UNIVERSITÄT DORTMUND

Sommersemester 2006

Mehrkriterielle Optimierung mit Metaheuristiken

(Vorlesung)

Prof. Dr. Günter Rudolph

Fachbereich Informatik

Lehrstuhl für Algorithm Engineering (LS XI)

Fachgebiet *Computational Intelligence*

Arbeitsdefinition:

Metaheuristik $=_{\text{def}}$

Algorithmischer Rahmen für eine Lösungsstrategie,
bei der viele Bestandteile initial unspezifiziert sind. ■

⇒ viele algorithmische Bestandteile können ausgetauscht werden, ohne die generelle Lösungsstrategie zu verändern

⇒ viele verschiedene Algorithmen

⇒ Aufbau einer Theorie mühsam

Beispiel: Monokriterieller (1+1) – EA („evolutionärer Algorithmus“)

wähle $x_0 \in X$ zufällig; setze $k = 0$

`repeat`

$y_k = x_k + m_k$ → Mutation

falls $f(y_k) \in B(f(x_k))$ dann $x_{k+1} = y_k$

sonst $x_{k+1} = x_k$

Selektion

$k = k + 1$

`until` Stoppkriterium erfüllt

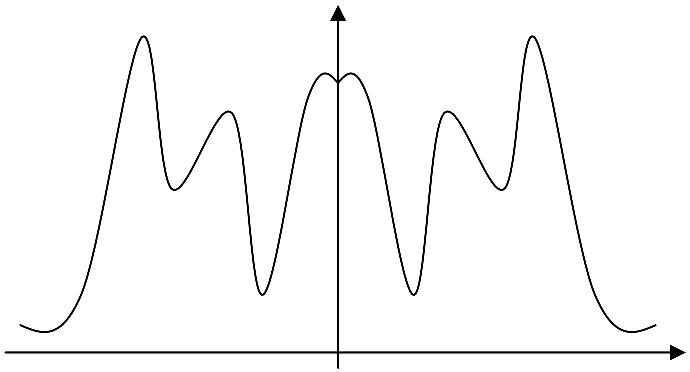
m_k : Zufallsvektor (hier: unspezifiziert)

$f : X \to \mathbb{R}$ Zielfunktion (→ min!)

$B(z^0) = \{z \in F : z < z^0\}$ Zielwerte, die besser als z^0 sind

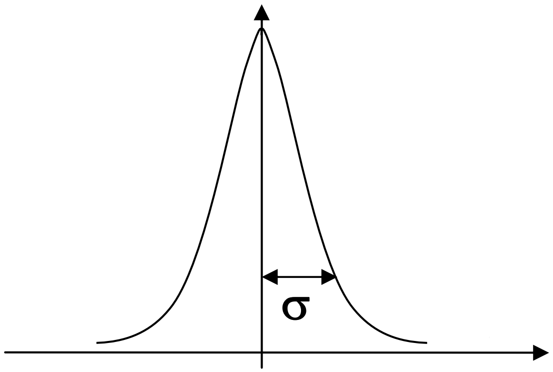
Forderungen an Such- / Mutationsverteilung von m_k

1. Keine Richtung ohne Grund bevorzugen → Symmetrie um 0
2. Kleine Änderungen wahrscheinlicher als große → Unimodal mit Modus 0
3. Steuerbar: Größe der Umgebung, Streuung → Parametrisierbar
4. Leicht erzeugbar
5. …



