



Adaptive Systeme

Evolutionäre Algorithmen:
Überlebenskampf und Evolutionäre Strategien

Prof. Dr. rer. nat. Nikolaus Wulff

Einleitung

Adaptive Filter

Künstliche neuronale Netze

Adaptive Vektorquantisierung

Evolutionäre Algorithmen

Genetische Algorithmen



Frühe Pioniere

Evolutionsstrategie

I. Rechenberg, H.-P. Schwefel

Evolutionäre Programmierung

L.J. Fogel

Genetische Algorithmen

J.H. Holland

Genetische Programmierung

J.R. Koza

Reelle Optimierungsprobleme

Lokales Optimum

Gegeben sei eine Funktion $f : \mathbb{R}^N \rightarrow \mathbb{R}$. Der Vektor $\vec{x}_o \in \mathbb{R}^N$ wird **lokales Minimum** genannt, wenn gilt:

$$\exists \epsilon > 0 : \forall \vec{x} \in \mathbb{R}^N : \|\vec{x} - \vec{x}_o\| < \epsilon \Rightarrow f(\vec{x}_o) \leq f(\vec{x})$$

Der Vektor $\vec{x}_o \in \mathbb{R}^N$ wird **lokales Maximum** genannt, wenn gilt:

$$\exists \epsilon > 0 : \forall \vec{x} \in \mathbb{R}^N : \|\vec{x} - \vec{x}_o\| < \epsilon \Rightarrow f(\vec{x}_o) \geq f(\vec{x})$$

Ist $\vec{x}_o \in \mathbb{R}^N$ ein lokales Minimum oder ein lokales Maximum, so wird \vec{x}_o ein **lokales Optimum** genannt.



Reelle Optimierungsprobleme

Beschränkte Optimierung

Gegeben sei eine Funktion $f : \mathbb{R}^N \rightarrow \mathbb{R}$. Optimiere f unter den M Nebenbedingungen $g_i(\vec{x}) \geq 0$ mit $1 \leq i \leq M$.

Die Funktionen $g_i : \mathbb{R}^N \rightarrow \mathbb{R}$ heißen **Beschränkungen** (*constraint*). An einem Punkt $\vec{x} \in \mathbb{R}^N$ ist die Beschränkung

erfüllt, wenn $g_i(\vec{x}) \geq 0$,

aktiv, wenn $g_i(\vec{x}) = 0$,

inaktiv, wenn $g_i(\vec{x}) > 0$,

verletzt, wenn $g_i(\vec{x}) < 0$.



Evolutionäre Algorithmen

Evolutionäre Algorithmen bilden in vereinfachter Form Mechanismen der **biologischen Evolution** und der **Molekulargenetik** nach, um adaptive Systeme zu realisieren oder Optimierungsprobleme zu lösen.

Evolutionäre Algorithmen sind in der Regel **stochastische Algorithmen**. Für die Implementierung sind somit geeignete Zufallszahlengeneratoren (insbesondere für gleichverteilte und normalverteilte Zufallszahlen und Zufallsvektoren) erforderlich.

N-dimensionale Normalverteilung

Normalverteilte Zufallsvariablen (mittelwertfrei)

Wahrscheinlichkeitsdichte einer normalverteilten Zufallsvariablen

$$p(z) = \frac{1}{\sqrt{2\pi \cdot \sigma^2}} \cdot \exp\left(-\frac{z^2}{2\sigma^2}\right)$$

Verbundwahrscheinlichkeitsdichte eines N -dimensionalen normalverteilten Zufallsvektors:

$$p(\vec{z}) = \frac{1}{\sqrt{2\pi \cdot \sigma^{2N}}} \cdot \exp\left(-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} z_n^2\right) \quad \vec{z} = \begin{pmatrix} z_0 \\ z_1 \\ \vdots \\ z_{N-1} \end{pmatrix}$$



