dat is beter dan bij de bovenstaande methode.

Meerdimensionale zoekmethoden (Boek H 12.5)

Steepest descent.

Het idee van deze methode is dat je vanuit een startpunt de steilste richting opzoekt en in deze richting een optimum zoekt. Vanuit dat punt kun je weer verder zoeken.

De steilste richting wordt gegeven door de gradiënt:

$$\nabla f(x) = f'(x)^T,$$

want $f(x+h) = f(x) + \nabla f(x)h + O(|h|^2)$

Kies een startpunt x_0

1. Vind de waarde t^* waarvoor $t \rightarrow f(x_k + tf'(x_k))$ maximaal is
2. $x_{k+1} := x_k + t^*f'(x_k)$
3. Als stopcriterium voldaan, stop, anders $k:=k+1$, ga naar 1

Voorbeeld:

$$f(x_1, x_2) = 2x_1x_2 + 2x_2 - x_1^2 - 2x_2^2$$

$$f'(x_1, x_2) = (2x_2 - 2x_1 \quad 2x_1 + 2 - 4x_2)$$

Startpunt $x_0 = (0,0)$

Iteratie 1: $f'(0,0) = (0,2)$

Vind het maximum van $f((0,0) + t(0,2)) = f(0, 2t) = 4t - 8t^2$

$$t^* = 1/4$$

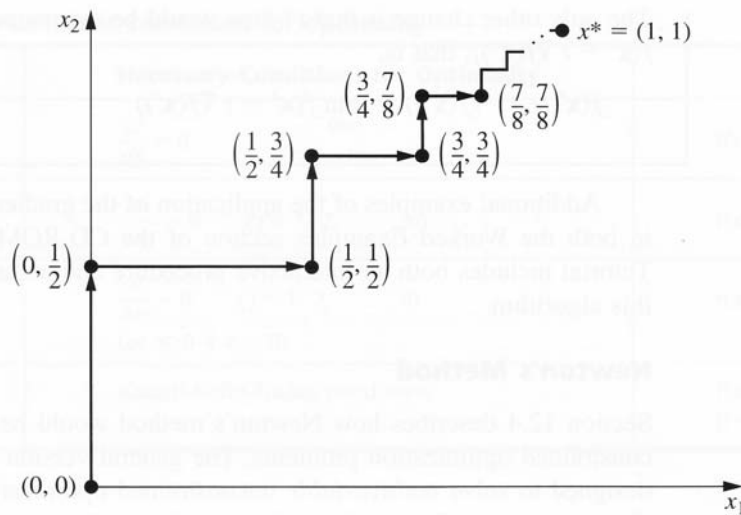
$$x_1 = (0,0) + 1/4(0,2) = (0,1/2)$$

Iteratie 2: $f'(0,1/2) = (1 \quad 0)$

Vind het maximum van $f((0,1/2) + t(1,0)) = f(t, 1/2) = 1/2 + t - t^2$

$$t^* = 1/2$$

$$x_1 = (0,1/2) + 1/2(1,0) = (1/2,1/2)$$



■ **FIGURE 12.14**

Illustration of the gradient search procedure when $f(x_1, x_2) = 2x_1x_2 + 2x_2 - x_1^2 - 2x_2^2$.

■ **TABLE 12.3** Application of the gradient search procedure to the example

Iteration	x'	$\nabla f(x')$	$x' + t \nabla f(x')$	$f(x' + t \nabla f(x'))$	t^*	$x' + t^* \nabla f(x')$
1	$(0, 0)$	$(0, 2)$	$(0, 2t)$	$4t - 8t^2$	$\frac{1}{4}$	$(0, \frac{1}{2})$
2	$(0, \frac{1}{2})$	$(1, 0)$	$(t, \frac{1}{2})$	$t - t^2 + \frac{1}{2}$	$\frac{1}{2}$	$(\frac{1}{2}, \frac{1}{2})$

Methode van Newton:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - (\nabla^2 f(\mathbf{x}_k))^{-1} \nabla f(\mathbf{x}_k)$$

Voorbeeld:

$$f(\mathbf{x}_1, \mathbf{x}_2) = 2\mathbf{x}_1\mathbf{x}_2 + 2\mathbf{x}_2 - \mathbf{x}_1^2 - 2\mathbf{x}_2^2$$

$$f'(\mathbf{x}_1, \mathbf{x}_2) = (2\mathbf{x}_2 - 2\mathbf{x}_1 \quad 2\mathbf{x}_1 + 2 - 4\mathbf{x}_2)$$

$$\nabla^2 f(x_1, x_2) = \begin{pmatrix} -2 & 2 \\ 2 & -4 \end{pmatrix}$$

Startpunt $\mathbf{x}_0 = (0,0)$

$$\mathbf{x}_1 = -\begin{pmatrix} -2 & 2 \\ 2 & -4 \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ 2 \end{pmatrix} = -\begin{pmatrix} -1 & -\frac{1}{2} \\ -\frac{1}{2} & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} 0 \\ 2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

De Newtonmethode is met één iteratie klaar (want een kwadratische functie!)

Nadeel van Newton: Per iteratie een afgeleide en een tweede afgeleide $\approx n + n^2$ functie-evaluaties.

Andere aanpak: werk met een goedkoper te berekenen inverse van tweede afgeleide: Quasi-Newtonmethoden.