

"Covariance Matrix Adaptation - Evolution Strategy"

Problem: $\left\{ \begin{array}{l} f: \mathbb{R}^n \rightarrow \mathbb{R} \\ x^0 \in \mathbb{R}^n \end{array} \right\}$ gegeben

"minimiere f ".

1. Ansatz

$\bar{x}^0 := x^0$, Fixiere $\lambda \in \mathbb{N}$

Für $k = 0, 1, 2, \dots$

generiere $x^1, \dots, x^\lambda$ i.i.d. $N(\bar{x}^k, I)$ -verteilt

ersetze x^0 durch $\operatorname{argmin}_{0 \leq i \leq \lambda} f(x^i)$

$\bar{x}^{k+1} := x^0$

end

Im Folgenden soll die Aktualisierung der Verteilung verbessert werden und I durch eine "Covariance-matrix" C ersetzt werden, die ebenfalls aktualisiert wird.

Vorbemerkung: Wenn $x_1, \dots, x_n \in \mathbb{R}$ unabhängig $N(0, 1)$ -verteilt erzeugt werden, so ist $x = (x_1, \dots, x_n)^T$ $N(0, I)$ -verteilt (mit $0 \in \mathbb{R}^n$).

Und wenn $B \in \mathbb{R}^{n \times n}$ fest gegeben

und nichtsingulär ist, so ist $y := Bx$

$N(0, BB^T)$ -verteilt und $C := BB^T$ positiv definit.

Modifizierungen

ii) Angefangen mit $C := I$ erhöhe nach jedem Schritt die Wahrscheinlichkeit, dass in Richtung $\pm(\bar{x}^k - \bar{x}^{k+1})$ gesucht wird (für $\bar{x}^{k+1} \neq \bar{x}^k$) und setze z.B. $c_1 > 0$ sowie

$$C := C \cdot (1 - c_1) + c_1 \cdot \frac{(\bar{x}^{k+1} - \bar{x}^k)(\bar{x}^{k+1} - \bar{x}^k)^T}{\|\bar{x}^{k+1} - \bar{x}^k\|_C^2}$$

⚠ c_1 "zu groß" $\Rightarrow \text{cond}(C) \xrightarrow{k \rightarrow \infty} \infty$.

iii) Nutze eine Schrittweite $\sigma > 0$, d.h.

generiere $\tilde{x}^1, \dots, \tilde{x}^\lambda$ i.i.d. $N(0, I)$ verteilt

$$\text{Setze } x^i := \bar{x}^k + \sigma C^{1/2} \tilde{x}^i \quad (1 \leq i \leq \lambda)$$

iii) Passe σ z.B. nach der $1/5$ -Regel an:

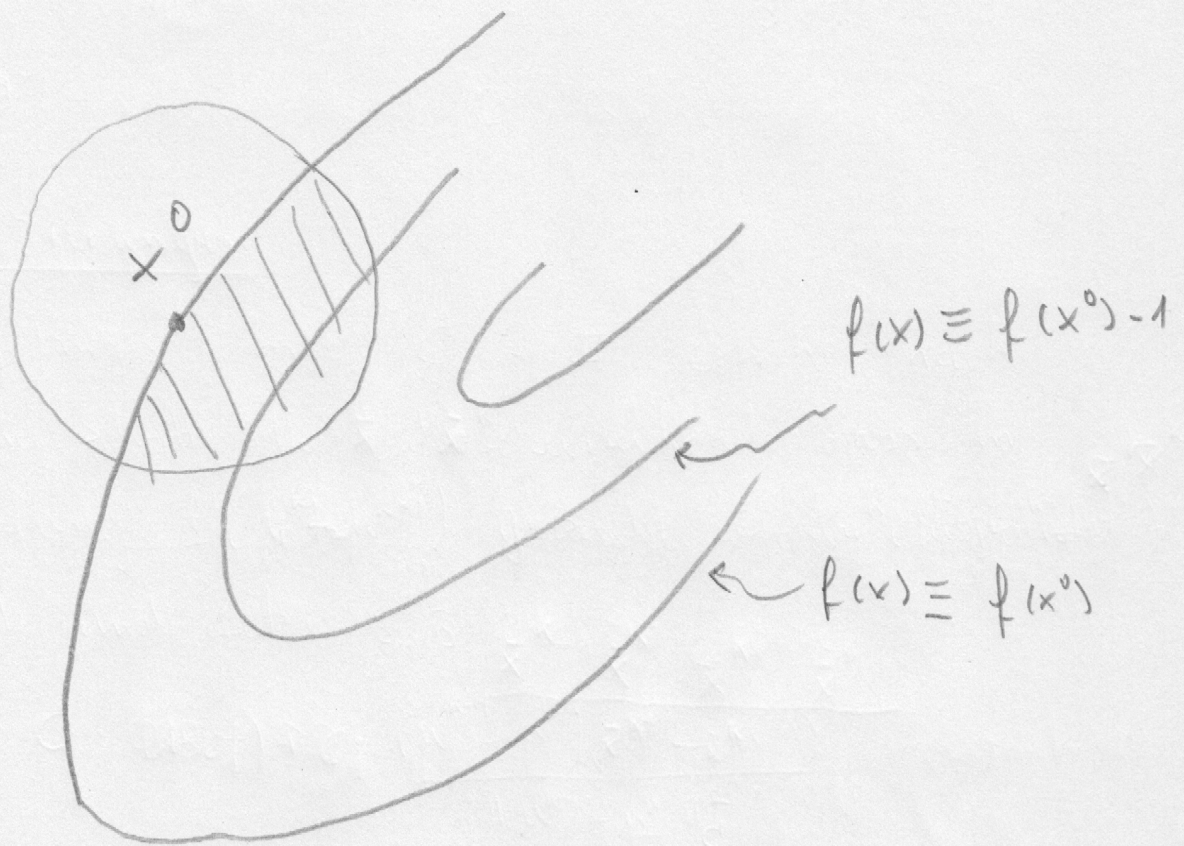
Wenn $|\{i \in \{1, \dots, \lambda\} \mid f(x^i) < f(x^0)\}| < \frac{\lambda}{5}$

