# Weighted Multirecombination Evolution Strategy with Mutation Strength Self-Adaptation on Quadratic Sphere

A. Melkozerov

Research Centre for Process and Product Engineering
Vorarlberg University of Applied Sciences

# Outline

# The $(\mu/\mu_I, \lambda)$-$\sigma$SA-ES

1. Initialize the parent state $(\sigma_p, \mathbf{y}_p) \leftarrow (\sigma_{init}, \mathbf{y}_{init})$
2. Generate $\lambda$ offspring according to

$$\forall l = 1, \ldots, \lambda : \begin{cases} \tilde{\sigma}_l \leftarrow \sigma_p \mathrm{e}^{\tau \mathrm{N}_l(0,1)} \\ \tilde{\mathbf{z}}_l \leftarrow \mathbf{N}_l(\mathbf{0}, \mathbf{I}) \\ \tilde{\mathbf{y}}_l \leftarrow \mathbf{y}_p + \tilde{\sigma}_l \tilde{\mathbf{z}}_l \\ \text{\pounds}_l \leftarrow f(\tilde{\mathbf{y}}_l). \end{cases}$$

3. Order $\lambda$ offspring according to its objective function values
4. Perform recombination

$$\langle \sigma \rangle \leftarrow \frac{1}{\mu} \sum_{m=1}^{\mu} \tilde{\sigma}_{m;\lambda}$$

$$\langle \mathbf{y} \rangle \leftarrow \frac{1}{\mu} \sum_{m=1}^{\mu} \tilde{\mathbf{y}}_{m;\lambda}$$

5. Create new parent state $(\sigma_p, \mathbf{y}_p) \leftarrow (\langle \sigma \rangle, \langle \mathbf{y} \rangle)$
6. Goto 2. until termination criterion fulfilled

## Learning Parameter

- $\tilde{\sigma}_l \leftarrow \sigma_p \mathrm{e}^{\tau \mathrm{N}_l(0,1)}$, where $\mathrm{e}^{\tau \mathrm{N}_l(0,1)}$ – the log-normal operator
- The learning parameter $\tau$ controls the self-adaptation rate
- For $(\mu/\mu_I, \lambda)$-$\sigma$SA-ES, sphere model:
  - The performance is dependent on $\tau$

$$\tau = \frac{\alpha}{\sqrt{N}}$$

  - $N \to \infty$, large populations $\Rightarrow$ optimal $\alpha = \frac{1}{\sqrt{2}}$ [2]

# Weighted Multirecombination

- $(\mu/\rho \stackrel{+}{,} \lambda)$-ES discards $(\lambda - \mu)$ worst individuals
- $(\lambda)_{\text{opt}}$-CSA-ES proposed by Arnold overcomes this weak point by means of weighted multirecombination
- Rank the individuals w.r.t. $f(\tilde{\mathbf{y}}_l)$ values
- Compute the weighted sum

$$\langle \mathbf{z} \rangle_\omega = \sum_{l=1}^{\lambda} \omega_{l,\lambda} \mathbf{z}^{(l;\lambda)}$$

  - $(l;\lambda)$ – the $l$th best of the $\lambda$ offspring
- Weights $\omega_{l,\lambda}$ are dependent on the rank of the individual

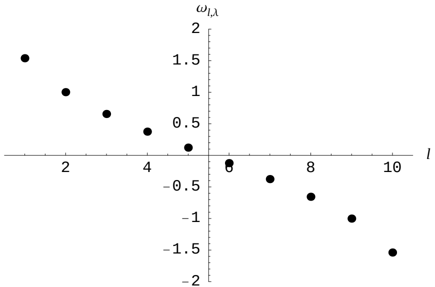
# Optimal Weighted Multirecombination

- Optimally chosen weights $\omega_{l,\lambda}$ maximize the quality gain $\triangle$

$$\omega_{l,\lambda} = E_{l,\lambda} \text{ for } l = 1, \ldots, \lambda,$$

where $E_{l,\lambda}$ – the expectation of the $(\lambda + 1 - l)$th order statistic

- The weighted multirecombination ES with optimally chosen weights – the $(\lambda)_{opt}$-ES



# Analysis of ES

- The quadratic sphere

$$f(\mathbf{y}) = \| \hat{\mathbf{y}} - \mathbf{y} \|^2$$

## Definitions

- $r^{(g)}$ – the distance of the centroid in the $g$th generation $\langle \mathbf{y} \rangle^{(g)}$ to the optimum $\hat{\mathbf{y}}$
- $s^{(g)} = \langle \sigma^{(g)} \rangle$ – the mean value of the parental $\sigma$

