Evaluation of the impact of various modifications to CMA-ES that facilitate its theoretical analysis

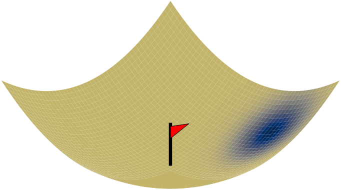
BBOB Workshop - GECCO '23 in Lisbon, Portugal

Armand Gissler

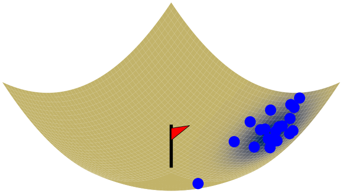
Sunday 16th July, 2023

CMAP, École polytechnique & Inria

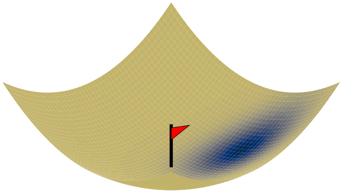
CMA-ES approximates the minimum by a normal distribution



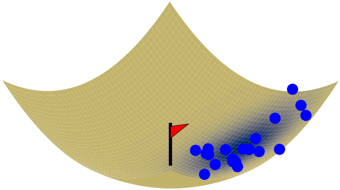
CMA-ES approximates the minimum by a normal distribution



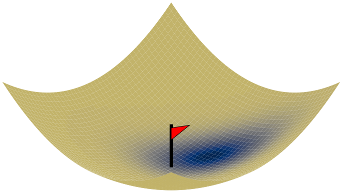
CMA-ES approximates the minimum by a normal distribution



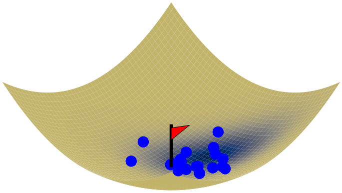
CMA-ES approximates the minimum by a normal distribution



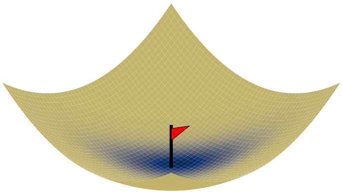
CMA-ES approximates the minimum by a normal distribution



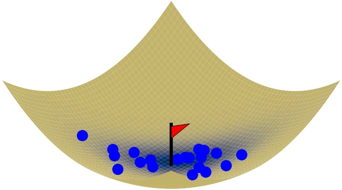
CMA-ES approximates the minimum by a normal distribution



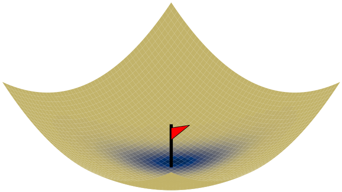
CMA-ES approximates the minimum by a normal distribution



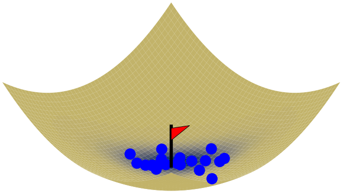
CMA-ES approximates the minimum by a normal distribution



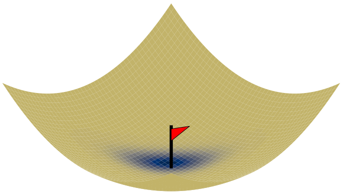
CMA-ES approximates the minimum by a normal distribution



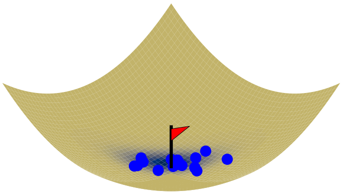
CMA-ES approximates the minimum by a normal distribution



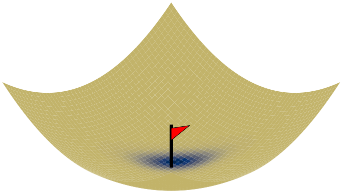
CMA-ES approximates the minimum by a normal distribution



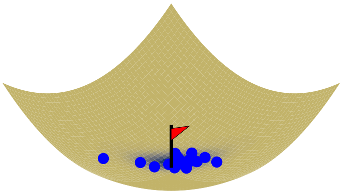
CMA-ES approximates the minimum by a normal distribution



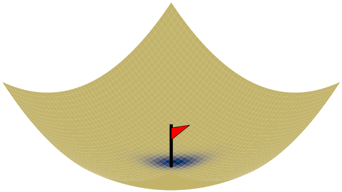
CMA-ES approximates the minimum by a normal distribution



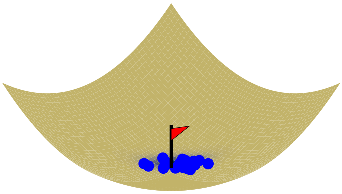
CMA-ES approximates the minimum by a normal distribution



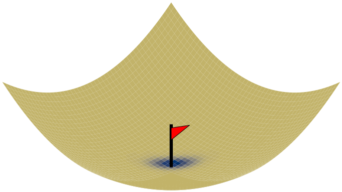
CMA-ES approximates the minimum by a normal distribution



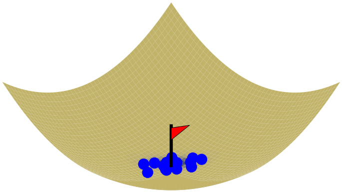
CMA-ES approximates the minimum by a normal distribution



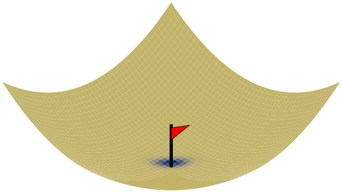
CMA-ES approximates the minimum by a normal distribution



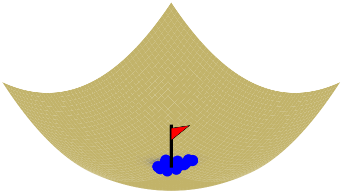
CMA-ES approximates the minimum by a normal distribution



CMA-ES approximates the minimum by a normal distribution

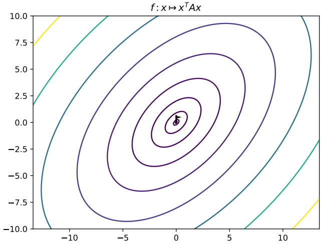


CMA-ES approximates the minimum by a normal distribution



CMA-ES

Algorithm 1 default-CMA-ES

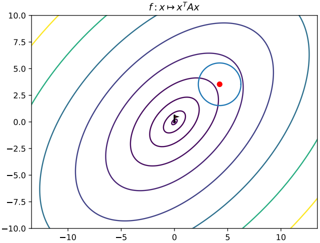


CMA-ES

Algorithm 1 default-CMA-ES

Given: $m_0 \in \mathbb{R}^d,\ \sigma_0 > 0,\ C_0 \in \mathcal{S}^d_{++},\ p^\sigma_0 \in \mathbb{R}^d,\ p^c_0 \in \mathbb{R}^d$

