DERIVATIVE FREE OPTIMIZATION 2025/2026

CLASS 4

On the scientific project:

Do	Don't
Allowed to copy paste but put into " " and cite source	→ Plagiarism
Borrow figures but cite the source	

Presenting numerical results:

1/ compare to state of the the art → if you want to prove that a new approach is competitive

2/ → report experimental setup to be able to redo experiments

→ quantify error

When describing results :

1/ Describe what is displayed on a graph

↳ on the y axis we have...

↳ ______ x ______ ______ ...

2/ Describe what you observe

3/ Interpret the observation.

Use the right scale to present graphs

↳ don't hesitate to use log scale or log-log scale

Last time we have motivated that to solve ill-conditioned problems, we need to sample with

$$\mathcal{N}(m_t, \sigma_t C_t)$$

where C_t is adapted.

Ideally we would like $C_t \propto H^{-1}$.

Covariance Matrix Adaptation

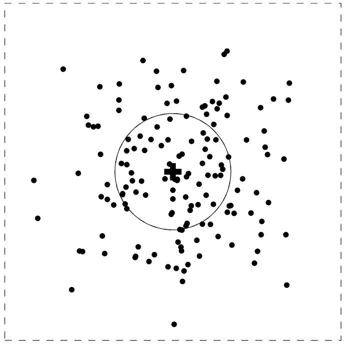
Covariance Matrix Adaptation

Evolution Strategies

Recalling

New search points are sampled normally distributed

$$\boldsymbol{x}_i \sim \boldsymbol{m} + \sigma\,\mathcal{N}_i(\mathbf{0}, \mathbf{C}) \qquad \text{for } i = 1, \ldots, \lambda$$



as perturbations of $\boldsymbol{m}$, where $\boldsymbol{x}_i, \boldsymbol{m} \in \mathbb{R}^n$, $\sigma \in \mathbb{R}_+$, $\mathbf{C} \in \mathbb{R}^{n \times n}$

where

- the mean vector $\boldsymbol{m} \in \mathbb{R}^n$ represents the favorite solution
- the so-called step-size $\sigma \in \mathbb{R}_+$ controls the *step length*
- the covariance matrix $\mathbf{C} \in \mathbb{R}^{n \times n}$ determines the shape of the distribution ellipsoid

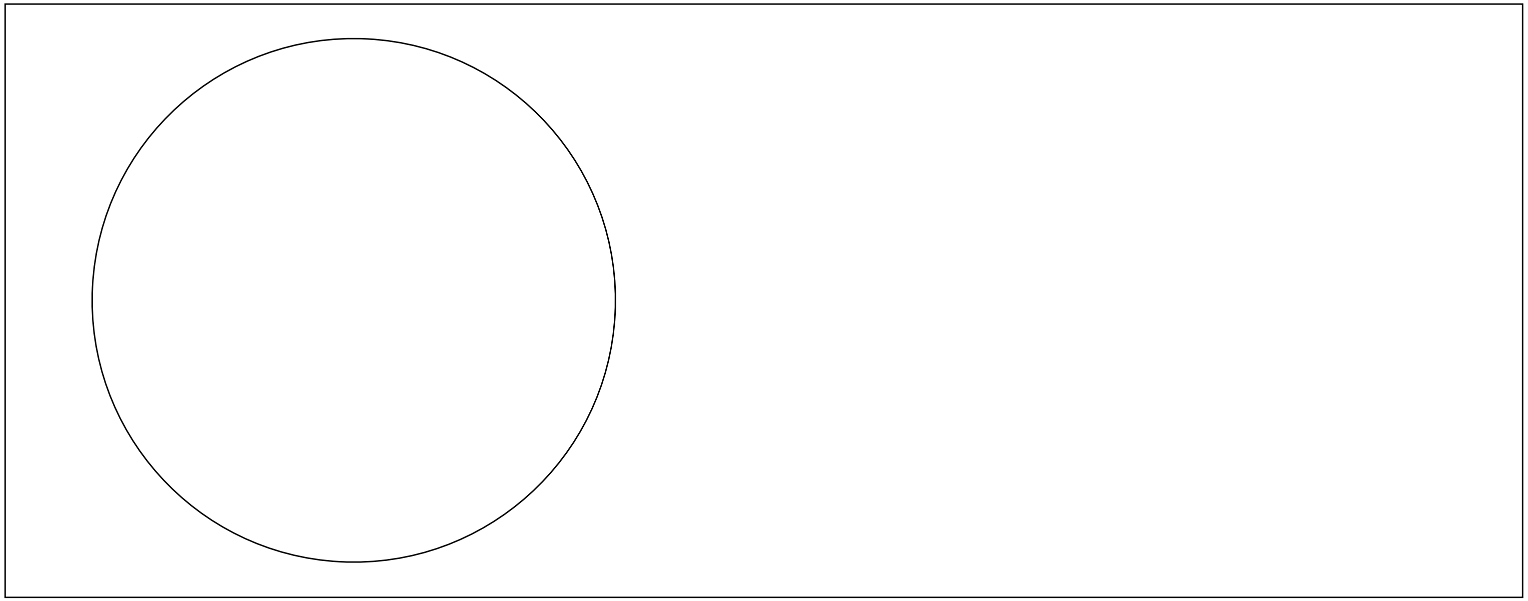
The remaining question is how to update $\mathbf{C}$.

