Emergent Proximo-Distal Maturation through Adaptive Exploration

Freek Stulp and Pierre-Yves Oudeyer

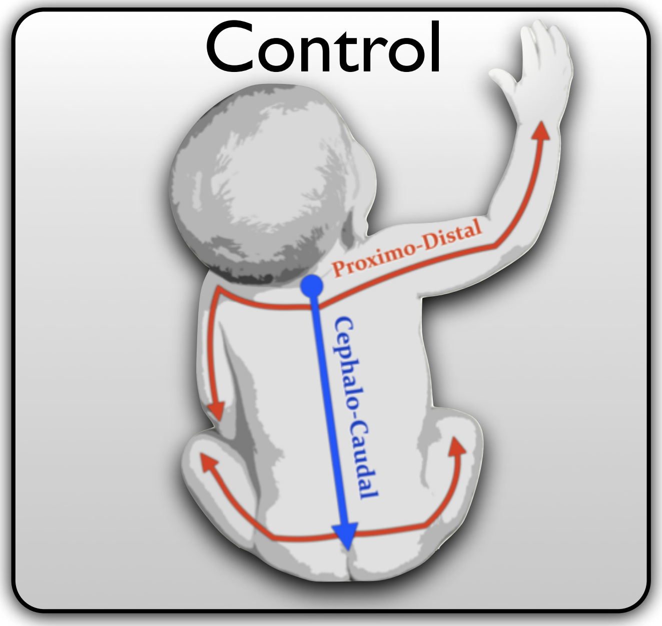
FLOWERS team

ENSTA-ParisTech and Inria, France

http://flowers.inria.fr

Maturational skill learning: freeing and freezing of *motor* DOFs in humans

Reaching



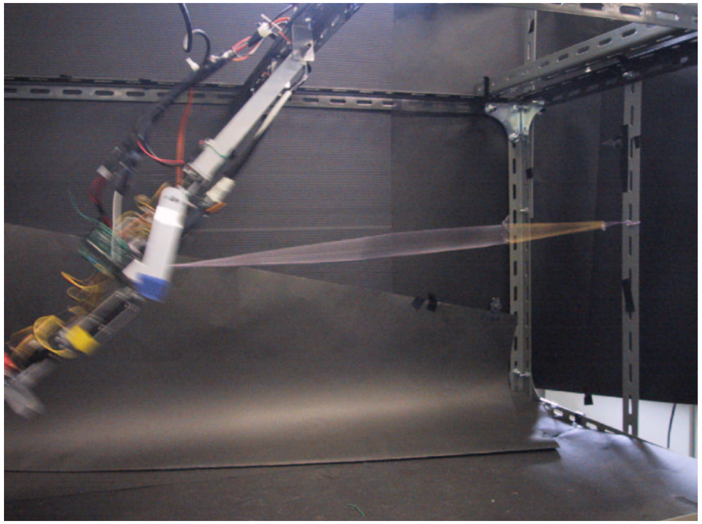
(Bjorklund, 1997; Turkewitz and Kenny, 1985)

Skiing

(Bernstein, 1967; Verijken et al., 1992)

Motor maturation for efficient skill learning in robots

One task: Swinging



Berthouze and Lungarella, 2004)

(See also Ivanchenko and Jacobs, 2003)

MATURATION:
Freeing and freezing of DOFs

Fixed or random

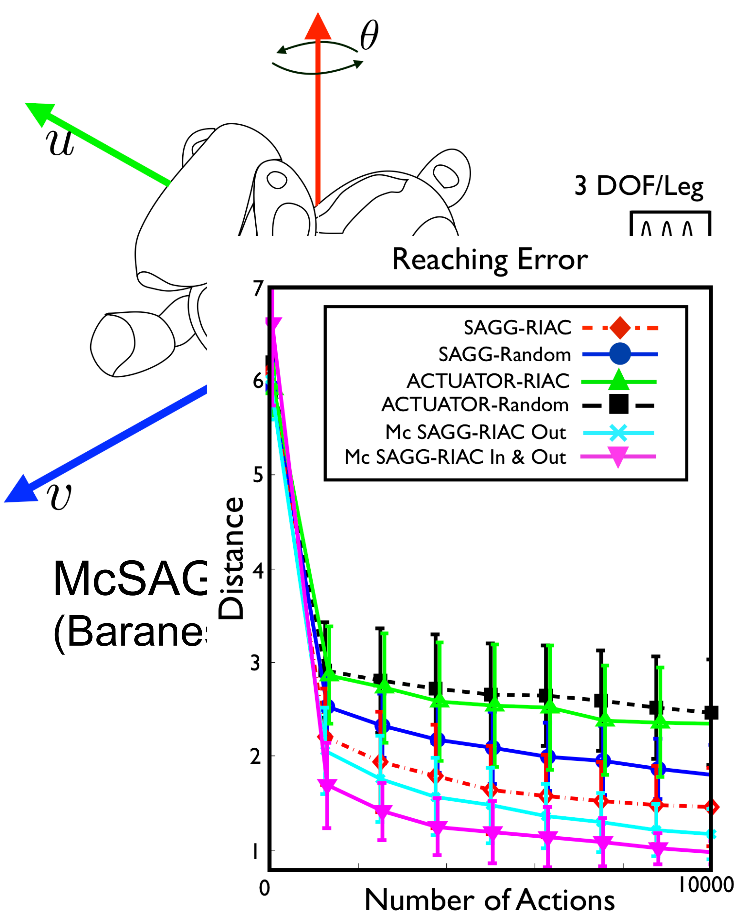
↓

OPTIMIZATION
Reinforcement learning

Learn one skill
Simple RL/Opt.

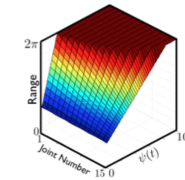
Adaptive maturation for skill learning in robots

Many tasks:
Omnidirectional locomotion



McSAG
(Baranes

MATURATION:
Adaptive freeing
of DOFs



Adaptive clock

INTRINSIC MOTIVATION
Competence progress

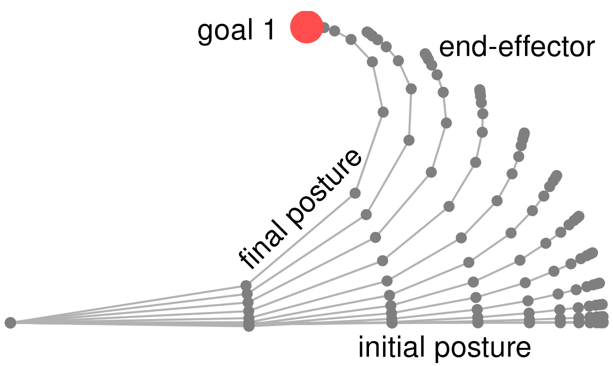
Learn field of skills

OPTIMIZATION
Reinforcement learning

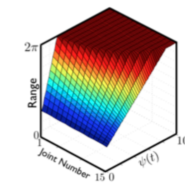
Simple local RL/Opt.

Initial goal:
Adaptive maturation controlled **by** PI^2_{CMA-ES}

One task: reaching



MATURATION:
Freeing and freezing of DOFs



Adaptive clock

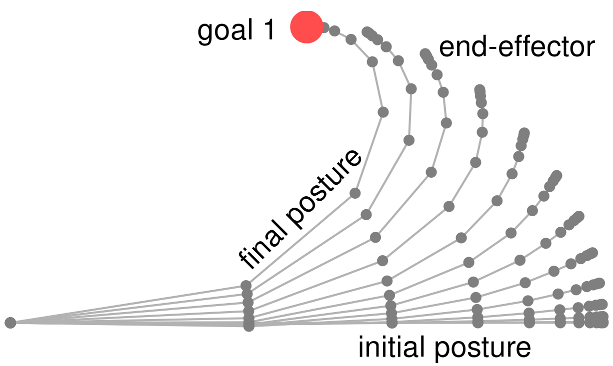
OPTIMIZATION
Reinforcement learning

SoA RL

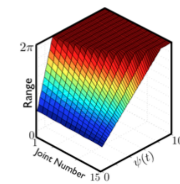
PI^2_{CMA-ES}

What we got:
Emergent maturation ***from*** PI^2_{CMA-ES}

One task: reaching



MATURATION:
Freeing and freezing of DOFs



Emergent Schedule

OPTIMIZATION
Reinforcement learning

SoA Episodic RL
PI^2_{CMA-ES}

PI^2_{CMA-ES} Policy Improvement with Path Integrals and Covariant Matrix Adaptation

information geometric optimization

CMA-ES

- reward-weighted averaging
- black-box optimization
- covariance matrix updating

(Hansen and Ostermeier, 2001)

stochastic optimal control

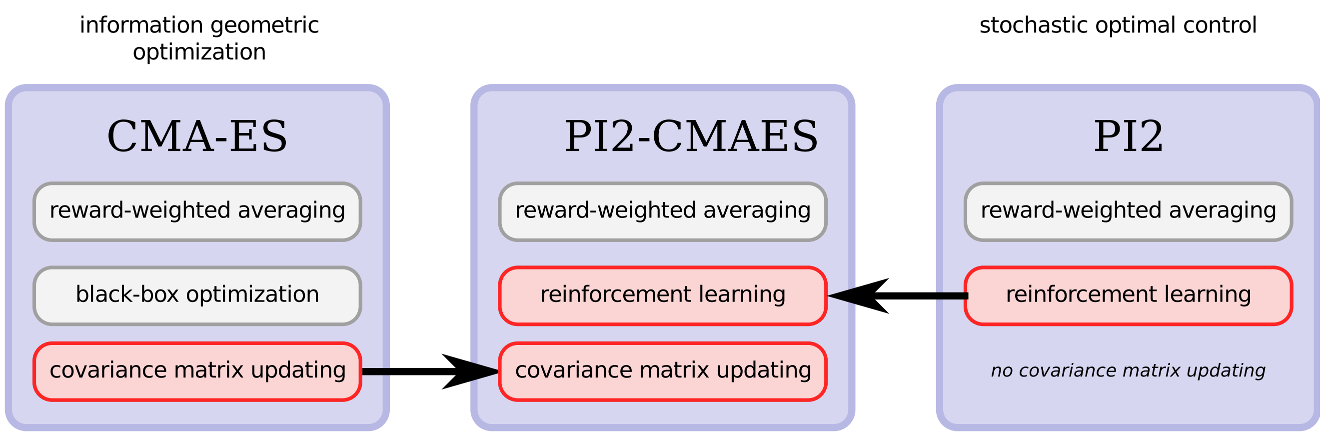
PI2

- reward-weighted averaging
- reinforcement learning

no covariance matrix updating

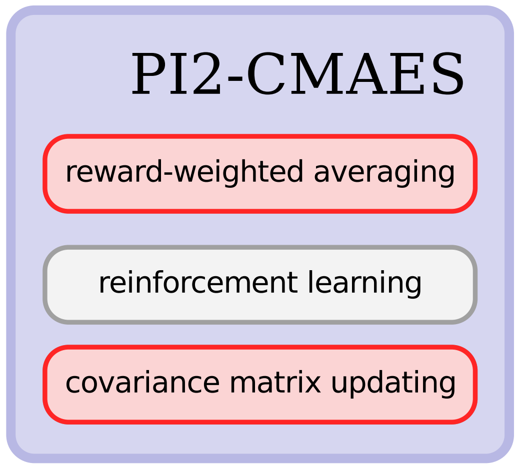
(Theodorou et al., 2010)

PI^2_{CMA-ES} Policy Improvement with Path Integrals and Covariant Matrix Adaptation



(Stulp and Sigaud, 2012)

PI^2_{CMA-ES} Policy Improvement with Path Integrals and Covariant Matrix Adaptation



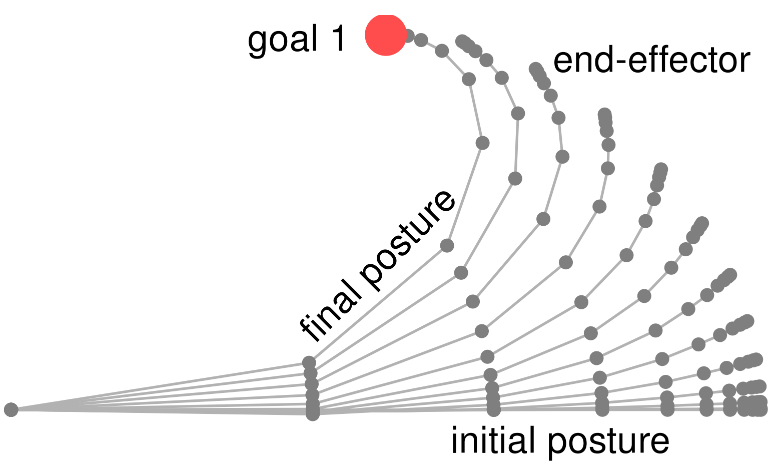
Application to reaching

- Task: reach to a goal in the workspace
- 10-DOF kinematically simulated 'arm' in 2-D plane
- Policy representation:

$$\ddot{q}_{m,t} = \mathbf{g}(t)^\intercal \boldsymbol{\theta}_m \qquad \text{Acc. of joint } m \qquad (1)$$

$$[\mathbf{g}(t)]_b = \frac{\Psi_b(t)}{\sum_{b=1}^{B} \Psi_b(t)} \text{ with } \Psi_b(t) = \exp\left(-(t-c_b)^2/w^2\right) \qquad \text{Basis functions} \qquad (2)$$

- Duration of movement is 0.5s
- Initialy, $\boldsymbol{\theta} = 0$ (no movement)



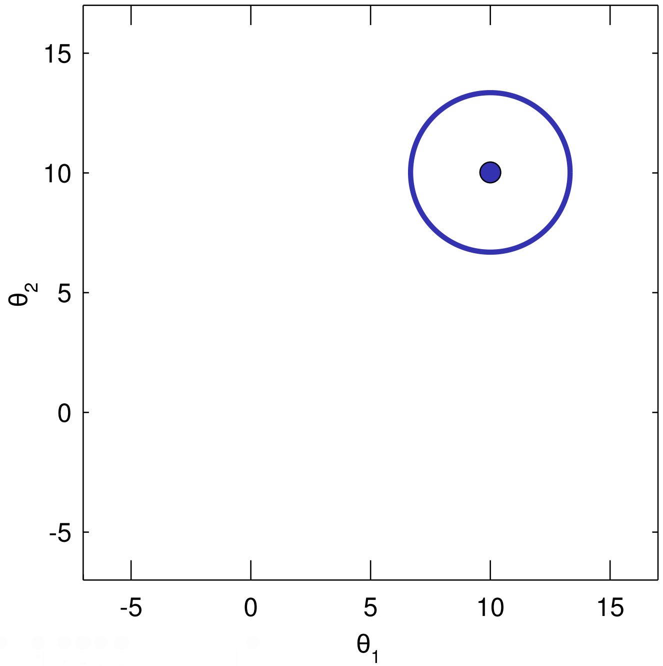
Cost function:

$$\phi_{t_N} = 10^4 ||\mathbf{x}_{t_N} - g||^2 + \max(\mathbf{q}_{t_N}) \qquad \text{Terminal cost} \qquad (1)$$

$$r_t = 10^{-5} \frac{\sum_{m=1}^{M} (M+1-m)(\ddot{q}_{t,m})^2}{\sum_{m=1}^{M} (M+1-m)} \qquad \text{Immediate cost} \qquad (2)$$

