Additive drift is all you need if you are an evolution strategy

FOGA

August 2025

Tobias Glasmachers
Ruhr-University Bochum

Additive drift is all you need if you are an evolution strategy

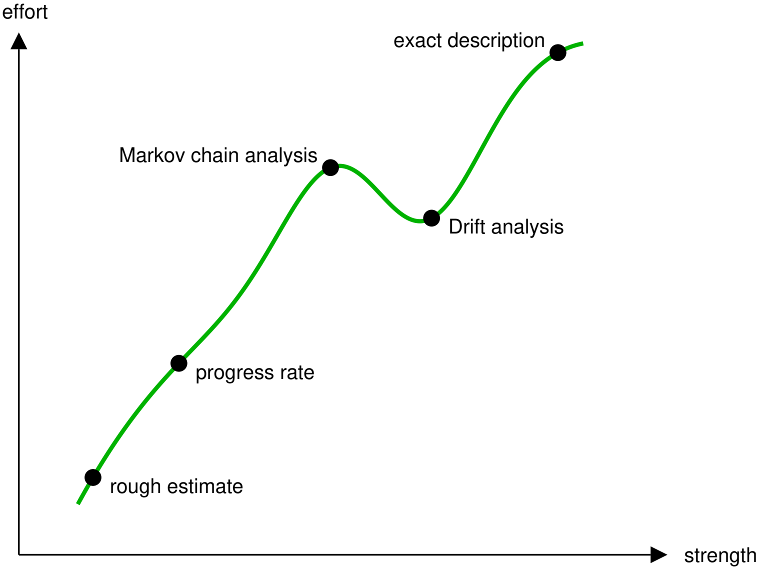
FOGA

August 2025

including two images
generated with AI!

Tobias Glasmachers
Ruhr-University Bochum

Runtime Analysis



Outline

- Evolution Strategies
- Analysis: Methods and Results
- The case of CMA-ES
- Construction of the Potential

Evolution Strategies in a Nutshell

position $m \in \mathbb{R}^n$
scale / step size $\sigma > 0$
shape $C \in \mathbb{R}^{n \times n}$
paths p_c, p_s

Evolution Strategies in a Nutshell

position $m \in \mathbb{R}^n$
scale / step size $\sigma > 0$
shape $C \in \mathbb{R}^{n \times n}$
paths p_c, p_s
repeat
    sample offspring $x_1, \dots, x_\lambda \sim \mathcal{N}(m, \sigma^2 C)$
    rank by fitness

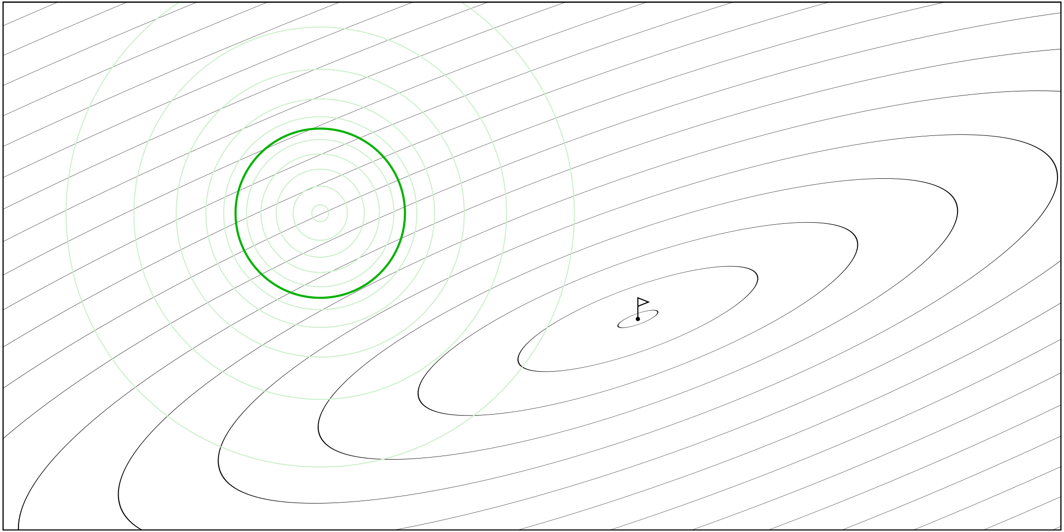
Evolution Strategies in a Nutshell

position $m \in \mathbb{R}^n$
scale / step size $\sigma > 0$
shape $C \in \mathbb{R}^{n \times n}$
paths p_c, p_s
repeat
 sample offspring $x_1, \dots, x_\lambda \sim \mathcal{N}(m, \sigma^2 C)$
 rank by fitness
 update evolution paths

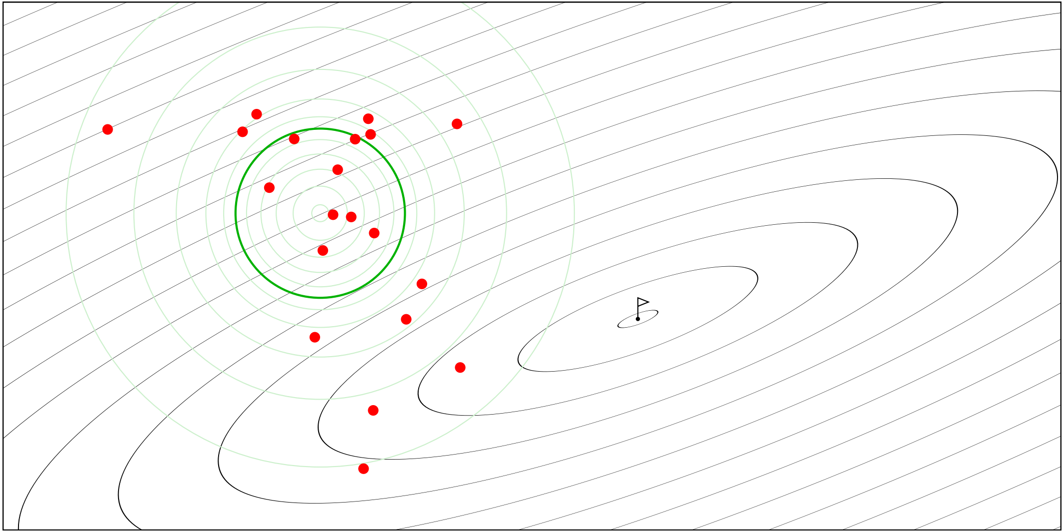
Evolution Strategies in a Nutshell

position $m \in \mathbb{R}^n$
scale / step size $\sigma > 0$
shape $C \in \mathbb{R}^{n \times n}$
paths p_c, p_s
repeat
    sample offspring $x_1, \dots, x_\lambda \sim \mathcal{N}(m, \sigma^2 C)$
    rank by fitness
    update evolution paths
    update mean m
    update step size σ
    (update covariance matrix C)
until stopping criterion is met