Covariance Matrix Adaptation

Rank-One Update

[Hansen, Ostermeier, 2001]

$$\boldsymbol{m} \leftarrow \boldsymbol{m} + \sigma \boldsymbol{y}_w, \quad \boldsymbol{y}_w = \sum_{i=1}^{\mu} w_i \, \boldsymbol{y}_{i:\lambda}, \quad \boldsymbol{y}_i \sim \mathcal{N}_i(\mathbf{0}, \mathbf{C})$$

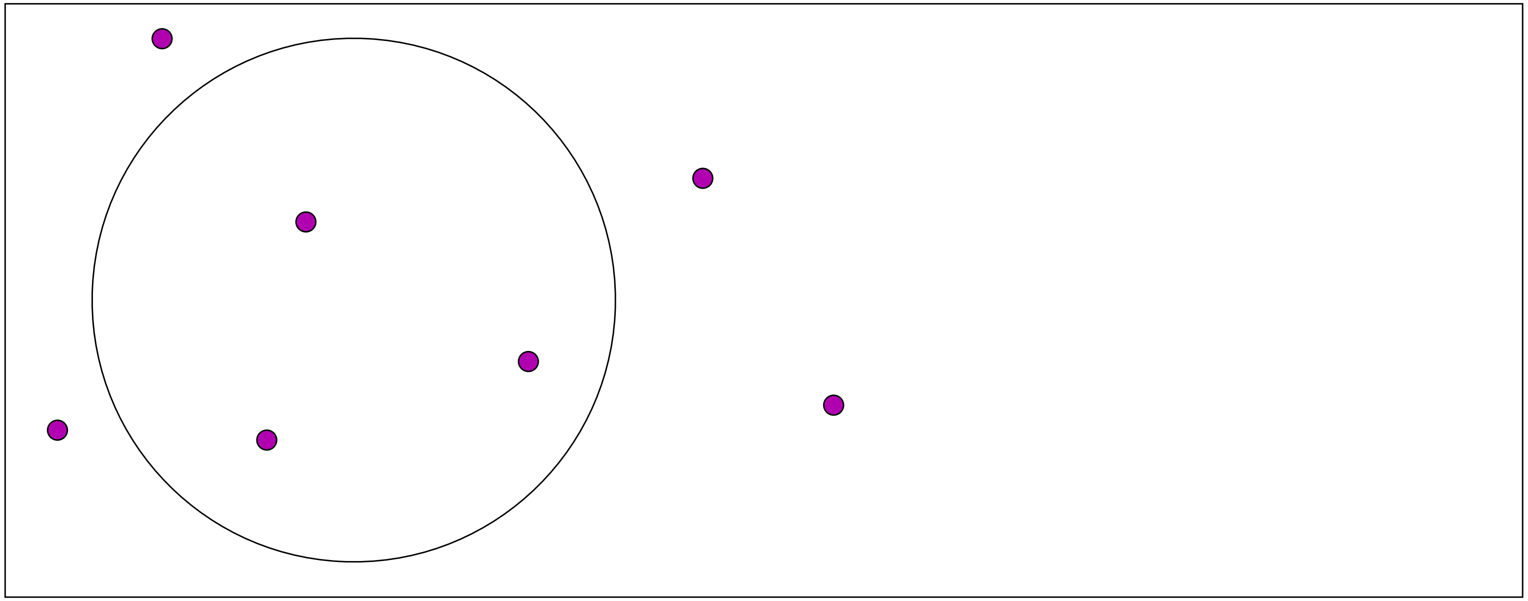


initial distribution, $\mathbf{C} = \mathbf{I}$

Covariance Matrix Adaptation

Rank-One Update

$$\boldsymbol{m} \leftarrow \boldsymbol{m} + \sigma \boldsymbol{y}_w, \quad \boldsymbol{y}_w = \sum_{i=1}^{\mu} w_i \, \boldsymbol{y}_{i:\lambda}, \quad \boldsymbol{y}_i \sim \mathcal{N}_i(\mathbf{0}, \mathbf{C})$$

