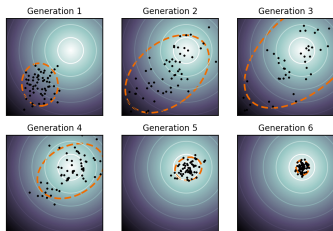


Optimization in Machine Learning

Evolutionary Algorithms

CMA-ES Algorithm



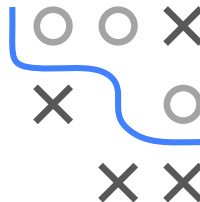
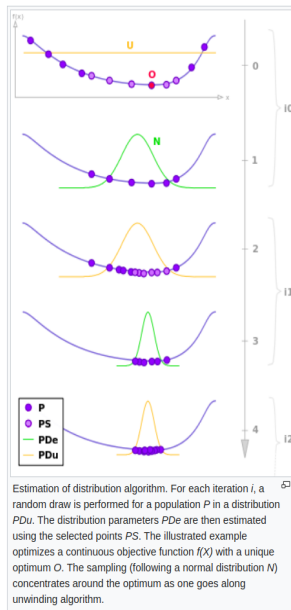
Learning goals

- CMA-ES strategy
- Estimation of distribution
- Step size control

ESTIMATION OF DISTRIBUTION ALGORITHM

- Instead of population, maintain distribution to sample offspring from

- 1 Draw λ offsprings $\mathbf{x}^{(i)}$ from $p(\cdot|\theta^{[t]})$
- 2 Evaluate fitness $f(\mathbf{x}^{(i)})$
- 3 Update $\theta^{[t+1]}$ with μ best offsprings



CMA-ES: BASIC METHOD - ITERATION 1

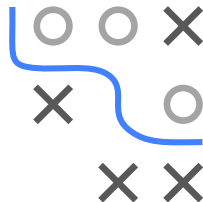
- 0 Initialize $\mathbf{m}^{[0]}$, $\sigma^{[0]}$ problem-dependent and $\mathbf{C}^{[0]} = \mathbf{I}_d$



CMA-ES: BASIC METHOD - ITERATION 1 / 2

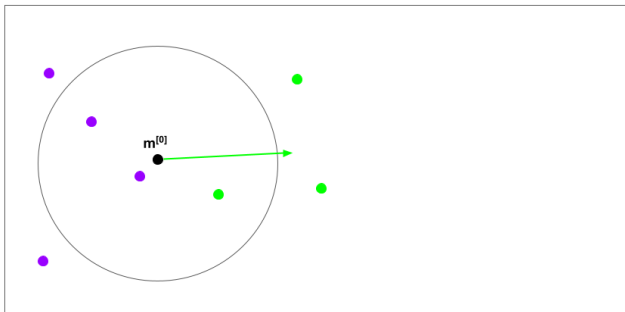
❶ **Sample** λ offsprings from distribution

$$\mathbf{x}^{[1](i)} = \mathbf{m}^{[0]} + \sigma^{[0]} \mathcal{N}(\mathbf{0}, \mathbf{C}^{[0]})$$



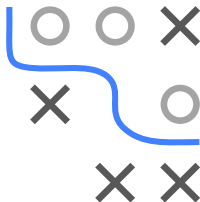
CMA-ES: BASIC METHOD - ITERATION 1 / 3

- ② **Selection and recombination** of $\mu < \lambda$ best-performing offspring using fixed weights $w_1 \geq \dots \geq w_\mu > 0, \sum_{i=1}^{\mu} w_i = 1$.
 $\mathbf{x}_{i:\lambda}$ is i -th ranked solution, ranked by $f(\mathbf{x}_{i:\lambda})$.



