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

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

Assumption: Best known problem encoding is used.

Randomized search: Every search point $x^{(t)}$ 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 $\mathcal{N}(\mu^{(t)}, \mathbf{I})$ as search distribution $P^{(t)}$:

FOR $t = 1, 2, \ldots$

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

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

ROF

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

The CMA-ES (Evolution Strategy with Covariance Matrix Adaptation)

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

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

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

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