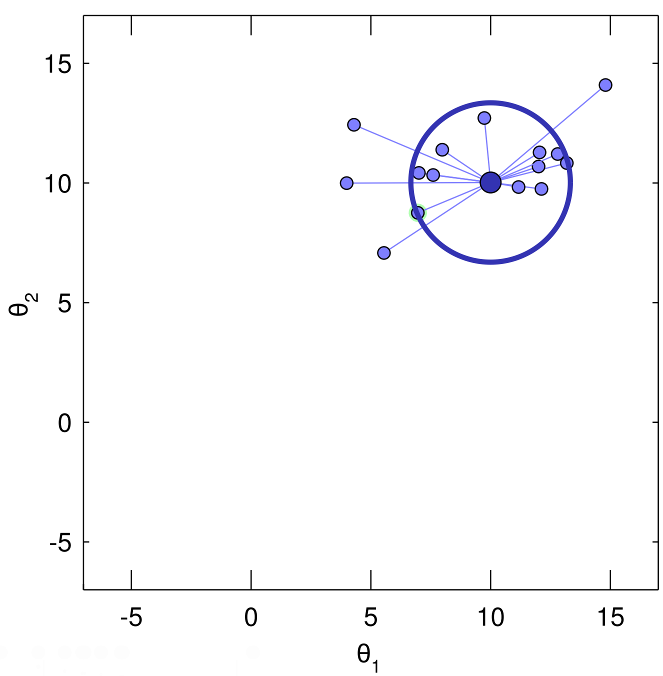
PI^2_{CMA-ES}: Reward-Weighted Averaging

$\mathcal{N}(\boldsymbol{\theta}, \Sigma)$



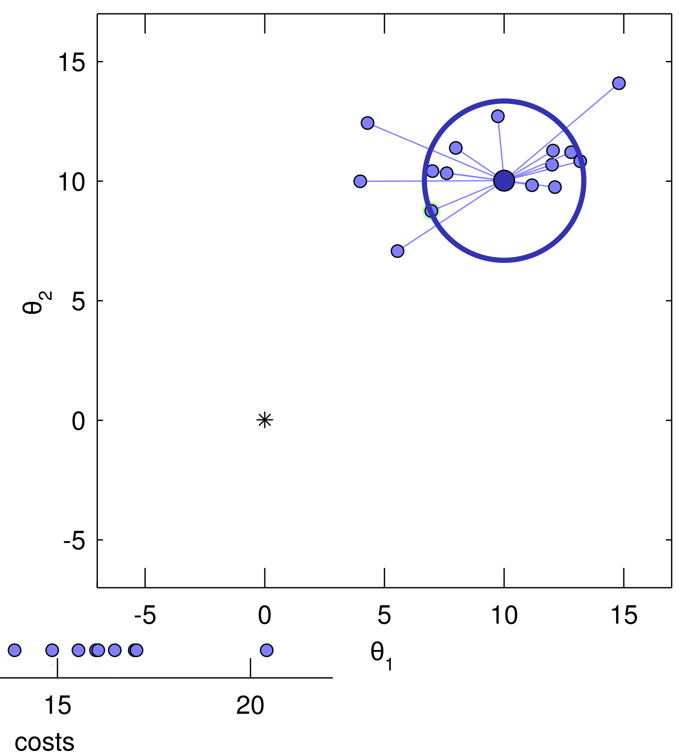
PI^2_{CMA-ES}: Reward-Weighted Averaging

$\boldsymbol{\theta}_{k=1\ldots K} \sim \mathcal{N}(\boldsymbol{\theta}, \Sigma)$



PI^2_{CMA-ES}: Reward-Weighted Averaging

$$\boldsymbol{\theta}_{k=1\dots K} \sim \mathcal{N}(\boldsymbol{\theta}, \Sigma)$$
$$\forall k \;\; J_k = J(\boldsymbol{\theta}_k)$$

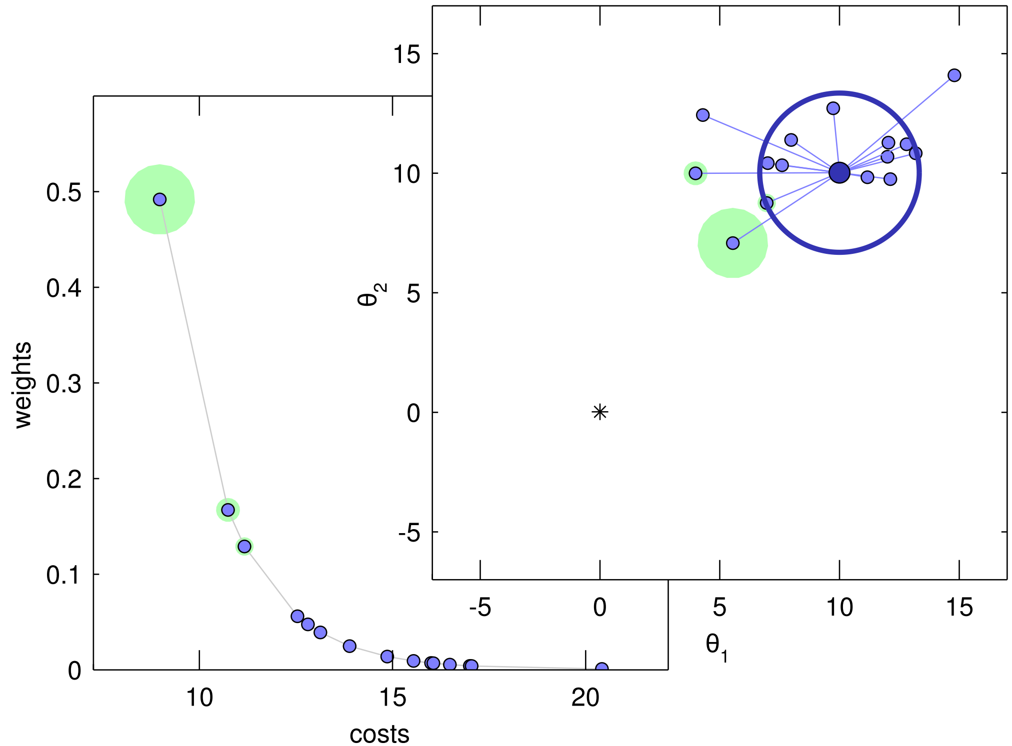


PI^2_{CMA-ES} : Reward-Weighted Averaging

$$\boldsymbol{\theta}_{k=1\dots K} \sim \mathcal{N}(\boldsymbol{\theta}, \Sigma)$$

$$\forall k \;\; J_k = J(\boldsymbol{\theta}_k)$$

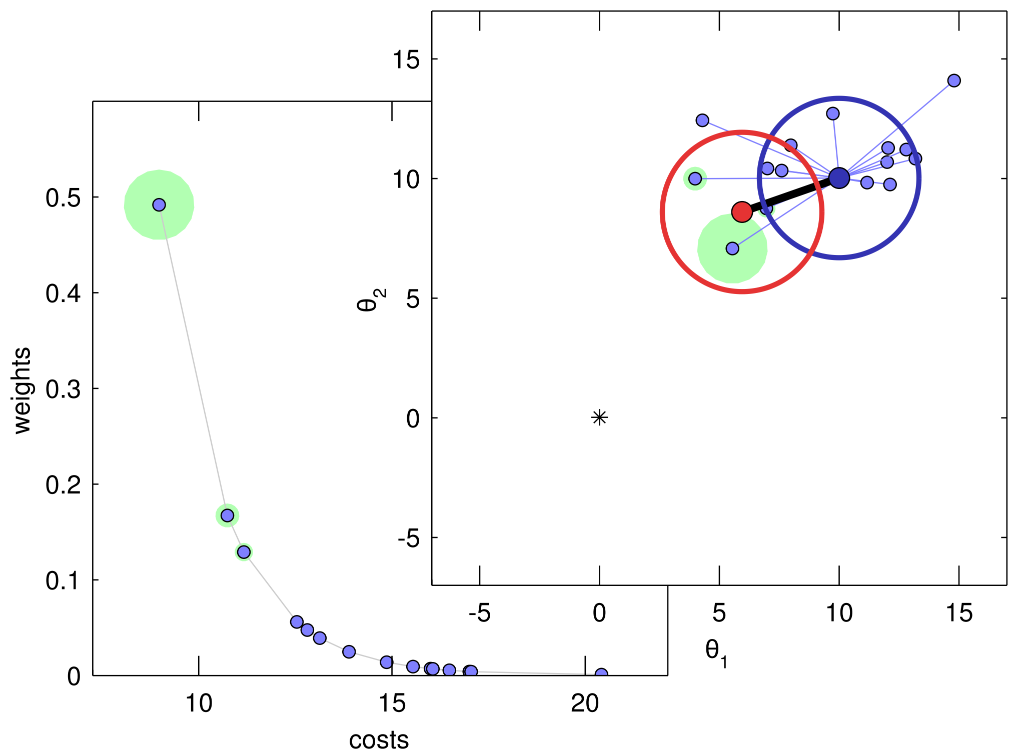
$$\forall k \;\; P_k = \frac{e^{\left(\frac{-h(J_k - \min(J_k))}{\max(J_k) - \min(J_k)}\right)}}{\sum_{l=1}^{K} e^{\left(\frac{-h(J_l - \min(J_l))}{\max(J_l) - \min(J_l)}\right)}}$$



PI^2_{CMA-ES} : Reward-Weighted Averaging

$$\boldsymbol{\theta}_{k=1\dots K} \sim \mathcal{N}(\boldsymbol{\theta}, \Sigma)$$
$$\forall k \;\; J_k = J(\boldsymbol{\theta}_k)$$
$$\forall k \;\; P_k = \frac{e^{\left(\frac{-h(J_k - \min(J_k))}{\max(J_k) - \min(J_k)}\right)}}{\sum_{l=1}^{K} e^{\left(\frac{-h(J_l - \min(J_l))}{\max(J_l) - \min(J_l)}\right)}}$$

$$\boldsymbol{\theta}^{new} = \sum_{k=1}^{K} P_k \boldsymbol{\theta}_k$$



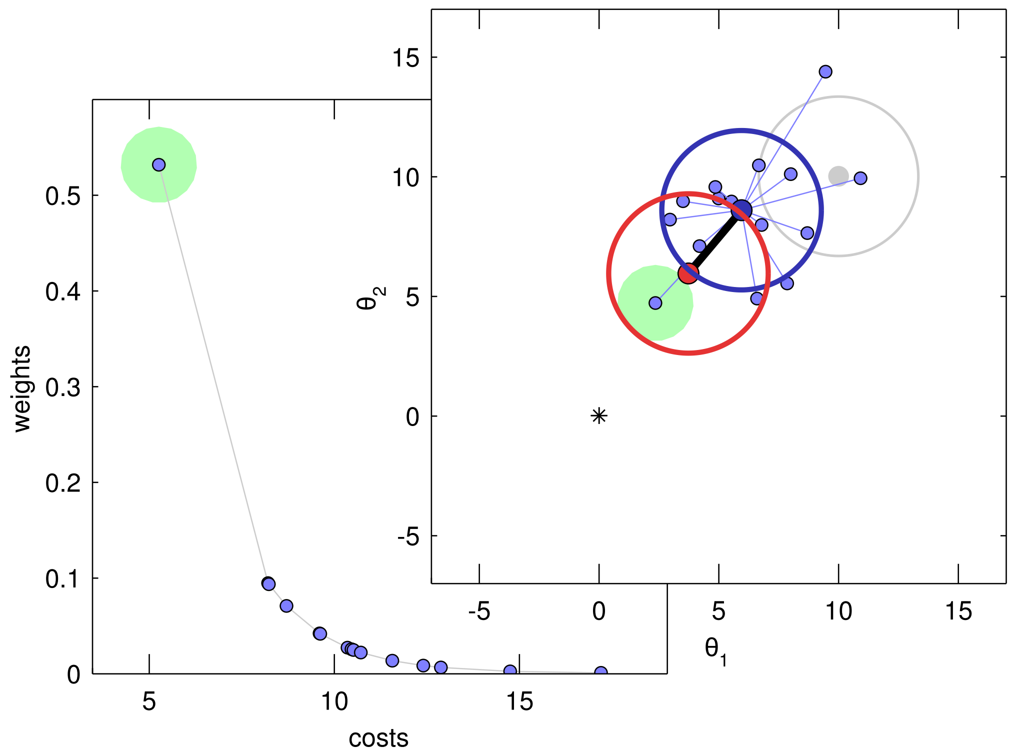
PI^2_{CMA-ES}: Reward-Weighted Averaging

$$\boldsymbol{\theta}_{k=1\ldots K} \sim \mathcal{N}(\boldsymbol{\theta}, \Sigma)$$

$$\forall k \; J_k = J(\boldsymbol{\theta}_k)$$

$$\forall k \; P_k = \frac{e^{\left(\frac{-h(J_k - \min(J_k))}{\max(J_k) - \min(J_k)}\right)}}{\sum_{l=1}^{K} e^{\left(\frac{-h(J_l - \min(J_l))}{\max(J_l) - \min(J_l)}\right)}}$$