Box-Muller-Methode

Erzeugung einer normalverteilten Zufallsvariablen

Aus den in dem Intervall $[0,1]$ gleichverteilten Zufallszahlen u und v werden die beiden normalverteilten Zufallszahlen x und y mit dem Mittelwert 0 und der Varianz $\sigma^2 = 1$ erzeugt gemäß

$$x = \sqrt{-2 \cdot \log_e(1-u)} \cdot \cos(2\pi \cdot v)$$

und

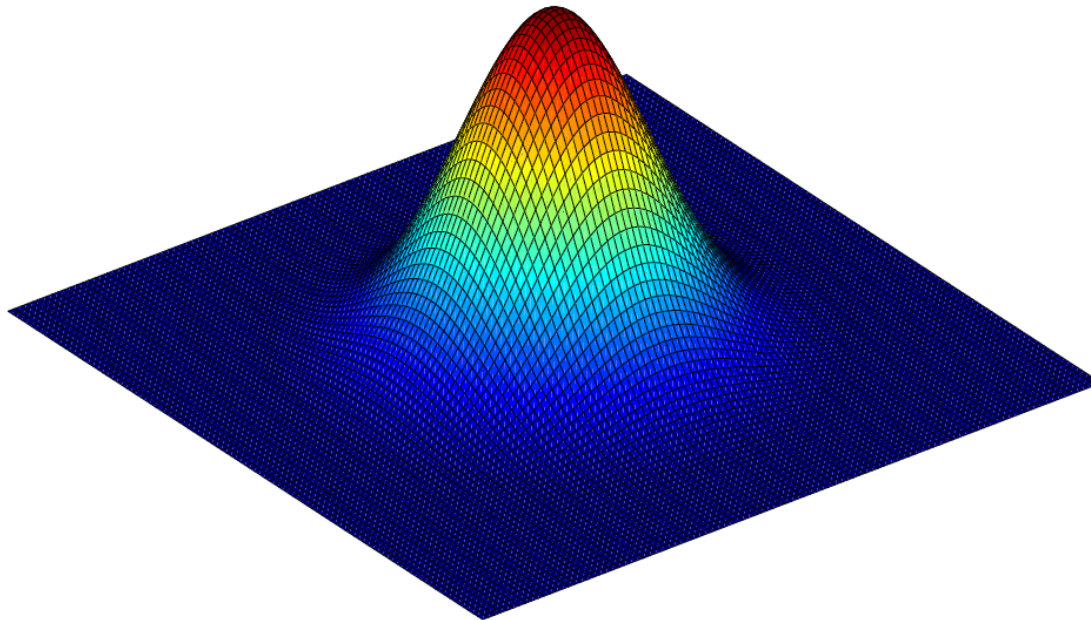
$$y = \sqrt{-2 \cdot \log_e(1-u)} \cdot \sin(2\pi \cdot v)$$

Erzeugung einer (μ, σ^2) verteilten Zufallsvariablen

Erzeuge x $(0,1)$ normalverteilt, dann ist $y = \mu + \sigma x$ nach dieser Translation und Skalierung (μ, σ^2) normal/Gauß-verteilt.

Normalverteilte Zufallsvariablen (mittelwertfrei)

Wahrscheinlichkeitsdichte einer zweidimensionalen normalverteilten Zufallsvariablen





Grundidee

- Individuen werden durch ein n -dimensionales Tupel $a \in \mathbb{R}^n$ charakterisiert.
- Eine Population $P(t) = \{a_0(t), a_1(t), \dots, a_{\mu-1}(t)\}$ aus μ Individuen zum Zeitpunkt t .
- Als Maß für die „Anpassung an die Umwelt“ wird den Individuen eine Fitness $f(a_k(t))$ zugewiesen.
- Aus den Besten werden zufällig oder nach anderen Kriterien „Eltern“ ausgewählt die „Kinder zeugen“.
 - Die Tupel der Kinder entstehen durch zufällige und/oder geeignete Variationen der Elter(n) Tupel.
 - Individuen haben eine gewisse Lebenszeit.



