

# Optimierverfahren für $n=1$

1. simultane Methoden
  - 1.1. Raster
  - 1.2. Monte-Carlo
2. sequentielle Methoden
  - 2.1. Intervallteilung
    - 2.1.0. Halbierungsverfahren
    - 2.1.1. Fibonacci
    - 2.1.2. goldener Schnitt
  - Einschachtelung
  - 2.2. Interpolation
    - 2.2.1. Regula falsi / Sekanten meth.
    - 2.2.2. Newton-Raphson
    - 2.2.3. Lagrange
    - 2.2.4. Hermite

keine Polstellen  
keine Singularitäten

Unimodalität

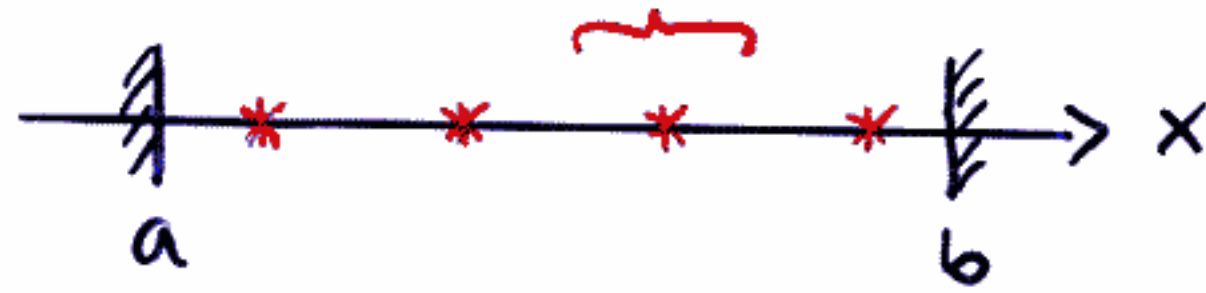
stetig, stetig diff.

Q1

Hinweis auf Bücher

Optimierverfahren für  $n=1$ 1. simultane Methoden (Intervall  $a \leq x \leq b$  geg.)

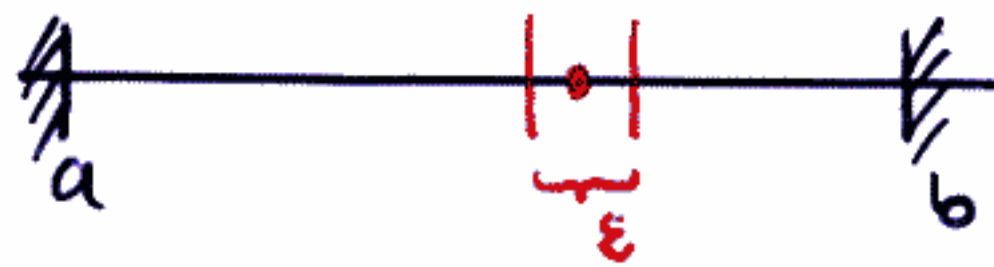
1.1. Gitter- bzw. Raster-Methode

4 Proben äquidistant: Restintervall  $\frac{1}{4}(b-a)$  $N$   $\frac{1}{N}(b-a)$ ↪ bei gegebener Genauigkeitsforderung  $\varepsilon$ 

$$N \leq \frac{b-a}{\varepsilon} + 1; N \text{ ganz}$$

1.2. Monte-Carlo-Methode

im Intervall gleichverteilte Zufallsproben

1 Versuch  $p_1 = \frac{\varepsilon}{b-a}$  Trefferwahrsch.

$$\bar{p}_1 = 1 - \frac{\varepsilon}{b-a}$$

N Versuche  $\bar{p}_N = \left(1 - \frac{\varepsilon}{b-a}\right)^N$ 

$$p_N = 1 - \left(1 - \frac{\varepsilon}{b-a}\right)^N$$

d.h. mit dieser Wahrsch. mind. 1 Versuch im  $\varepsilon$ -Intervall

$$N = \frac{\ln(1 - p_N)}{\ln\left(1 - \frac{\varepsilon}{b-a}\right)} \approx -\frac{b-a}{\varepsilon} \ln(1 - p_N)$$

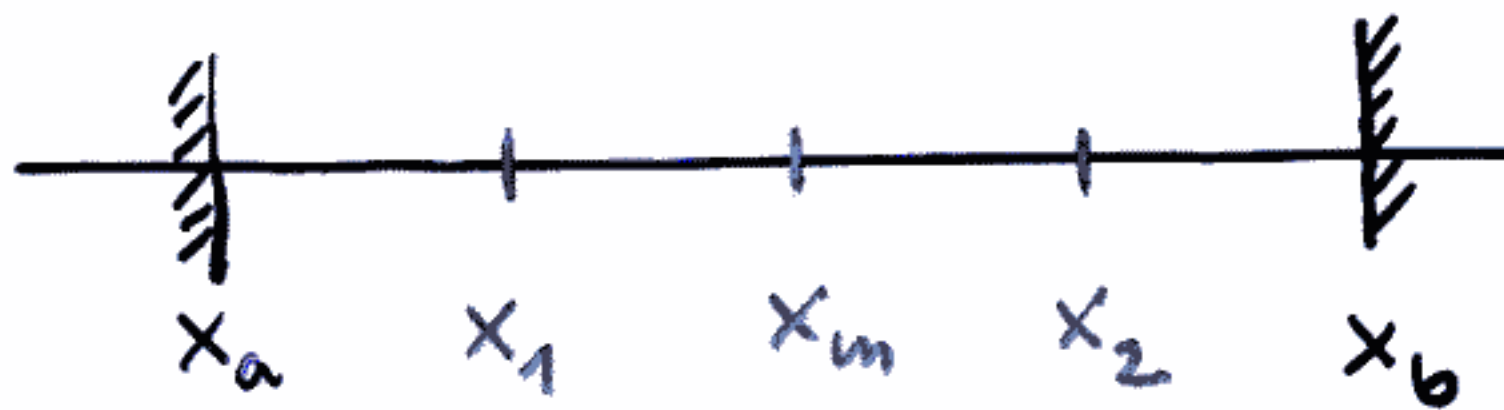
↑ für  $\varepsilon \ll (b-a)$ :  $\ln(1-x) \approx -x$

$$p_N = 0.63 \quad N = 1 \cdot \frac{b-a}{\varepsilon}$$

$$p_N = 0.90 \quad N = 2.3 \cdot \frac{b-a}{\varepsilon}$$

2.1 sequentielle Methoden für  $n=1$ 

## 2.1.0 Intervall-Halbierungsmethode



## Schritt

- 1 Setze  $x_m = \frac{1}{2}(x_a + x_b)$ ;  $l = x_b - x_a$ ; berechne  $f(x_m)$
- 2 Setze  $x_1 = x_a + \frac{l}{4}$  und  $x_2 = x_b - \frac{l}{4}$   
und berechne  $f(x_1)$  und  $f(x_2)$
- 3 Vergleiche  $f(x_1)$  mit  $f(x_m)$   
Wenn  $f(x_1) < f(x_m)$ , eliminiere  $(x_m, x_b)$   
und setze  $x'_b = x_m$ ;  $x'_m = x_1$   
und gehe zu Schritt 5.
- 4 Vergleiche  $f(x_2)$  mit  $f(x_m)$   
Wenn  $f(x_2) < f(x_m)$ , eliminiere  $(x_a, x_m)$   
und setze  $x'_a = x_m$ ;  $x'_m = x_2$   
und gehe zu Schritt 5,  
sonst eliminiere  $(x_a, x_1)$  und  $(x_2, x_b)$   
und setze  $x'_a = x_1$ ,  $x'_b = x_2$ ,  $x'_m = x_m$ .
- 5 Falls  $l' = x'_b - x'_a < \varepsilon$ , beende die Suche,  
sonst gehe zu Schritt 2 unter Verwendung  
der  $x'$  Größen als neue  $x$ .

Analyse:

Nutzen: in jedem Fall wird das Intervall halbiert

$$l' = \frac{1}{2} l$$

Kosten: außer im 1. Schritt werden pro Iteration

$N = 2$  ZFA (Zielfunktionsauswertungen) benötigt

Intervallreduktion pro ZFA:  $\left(\frac{1}{2}\right)^{N/2}$

$$\left(\frac{1}{2}\right)^{N/2} \stackrel{!}{=} \frac{l^*}{l}$$

Bei vorgegebener Genauigkeit  $\varepsilon$  werden

also etwa

$$N = \frac{2}{\log 2} \log \left( \frac{x_b - x_a}{\varepsilon} \right)$$

$N$  aufrunden  
 $N$  gerade

Auswertungen benötigt.

↑ äquidistante 3-Punkt-Suche

Voraussetzungen: Unimodalität

nicht: Stetigkeit

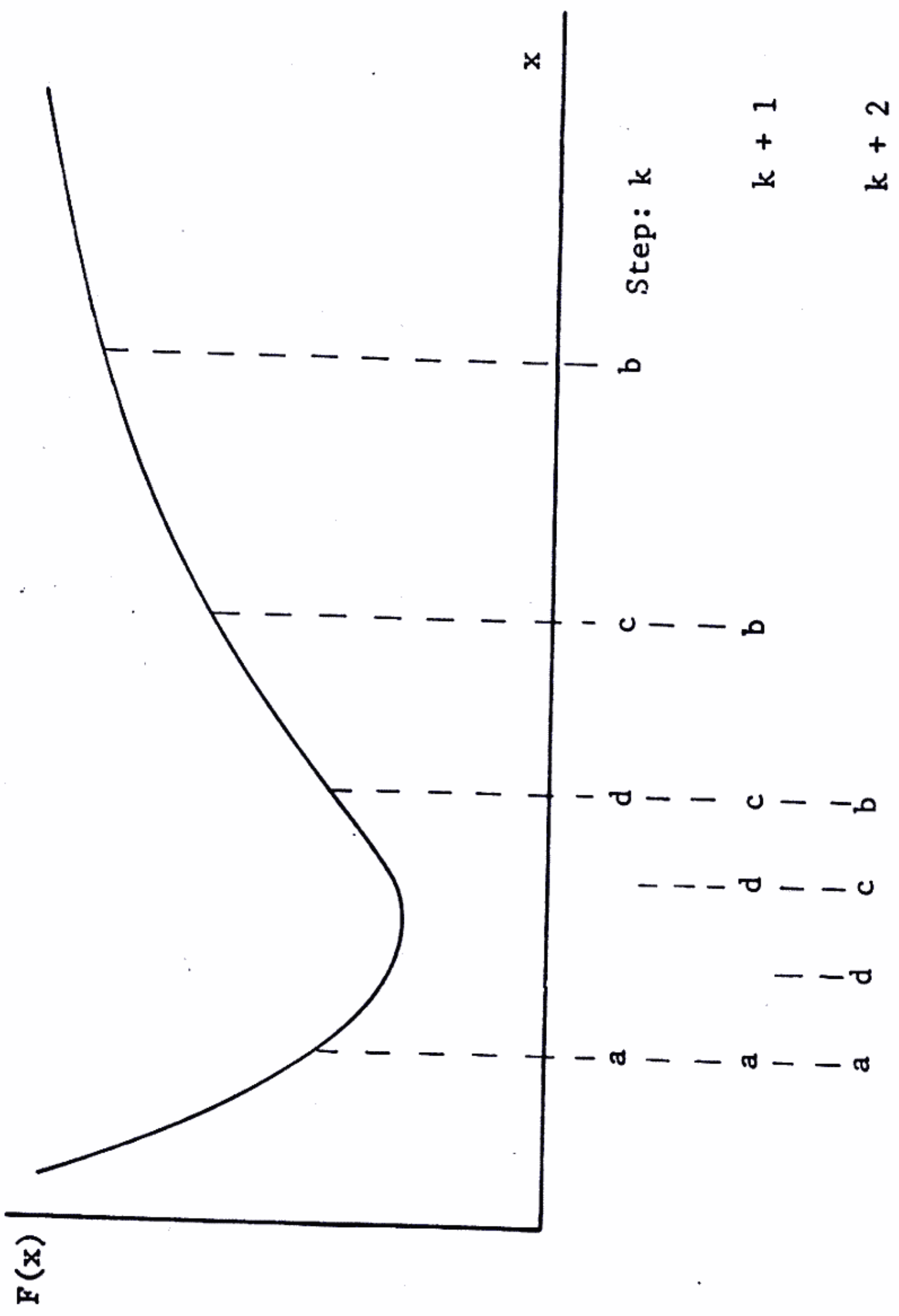
Differenzierbarkeit

Kiefer, 1957: Unter allen sequentiellen äquidistanten

$n$ -Punkt Intervallteilungsverfahren ist Strategie

mit  $n=3$  optimal

# Interval division in the Fibonacci search



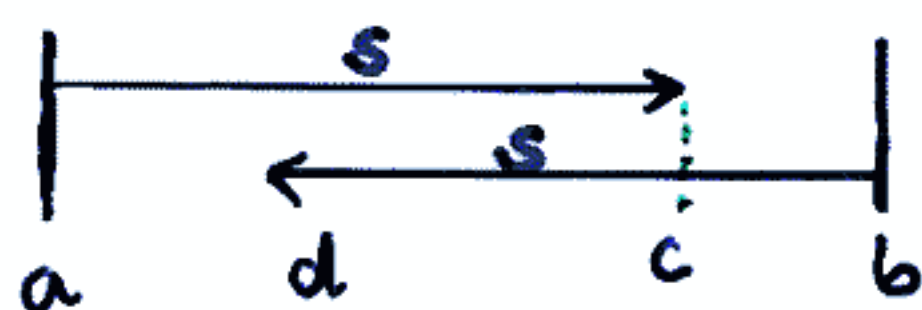
## 2. sequentielle Verfahren

## 2.1 Intervallteilungsverfahren

z. B. Fibonacci - Methode

$$f_N = f_{N-1} + f_{N-2} \quad ; \quad f_0 = f_1 = 1$$

$$\rightarrow 1, 1, 2, 3, 5, 8, 13, 21, \dots$$



$$c^{(k)} = a^{(k)} + s^{(k)}$$

$$d^{(k)} = b^{(k)} - s^{(k)}$$

$$s^{(k)} = t^{(k)} \cdot (b^{(k)} - a^{(k)}) = b^{(k+1)} - a^{(k+1)}$$

$$t^{(k)} = \frac{f_{N-k-1}}{f_{N-k}}$$

$N$  = Zahl der Int.teilungen  
muß zuvor festgelegt werden

$$\text{wenn } F(d^{(k)}) < F(c^{(k)}) \quad : \quad \begin{aligned} a^{(k+1)} &= a^{(k)} \\ b^{(k+1)} &= c^{(k)} \end{aligned}$$

$$> \quad : \quad \begin{aligned} a^{(k+1)} &= d^{(k)} \\ b^{(k+1)} &= b^{(k)} \end{aligned}$$

$$N \sim \log \frac{b^{(0)} - a^{(0)}}{\varepsilon}$$

$$\text{oft vorzeitig } F(d^{(k)}) = F(c^{(k)})$$

$$\text{dann besser: } t^{(k)} = \frac{2}{1 + \sqrt{5}} \approx 0.618$$

Teilung nach dem goldenen Schnitt

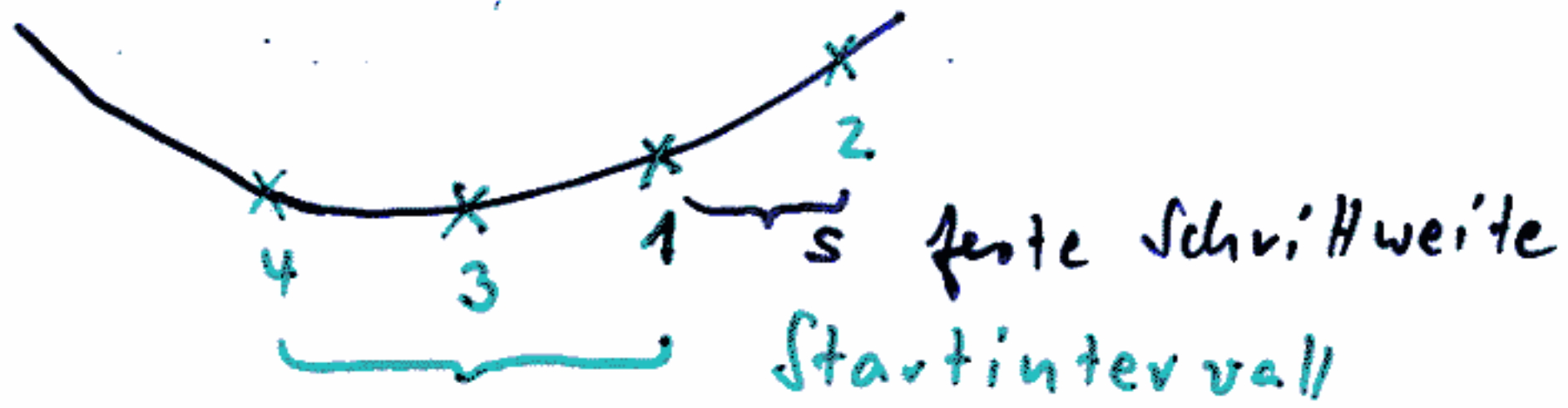
$$[\text{aus } t^2 + t = 1]$$

max. 17% Versuche mehr; oft weniger

Aufangsintervall muß vorgegeben werden (Einschachtelungsverfahren)  
Unimodalität wird vorausgesetzt  
(nicht: Stetigkeit, Differenzierbarkeit)

nur 1 neue Stützstelle  
pro Iteration

## Einschachtelungsverfahren (für Start-Intervall)



evtl. Schrittweite verdoppeln, wenn es 'abwärts' geht  
nach Richtungsumkehr halbieren

## 2.2 Interpolationsverfahren

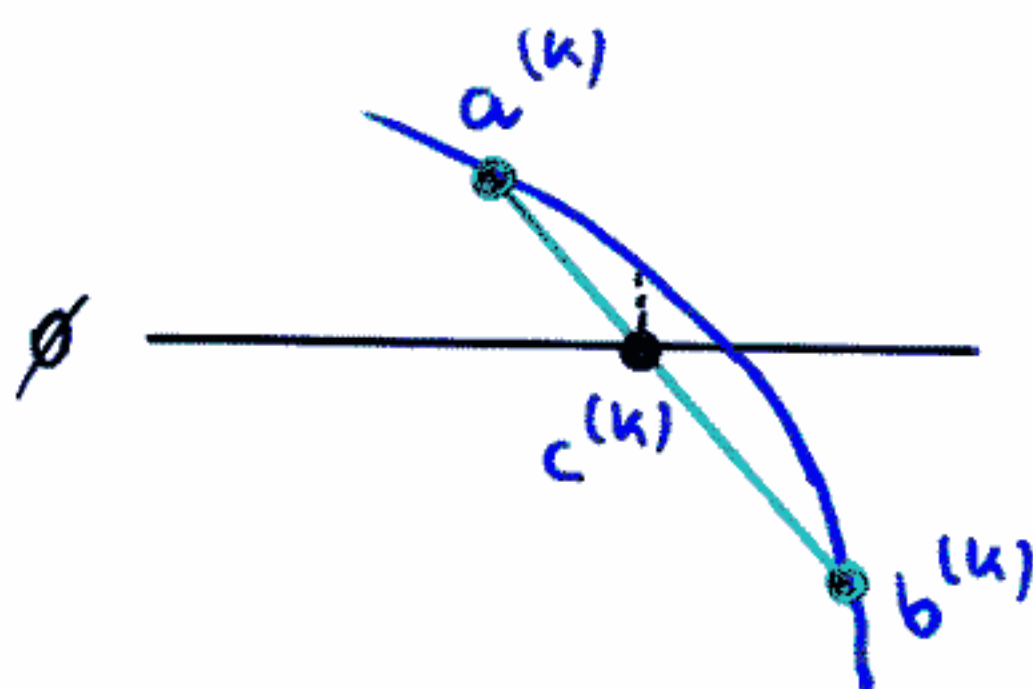
benötigen kein Startintervall

Konvergenz meist nur dann gesichert, wenn  
Start nahe Minimum

Verfahren wie bei Nullstellen - Bestimmung

## 2.2.1 Regula falsi

a. für Nullstelle



2 Stützpunkte

2 Funktionswerte

$a^{(k)}, b^{(k)}$

$F(a^{(k)}), F(b^{(k)})$

$$c^{(k)} = a^{(k)} - F(a^{(k)}) \frac{b^{(k)} - a^{(k)}}{F(b^{(k)}) - F(a^{(k)})}$$

