Search

Given : Objective function $f : \mathbb{R}^n \to \mathbb{R}$

Objective : Generate sequence $\left(\boldsymbol{x}^{(t)}\right)_{t=1,2,\ldots} \in \mathbb{R}^n$ where $f(\boldsymbol{x}^{(t)})$ as small as possible for all $t = 1, 2, \ldots$

Assumption : Best known problem encoding is used.

Randomized search : Every search point $\boldsymbol{x}^{(t)}$ is drawn from a given distribution $P^{(t)}$.

Why Randomize Search?

- Black box optimization (derivatives not available)
- The typical difference quotients are not useful
- Discontinuities
- Noise and outlier
- Many local optima

In summary: Black box optimization in a rough or rugged landscape.

Randomized Search Example

Use normal distribution $\mathcal{N}\left(\boldsymbol{\mu}^{(t)}, \boldsymbol{I}\right)$ as search distribution $P^{(t)}$:

FOR $t = 1, 2, \ldots$

Sample $P^{(t)} = \mathcal{N}\left(\boldsymbol{\mu}^{(t)}, \boldsymbol{I}\right) \rightarrow \boldsymbol{x}_1^{(t)}, \ldots, \boldsymbol{x}_\lambda^{(t)} \in \mathbb{R}^n$

Set $\boldsymbol{\mu}^{(t+1)} = \boldsymbol{x}_{\text{selected}}^{(t)} := \arg\min_k \left(f(\boldsymbol{x}_k^{(t)})\right)$

ROF

implements the so-called $(1, \lambda)$-ES (evolution strategy) without strategy parameter adaptation.

The CMA-ES (**E**volution **S**trategy with **C**ovariance **M**atrix **A**daptation)

Consider $P^{(t)} = \mathcal{N}\left(\boldsymbol{\mu}^{(t)}, {\sigma^{(t)}}^2 \boldsymbol{C}^{(t)}\right)$ where $\boldsymbol{\mu}^{(t)} \in \mathbb{R}^n$, $\sigma^{(t)} \in \mathbb{R}_+$, $\boldsymbol{C}^{(t)} \in \mathbb{R}^{n \times n}$

- $\boldsymbol{\mu}^{(t)} \rightarrow \boldsymbol{\mu}^{(t+1)}$: Maximum likelihood update, i.e. $P(\boldsymbol{x}_{\text{selected}}^{(t)} | \boldsymbol{\mu}^{(t+1)}) \rightarrow \max$
- $\boldsymbol{C}^{(t)} \rightarrow \boldsymbol{C}^{(t+1)}$: Maximum likelihood update, i.e. $P(\frac{\boldsymbol{x}_{\text{selected}}^{(t)} - \boldsymbol{\mu}^{(t)}}{\sigma^{(t)}} | \boldsymbol{C}^{(t+1)}) \rightarrow \max$, under consideration of prior $\boldsymbol{C}^{(t)}$ (otherwise $\boldsymbol{C}^{(t+1)}$ becomes singular).
- $\sigma^{(t)} \rightarrow \sigma^{(t+1)}$: Update to achieve conjugate perpendicularity, i.e. conceptually $(\boldsymbol{\mu}^{(t+2)} - \boldsymbol{\mu}^{(t+1)})^{\mathrm{T}} {\boldsymbol{C}^{(t)}}^{-1} (\boldsymbol{\mu}^{(t+1)} - \boldsymbol{\mu}^{(t)}) / {\sigma^{(t+1)}}^2 \rightarrow 0$

Remark: It follows that $\boldsymbol{\mu}^{(t+1)} = \boldsymbol{x}_{\text{selected}}^{(t)}$.

Remark: The update of $\boldsymbol{C}$ roughly results in a PCA of the points $\frac{\boldsymbol{x}_{\text{selected}}^{(t)} - \boldsymbol{\mu}^{(t)}}{\sigma^{(t)}}, \frac{\boldsymbol{x}_{\text{selected}}^{(t-1)} - \boldsymbol{\mu}^{(t-1)}}{\sigma^{(t-1)}}, \ldots, \frac{\boldsymbol{x}_{\text{selected}}^{(t-n^2)} - \boldsymbol{\mu}^{(t-n^2)}}{\sigma^{(t-n^2)}}$ assuming their empirical mean to be zero.