Evolutionäre Algorithmen

Pseudocode eines Evolutionären Algorithmus

```

t = 0 ;
Initialisiere Population  $P(t) = \{a_0(t), a_1(t), \dots, a_{r-1}(t)\}$ ;
Berechne Fitness  $f(a_k(t))$  von  $a_k(t)$  für  $0 \leq k \leq r-1$ ;
while end of adaptation  $\neq$  true do
    for ( $k=0$ ;  $k \leq r-1$ ;  $k++$ )
        Selektiere Eltern aus der Population  $P(t)$ ;
        Bilde Nachkommen  $a_k(t+1)$ ;
        Berechne Fitness  $f(a_k(t+1))$  von  $a_k(t+1)$ ;
        Bilde neue Population  $P(t+1)$  passend zur Fitness
    end
    t ++ ;
end

```



Evolutionäre Algorithmen

Mögliche Überlebens- und Vermehrungsstrategien:

$(1+1)$ -Evolutionstrategie: Einer von Elter oder Kind

$(1+\lambda)$ -Evolutionstrategie: Einer von Elter oder λ Kindern

$(1, \lambda)$ -Evolutionstrategie: Einer von λ Kindern

$(\mu+\lambda)$ -Evolutionstrategie: μ von μ Elter und λ Kindern

(μ, λ) -Evolutionstrategie: μ von λ Kindern

$(\mu\#\lambda)$ -Evolutionstrategie # ist '+' oder ','

$(\mu/\rho\#\lambda)$ -Evolutionstrategie: ρ Eltern zeugen λ Kinder

Selbstadaption der freien Parameter (welcher?)



$(\mu/\rho\#\lambda)$ -Evolutionstrategie

Bemerkungen:

- Bei $\mu=1$ Strategien überlebt nur der jeweils Fitteste.
- In der '+' Variante kann sich dieser dominant in allen Populationen durchsetzen => Gefahr des Hängens in falschen lokalen Optima!
- Die $(*, \lambda)$ Variante verhindert dies, birgt aber die Gefahr, das ein gefundenes Optimum ausstirbt.
- Bei $\rho=1$ findet kein Austausch von Tupelmerkmalen verschiedener Eltern statt, sondern nur statistische Variation/**Mutation**, erst $\rho > 1$ erlaubt die **sexuelle Rekombination** von Erbinformationen.



(1+1)-Evolutionsstrategie

Ablauf:

Initialisierung

Ein N -dimensionaler Vektor $\vec{x}_E(0) \in \mathbb{R}^N$ wird zufällig gewählt. Setze $t=0$.

Mutation

Der Elter $\vec{x}_E(t)$ in der Generation t erzeugt einen Nachkommen $\vec{x}_N(t)$, der sich von dem Elter ein wenig unterscheidet.

Selektion

Die beiden Individuen $\vec{x}_E(t)$ und $\vec{x}_N(t)$ werden hinsichtlich ihrer Fitness $f(\vec{x}_E(t))$ und $f(\vec{x}_N(t))$ verglichen; das bessere wird als Elter der nächsten Generation $t+1$ verwendet.



(1+1)-Evolutionstrategie

Eigenschaften:

- Die Größe der Population ist konstant.
- Ein Individuum besitzt prinzipiell eine unendliche lange Lebensdauer und kann unendlich viele Nachkommen erzeugen.
- Nur Mutationen treten auf.
- Die Umgebung und somit die Fitness sind konstant und ändern sich nicht mit der Zeit.
- Es werden keine Mechanismen der modernen synthetischen Evolutionstheorie wie Chromosomenmutationen, Rekombination, Diploidie, Dominanz und Rezessivität, Migration, ... verwendet.

(1+1)-Evolutionstrategie

Algorithmus Idee

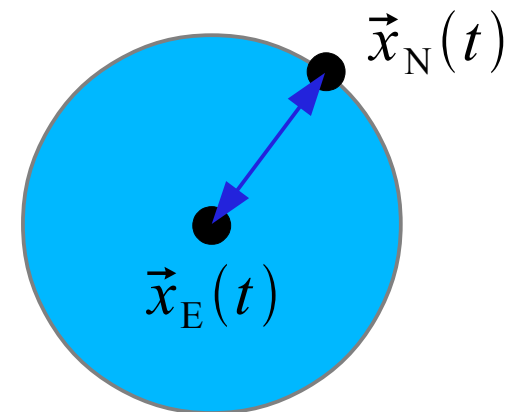
Erzeugung eines Nachkommens durch **Mutation**

$$\vec{x}_N(t) = \vec{x}_E(t) + \vec{z}(t)$$

mit dem mittelwertfreien und unkorrelierten normalverteilten Zufallsvektor $\vec{z}(t)$ mit der Varianz σ^2