# The Progress Rate and the SAR

## Definitions

- The progress rate

$$\varphi\left(s^{(g)}, r^{(g)}\right) = \mathrm{E}\left[r^{(g)} - r^{(g+1)}\right]$$

- The self-adaptation response (SAR)

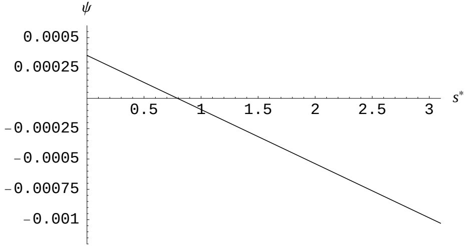
$$\psi\left(s^{(g)}, r^{(g)}\right) = \mathrm{E}\left[\frac{s^{(g+1)} - s^{(g)}}{s^{(g)}}\right]$$

- Independent of the position in the search space

$$\varphi^* = \varphi \frac{N}{r^{(g)}} \text{ and } s^{*(g)} = s^{(g)} \frac{N}{r^{(g)}}$$

- An approximate SAR formula ($\tau \to 0$, $N \to \infty$, Taylor series expansions) [3]

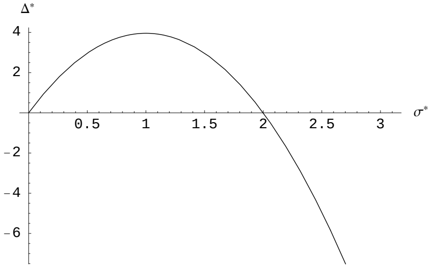
$$\psi(s^{*(g)}) \approx \tau^2 \left[ \frac{1}{2} + e^{1,1}_{\mu,\lambda} - s^{*(g)} c_{\mu/\mu,\lambda} \right]$$



# Analysis of the $(\lambda)_{opt}$-ES

- The normalized quality gain $\triangle^* = \triangle \frac{N}{2\left(r^{(g)}\right)^2}$
- Using simplifications ($N \to \infty$, $\sigma^*$ is of $\mathcal{O}(1)$, Taylor series expansion) [1]

$$\triangle^*(\sigma^{*(g)}) = W_\lambda \left( \sigma^{*(g)} - \frac{(\sigma^{*(g)})^2}{2} \right)$$



# The $(\lambda)_{\text{opt}}$-$\sigma$SA-ES

- The idea – to design a self-adaptive ES that performs comparably well when using weighted multirecombination on the object parameters
- The weighted multirecombination can be introduced in the $(\mu/\mu_I, \lambda)$-$\sigma$SA-ES algorithm
- The result – a new $(\lambda)_{\text{opt}}$-$\sigma$SA-ES

# The $(\lambda)_{\text{opt}}$-$\sigma$SA-ES

1. Initialize the parent state
2. Generate $\lambda$ offspring
3. Order $\lambda$ offspring according to its objective function values
4. Perform recombination of mutation strengths according to

$$\langle \sigma \rangle \leftarrow \frac{1}{\mu} \sum_{m=1}^{\mu} \tilde{\sigma}_{m;\lambda}$$

5. Compute the weighted sum $\langle \mathbf{z} \rangle_\omega$ according to

$$\langle \mathbf{z} \rangle_\omega \leftarrow \sum_{l=1}^{\lambda} \omega_{l;\lambda} \tilde{\mathbf{z}}_{l;\lambda}$$

6. Create new parent state

$$\sigma_p \leftarrow \langle \sigma \rangle$$

$$\mathbf{y}_p \leftarrow \mathbf{y}_p + \langle \sigma \rangle \langle \mathbf{z} \rangle_\omega$$

7. Goto 2. until termination criterion fulfilled

# Analysis of the $(\lambda)_{\text{opt}}$-$\sigma$SA-ES

- Assumptions for the sphere environment:
  - $s^{(g)} \overset{N\to\infty}{=} \sigma_p^{(g-1)}$
  - $\triangle^*\left(s^{*(g)}\right) \overset{N\to\infty}{=} \varphi^*(s^{*(g)})$
- The normalized progress rate

$$\varphi^*(s^{*(g)}) = W_\lambda \left( s^{*(g)} - \frac{(s^{*(g)})^2}{2} \right)$$