For $t = 0, 1, 2, \ldots$:



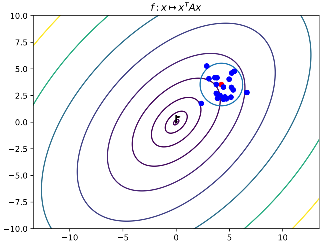
CMA-ES

Algorithm 1 default-CMA-ES

Given: $m_0 \in \mathbb{R}^d,\ \sigma_0 > 0,\ C_0 \in \mathcal{S}^d_{++},\ p_0^\sigma \in \mathbb{R}^d,\ p_0^c \in \mathbb{R}^d$

For $t = 0, 1, 2, \ldots :$

- $U^1_{t+1}, \ldots, U^\lambda_{t+1} \sim \mathcal{N}(0, I_d)$



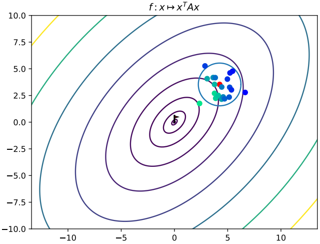
CMA-ES

Algorithm 1 default-CMA-ES

Given: $m_0 \in \mathbb{R}^d$, $\sigma_0 > 0$, $C_0 \in \mathcal{S}^d_{++}$, $p_0^\sigma \in \mathbb{R}^d$, $p_0^c \in \mathbb{R}^d$

For $t = 0, 1, 2, \ldots :$

- $U^1_{t+1}, \ldots, U^\lambda_{t+1} \sim \mathcal{N}(0, I_d)$
- $\mathrm{sort}\ f(m_t + \sigma_t C_t^{1/2} U^i_{t+1})$



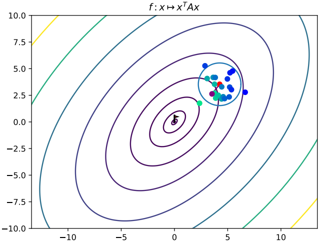
CMA-ES

Algorithm 1 default-CMA-ES

Given: $m_0 \in \mathbb{R}^d,\ \sigma_0 > 0,\ C_0 \in \mathcal{S}^d_{++},\ p_0^\sigma \in \mathbb{R}^d,\ p_0^c \in \mathbb{R}^d$

For $t = 0, 1, 2, \ldots$:

- $U^1_{t+1}, \ldots, U^\lambda_{t+1} \sim \mathcal{N}(0, I_d)$
- sort $f(m_t + \sigma_t C_t^{1/2} U^i_{t+1})$
- $m_{t+1} = m_t + \sigma_t C_t^{1/2} \sum_{i=1}^{\mu} w_i U^{i:\lambda}_{t+1}$



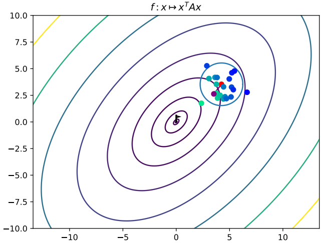
CMA-ES

Algorithm 1 default-CMA-ES

Given: $m_0 \in \mathbb{R}^d,\ \sigma_0 > 0,\ C_0 \in \mathcal{S}^d_{++},\ p_0^\sigma \in \mathbb{R}^d,\ p_0^c \in \mathbb{R}^d$

For $t = 0, 1, 2, \ldots$:

- $U^1_{t+1}, \ldots, U^\lambda_{t+1} \sim \mathcal{N}(0, I_d)$
- sort $f(m_t + \sigma_t C_t^{1/2} U^i_{t+1})$
- $m_{t+1} = m_t + \sigma_t C_t^{1/2} \sum_{i=1}^{\mu} w_i U^{i:\lambda}_{t+1}$
- $p^\sigma_{t+1} = (1 - c_\sigma) p^\sigma_t + \alpha_{c_\sigma} \sum_{i=1}^{\mu} w_i U^{i:\lambda}_{t+1}$



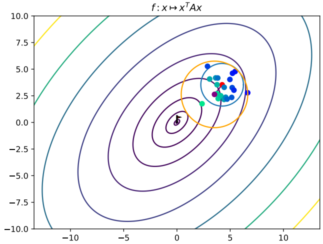
CMA-ES

Algorithm 1 default-CMA-ES

Given: $m_0 \in \mathbb{R}^d,\ \sigma_0 > 0,\ C_0 \in \mathcal{S}^d_{++},\ p_0^\sigma \in \mathbb{R}^d,\ p_0^c \in \mathbb{R}^d$

For $t = 0, 1, 2, \ldots$:

- $U^1_{t+1}, \ldots, U^\lambda_{t+1} \sim \mathcal{N}(0, I_d)$
- sort $f(m_t + \sigma_t C_t^{1/2} U^i_{t+1})$
- $m_{t+1} = m_t + \sigma_t C_t^{1/2} \sum_{i=1}^{\mu} w_i U^{i:\lambda}_{t+1}$
- $p^\sigma_{t+1} = (1 - c_\sigma) p^\sigma_t + \alpha_{c_\sigma} \sum_{i=1}^{\mu} w_i U^{i:\lambda}_{t+1}$
- $\sigma_{t+1} = \sigma_t \exp\left(\frac{c_\sigma}{d_\sigma}\left(\frac{\|p^\sigma_{t+1}\|}{\mathbb{E}\|\mathcal{N}(0, I_d)\|} - 1\right)\right)$



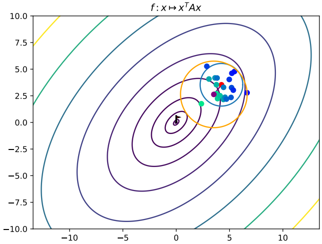
CMA-ES

Algorithm 1 default-CMA-ES

Given: $m_0 \in \mathbb{R}^d,\ \sigma_0 > 0,\ C_0 \in \mathcal{S}^d_{++},\ p_0^\sigma \in \mathbb{R}^d,\ p_0^c \in \mathbb{R}^d$

For $t = 0, 1, 2, \ldots$:

- $U^1_{t+1}, \ldots, U^\lambda_{t+1} \sim \mathcal{N}(0, I_d)$
- sort $f(m_t + \sigma_t C_t^{1/2} U^i_{t+1})$
- $m_{t+1} = m_t + \sigma_t C_t^{1/2} \sum_{i=1}^{\mu} w_i U^{i:\lambda}_{t+1}$
- $p^\sigma_{t+1} = (1 - c_\sigma) p^\sigma_t + \alpha_{c_\sigma} \sum_{i=1}^{\mu} w_i U^{i:\lambda}_{t+1}$
- $\sigma_{t+1} = \sigma_t \exp\left(\frac{c_\sigma}{d_\sigma}\left(\frac{\|p^\sigma_{t+1}\|}{\mathbb{E}\|\mathcal{N}(0, I_d)\|} - 1\right)\right)$
- $p^c_{t+1} = (1 - c_c) p^c_t + \alpha_{c_c} C_t^{1/2} \sum_{i=1}^{\mu} w_i U^{i:\lambda}_{t+1}$



CMA-ES

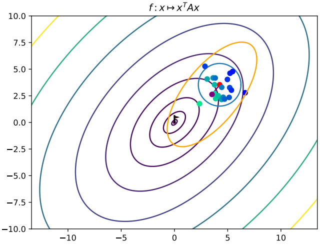
Algorithm 1 default-CMA-ES

Given: $m_0 \in \mathbb{R}^d,\ \sigma_0 > 0,\ C_0 \in \mathcal{S}^d_{++},\ p_0^\sigma \in \mathbb{R}^d,\ p_0^c \in \mathbb{R}^d$

For $t = 0, 1, 2, \dots$:

- $U^1_{t+1}, \dots, U^\lambda_{t+1} \sim \mathcal{N}(0, I_d)$
- sort $f(m_t + \sigma_t C_t^{1/2} U^i_{t+1})$
- $m_{t+1} = m_t + \sigma_t C_t^{1/2} \sum_{i=1}^\mu w_i U^{i:\lambda}_{t+1}$
- $p^\sigma_{t+1} = (1 - c_\sigma) p^\sigma_t + \alpha_{c_\sigma} \sum_{i=1}^\mu w_i U^{i:\lambda}_{t+1}$
- $\sigma_{t+1} = \sigma_t \exp\left(\frac{c_\sigma}{d_\sigma}\left(\frac{\|p^\sigma_{t+1}\|}{\mathbb{E}\|\mathcal{N}(0,I_d)\|} - 1\right)\right)$
- $p^c_{t+1} = (1 - c_c) p^c_t + \alpha_{c_c} C_t^{1/2} \sum_{i=1}^\mu w_i U^{i:\lambda}_{t+1}$
- $C_{t+1} = (1 - c_1 - c_\mu) C_t + c_1 (p^c_{t+1})(p^c_{t+1})^T$

  $+ c_\mu \sum_{i=1}^\mu w_i \left(C_t^{1/2} U^{s_{t+1}(i)}_{t+1}\right)\left(C_t^{1/2} U^{s_{t+1}(i)}_{t+1}\right)^T$



$\mathcal{C}^1$ variant of CMA-ES

Algorithm 2 d-CMA-ES

Given: $m_0 \in \mathbb{R}^d,\ \sigma_0 > 0,\ C_0 \in \mathcal{S}^d_{++},\ p_0^\sigma \in \mathbb{R}^d,\ p_0^c \in \mathbb{R}^d$

For $t = 0, 1, 2, \ldots$:

- $U^1_{t+1}, \ldots, U^\lambda_{t+1} \sim \mathcal{N}(0, I_d)$
- sort $f(m_t + \sigma_t C_t^{1/2} U^i_{t+1})$
- $m_{t+1} = m_t + \sigma_t C_t^{1/2} \sum_{i=1}^{\mu} w_i U^{i:\lambda}_{t+1}$
- $p^\sigma_{t+1} = (1 - c_\sigma) p^\sigma_t + \alpha_{c_\sigma} \sum_{i=1}^{\mu} w_i U^{i:\lambda}_{t+1}$
- $\sigma_{t+1} = \sigma_t \exp\left(\frac{c_\sigma}{d_\sigma}\left(\frac{\|p^\sigma_{t+1}\|}{\mathbb{E}\|\mathcal{N}(0, I_d)\|} - 1\right)\right)$
- $p^c_{t+1} = (1 - c_c) p^c_t + \alpha_{c_c} C_t^{1/2} \sum_{i=1}^{\mu} w_i U^{i:\lambda}_{t+1}$
- $C_{t+1} = (1 - c_1 - c_\mu) C_t + c_1 (p^c_{t+1})(p^c_{t+1})^T$
  $+ c_\mu \sum_{i=1}^{\mu} w_i \left(C_t^{1/2} U^{s_{t+1}(i)}_{t+1}\right)\left(C_t^{1/2} U^{s_{t+1}(i)}_{t+1}\right)^T$

$\mathcal{C}^1$ variant of CMA-ES

Algorithm 2 d-CMA-ES

Given: $m_0 \in \mathbb{R}^d,\ \sigma_0 > 0,\ C_0 \in \mathcal{S}^d_{++},\ p^\sigma_0 \in \mathbb{R}^d,\ p^c_0 \in \mathbb{R}^d$

For $t = 0, 1, 2, \ldots$:

- $U^1_{t+1}, \ldots, U^\lambda_{t+1} \sim \mathcal{N}(0, I_d)$
- sort $f(m_t + \sigma_t C_t^{1/2} U^i_{t+1})$
- $m_{t+1} = m_t + \sigma_t C_t^{1/2} \sum_{i=1}^\mu w_i U^{i:\lambda}_{t+1}$
- $p^\sigma_{t+1} = (1 - c_\sigma) p^\sigma_t + \alpha_{c_\sigma} \sum_{i=1}^\mu w_i U^{i:\lambda}_{t+1}$
- $\sigma_{t+1} = \sigma_t \exp\left(\frac{c_\sigma}{2 d_\sigma}\left(\frac{\|p^\sigma_{t+1}\|^2}{\mathbf{d}} - 1\right)\right)$
- $p^c_{t+1} = (1 - c_c) p^c_t + \alpha_{c_c} C_t^{1/2} \sum_{i=1}^\mu w_i U^{i:\lambda}_{t+1}$
- $C_{t+1} = (1 - c_1 - c_\mu) C_t + c_1 (p^c_{t+1})(p^c_{t+1})^T$
  $+ c_\mu \sum_{i=1}^\mu w_i \left(C_t^{1/2} U^{s_{t+1}(i)}_{t+1}\right)\left(C_t^{1/2} U^{s_{t+1}(i)}_{t+1}\right)^T$