Selektion

$$\vec{x}_E(t+1) = \begin{cases} \vec{x}_N(t) & , \quad f(\vec{x}_N(t)) \geq f(\vec{x}_E(t)) \\ \vec{x}_E(t) & , \quad \text{sonst} \end{cases}$$





(1+1)-Evolutionstrategie

Pseudocode der (1+1)-Strategie

$t = 0$;

Initialisiere $\vec{x}_E(0) = (x_{E,0}(0), x_{E,1}(0), \dots, x_{E,N-1}(0))^T$;

while end of adaptation \neq true **do**

Bilde Nachkomme $\vec{x}_N(t) = \vec{x}_E(t) + \vec{z}(t)$ durch Addition
eines normalverteilten Zufallsvektors $\vec{z}(t)$;

Selektiere $\vec{x}_E(t+1) = \begin{cases} \vec{x}_N(t) & , \quad f(\vec{x}_N(t)) \geq f(\vec{x}_E(t)) \\ \vec{x}_E(t) & , \quad \text{sonst} \end{cases}$;

$t++$;

end



Evolutionäre Algorithmen

- Während die Grundidee eines EA simpel und einfach zu programmieren ist, erfordert die Vorhersage seiner Eigenschaften ein gehöriges Maß an Statistik und Mathematik.
- Am einfachsten lässt sich das Verhalten des (1+1) Algorithmus bestimmen.

Die zentralen Fragen sind

1. wie groß ist der **Fortschritt** der Entwicklung der Populationen in Richtung Optimum und
2. mit welcher **Erfolgswahrscheinlichkeit** passiert dies in Abhängigkeit von freien Parametern (welchen?).

Evolutionäre Algorithmen

Zwei Fitness Modelle:

Kugelmodell

$$f(\vec{x}) = - \sum_{n=0}^{N-1} x_n^2$$

Korridormodell

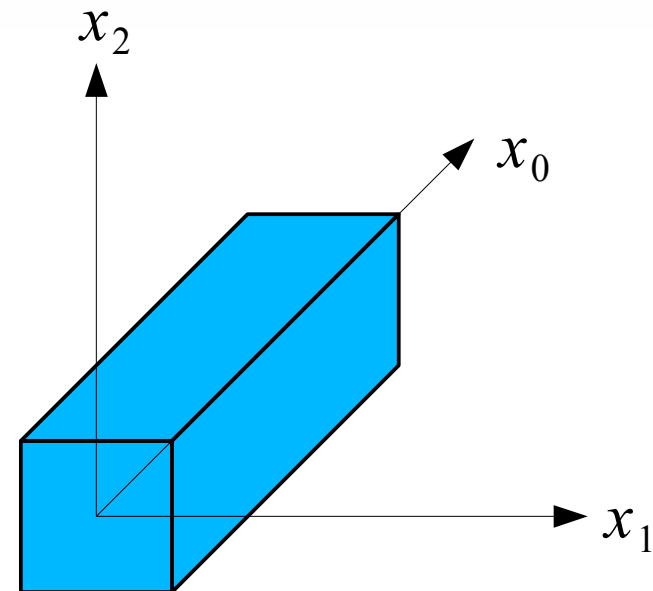
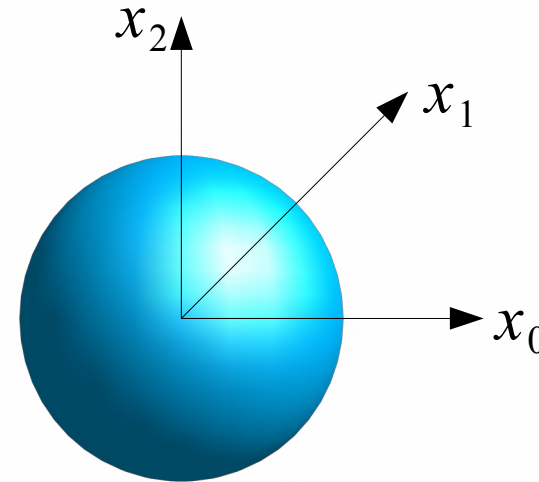
$$f(\vec{x}) = x_0$$

mit den Beschränkungen

$$|x_i| \leq B \quad \text{für } 1 \leq i \leq N-1$$

bzw.