## The Steady State of the $(\lambda)_{\text{opt}}$-$\sigma$SA-ES

- Correctly working $\sigma$-SA $\Rightarrow$

$$\lim_{g\to\infty} s^{*(g)} = s^*_{st}$$

- The steady state condition

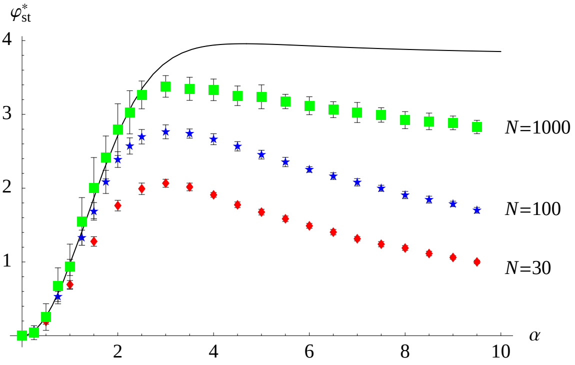
$$\frac{\varphi^*(s^{*(g)}_{st})}{N} = -\psi\left(s^{*(g)}_{st}\right)$$

- Inserting $\varphi^*(s^{*(g)})$ and $\psi\left(s^{*(g)}\right)$, solving the equation $\Rightarrow$ the stationary mutation strength

$$s^*_{st} = 1 - \frac{c_{\mu/\mu,\lambda}\,\alpha^2}{W_\lambda} + K$$

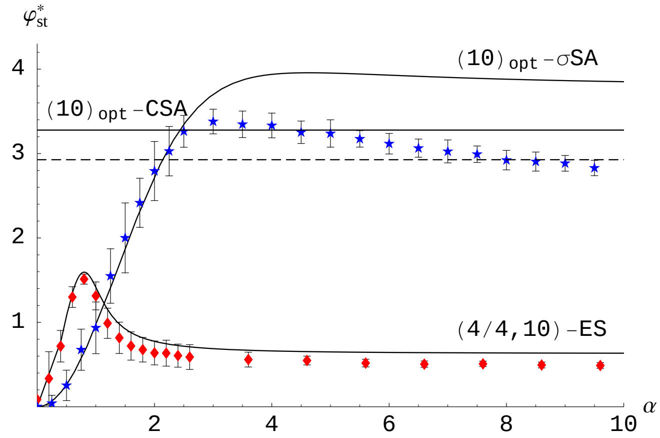
# The Stationary Progress Rate

- Insert $s^*_{st}$ into $\varphi^*(s^{*(g)}) \Rightarrow \varphi^*_{st}(\alpha) = \frac{W_\lambda}{2}\left[1 - \left(\frac{c_{\mu/\mu,\lambda}\alpha^2}{W_\lambda} - K\right)^2\right]$
- The stationary progress rate of the $(\lambda)_{\text{opt}}$-$\sigma$SA-ES ($\mu = 4$, $\lambda = 10$, $\mathbf{y}^{(0)} = \mathbf{1000}$, $\sigma^{(0)} = 1$, averaged over 30 runs)



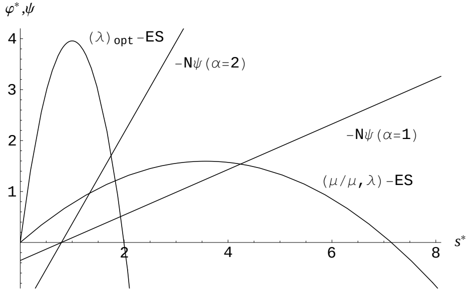
# Comparison with Other ES Versions

- Experimental results for $N = 1000$, $\mu = 4$, $\lambda = 10$, $\mathbf{y}^{(0)} = \mathbf{1000}$, $\sigma^{(0)} = 1$, averaged over 30 runs.