so setze $\sigma \rightarrow \sigma/2$

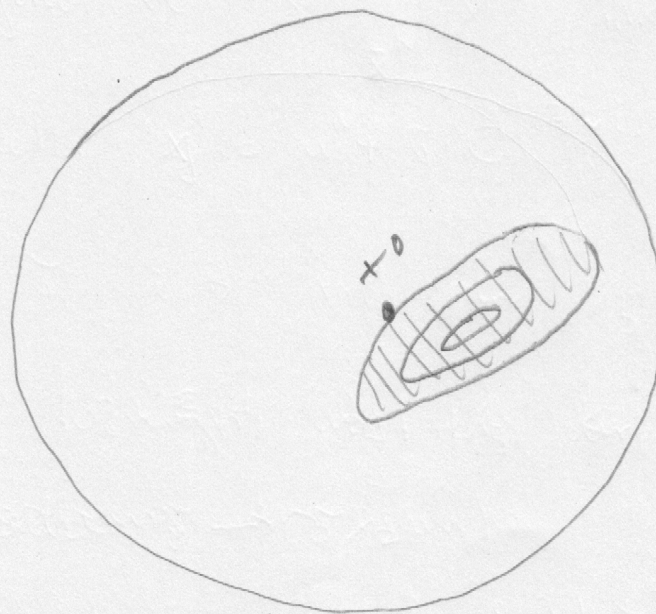
Wenn $|\{i \mid f(x^i) < f(x^0)\}| > \frac{2\lambda}{5}$

so setze $\sigma \rightarrow 2\sigma$ (grob vereinfacht)

Skizze für differenzierbares f :



Kurzer Schritt $\approx 50\%$ der x^i liegen
in $\{x \mid f(x) < f(x^0)\}$ (schraffierter Bereich)



(zu) langer Schritt: nur wenige der x^i
liegen in $\{x \mid f(x) \leq f(x^0)\}$.

iv)) Setze z.B. $\mu := \lceil 0,3 \lambda \rceil$ und

sortiere $f(x^0) \leq f(x^1) \leq \dots \leq f(x^{\mu-1})$

Setze $w_i := \mu - i$ für $0 \leq i \leq \mu - 1$

und $\bar{w} := \sum_{i=0}^{\mu-1} w_i$ und

aktualisiere $C := (1 - \frac{\mu}{n^2}) C + \frac{\mu}{n^2} \sum_{i=0}^{\mu-1} \frac{w_i}{\bar{w}} \frac{x^i (x^i)^T}{\sigma^2}$

v)) Weitere Modifizierungen:

"akkumulierter Pfad der letzten Schritte", z.B.

$$p^0 = 0, \quad p^{k+1} = 0,9 p^k + 0,1 (\bar{x}^{k+1} - \bar{x}^k)$$

mit zusätzlicher Aktualisierung $C := (1 - c_1) C + c_1 p^k (p^k)^T$

Zahlreiche weitere (etwas andere) Modifizierungen

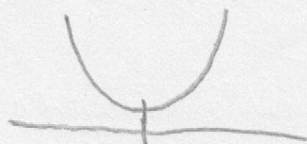
Zwischen 2001 und 2023 $\Rightarrow$ CMA-ES

(public domain, python, matlab, ...)

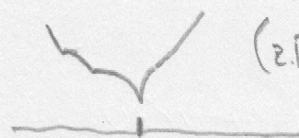
Invarianzeigenschaften

- gegenüber Verschiebungen & Rotationen
- gegenüber monotonen Transformationen!

(z.B. $\|x\|_2^2$)



$\leftrightarrow$



(z.B. $\sqrt{|x|}$ u.a.)

- $C \rightsquigarrow \approx \nabla^2 f(x)^{-1}$ im konvex quadratischen Fall?
(Auger & Hansen '2013, p. 49)

⚠ Die Konvergenz hängt stark von der Wahl der Parameter ab

Bsp mit konstantem $C \equiv \nabla^2 f(x) \equiv I$.

($n = 2, 4, 8, \dots$)

$x = \text{ones}(n, 1) / \text{sqrt}(n);$

$it = 0;$

While $\text{norm}(x) > 1e-100$

$it = it + 1;$

$r = \text{norm}(x);$ % length of ideal step

$x_p = x + \text{randn}(n, 1) * r / n;$ % scaling matters!

if $\text{norm}(x_p) < \text{norm}(x)$

$x = x_p;$

end

end

$it / 100$ % average number of iterations per digit

n	2	4	8	16	32	64	128
it/100	1167	3416	8129	1716	363.5	728.0	1478

In diesem Idealfall:

Etwas mehr als $10n$ Iterationen pro gültige Stelle.