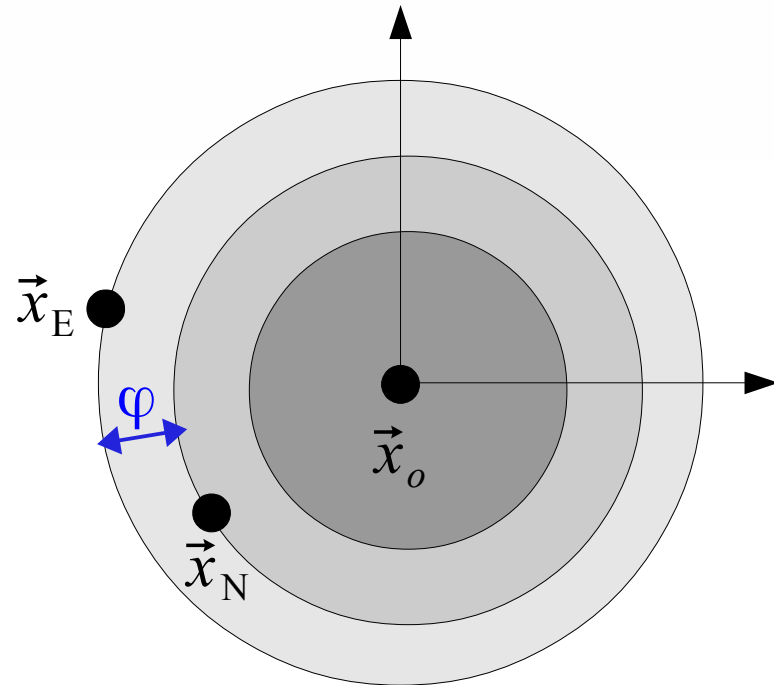
$$g_i(\vec{x}) = |x_i| \quad \text{für } 1 \leq i \leq N-1$$



Evolutionäre Algorithmen

Lokaler Fortschritt

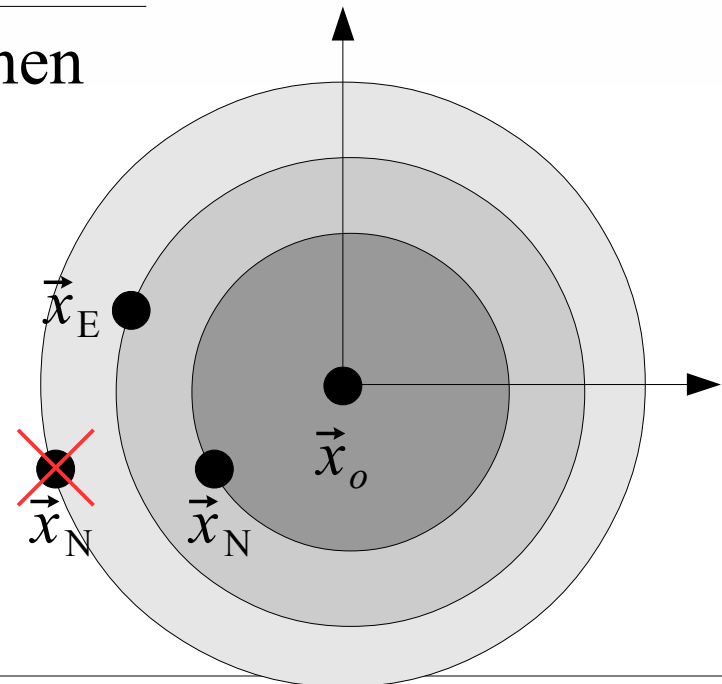
Der **lokale Fortschritt** φ ist definiert als die mittlere Strecke, die in einer Generation in Richtung des Optimums \vec{x}_o zurückgelegt wird.