# Why the $(\mu/\mu_I, \lambda)$-ES Outperforms the $(\lambda)_{\text{opt}}$-$\sigma$SA-ES for Small $\alpha$?

- Intersections of the progress rate curves and the SAR lines correspond to the solutions of $\frac{\varphi^*(s_{st}^{*(g)})}{N} = -\psi\left(s_{st}^{*(g)}\right)$
- Small $\alpha$: $s_{st}^*$ larger, $\varphi_{st}^*$ larger for $(\mu/\mu_I, \lambda)$-ES



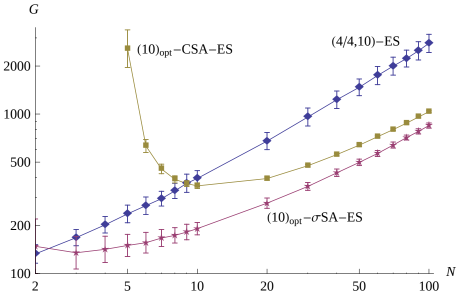
# Optimal Parameter $\alpha$

- $\varphi^*(s^{*(g)}) = W_\lambda \left( s^{*(g)} - \frac{(s^{*(g)})^2}{2} \right) \Rightarrow \varphi^*_{\text{max}} = \frac{W_\lambda}{2}$ at $s_{\varphi^*_{\text{max}}} = 1$
- Maximal progress in the stationary state $\Rightarrow s^*_{st} = s_{\varphi^*_{\text{max}}} = 1$

$$\Rightarrow \alpha_{opt} = \sqrt{\frac{W_\lambda}{2c_{\mu/\mu,\lambda} - 2e^{1,1}_{\mu,\lambda} - 1}}$$

# Comparison with other ES for Different $N$

- $\alpha$: $\alpha = \alpha_{opt}$ for the $(\lambda)_{\text{opt}}$-$\sigma$SA-ES, $\alpha = \frac{1}{\sqrt{2}}$ for the $(\mu/\mu_I, \lambda)$-ES
- Number of generations $G$ required to reach an objective function value of $f(\mathbf{y}) = 10^{-10}$ ($\mu = 4$, $\lambda = 10$, $\mathbf{y}^{(0)} = \mathbf{1000}$, $\sigma^{(0)} = 1$, averaged over 300 runs)



# Investigation of the Noisy Case

- The noisy sphere model

$$f(\mathbf{y}) = \| \hat{\mathbf{y}} - \mathbf{y} \|^2 + \varepsilon,$$

  $\varepsilon \sim \mathrm{N}(0, s_\varepsilon)$, where $s_\varepsilon$ – *noise strength*
- The constant normalized noise strength case $s_\varepsilon^* = \mathrm{const}$, where $s_\varepsilon^* = \frac{N}{2\left(r^{(g)}\right)^2} s_\varepsilon$
- Scaled optimal weights $\omega_{l,\lambda} = E_{l,\lambda}/\kappa$, $\kappa > 0$:
  - decrease the maximal possible progress rate in the non-noisy fitness case
  - the strategy will work with larger mutations $\Rightarrow$ can be beneficial in the noisy fitness case

# The Steady State in the Noisy Case

- $\varphi^*(s^{*(g)}, s_\varepsilon^{*(g)}) = \frac{W_\lambda}{\kappa}\left(\frac{(s^{*(g)})^2}{\sqrt{(s^{*(g)})^2+(s_\varepsilon^{*(g)})^2}} - \frac{(s^{*(g)})^2}{2\kappa}\right)$ [1]
- $\psi(s^{*(g)}, s_\varepsilon^{*(g)}) \approx$
  $\tau^2\left(\frac{1}{2} + \frac{(s^{*(g)})^2}{(s^{*(g)})^2+(s_\varepsilon^{*(g)})^2} e_{\mu,\lambda}^{1,1} - \frac{(s^{*(g)})^2}{\sqrt{(s^{*(g)})^2+(s_\varepsilon^{*(g)})^2}} c_{\mu/\mu,\lambda}\right)$ [2]
- Equilibrium condition $\frac{\varphi^*(s_{st}^{*(g)}, s_\varepsilon^{*(g)})}{N} = -\psi\left(s_{st}^{*(g)}, s_\varepsilon^{*(g)}\right) \Rightarrow$ numerical solving

# $\alpha_{opt}$ in the Noisy Case

1. Numerical maximization of the numerically obtained stationary progress rate
2. The non-noisy case formula for scaled weights

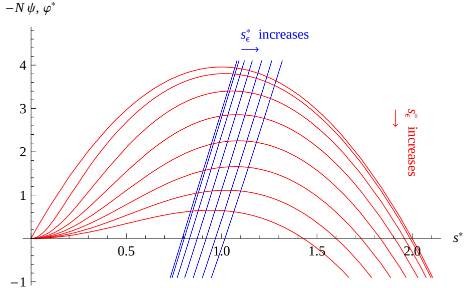
$$\alpha_{opt} = \sqrt{\frac{W_\lambda}{2\kappa c_{\mu/\mu,\lambda} - 2e^{1,1}_{\mu,\lambda} - 1}}$$