$\mathcal{C}^1$ variant of CMA-ES

Algorithm 2 d-CMA-ES

Given: $m_0 \in \mathbb{R}^d,\ \sigma_0 > 0,\ C_0 \in \mathcal{S}^d_{++},\ p^\sigma_0 \in \mathbb{R}^d,\ p^c_0 \in \mathbb{R}^d$

For $t = 0, 1, 2, \dots$:

- $U^1_{t+1}, \dots, U^\lambda_{t+1} \sim \mathcal{N}(0, I_d)$
- sort $f(m_t + \sigma_t C_t^{1/2} U^i_{t+1})$
- $m_{t+1} = m_t + \sigma_t C_t^{1/2} \sum_{i=1}^{\mu} w_i U^{i:\lambda}_{t+1}$
- $p^\sigma_{t+1} = (1 - c_\sigma) p^\sigma_t + \alpha_{c_\sigma} \sum_{i=1}^{\mu} w_i U^{i:\lambda}_{t+1}$
- $\sigma_{t+1} = \sigma_t \exp\left(\frac{c_\sigma}{2 d_\sigma} \left(\frac{\|p^\sigma_{t+1}\|^2}{\mathbf{d}} - 1 \right) \right)$ **smooth update**
- $p^c_{t+1} = (1 - c_c) p^c_t + \alpha_{c_c} C_t^{1/2} \sum_{i=1}^{\mu} w_i U^{i:\lambda}_{t+1}$
- $C_{t+1} = (1 - c_1 - c_\mu) C_t + c_1 (p^c_{t+1})(p^c_{t+1})^T$
  $+ c_\mu \sum_{i=1}^{\mu} w_i \left(C_t^{1/2} U^{s_{t+1}(i)}_{t+1} \right) \left(C_t^{1/2} U^{s_{t+1}(i)}_{t+1} \right)^T$

One path variant of CMA-ES

Algorithm 3 a-CMA-ES

Given: $m_0 \in \mathbb{R}^d,\ \sigma_0 > 0,\ C_0 \in \mathcal{S}^d_{++},\ p^\sigma_0 \in \mathbb{R}^d,\ p^c_0 \in \mathbb{R}^d$

For $t = 0, 1, 2, \ldots$:

- $U^1_{t+1}, \ldots, U^\lambda_{t+1} \sim \mathcal{N}(0, I_d)$
- $\text{sort } f(m_t + \sigma_t C_t^{1/2} U^i_{t+1})$
- $m_{t+1} = m_t + \sigma_t C_t^{1/2} \sum_{i=1}^{\mu} w_i U^{i:\lambda}_{t+1}$
- ~~$p^\sigma_{t+1} = (1 - c_\sigma) p^\sigma_t + \alpha_{c_\sigma} \sum_{i=1}^{\mu} w_i U^{i:\lambda}_{t+1}$~~
- ~~$\sigma_{t+1} = \sigma_t \exp\left(\frac{c_\sigma}{d_\sigma}\left(\frac{\|p^\sigma_{t+1}\|}{\mathbb{E}\|\mathcal{N}(0, I_d)\|} - 1\right)\right)$~~
- $p^c_{t+1} = (1 - c_c) p^c_t + \alpha_{c_c} C_t^{1/2} \sum_{i=1}^{\mu} w_i U^{i:\lambda}_{t+1}$
- $C_{t+1} = (1 - c_1 - c_\mu) C_t + c_1 (p^c_{t+1})(p^c_{t+1})^T$
  $+ c_\mu \sum_{i=1}^{\mu} w_i \left(C_t^{1/2} U^{s_{t+1}(i)}_{t+1}\right)\left(C_t^{1/2} U^{s_{t+1}(i)}_{t+1}\right)^T$

One path variant of CMA-ES

Algorithm 3 a-CMA-ES

Given: $m_0 \in \mathbb{R}^d,\ \sigma_0 > 0,\ C_0 \in \mathcal{S}^d_{++},\ p^\sigma_0 \in \mathbb{R}^d,\ p^c_0 \in \mathbb{R}^d$

For $t = 0, 1, 2, \dots:$

- $U^1_{t+1}, \dots, U^\lambda_{t+1} \sim \mathcal{N}(0, I_d)$
- sort $f(m_t + \sigma_t C_t^{1/2} U^i_{t+1})$
- $m_{t+1} = m_t + \sigma_t C_t^{1/2} \sum_{i=1}^\mu w_i U^{i:\lambda}_{t+1}$
- ~~$p^\sigma_{t+1} = (1 - c_\sigma) p^\sigma_t + \alpha_{c_\sigma} \sum_{i=1}^\mu w_i U^{i:\lambda}_{t+1}$~~
- $\sigma_{t+1} = \sigma_t \exp\left(\frac{c_\sigma}{d_\sigma}\left(\frac{\|\mathbf{C_t^{-1/2} p^c_{t+1}}\|}{\mathbb{E}\|\mathcal{N}(0,I_d)\|} - 1\right)\right)$
- $p^c_{t+1} = (1 - c_c) p^c_t + \alpha_{c_c} C_t^{1/2} \sum_{i=1}^\mu w_i U^{i:\lambda}_{t+1}$
- $C_{t+1} = (1 - c_1 - c_\mu) C_t + c_1 (p^c_{t+1})(p^c_{t+1})^T$
  $+ c_\mu \sum_{i=1}^\mu w_i \left(C_t^{1/2} U^{s_{t+1}(i)}_{t+1}\right)\left(C_t^{1/2} U^{s_{t+1}(i)}_{t+1}\right)^T$

One path variant of CMA-ES

Algorithm 3 a-CMA-ES

Given: $m_0 \in \mathbb{R}^d,\ \sigma_0 > 0,\ C_0 \in \mathcal{S}^d_{++},\ p^\sigma_0 \in \mathbb{R}^d,\ p^c_0 \in \mathbb{R}^d$

For $t = 0, 1, 2, \ldots$:

- $U^1_{t+1}, \ldots, U^\lambda_{t+1} \sim \mathcal{N}(0, I_d)$
- sort $f(m_t + \sigma_t C_t^{1/2} U^i_{t+1})$
- $m_{t+1} = m_t + \sigma_t C_t^{1/2} \sum_{i=1}^{\mu} w_i U^{i:\lambda}_{t+1}$
- ~~$p^\sigma_{t+1} = (1 - c_\sigma) p^\sigma_t + \alpha_{c_\sigma} \sum_{i=1}^{\mu} w_i U^{i:\lambda}_{t+1}$~~
- $\sigma_{t+1} =$
  $\sigma_t \exp\left(\frac{c_\sigma}{d_\sigma}\left(\frac{\|\mathbf{C_t}^{-1/2} p^{\mathbf{c}}_{t+1}\|}{\mathbb{E}\|\mathcal{N}(0, I_d)\|} - 1\right)\right)$ **only one path is used**
- $p^c_{t+1} = (1 - c_c) p^c_t + \alpha_{c_c} C_t^{1/2} \sum_{i=1}^{\mu} w_i U^{i:\lambda}_{t+1}$
- $C_{t+1} = (1 - c_1 - c_\mu) C_t + c_1 (p^c_{t+1})(p^c_{t+1})^T$
  $+ c_\mu \sum_{i=1}^{\mu} w_i \left(C_t^{1/2} U^{s_{t+1}(i)}_{t+1}\right)\left(C_t^{1/2} U^{s_{t+1}(i)}_{t+1}\right)^T$

CMA-ES without cumulation on the stepsize

Algorithm 4 s-CMA-ES

Given: $m_0 \in \mathbb{R}^d,\ \sigma_0 > 0,\ C_0 \in \mathcal{S}^d_{++},\ p_0^\sigma \in \mathbb{R}^d,\ p_0^c \in \mathbb{R}^d$

For $t = 0, 1, 2, \ldots$:

- $U^1_{t+1}, \ldots, U^\lambda_{t+1} \sim \mathcal{N}(0, I_d)$
- sort $f(m_t + \sigma_t C_t^{1/2} U^i_{t+1})$
- $m_{t+1} = m_t + \sigma_t C_t^{1/2} \sum_{i=1}^{\mu} w_i U^{i:\lambda}_{t+1}$
- $p^\sigma_{t+1} = \xcancel{(1 - c_\sigma) p^\sigma_t} + \alpha_{c_\sigma} \sum_{i=1}^{\mu} w_i U^{i:\lambda}_{t+1}$
- $\sigma_{t+1} = \sigma_t \exp\left(\frac{c_\sigma}{d_\sigma}\left(\frac{\|p^\sigma_{t+1}\|}{\mathbb{E}\|\mathcal{N}(0,I_d)\|} - 1\right)\right)$
- $p^c_{t+1} = (1 - c_c) p^c_t + \alpha_{c_c} C_t^{1/2} \sum_{i=1}^{\mu} w_i U^{i:\lambda}_{t+1}$
- $C_{t+1} = (1 - c_1 - c_\mu) C_t + c_1 (p^c_{t+1})(p^c_{t+1})^T$
  $+ c_\mu \sum_{i=1}^{\mu} w_i \left(C_t^{1/2} U^{s_{t+1}(i)}_{t+1}\right)\left(C_t^{1/2} U^{s_{t+1}(i)}_{t+1}\right)^T$

Algorithm 4 s-CMA-ES

Given: $m_0 \in \mathbb{R}^d,\ \sigma_0 > 0,\ C_0 \in \mathcal{S}^d_{++},\ p^\sigma_0 \in \mathbb{R}^d,\ p^c_0 \in \mathbb{R}^d$

For $t = 0, 1, 2, \dots :$

- $U^1_{t+1}, \dots, U^\lambda_{t+1} \sim \mathcal{N}(0, I_d)$
- sort $f(m_t + \sigma_t C_t^{1/2} U^i_{t+1})$
- $m_{t+1} = m_t + \sigma_t C_t^{1/2} \sum_{i=1}^{\mu} w_i U^{i:\lambda}_{t+1}$
- $p^\sigma_{t+1} = \qquad \alpha_1 \sum_{i=1}^{\mu} w_i U^{i:\lambda}_{t+1}$
- $\sigma_{t+1} = \sigma_t \exp\left(\frac{c_\sigma}{d_\sigma}\left(\frac{\|p^\sigma_{t+1}\|}{\mathbb{E}\|\mathcal{N}(0,I_d)\|} - 1\right)\right)$
- $p^c_{t+1} = (1 - c_c)p^c_t + \alpha_{c_c} C_t^{1/2} \sum_{i=1}^{\mu} w_i U^{i:\lambda}_{t+1}$
- $C_{t+1} = (1 - c_1 - c_\mu)C_t + c_1(p^c_{t+1})(p^c_{t+1})^T$
  $+c_\mu \sum_{i=1}^{\mu} w_i \left(C_t^{1/2} U^{s_{t+1}(i)}_{t+1}\right)\left(C_t^{1/2} U^{s_{t+1}(i)}_{t+1}\right)^T$

CMA-ES without cumulation on the stepsize

Algorithm 4 s-CMA-ES

Given: $m_0 \in \mathbb{R}^d$, $\sigma_0 > 0$, $C_0 \in \mathcal{S}^d_{++}$, $p^\sigma_0 \in \mathbb{R}^d$, $p^c_0 \in \mathbb{R}^d$

For $t = 0, 1, 2, \ldots$:

- $U^1_{t+1}, \ldots, U^\lambda_{t+1} \sim \mathcal{N}(0, I_d)$
- sort $f(m_t + \sigma_t C_t^{1/2} U^i_{t+1})$
- $m_{t+1} = m_t + \sigma_t C_t^{1/2} \sum_{i=1}^\mu w_i U^{i:\lambda}_{t+1}$
- $p^\sigma_{t+1} = \alpha_1 \sum_{i=1}^\mu w_i U^{i:\lambda}_{t+1}$ **no cumulation on the stepsize**
- $\sigma_{t+1} = \sigma_t \exp\left(\frac{c_\sigma}{d_\sigma}\left(\frac{\|p^\sigma_{t+1}\|}{\mathbb{E}\|\mathcal{N}(0,I_d)\|} - 1\right)\right)$
- $p^c_{t+1} = (1 - c_c)p^c_t + \alpha_{c_c} C_t^{1/2} \sum_{i=1}^\mu w_i U^{i:\lambda}_{t+1}$
- $C_{t+1} = (1 - c_1 - c_\mu)C_t + c_1(p^c_{t+1})(p^c_{t+1})^T$
  $+c_\mu \sum_{i=1}^\mu w_i \left(C_t^{1/2} U^{s_{t+1}(i)}_{t+1}\right)\left(C_t^{1/2} U^{s_{t+1}(i)}_{t+1}\right)^T$

CMA-ES without cumulation on the covariance matrix

Algorithm 5 d-CMA-ES

Given: $m_0 \in \mathbb{R}^d,\ \sigma_0 > 0,\ C_0 \in \mathcal{S}^d_{++},\ p^\sigma_0 \in \mathbb{R}^d,\ p^c_0 \in \mathbb{R}^d$