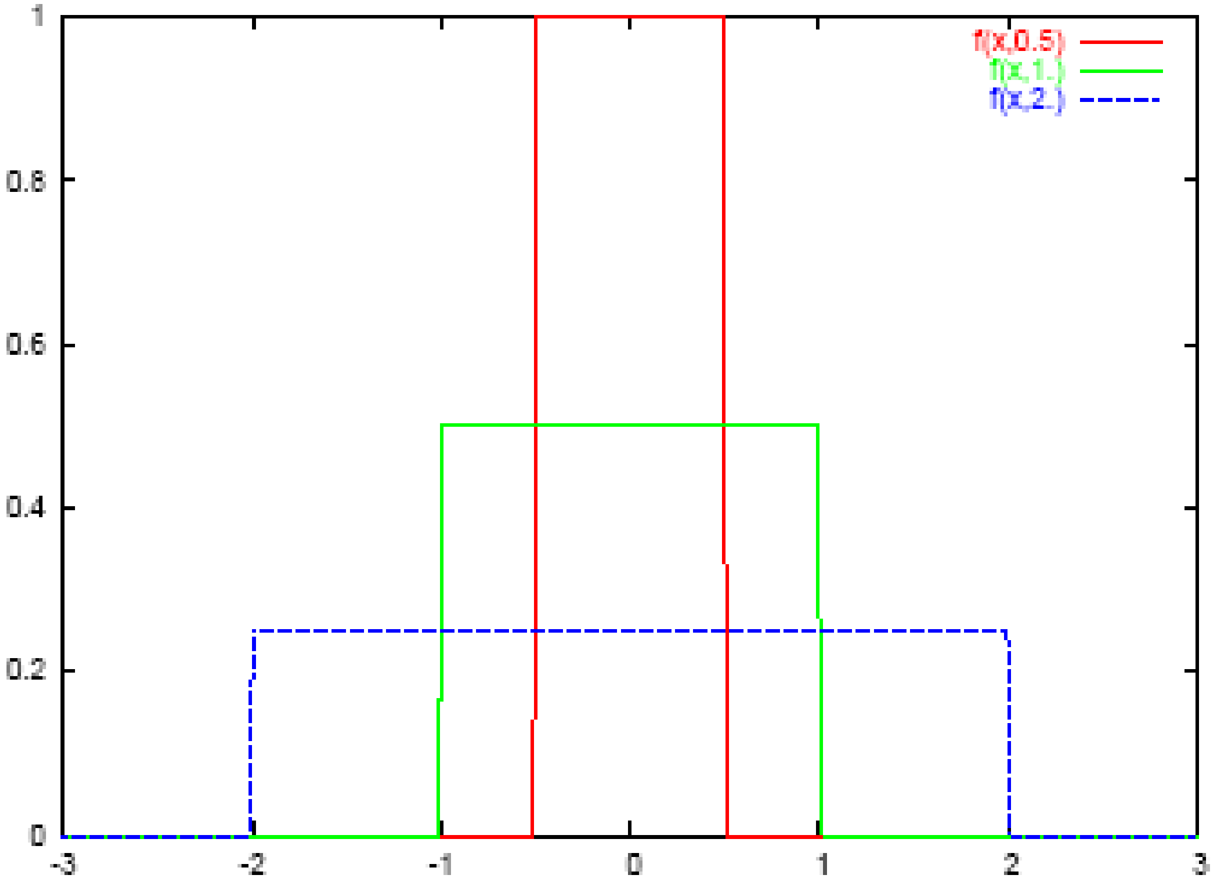
symmetrisch, multimodal

$\Rightarrow$



symmetrisch, unimodal

Gleichverteilung $$f_m(x) = \frac{1}{2\,r} \cdot 1_{(-r,r)}(x)$$

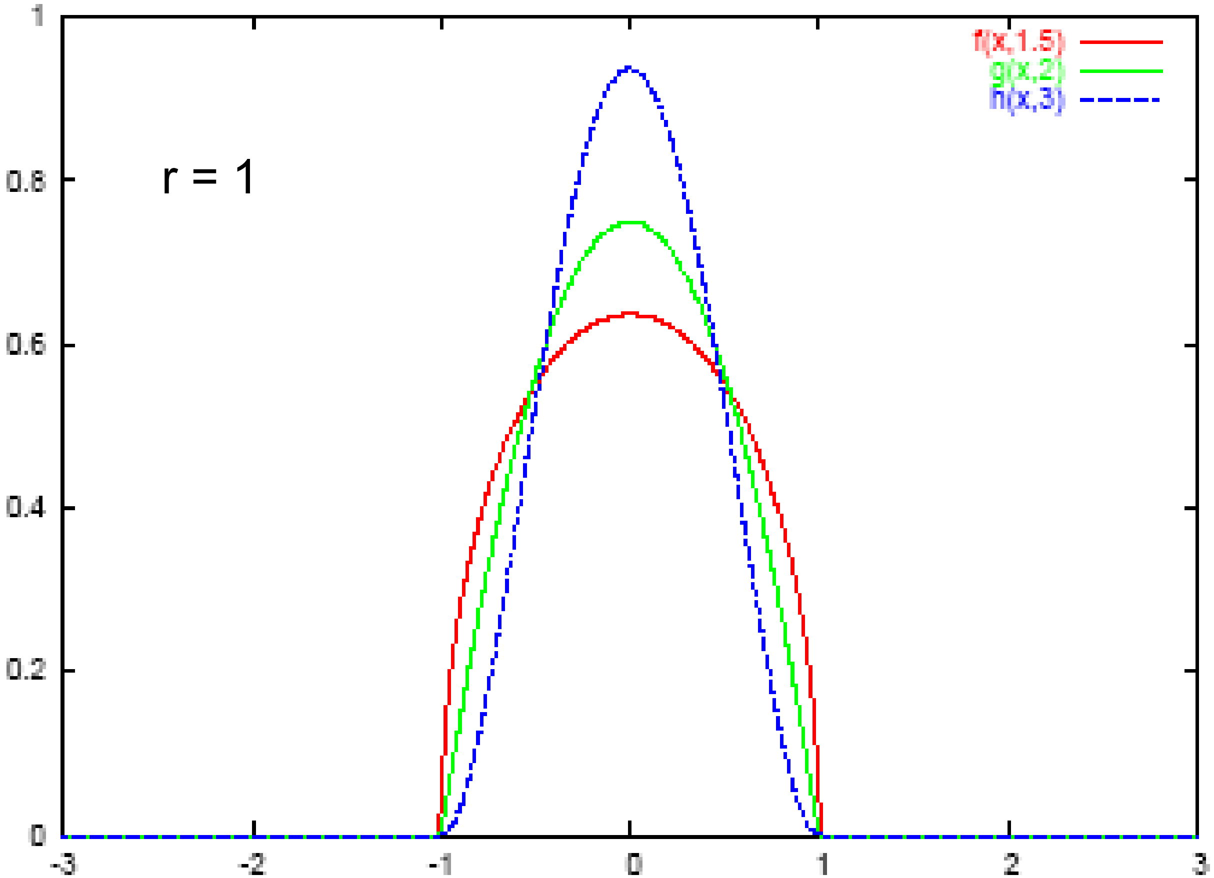


- symmetrisch
- unimodal
- steuerbar → r
- leicht erzeugbar:

m = r (2 u – 1)

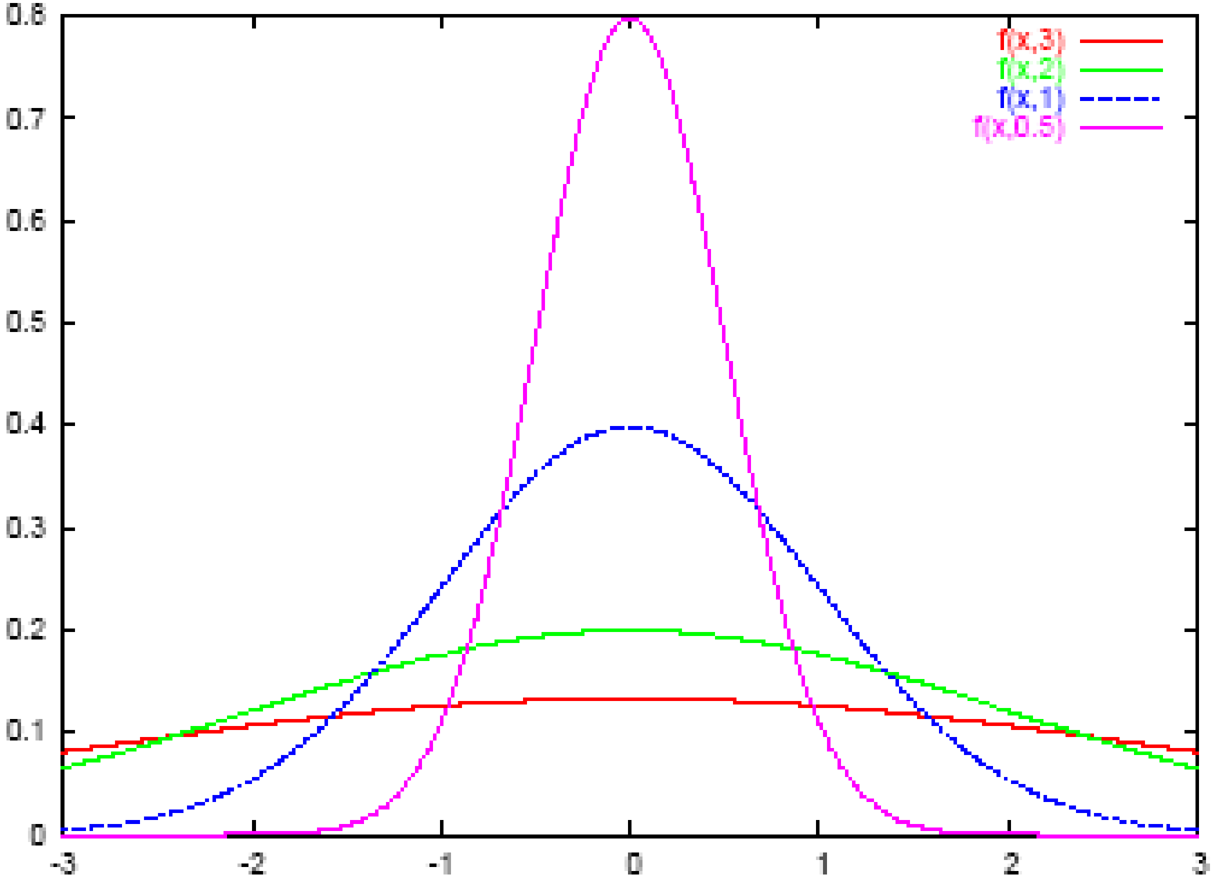
wobei u ∈ [0,1) gleichverteilt (aus Bibliothek)