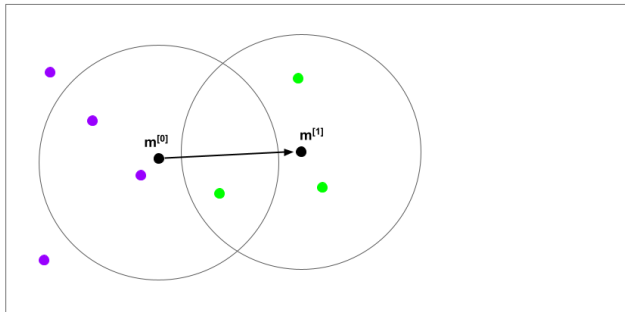
Calculation of auxiliary variables ($\mu = 3$ points)

$$\mathbf{y}_w^{[1]} := \sum_{i=1}^{\mu} w_i (\mathbf{x}_{i:\lambda}^{[1]} - \mathbf{m}^{[0]}) / \sigma^{[0]} := \sum_{i=1}^{\mu} w_i \mathbf{y}_{i:\lambda}^{[1]}$$



CMA-ES: BASIC METHOD - ITERATION 1 / 4

3 Update mean



Movement towards the new distribution with mean

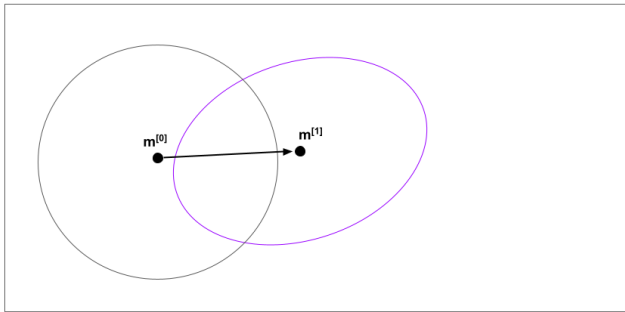
$$\mathbf{m}^{[1]} = \mathbf{m}^{[0]} + \sigma^{[0]} \mathbf{y}_w^{[1]}.$$

CMA-ES: BASIC METHOD - ITERATION 1 / 5

4 Update covariance matrix

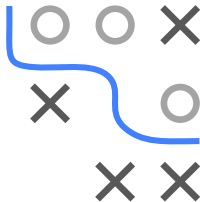
Roughly: elongate density ellipsoid in direction of successful steps.

$\mathbf{C}^{[1]}$ reproduces successful points with higher probability than $\mathbf{C}^{[0]}$.



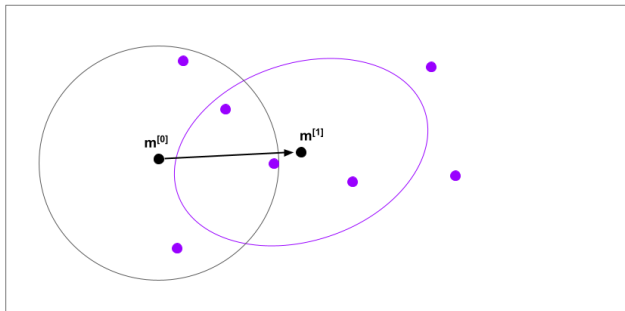
Update $\mathbf{C}^{[0]}$ using sum of outer products and parameter c_μ :

$$\mathbf{C}^{[1]} = (1 - c_\mu)\mathbf{C}^{[0]} + c_\mu \sum_{i=1}^{\mu} w_i \mathbf{y}_{i:\lambda}^{[1]} (\mathbf{y}_{i:\lambda}^{[1]})^\top \text{ (rank-}\mu \text{ update).}$$



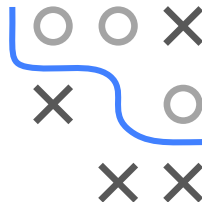
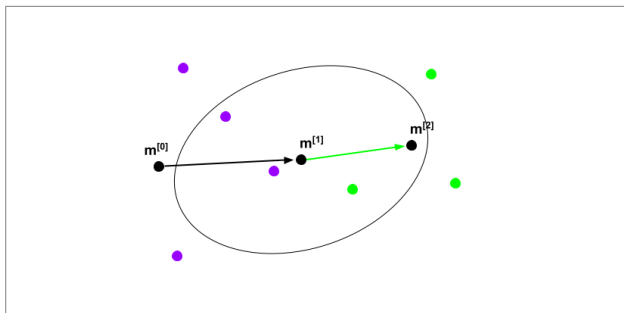
CMA-ES: BASIC METHOD - ITERATION 2