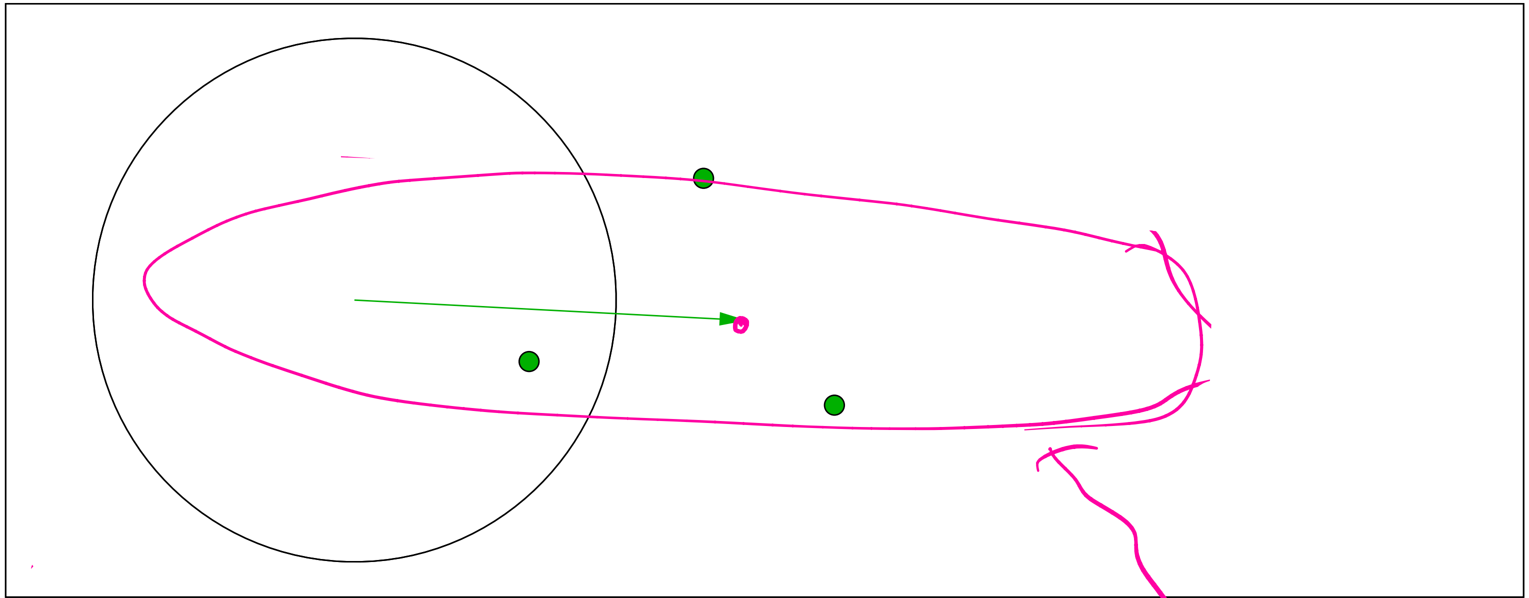


initial distribution, $\mathbf{C} = \mathbf{I}$

Covariance Matrix Adaptation

Rank-One Update

$$\boldsymbol{m} \leftarrow \boldsymbol{m} + \sigma \boldsymbol{y}_w, \quad \boldsymbol{y}_w = \sum_{i=1}^{\mu} w_i \, \boldsymbol{y}_{i:\lambda}, \quad \boldsymbol{y}_i \sim \mathcal{N}_i(\mathbf{0}, \mathbf{C})$$



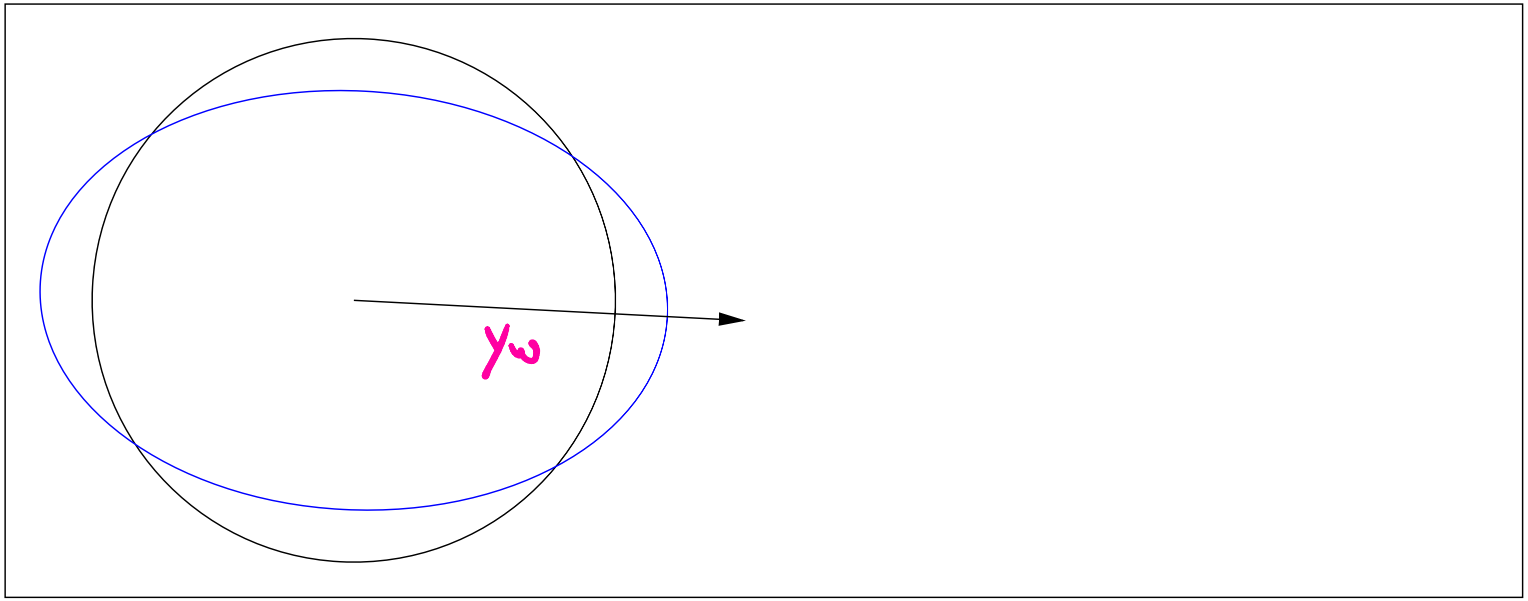
$\boldsymbol{y}_w$, movement of the population mean $\boldsymbol{m}$ (disregarding σ)