$c^{(k)}$  ersetzt  $a^{(k)}$  oder  $b^{(k)}$  so daß

$$\text{sign}(F(c^{(k)})) \neq \text{sign} \begin{cases} F(a^{(k)}) \\ F(b^{(k)}) \end{cases}$$

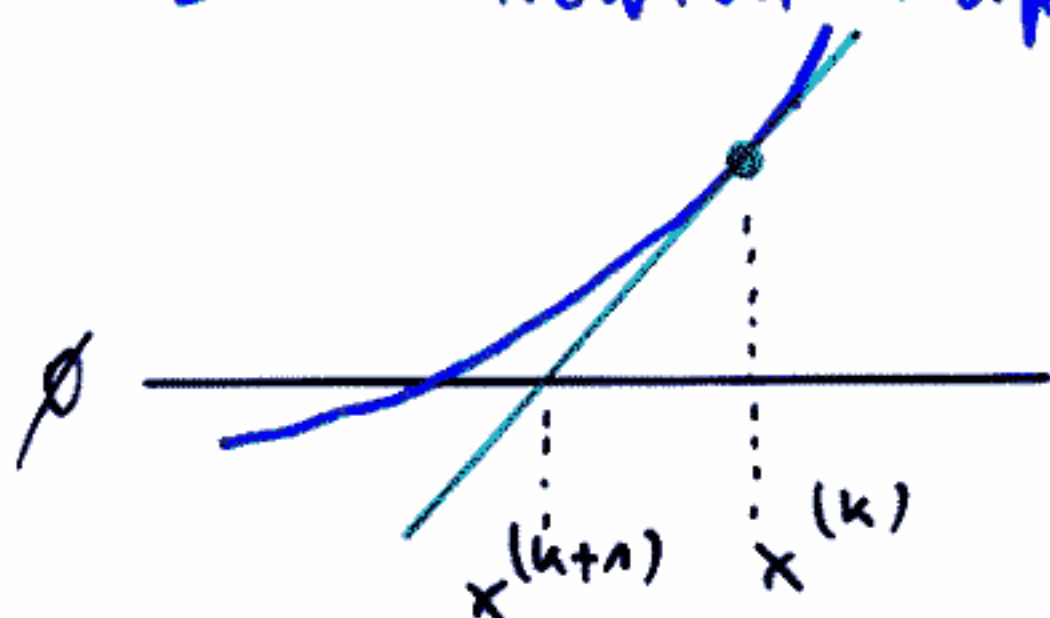
Wenn  $\text{sign}(F(a^{(k)})) = \text{sign}(F(b^{(k)}))$ : Sekantenmeth.

Abbruch, wenn  
 $F(x^{(k)}) = 0$  bzw.  
 $|F(x^{(k)})| \leq \epsilon > 0$

b. Regula falsi für Max. bzw. Min. suche:

$$c^{(k)} = a^{(k)} - \frac{F_x(a^{(k)})}{\frac{F_x(b^{(k)}) - F_x(a^{(k)})}{b^{(k)} - a^{(k)}}}$$

### 2.2.2 Newton-Raphson



- 1 Stützstelle
- 1 Funktionswert
- 1 erste Ableitung

$$x^{(k+1)} = x^{(k)} - \frac{F(x^{(k)})}{F_x(x^{(k)})} \quad \text{für Nullstelle}$$

$$x^{(k+1)} = x^{(k)} - \frac{F_x(x^{(k)})}{F_{xx}(x^{(k)})} \quad \text{für Extremum}$$

Wenn  $F(x)$  quadratisch, dann Lösung in einem Schritt

z. B.  $F(x) = (x-3)^2 = x^2 - 6x + 9$  / Max. Min.  
 $F_x(x) = 2x - 6$   
 $F_{xx}(x) = 2$

$$x_0 = 1 \quad x_1 = 1 - \frac{-4}{2} = 1 + 2 = 3$$

z. B.  $F(z) = z^3 - 1$  / Nullst.  
 $F_z(z) = 3z^2$

$$z_{k+1} = z_k - \frac{z_k^3 - 1}{3z_k^2} = \frac{2}{3}z_k + \frac{1}{3z_k^2}$$

drei Nullstellen mit unterschiedl. Einzugsbereichen  
 Grenzen der Einzugsbereiche fraktal!



Example

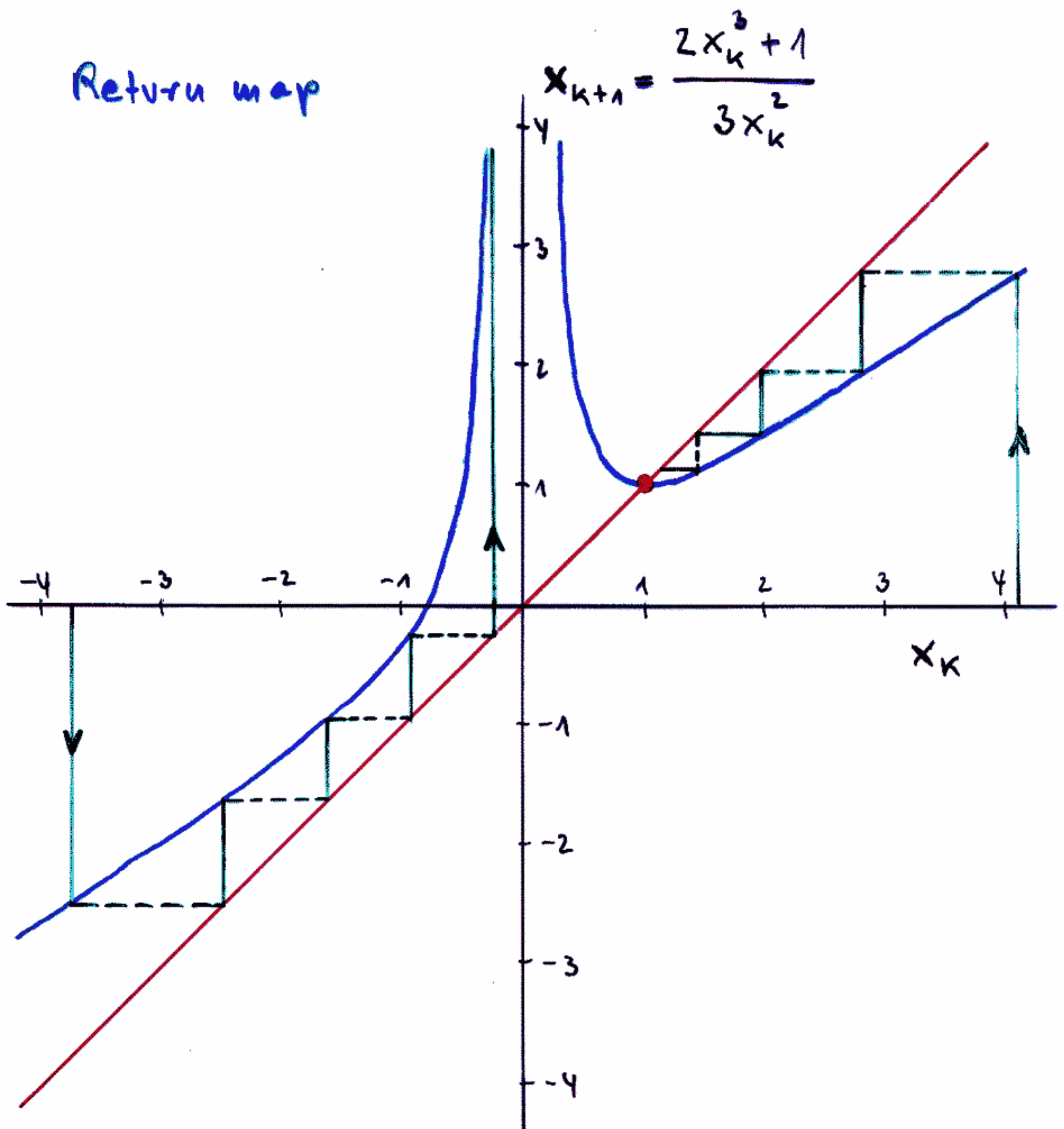
$$q(x) = \frac{1}{4}x^4 - x \rightarrow \min$$

$$p(x) = q'(x) = x^3 - 1 \rightarrow 0 \quad \text{necessary condition}$$

$$\text{Newton-Iteration: } x_{k+1} = x_k - \frac{p(x)}{p'(x)}$$

$$x_{k+1} = x_k - \frac{q'(x)}{q''(x)}$$

Return map



Newton-Raphson

- 101 -

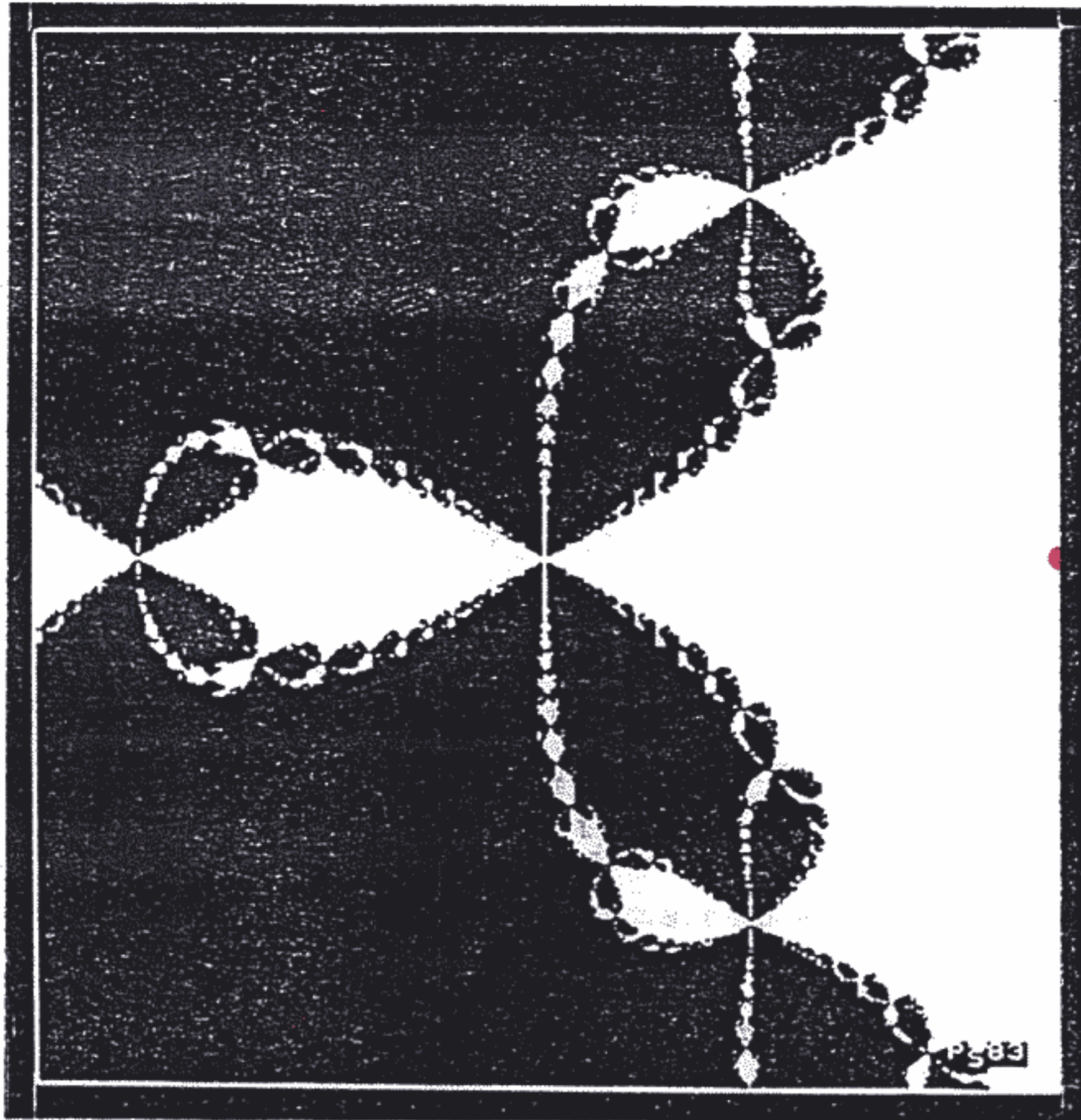
für  $z^3 = 1$ 

Figure 1: Spatial pattern of the town RGB. The customers of shop R are shown in white

### Cayley's Problem

In 1879 A.Cayley /Ca1/ suggests the extension of what he calls the Newton-Fourier Method

$$N(z_k) := z_{k+1} = z_k - p(z_k)/p'(z_k) \quad (1)$$

to complex roots of a polynomial  $p$ : "... In connexion herewith, throwing aside the restrictions as to reality, we have what I call the Newton-Fourier Imaginary Problem...". Furthermore, he suggests to study the problem globally: "... The problem is to determine the regions of the plane, such that  $P$  (initial point) being taken at pleasure anywhere within one region we arrive ultimately at the point  $A$  (a root of the problem) ...". In two notes published in 1879 /Ca2/ and 1890 /Ca3/ he takes up the problem

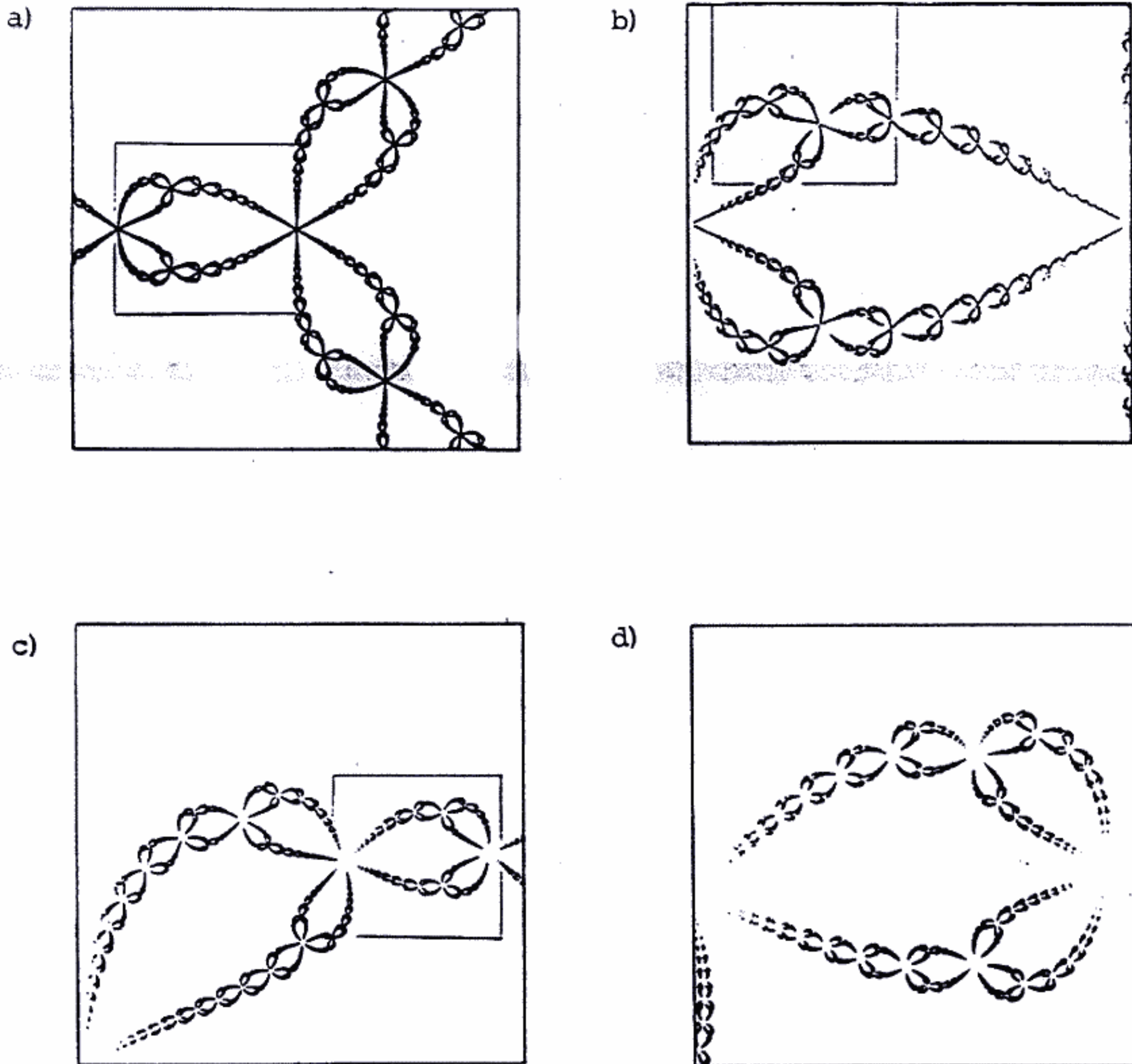
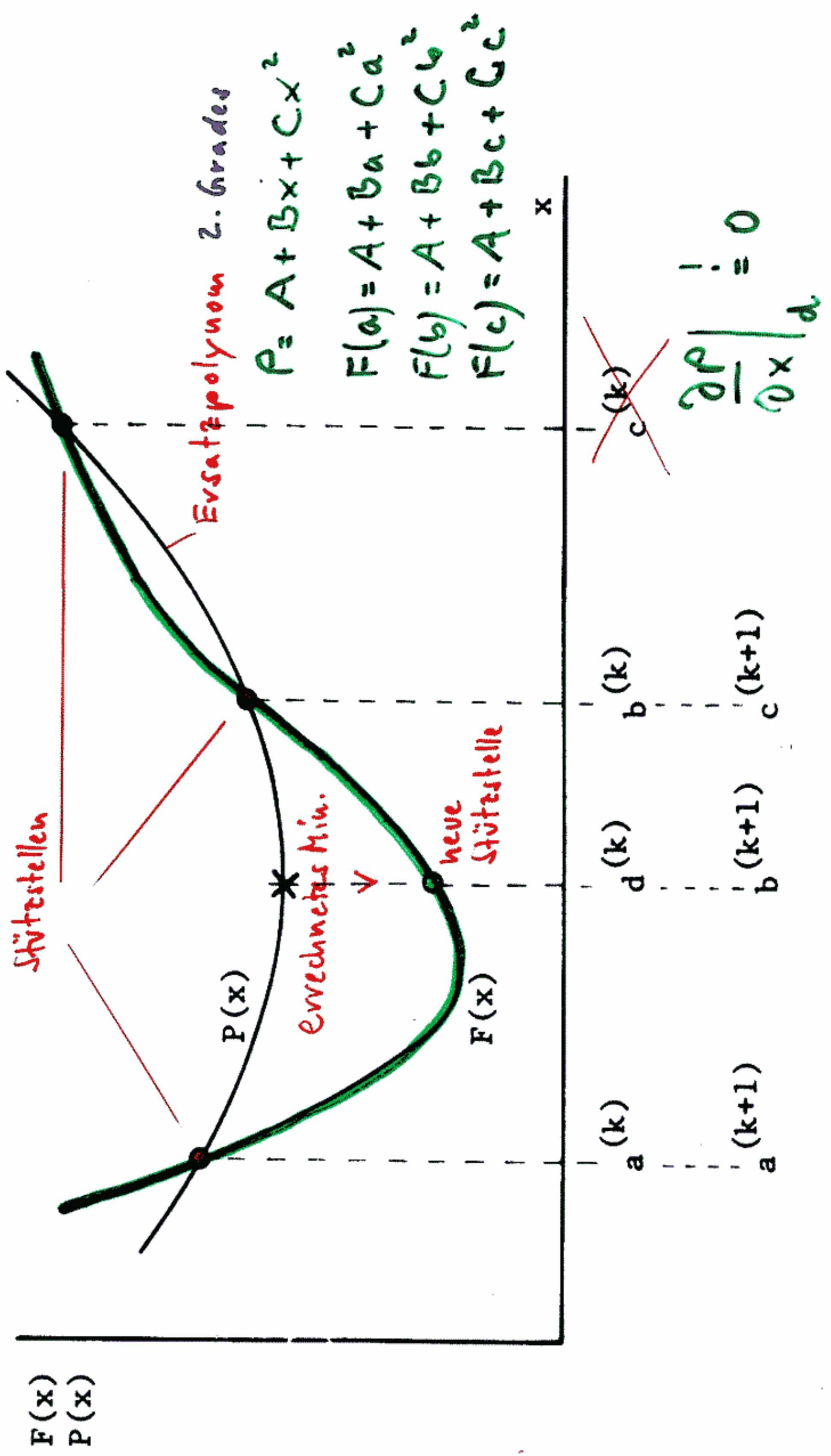