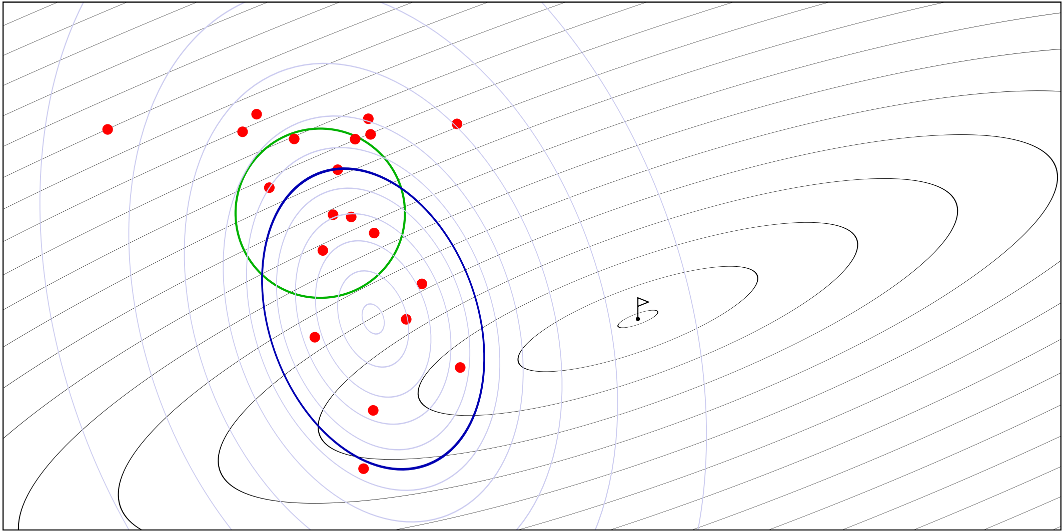
Behavior and Invariance



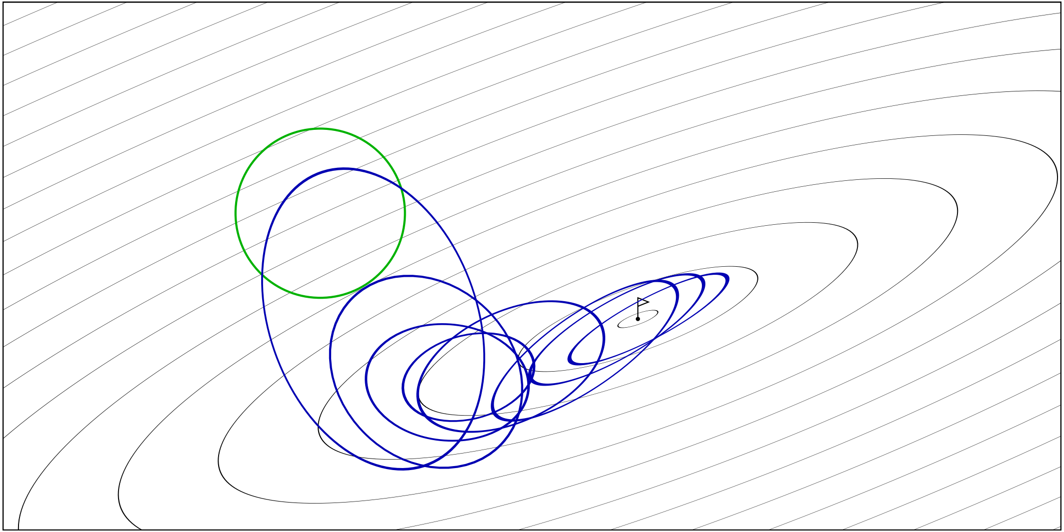
Behavior and Invariance



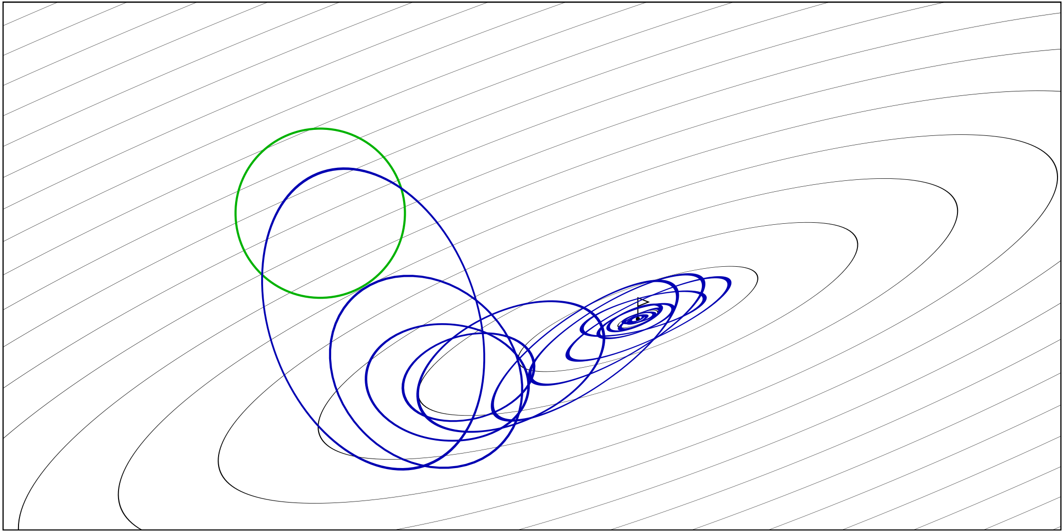
Behavior and Invariance



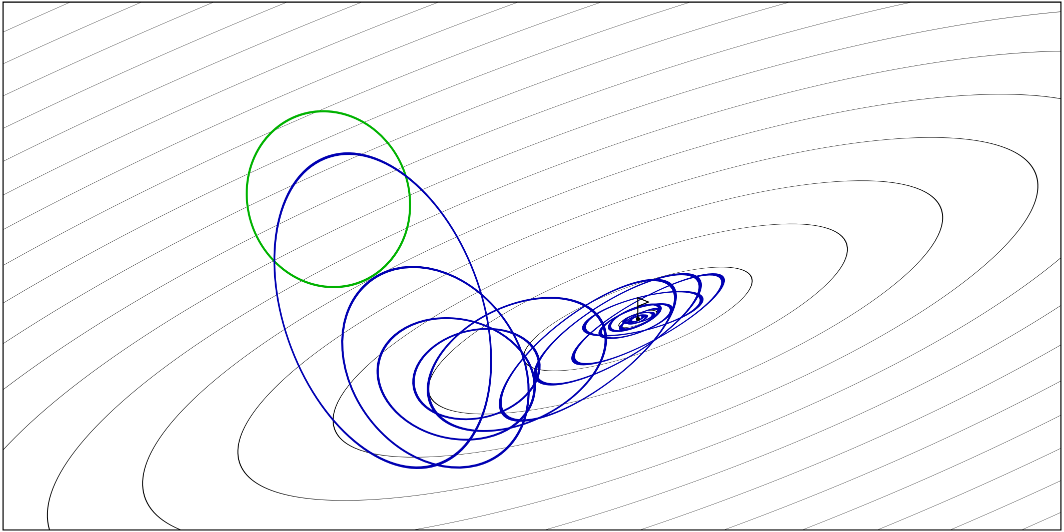
Behavior and Invariance



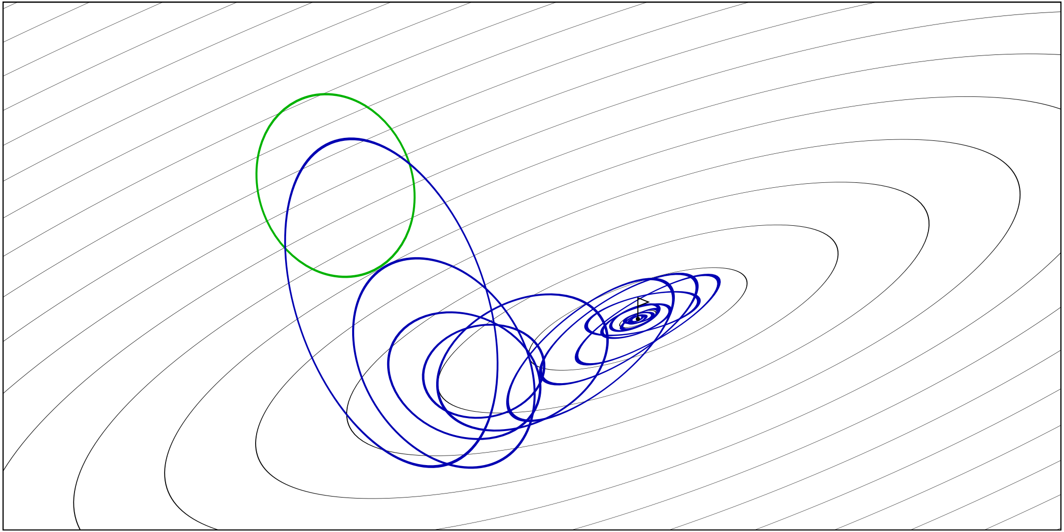
Behavior and Invariance



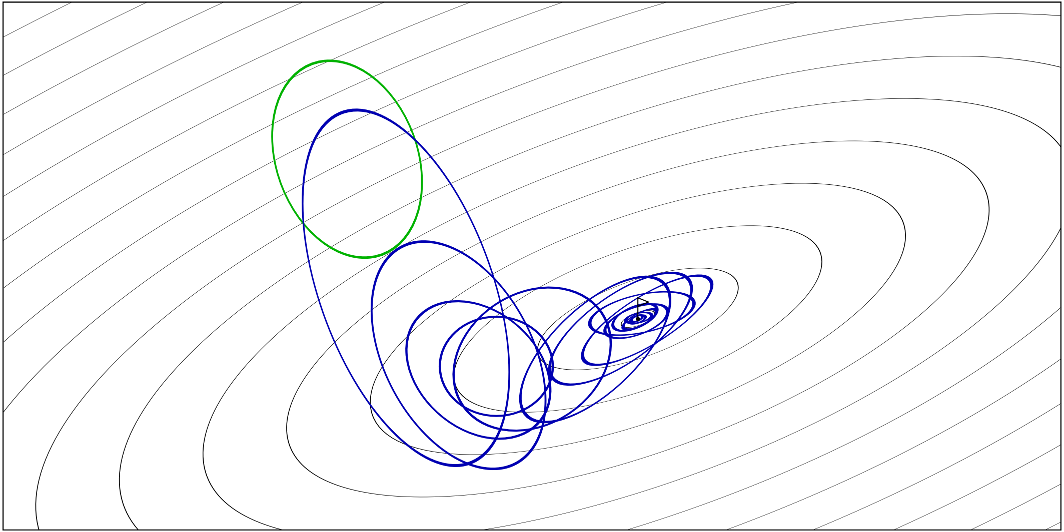
Behavior and Invariance



Behavior and Invariance

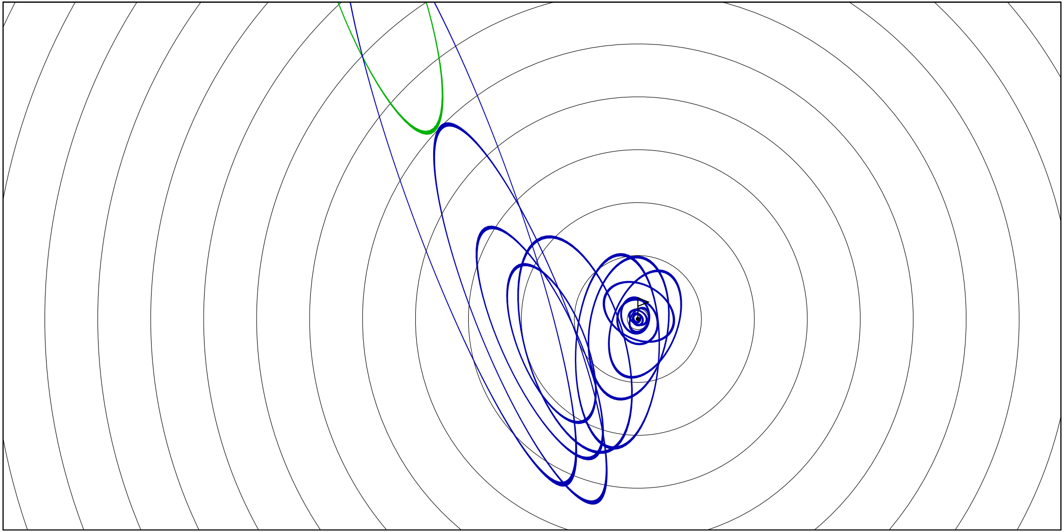


Behavior and Invariance

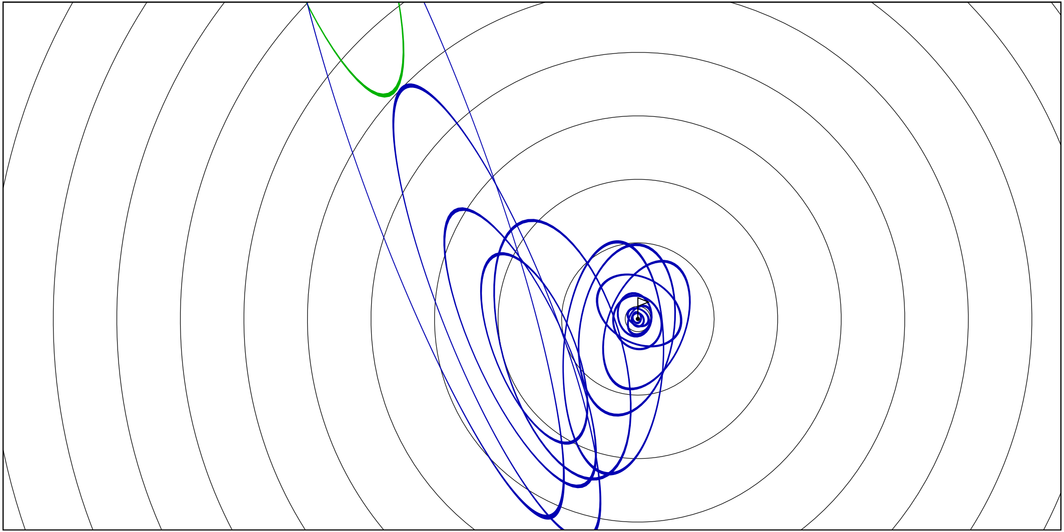


Behavior and Invariance

Behavior and Invariance

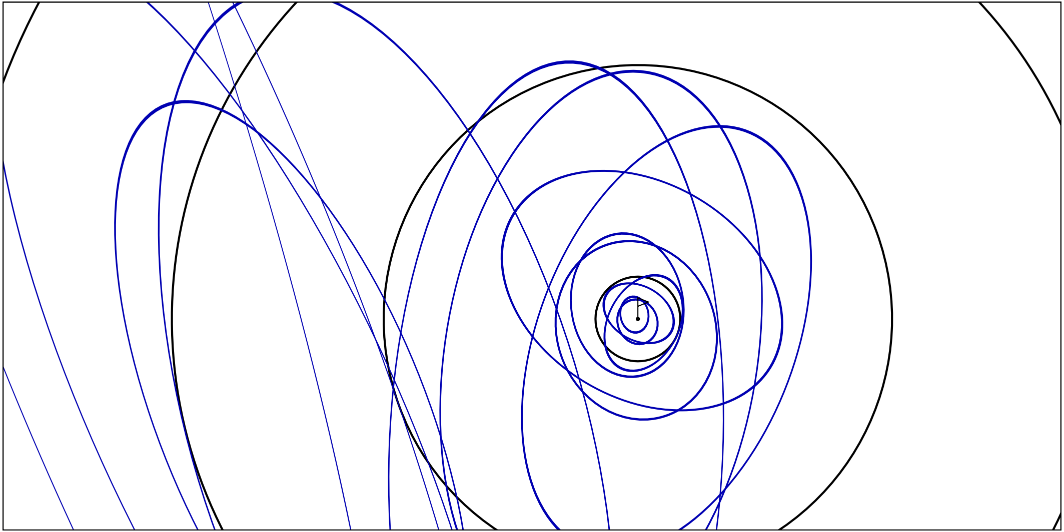


Behavior and Invariance

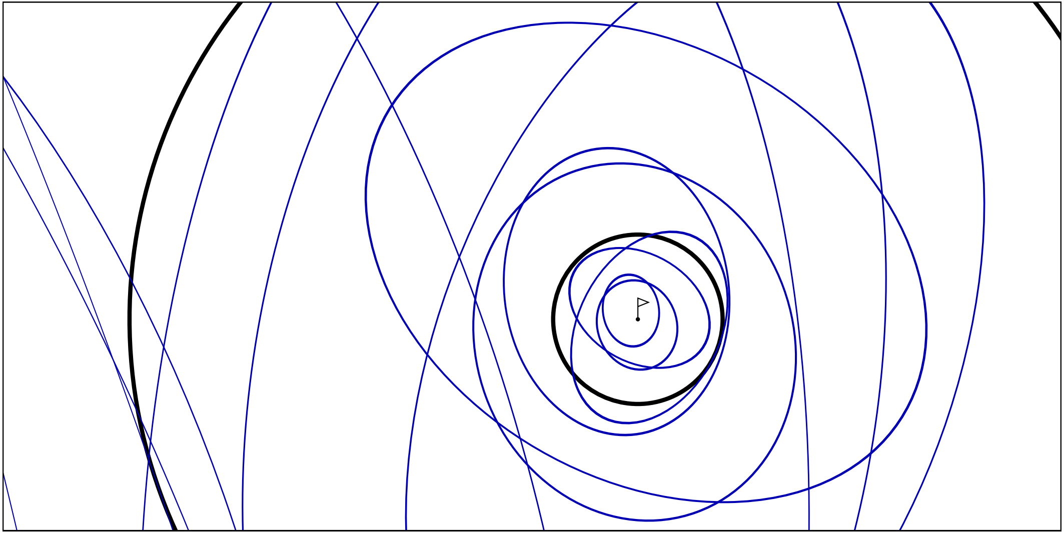


Behavior and Invariance

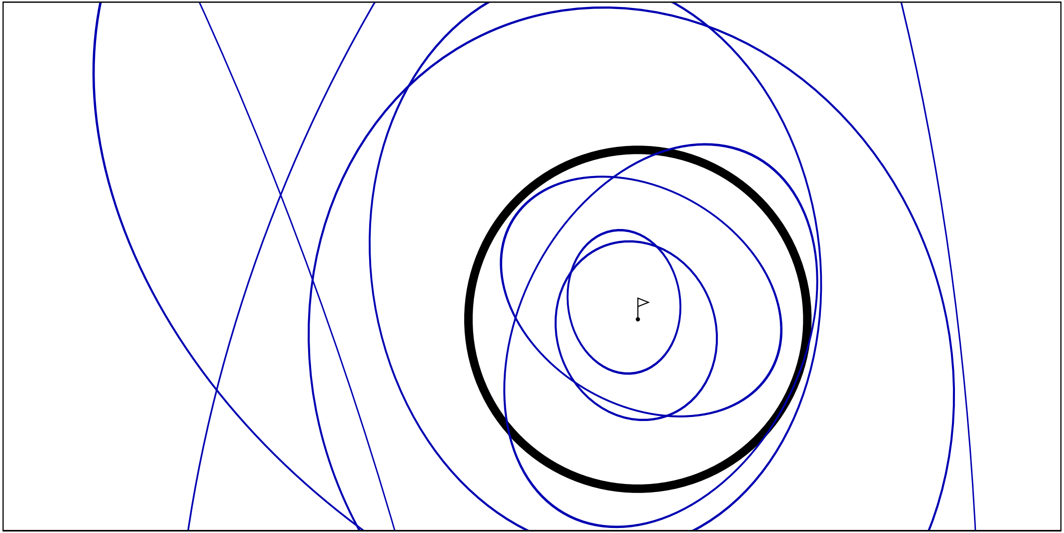
Behavior and Invariance



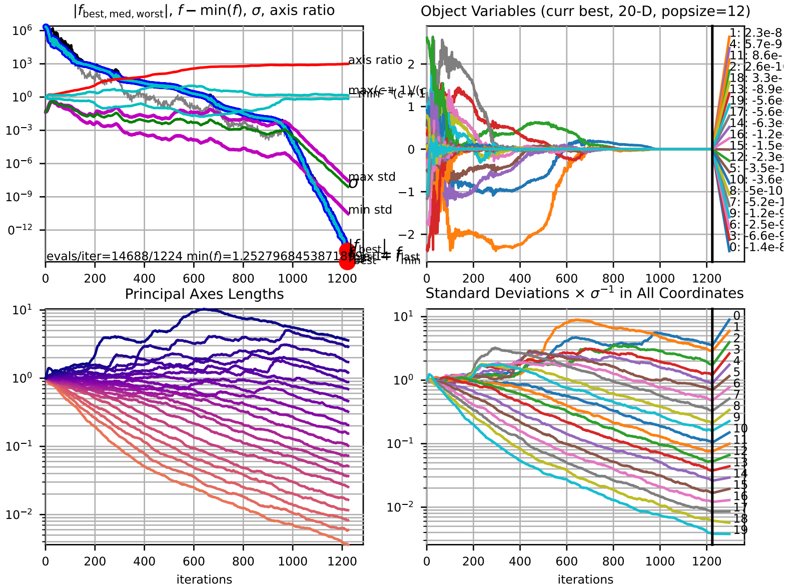
Behavior and Invariance



Behavior and Invariance



An Actual Run



Evolution Strategies in a Nutshell

- Local optima of a $\mathcal{C}^2$ function are approximately convex quadratic.

Evolution Strategies in a Nutshell

- Local optima of a $\mathcal{C}^2$ function are approximately convex quadratic.
- Therefore, analyzing the Sphere function is very meaningful.

Evolution Strategies in a Nutshell

- Local optima of a $\mathcal{C}^2$ function are approximately convex quadratic.