I would like to have this

Covariance Matrix Adaptation

Rank-One Update

$$\boldsymbol{m} \leftarrow \boldsymbol{m} + \sigma \boldsymbol{y}_w, \quad \boldsymbol{y}_w = \sum_{i=1}^{\mu} w_i \, \boldsymbol{y}_{i:\lambda}, \quad \boldsymbol{y}_i \sim \mathcal{N}_i(\mathbf{0}, \mathbf{C})$$



mixture of distribution $\mathbf{C}$ and step $\boldsymbol{y}_w$,

$$\mathbf{C} \leftarrow 0.8 \times \mathbf{C} + 0.2 \times \underbrace{\boldsymbol{y}_w \boldsymbol{y}_w^{\mathrm{T}}}_{\text{semi-definite symmetric}}$$

$y_w y_w^T$ is a rank-one matrix

$$\begin{pmatrix} y_w \end{pmatrix} \begin{pmatrix} & y_w^T & \\ & & \end{pmatrix} = \begin{pmatrix} (y_w)_1 y_w, (y_w)_2 y_w, \dots, (y_w)_n y_w \end{pmatrix}$$

$$\mathrm{Vect}\left((y_w)_1 y_w, \dots, (y_w)_n y_w \right) = \mathrm{Vect}(y_w)$$

$$\mathrm{rank}(y_w y_w^T) = 1$$

$\mathrm{Ker}(y_w y_w^T)$ has dimension $n-1$.

Check that y_w is eigenvector associated to non-zero eigenvalue:

$$y_w \underbrace{y_w^T y_w}_{\|y_w\|^2} = \|y_w\|^2 y_w$$

$\rightarrow$ y_w is eigenvector associated to the eigenvalue $\|y_w\|^2$.

Deduce eigenvectors and eigenvalue of :

$$(1-\mathrm{cov})\,Id + \mathrm{cov}\; y_w y_w^T$$

1/ y_w is also eigenvector of $(1-\mathrm{cov})\,Id + \mathrm{cov}\; y_w y_w^T$:

$$\left[(1-\mathrm{cov})\,Id + \mathrm{cov}\; y_w y_w^T\right] y_w = (1-\mathrm{cov})\,y_w + \mathrm{cov}\,\|y_w\|^2\, y_w$$

$$= (1-\mathrm{cov}) + \mathrm{cov}\,\|y_w\|^2\, y_w$$

I can complement $(y_w, \underbrace{y_2, \dots, y_n}_{\mathrm{Ker}(y_w y_w^T)})$ into orthogonal basis.

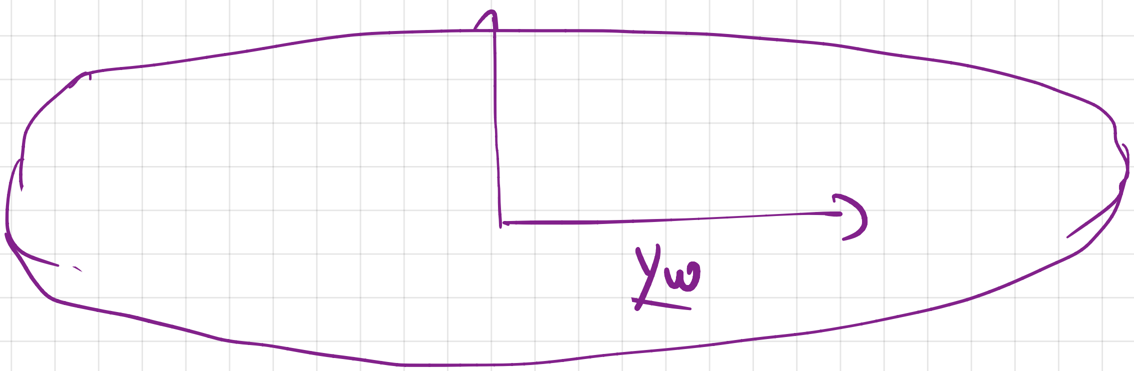
For y_i vector $\left(y_w y_w^T\right) y_i = 0$

$$\left[(1-\mathrm{cov})\,Id + \mathrm{cov}\,(y_w y_w^T)\right] y_i = (1-\mathrm{cov})\,y_i$$