$$\boldsymbol{\theta}^{new} = \sum_{k=1}^{K} P_k \boldsymbol{\theta}_k$$



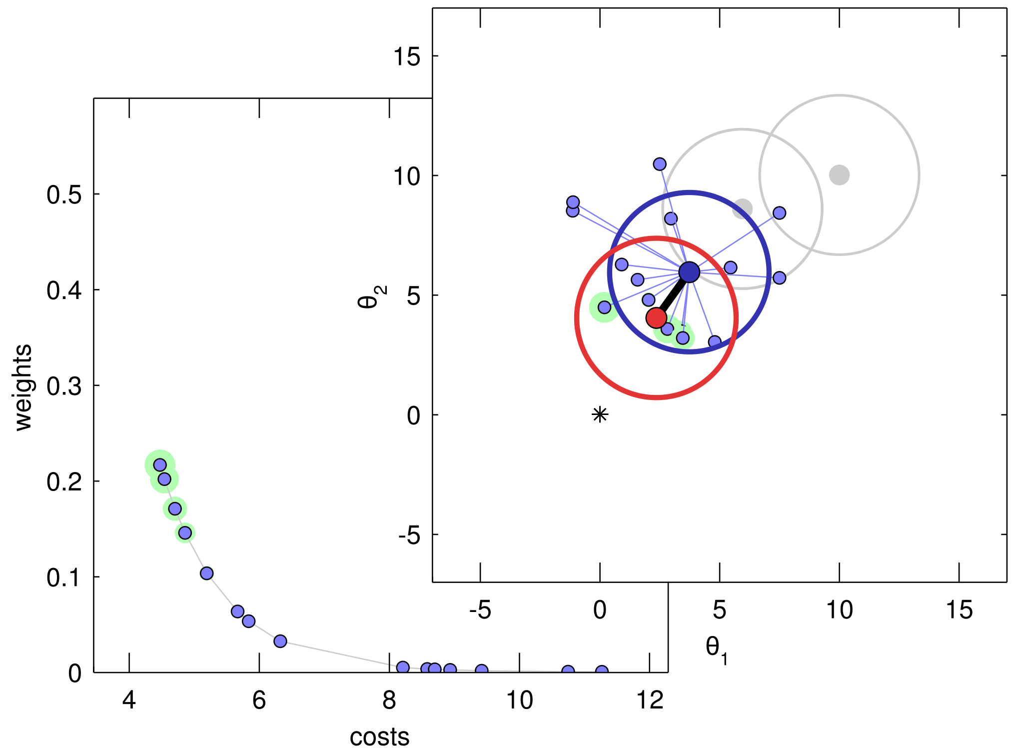
PI^2_{CMA-ES}: Reward-Weighted Averaging

$$\boldsymbol{\theta}_{k=1\ldots K} \sim \mathcal{N}(\boldsymbol{\theta}, \Sigma)$$

$$\forall k \;\; J_k = J(\boldsymbol{\theta}_k)$$

$$\forall k \;\; P_k = \frac{e^{\left(\frac{-h(J_k - \min(J_k))}{\max(J_k) - \min(J_k)}\right)}}{\sum_{l=1}^{K} e^{\left(\frac{-h(J_l - \min(J_l))}{\max(J_l) - \min(J_l)}\right)}}$$

$$\boldsymbol{\theta}^{new} = \sum_{k=1}^{K} P_k \boldsymbol{\theta}_k$$



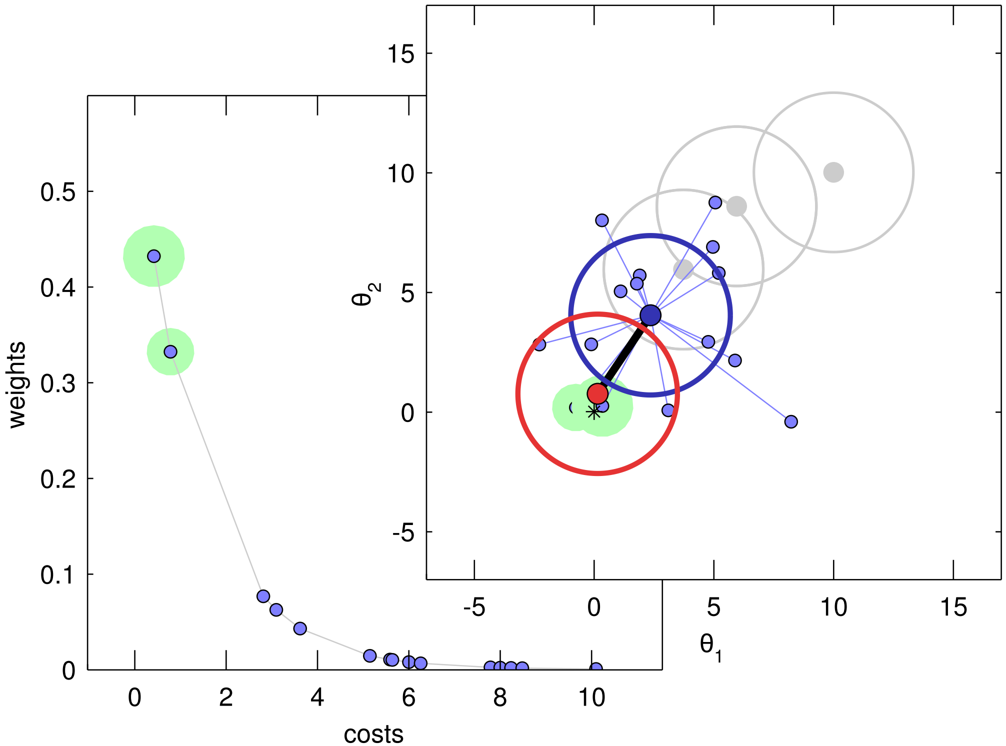
PI^2_{CMA-ES}: Reward-Weighted Averaging

$$\boldsymbol{\theta}_{k=1\dots K} \sim \mathcal{N}(\boldsymbol{\theta}, \Sigma)$$

$$\forall k \;\; J_k = J(\boldsymbol{\theta}_k)$$

$$\forall k \;\; P_k = \frac{e^{\left(\frac{-h(J_k - \min(J_k))}{\max(J_k) - \min(J_k)}\right)}}{\sum_{l=1}^{K} e^{\left(\frac{-h(J_l - \min(J_l))}{\max(J_l) - \min(J_l)}\right)}}$$

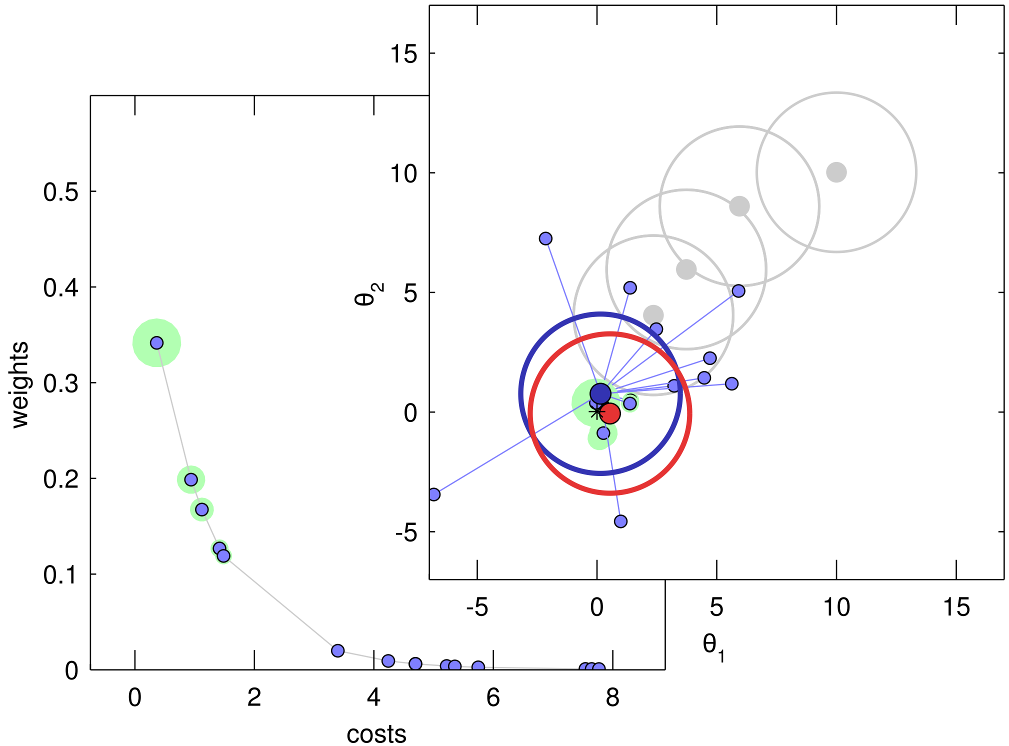
$$\boldsymbol{\theta}^{new} = \sum_{k=1}^{K} P_k \boldsymbol{\theta}_k$$



PI^2_{CMA-ES}: Reward-Weighted Averaging

$$\boldsymbol{\theta}_{k=1\ldots K} \sim \mathcal{N}(\boldsymbol{\theta}, \Sigma)$$
$$\forall k \; J_k = J(\boldsymbol{\theta}_k)$$
$$\forall k \; P_k = \frac{e^{\left(\frac{-h(J_k - \min(J_k))}{\max(J_k) - \min(J_k)}\right)}}{\sum_{l=1}^{K} e^{\left(\frac{-h(J_l - \min(J_l))}{\max(J_l) - \min(J_l)}\right)}}$$

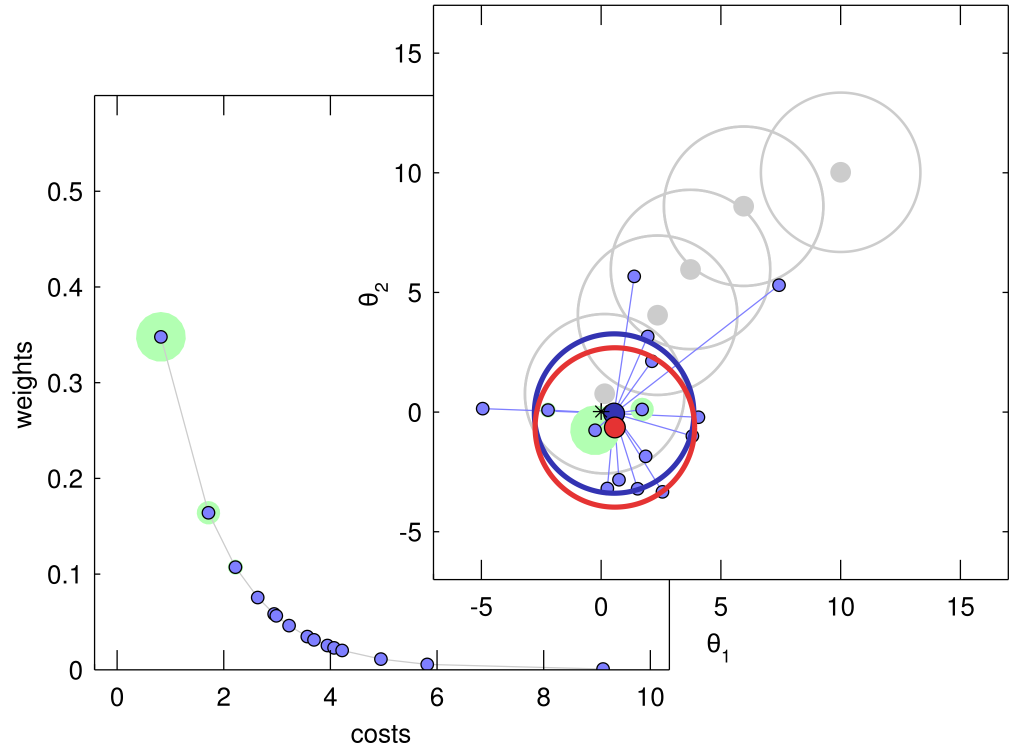
$$\boldsymbol{\theta}^{new} = \sum_{k=1}^{K} P_k \boldsymbol{\theta}_k$$



PI^2_{CMA-ES}: Reward-Weighted Averaging

$$\boldsymbol{\theta}_{k=1\ldots K} \sim \mathcal{N}(\boldsymbol{\theta}, \Sigma)$$
$$\forall k \; J_k = J(\boldsymbol{\theta}_k)$$
$$\forall k \; P_k = \frac{e^{\left(\frac{-h(J_k - \min(J_k))}{\max(J_k) - \min(J_k)}\right)}}{\sum_{l=1}^{K} e^{\left(\frac{-h(J_l - \min(J_l))}{\max(J_l) - \min(J_l)}\right)}}$$

$$\boldsymbol{\theta}^{new} = \sum_{k=1}^{K} P_k \boldsymbol{\theta}_k$$



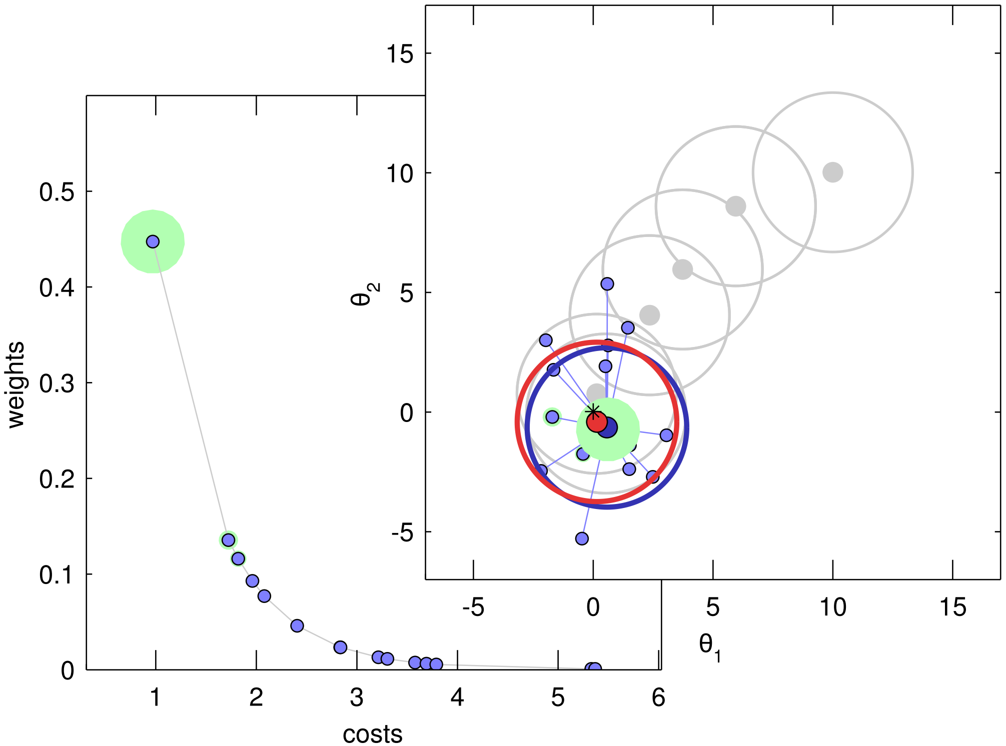
PI^2_{CMA-ES}: Reward-Weighted Averaging

$$\boldsymbol{\theta}_{k=1\ldots K} \sim \mathcal{N}(\boldsymbol{\theta}, \Sigma)$$

$$\forall k \;\; J_k = J(\boldsymbol{\theta}_k)$$

$$\forall k \;\; P_k = \frac{e^{\left(\frac{-h(J_k - \min(J_k))}{\max(J_k) - \min(J_k)}\right)}}{\sum_{l=1}^{K} e^{\left(\frac{-h(J_l - \min(J_l))}{\max(J_l) - \min(J_l)}\right)}}$$