Figure 7a: Four views of sections of the Julia set of Newton's Method for  $z^3 - 1 = 0$ . In part a) the section  $[-1, 1] \times [-1, 1]$  of  $\mathcal{C}$  is shown, and in b), c) and d) subsequent closeups are displayed. Moreover, we note the "self-similarity" in the pictures. The crab-like structures seem to repeat on each scale.

Lagrangian quadratic interpolation



## 2.2.3 Lagrange'sche (quadratische) Interpolation

für Extr. :  $p+1$  Stützstellen (3) $p+1$  Funktionswerte (3)

(also keine Ableitungen)

mit Ersatzpolynom  $p$ -ten Grades (Parabel!)  
 $p=2$  $p=2$  ; Stützstellen  $a, b, c$ neue Stützstelle  $d$ 

$$d = \frac{(b^2 - c^2)F(a) + (c^2 - a^2)F(b) + (a^2 - b^2)F(c)}{(b - c)F(a) + (c - a)F(b) + (a - b)F(c)}$$

Berechnung  $F(d)$ 

Entscheidung

$a < d < b$		$b < d < c$
$F(d) < F(b)$	$F(d) > F(b)$	(analog)
$\curvearrowright$	$\curvearrowright$	
$a, d, b$	$d, b, c$	
als neue Stützstellen		

bis  $d$  in  $\varepsilon$ -Nähe zu einer der Stützstellen
 $\uparrow$   
 zuvor geforderte Genauigkeit

## 2.2.4 Hermite'sche Interpolation (kubisch)

für Extr.: 2 Stützstellen

2 Werte  $F$ 2 Werte  $F_x$ 

(Ersatzpolynom 3. Grades)

Stützstellen  $a, b$ 

neue Stützstelle

$$c = a + (b-a) \frac{w - F_x(a) - z}{2w + F_x(b) - F_x(a)}$$

$$\text{mit } w = + \sqrt{z^2 - F_x(a) \cdot F_x(b)}$$

$$z = \frac{3}{a-b} [F(a) - F(b)] - F_x(a) - F_x(b)$$

$c$  ersetzt  $a$  oder  $b$ , so daß  $a^{(k)}$  und  $b^{(k)}$  stets das Minimum einschließen

Abbruch, wenn  $|a-b| \leq \varepsilon > 0$

2.2.3a <sup>Lagrange'sche</sup> kubische Interpolation ohne Ableitungen4 Stützstellen, 4 Funktionswerte ( $\rho=3$ )

quadrat. Gleichung für Extrema

2.2.3b Lagrange'sche Int. mit  $\rho=4$ 

5 Stützstellen, 5 Funktionswerte

kubische Gleichung für Extrema

(Cardanische Formeln)

$\rho \geq 3$ : mögliches Entkommen aus nur lokalen Optima

Information

Stützstellen

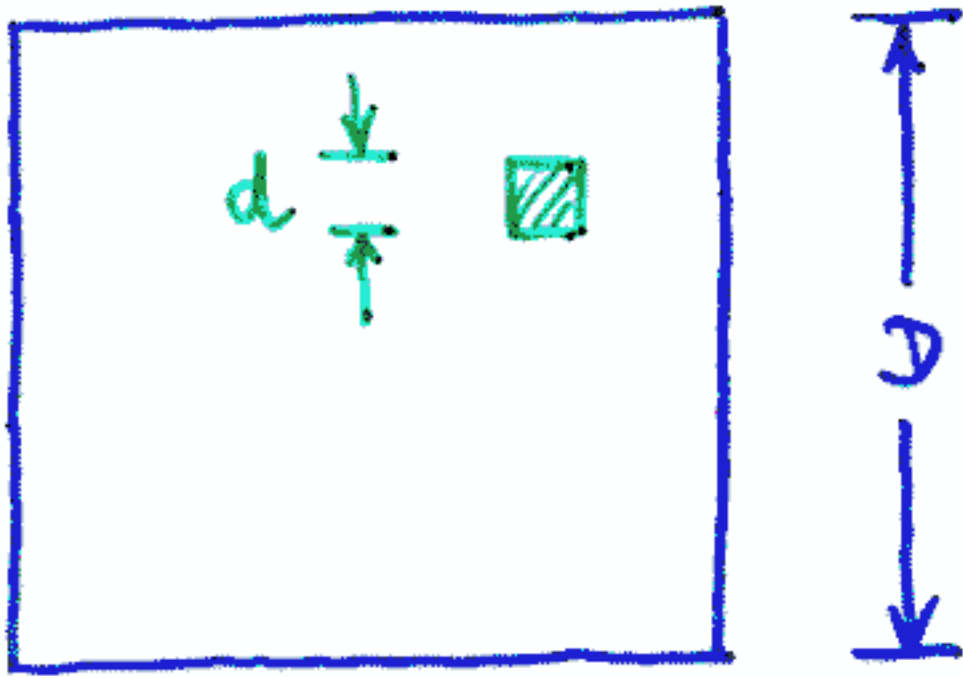
Regula falsi	2	$F_x$
Newton - Raphson	1	$F_x$ $F_{xx}$
Laqrange	$p+1$	$F$
- quadratisch	3	$F$
- Kubisch	4.	$F$
Hermite	2	$F$ $F_x$

## Optimierverfahren für $n > 1$

- Raster
- Monte-Carlo
- Koordinaten
- Gradienten
  - steilster Abstieg / Anstieg
  - konjug. Grad.
  - konjug. Richt.
- Newton
  - variable Metrik
- direkte Suchverfahren
  - Hooke + Jeeves
  - Rosenbrock
  - Nelder + Mead
  - M. J. Box
- Evolutionsstrategien



## Monte-Carlo-Strategie

(reine Zufallsuche)  
Ashby: Homöostat $n$  = Zahl der Variablen(Bild:  $n=2$ )gleichverteilte Zufallstreffer  
in fest vorgeg. endl. Bereich

Trefferwahrscheinlichkeit bei einem Versuch

$$n=1 \quad p_1 = \frac{d}{D}$$

$$n \text{ allgem.} \quad p_n = \left(\frac{d}{D}\right)^n$$

$$\text{Nicht-Treffer 1 Versuch} \quad \bar{p}_1 = 1 - \left(\frac{d}{D}\right)^n$$

$$\text{Nicht-Treffer } N \text{ Versuche} \quad \bar{p}_N = \left[1 - \left(\frac{d}{D}\right)^n\right]^N$$

$$\text{mind. 1 Treffer } N \text{ Versuche} \quad p_N = 1 - \left[1 - \left(\frac{d}{D}\right)^n\right]^N$$

↗ Versuche für vorgegebenes  $p_N$ :

$$N = \frac{\log(1-p_N)}{\log\left[1 - \left(\frac{d}{D}\right)^n\right]} \approx -\ln(1-p_N) \cdot \left(\frac{D}{d}\right)^n$$

$\uparrow_{n \gg 1}$

$$p_N = 0.63$$

$$N \approx 1 \cdot \left(\frac{D}{d}\right)^n$$

← äquidistante  
Rastersuche

$$p_N = 0.9$$

$$N \approx 2,3 \cdot \left(\frac{D}{d}\right)^n$$

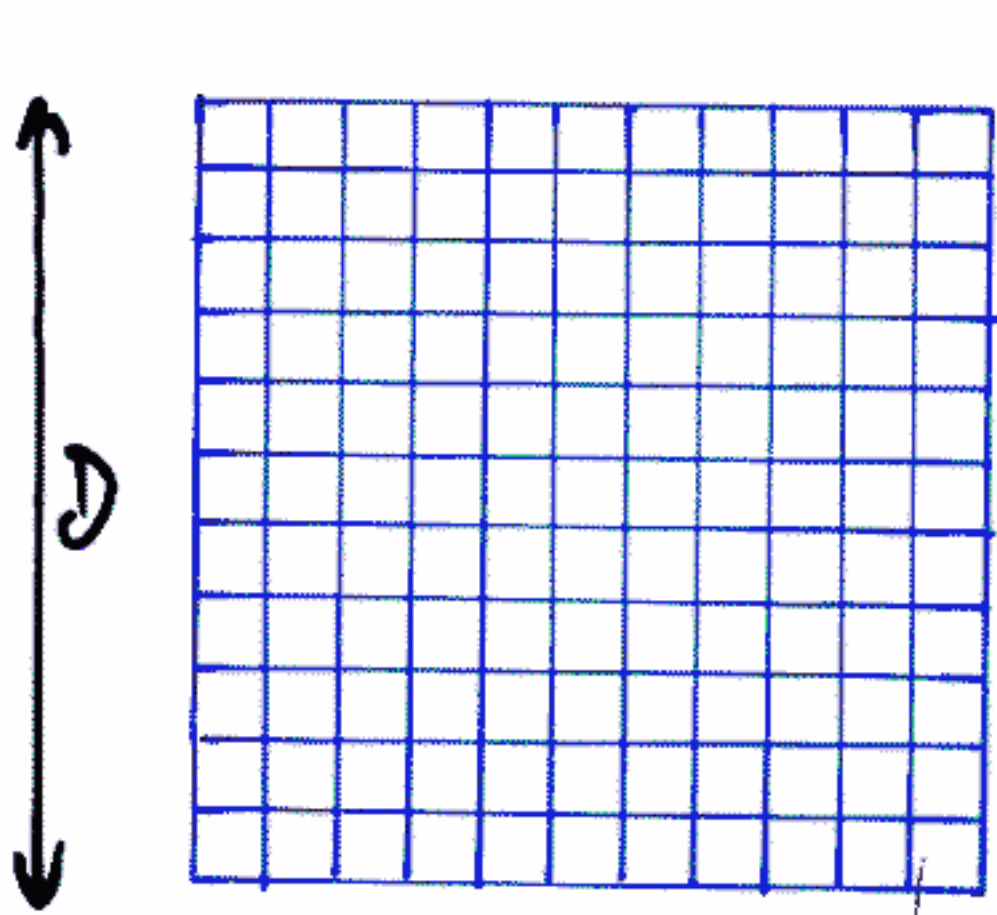
Brooks:

~

$$\frac{1}{\sqrt{n}}$$

unabh. von  $n$

# äquidistante Rastersuche (totale Enumeration)



↓  
↑  
Marschenweite  $d$

$$\text{Aufwand } N = 1 \cdot \left(\frac{D}{d}\right)^n$$

Warum 'besser' als MC?

Weil Zufallsmethoden grundsätzlich schlechter?

Nein! Bei Rastermeth. können Versuche in beliebiger Reihenfolge (also auch zufällig) ausgeführt werden.

MC schlechter, weil Wiederholungen bzw. unnötig dicht benachbarte Versuche nicht ausgeschl. werden.

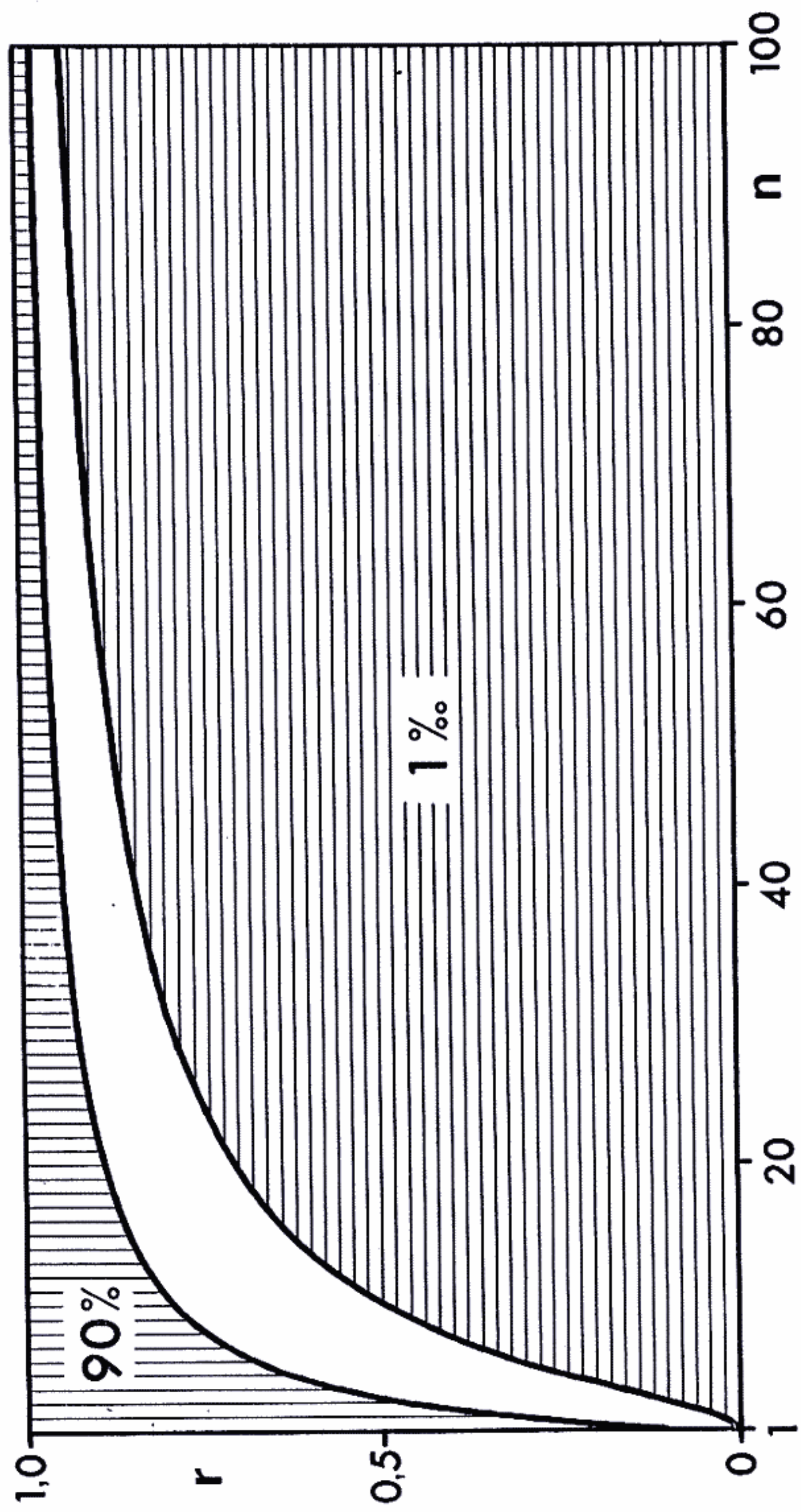
## Weiterentwickelte MC-Methoden

- Suche allmählich auf aussichtsreiche Teilregionen beschränken
- Wahrscheinlichkeitsdichtefunktion an Erfolge anpassen

Folge: nicht mehr voll parallelisierbar

Behauptung Brooks 1958: Aufwand unabh. von  $n$

$$N \sim \frac{V}{v} = \frac{\text{Gesamtvolumen}}{\text{gesuchtes Volumenelement}}$$



Volumen - Verteilung  $n$ -dim. Kugel

Strecken- vs. Volumenverhältnisse für  $n \gg 1$

Sei  $\frac{d}{D} = \varepsilon$        $\frac{v}{V} = \eta = \varepsilon^n$

geg.  $\varepsilon = 0.1$

$n$        $\eta$

2      0.01

10       $10^{-10}$

100       $10^{-100}$

geg.  $\eta = 0.1$

$n$        $\varepsilon$

2      0.32

10      0.79

100      0.98

Sei  $p = 0.63$  ;  $\varepsilon = 0.1$  ;  $n = 17$

dann  $N = 1 \cdot 10^{17}$  Versuche erforderlich

z.B. 1s/Versuch : 3 Mrd. Jahre !