Betaverteilung $$f_m(x) = \frac{r^{1-2p}}{\sqrt{\pi}} \cdot \frac{\Gamma(p+\frac{1}{2})}{\Gamma(p)} (1-x^2)^{p-1} \cdot 1_{(-r,r)}(x)$$



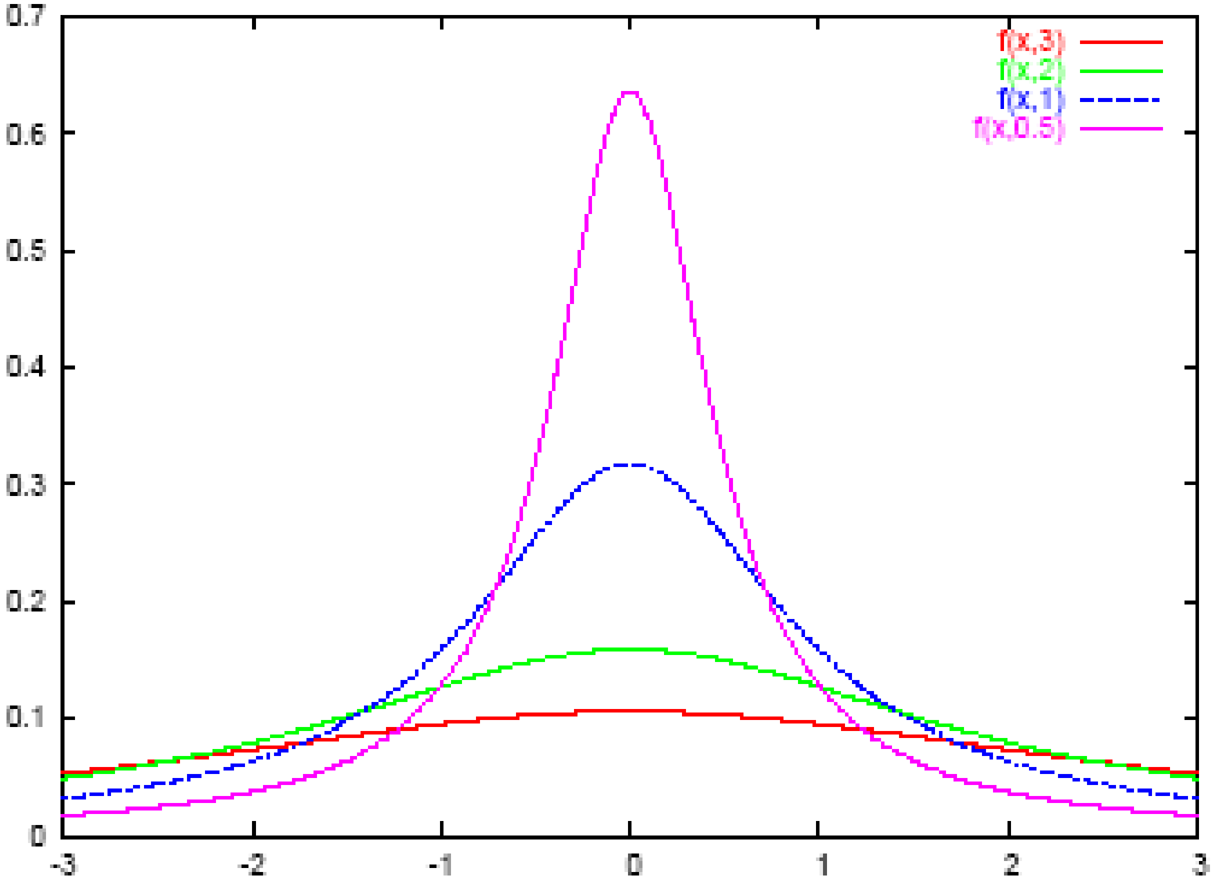
- symmetrisch
- unimodal
- steuerbar → r, p
- leicht erzeugbar (Bibliothek)

Normalverteilung $$f_m(x) = \frac{1}{\sigma\sqrt{2\,\pi}} \exp\left(-\frac{x^2}{2\,\sigma^2}\right)$$



- symmetrisch
- unimodal
- steuerbar → σ
- nicht ganz so leicht erzeugbar (Bibliothek)

Cauchyverteilung $$f_m(x) = \frac{1}{c\,\pi} \cdot \frac{1}{1 + (1 + \frac{x}{c})^2}$$



- symmetrisch
- unimodal
- steuerbar → c
- leicht erzeugbar (Bibliothek)

Besonderheit:

unendliche Varianz

Höherdimensionale Suchräume: Symmetrie? Unimodalität? Steuerbarkeit?

↓

Rotationssymmetrie

Definition 5.1:

Sei T eine (n x n)-Matrix mit T'T = I_n. (I_n: n-dim. Einheitsmatrix)

T heißt ***orthogonale Matrix*** oder ***Rotationsmatrix***. ■

Beispiel:

$$T = \begin{pmatrix} \cos\omega & -\sin\omega \\ \sin\omega & \cos\omega \end{pmatrix}$$

$y = T'x \quad \Rightarrow$ Vektor x wurde um Winkel ω gedreht

Definition 5.2:

n-dimensionaler Zufallsvektor x heißt

sphärisch symmetrisch oder ***rotationssymmetrisch***

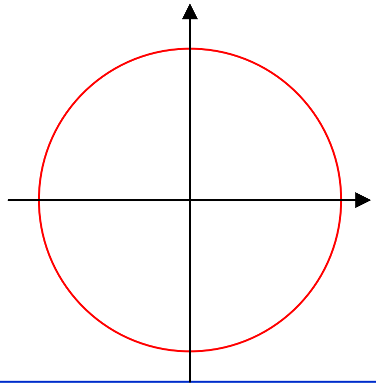
$\Leftrightarrow$ $x \stackrel{d}{=} T'x$ für jede orthogonale Matrix T. ■

$x \stackrel{d}{=} y$ bedeutet: x hat die gleiche Verteilung wie y

Beispiel: Gleichverteilung auf Kreis (Hyperkugel der Dimension n = 2)

u gleichverteilt in [0,1] $\Rightarrow$ $\omega = 2\pi u$

$$x \stackrel{d}{=} \begin{pmatrix} \cos\omega \\ \sin\omega \end{pmatrix}$$



Satz 5.1:

Zufallsvektor x rotationssymmetrisch ⇔ $x \overset{d}{=} r\, u^{(n)}$, wobei

r nichtnegative Zufallsvariable und

$u^{(n)}$ Zufallsvektor mit Gleichverteilung auf n-dim. Hyperkugelrand mit Radius 1. ■

Bemerkung:

r und $u^{(n)}$ sind stochastisch unabhängig, $u^{(n)} \overset{d}{=} \dfrac{x}{\| x \|}$

Erzeugung von rotationssymmetrischen Zufallsvektoren:

1. Wähle zufällige Richtung $u^{(n)}$
2. Wähle zufällige Schrittlänge r
3. Multiplikation: $x = r\, u^{(n)}$

Beispiel: Multivariate Normalverteilung

Zufallsvektor m erzeugbar via

1. $m = \sigma \cdot (m_1, m_2, \ldots, m_n)$,
   wobei $m_i \sim N(0,1)$ stoch. unabh., oder

2. $m = r \cdot u$, wobei $r \sim \chi_n(\sigma)$, $u \sim U(\partial S_n(1))$.

χ–Verteilung mit n Freiheitsgraden

Gleichverteilung auf Hyperkugelrand

$\partial S_n(r) = \{ x \in \mathbb{R}^n : || x || = r \}$ Hyperkugelrand

Beispiel: Multivariate Cauchyverteilung

Zufallsvektor m erzeugbar via

1. $m = \sigma \cdot (m_1, m_2, \ldots, m_n)/m_0$,
   wobei $m_i \sim N(0,1)$ stoch. unabh., oder

2. $m = r \cdot u$, wobei $r/n \sim F_{n,1}$, $u \sim U(\partial S_n(1))$.

F–Verteilung mit (n,1) Freiheitsgraden

Gleichverteilung auf Hyperkugelrand

Achtung:
Zufallsvektor aus n unabh. Cauchy-Zufallsvariablen <u>nicht</u> rotationssymmetrisch!

Multikriterieller (1+1)-EA: Was wäre zu ändern?

wähle $x_0 \in X$ zufällig; setze $k = 0$

`repeat`

$y_k = x_k + m_k$

falls $f(y_k) \in B(f(x_k))$ dann $x_{k+1} = y_k$

sonst $x_{k+1} = x_k$

$k = k + 1$

`until` Stoppkriterium erfüllt

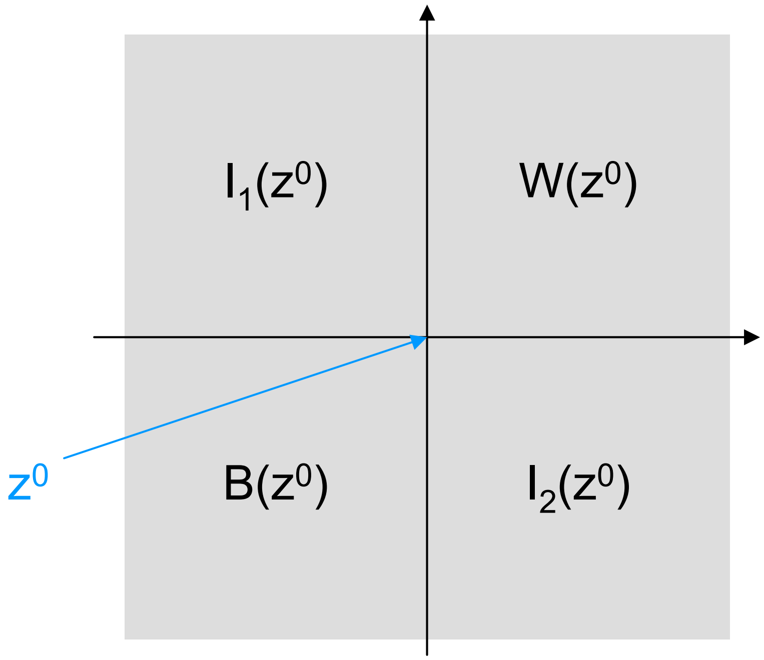
1. Mutationsvektoren $\rightarrow$ keine Änderung nötig!
2. Selektion $\rightarrow$ Neudefinition: $B(z^0) = \{ z \in F : z \preceq z^0 \}$

$\Rightarrow$ keine Änderung an der Rahmenstruktur des Algorithmus!

$B(z^0) = \{ z \in F : z \preceq z^0 \}$ “better than“ z^0

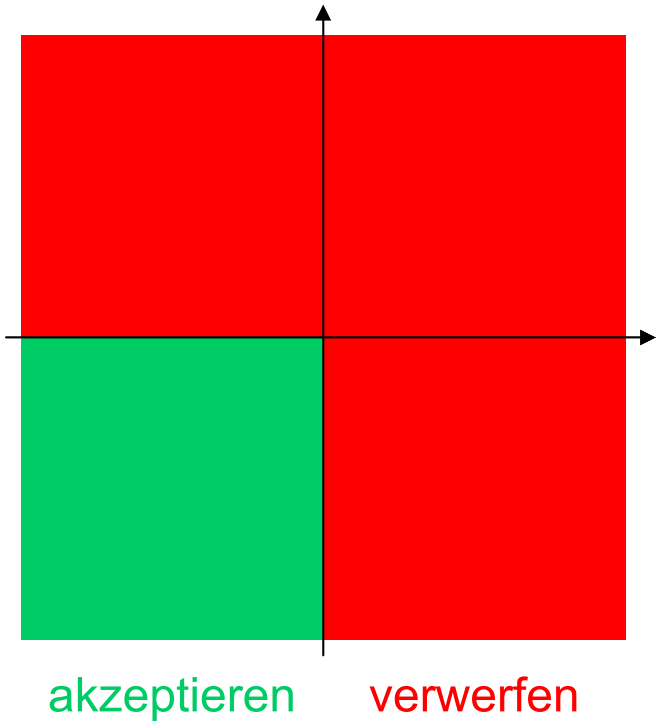
$I(z^0) = \{ z \in F : z \parallel z^0 \}$ “incomparable to“ z^0

$W(z^0) = \{ z \in F : z^0 \preceq z \}$ “worse than“ z^0



$\Rightarrow$

Multikriterieller (1+1)-EA:



Beispiel: Monokriterielles Threshold Accepting (TA)

wähle $x_0 \in X$ zufällig; setze $k = 0$

```
repeat
```

$y_k = x_k + m_k$

falls $f(y_k) \in B(f(x_k) + \boxed{T_k})$ dann $x_{k+1} = y_k$

sonst $x_{k+1} = x_k$

$\boxed{T_{k+1} = \gamma(T_k),}$ $k = k + 1$

`until` Stoppkriterium erfüllt

m_k : Zufallsvektor (hier: unspezifiziert)

$f : X \to \mathbb{R}$ Zielfunktion (→ min!)