For $t = 0, 1, 2, \ldots:$

- $U^1_{t+1}, \ldots, U^\lambda_{t+1} \sim \mathcal{N}(0, I_d)$
- sort $f(m_t + \sigma_t C_t^{1/2} U^i_{t+1})$
- $m_{t+1} = m_t + \sigma_t C_t^{1/2} \sum_{i=1}^\mu w_i U^{i:\lambda}_{t+1}$
- $p^\sigma_{t+1} = (1 - c_\sigma) p^\sigma_t + \alpha_{c_\sigma} \sum_{i=1}^\mu w_i U^{i:\lambda}_{t+1}$
- $\sigma_{t+1} = \sigma_t \exp\left(\frac{c_\sigma}{d_\sigma}\left(\frac{\|p^\sigma_{t+1}\|}{\mathbb{E}\|\mathcal{N}(0,I_d)\|} - 1\right)\right)$
- $p^c_{t+1} = \cancel{(1 - c_c) p^c_t} + \alpha_{c_c} C_t^{1/2} \sum_{i=1}^\mu w_i U^{i:\lambda}_{t+1}$
- $C_{t+1} = (1 - c_1 - c_\mu) C_t + c_1 (p^c_{t+1})(p^c_{t+1})^T$
  $+ c_\mu \sum_{i=1}^\mu w_i \left(C_t^{1/2} U^{s_{t+1}(i)}_{t+1}\right)\left(C_t^{1/2} U^{s_{t+1}(i)}_{t+1}\right)^T$

CMA-ES without cumulation on the covariance matrix

Algorithm 5 d-CMA-ES

Given: $m_0 \in \mathbb{R}^d,\ \sigma_0 > 0,\ C_0 \in \mathcal{S}^d_{++},\ p^\sigma_0 \in \mathbb{R}^d,\ p^c_0 \in \mathbb{R}^d$

For $t = 0, 1, 2, \ldots:$

- $U^1_{t+1}, \ldots, U^\lambda_{t+1} \sim \mathcal{N}(0, I_d)$
- sort $f(m_t + \sigma_t C_t^{1/2} U^i_{t+1})$
- $m_{t+1} = m_t + \sigma_t C_t^{1/2} \sum_{i=1}^{\mu} w_i U^{i:\lambda}_{t+1}$
- $p^\sigma_{t+1} = (1 - c_\sigma) p^\sigma_t + \alpha_{c_\sigma} \sum_{i=1}^{\mu} w_i U^{i:\lambda}_{t+1}$
- $\sigma_{t+1} = \sigma_t \exp\left(\frac{c_\sigma}{d_\sigma}\left(\frac{\|p^\sigma_{t+1}\|}{\mathbb{E}\|\mathcal{N}(0,I_d)\|} - 1\right)\right)$
- $p^c_{t+1} = \alpha_{\mathbf{1}} C_t^{1/2} \sum_{i=1}^{\mu} w_i U^{i:\lambda}_{t+1}$
- $C_{t+1} = (1 - c_1 - c_\mu) C_t + c_1 (p^c_{t+1})(p^c_{t+1})^T$
  $+ c_\mu \sum_{i=1}^{\mu} w_i \left(C_t^{1/2} U^{s_{t+1}(i)}_{t+1}\right)\left(C_t^{1/2} U^{s_{t+1}(i)}_{t+1}\right)^T$

CMA-ES without cumulation on the covariance matrix

Algorithm 5 d-CMA-ES

Given: $m_0 \in \mathbb{R}^d$, $\sigma_0 > 0$, $C_0 \in \mathcal{S}^d_{++}$, $p^\sigma_0 \in \mathbb{R}^d$, $p^c_0 \in \mathbb{R}^d$

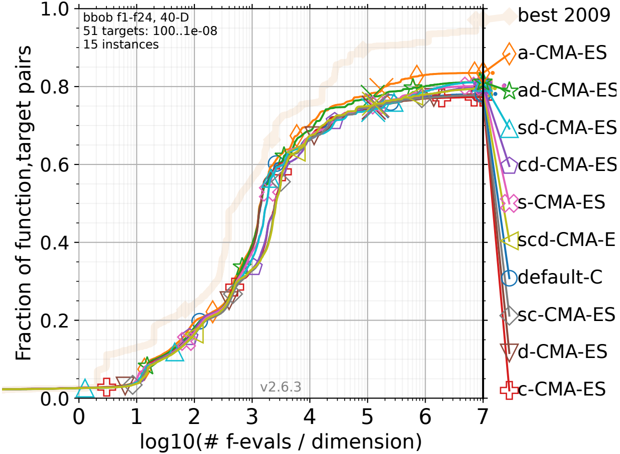
For $t = 0, 1, 2, \ldots$:

- $U^1_{t+1}, \ldots, U^\lambda_{t+1} \sim \mathcal{N}(0, I_d)$
- sort $f(m_t + \sigma_t C_t^{1/2} U^i_{t+1})$
- $m_{t+1} = m_t + \sigma_t C_t^{1/2} \sum_{i=1}^{\mu} w_i U^{i:\lambda}_{t+1}$
- $p^\sigma_{t+1} = (1 - c_\sigma) p^\sigma_t + \alpha_{c_\sigma} \sum_{i=1}^{\mu} w_i U^{i:\lambda}_{t+1}$
- $\sigma_{t+1} = \sigma_t \exp\left(\frac{c_\sigma}{d_\sigma}\left(\frac{\|p^\sigma_{t+1}\|}{\mathbb{E}\|\mathcal{N}(0, I_d)\|} - 1\right)\right)$
- $p^c_{t+1} = \alpha_{\mathbf{1}} C_t^{1/2} \sum_{i=1}^{\mu} w_i U^{i:\lambda}_{t+1}$ **no cumulation on $\mathbf{C_t}$**
- $C_{t+1} = (1 - c_1 - c_\mu) C_t + c_1 (p^c_{t+1})(p^c_{t+1})^T$
  $+ c_\mu \sum_{i=1}^{\mu} w_i \left(C_t^{1/2} U^{s_{t+1}(i)}_{t+1}\right)\left(C_t^{1/2} U^{s_{t+1}(i)}_{t+1}\right)^T$

10 variants of CMA-ES

- default-CMA-ES
- d-CMA-ES: with a smooth update on the stepsize
- a-CMA-ES: using one path for both cumulations
- s-CMA-ES: no cumulation on the stepsize
- c-CMA-ES: no cumulation on the covariance matrix
- ad-CMA-ES, sd-CMA-ES, cd-CMA-ES, sc-CMA-ES, scd-CMA-ES