Sei  $p = 0.63$  ;  $\varepsilon = \frac{1}{2}$  ;  $n = 10^5$

dann  $N = 1 \cdot 2^{100000} \approx 10^{30000} \gg 10^{120}$

praktisch  $\infty$

$10^{80}$  Elementarbausteine im Kosmos

$10^{60}$  Elementarzeitschritte seit Urknall

$10^{140}$  kosmolog. Obergrenze für Ereignisse  
z.B. Zahl der Versuche

Parallelität allein hilft kaum:

$10^{30000} / 10^{80} = 10^{29920}$

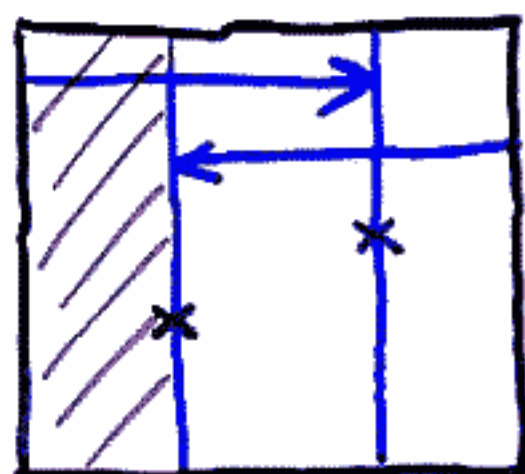
Generationen

mind. Stunden

(Bakterien, Hefen)

2. sequentielle Methoden für  $n > 1$ 

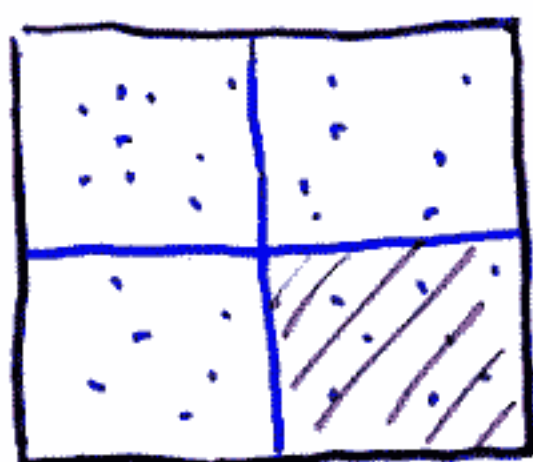
## 2.1 Fibonacci



sukzessives Wegschneiden vom Ausgangsintervall

$$N \sim \left( \log \left( \frac{D}{d} \right) \right)^n$$

## 2.2 M-C mit Intervallreduktion



## 2.3 Verfahren, die einen linienförmigen Weg zum Optimum hin verfolgen

Abstand zweier Zufallspunkte in einem  $n$ -dimensionalen Würfel mit Kantenlänge  $L$ :

$$l = L \cdot \sqrt{\frac{n}{6}} \quad \leadsto \text{Aufwand} \sim \sqrt{n}$$

allgemeine Iterationsvorschrift:

$$x^{(k+1)} = x^{(k)} + \underbrace{s^{(k)}}_{\text{Schrittweite}} \cdot \underbrace{v^{(k)}}_{\text{Richtung}} \quad (\|v^{(k)}\| = 1)$$

einfachste Regel:  $s^{(k)}$  fest

$$v^{(k)} = \{e_1, e_2, \dots, e_n\}$$

zyklisch

allg. Iterationsregel

sequentielle Optimierung

$$\underline{x}^{(k)} = \underline{x}^{(k-1)} + s^{(k)} \cdot \underline{v}^{(k)}$$

$s^{(k)}$  Schrittweite       $\underline{v}^{(k)}$  Richtung       $|\underline{v}^{(k)}| = 1$

$$\underline{x} = \{x_1, x_2, \dots, x_n\}$$

Richtung: z. B. Koordinatenrichtungen, zyklisch  
 Gradientenrichtung  
 (lokal optimal)  
 Zufallsrichtung gleichverteilt

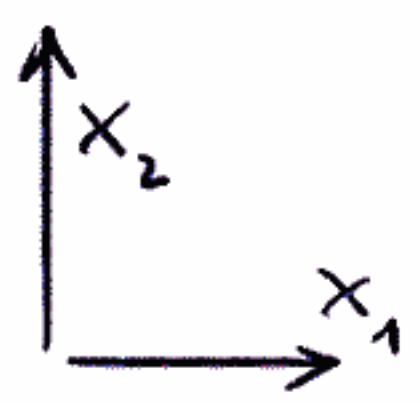
Schrittweite: z. B. fix  
 maximal (bis zum relativen Opt.)  
 Eindim. (Linien-)Suche  
 zufällig normalverteilt

gute Algorithmen "erlernen"  $s$  und  $v$   
 adaptiv durch "learning by doing"  
 Versuch und Irrtum

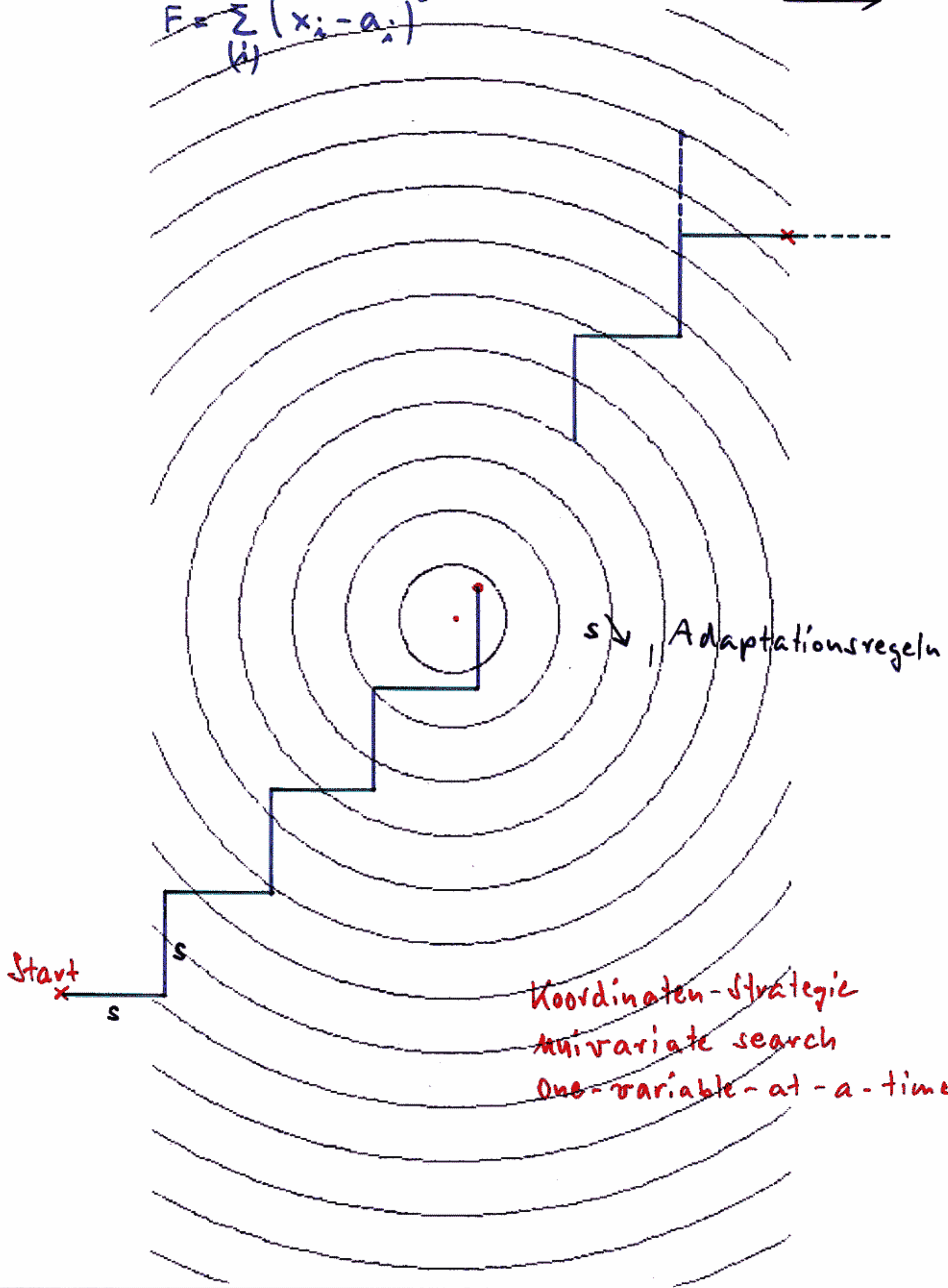
Evolutionstrat.:

$v$  zufällig, gleichverteilt  
 $\Delta x_i = s_i$  zufällig, normalverteilt ( $\mu = 0, \sigma_i$ )  
 $s_{ges}$  quasi fix für  $n \gg 1$   
 (Kugelrandverteilung)

Graß-Weidel-Strategie I  
Feste Schrittweite



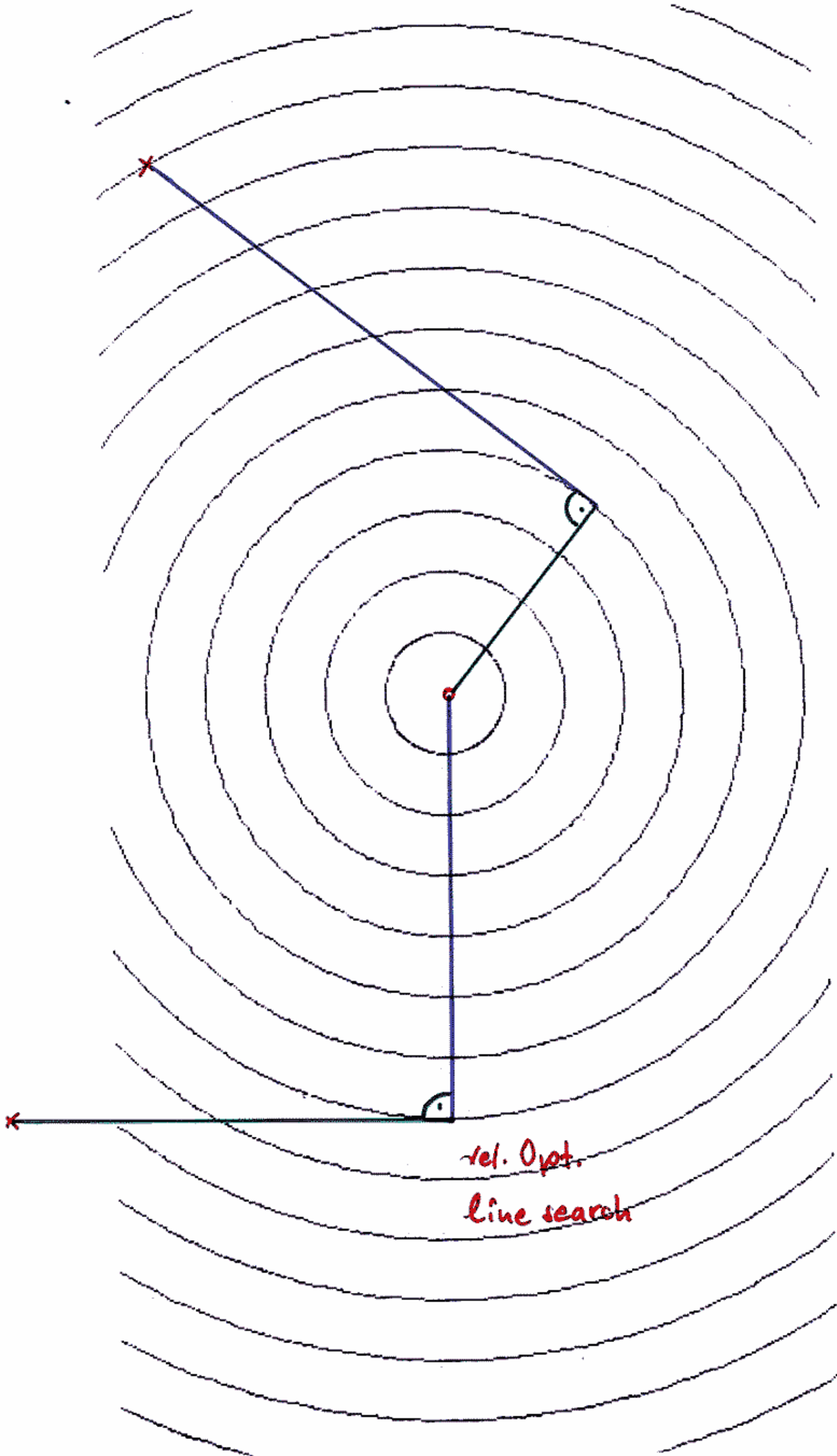
$$F = \sum_{(i)} (x_i - a_i)^2$$



Koordinaten-Strategie  
univariate search  
one-variable-at-a-time

# Graß-Seidel-Strategie II

mit line search

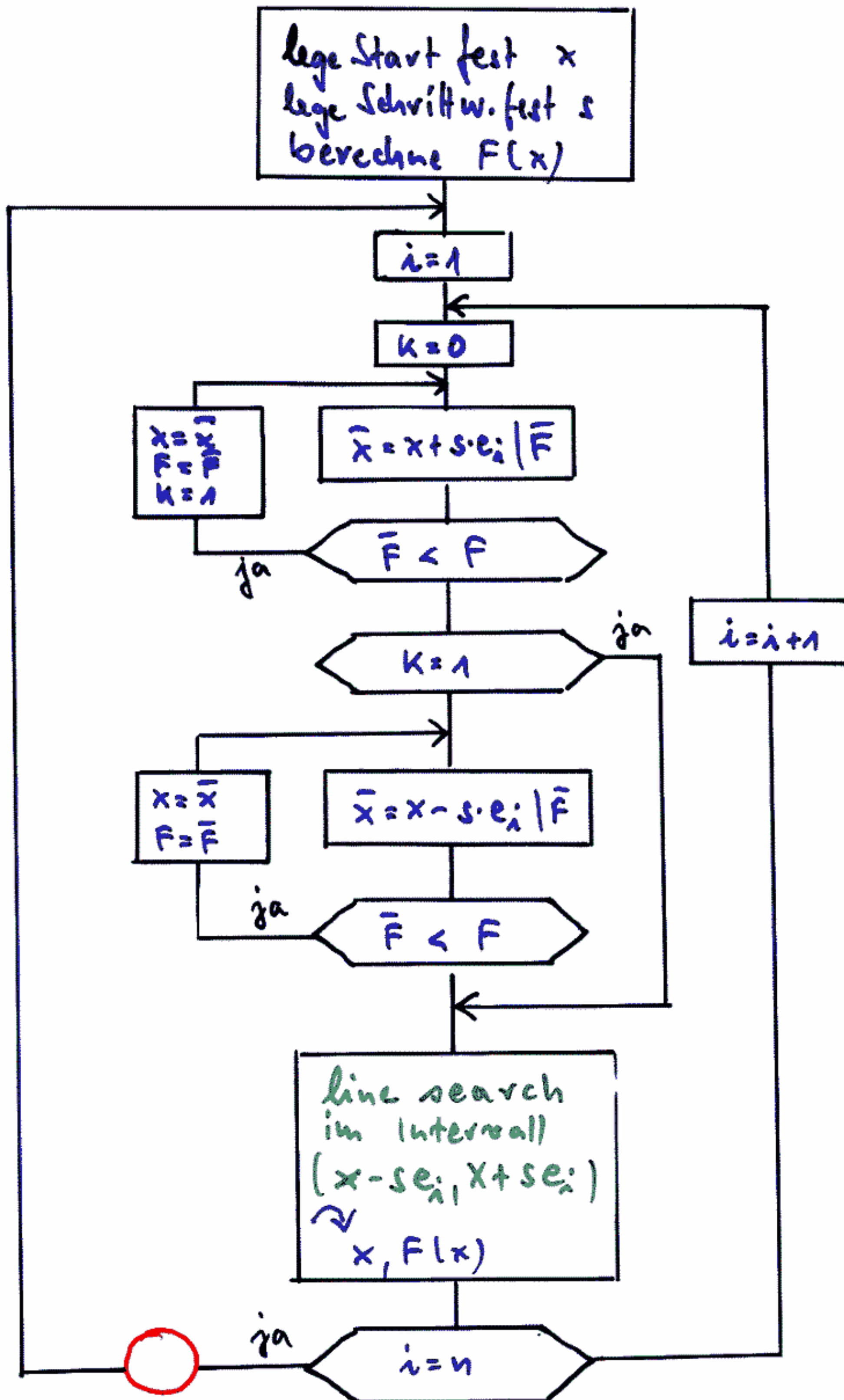


$$F(x) = \sum_{(i)} f_i(x_i)$$

weil Problem dekomponierbar  
Zielfunktion separabel



# Graß-Seidel-Strategie



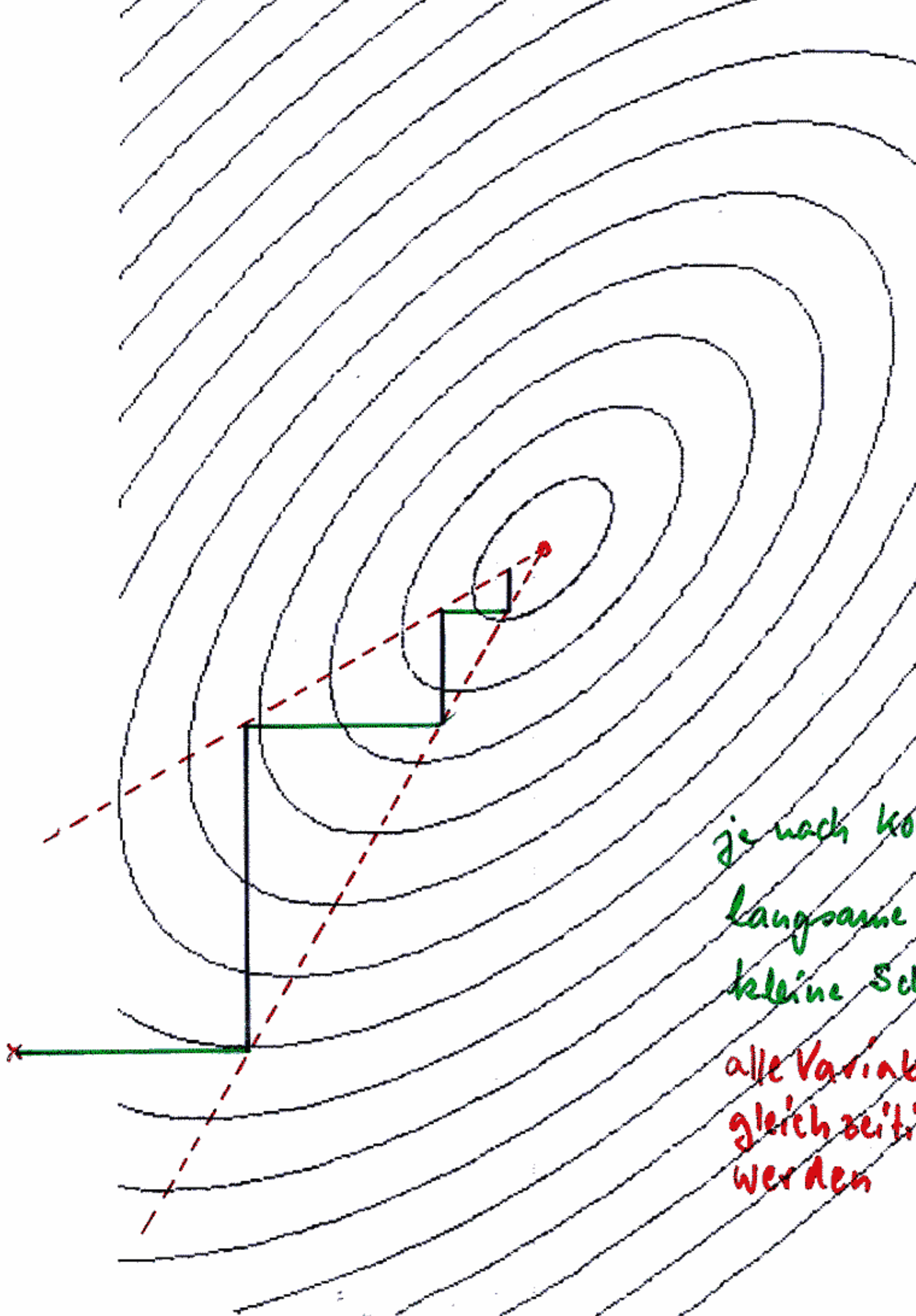
Abschaltregel:

wenn über 1 Zyklus keine Verbesserung, dann  $s$  halbieren  
wenn  $s < \epsilon$  dann Stop