$B(z^0) = \{z \in F : z < z^0\}$ Zielwerte, die besser als z^0 sind

Threshold $T_k \geq 0$

T_k monoton fallend → 0

Idee bei monokriteriellem Threshold Accepting (TA)

- Akzeptiere jede Verbesserung
- Akzeptiere zu Anfang der Suche auch große Verschlechterungen
- Akzeptiere danach immer geringere Verschlechterungen

⇒ am Ende nahezu keine Akzeptanz von Verschlechterungen mehr

⇒ Verlassen von lokalen Optima durch Akzeptanz von Verschlechterungen

Regeln für T_k:

$$\left.\begin{array}{l} T_k = c\, T_{k-1},\ c \in (0,1) \\ T_k = T_0 \,/\, (k+1)^{\beta}\ ,\ \beta > 0 \\ T_k = T_0 \,/\, \log(k+1) \end{array}\right\}$$

weitere Regeln denkbar
und wohl auch im Einsatz

Multikriterielles Threshold Accepting: Was wäre zu ändern?

wähle $x_0 \in X$ zufällig; setze $k = 0$

`repeat`

$y_k = x_k + m_k$

falls $f(y_k) \in B(f(x_k) + T_k)$ dann $x_{k+1} = y_k$

sonst $x_{k+1} = x_k$

$T_{k+1} = \gamma(T_k)$, $k = k + 1$

`until` Stoppkriterium erfüllt

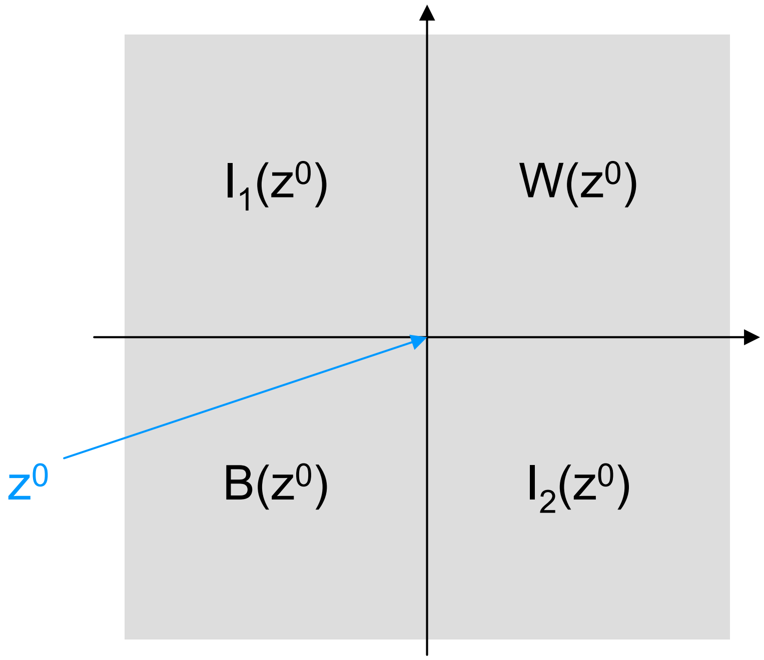
1. T_k wird m-dimensionaler Vektor (ein Threshold je Zielgröße oder einer für alle)

2. ggf. m verschiedene T_k – Verringerungsregeln

$B(z^0) = \{ z \in F : z \preceq z^0 \}$ “better than“ z^0

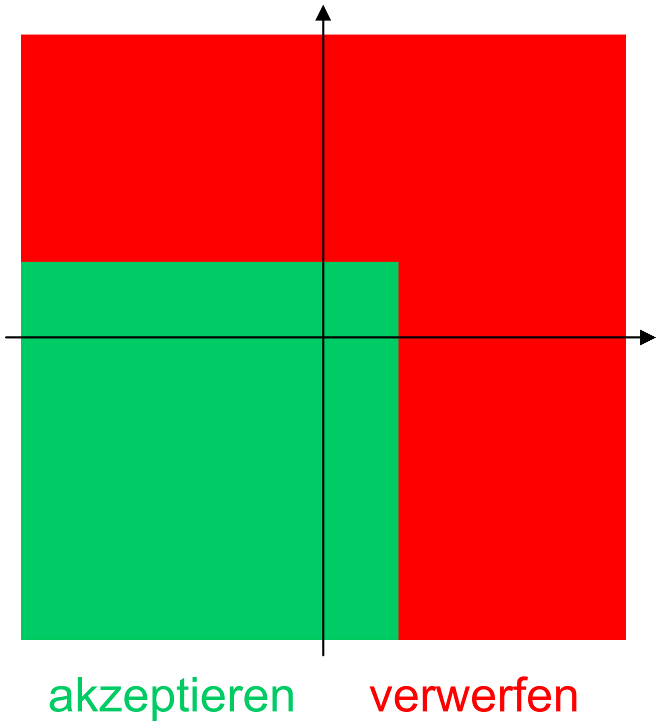
$I(z^0) = \{ z \in F : z \parallel z^0 \}$ “incomparable to“ z^0

$W(z^0) = \{ z \in F : z^0 \preceq z \}$ “worse than“ z^0



$\Rightarrow$

<u>Multikriterielles TA:</u>



Alternative Akzeptanzregeln für multikriterielles TA

Sei $\delta_i = f_i(y_k) - f_i(x_k)$.

$\delta_i > 0 \Rightarrow$ Verschlechterung bzgl. Ziel i

$\delta_i < 0 \Rightarrow$ Verbesserung bzgl. Ziel i

1. $\max\{ \delta_i : i = 1,\ldots,m\} \leq T_k$ $\longrightarrow$ $f(y_k) \in B(f(x_k) + T_k \cdot (1,\ldots,1)')$
2. $\sum_i w_i \max\{ \delta_i, 0 \} \leq T_k$
3. $\sum_i w_i \delta_i \leq T_k$
4. …