$(y_w, y_2, \dots, y_m)$ forms an orthogonal basis of eigenvectors $(1-\mathrm{cov})\,\mathrm{Id} + \mathrm{cov}\,(y_w y_w^T)$ associated to the eigenvalue :

$$\left((1-\mathrm{cov}) + \mathrm{cov}\,\|y_w\|^2,\ (1-\mathrm{cov}), \dots, (1-\mathrm{cov})\right)$$

The ellipsoid associated to the density related to $C = (1-\mathrm{cov})\,I_d + \mathrm{cov}\,(y_w y_w^T)$

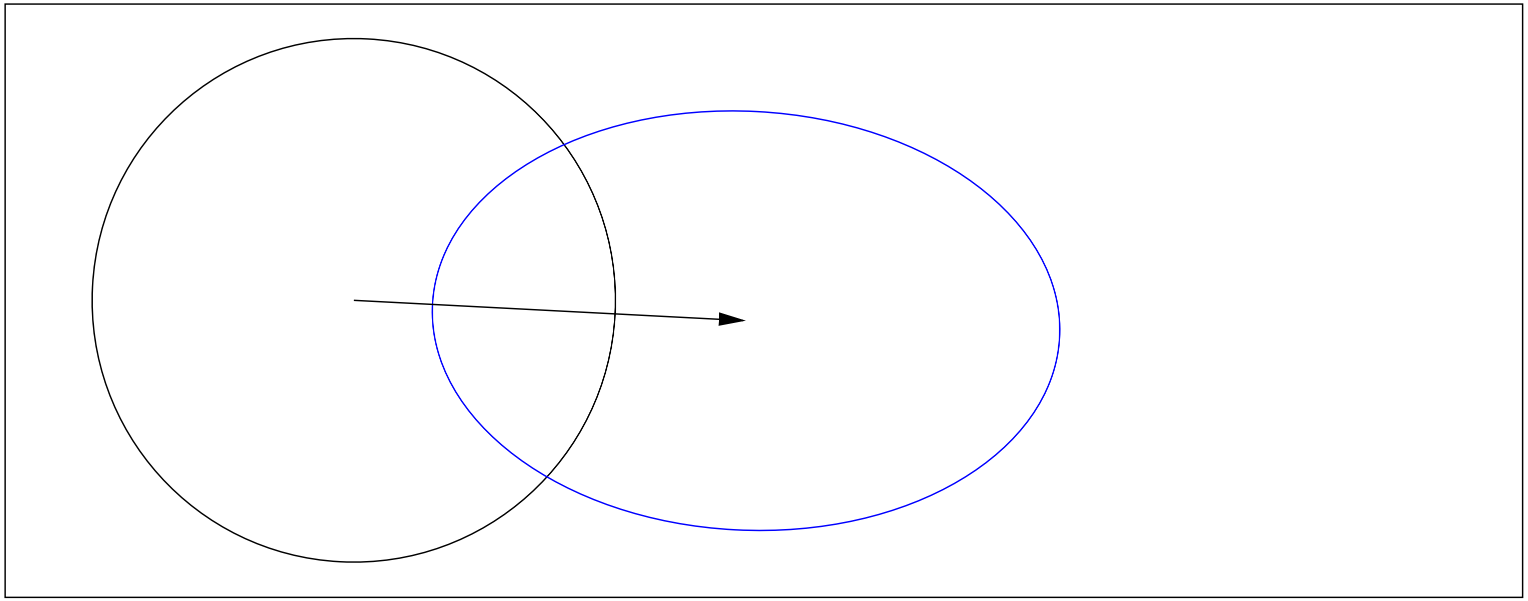


Conclusion: The rank-one update elongates the iso-density ellipsoid in the direction y_w.

Covariance Matrix Adaptation

Rank-One Update

$$\boldsymbol{m} \leftarrow \boldsymbol{m} + \sigma \boldsymbol{y}_w, \quad \boldsymbol{y}_w = \sum_{i=1}^{\mu} w_i \, \boldsymbol{y}_{i:\lambda}, \quad \boldsymbol{y}_i \sim \mathcal{N}_i(\mathbf{0}, \mathbf{C})$$