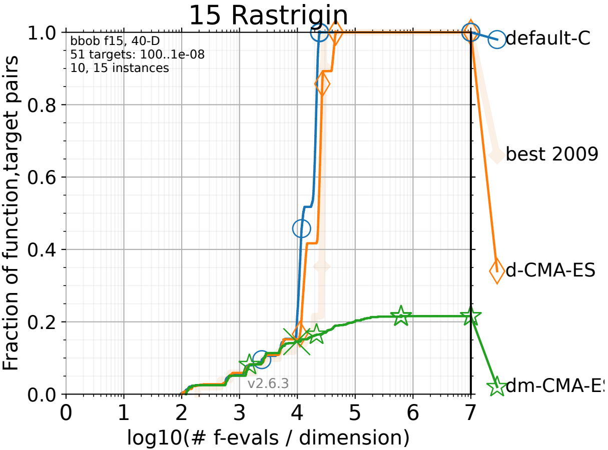
10 variants of CMA-ES



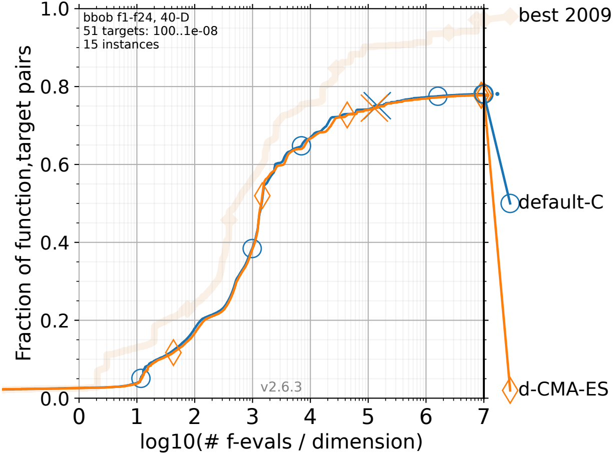
10 **variants of CMA-ES**

- default-CMA-ES
- **d-CMA-ES**: with a smooth update on the stepsize
- a-CMA-ES: using one path for both cumulations
- s-CMA-ES: no cumulation on the stepsize
- c-CMA-ES: no cumulation on the covariance matrix
- ad-CMA-ES, sd-CMA-ES, cd-CMA-ES, sc-CMA-ES, scd-CMA-ES

d-CMA-ES: the parameter d_σ has to be scaled correctly



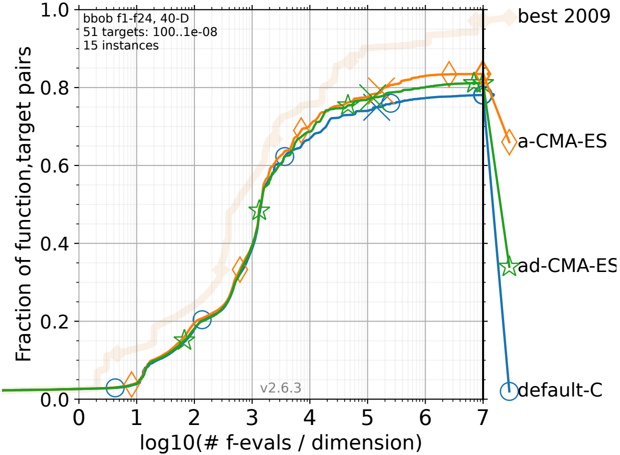
d-CMA-ES: no significant differences



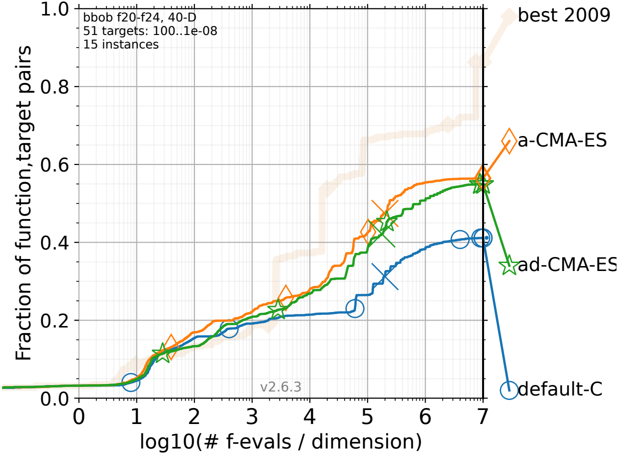
10 **variants of CMA-ES**

- default-CMA-ES
- d-CMA-ES: with a smooth update on the stepsize
- **a-CMA-ES**: using one path for both cumulations
- s-CMA-ES: no cumulation on the stepsize
- c-CMA-ES: no cumulation on the covariance matrix
- **ad-CMA-ES**, sd-CMA-ES, cd-CMA-ES, sc-CMA-ES, scd-CMA-ES

a-CMA-ES

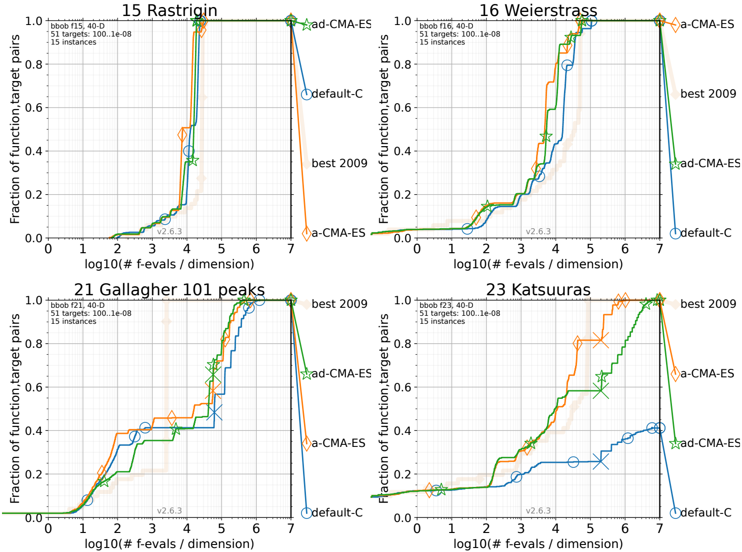

bbob f1-f24, 40-D
51 targets: 100..1e-08
15 instances
v2.6.3
best 2009
a-CMA-ES
ad-CMA-ES
default-C
Fraction of function,target pairs
log10(# f-evals / dimension)
1.0
0.8
0.6
0.4
0.2
0.0
0
1
2
3
4
5
6
7

a-CMA-ES



bbob f20-f24, 40-D
51 targets: 100..1e-08
15 instances
Fraction of function,target pairs
log10(# f-evals / dimension)
v2.6.3
best 2009
a-CMA-ES
ad-CMA-ES
default-C