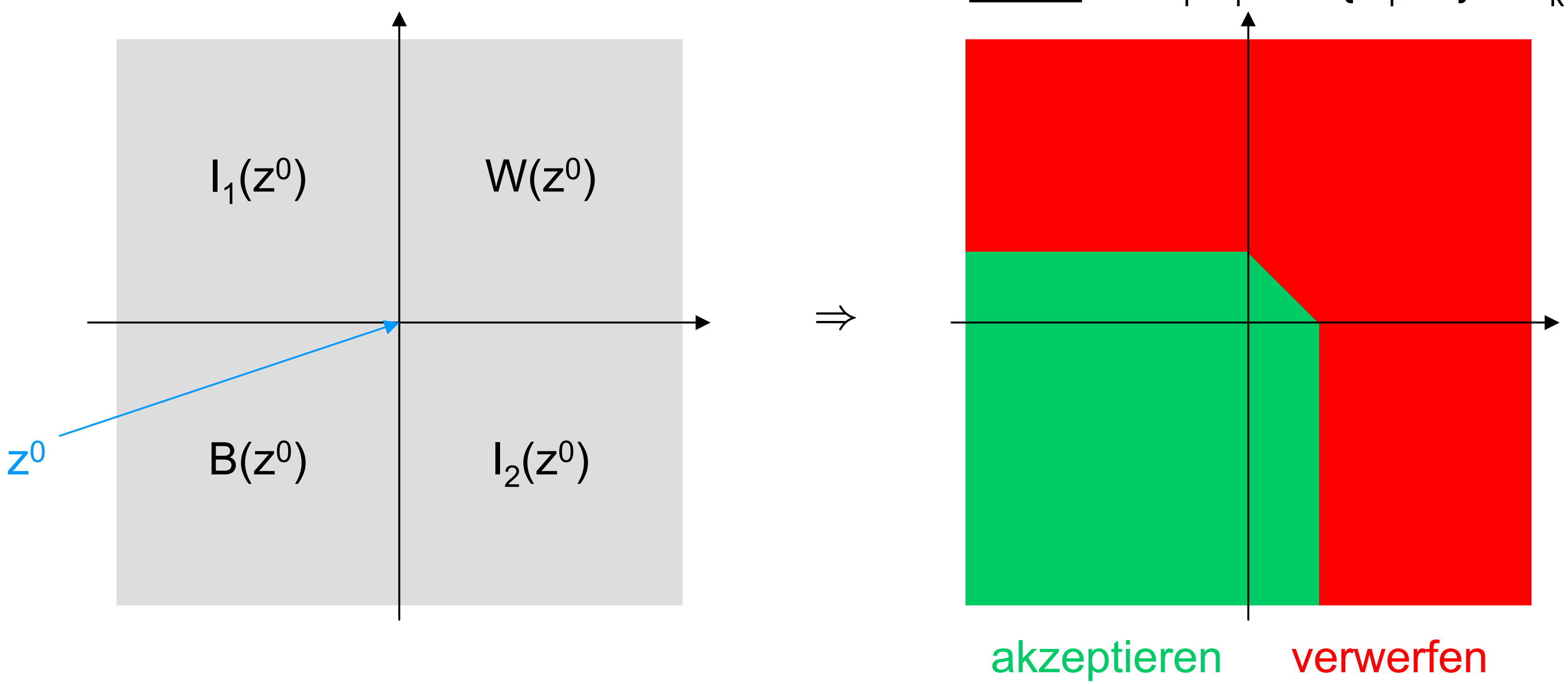
(2., 3.:) Konvexkombination $w_1,\ldots w_m$

$B(z^0) = \{ z \in F : z \preceq z^0 \}$ "better than" z^0

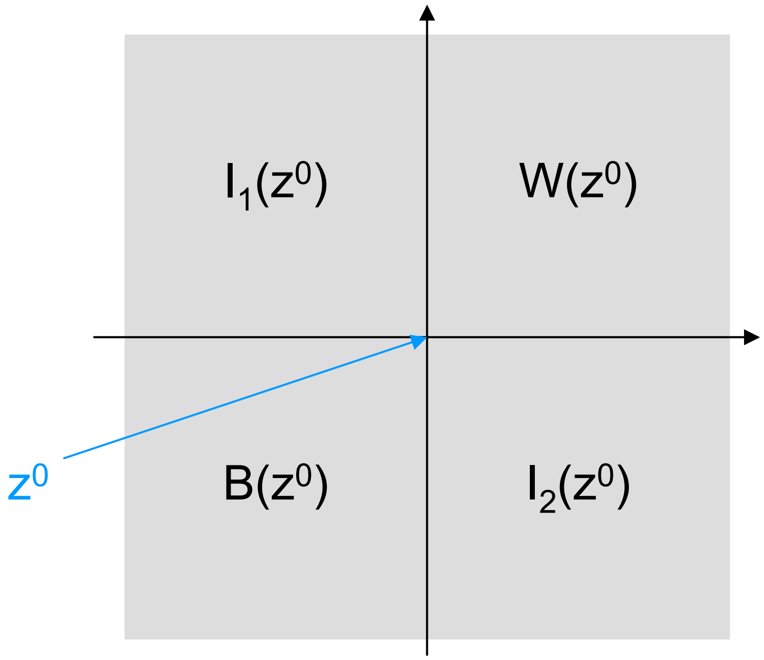
$I(z^0) = \{ z \in F : z \,||\, z^0 \}$ "incomparable to" z^0

$W(z^0) = \{ z \in F : z^0 \preceq z \}$ "worse than" z^0

MTA: $\sum_i w_i \max\{ \delta_i, 0 \} \le T_k$

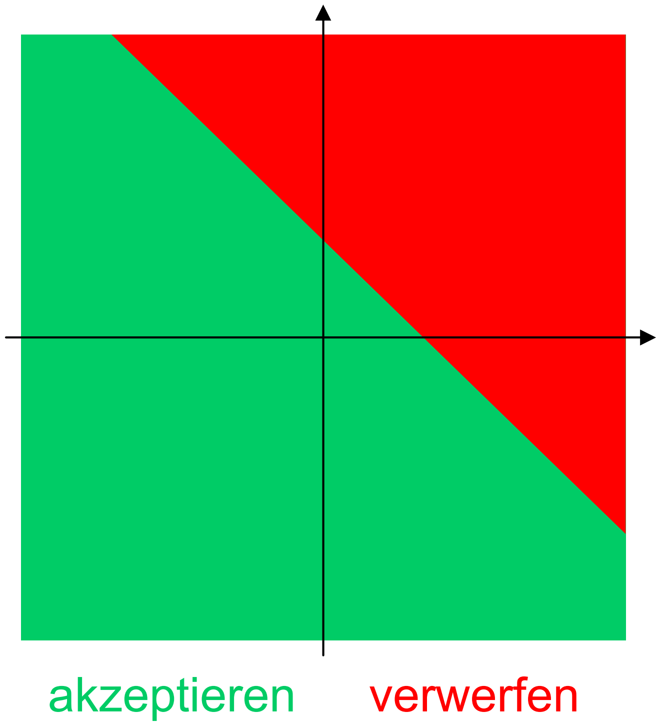


$B(z^0)$	$= \{ z \in F : z \preceq z^0 \}$	"better than" z^0
$I(z^0)$	$= \{ z \in F : z \parallel z^0 \}$	"incomparable to" z^0
$W(z^0)$	$= \{ z \in F : z^0 \preceq z \}$	"worse than" z^0



$\Rightarrow$

MTA: $\sum_i w_i \delta_i \leq T_k$



Beispiel: Monokriterielles Simulated Annealing (SA)

wähle $x_0 \in X$ zufällig; setze $k = 0$

`repeat`

$y_k = x_k + m_k$

falls $f(y_k) \in B(f(x_k))$ dann $x_{k+1} = y_k$

falls $f(y_k) \in W(f(x_k))$

falls $u < exp(-\Delta f_k / T_k)$ dann $x_{k+1} = y_k$

sonst $x_{k+1} = x_k$

$T_{k+1} = \gamma(T_k)$; $k = k + 1$

`until` Stoppkriterium erfüllt

m_k : Zufallsvektor

u ~ U[0,1], $T_k \to 0$ monoton fallend

$B(z^0) = \{z \in F : z < z^0\}$ Zielwerte, die besser als z^0 sind

$W(z^0) = \{z \in F : z > z^0\}$ Zielwerte, die schlechter als z^0 sind

Idee beim monokriteriellem Simulated Annealing:

Akzeptiere auch schlechtere Lösungen mit abnehmender Wahrscheinlichkeit, um aus lokalen Optima zu entkommen!