Evolutionäre Algorithmen

Die Erfolgswahrscheinlichkeit P_{Erfolg} ist definiert als die relative Häufigkeit der erfolgreichen Mutationen.

$$P_{\text{Erfolg}} = \frac{\text{Anzahl erfolgreicher Mutationen}}{\text{Gesamtzahl der Mutationen}}$$



(1+1)-Evolutionstrategie

Berechnung des Fortschritts

Erzeugung eines Nachkommen

$$\vec{x}_N(t) = \vec{x}_E(t) + \vec{z}(t)$$

Bedingte Wahrscheinlichkeitsdichte

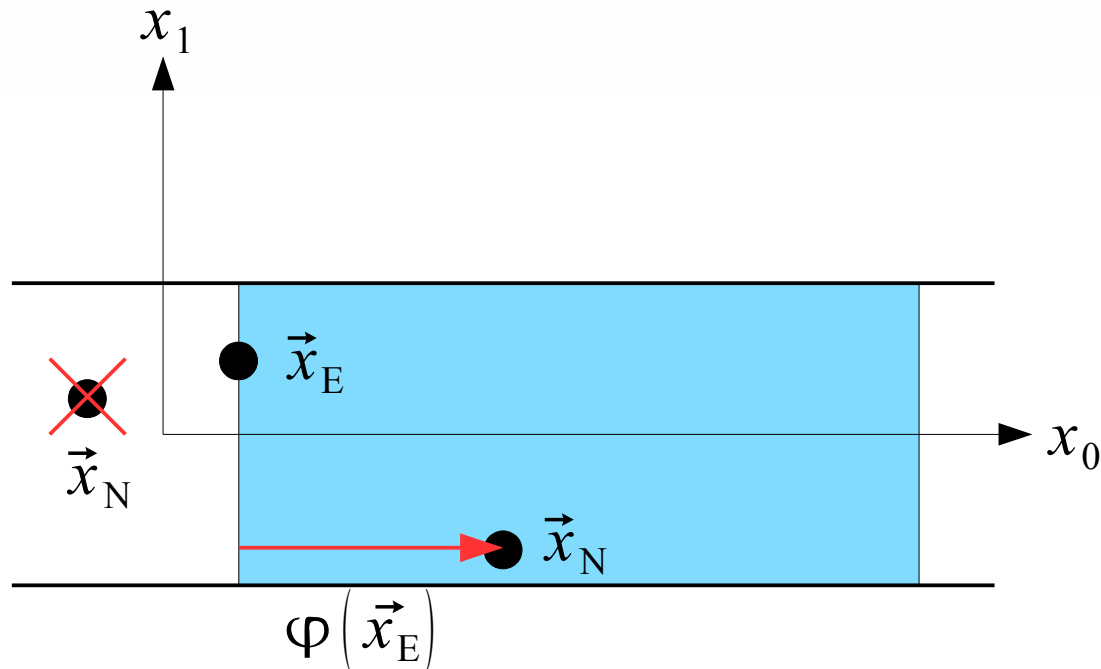
$$\begin{aligned} p(\vec{x}_N | \vec{x}_E) &= \frac{1}{\sqrt{2\pi \cdot \sigma^{2N}}} \cdot \exp\left(-\frac{\|\vec{x}_N - \vec{x}_E\|^2}{2\sigma^2}\right) \\ &= \frac{1}{\sqrt{2\pi \cdot \sigma^{2N}}} \cdot \exp\left(-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x_{N,n} - x_{E,n})^2\right) \end{aligned}$$

(1+1)-Evolutionstrategie

Fortschritt im Korridormodell

Betrachte Korridormodell mit $N = 2$

$$f(\vec{x}) = f(x_0, x_1) = x_0 \quad \text{und} \quad |x_1| \leq B$$





(1+1)-Evolutionstrategie

Fortschritt im Korridormodell

Bedingte Wahrscheinlichkeitsdichte mit $N = 2$

$$\begin{aligned}
 p(\vec{x}_N | \vec{x}_E) &= p(x_{N,0}, x_{N,1} | x_{E,0}, x_{E,1}) \\
 &= \frac{1}{2\pi\sigma^2} \cdot \exp\left(-\frac{(x_{N,0} - x_{E,0})^2 + (x_{N,1} - x_{E,1})^2}{2\sigma^2}\right) \\
 &= \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left(-\frac{(x_{N,0} - x_{E,0})^2}{2\sigma^2}\right) \\
 &\quad \cdot \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left(-\frac{(x_{N,1} - x_{E,1})^2}{2\sigma^2}\right)
 \end{aligned}$$