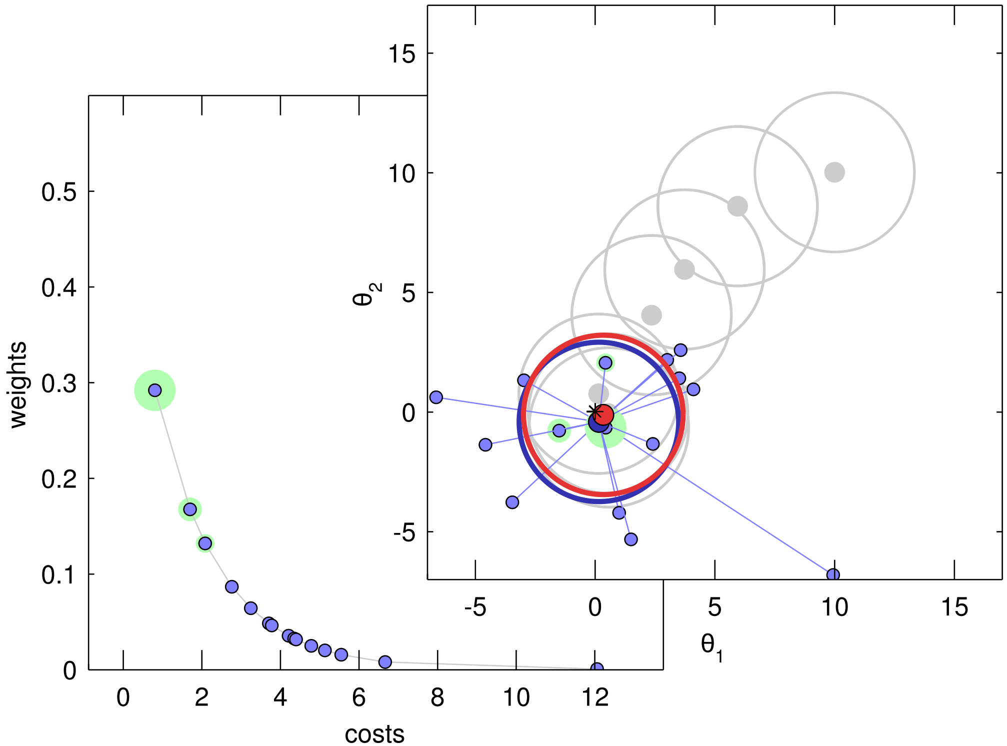
$$\boldsymbol{\theta}^{new} = \sum_{k=1}^{K} P_k \boldsymbol{\theta}_k$$



PI^2_{CMA-ES}: Reward-Weighted Averaging

$$\boldsymbol{\theta}_{k=1\ldots K} \sim \mathcal{N}(\boldsymbol{\theta}, \Sigma)$$
$$\forall k \; J_k = J(\boldsymbol{\theta}_k)$$
$$\forall k \; P_k = \frac{e^{\left(\frac{-h(J_k - \min(J_k))}{\max(J_k) - \min(J_k)}\right)}}{\sum_{l=1}^{K} e^{\left(\frac{-h(J_l - \min(J_l))}{\max(J_l) - \min(J_l)}\right)}}$$

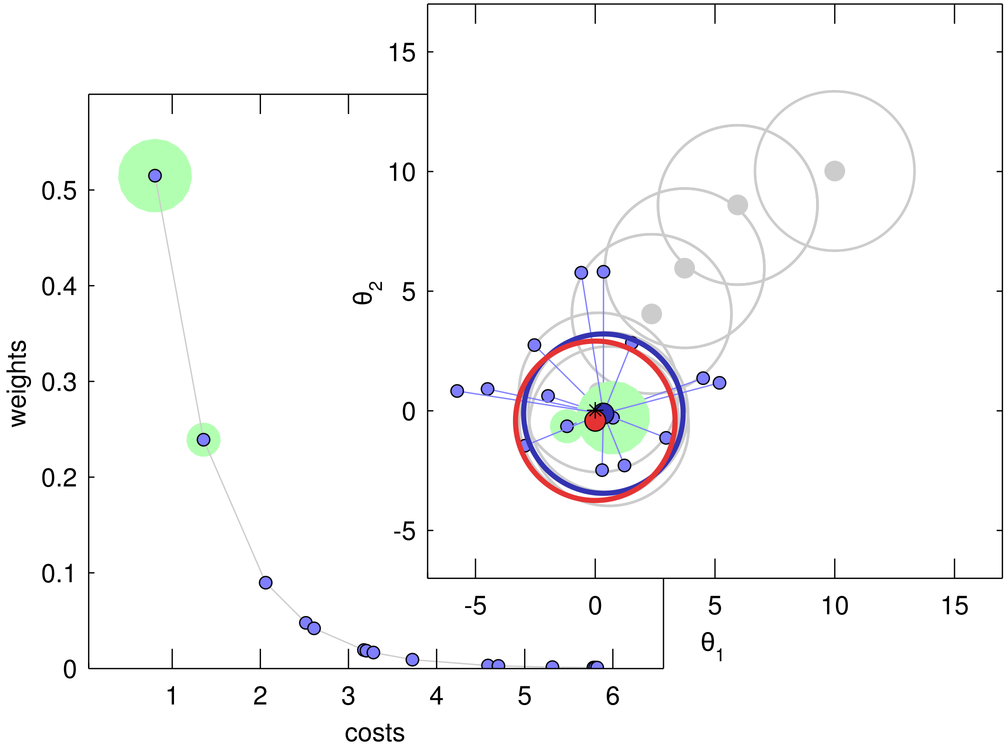
$$\boldsymbol{\theta}^{new} = \sum_{k=1}^{K} P_k \boldsymbol{\theta}_k$$



PI^2_{CMA-ES}: Reward-Weighted Averaging

$$\boldsymbol{\theta}_{k=1\dots K} \sim \mathcal{N}(\boldsymbol{\theta}, \Sigma)$$
$$\forall k \; J_k = J(\boldsymbol{\theta}_k)$$
$$\forall k \; P_k = \frac{e^{\left(\frac{-h(J_k - \min(J_k))}{\max(J_k) - \min(J_k)}\right)}}{\sum_{l=1}^{K} e^{\left(\frac{-h(J_l - \min(J_l))}{\max(J_l) - \min(J_l)}\right)}}$$

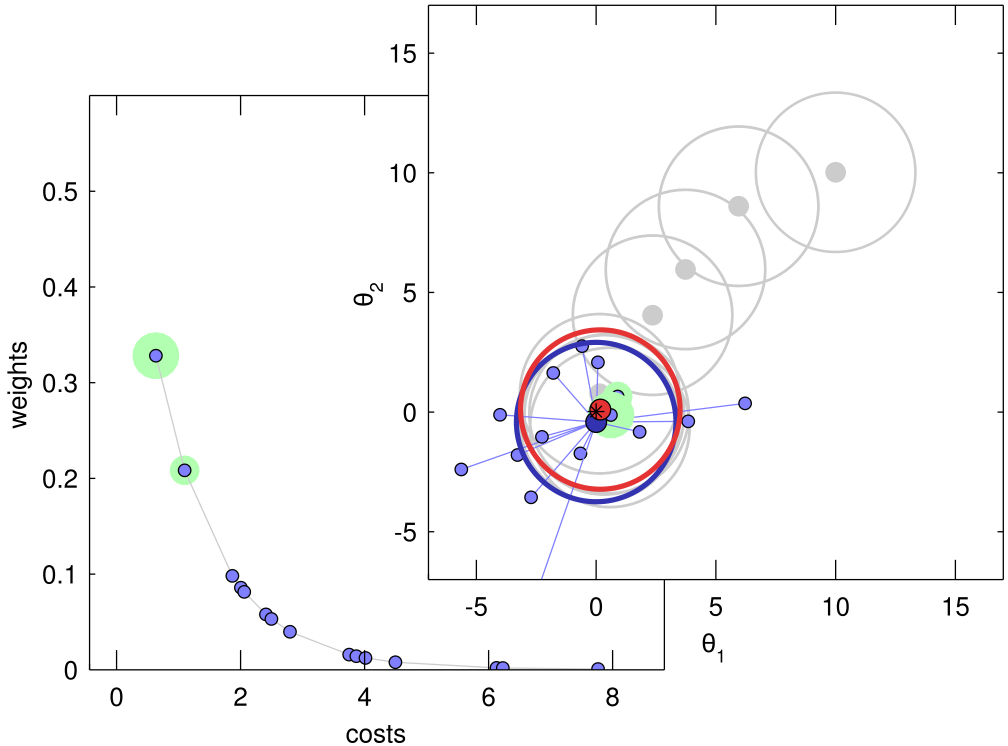
$$\boldsymbol{\theta}^{new} = \sum_{k=1}^{K} P_k \boldsymbol{\theta}_k$$



PI^2_{CMA-ES}: Reward-Weighted Averaging

$$\boldsymbol{\theta}_{k=1\dots K} \sim \mathcal{N}(\boldsymbol{\theta}, \Sigma)$$
$$\forall k \;\; J_k = J(\boldsymbol{\theta}_k)$$
$$\forall k \;\; P_k = \frac{e^{\left(\frac{-h(J_k - \min(J_k))}{\max(J_k)-\min(J_k)}\right)}}{\sum_{l=1}^{K} e^{\left(\frac{-h(J_l - \min(J_l))}{\max(J_l)-\min(J_l)}\right)}}$$

$$\boldsymbol{\theta}^{new} = \sum_{k=1}^{K} P_k \boldsymbol{\theta}_k$$



PI^2_{CMA-ES} Reward-Weighted Averaging and Covariance Matrix Adaptation

$$\boldsymbol{\theta}_{k=1\dots K} \sim \mathcal{N}(\boldsymbol{\theta}, \Sigma)$$

$$\forall k \; J_k = J(\boldsymbol{\theta}_k)$$

$$\forall k \; P_k = \frac{e^{\left(\frac{-h(J_k - \min(J_k))}{\max(J_k) - \min(J_k)}\right)}}{\sum_{l=1}^{K} e^{\left(\frac{-h(J_l - \min(J_l))}{\max(J_l) - \min(J_l)}\right)}}$$

$$\boldsymbol{\theta}^{new} = \sum_{k=1}^{K} P_k \boldsymbol{\theta}_k$$

$$\Sigma^{new} = \sum_{k=1}^{K} P_k (\boldsymbol{\theta}_k - \boldsymbol{\theta})(\boldsymbol{\theta}_k - \boldsymbol{\theta})^\intercal$$

PI^2_{CMA-ES} Reward-Weighted Averaging and Covariance Matrix Adaptation

$$\boldsymbol{\theta}_{k=1\dots K} \sim \mathcal{N}(\boldsymbol{\theta}, \Sigma)$$

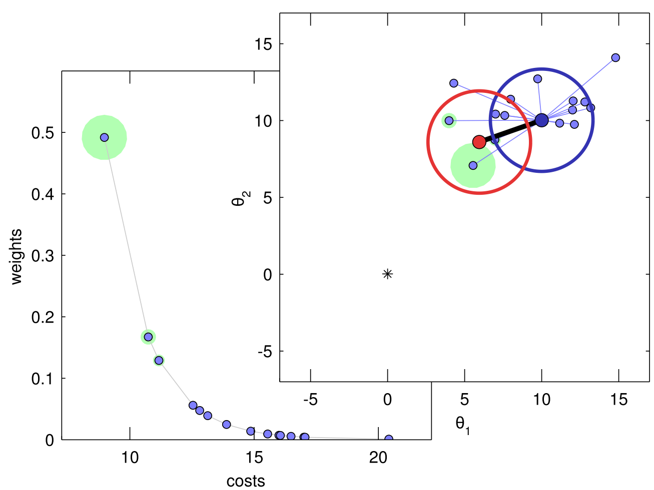
$$\forall k \;\; J_k = J(\boldsymbol{\theta}_k)$$

$$\forall k \;\; P_k = \frac{e^{\left(\frac{-h(J_k - \min(J_k))}{\max(J_k) - \min(J_k)}\right)}}{\sum_{l=1}^{K} e^{\left(\frac{-h(J_l - \min(J_l))}{\max(J_l) - \min(J_l)}\right)}}$$

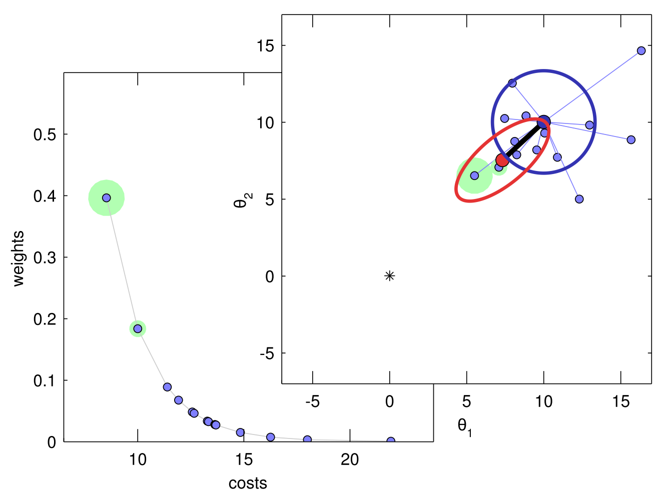
$$\boldsymbol{\theta}^{new} = \sum_{k=1}^{K} P_k \boldsymbol{\theta}_k$$

$$\Sigma^{new} = \sum_{k=1}^{K} P_k (\boldsymbol{\theta}_k - \boldsymbol{\theta})(\boldsymbol{\theta}_k - \boldsymbol{\theta})^{\mathsf{T}}$$