Satz (Hayek 1988, Haario & Saksman 1991)

$T_k = T_0 / \log(k + 1)$, m_k mit beschränktem Träger
$\Rightarrow$ SA konvergiert gegen globales Optimum mit W'keit 1 ■

$\rightarrow$ sehr langsame Temperaturverringerung!

$\rightarrow$ Praxis: $T_{k+1} = c \cdot T_k$ mit $c \in (0,1)$

Satz (Belisle 1992)

$T_k = c^k \cdot T_0$, $c \in (0,1)$ und $\text{supp}(m_k) = \mathbb{R}^n$
$\Rightarrow$ SA konvergiert zum globalen Optimum mit W'keit 1 ■

Multikriterielles Simulated Annealing: Was wäre zu ändern?

wähle $x_0 \in X$ zufällig; setze $k = 0$

repeat

$y_k = x_k + m_k$

falls $f(y_k) \in B(f(x_k))$ dann $x_{k+1} = y_k$

falls $f(y_k) \in W(f(x_k)) \cup I(f(x_k))$

falls $u < exp(-\Delta_k / T_k)$ dann $x_{k+1} = y_k$

sonst $x_{k+1} = x_k$

$T_{k+1} = \gamma(T_k)$; $k = k + 1$

until Stoppkriterium erfüllt

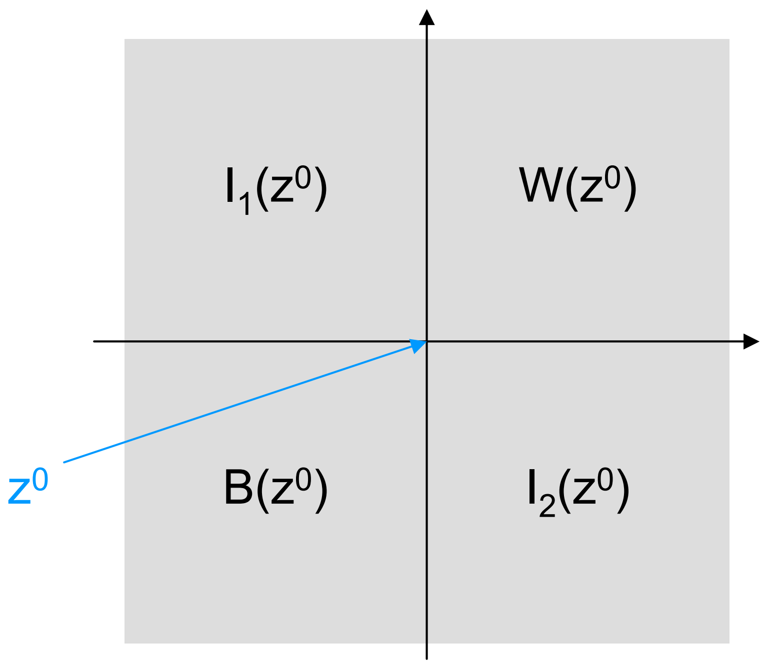
$\delta_i(k) = f_i(y_k) - f_i(x_k)$

$\Delta_k = \max\{ \delta_i(k) : i = 1,..m \}$

$B(z^0) = \{z \in F : z \preceq z^0\}$ Zielwerte, die besser als z^0 sind

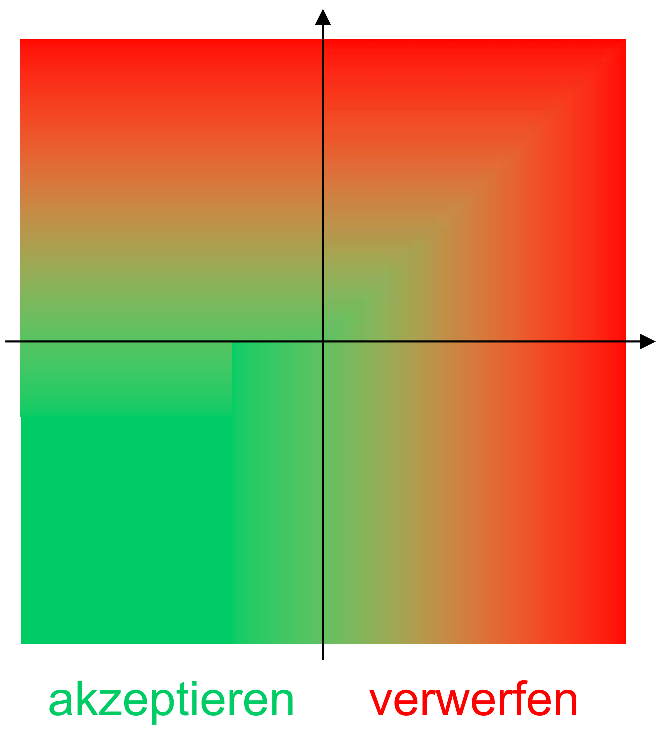
$W(z^0) = \{z \in F : z \succeq z^0\}$ Zielwerte, die schlechter als z^0 sind

$B(z^0)$	$= \{ z \in F : z \preceq z^0 \}$	"better than" z^0
$I(z^0)$	$= \{ z \in F : z \parallel z^0 \}$	"incomparable to" z^0
$W(z^0)$	$= \{ z \in F : z^0 \preceq z \}$	"worse than" z^0



$\Rightarrow$

Multikriterielles SA:



Multikriterielles Simulated Annealing: Serafini (1992)

wähle $x_0 \in X$ zufällig; setze $k = 0$

repeat

$y_k = x_k + m_k$

falls $f(y_k) \in B(f(x_k))$ dann $x_{k+1} = y_k$

sonst

falls $u < \boxed{\vartheta(\delta(k), T_k)}$ dann $x_{k+1} = y_k$

sonst $x_{k+1} = x_k$

$T_{k+1} = \gamma(T_k)$; $k = k + 1$

until Stoppkriterium erfüllt

$\vartheta: \mathbb{R}^m \times \mathbb{R}_+ \rightarrow [0,1]$

$\delta(k) = (\delta_1(k), \ldots, \delta_m(k))'$ mit $\delta_i(k) = f_i(y_k) - f_i(x_k)$