(1+1)-Evolutionstrategie

Fortschritt im Korridormodell

Lokaler Fortschritt für Korridormodell mit $N = 2$

$$\begin{aligned}
 & \Phi(x_{E,0}, x_{E,1}) \\
 &= \int_{x_{N,0}=x_{E,0}}^{\infty} \int_{x_{N,1}=-B}^B (x_{N,0} - x_{E,0}) \cdot p(x_{N,0}, x_{N,1} | x_{E,0}, x_{E,1}) dx_{N,0} dx_{N,1} \\
 &= \frac{1}{\sqrt{2\pi\sigma}} \int_{x_{N,0}=x_{E,0}}^{\infty} (x_{N,0} - x_{E,0}) \cdot \exp\left(-\frac{(x_{N,0} - x_{E,0})^2}{2\sigma^2}\right) dx_{N,0} \\
 & \quad \cdot \frac{1}{\sqrt{2\pi\sigma}} \int_{x_{N,1}=-B}^B \exp\left(-\frac{(x_{N,1} - x_{E,1})^2}{2\sigma^2}\right) dx_{N,1}
 \end{aligned}$$



(1+1)-Evolutionstrategie

Fortschritt im Korridormodell

Lokaler Fortschritt für Korridormodell mit $N = 2$

$$\begin{aligned}
 & \Phi(x_{E,0}, x_{E,1}) \\
 &= \frac{\sqrt{2}\sigma}{\sqrt{\pi}} \underbrace{\int_0^{\infty} \xi \cdot e^{-\xi^2} d\xi}_{=\frac{1}{2}} \cdot \frac{1}{\sqrt{\pi}} \int_{(-B-x_{E,1})/(\sqrt{2}\sigma)}^{(B-x_{E,1})/(\sqrt{2}\sigma)} e^{-\xi^2} d\xi \\
 &= \frac{\sigma}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{\pi}} \int_{(-B-x_{E,1})/(\sqrt{2}\sigma)}^{(B-x_{E,1})/(\sqrt{2}\sigma)} e^{-\xi^2} d\xi \\
 &= \frac{\sigma}{\sqrt{2\pi}} \cdot \frac{1}{2} \cdot \left\{ \operatorname{erf}\left(\frac{B-x_{E,1}}{\sqrt{2}\sigma}\right) + \operatorname{erf}\left(\frac{B+x_{E,1}}{\sqrt{2}\sigma}\right) \right\}
 \end{aligned}$$

Fehlerfunktion

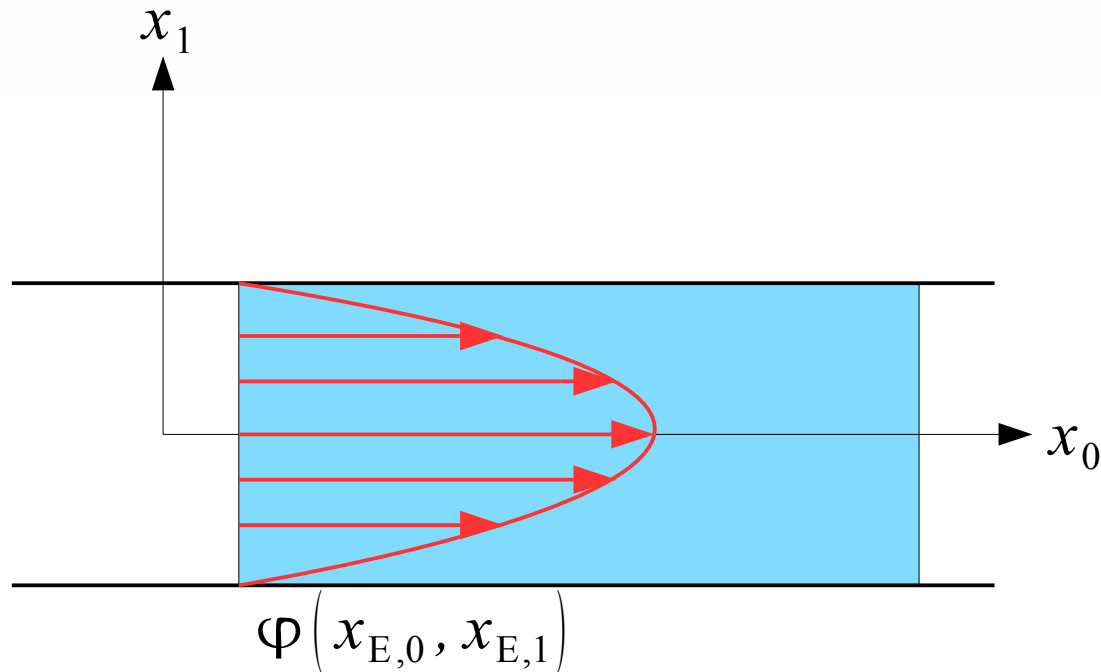
$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-\xi^2} d\xi$$