Without CMA



With CMA



PI^2_{CMA-ES} Reward-Weighted Averaging and Covariance Matrix Adaptation

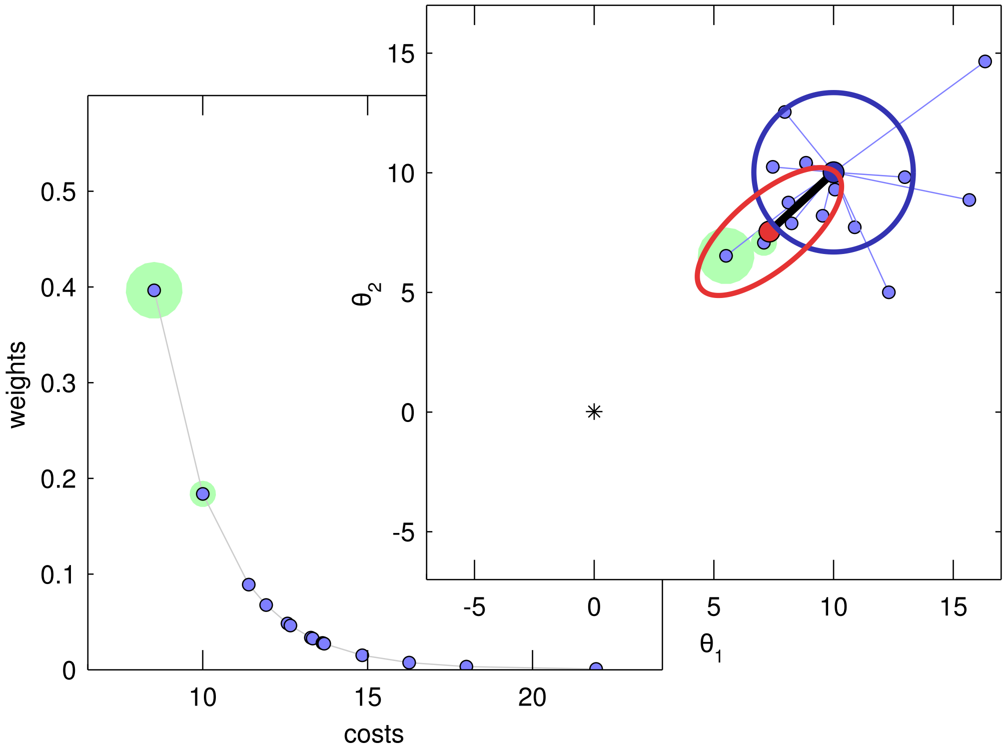
$$\boldsymbol{\theta}_{k=1\ldots K} \sim \mathcal{N}(\boldsymbol{\theta}, \Sigma)$$

$$\forall k \;\; J_k = J(\boldsymbol{\theta}_k)$$

$$\forall k \;\; P_k = \frac{e^{\left(\frac{-h(J_k - \min(J_k))}{\max(J_k) - \min(J_k)}\right)}}{\sum_{l=1}^{K} e^{\left(\frac{-h(J_l - \min(J_l))}{\max(J_l) - \min(J_l)}\right)}}$$

$$\boldsymbol{\theta}^{new} = \sum_{k=1}^{K} P_k \boldsymbol{\theta}_k$$

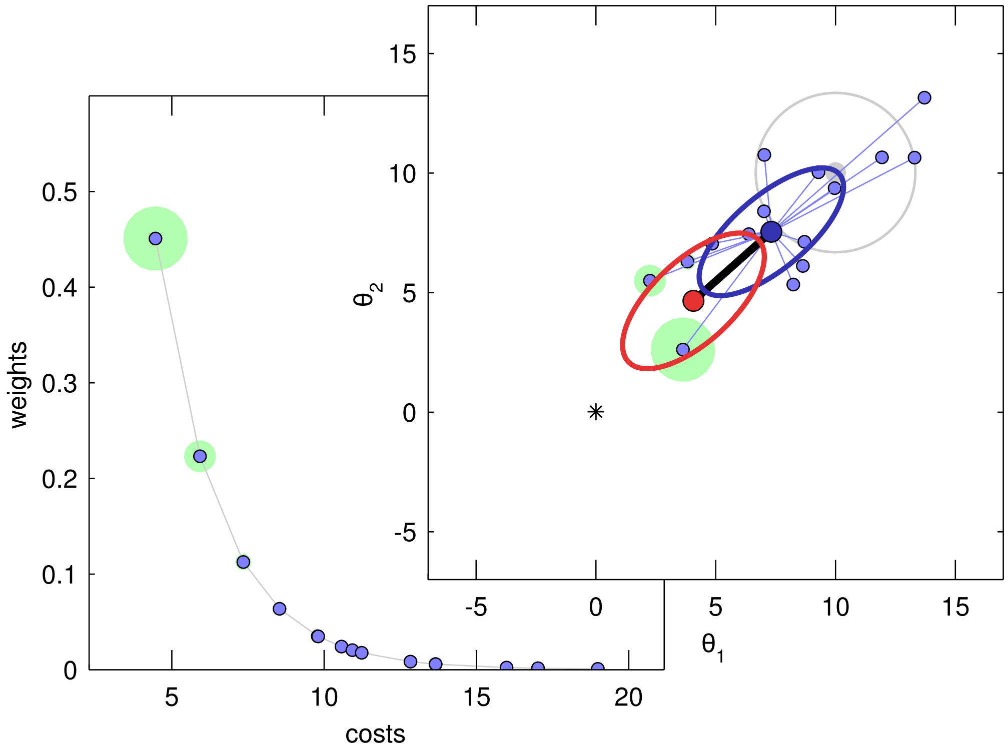
$$\Sigma^{new} = \sum_{k=1}^{K} P_k (\boldsymbol{\theta}_k - \boldsymbol{\theta})(\boldsymbol{\theta}_k - \boldsymbol{\theta})^\intercal$$



PI^2_{CMA-ES} Reward-Weighted Averaging and Covariance Matrix Adaptation

$$\boldsymbol{\theta}_{k=1\dots K} \sim \mathcal{N}(\boldsymbol{\theta}, \Sigma)$$
$$\forall k \;\; J_k = J(\boldsymbol{\theta}_k)$$
$$\forall k \;\; P_k = \frac{e^{\left(\frac{-h(J_k - \min(J_k))}{\max(J_k)-\min(J_k)}\right)}}{\sum_{l=1}^{K} e^{\left(\frac{-h(J_l - \min(J_l))}{\max(J_l)-\min(J_l)}\right)}}$$

$$\boldsymbol{\theta}^{new} = \sum_{k=1}^{K} P_k \boldsymbol{\theta}_k$$
$$\Sigma^{new} = \sum_{k=1}^{K} P_k (\boldsymbol{\theta}_k - \boldsymbol{\theta})(\boldsymbol{\theta}_k - \boldsymbol{\theta})^{\intercal}$$



PI^2_{CMA-ES} Reward-Weighted Averaging and Covariance Matrix Adaptation

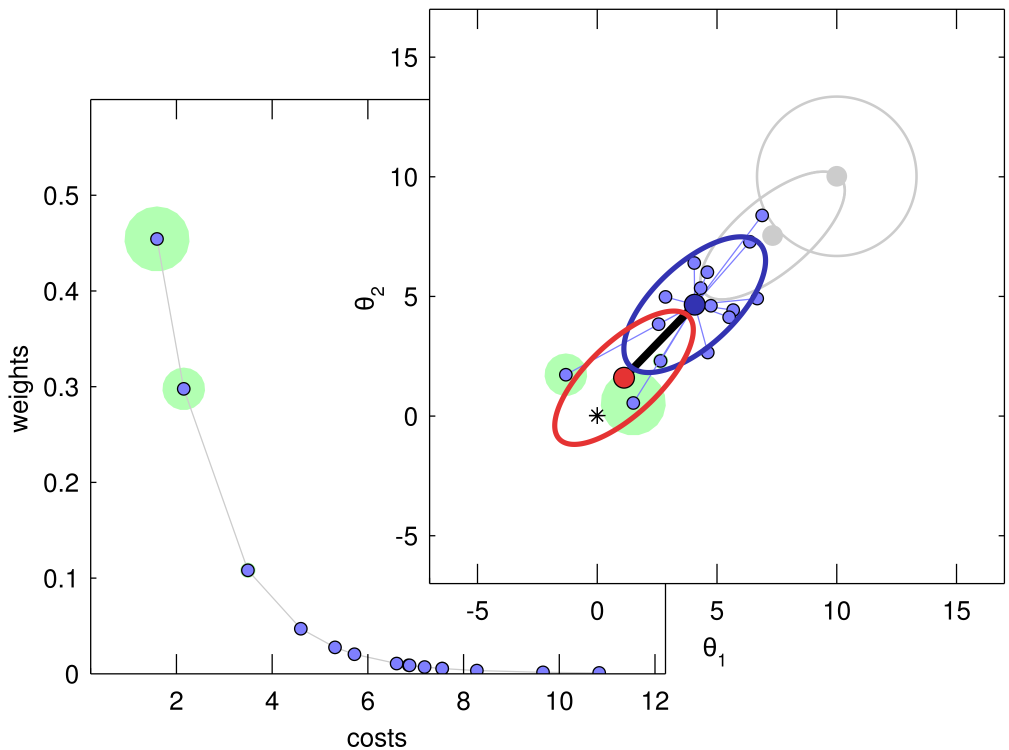
$$\boldsymbol{\theta}_{k=1\dots K} \sim \mathcal{N}(\boldsymbol{\theta}, \Sigma)$$

$$\forall k \; J_k = J(\boldsymbol{\theta}_k)$$

$$\forall k \; P_k = \frac{e^{\left(\frac{-h(J_k - \min(J_k))}{\max(J_k) - \min(J_k)}\right)}}{\sum_{l=1}^{K} e^{\left(\frac{-h(J_l - \min(J_l))}{\max(J_l) - \min(J_l)}\right)}}$$

$$\boldsymbol{\theta}^{new} = \sum_{k=1}^{K} P_k \boldsymbol{\theta}_k$$

$$\Sigma^{new} = \sum_{k=1}^{K} P_k (\boldsymbol{\theta}_k - \boldsymbol{\theta})(\boldsymbol{\theta}_k - \boldsymbol{\theta})^{\intercal}$$



PI^2_{CMA-ES} Reward-Weighted Averaging and Covariance Matrix Adaptation

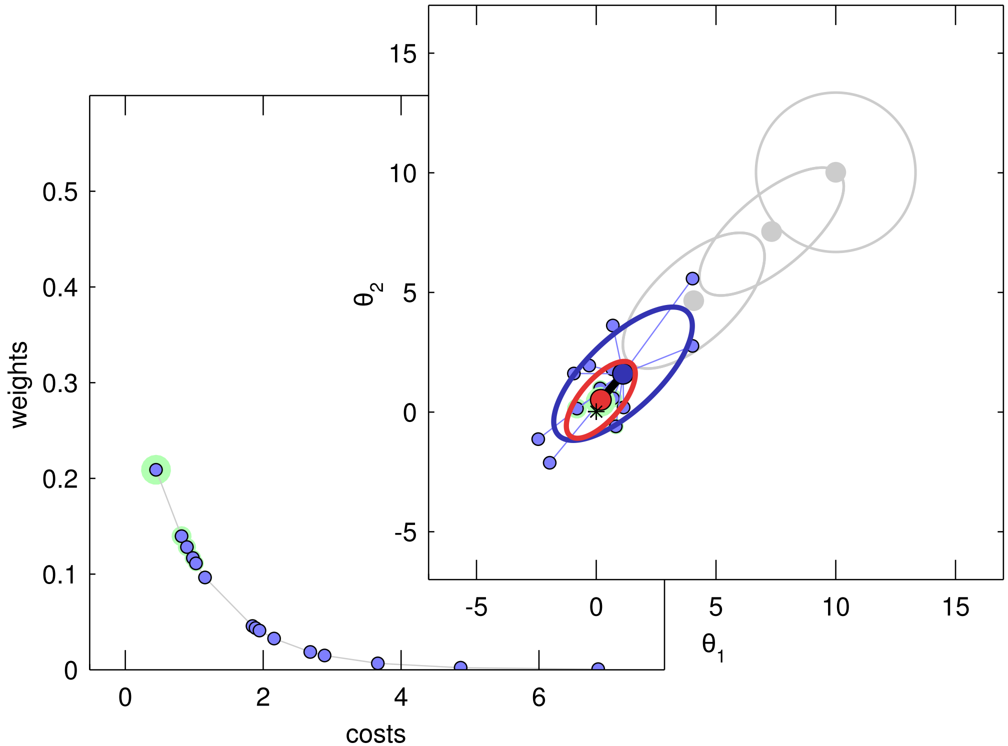
$$\boldsymbol{\theta}_{k=1\ldots K} \sim \mathcal{N}(\boldsymbol{\theta}, \Sigma)$$

$$\forall k \; J_k = J(\boldsymbol{\theta}_k)$$

$$\forall k \; P_k = \frac{e^{\left(\frac{-h(J_k - \min(J_k))}{\max(J_k) - \min(J_k)}\right)}}{\sum_{l=1}^{K} e^{\left(\frac{-h(J_l - \min(J_l))}{\max(J_l) - \min(J_l)}\right)}}$$

$$\boldsymbol{\theta}^{new} = \sum_{k=1}^{K} P_k \boldsymbol{\theta}_k$$

$$\Sigma^{new} = \sum_{k=1}^{K} P_k (\boldsymbol{\theta}_k - \boldsymbol{\theta})(\boldsymbol{\theta}_k - \boldsymbol{\theta})^{\intercal}$$