# Gauß-Seidel Strategie III

$$F(x) = x^T A x + b x + c$$

nicht separierbar



je nach kondition von A:  
langsame Konvergenz,  
kleine Schritte

alle Variablen müssen  
gleichzeitig geändert  
werden

## Gradientenrichtung

$$F(x) \rightarrow \min$$

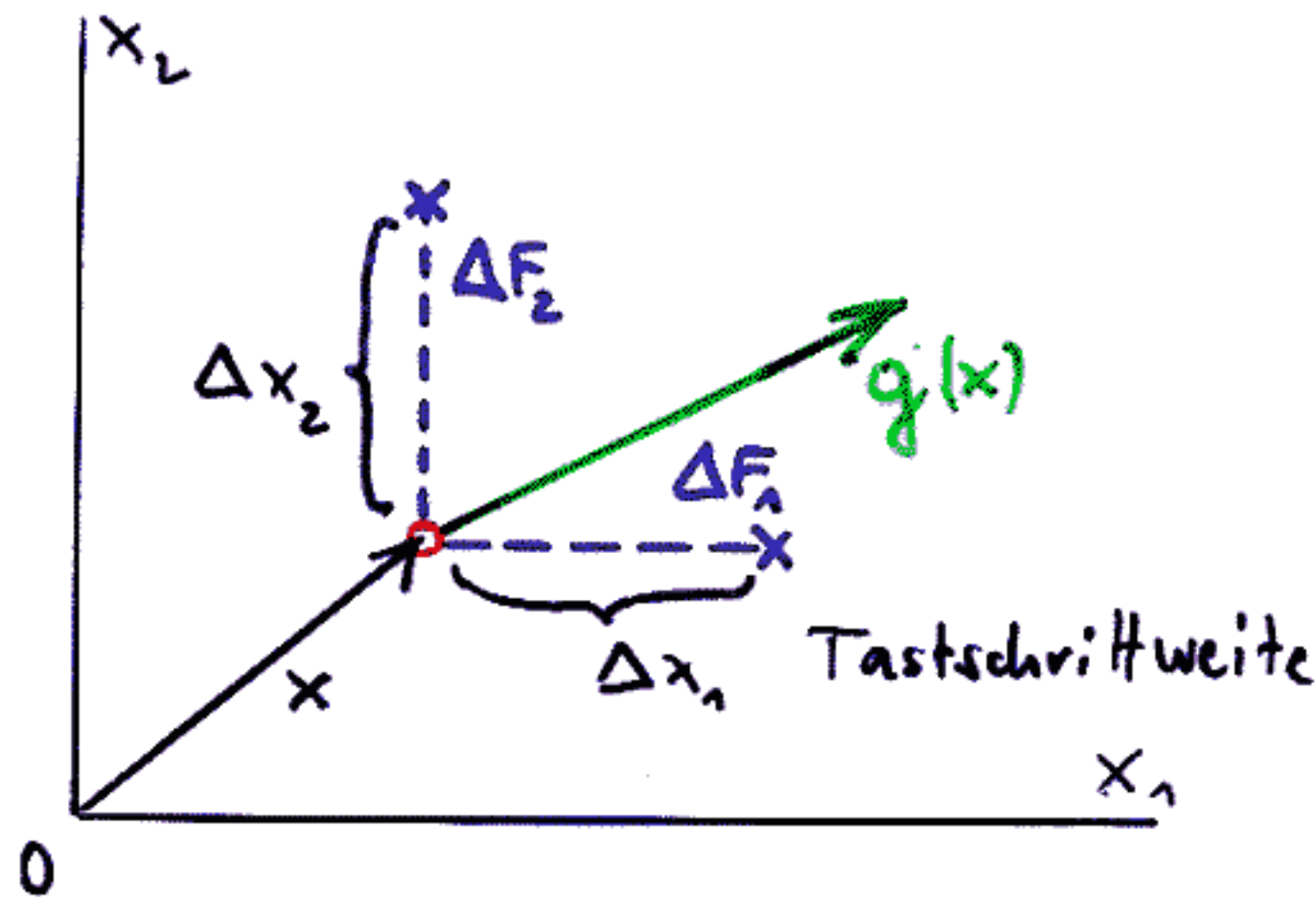
$$x = \begin{Bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{Bmatrix} = \{x_1, x_2, \dots, x_n\}^T = x_1 e_1 + x_2 e_2 + \dots + x_n e_n$$

$$\nabla F = \left\{ \frac{\partial F}{\partial x_1}, \frac{\partial F}{\partial x_2}, \dots, \frac{\partial F}{\partial x_n} \right\}^T \quad \text{Nabla - Operator}$$

$$g(x) = \frac{\partial F}{\partial x_1}(x) e_1 + \frac{\partial F}{\partial x_2}(x) e_2 + \dots + \frac{\partial F}{\partial x_n}(x) e_n$$

$$\text{bzw. } g'(x) = \frac{g(x)}{\|g(x)\|}$$

⇒ n erste partielle Ableitungen der Zielfunktion müssen bekannt sein oder näherungsweise ermittelt werden



$$\frac{\partial F}{\partial x_i} \approx \frac{\Delta F}{\Delta x_i}$$

$$x^{(k+1)} = x^{(k)} \pm s \cdot g^{(k)}$$



line search

$k$  = Iterationszähler

Gradientenverfahren (für Min. Suche)

$$v^{(k)} = -g(x^{(k)}) = -\frac{\nabla F(x^{(k)})}{\|\nabla F(x^{(k)})\|} \quad ; \quad \|v^{(k)}\| = 1$$

$s^{(k)}$  = „optimal“

d.h. line search entlang  $v^{(k)}$

z.B. mit Lagrange'scher Interpolation  
oder Hermite'scher Interpolation

↳ Optimom gradient search

= steepest descent (ascent)

steilster Abstieg (Anstieg)

d.h.  $n$  partielle Ableitungen auswerten

oder  $n$  Tastschritte ausführen

Aufwand je zurück gelegter Strecke  $\sim \frac{n+1}{s}$

bzw. Konvergenzgeschwindigkeit  $\sim \frac{s}{n+1}$

Gradientenkonzept entspricht

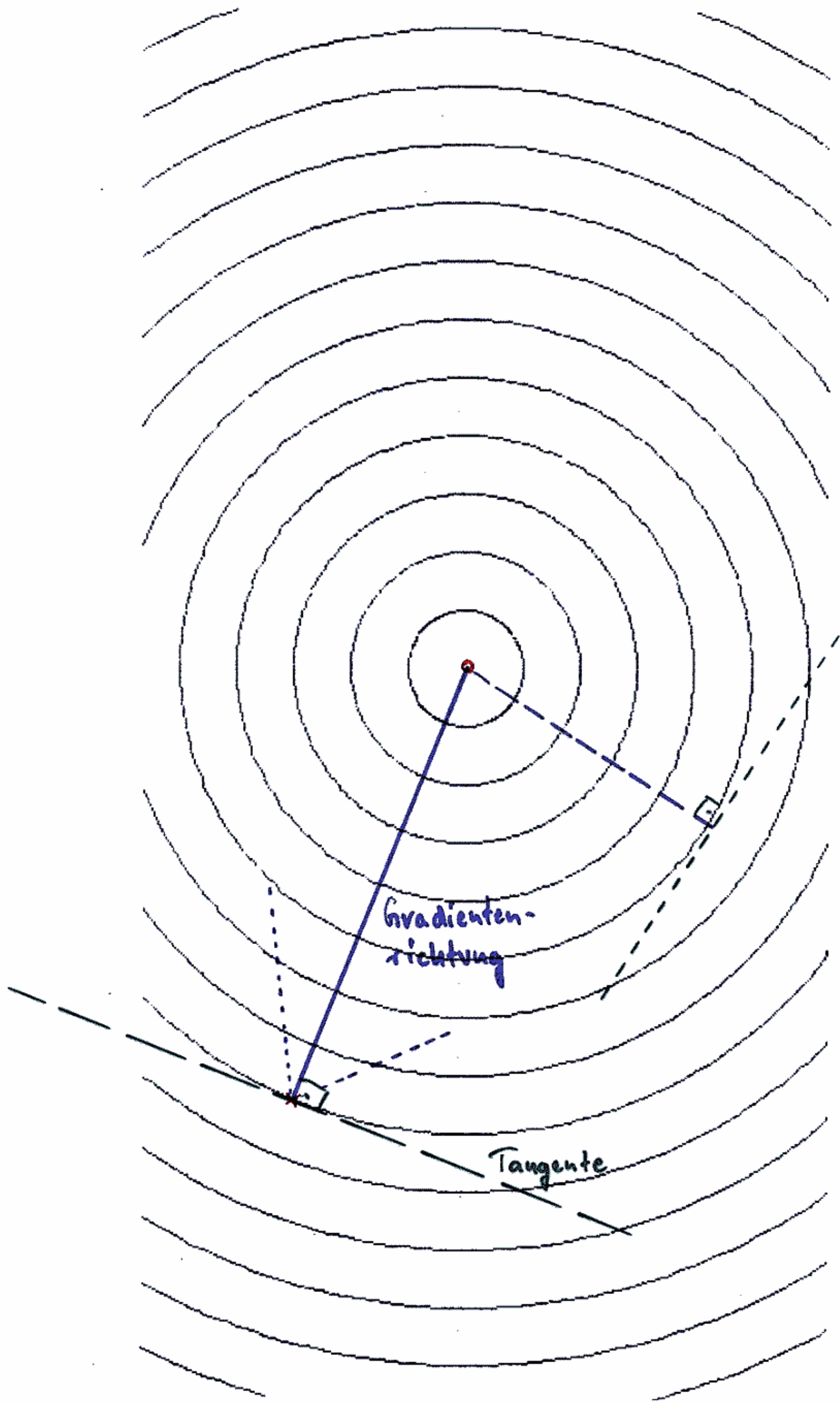
linearem „inneren Modell“ über Zielfunktion

d.h. im linearen Fall ist Gradientenrichtung  
global optimal

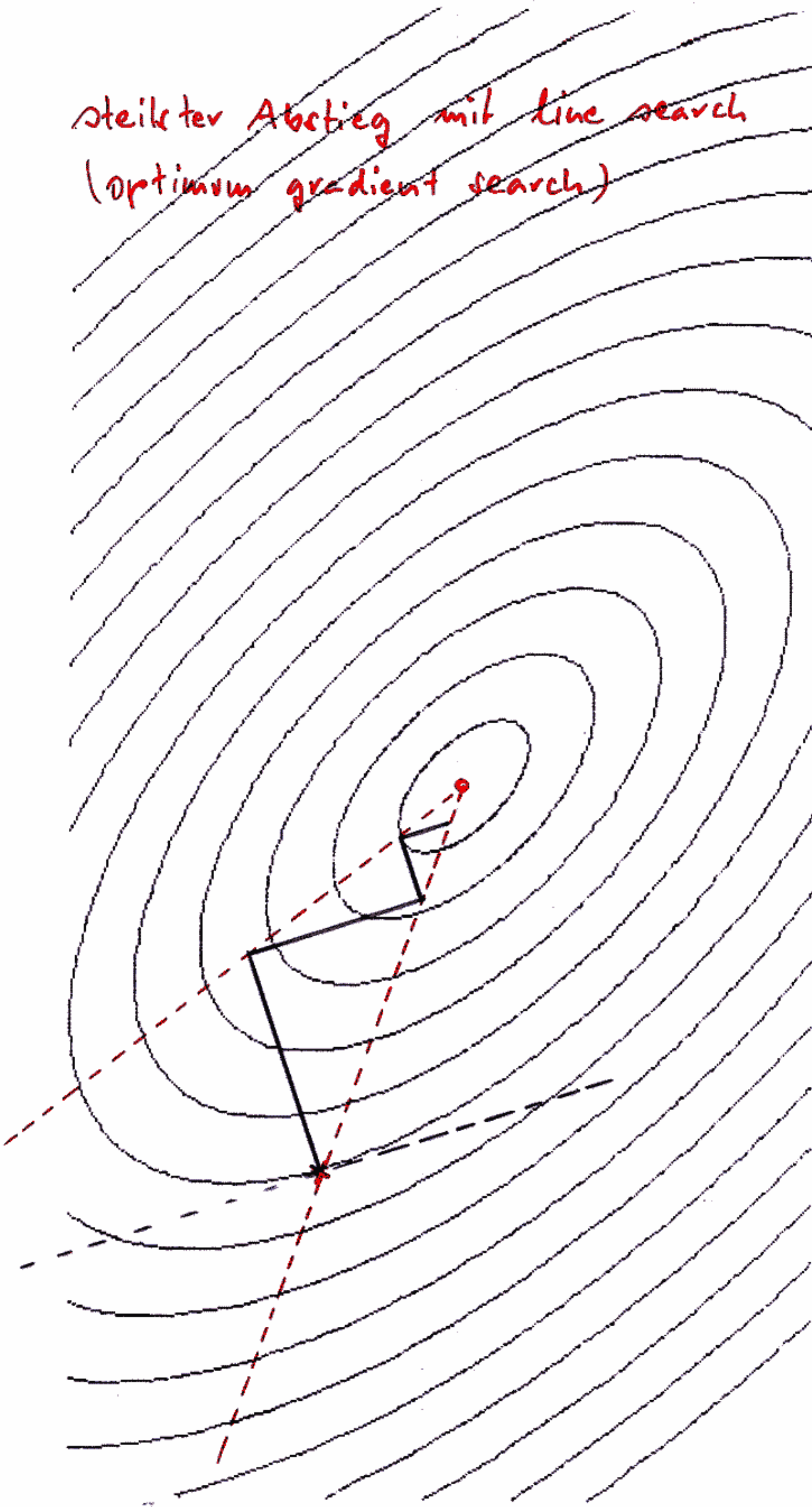
sonst nur lokal optimale Richtung

Steilster Abstieg  
(Anstieg)

Gradient

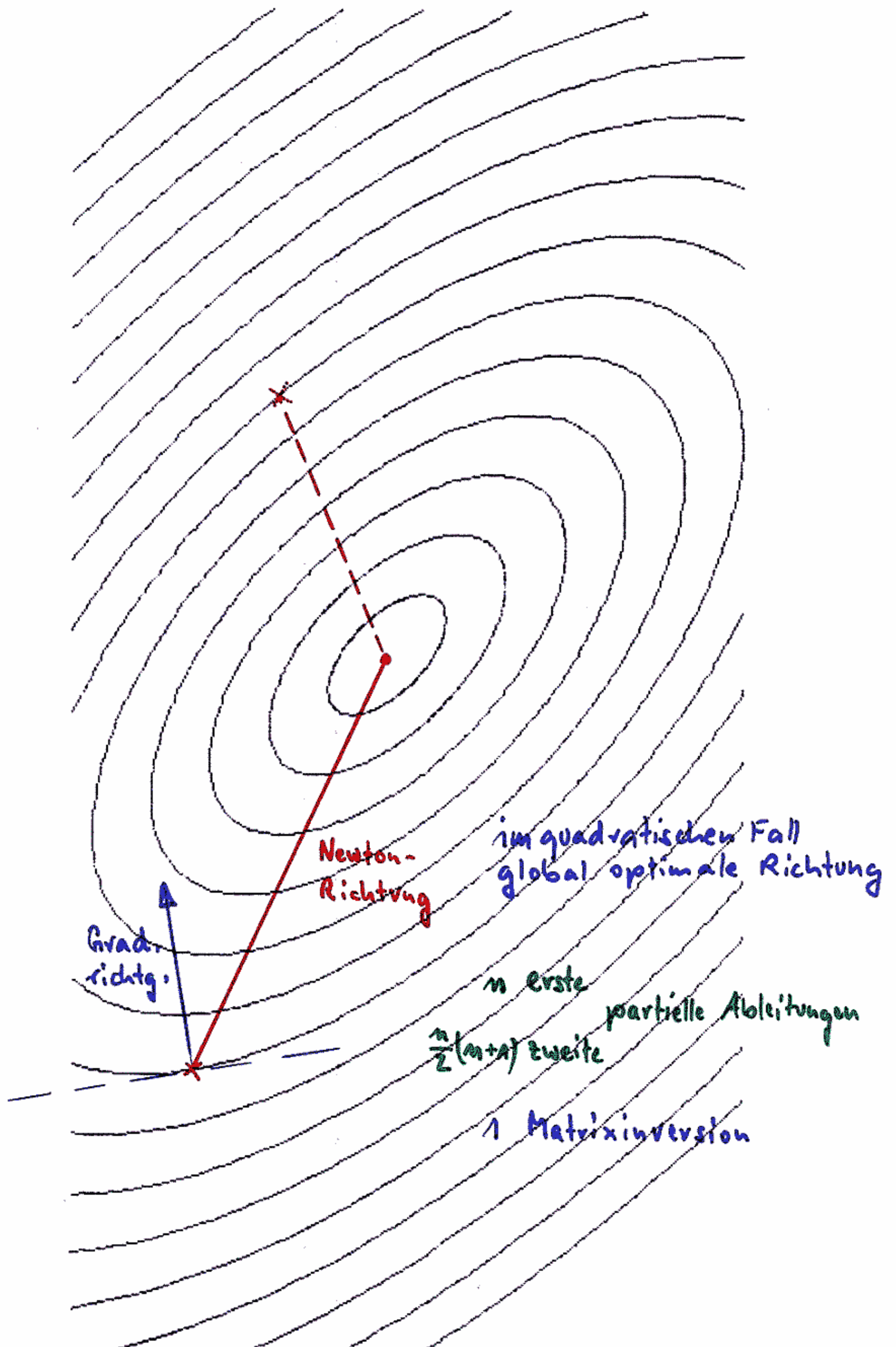


steilster Abstieg mit line search  
(optimum gradient search)



nicht besser als haupt-Seidel  
1. Ableitungen benötigt  
lineares internes Modell