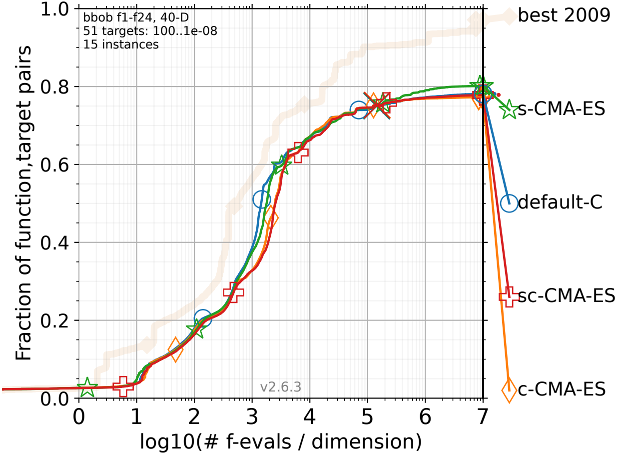
a-CMA-ES


15 Rastrigin
bbob f15, 40-D
51 targets: 100..1e-08
15 instances
Fraction of function,target pairs
log10(# f-evals / dimension)
ad-CMA-ES
default-C
best 2009
a-CMA-ES
v2.6.3
16 Weierstrass
bbob f16, 40-D
51 targets: 100..1e-08
15 instances
a-CMA-ES
best 2009
ad-CMA-ES
default-C
21 Gallagher 101 peaks
bbob f21, 40-D
51 targets: 100..1e-08
15 instances
best 2009
ad-CMA-ES
a-CMA-ES
default-C
23 Katsuuras
bbob f23, 40-D
51 targets: 100..1e-08
15 instances
best 2009
a-CMA-ES
ad-CMA-ES
default-C

10 **variants of CMA-ES**

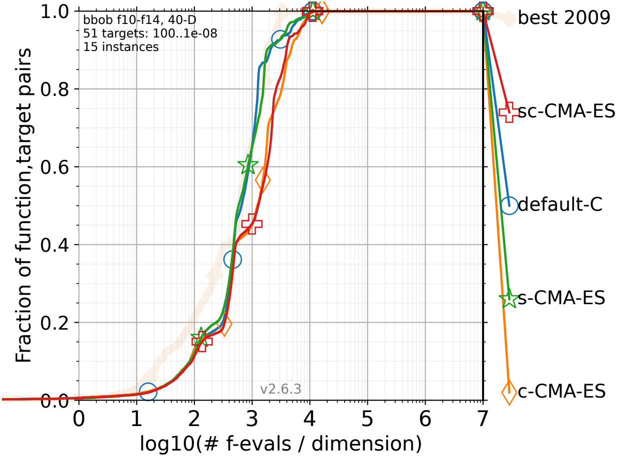
- default-CMA-ES
- d-CMA-ES: with a smooth update on the stepsize
- a-CMA-ES: using one path for both cumulations
- **s-CMA-ES**: no cumulation on the stepsize
- **c-CMA-ES**: no cumulation on the covariance matrix
- ad-CMA-ES, sd-CMA-ES, cd-CMA-ES, **sc-CMA-ES**, scd-CMA-ES

s,c,sc-CMA-ES



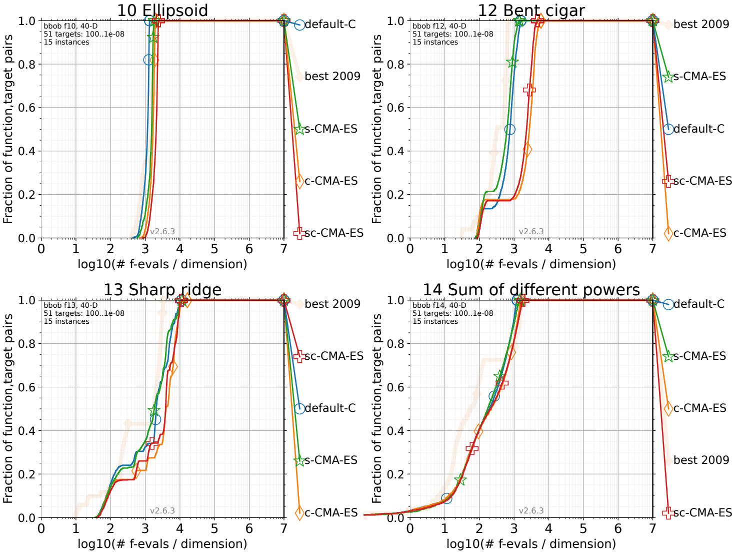
bbob f1-f24, 40-D
51 targets: 100..1e-08
15 instances
Fraction of function,target pairs
log10(# f-evals / dimension)
1.0
0.8
0.6
0.4
0.2
0.0
0
1
2
3
4
5
6
7
v2.6.3
best 2009
s-CMA-ES
default-C
sc-CMA-ES
c-CMA-ES

s,c,sc-CMA-ES



bbob f10-f14, 40-D
51 targets: 100..1e-08
15 instances
v2.6.3
Fraction of function,target pairs
log10(# f-evals / dimension)
1.0
0.8
0.6
0.4
0.2
0.0
0
1
2
3
4
5
6
7
best 2009
sc-CMA-ES
default-C
s-CMA-ES
c-CMA-ES

s,c,sc-CMA-ES


10 Ellipsoid
bbob f10, 40-D
51 targets: 100..1e-08
15 instances
default-C
best 2009
s-CMA-ES
c-CMA-ES
sc-CMA-ES
12 Bent cigar
bbob f12, 40-D
51 targets: 100..1e-08
15 instances
best 2009
s-CMA-ES
default-C
sc-CMA-ES
c-CMA-ES
13 Sharp ridge
bbob f13, 40-D
51 targets: 100..1e-08
15 instances
best 2009
sc-CMA-ES
default-C
s-CMA-ES
c-CMA-ES
14 Sum of different powers
bbob f14, 40-D
51 targets: 100..1e-08
15 instances
default-C
s-CMA-ES
c-CMA-ES
best 2009
sc-CMA-ES
Fraction of function,target pairs
log10(# f-evals / dimension)
v2.6.3

Conclusions

- For **theoretical** purposes, we introduce several modifications to CMA-ES
- None of them **break** the algorithm – in the sense that the algorithm still converges to the solution of the problems it already solved
- **Cumulation on the covariance matrix** is important for the performances of the algorithm on highly **ill-conditionned** problems
- Otherwise there is no significantly remarkable loss of performances due to these modifications

Thank you!