PI^2_{CMA-ES} Reward-Weighted Averaging and Covariance Matrix Adaptation

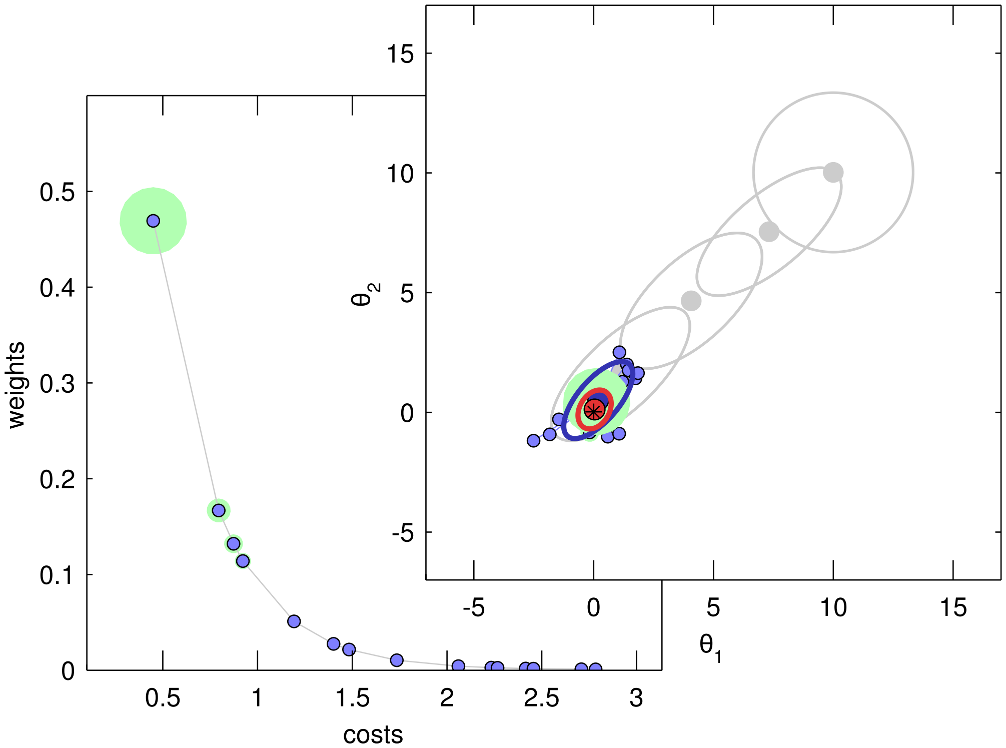
$$\boldsymbol{\theta}_{k=1\dots K} \sim \mathcal{N}(\boldsymbol{\theta}, \Sigma)$$

$$\forall k \; J_k = J(\boldsymbol{\theta}_k)$$

$$\forall k \; P_k = \frac{e^{\left(\frac{-h(J_k - \min(J_k))}{\max(J_k) - \min(J_k)}\right)}}{\sum_{l=1}^{K} e^{\left(\frac{-h(J_l - \min(J_l))}{\max(J_l) - \min(J_l)}\right)}}$$

$$\boldsymbol{\theta}^{new} = \sum_{k=1}^{K} P_k \boldsymbol{\theta}_k$$

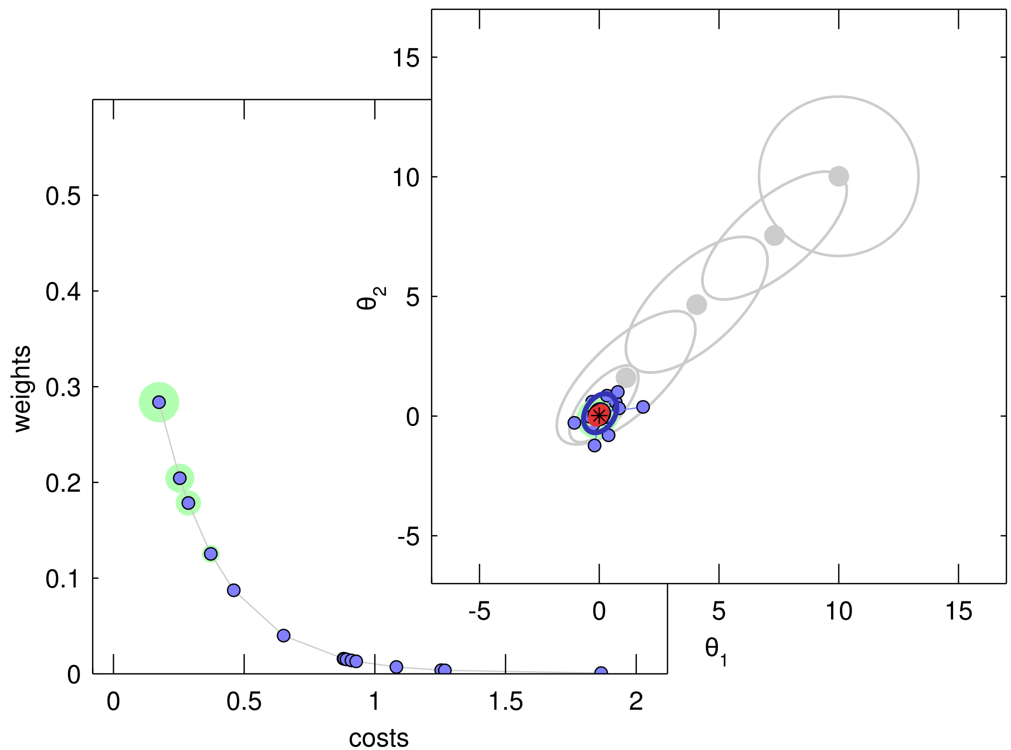
$$\Sigma^{new} = \sum_{k=1}^{K} P_k (\boldsymbol{\theta}_k - \boldsymbol{\theta})(\boldsymbol{\theta}_k - \boldsymbol{\theta})^\intercal$$



PI^2_{CMA-ES} Reward-Weighted Averaging and Covariance Matrix Adaptation

$$\boldsymbol{\theta}_{k=1\ldots K} \sim \mathcal{N}(\boldsymbol{\theta}, \Sigma)$$
$$\forall k \;\; J_k = J(\boldsymbol{\theta}_k)$$
$$\forall k \;\; P_k = \frac{e^{\left(\frac{-h(J_k - \min(J_k))}{\max(J_k) - \min(J_k)}\right)}}{\sum_{l=1}^{K} e^{\left(\frac{-h(J_l - \min(J_l))}{\max(J_l) - \min(J_l)}\right)}}$$

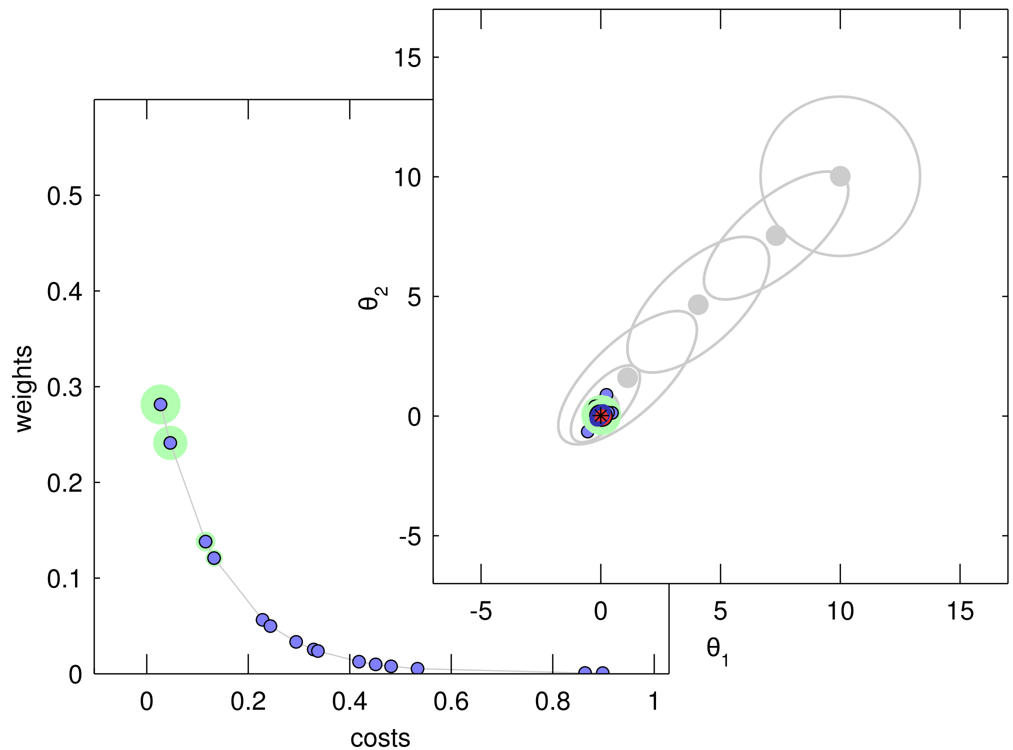
$$\boldsymbol{\theta}^{new} = \sum_{k=1}^{K} P_k \boldsymbol{\theta}_k$$
$$\Sigma^{new} = \sum_{k=1}^{K} P_k (\boldsymbol{\theta}_k - \boldsymbol{\theta})(\boldsymbol{\theta}_k - \boldsymbol{\theta})^{\mathsf{T}}$$



PI^2_{CMA-ES} Reward-Weighted Averaging and Covariance Matrix Adaptation

$$\boldsymbol{\theta}_{k=1\dots K} \sim \mathcal{N}(\boldsymbol{\theta}, \Sigma)$$
$$\forall k \;\; J_k = J(\boldsymbol{\theta}_k)$$
$$\forall k \;\; P_k = \frac{e^{\left(\frac{-h(J_k - \min(J_k))}{\max(J_k) - \min(J_k)}\right)}}{\sum_{l=1}^{K} e^{\left(\frac{-h(J_l - \min(J_l))}{\max(J_l) - \min(J_l)}\right)}}$$

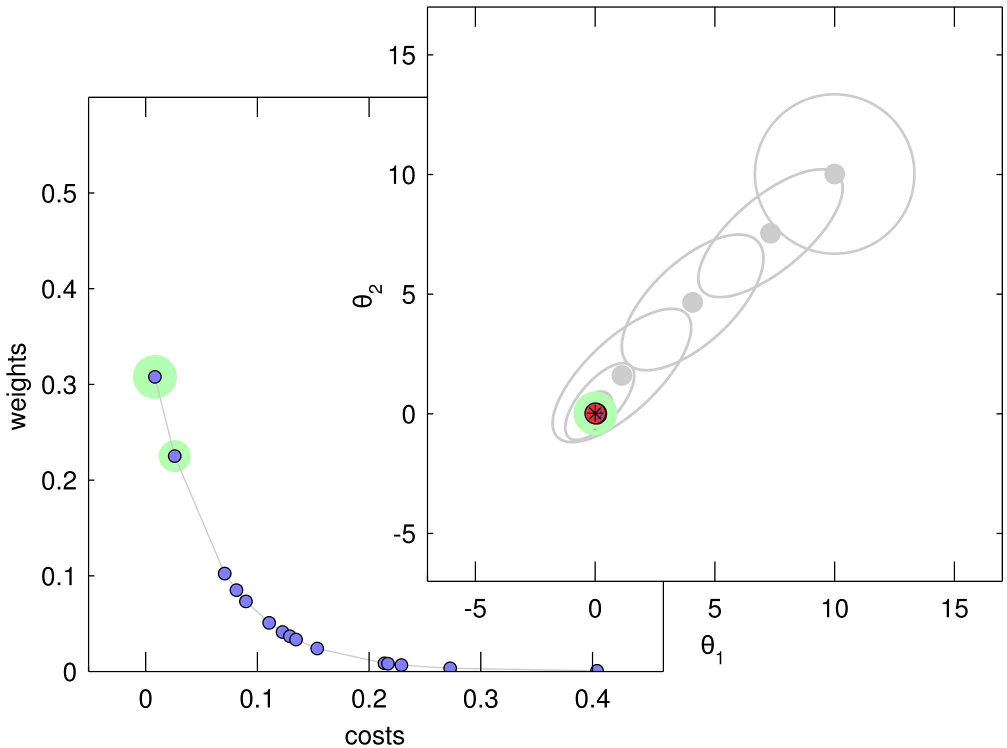
$$\boldsymbol{\theta}^{new} = \sum_{k=1}^{K} P_k \boldsymbol{\theta}_k$$
$$\Sigma^{new} = \sum_{k=1}^{K} P_k (\boldsymbol{\theta}_k - \boldsymbol{\theta})(\boldsymbol{\theta}_k - \boldsymbol{\theta})^{\intercal}$$



PI^2_{CMA-ES} Reward-Weighted Averaging and Covariance Matrix Adaptation

$$\boldsymbol{\theta}_{k=1\ldots K} \sim \mathcal{N}(\boldsymbol{\theta}, \Sigma)$$
$$\forall k \; J_k = J(\boldsymbol{\theta}_k)$$
$$\forall k \; P_k = \frac{e^{\left(\frac{-h(J_k - \min(J_k))}{\max(J_k) - \min(J_k)}\right)}}{\sum_{l=1}^{K} e^{\left(\frac{-h(J_l - \min(J_l))}{\max(J_l) - \min(J_l)}\right)}}$$

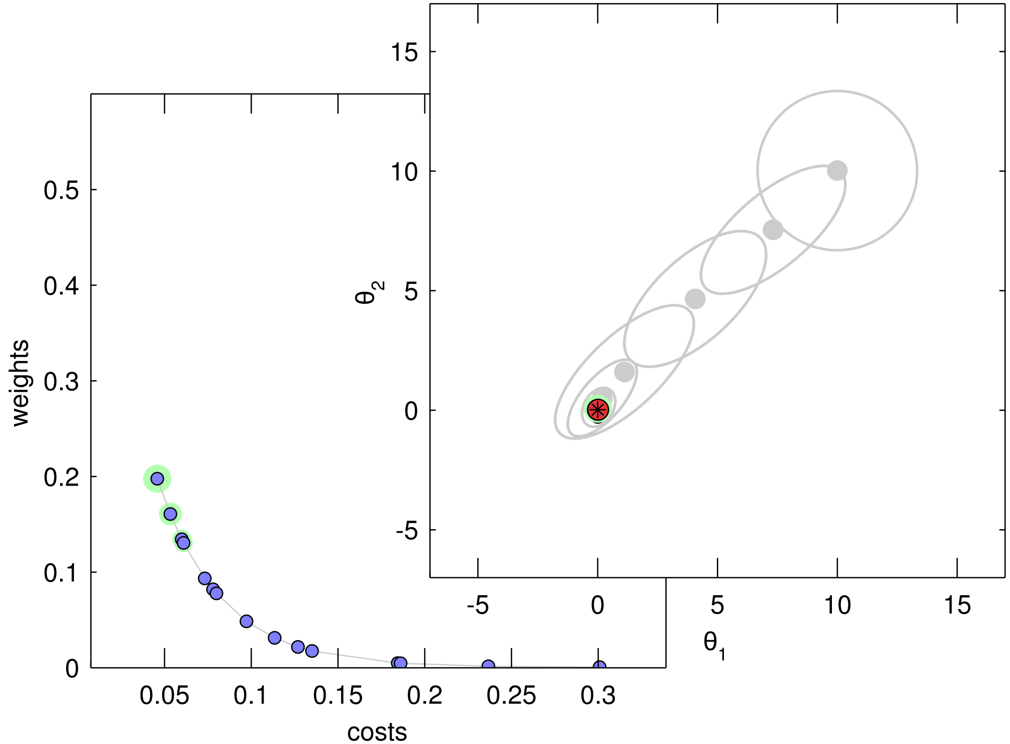
$$\boldsymbol{\theta}^{new} = \sum_{k=1}^{K} P_k \boldsymbol{\theta}_k$$
$$\Sigma^{new} = \sum_{k=1}^{K} P_k (\boldsymbol{\theta}_k - \boldsymbol{\theta})(\boldsymbol{\theta}_k - \boldsymbol{\theta})^\intercal$$



PI^2_{CMA-ES} Reward-Weighted Averaging and Covariance Matrix Adaptation

$$\boldsymbol{\theta}_{k=1\ldots K} \sim \mathcal{N}(\boldsymbol{\theta}, \Sigma)$$
$$\forall k \;\; J_k = J(\boldsymbol{\theta}_k)$$
$$\forall k \;\; P_k = \frac{e^{\left(\frac{-h(J_k - \min(J_k))}{\max(J_k) - \min(J_k)}\right)}}{\sum_{l=1}^{K} e^{\left(\frac{-h(J_l - \min(J_l))}{\max(J_l) - \min(J_l)}\right)}}$$

$$\boldsymbol{\theta}^{new} = \sum_{k=1}^{K} P_k \boldsymbol{\theta}_k$$
$$\Sigma^{new} = \sum_{k=1}^{K} P_k (\boldsymbol{\theta}_k - \boldsymbol{\theta})(\boldsymbol{\theta}_k - \boldsymbol{\theta})^\intercal$$