# Newton - Raphson Strategie



$$\left\{ \begin{array}{ccc} \frac{\partial^2 F}{\partial x_1^2} & \frac{\partial^2 F}{\partial x_1 \partial x_2} & \dots & \frac{\partial^2 F}{\partial x_1 \partial x_n} \\ \frac{\partial^2 F}{\partial x_2 \partial x_1} & \dots & \dots & \dots \\ \vdots & & & \\ \frac{\partial^2 F}{\partial x_n \partial x_1} & \dots & \dots & \frac{\partial^2 F}{\partial x_n^2} \end{array} \right\} = \nabla^2 F$$

$$H(x) = \nabla^2 F(x)$$

Hesse'sche Matrix  
inverse H

$$x^{(k+1)} = x^{(k)} - \underbrace{\left[ \nabla^2 F(x^{(k)}) \right]^{-1} \nabla F(x^{(k)})}_{\text{Schnittpunkt}}$$

$$= x^{(k)} + s^{(k)} \cdot v^{(k)}$$

Schriftweite      Richtung

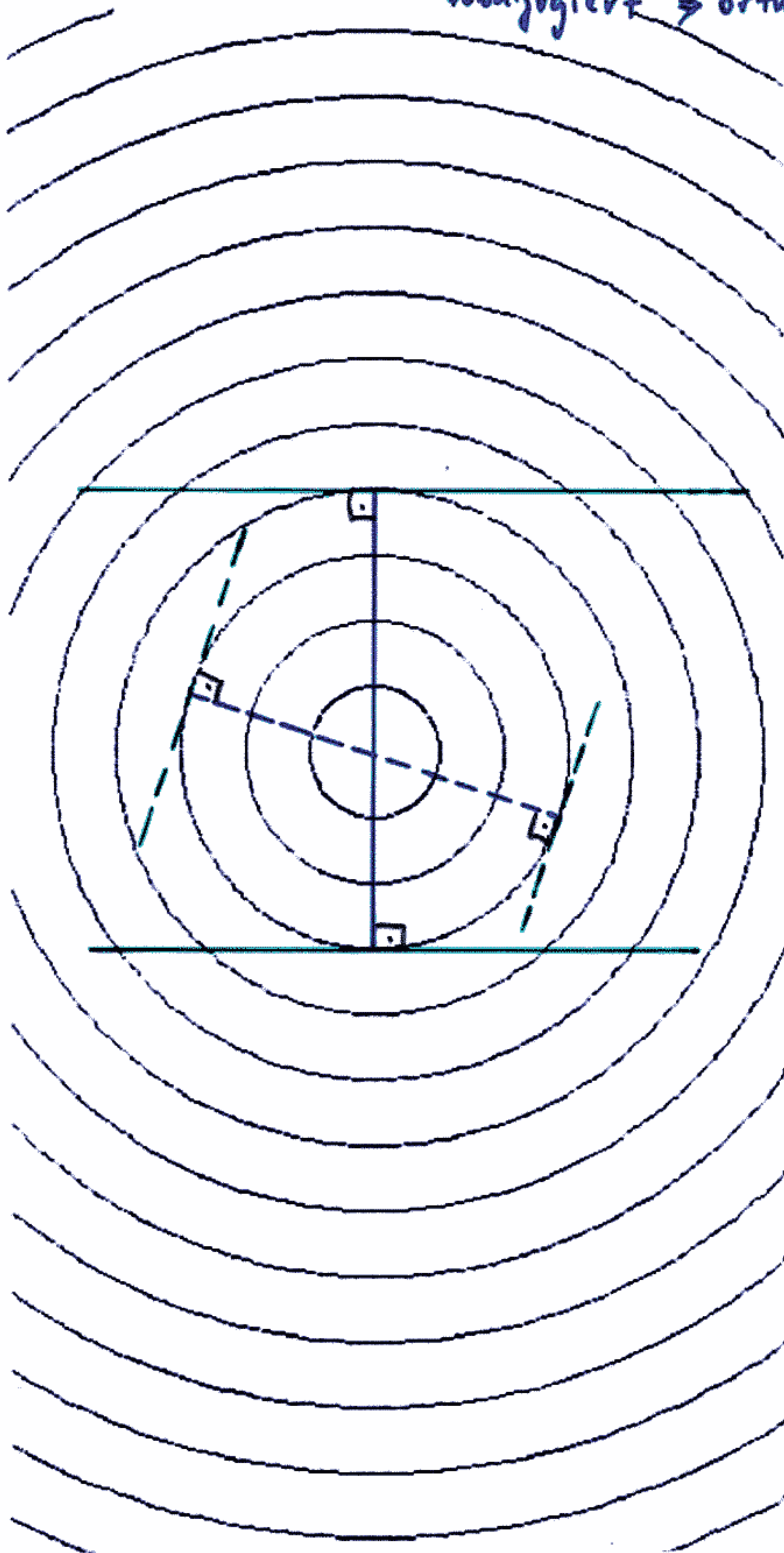
im quadrat. Fall keine line search nötig  
nur 1 Iteration  $\Rightarrow$  Q1



nach linearer Koordinatentransformation

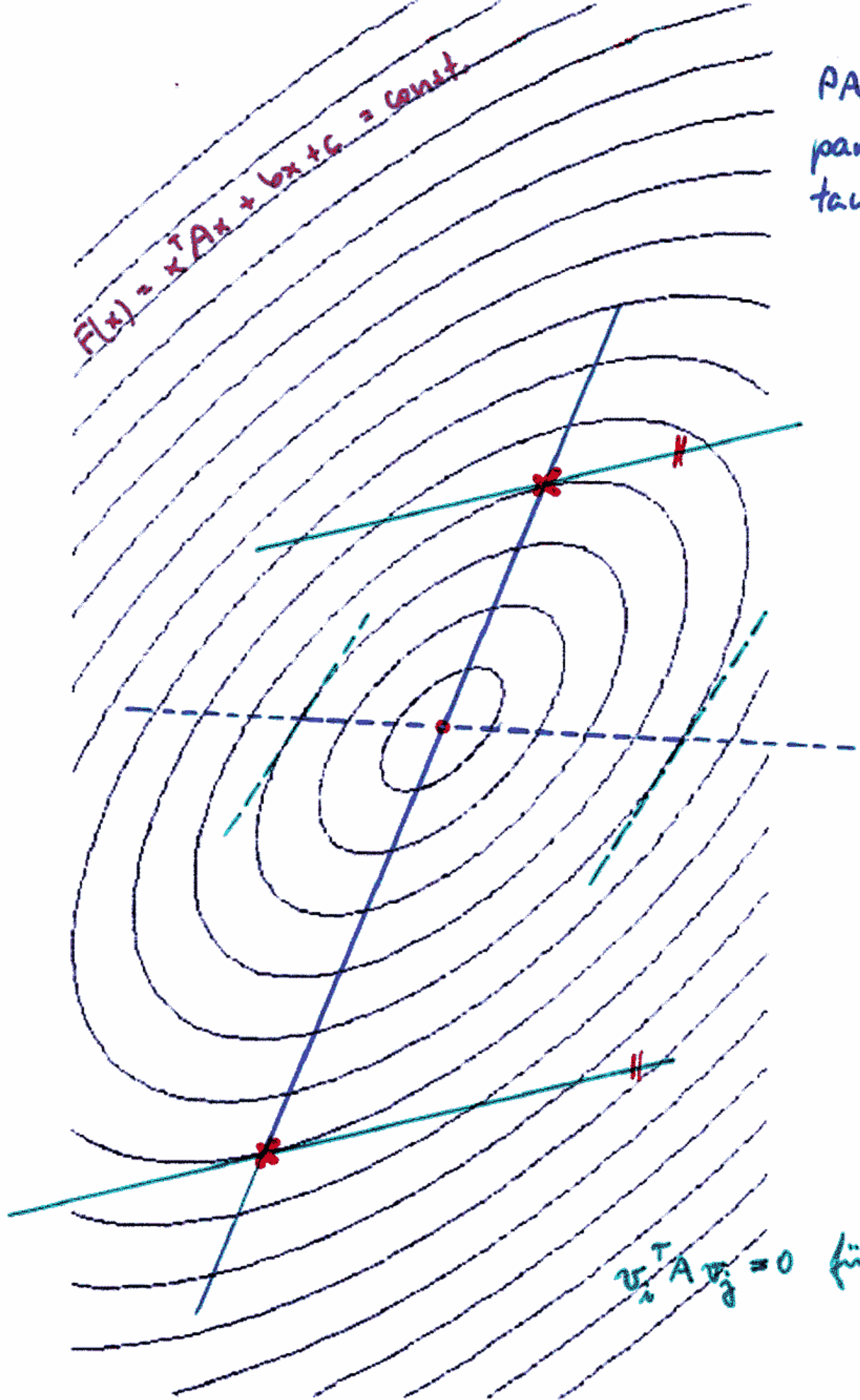
$$A \Rightarrow I$$

kongugiert  $\Rightarrow$  orthogonal



PARTAN  
parallel  
tangents

$F(x) = x^T A x + b x + c = \text{const.}$



$v_i^T A v_j = 0$  für  $1 \leq i \leq n$   
 $1 \leq j \leq n$   
 $i \neq j$

konjugierte Richtungen / Hestenes 1956

$O(n)$  line searches  $\Rightarrow O(n)$  für PARTAN

## Verfahren der konjugierten Gradienten (Fletcher + Reeves, 1964)

$$v^{(0)} = -\nabla F(x^{(0)})$$

$$v^{(k)} = \alpha v^{(k-1)} - \nabla F(x^{(k)}) \quad 1 \leq k \leq n$$

$$\alpha = \frac{\{\nabla F(x^{(k)})\}^T \nabla F(x^{(k)})}{\{\nabla F(x^{(k-1)})\}^T \nabla F(x^{(k-1)})} \quad \text{Korrektur-} \\ \text{faktor}$$

↳ im quadratischen Fall

n Iterationen mit ~~Ihr Text~~ line search

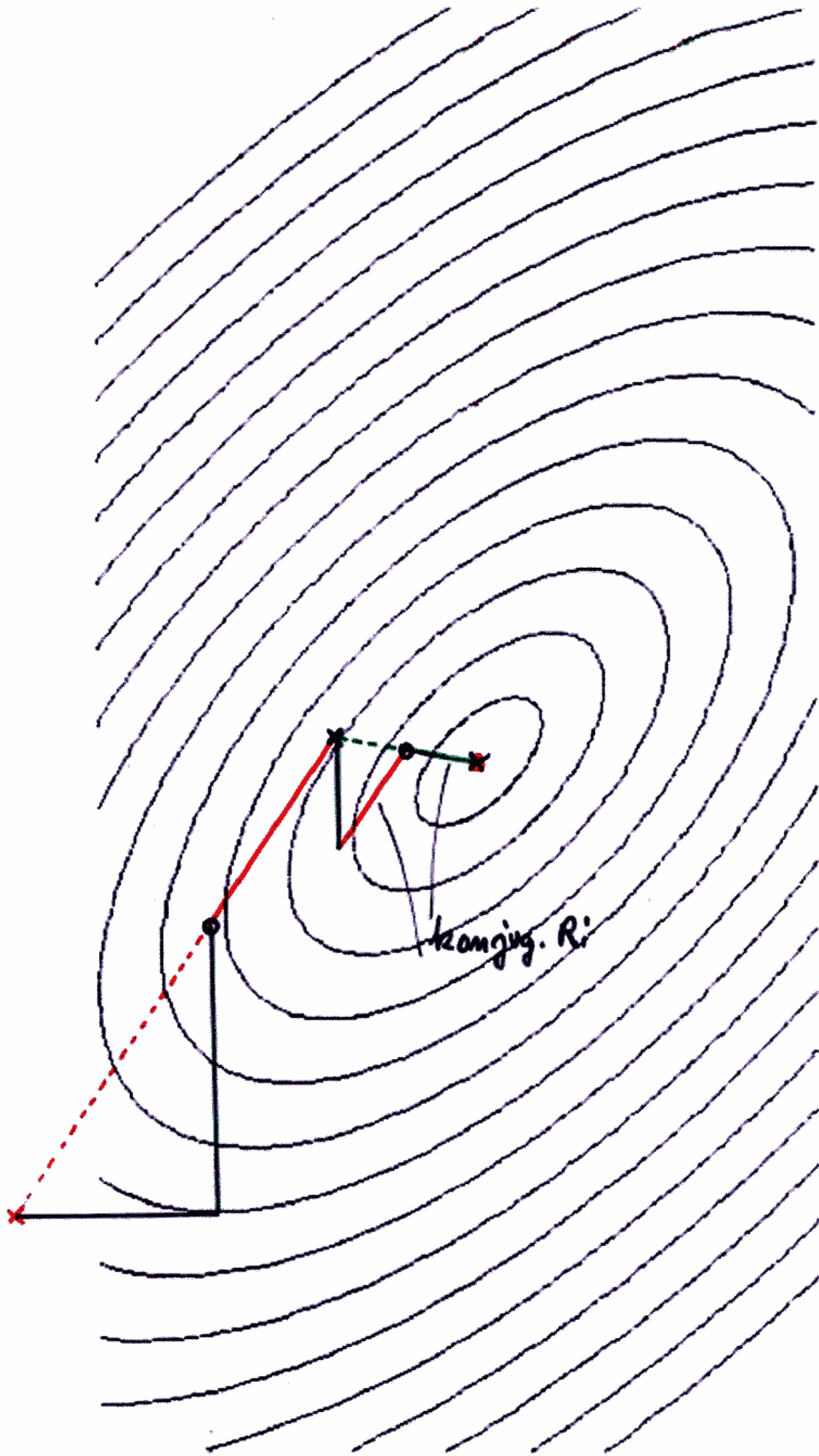
Qu

(muß exakt sein)

z.B. Hermite'sche Interpolation

nicht quadratisch:

"Neustart" mit einfacher Gradientenrichtung  
nach jeweils n Iterationen



Powell - Verfahren der konjug. Richtungen (ableitungsfrei) 1964

$n(n+1)$  Line searches.

$$T \sim n^2 \cdot k \cdot n^f \quad f \geq 1$$

$$\Rightarrow O(n^{2+f})$$

## Quasi-Newton-Verfahren

variable Metrik  
z.B. Davidon  
1959

$$x^{(k+1)} = x^{(k)} + s^{(k)} \cdot H^{(k)T} \nabla F(x^{(k)})$$

line search

Näherung für inverse  
Hesse'sche Matrix

$$H^{(k+1)} = H^{(k)} + A^{(k)}$$

$$H^{(0)} = I$$

Korrekturterm [wird "erlernt" aus  
Erfahrungen der  
vorhergehenden  
Iterationen]

im quadratischen Fall:  $\mathcal{O}(n)$  Iterationen

$\Rightarrow$  Eigenschaft  $\mathcal{O}(n)$

$$T \sim n \cdot [(1+n) \cdot n] = \mathcal{O}(n^3)$$

## DFP Davidon, Fletcher, Powell

$$A_{DFP}^{(k)} = \frac{y^{(k)} y^{(k)T}}{y^{(k)T} z^{(k)}} - \frac{H^{(k)} z^{(k)} \{H^{(k)} z^{(k)}\}^T}{z^{(k)T} H^{(k)} z^{(k)}} \quad \text{Rang 2}$$

wobei:  $y^{(k)} = x^{(k)} - x^{(k-1)}$

$$z^{(k)} = \nabla F(x^{(k)}) - \nabla F(x^{(k-1)})$$

## BFGS Broyden, Fletcher, Goldfarb, Shanno

$$A_{BFGS}^{(k)} = \frac{\{y^{(k)} - H^{(k)} z^{(k)}\} \{y^{(k)} - H^{(k)} z^{(k)}\}^T}{\{y^{(k)} - H^{(k)} z^{(k)}\} z^{(k)T}} \quad \text{Rang 1}$$

(rank one or variance method(s))

Vorteil: auf line search kann verzichtet werden

Nachteil: sogar im quadrat. Fall kann  $H^{(k)}$  positive Definitheit verlieren

(im allg. Fall sowieso!)

Mischformel:

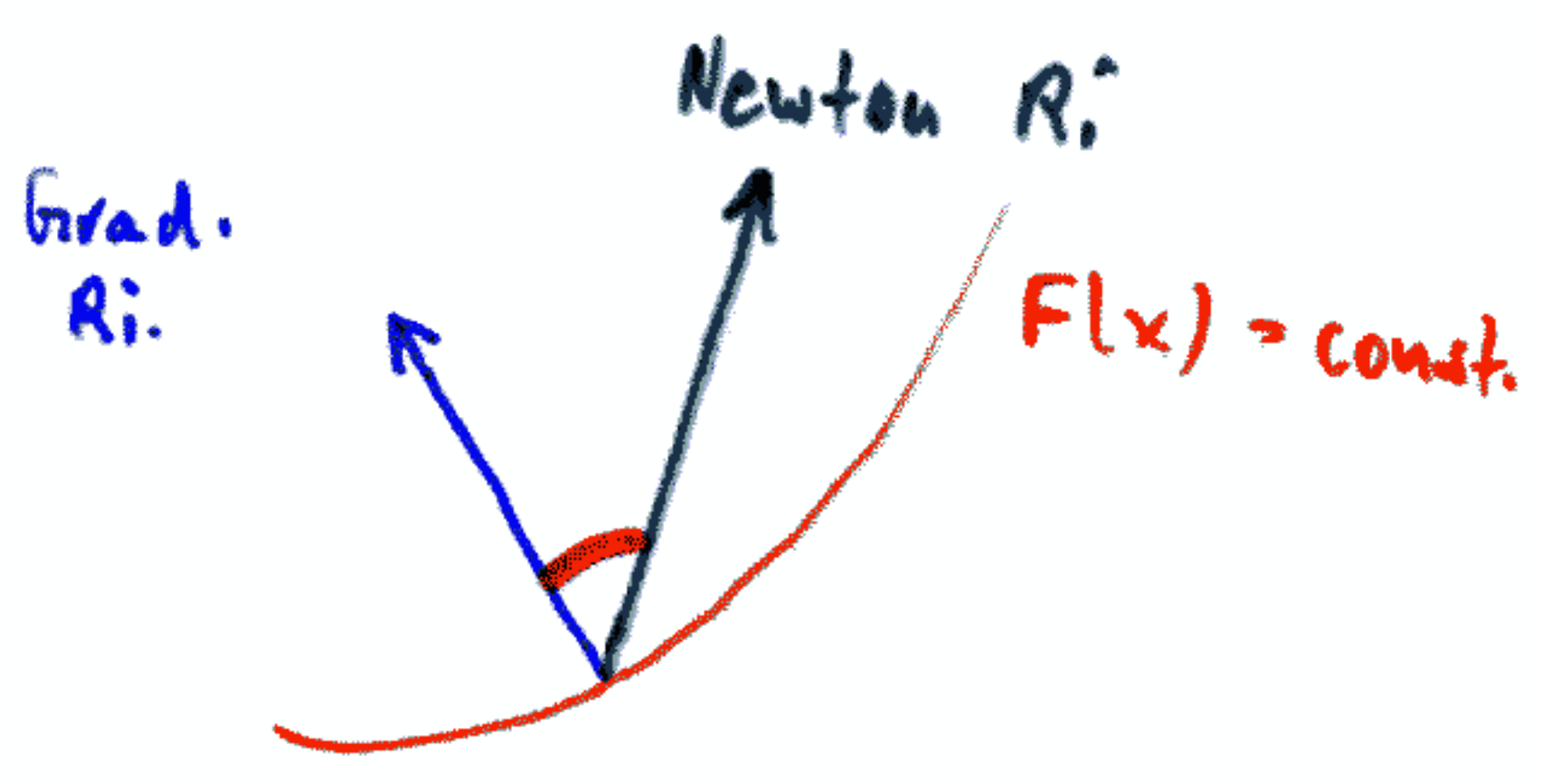
$$H^{(k+1)} = H^{(k)} + A_{DFP}^{(k)} + \alpha^{(k)} A_{BFGS}^{(k)}$$

$$\alpha^{(k)} > 0 \text{ beliebig}$$

n. v. a. m.