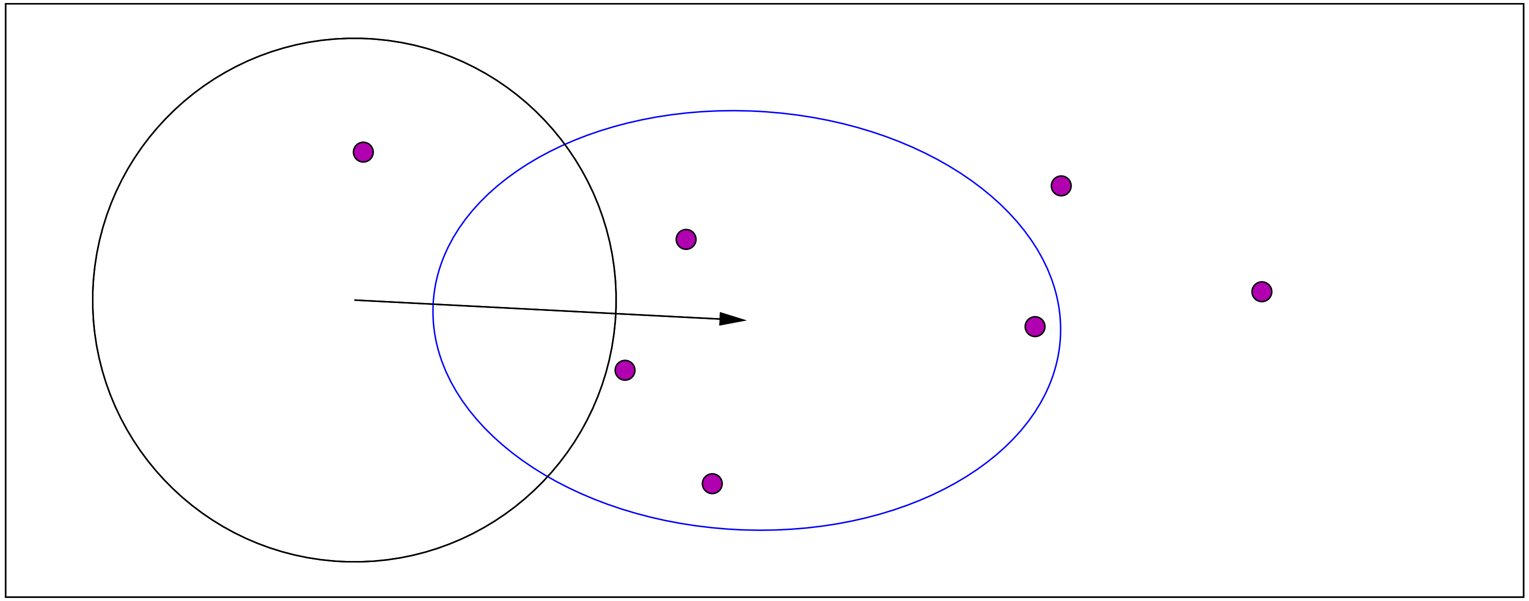


new distribution (disregarding σ)

Covariance Matrix Adaptation

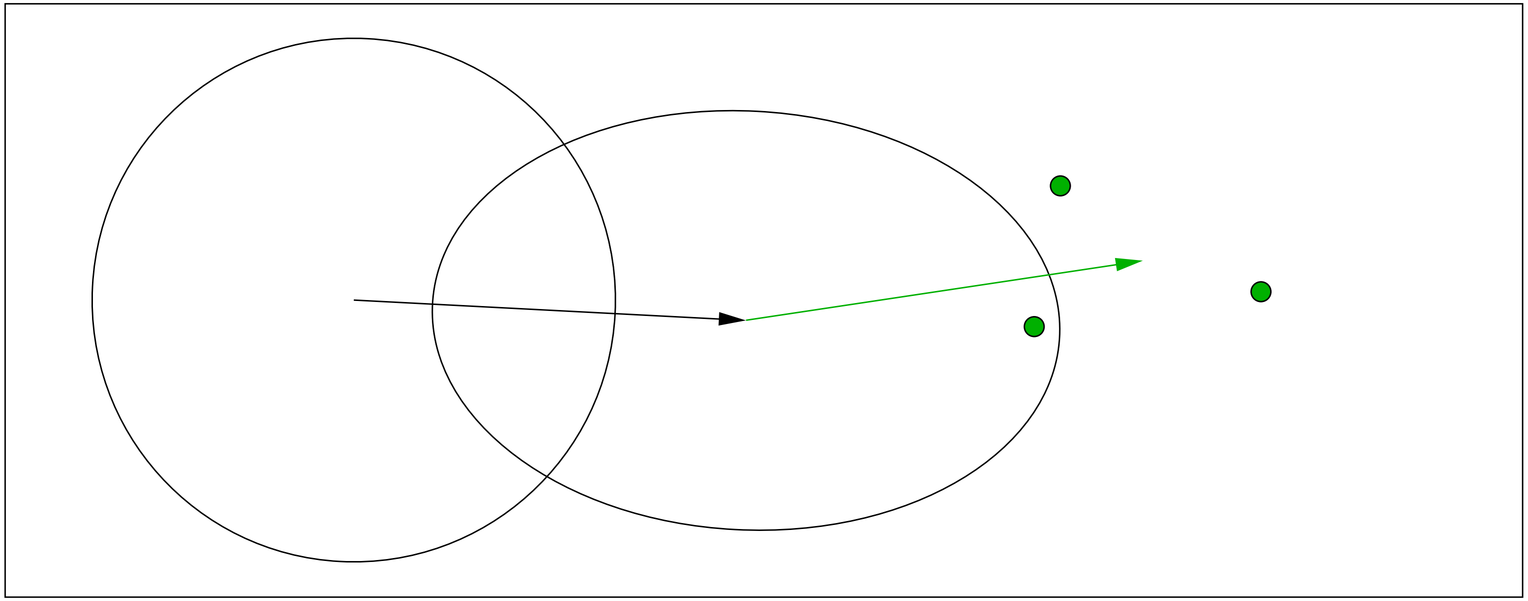
Rank-One Update

$$\boldsymbol{m} \leftarrow \boldsymbol{m} + \sigma \boldsymbol{y}_w, \quad \boldsymbol{y}_w = \sum_{i=1}^{\mu} w_i \, \boldsymbol{y}_{i:\lambda}, \quad \boldsymbol{y}_i \sim \mathcal{N}_i(\mathbf{0}, \mathbf{C})$$