PI^2_{CMA-ES} Reward-Weighted Averaging and Covariance Matrix Adaptation

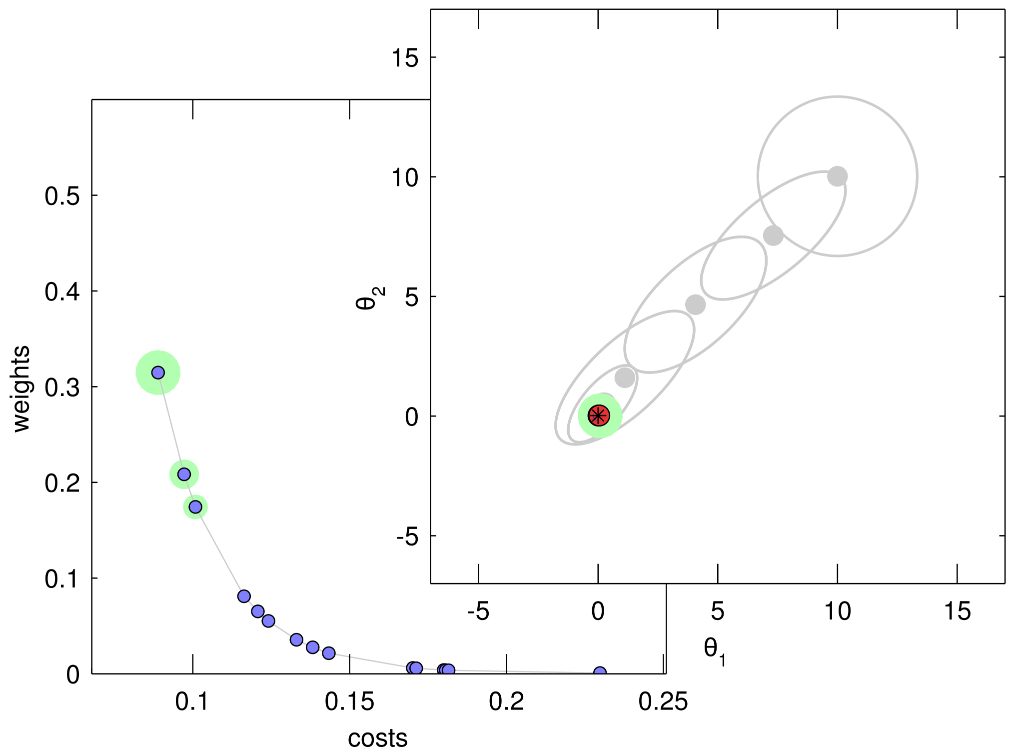
$$\boldsymbol{\theta}_{k=1\dots K} \sim \mathcal{N}(\boldsymbol{\theta}, \Sigma)$$

$$\forall k \; J_k = J(\boldsymbol{\theta}_k)$$

$$\forall k \; P_k = \frac{e^{\left(\frac{-h(J_k - \min(J_k))}{\max(J_k) - \min(J_k)}\right)}}{\sum_{l=1}^{K} e^{\left(\frac{-h(J_l - \min(J_l))}{\max(J_l) - \min(J_l)}\right)}}$$

$$\boldsymbol{\theta}^{new} = \sum_{k=1}^{K} P_k \boldsymbol{\theta}_k$$

$$\Sigma^{new} = \sum_{k=1}^{K} P_k (\boldsymbol{\theta}_k - \boldsymbol{\theta})(\boldsymbol{\theta}_k - \boldsymbol{\theta})^{\mathsf{T}}$$

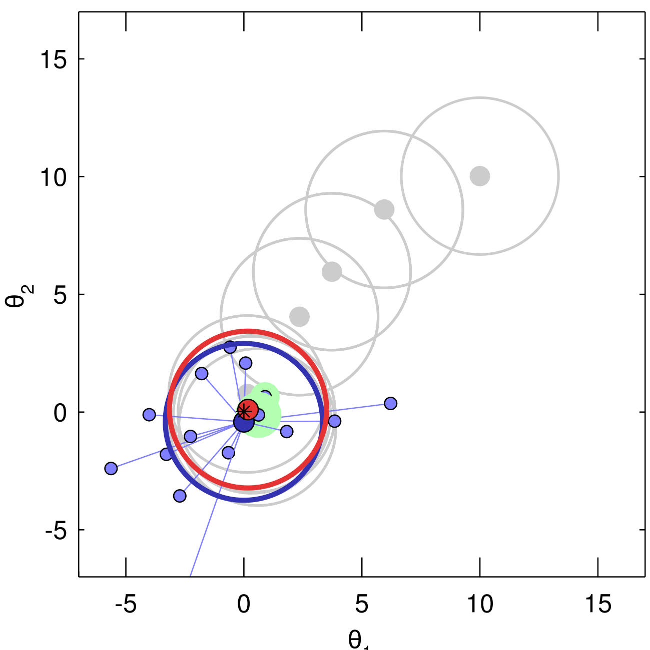


PI^2_{CMA-ES} Reward-Weighted Averaging and Covariance Matrix Adaptation

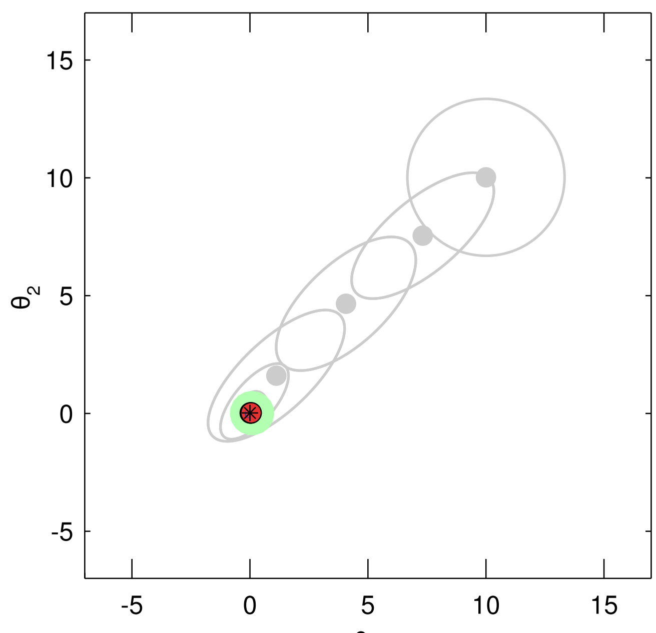
$$\boldsymbol{\theta}_{k=1\dots K} \sim \mathcal{N}(\boldsymbol{\theta}, \Sigma)$$
$$\forall k \;\; J_k = J(\boldsymbol{\theta}_k)$$
$$\forall k \;\; P_k = \frac{e^{\left(\frac{-h(J_k - \min(J_k))}{\max(J_k) - \min(J_k)}\right)}}{\sum_{l=1}^{K} e^{\left(\frac{-h(J_l - \min(J_l))}{\max(J_l) - \min(J_l)}\right)}}$$

$$\boldsymbol{\theta}^{new} = \sum_{k=1}^{K} P_k \boldsymbol{\theta}_k$$
$$\Sigma^{new} = \sum_{k=1}^{K} P_k (\boldsymbol{\theta}_k - \boldsymbol{\theta})(\boldsymbol{\theta}_k - \boldsymbol{\theta})^{\intercal}$$

Without CMA



With CMA



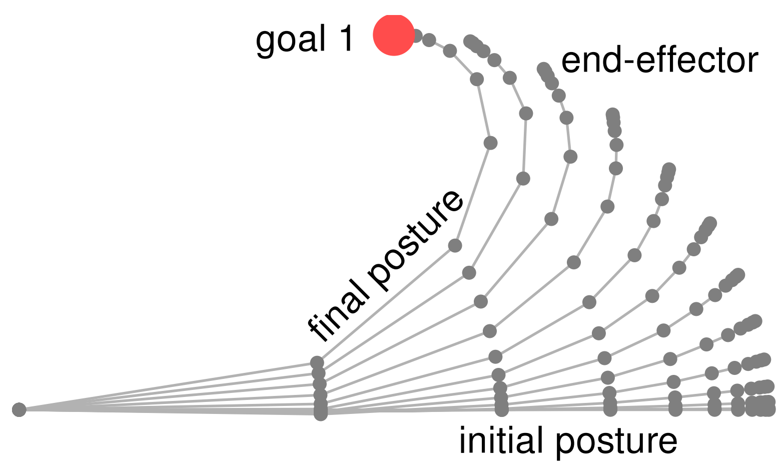
Application to reaching

- Task: reach to a goal in the workspace
- 10-DOF kinematically simulated 'arm' in 2-D plane
- Policy representation:

$$\ddot{q}_{m,t} = \mathbf{g}(t)^\intercal \boldsymbol{\theta}_m \qquad \text{Acc. of joint } m \qquad (1)$$

$$[\mathbf{g}(t)]_b = \frac{\Psi_b(t)}{\sum_{b=1}^{B} \Psi_b(t)} \text{ with } \Psi_b(t) = \exp\left(-(t-c_b)^2/w^2\right) \qquad \text{Basis functions} \qquad (2)$$

- Duration of movement is 0.5s
- Initialy, $\boldsymbol{\theta} = 0$ (no movement)



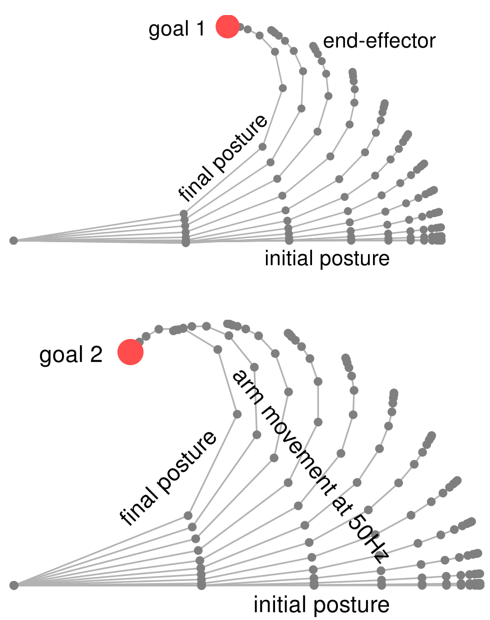
Cost function:

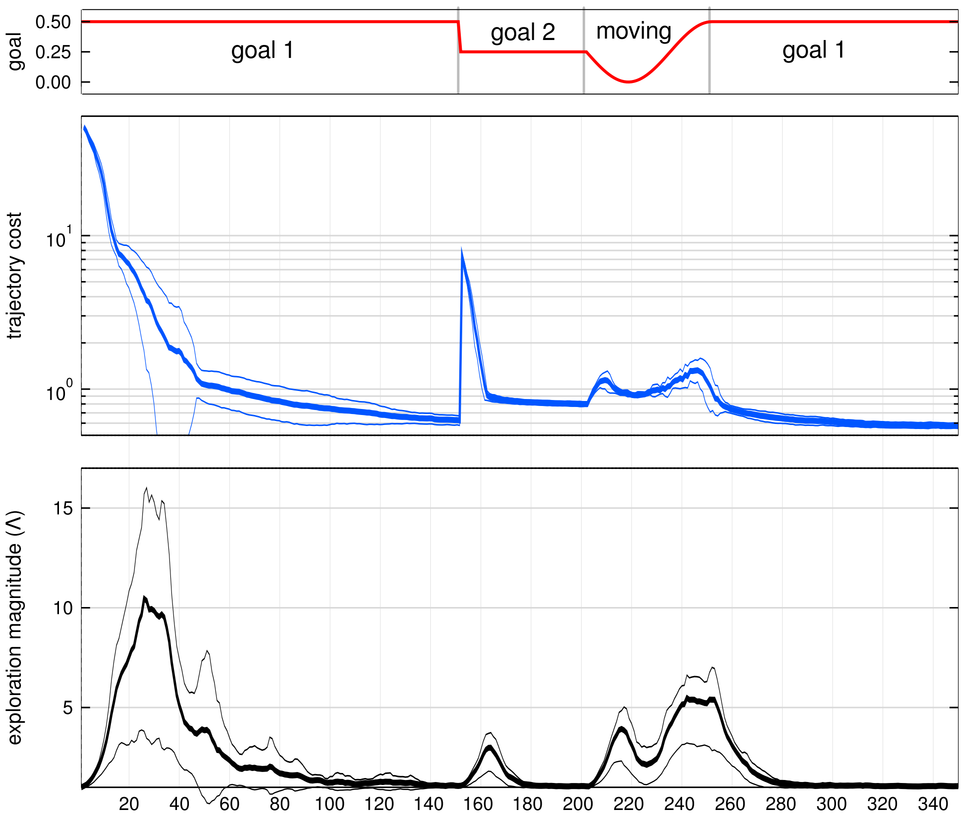
$$\phi_{t_N} = 10^4 ||\mathbf{x}_{t_N} - g||^2 + \max(\mathbf{q}_{t_N}) \qquad \text{Terminal cost} \qquad (1)$$

$$r_t = 10^{-5} \frac{\sum_{m=1}^{M} (M+1-m)(\ddot{q}_{t,m})^2}{\sum_{m=1}^{M} (M+1-m)} \qquad \text{Immediate cost} \qquad (2)$$

- $\text{PI}^2_{\text{CMAES}}$ parameters: $K = 20$

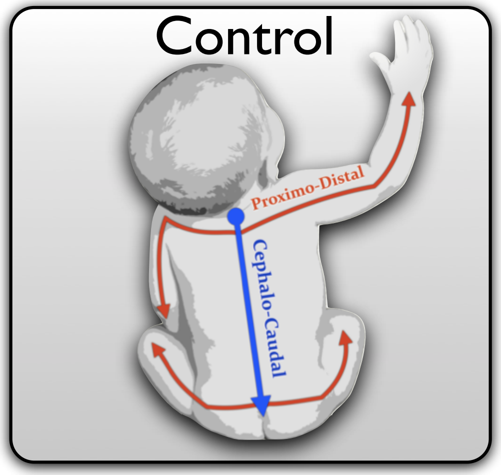
Results: Re-Adadaptation to Changing Tasks