Strategie	<u>Zahl der</u> Iterationen	Teiloperationen je Iteration				
		Funktionsauswertungen			einfachen Operationen	
		F	VF	V <sup>2</sup> F		
Newton z.B. Newton-Raphson	$n^0$	-	$n^0$	$n^0$	$n^3$	
variable Metrik z.B. Davidon	$n^1$	$n^0$	$n^0$	-	$n^2$	
konjugierte Gradienten z.B. Fletcher-Reeves	$n^1$	$n^0$	$n^0$	-	$n^1$	
konjugierte Richtungen z.B. Powell	$n^2$	$n^0$	-	-	$n^1$	
		$n^0$	$n^1$	$n^2$		
		Gewichtsfaktoren				

$O(n^3)$  trotz Q1



allgem. Fall :  $\Delta$  oft nahezu  $\perp$   
 'beste' Richtung evtl. außerhalb  $\Delta$

- ↪ scheinbar simple Konzepte miteinander erfolgreicher
- hill climbing
- trial and error
- heuristische (direkte) Suchverfahren



Hooke + Jeeves pattern search (1961)

Grobschema:

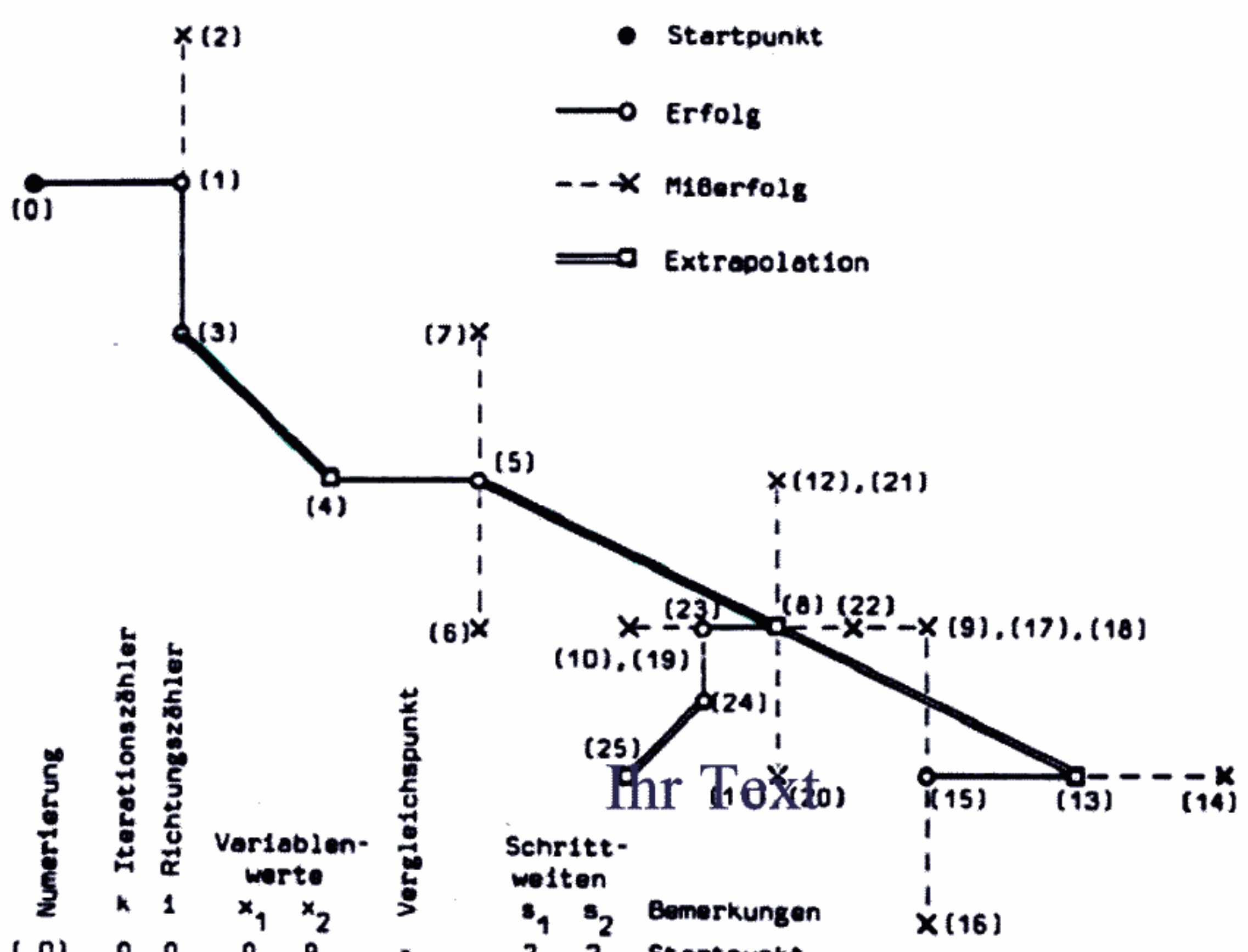
- ①  $r^{(k)}$  zyklisch in Koord. richtungen  
 $s^{(k)}$  diskret, fest  
 bei Erfolg: sofort realisieren  
 bei Mißerfolg: Gegenrichtung probieren
- ② nach Zyklus mit mind. einem Erfolg  
 Extrapolation in Gesamterfolgsrichtung mit  
 Schrittweite entspr. Summe der erfolgr. Schritte  
 keine Erfolgskontrolle  
 weiter mit ①  
 anderenfalls: letzte Extrapolation rück-  
 gängig machen und weiter mit ①;  
 wenn auch dies erfolglos, dann  
 Schrittweite halbieren, solange  $s \geq \epsilon$
- ③ Stop, wenn  $s < \epsilon$

Abwandlungen: razor search (mit Neb.)

line search für Extrapolation

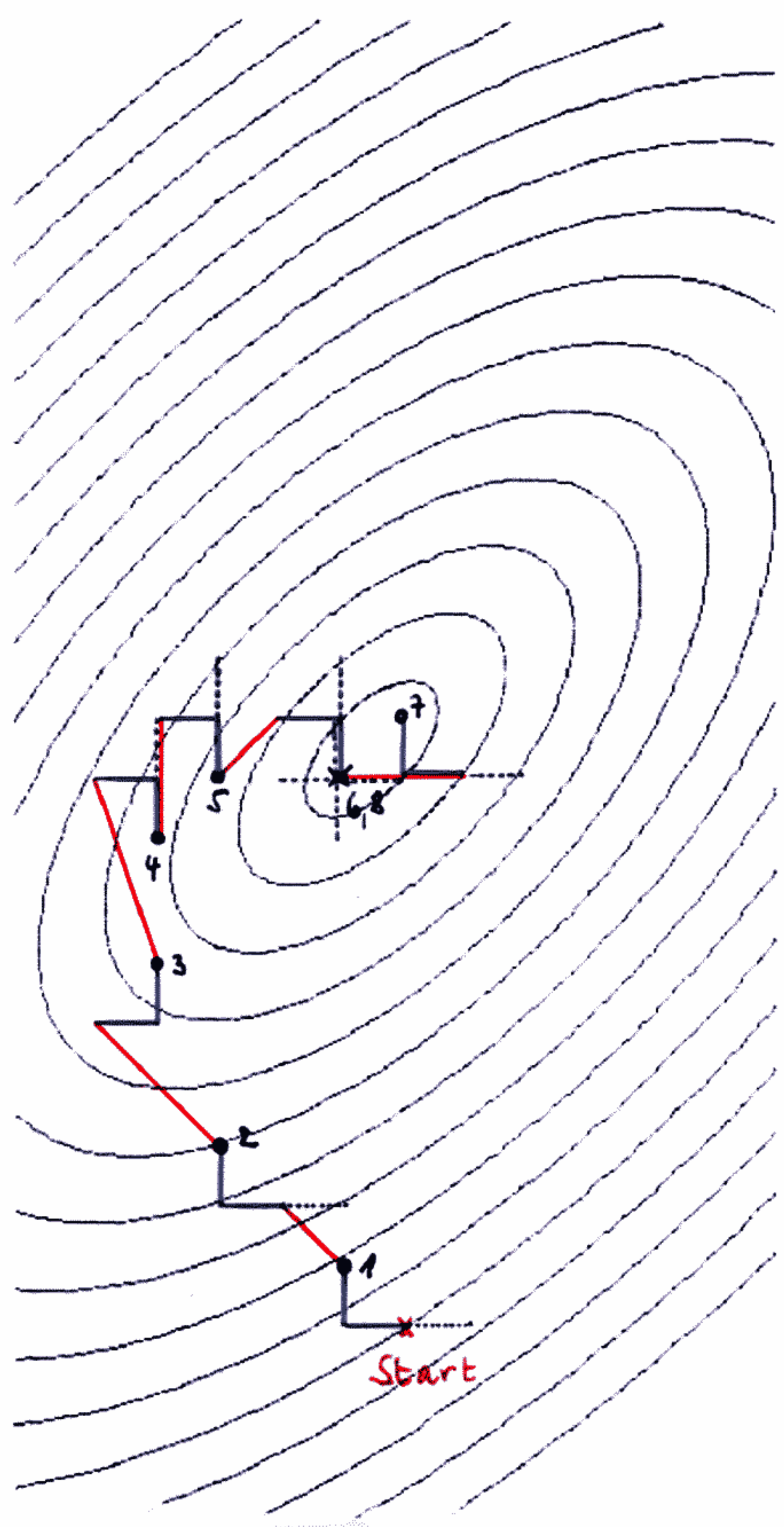
pattern search 1961

Strategie von Hooke und Jeeves



Numerierung	Iterationen	Richtung	Variablenwerte		Vergleichspunkt	Schrittweiten		Bemerkungen
			$x_1$	$x_2$		$s_1$	$s_2$	
(0)	0	0	0	9	-	2	2	Startpunkt
(1)	0	1	2	9	(0)			Erfolg
(2)	0	2	2	11	(1)			Mißerfolg
(3)	0	2	2	7	(1)			Erfolg
(4)	1	0	4	5	-	2	-2	Extrapolation
(5)	1	1	6	5	(4), (3)			Erfolg, Erfolg
(6)	1	2	6	3	(5)			Mißerfolg
(7)	1	2	6	7	(5)			Mißerfolg
(8)	2	0	10	3	(5)	2	-2	Extrapolation, Erfolg
(9)	2	1	12	3	(8)			Mißerfolg
(10)	2	1	8	3	(8)			Mißerfolg
(11)	2	2	10	1	(8)			Mißerfolg
(12)	2	2	10	5	(8)			Mißerfolg
(13)	3	0	14	1	-	2	-2	Extrapolation
(14)	3	1	16	1	(13)			Mißerfolg
(15)	3	1	12	1	(13), (8)			Erfolg, Mißerfolg
(16)	3	2	12	-1	(15)			Mißerfolg
(17)	3	2	12	3	(15)			Mißerfolg
(18)	4	0	10	3	-	2	-2	Rücksetzung
(19)	4	1	12	3	(8)			Mißerfolg
(20)	4	1	8	3	(8)			Mißerfolg
(21)	4	2	10	1	(8)			Mißerfolg
(22)	4	2	10	5	(8)			Mißerfolg
(23)	5	0	10	3	-	1	-1	Schrittweiten halbiert
(24)	5	1	11	3	(8)			Mißerfolg
(25)	5	1	9	3	(8)			Erfolg
(26)	5	2	9	2	(23), (8)			Erfolg, Erfolg
(27)	6	0	8	1	-	-1	-1	Extrapolation

Strategie von Hooke + Jeeves  
(pattern search) 1961



# Rosenbrock rotating coordinates (1960)

## Grobschema

- ① Start mit Koord.  $v_i$   $v_i^{(0)} = e_i$   
 feste Schrittweite

bei Erfolg: sofort realisieren und Schrittweite für nächste Runde vergrößern  
 $s_i^{(k+1)} = 3 s_i^{(k)}$

bei Mißerfolg: kein Versuch in Gegenri.  
 $s_i^{(k+1)} = -\frac{1}{2} s_i^{(k)}$

- ② Sobald in jeder  $R_i$  mindestens ein Erfolg und danach ein Mißerfolg registriert wurde (in der Regel erst nach mehreren Zyklen):

Orthogonalisierung: neuer Set von Richtungen

Sei  $d_j^{(k)}$  der Gesamtschritt in  $R_i$   $v_j^{(k)}$

$$\text{und } a_i = \sum_{j=i}^n d_j^{(k)} v_j^{(k)}$$

$$\text{dann } w_1 = a_1$$

$$w_i = a_i - \sum_{j=1}^{i-1} (a_i^T v_j^{(k+1)}) v_j^{(k+1)} \quad \forall i=2, \dots, n$$

$$v_i^{(k+1)} = w_i / \|w_i\|$$

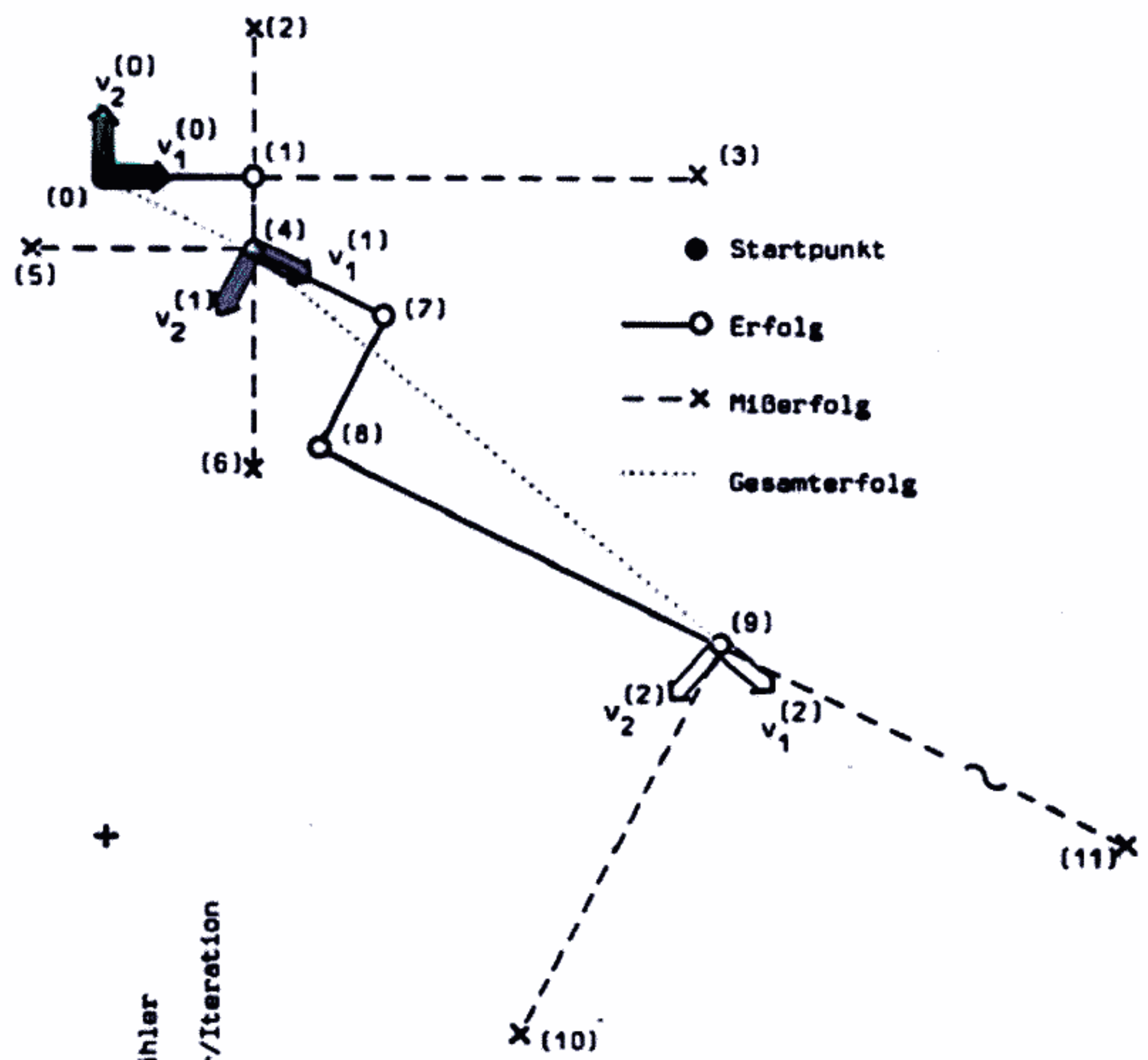
Gram-Schmidt-Orthon.

$O(n^3)$

- ③ Stop, wenn  $\|a_n\| < \varepsilon$  und  $\|a_2\| > 0.3 \|a_1\|$

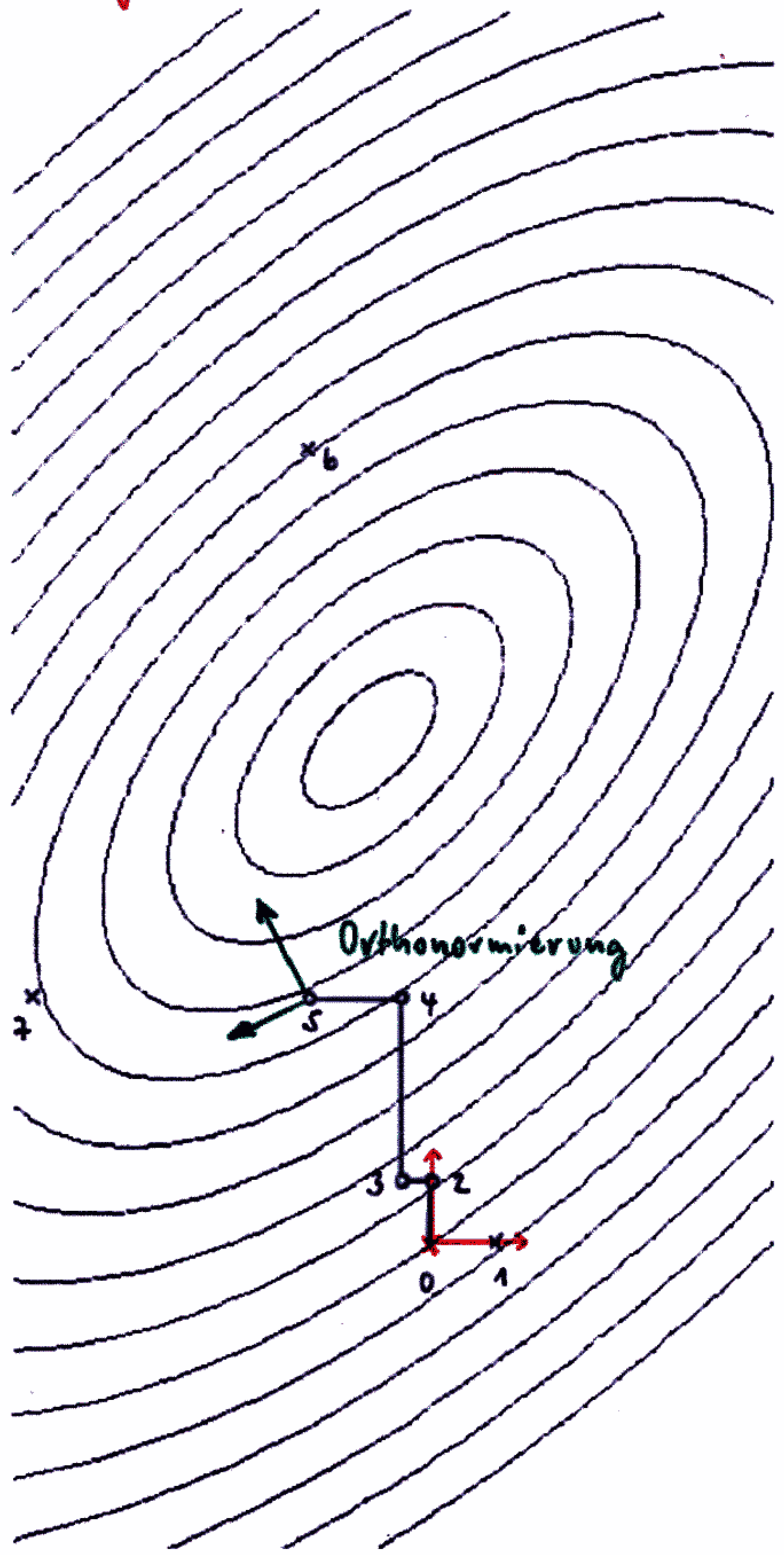
rotating coordinates  
1960

Skizze 3.7  
Strategie von Rosenbrock



Numerierung	Iterationszähler		Variablenwerte		Schrittweiten		Bemerkungen
	k	nl+1	x <sub>1</sub>	x <sub>2</sub>	s <sub>1</sub>	s <sub>2</sub>	
( 0)	0	0	0	9	2	2	Startpunkt
( 1)	0	1	2	9	2		Erfolg
( 2)	0	2	2	11		2	Mißerfolg
( 3)	0	3	8	9	6		Mißerfolg
( 4)	0	4	2	8		-1	Erfolg
( 5)	0	5	-1	8	-3		Mißerfolg
( 6)	0	6	2	5		-3	Mißerfolg
( 4)	1	0	2	8	2	2	Orthonormierung und Übertrag
( 7)	1	1	3.8	7.1	2		Erfolg
( 8)	1	2	2.9	5.3		2	Erfolg
( 9)	1	3	8.3	2.6	6		Erfolg
(10)	1	4	5.6	-2.7		6	Mißerfolg
(11)	1	5	24.4	-5.4	18		Mißerfolg
( 9)	2	0	8.3	2.6	2	2	Orthonormierung und Übertrag