- Therefore, analyzing the Sphere function is very meaningful.
- However, there is more than that!

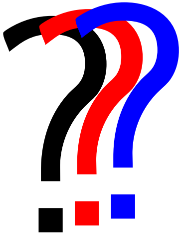
Evolution Strategies in a Nutshell

- Local optima of a $\mathcal{C}^2$ function are approximately convex quadratic.
- Therefore, analyzing the Sphere function is very meaningful.
- However, there is more than that!
- Ignored: multi-modality, asymmetry, plateaus, noise, ...

Evolution Strategies in a Nutshell

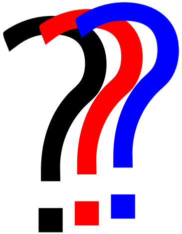
- Local optima of a $\mathcal{C}^2$ function are approximately convex quadratic.
- Therefore, analyzing the Sphere function is very meaningful.
- However, there is more than that!
- Ignored: multi-modality, asymmetry, plateaus, noise, ...
- First things first: this talk is mainly about the Sphere function.

Questions to our Runtime Analysis



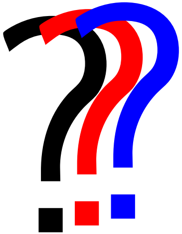
Questions to our Runtime Analysis

- scaling with problem dimension



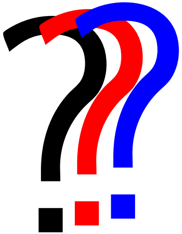
Questions to our Runtime Analysis

- scaling with problem dimension
- dependence on problem instance



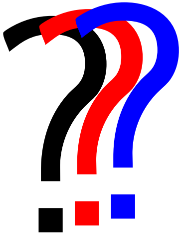
Questions to our Runtime Analysis

- scaling with problem dimension
- dependence on problem instance
- convergence rate



Questions to our Runtime Analysis