3. The approximate analytic formula

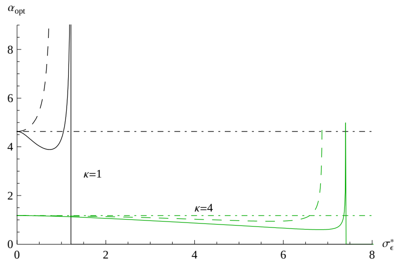
$$\alpha_{opt} \approx \sqrt{\frac{W_\lambda}{\kappa^2} \frac{\kappa^2 + \frac{3}{4}\left(s^*_\varepsilon\right)^2 - \kappa\sqrt{4\kappa^2 + 3\left(s^*_\varepsilon\right)^2}}{1 + 2e^{1,1}_{\mu,\lambda} - 4\frac{\left(s^*_\varepsilon\right)^2}{\left(s^*_\varepsilon\right)^2 - 4\kappa^2} - c_{\mu/\mu,\lambda}\sqrt{4\kappa^2 + 3\left(s^*_\varepsilon\right)^2}}}$$

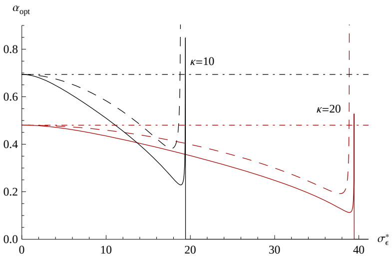
# The Approximate Analytic Formula

- Assumption: the progress rate curves can be approximated by parabolas
- Second zero of the progress rate $s^*_{\varphi^*_0} = \sqrt{4\kappa^2 - s^{*2}_\varepsilon}$
- The optimum of the progress rate $s^*_{opt} \approx s^*_{\varphi^*_0}/2 \Rightarrow$ resolve the steady state condition for $s^*_{st} = s^*_{opt}$

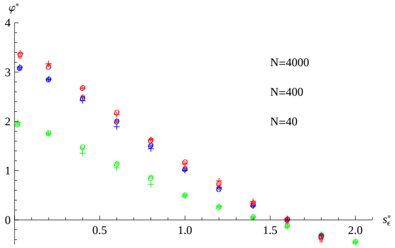


- Solid curves – numerically obtained $\alpha_{opt}$. Dashed curves – the approximate formula. Dot-dashed lines – the non-noisy case formula

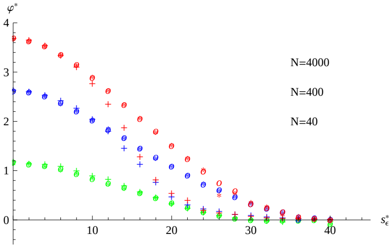


- The stationary progress rate of the $(\lambda)_{\rm opt}$-$\sigma$SA-ES ($\mu = 4$, $\lambda = 10$, $\mathbf{y}^{(0)} = \mathbf{10}$, $\sigma^{(0)} = 1$)
- Circles – numerically obtained $\alpha_{opt}$. Stars – the approximate formula. Crosses – the non-noisy case formula



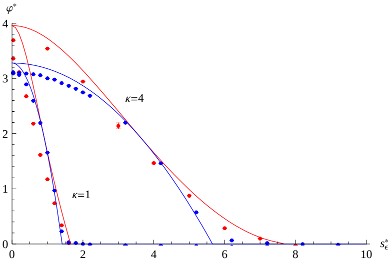

(a) $\kappa = 1$

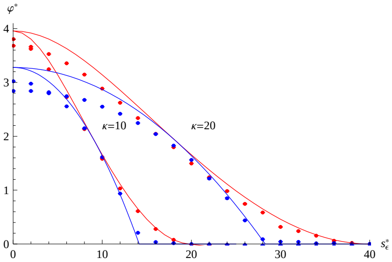



(b) $\kappa = 20$

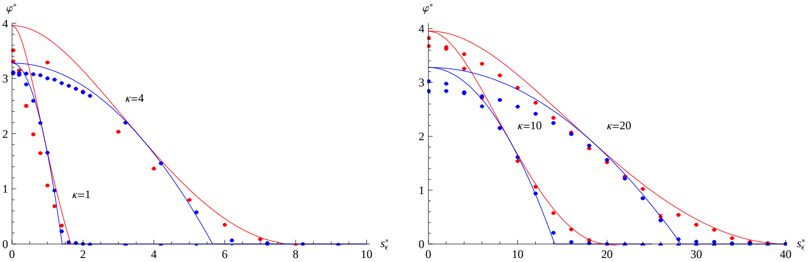
# Comparison with the $(\lambda)_{\text{opt}}$-CSA-ES