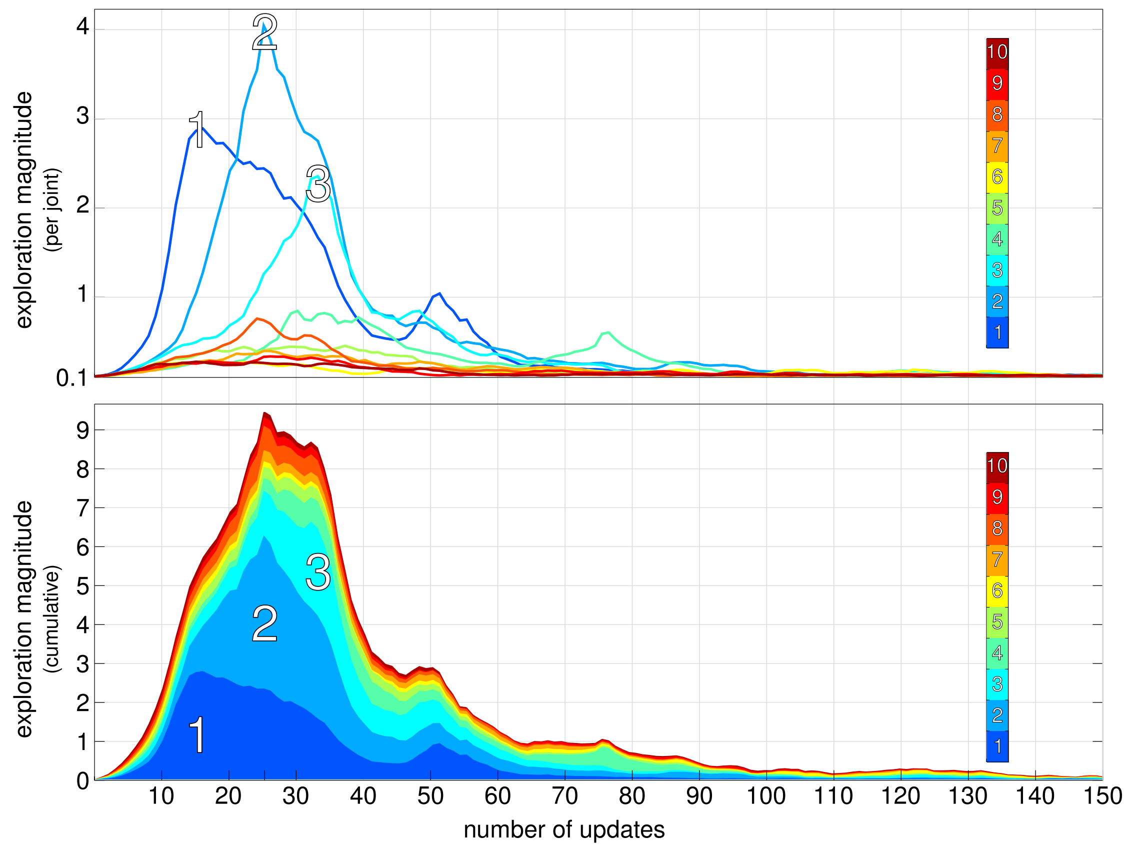




⇒ Life-long continual reinforcement learning with automatic exploration/exploitation trade-off

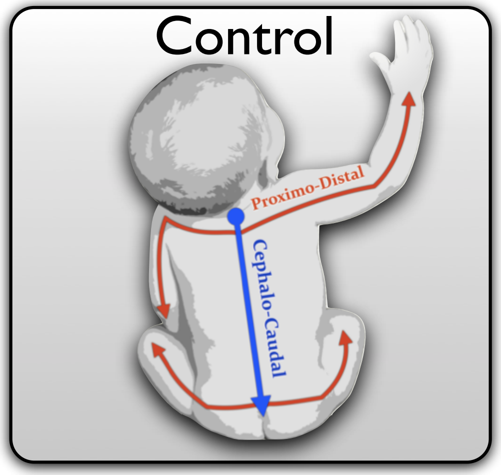
Results: Emergent Proximo-Distal Maturation

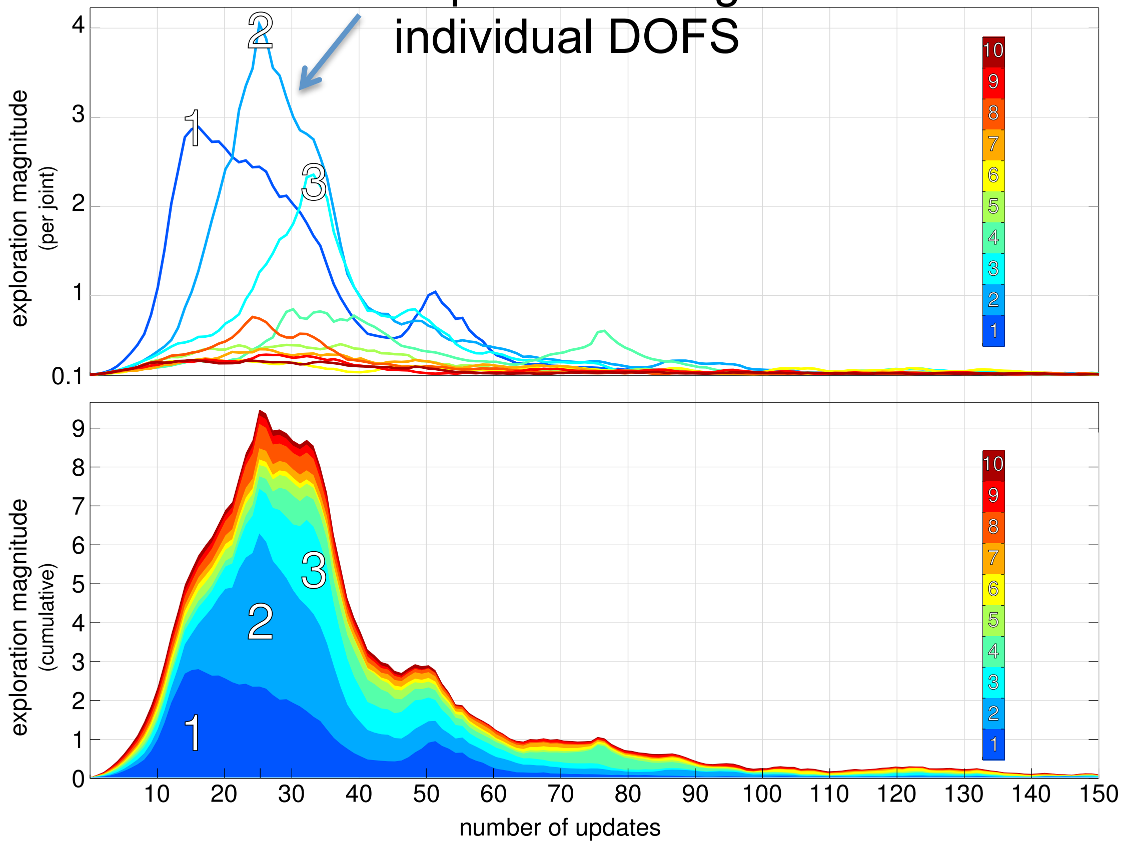




⇒ Emergent Proximo-Distal Maturation

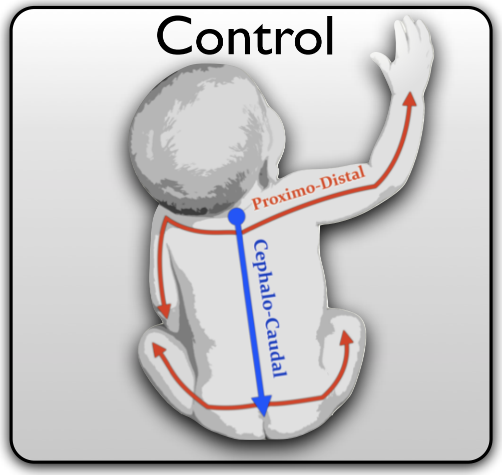
Results: Emergent Proximo-Distal Maturation

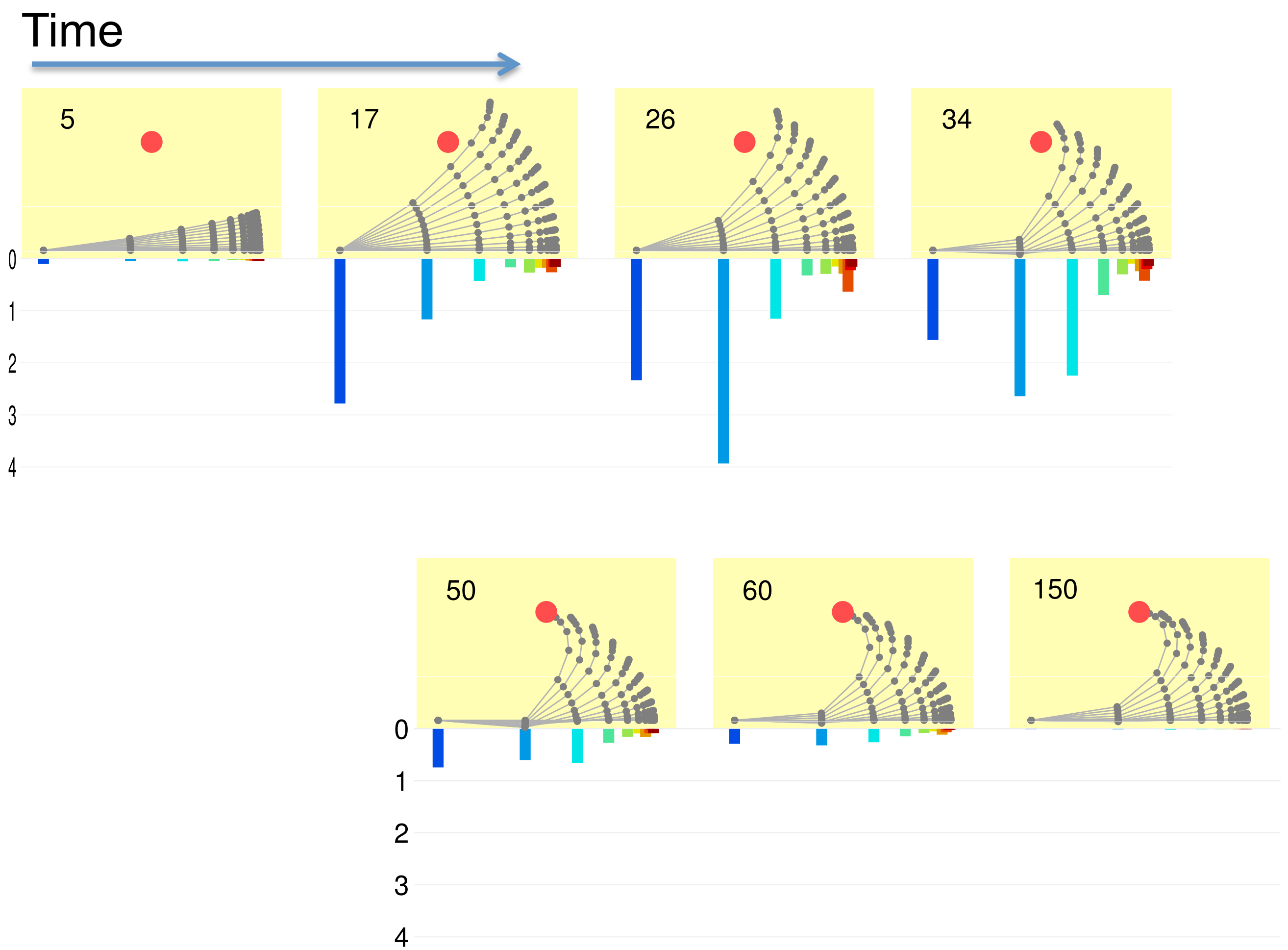




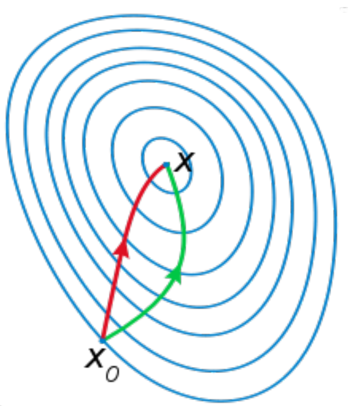
⇒ Emergent Proximo-Distal Maturation

Results: Emergent Proximo-Distal Maturation

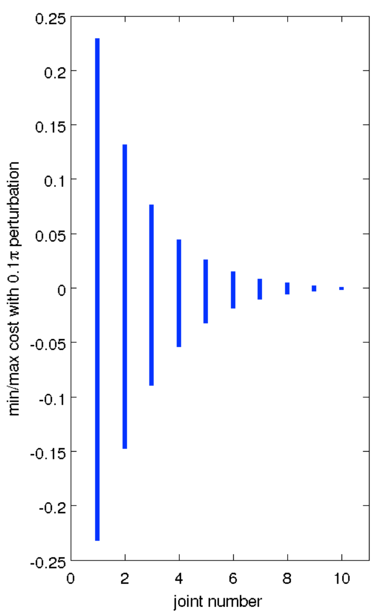




Interpretation



- ***Stochastic optimization*** directs itself by following an ***approximated smoothed gradient/curvature***
  i.e.
  by fostering exploration in directions where impact on cost function is big, and fostering ignorance of directions where impact is less



- Arm structure is such that
1) initially ***proximal joints have more impact on cost function than distal ones***
2) This ***relative impact changes as one gets closer to the maximum of the cost function***

➔ Emergent maturation is a property of the combination between the structure of cost function (dep. on body structure) and adaptive exploration in stochastic optimization

References

Baranes, A., Oudeyer, P-Y. (2011) The Interaction of Maturational Constraints and Intrinsic Motivations in Active Motor Development, in proceedings of the IEEE International Conference on Development and Learning (ICDL-Epirob) , Frankfurt, Germany.

Berthouze, L. and Lungarella,M. (2004) Motor skill acquisition under environmental perturbations: On the necessity of alternate freezing and freeing degrees of freedom. Adaptive Behavior, 12(1):47–63, 2004.

Hansen, N. and Ostermeier, A. (2001) Completely derandomized self-adaptation in evolution strategies. Evolutionary Computation, 9(2):159–195.

Ivanchenko, V. and Jacobs, R.A. (2003) A developmental approach aids motor learning, Neural Computation, 15, 2051-2065.

Stulp, F. and Oudeyer, P.-Y. (2012) Emergent proximo-distal maturation through adaptive exploration. In International Conference on Development and Learning (ICDL). Accepted for publication.

Stulp, F. and Sigaud, O. (2012a) Path integral policy improvement with covariance matrix adaptation. In Proceedings of the 29th International Conference on Machine Learning (ICML).

Theodorou, E., Buchli, J., and Schaal, S. (2010) A generalized path integral control approach to reinforcement learning. J. of Machine Learning Research, 11:3137–3181, 2010.