- scaling with problem dimension
- dependence on problem instance
- convergence rate
- duration of transient adaptation phase



Markov Chains

- Auger, Anne. "Convergence results for the $(1, \lambda)$-SA-ES using the theory of ϕ-irreducible Markov chains." Theoretical Computer Science 334.1-3 (2005): 35-69.

Markov Chains

- Auger, Anne. "Convergence results for the $(1, \lambda)$-SA-ES using the theory of ϕ-irreducible Markov chains." Theoretical Computer Science 334.1-3 (2005): 35-69.
- Chotard, Alexandre, and Anne Auger. "Verifiable conditions for the irreducibility and aperiodicity of Markov chains by analyzing underlying deterministic models." Bernoulli 25(1) (2019): 112-147.

Markov Chains

- Auger, Anne. “Convergence results for the $(1, \lambda)$-SA-ES using the theory of ϕ-irreducible Markov chains.” Theoretical Computer Science 334.1-3 (2005): 35-69.
- Chotard, Alexandre, and Anne Auger. “Verifiable conditions for the irreducibility and aperiodicity of Markov chains by analyzing underlying deterministic models.” Bernoulli 25(1) (2019): 112-147.
- Gissler, Armand. Linear convergence of evolution strategies with covariance matrix adaptation. Diss. École polytechnique, 2024.

The Normalized Chain

- Convergence on the Sphere is a scale-invariant process: $\sigma \propto \|m\|$.