(1+1)-Evolutionstrategie

Fortschritt im Korridormodell

Betrachte Korridormodell mit $N = 2$

$$f(\vec{x}) = f(x_0, x_1) = x_0 \quad \text{und} \quad |x_1| \leq B$$





(1+1)-Evolutionstrategie

Fortschritt im Korridormodell

Lokaler Fortschritt für Korridormodell mit $N = 2$

$$\begin{aligned}
 \varphi &= \frac{1}{2B} \int_{x_{E,1}=-B}^B \varphi(x_{E,0}, x_{E,1}) dx_{E,1} \\
 &= \frac{\sigma}{\sqrt{2\pi}} \cdot \frac{1}{4B} \int_{x_{E,1}=-B}^B \left\{ \operatorname{erf}\left(\frac{B-x_{E,1}}{\sqrt{2}\sigma}\right) + \operatorname{erf}\left(\frac{B+x_{E,1}}{\sqrt{2}\sigma}\right) \right\} dx_{E,1} \\
 &= \frac{\sigma}{\sqrt{2\pi}} \cdot \left\{ \operatorname{erf}\left(\frac{\sqrt{2}B}{\sigma}\right) - \frac{\sigma}{\sqrt{2\pi}B} \cdot \left(1 - e^{-2B^2/\sigma^2}\right) \right\}
 \end{aligned}$$

(1+1)-Evolutionstrategie

Fortschritt im Korridormodell

Lokaler Fortschritt für Korridormodell mit $N = 2$

$$\operatorname{erf}(x) \approx 1 - \frac{1}{\sqrt{\pi}} \cdot \frac{e^{-x^2}}{x} \quad \text{für } x \gg 1$$

$$\Rightarrow \varphi = \frac{\sigma}{\sqrt{2\pi}} \cdot \left(1 - \frac{1}{\sqrt{2\pi}} \cdot \frac{\sigma}{B} \right) \quad \text{für } \sigma \ll B$$

Lokaler Fortschritt für Korridormodell mit beliebigem N

$$\Rightarrow \varphi = \frac{\sigma}{\sqrt{2\pi}} \cdot \left(1 - \frac{1}{\sqrt{2\pi}} \cdot \frac{\sigma}{B} \right)^{N-1} \quad \text{für } \sigma \ll B$$



Übung

- Der lokale Fortschritt der (1+1)-Evolutionstrategie beim Korridormodell ist gegeben durch die folgende Formel.

$$\varphi = \frac{\sigma}{\sqrt{2\pi}} \cdot \left(1 - \frac{1}{\sqrt{2\pi}} \cdot \frac{\sigma}{B} \right)^{N-1} \quad \text{für } \sigma \ll B$$

Bestimmen Sie die bei der Mutation verwendete **optimale Standardabweichung σ_{opt}** , die zu einem maximalen lokalen Fortschritt führt.

- Wie kann bei **ermittelter Erfolgswahrscheinlichkeit $P(t)$** die Standardabweichung σ adaptiv während des Algorithmus angepasst werden?