new distribution (disregarding σ)

$$\boldsymbol{m} \leftarrow \boldsymbol{m} + \sigma \boldsymbol{y}_w, \quad \boldsymbol{y}_w = \sum_{i=1}^{\mu} w_i \, \boldsymbol{y}_{i:\lambda}, \quad \boldsymbol{y}_i \sim \mathcal{N}_i(\mathbf{0}, \mathbf{C})$$

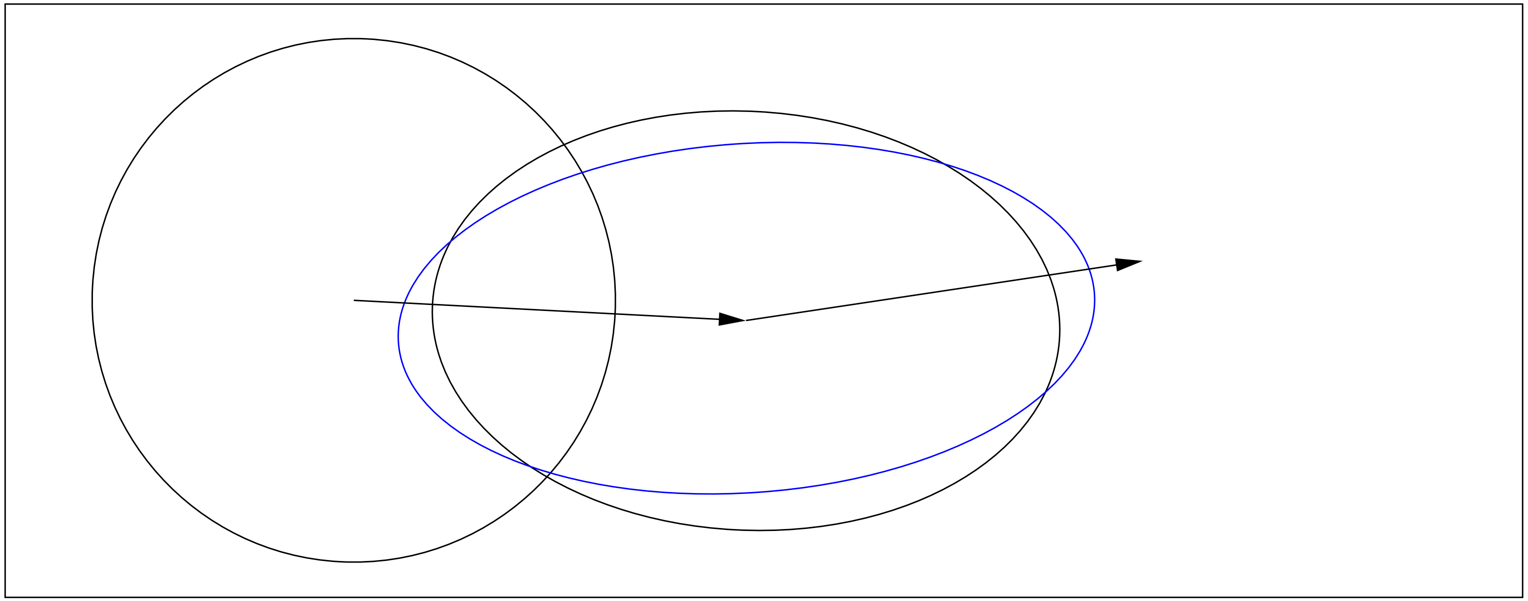


movement of the population mean $\boldsymbol{m}$

Covariance Matrix Adaptation

Rank-One Update

$$\boldsymbol{m} \leftarrow \boldsymbol{m} + \sigma \boldsymbol{y}_w, \quad \boldsymbol{y}_w = \sum_{i=1}^{\mu} w_i \, \boldsymbol{y}_{i:\lambda}, \quad \boldsymbol{y}_i \sim \mathcal{N}_i(\mathbf{0}, \mathbf{C})$$