$\mathrm{supp}(m_k)$ kann beschränkt sein

MSA (Serafini 1992)

Sei $f(y_k) \notin B(f(x_k))$, so dass $\exists i : \delta_i(k) > 0$.

$$\vartheta(\delta(k), T_k) = \begin{cases} \exp\left(\sum\limits_{i=1,\delta_i(k)>0}^{m} w_i \, \delta_i(k) / T_k \right) \\ \prod\limits_{i=1,\delta_i(k)>0}^{m} \exp(-\delta_i(k)/T_k) \\ \min\limits_{i=1,\ldots,m} \{\exp(-w_i \, \delta_i(k)/T_k) : \delta_i(k) > 0\} \\ \max\limits_{i=1,\ldots,m} \{\exp(-w_i \, \delta_i(k)/T_k) : \delta_i(k) > 0\} \end{cases}$$

zzgl. zusammengesetzte Regeln: $\vartheta(\cdot) = \alpha \, \vartheta_1(\cdot) + (1-\alpha) \, \vartheta_2(\cdot)$

MSA (Engrand et al. 1998)

Ersatzzielfunktion $$f(x) = \sum_{i=1}^{m} \log f_i(x)$$

Skalarisierung!

$\Downarrow$

vermutlich nicht alle Lösungen erreichbar!

→ erfordert ∀ i : ∀ x: $f_i(x) > 0$

$$\begin{aligned}
\Delta f_k &= f(y) - f(x) \\
&= \sum_{i=1}^{m} \log f_i(y) - \sum_{i=1}^{m} \log f_i(x) \\
&= \sum_{i=1}^{m} (\log f_i(y) - \log f_i(x)) \\
&= \sum_{i=1}^{m} \log \left(\frac{f_i(y)}{f_i(x)} \right)
\end{aligned}$$

MSA (Engrand et al. 1998)

Ersatzzielfunktion $$f(x) = \sum_{i=1}^{m} \log f_i(x)$$

↓

Ansatz: $g_i(x) = \exp(f_i(x))$ ⟶ $\forall\, i : \forall\, x : g_i(x) > 0$

↓

lokale und globale Minimalstellen identisch!

Ersatzzielfunktion $$g(x) = \sum_{i=1}^{m} g_i(x) = m \sum_{i=1}^{m} w_i\, g_i(x)$$ mit $w_i = 1/m$

⇒ kann nur Pareto-optimale Lösungen im konvexen Zielbereich finden!

Beispiel: Monokriterielles Tabu Search (TS)

wähle $x_0 \in X$ zufällig; setze $L_k = ()$ und $k = 0$

```
repeat
```

$$x_{k+1} = \text{argmin}\{f(x) : x \in N(x_k) \setminus L_k\}$$

$$L_{k+1} = \text{tail}(\text{append}(L_k, x_k), \ell)$$

$$k = k + 1$$

```
until
```
Stoppkriterium erfüllt

Tabu-Liste L_k mit max. Länge ℓ

Nachbarschaft N(x) von $x \in X$

zugeschnitten für diskrete Probleme

→ auch schlechtere neue Punkte werden akzeptiert!

Bisher: Ein Lauf → ein Lösungskandidat

Erwünscht: Approximation der Paretomenge

Ansatz: Einsammeln nicht-dominierter Lösungen

→ Knowles & Corne (2000): Pareto-Archived ES (PAES)

(1+1)-PAES (Knowles & Corne 2000)

wähle $x_0 \in X$ zufällig; $A_0 = \{x_0\}$; setze $k = 0$

`repeat`

$y_k = x_k + m_k$

falls $f(y_k) \in B(f(x_k))$ dann $x_{k+1} = y_k$; $A_k = A_k \cup \{y_k\}$

sonst falls $f(y_k) \in W(f(x_k))$ dann $x_{k+1} = x_k$

sonst falls $\exists a \in A_k : f(y_k) \in W(f(a))$ dann $x_{k+1} = x_k$

sonst $x_{k+1} = \texttt{replace}(A_k, x_k, y_k)$

$k = k + 1$

`until` Stoppkriterium erfüllt

A_k : Archiv zum Zeitpunkt k

(1+1)-PAES (Knowles & Corne 2000)

$\texttt{replace}(A_k, x_k, y_k) :=$

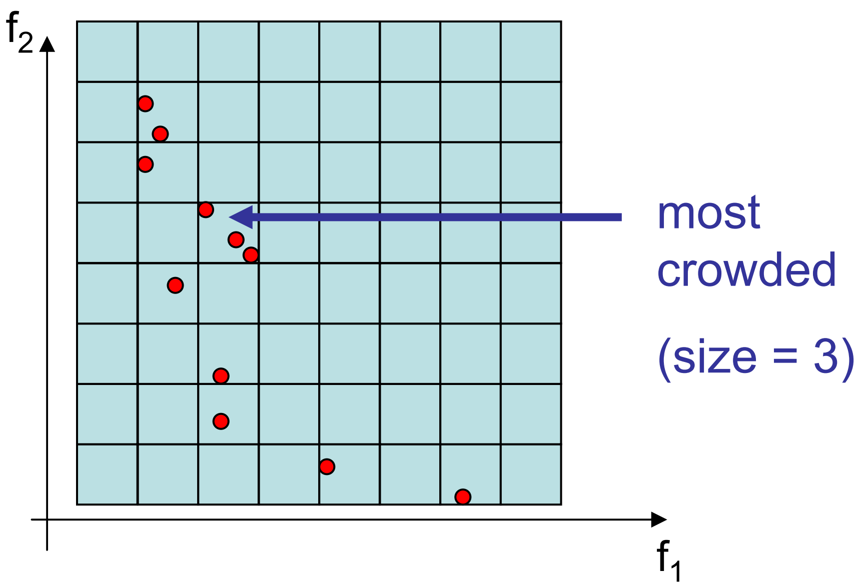
falls A_k nicht voll
    $A_k = A_k \cup \{y_k\}$
    falls $\texttt{lessCrowded}(y_k, x_k; A_k)$ dann $\texttt{return}\ y_k$
    sonst $\texttt{return}\ x_k$
sonst
    falls $\exists a \in A_k : \texttt{lessCrowded}(y_k, a; A_k)$
        $A_k = A_k \cup \{y_k\}; A_k = A_k \setminus \{a\}$
    falls $\texttt{lessCrowded}(y_k, x_k; A_k)$ dann $\texttt{return}\ y_k$
    sonst $\texttt{return}\ x_k$

(1+1)-PAES (Knowles & Corne 2000)

boolean lessCrowded(y, x; A):=

  return cell(y; A).size() $\leq$ cell(x; A).size();

Idee: Zielraum unterteilen in Raster mit $2^{d \cdot s}$ Zellen (cells); d = dim(F)



Speichereffiziente Implementierung möglich
→ **quadtree**