The Normalized Chain

- Convergence on the Sphere is a scale-invariant process: $\sigma \propto \|m\|$.
- Problem decomposition:
  - behavior of $\|m\|$ $\longrightarrow$ optimization progress
  - behavior of $\theta = m/\sigma$ $\longrightarrow$ stability

The Normalized Chain

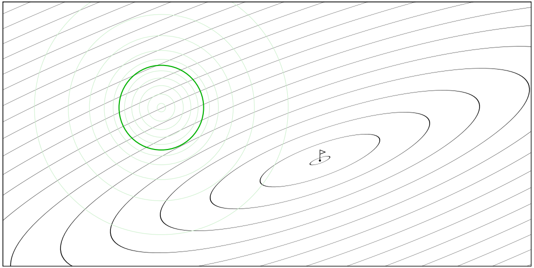
- Convergence on the Sphere is a scale-invariant process: $\sigma \propto \|m\|$.
- Problem decomposition:
  - behavior of $\|m\|$ $\longrightarrow$ optimization progress
  - behavior of $\theta = m/\sigma$ $\longrightarrow$ stability
- Normalized state with CMA:

$$\theta = \left(\frac{m}{\sqrt{\lambda_{\min}(\sqrt{H} C \sqrt{H})} \cdot \sigma}, \frac{\sqrt{H} C \sqrt{H}}{\lambda_{\min}(\sqrt{H} C \sqrt{H})} \right)$$

The Normalized Chain

- Convergence on the Sphere is a scale-invariant process: $\sigma \propto \|m\|$.
- Problem decomposition:
  - behavior of $\|m\|$ $\longrightarrow$ optimization progress
  - behavior of $\theta = m/\sigma$ $\longrightarrow$ stability
- Normalized state with CMA:

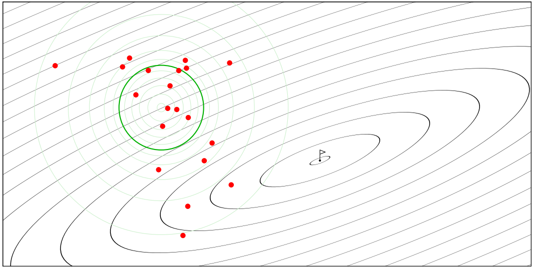
$$\theta = \left(\frac{m}{\sqrt{\lambda_{\min}(\sqrt{H} C \sqrt{H})} \cdot \sigma}, \frac{\sqrt{H} C \sqrt{H}}{\lambda_{\min}(\sqrt{H} C \sqrt{H})} \right)$$



The Normalized Chain

- Convergence on the Sphere is a scale-invariant process: $\sigma \propto \|m\|$.
- Problem decomposition:
  - behavior of $\|m\|$ $\longrightarrow$ optimization progress
  - behavior of $\theta = m/\sigma$ $\longrightarrow$ stability
- Normalized state with CMA:

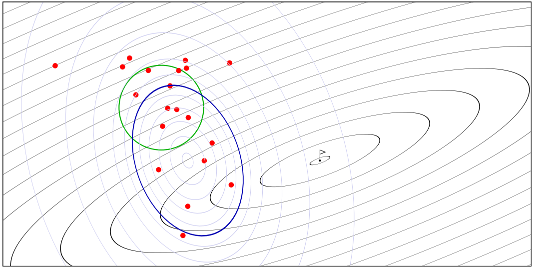
$$\theta = \left(\frac{m}{\sqrt{\lambda_{\min}(\sqrt{H}C\sqrt{H})} \cdot \sigma}, \frac{\sqrt{H}C\sqrt{H}}{\lambda_{\min}(\sqrt{H}C\sqrt{H})} \right)$$



The Normalized Chain

- Convergence on the Sphere is a scale-invariant process: $\sigma \propto \|m\|$.
- Problem decomposition:
  - behavior of $\|m\|$ $\longrightarrow$ optimization progress
  - behavior of $\theta = m/\sigma$ $\longrightarrow$ stability
- Normalized state with CMA:

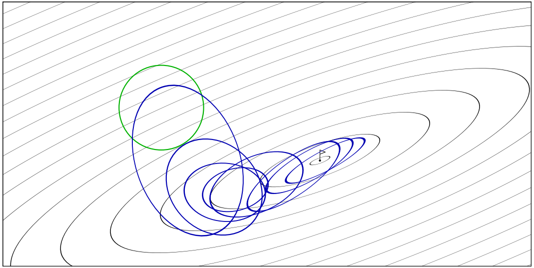
$$\theta = \left(\frac{m}{\sqrt{\lambda_{\min}(\sqrt{H}C\sqrt{H}) \cdot \sigma}}, \frac{\sqrt{H}C\sqrt{H}}{\lambda_{\min}(\sqrt{H}C\sqrt{H})} \right)$$



The Normalized Chain

- Convergence on the Sphere is a scale-invariant process: $\sigma \propto \|m\|$.
- Problem decomposition:
  - behavior of $\|m\|$ $\longrightarrow$ optimization progress
  - behavior of $\theta = m/\sigma$ $\longrightarrow$ stability
- Normalized state with CMA:

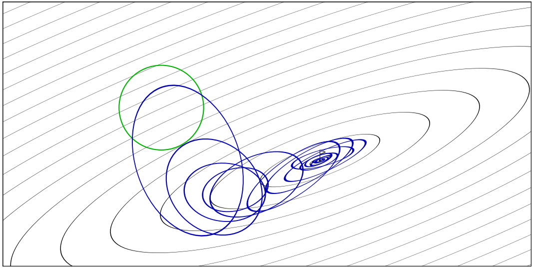
$$\theta = \left(\frac{m}{\sqrt{\lambda_{\min}(\sqrt{H}C\sqrt{H})} \cdot \sigma}, \frac{\sqrt{H}C\sqrt{H}}{\lambda_{\min}(\sqrt{H}C\sqrt{H})} \right)$$



The Normalized Chain

- Convergence on the Sphere is a scale-invariant process: $\sigma \propto \|m\|$.
- Problem decomposition:
  - behavior of $\|m\|$ $\longrightarrow$ optimization progress
  - behavior of $\theta = m/\sigma$ $\longrightarrow$ stability
- Normalized state with CMA:

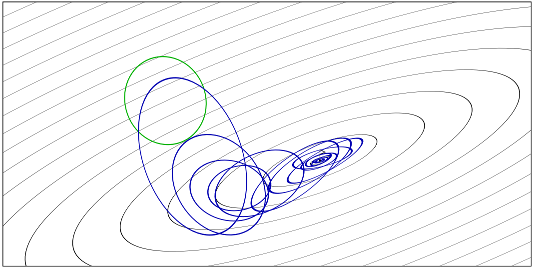
$$\theta = \left(\frac{m}{\sqrt{\lambda_{\min}(\sqrt{H} C \sqrt{H})} \cdot \sigma}, \frac{\sqrt{H} C \sqrt{H}}{\lambda_{\min}(\sqrt{H} C \sqrt{H})} \right)$$



The Normalized Chain

- Convergence on the Sphere is a scale-invariant process: $\sigma \propto \|m\|$.
- Problem decomposition:
  - behavior of $\|m\|$ $\longrightarrow$ optimization progress
  - behavior of $\theta = m/\sigma$ $\longrightarrow$ stability
- Normalized state with CMA:

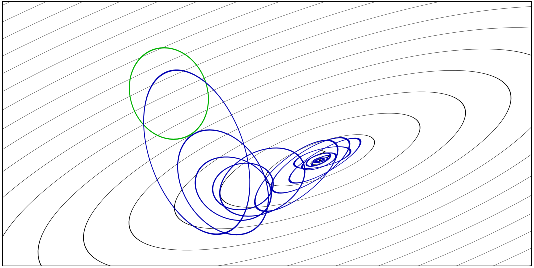
$$\theta = \left(\frac{m}{\sqrt{\lambda_{\min}(\sqrt{H} C \sqrt{H})} \cdot \sigma}, \frac{\sqrt{H} C \sqrt{H}}{\lambda_{\min}(\sqrt{H} C \sqrt{H})} \right)$$



The Normalized Chain

- Convergence on the Sphere is a scale-invariant process: $\sigma \propto \|m\|$.
- Problem decomposition:
  - behavior of $\|m\|$ $\longrightarrow$ optimization progress
  - behavior of $\theta = m/\sigma$ $\longrightarrow$ stability
- Normalized state with CMA:

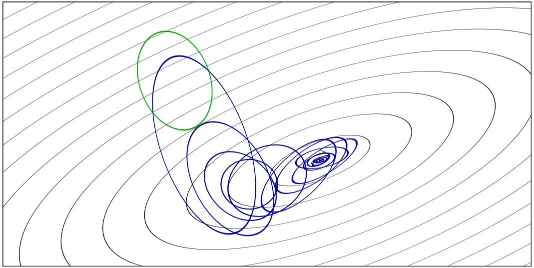
$$\theta = \left(\frac{m}{\sqrt{\lambda_{\min}(\sqrt{H}C\sqrt{H})} \cdot \sigma}, \frac{\sqrt{H}C\sqrt{H}}{\lambda_{\min}(\sqrt{H}C\sqrt{H})} \right)$$