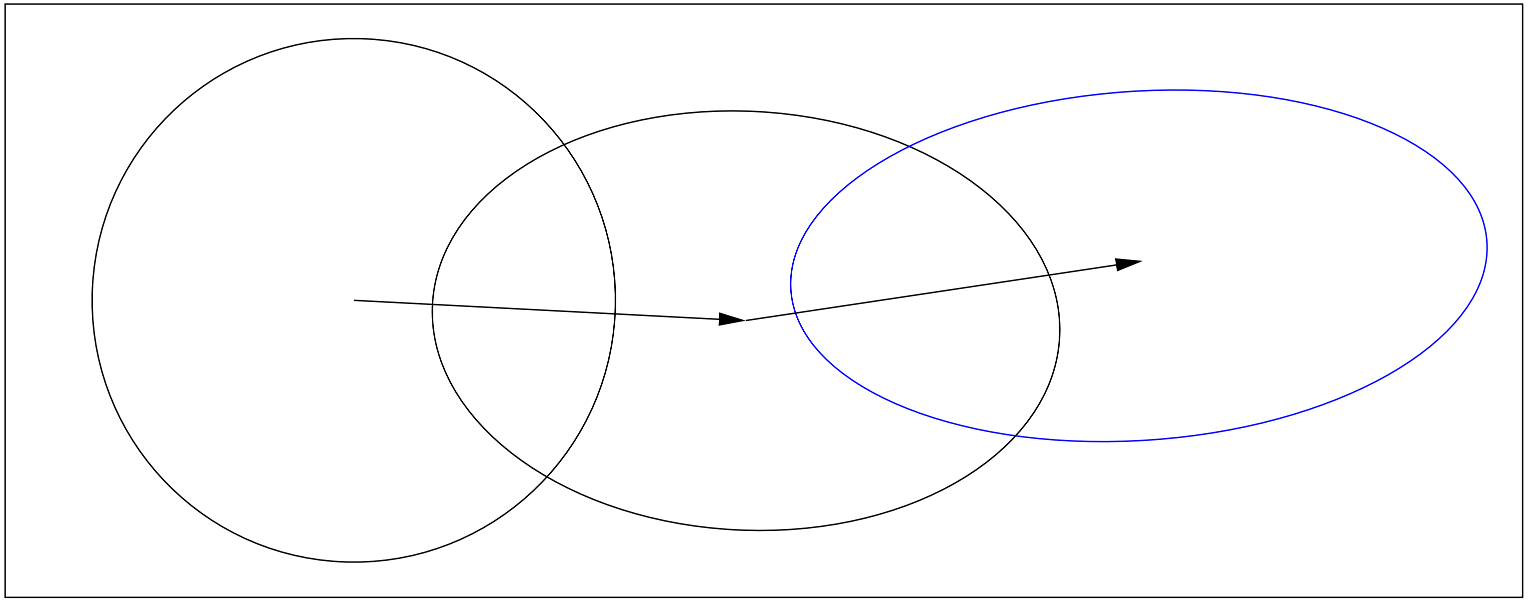
mixture of distribution $\mathbf{C}$ and step $\boldsymbol{y}_w$,

$\mathbf{C} \leftarrow 0.8 \times \mathbf{C} + 0.2 \times \boldsymbol{y}_w \boldsymbol{y}_w^{\mathrm{T}}$

Covariance Matrix Adaptation

Rank-One Update

$$\boldsymbol{m} \leftarrow \boldsymbol{m} + \sigma \boldsymbol{y}_w, \quad \boldsymbol{y}_w = \sum_{i=1}^{\mu} w_i \, \boldsymbol{y}_{i:\lambda}, \quad \boldsymbol{y}_i \sim \mathcal{N}_i(\mathbf{0}, \mathbf{C})$$



new distribution,

$\mathbf{C} \leftarrow 0.8 \times \mathbf{C} + 0.2 \times \boldsymbol{y}_w \boldsymbol{y}_w^{\mathrm{T}}$

the ruling principle: the adaptation increases the likelihood of successful steps, $\boldsymbol{y}_w$, to appear again

Covariance Matrix Adaptation

Rank-One Update

Initialize $\boldsymbol{m} \in \mathbb{R}^n$, and $\mathbf{C} = \mathbf{I}$, set $\sigma = 1$, learning rate $c_{\mathrm{cov}} \approx 2/n^2$
While not terminate

$$
\begin{aligned}
\boldsymbol{x}_i &= \boldsymbol{m} + \sigma\, \boldsymbol{y}_i, \qquad \boldsymbol{y}_i \sim \mathcal{N}_i(\mathbf{0}, \mathbf{C}), \\
\boldsymbol{m} &\leftarrow \boldsymbol{m} + \sigma\, \boldsymbol{y}_w \qquad \text{where } \boldsymbol{y}_w = \sum_{i=1}^{\mu} w_i\, \boldsymbol{y}_{i:\lambda} \\
\mathbf{C} &\leftarrow (1 - c_{\mathrm{cov}})\mathbf{C} + c_{\mathrm{cov}} \mu_w \underbrace{\boldsymbol{y}_w \boldsymbol{y}_w^{\mathrm{T}}}_{\text{rank-one}} \qquad \text{where } \mu_w = \frac{1}{\sum_{i=1}^{\mu} {w_i}^2} \geq 1
\end{aligned}
$$

REST OF THE CLASS: Exercice "Running and Understanding CNA-Es" (see exercice sheet)