Strategie von Rosenbrock  
(rotating coordinates search) 1960



# rotating coordinates II

vereinfachte Orthonormierung nach Palmer  $\sigma(m^2)$

mit line searcher :

Davies - Swann - Campsey (DSC)

Behandlung von Neb. in Form von Ungleichungen  
partielle innere Straffunktion

$$F'(x) = F(x) + \sum_{j=1}^m \phi_j(x) [f_j - F(x)]$$

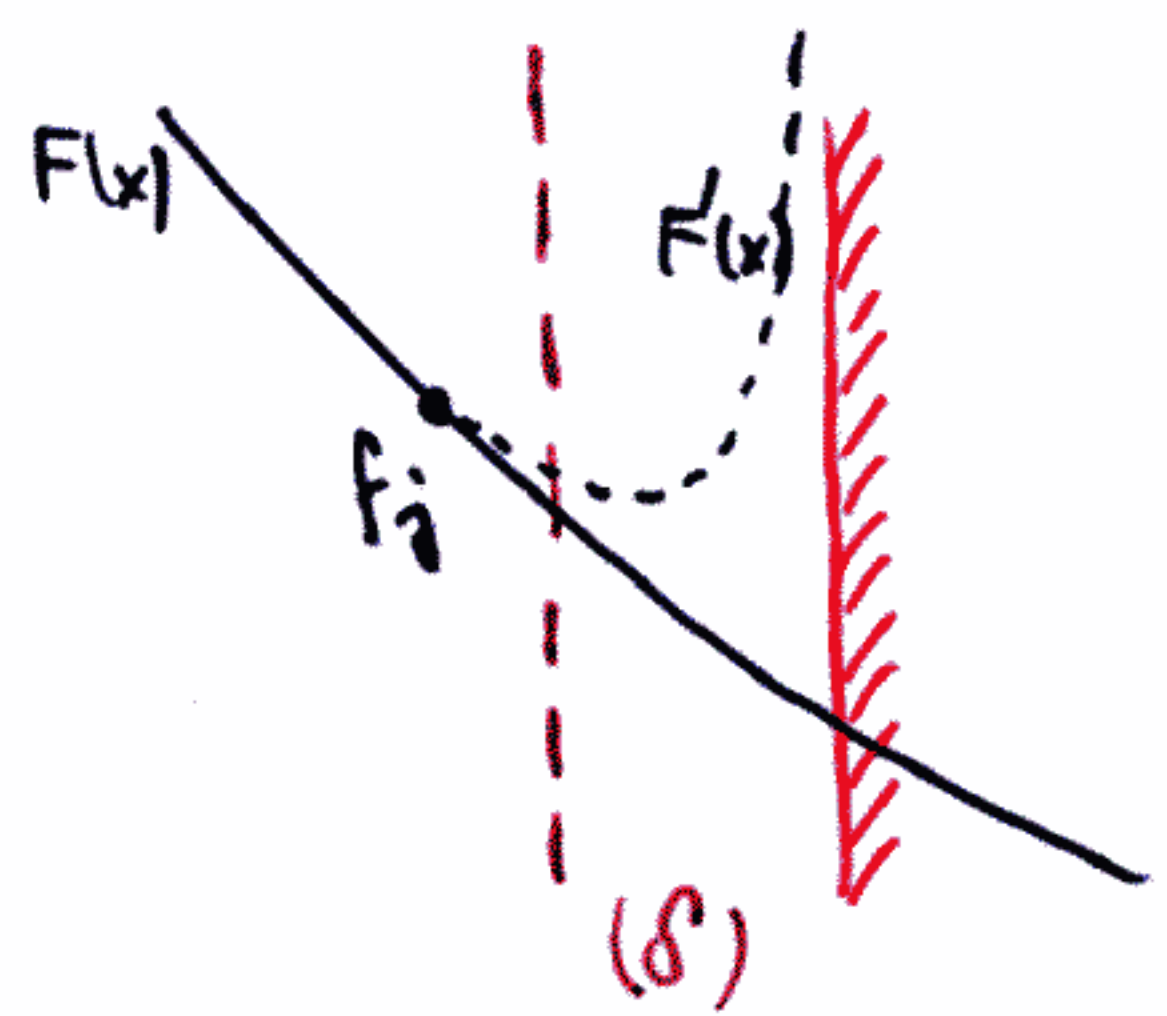
wobei

$$\phi_j(x) = \begin{cases} 0 & \text{wenn } g_j(x) \geq \delta \\ 3\eta - 4\eta^2 + 2\eta^3 & \text{wenn } 0 < g_j(x) < \delta \\ 1 & \text{wenn } g_j(x) \leq 0 \end{cases}$$

wenn  $g_j(x) \geq \delta$   
 $0 < g_j(x) < \delta$   
 $g_j(x) \leq 0$

$$\eta = 1 - \frac{1}{\delta} g_j(x)$$

$f_j = F(x)$  bei letztem Erfolg im Nicht-Grenzbereich der Restv.  $g_j(x)$

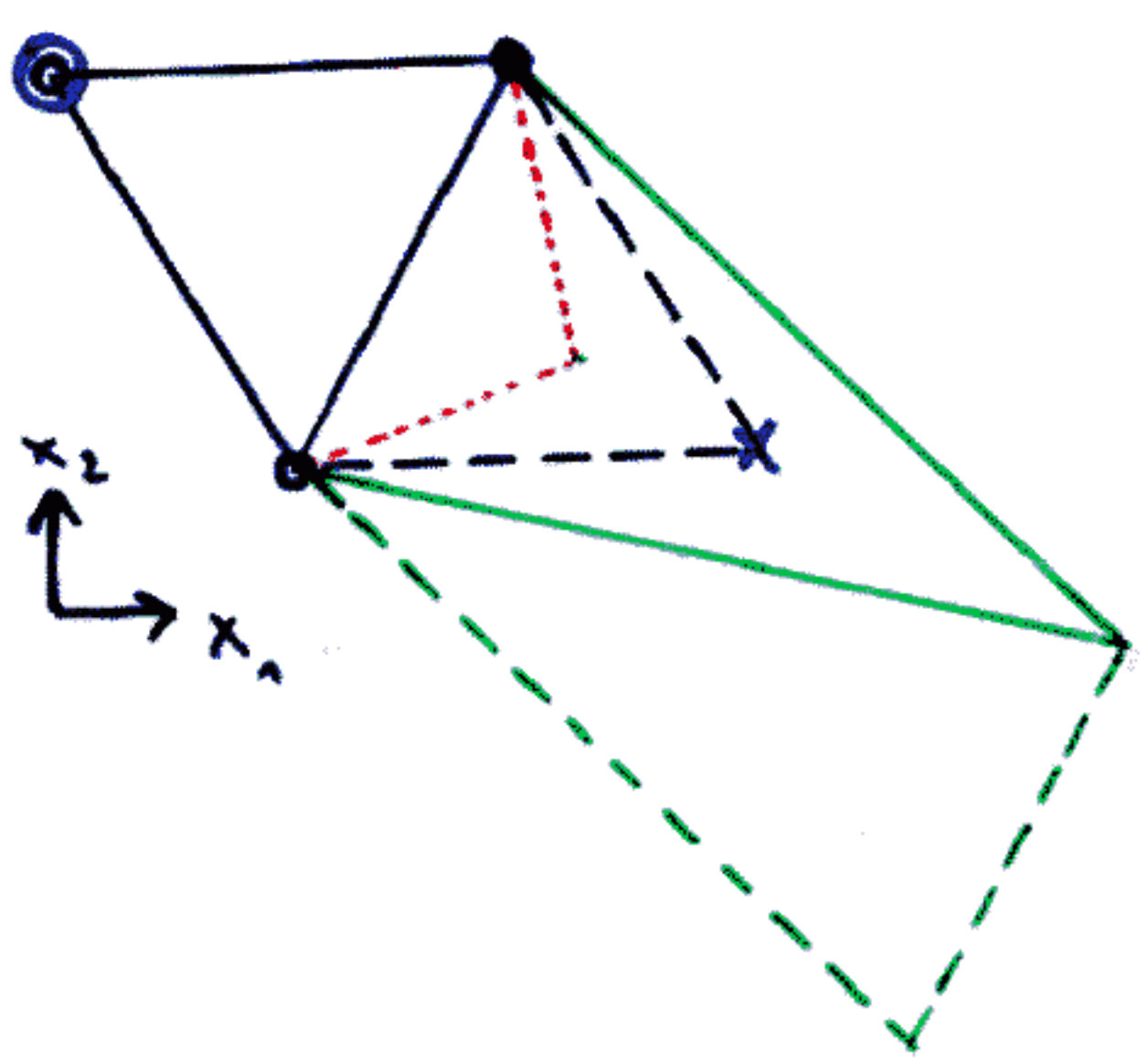


z.B.  $\delta = 10^{-4} f_j$

# Simplex - Strategie (Polyeder-M.)

Nelder + Mead  
M. J. Box

( $\Rightarrow$ )  $m+1$  Eckpunkte (N)



 Startsimplex

o schlechteste Ecke

x Reflektion  
(ggf. Expansion)  
(ggf. Kontraktion)

 neuer Simplex

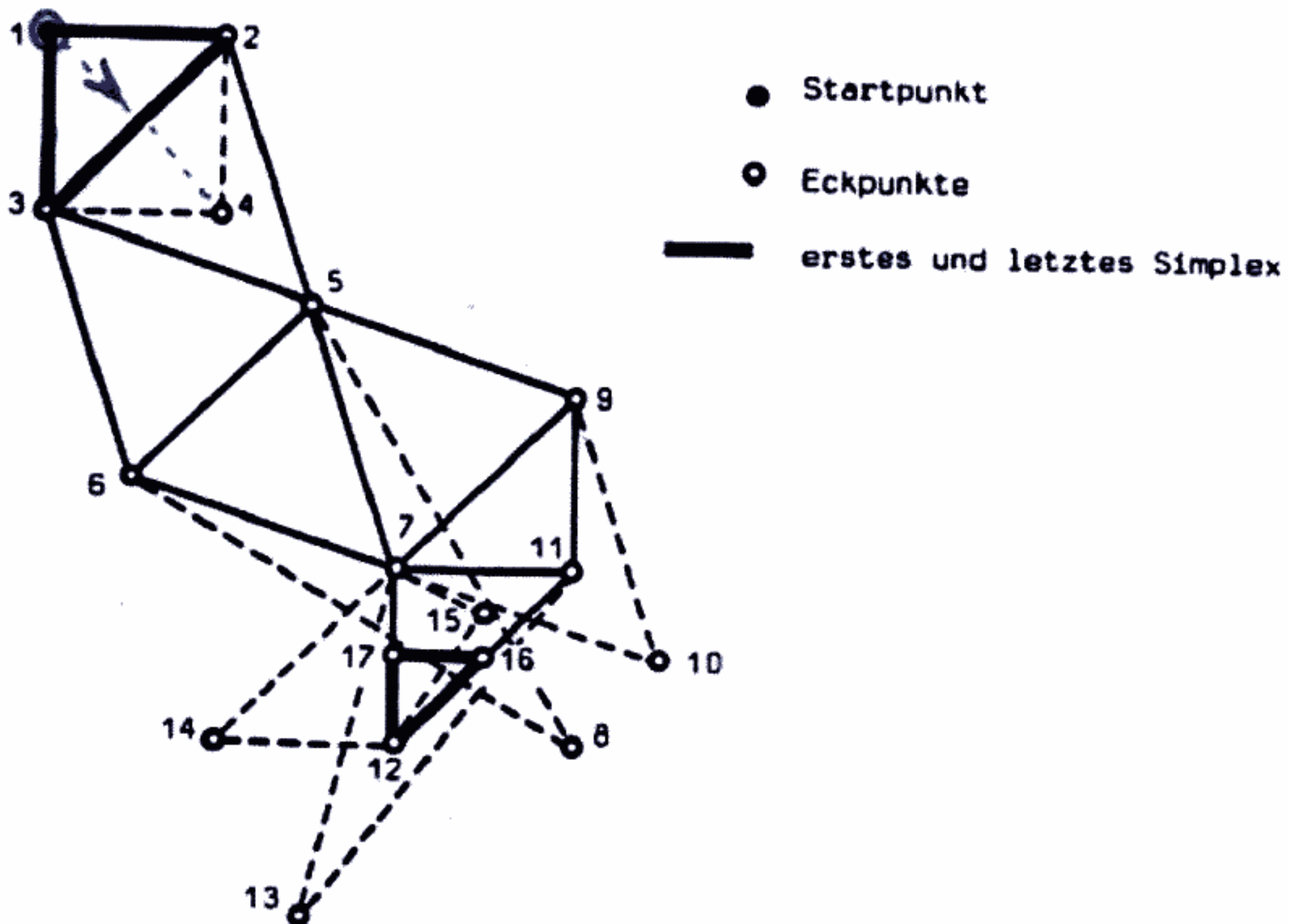
rein sequentielle Vorgehensweise  
mit Anpassung der Variabilität an Topologie

( $m+1$ ) - Strategie

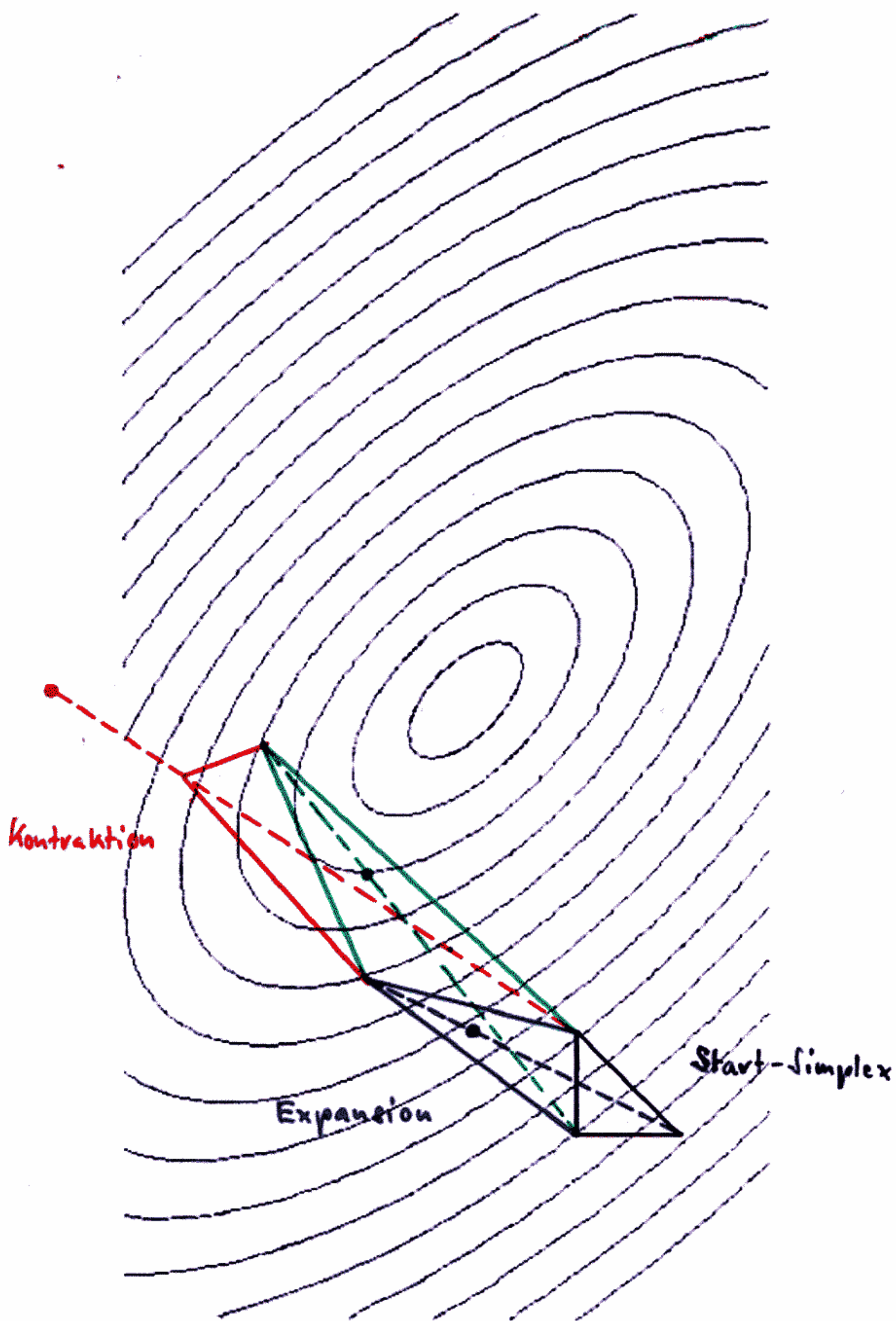


Simplex-Strategie von Nelder und Mead 1965

1965



Iterationszähler	Simplexecken				neue Eckpunkte
	schlechteste			beste	
0	1	2	3		Startsimplex
				4	Reflexion
1	2	3	5	5	Expansion (erfolgreich)
2	3	6	5		Reflexion
				7	Reflexion
3	6	5	7		Expansion (erfolglos)
4	5	9	7		Reflexion
		10			Reflexion
5	9	11	7		partielle Kontraktion außen
				12	Reflexion
6	11	7	12		Expansion (erfolglos)
	14				Reflexion
	15				partielle Kontraktion innen
7	17	16	12		totale Kontraktion



Strategie von Nelder + Mead

(Simplex-Strategie) 1965