The Normalized Chain

- Convergence on the Sphere is a scale-invariant process: $\sigma \propto \|m\|$.
- Problem decomposition:
  - behavior of $\|m\|$ $\longrightarrow$ optimization progress
  - behavior of $\theta = m/\sigma$ $\longrightarrow$ stability
- Normalized state with CMA:

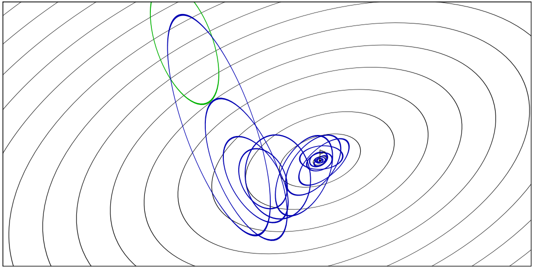
$$\theta = \left(\frac{m}{\sqrt{\lambda_{\min}(\sqrt{H}C\sqrt{H})}\cdot\sigma}, \frac{\sqrt{H}C\sqrt{H}}{\lambda_{\min}(\sqrt{H}C\sqrt{H})} \right)$$



The Normalized Chain

- Convergence on the Sphere is a scale-invariant process: $\sigma \propto \|m\|$.
- Problem decomposition:
  - behavior of $\|m\|$ $\longrightarrow$ optimization progress
  - behavior of $\theta = m/\sigma$ $\longrightarrow$ stability
- Normalized state with CMA:

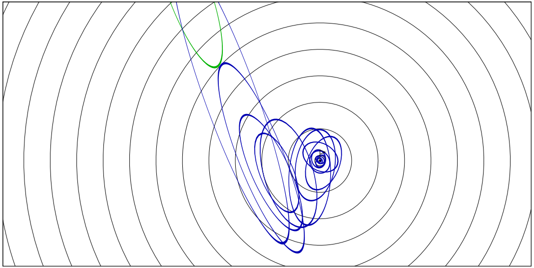
$$\theta = \left(\frac{m}{\sqrt{\lambda_{\min}(\sqrt{H} C \sqrt{H})} \cdot \sigma}, \frac{\sqrt{H} C \sqrt{H}}{\lambda_{\min}(\sqrt{H} C \sqrt{H})} \right)$$



The Normalized Chain

- Convergence on the Sphere is a scale-invariant process: $\sigma \propto \|m\|$.
- Problem decomposition:
  - behavior of $\|m\|$ $\longrightarrow$ optimization progress
  - behavior of $\theta = m/\sigma$ $\longrightarrow$ stability
- Normalized state with CMA:

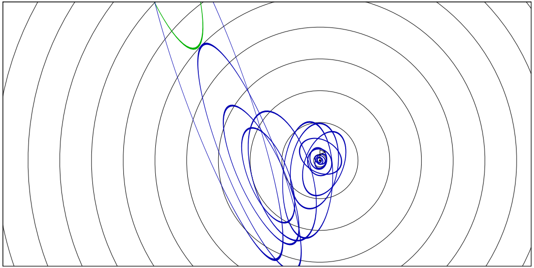
$$\theta = \left(\frac{m}{\sqrt{\lambda_{\text{min}}(\sqrt{H} C \sqrt{H})} \cdot \sigma}, \frac{\sqrt{H} C \sqrt{H}}{\lambda_{\text{min}}(\sqrt{H} C \sqrt{H})} \right)$$



The Normalized Chain

- Convergence on the Sphere is a scale-invariant process: $\sigma \propto \|m\|$.
- Problem decomposition:
  - behavior of $\|m\|$ $\longrightarrow$ optimization progress
  - behavior of $\theta = m/\sigma$ $\longrightarrow$ stability
- Normalized state with CMA:

$$\theta = \left(\frac{m}{\sqrt{\lambda_{\text{min}}(\sqrt{H}C\sqrt{H})} \cdot \sigma}, \frac{\sqrt{H}C\sqrt{H}}{\lambda_{\text{min}}(\sqrt{H}C\sqrt{H})} \right)$$



The Normalized Chain

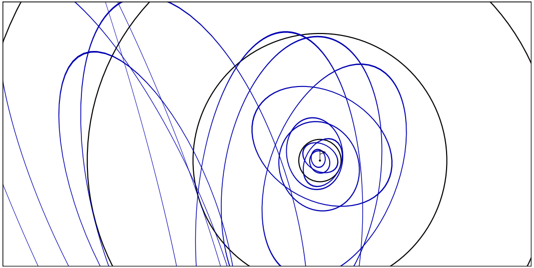
- Convergence on the Sphere is a scale-invariant process: $\sigma \propto \|m\|$.
- Problem decomposition:
  - behavior of $\|m\|$ $\longrightarrow$ optimization progress
  - behavior of $\theta = m/\sigma$ $\longrightarrow$ stability
- Normalized state with CMA:

$$\theta = \left(\frac{m}{\sqrt{\lambda_{\min}(\sqrt{H} C \sqrt{H})} \cdot \sigma}, \frac{\sqrt{H} C \sqrt{H}}{\lambda_{\min}(\sqrt{H} C \sqrt{H})} \right)$$

The Normalized Chain

- Convergence on the Sphere is a scale-invariant process: $\sigma \propto \|m\|$.
- Problem decomposition:
  - behavior of $\|m\|$ $\longrightarrow$ optimization progress
  - behavior of $\theta = m/\sigma$ $\longrightarrow$ stability
- Normalized state with CMA:

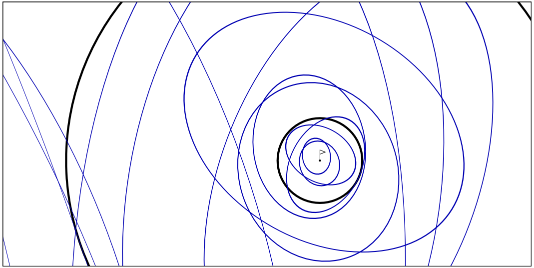
$$\theta = \left(\frac{m}{\sqrt{\lambda_{\min}(\sqrt{H} C \sqrt{H})} \cdot \sigma}, \frac{\sqrt{H} C \sqrt{H}}{\lambda_{\min}(\sqrt{H} C \sqrt{H})} \right)$$



The Normalized Chain

- Convergence on the Sphere is a scale-invariant process: $\sigma \propto \|m\|$.
- Problem decomposition:
  - behavior of $\|m\|$ $\longrightarrow$ optimization progress
  - behavior of $\theta = m/\sigma$ $\longrightarrow$ stability
- Normalized state with CMA:

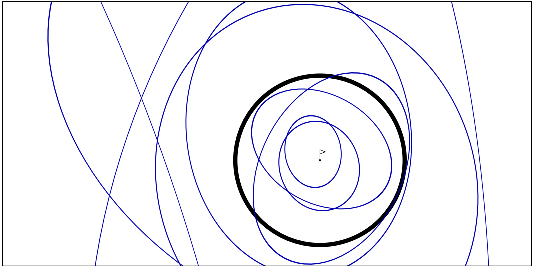
$$\theta = \left(\frac{m}{\sqrt{\lambda_{\min}(\sqrt{H} C \sqrt{H}) \cdot \sigma}}, \frac{\sqrt{H} C \sqrt{H}}{\lambda_{\min}(\sqrt{H} C \sqrt{H})} \right)$$



The Normalized Chain

- Convergence on the Sphere is a scale-invariant process: $\sigma \propto \|m\|$.
- Problem decomposition:
  - behavior of $\|m\|$ $\longrightarrow$ optimization progress
  - behavior of $\theta = m/\sigma$ $\longrightarrow$ stability
- Normalized state with CMA:

$$\theta = \left(\frac{m}{\sqrt{\lambda_{\min}(\sqrt{H}C\sqrt{H})} \cdot \sigma}, \frac{\sqrt{H}C\sqrt{H}}{\lambda_{\min}(\sqrt{H}C\sqrt{H})} \right)$$



Proof Logic

1. Stability $\Leftrightarrow$ θ follows a stable limit distribution.

Proof Logic

1. Stability $\Leftrightarrow$ θ follows a stable limit distribution.

   Task: prove irreducibility, aperidocity, (Harris-)recurrence, geometric ergodicity.