❶ **Sample** from distribution for new generation



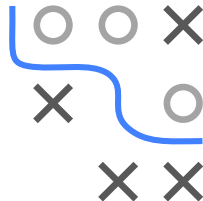
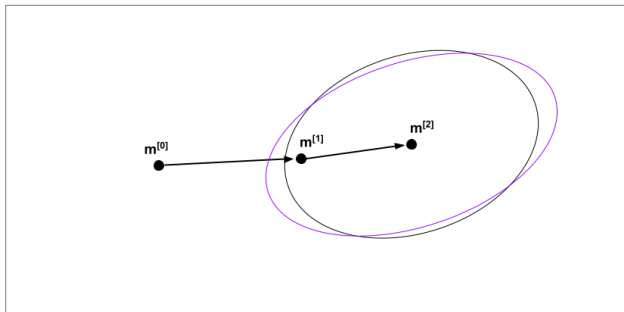
CMA-ES: BASIC METHOD - ITERATION 2 / 2

- ② Selection and recombination of $\mu < \lambda$ best-performing offspring
- ③ Update mean



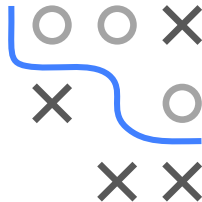
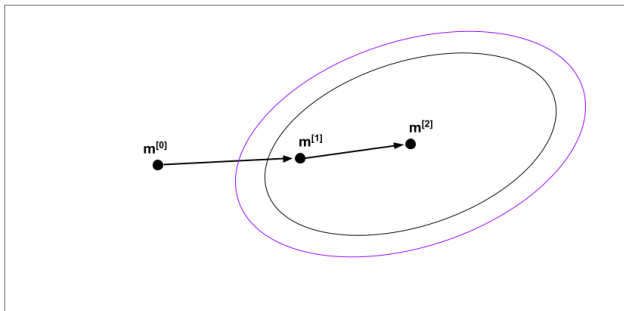
CMA-ES: BASIC METHOD - ITERATION 2 / 3

④ Update covariance matrix



CMA-ES: BASIC METHOD - ITERATION 2 / 4

- Update step-size** exploiting correlation in history of steps.
 - steps point in similar direction \implies increase step-size
 - steps cancel out \implies decrease step-size



UPDATING C: FULL UPDATE

Full CMA update of **C** combines rank- μ update with a rank-1 update using exponentially smoothed evolution path $\mathbf{p}_c \in \mathbb{R}^d$ of successive steps and learning rate c_1 :

$$\mathbf{p}_c^{[0]} = \mathbf{0}, \quad \mathbf{p}_c^{[t+1]} = (1 - c_1)\mathbf{p}_c^{[t]} + \sqrt{\frac{c_1(2 - c_1)}{\sum_{i=1}^{\mu} w_i^2}} \mathbf{y}_w$$

Final update of **C** is

$$\mathbf{C}^{[t+1]} = (1 - c_1 - c_{\mu} \sum_j w_j) \mathbf{C}^{[t]} + \underbrace{c_1 \mathbf{p}_c^{[t+1]} (\mathbf{p}_c^{[t+1]})^{\top}}_{\text{rank-1}} + \underbrace{c_{\mu} \sum_{i=1}^{\mu} w_i \mathbf{y}_{i:\lambda}^{[t+1]} (\mathbf{y}_{i:\lambda}^{[t+1]})^{\top}}_{\text{rank-}\mu}$$

- Correlation between generations used in rank-1 update
- Information from entire population is used in rank- μ update



UPDATING σ : METHODS STEP-SIZE CONTROL

- **1/5-th success rule**: increases the step-size if more than 20 % of the new solutions are successful, decrease otherwise
- **σ -self-adaptation**: mutation is applied to the step-size and the better - according to the objective function value - is selected
- **Path length control via cumulative step-size adaptation (CSA)**

Intuition:

- Short cumulative step-size \triangleq steps cancel \rightarrow decrease $\sigma^{[t+1]}$
- Long cumulative step-size \triangleq corr. steps \rightarrow increase $\sigma^{[t+1]}$