- The stationary progress rate of the $(\lambda)_{\text{opt}}$-$\sigma$SA-ES (red points) and $(\lambda)_{\text{opt}}$-CSA-ES (blue points) for $N = 4000$, $\mu = 4$, $\lambda = 10$, $\mathbf{y}^{(0)} = \mathbf{10}$, $\sigma^{(0)} = 1$
- **Numerically obtained $\alpha_{opt}$**

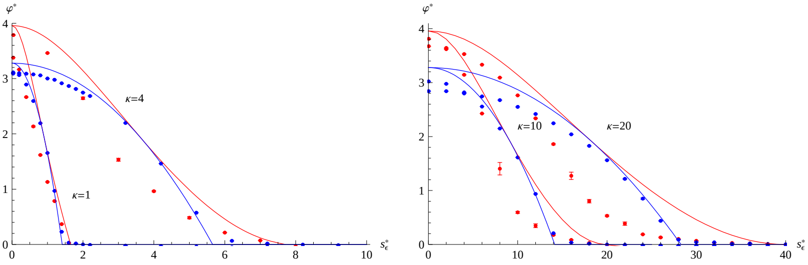


- The stationary progress rate of the $(\lambda)_{\text{opt}}$-$\sigma$SA-ES (red points) and $(\lambda)_{\text{opt}}$-CSA-ES (blue points) for $N = 4000$, $\mu = 4$, $\lambda = 10$, $\mathbf{y}^{(0)} = \mathbf{10}$, $\sigma^{(0)} = 1$
- $\alpha_{opt}$: **the approximate formula**



- The stationary progress rate of the $(\lambda)_{\text{opt}}$-$\sigma$SA-ES (red points) and $(\lambda)_{\text{opt}}$-CSA-ES (blue points) for $N = 4000$, $\mu = 4$, $\lambda = 10$, $\mathbf{y}^{(0)} = \mathbf{10}$, $\sigma^{(0)} = 1$
- $\alpha_{opt}$: **the non-noisy case formula**



## Summary

- The $(\lambda)_{\text{opt}}$-$\sigma$SA-ES – a new ES combining weighted multirecombination and $\sigma$-self-adaptation
- The $(\lambda)_{\text{opt}}$-$\sigma$SA-ES outperforms the $(\lambda)_{\text{opt}}$-CSA-ES in the non-noisy case
- Theoretical formulas derived in the limit $N \to \infty$ predicted this result
- Main theoretical outcome: $\alpha_{opt}$ formula allows to calculate the optimal learning parameter $\Rightarrow$ maximum performance of the $(\lambda)_{\text{opt}}$-$\sigma$SA-ES in the non-noisy case
- Future work – more complex test functions (PDQFs, general quadratic models)

# Summary

- In the presence of noise:
  - The choice of scaled weights contributes positively to the $(\lambda)_{\text{opt}}$-ES's robustness
  - The $(\lambda)_{\text{opt}}$-$\sigma$SA-ES using numerically calculated $\alpha_{opt}$ values outperforms the $(\lambda)_{\text{opt}}$-CSA-ES
  - An approximate analytic formula for $\alpha_{opt}$ was obtained, but it depends on information about $s^*_\varepsilon$
  - The non-noisy case $\alpha_{opt}$ can be used in practice, but the obtained performance is not optimal for $\kappa > 1$

- Future work:
  - analytic formula for $\alpha_{opt}$ which does not depend on $s^*_\varepsilon$
  - investigation of the constant non-normalized noise strength case

Thank you
for your attention!

# References


D. V. Arnold.
Weighted multirecombination evolution strategies.
*Theoretical computer science*, 361(1):18–37, 2006.

S. Meyer-Nieberg.
*Self-Adaptation in Evolution Strategies*.
PhD thesis, University of Dortmund, CS Department, Dortmund, Germany, 2007.

S. Meyer-Nieberg and H.-G. Beyer.
On the Analysis of Self-Adaptive Recombination Strategies: First Results.
In *Proceedings of the CEC'05 Conference*, pages 2341–2348, Piscataway, NJ, 2005. IEEE.