Proof Logic

1. Stability $\Leftrightarrow$ θ follows a stable limit distribution.

   Task: prove irreducibility, aperidocity, (Harris-)recurrence, geometric ergodicity.

2. For fixed θ, the change of $\|m\|$ is scale-invariant, resulting in an additive change of $\log(\|m\|)$. It exhibits either linear convergence or divergence.

Proof Logic

1. Stability $\Leftrightarrow$ θ follows a stable limit distribution.

   Task: prove irreducibility, aperidocity, (Harris-)recurrence, geometric ergodicity.

2. For fixed θ, the change of $\|m\|$ is scale-invariant, resulting in an additive change of $\log(\|m\|)$. It exhibits either linear convergence or divergence.

   The same holds for the expectation over θ.

Achievements

- Linear convergence of CMA-ES at a rate that is independent of the problem instance (independent of the Hessian H).

Achievements

- Linear convergence of CMA-ES at a rate that is independent of the problem instance (independent of the Hessian H).

  CMA helps, it achieves what it was designed for.

Achievements

- Linear convergence of CMA-ES at a rate that is independent of the problem instance (independent of the Hessian H).

  CMA helps, it achieves what it was designed for.

- No control over the limit distribution and over

$$\mathbb{E}[\log(\|m^{(t)}\|) - \log(\|m^{(t+1)}\|)].$$

Achievements

- Linear convergence of CMA-ES at a rate that is independent of the problem instance (independent of the Hessian H).

  CMA helps, it achieves what it was designed for.

- No control over the limit distribution and over

$$\mathbb{E}[\log(\|m^{(t)}\|) - \log(\|m^{(t+1)}\|)].$$

  No convergence rate, no dependency on the dimension.

Achievements

- Linear convergence of CMA-ES at a rate that is independent of the problem instance (independent of the Hessian H).

  CMA helps, it achieves what it was designed for.

- No control over the limit distribution and over

$$\mathbb{E}[\log(\|m^{(t)}\|) - \log(\|m^{(t+1)}\|)].$$

  No convergence rate, no dependency on the dimension.

- No control over the time to reach the stable limit distribution.

Drift Analysis

- Jägersküpper, Jens. "How the (1+1)-ES using isotropic mutations minimizes positive definite quadratic forms." Theoretical Computer Science 361.1 (2006): 38-56.

Drift Analysis

- Jägersküpper, Jens. "How the (1+1)-ES using isotropic mutations minimizes positive definite quadratic forms." Theoretical Computer Science 361.1 (2006): 38-56.

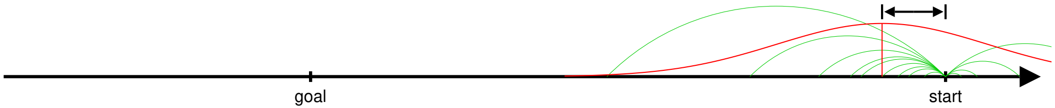
- Akimoto, Youhei, Anne Auger, and Tobias Glasmachers. "Drift theory in continuous search spaces: expected hitting time of the (1+1)-ES with 1/5 success rule." Proceedings of the Genetic and Evolutionary Computation Conference. 2018.

Drift Analysis

- Jägersküpper, Jens. "How the (1+1)-ES using isotropic mutations minimizes positive definite quadratic forms." Theoretical Computer Science 361.1 (2006): 38-56.
- Akimoto, Youhei, Anne Auger, and Tobias Glasmachers. "Drift theory in continuous search spaces: expected hitting time of the (1+1)-ES with 1/5 success rule." Proceedings of the Genetic and Evolutionary Computation Conference. 2018.
- Morinaga, Daiki, and Youhei Akimoto. "Generalized drift analysis in continuous domain: linear convergence of (1+1)-ES on strongly convex functions with Lipschitz continuous gradients." Proceedings of the 15th ACM/SIGEVO Conference on Foundations of Genetic Algorithms. 2019.

Drift Analysis

- Jägersküpper, Jens. "How the (1+1)-ES using isotropic mutations minimizes positive definite quadratic forms." Theoretical Computer Science 361.1 (2006): 38-56.
- Akimoto, Youhei, Anne Auger, and Tobias Glasmachers. "Drift theory in continuous search spaces: expected hitting time of the (1+1)-ES with 1/5 success rule." Proceedings of the Genetic and Evolutionary Computation Conference. 2018.
- Morinaga, Daiki, and Youhei Akimoto. "Generalized drift analysis in continuous domain: linear convergence of (1+1)-ES on strongly convex functions with Lipschitz continuous gradients." Proceedings of the 15th ACM/SIGEVO Conference on Foundations of Genetic Algorithms. 2019.
- Morinaga, Daiki, et al. "Convergence rate of the (1+1)-evolution strategy with success-based step-size adaptation on convex quadratic functions." Proceedings of the Genetic and Evolutionary Computation Conference. 2021.



Theorem (Additive Drift, Upper Bound with Overshooting; Doerr and Kötzing 2021)

Let $(X^{(t)})_{t \in \mathbb{N}}$ *be an integrable process over* $\mathbb{R}$*, and let*

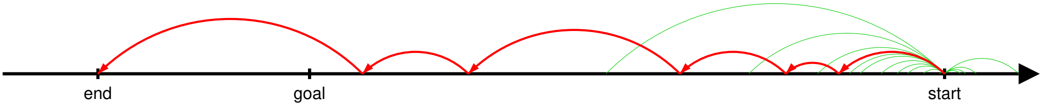
$$T = \inf\left\{t \in \mathbb{N} \,\middle|\, X^{(t)} \leq \beta\right\}$$

for some $\beta \in \mathbb{R}$*. Furthermore, there is a* $\delta > 0$ *such that, for all* $t < T$*, it holds that*

$$0 < \delta \leq \mathbb{E}\left[X^{(t)} - X^{(t+1)} \,\middle|\, X^{(0)}, \dots, X^{(t)}\right].$$

Then

$$\mathbb{E}[T] \leq \frac{\mathbb{E}[X^{(0)}] - \mathbb{E}[X^{(T)}]}{\delta}.$$



Theorem (Additive Drift, Upper Bound with Overshooting; Doerr and Kötzing 2021)

Let $(X^{(t)})_{t \in \mathbb{N}}$ be an integrable process over $\mathbb{R}$, and let

$$T = \inf \left\{ t \in \mathbb{N} \,\middle|\, X^{(t)} \leq \beta \right\}$$

for some $\beta \in \mathbb{R}$. Furthermore, there is a $\delta > 0$ such that, for all $t < T$, it holds that

$$0 < \delta \leq \mathbb{E}\left[X^{(t)} - X^{(t+1)} \,\middle|\, X^{(0)}, \dots, X^{(t)} \right].$$

Then

$$\mathbb{E}[T] \leq \frac{\mathbb{E}[X^{(0)}] - \mathbb{E}[X^{(T)}]}{\delta}.$$

Proof Logic

- Same problem decomposition as for Markov chains.

Proof Logic

- Same problem decomposition as for Markov chains.
- Apply additive drift to a potential function

$$V(\|m\|, \theta) = \log(\|m\|) + \Psi(\theta)$$

with $\Psi \geq 0$.

Proof Logic

- Same problem decomposition as for Markov chains.
- Apply additive drift to a potential function

$$V(\|m\|, \theta) = \log(\|m\|) + \Psi(\theta)$$

  with $\Psi \geq 0$.

- If V decays by $\delta > 0$ in expectation then at some point $\log(\|m\|)$ decays by $\delta > 0$. We obtain linear convergence at rate $e^{-\delta} < 1$.

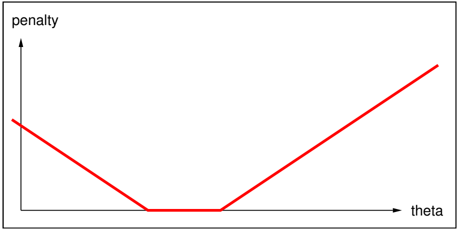
Proof Logic

- Same problem decomposition as for Markov chains.
- Apply additive drift to a potential function

$$V(\|m\|, \theta) = \log(\|m\|) + \Psi(\theta)$$

  with $\Psi \geq 0$.
- If V decays by $\delta > 0$ in expectation then at some point $\log(\|m\|)$ decays by $\delta > 0$. We obtain linear convergence at rate $e^{-\delta} < 1$.
- Note: V is unbounded, and single-iteration changes are unbounded. Calls for additive drift with overshooting.

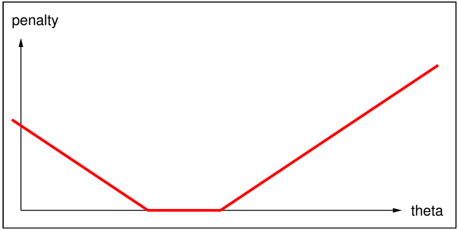
Proof Logic (continued)



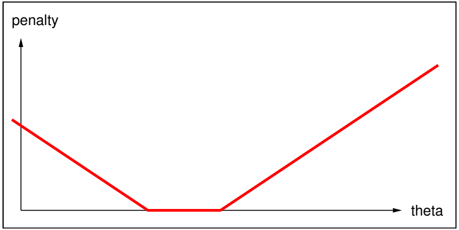
- Task: construct a suitable $\Psi(\theta)$.

Proof Logic (continued)

- Task: construct a suitable $\Psi(\theta)$.
- Task: show that a drift $\delta > 0$ exists.

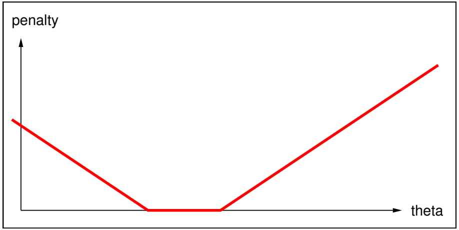


Proof Logic (continued)



- Task: construct a suitable $\Psi(\theta)$.
- Task: show that a drift $\delta > 0$ exists.
- The form of $\Psi(\theta)$ with CMA is an open problem.

Proof Logic (continued)



- Task: construct a suitable $\Psi(\theta)$.
- Task: show that a drift $\delta > 0$ exists.
- The form of $\Psi(\theta)$ with CMA is an open problem.
- The extension to strongly convex functions with Lipschitz continuous gradients does not change the overall proof logic too much.

Achievements

- Correct dependency on dimension and problem characteristics.

Achievements

- Correct dependency on dimension and problem characteristics.
- Convergence rate and duration of the adaptation phase can be bounded.

Achievements

- Correct dependency on dimension and problem characteristics.
- Convergence rate and duration of the adaptation phase can be bounded.
- No analysis of CMA (as of today).


The Quest for the Penalty Term

The Route

- How to approach this problem?
  Pull a penalty term Ψ out of the hat and then...?
  Options:

The Route

- How to approach this problem?
  Pull a penalty term Ψ out of the hat and then...?
  Options:
  1. **Deductive approach**
     Try to make the proof work.

The Route

- How to approach this problem?
  Pull a penalty term Ψ out of the hat and then...?
  Options:
  1. **Deductive approach**
     Try to make the proof work.
  2. **Inductive approach**
     Simulate/approximate the drift to empirically verify that $\delta > 0$ exists.
     Then think about a proof, otherwise refine.

The Route

- How to approach this problem?
  Pull a penalty term Ψ out of the hat and then...?
  Options:
  1. **Deductive approach**
     Try to make the proof work.
  2. **Inductive approach**
     Simulate/approximate the drift to empirically verify that $\delta > 0$ exists.
     Then think about a proof, otherwise refine.
- Which route should we take?

The Route

- How to approach this problem?
  Pull a penalty term Ψ out of the hat and then...?
  Options:
  1. **Deductive approach**
     Try to make the proof work.
  2. **Inductive approach**
     Simulate/approximate the drift to empirically verify that $\delta > 0$ exists.
     Then think about a proof, otherwise refine.
- Which route should we take?

"It depends on how fast you think."

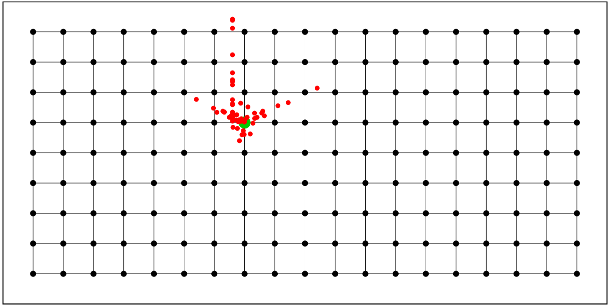
The Route

- How to approach this problem?
  Pull a penalty term Ψ out of the hat and then...?
  Options:
  1. **Deductive approach**
     Try to make the proof work.
  2. **Inductive approach**
     Simulate/approximate the drift to empirically verify that $\delta > 0$ exists.
     Then think about a proof, otherwise refine.
- Which route should we take?

  "It depends on how fast you think."

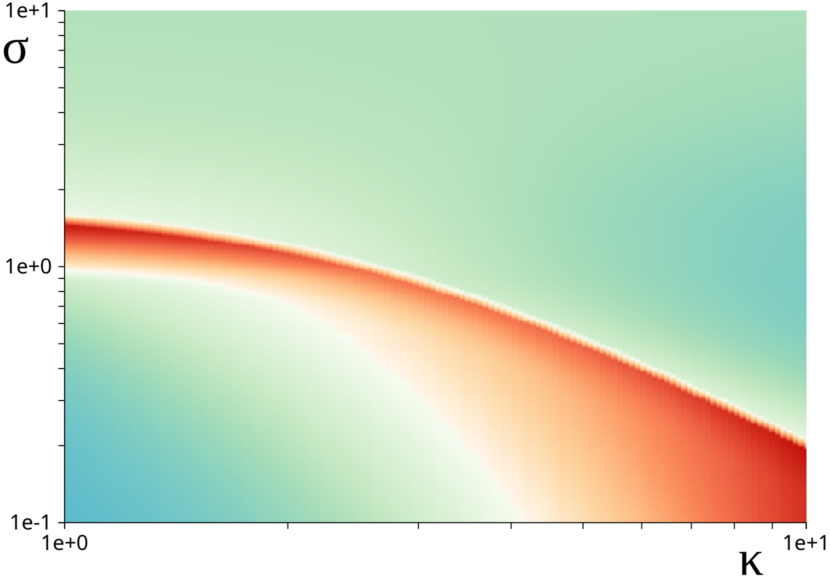
- I decided to take the inductive route.

Grid Interpolation

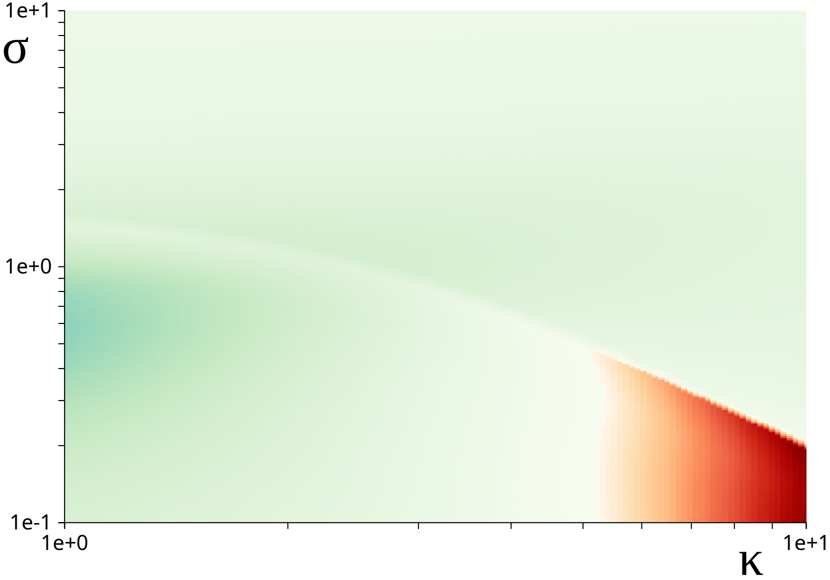


1. guess a potential function
2. estimate drift on a grid in state space (Monte Carlo simulation)
3. visualize the result

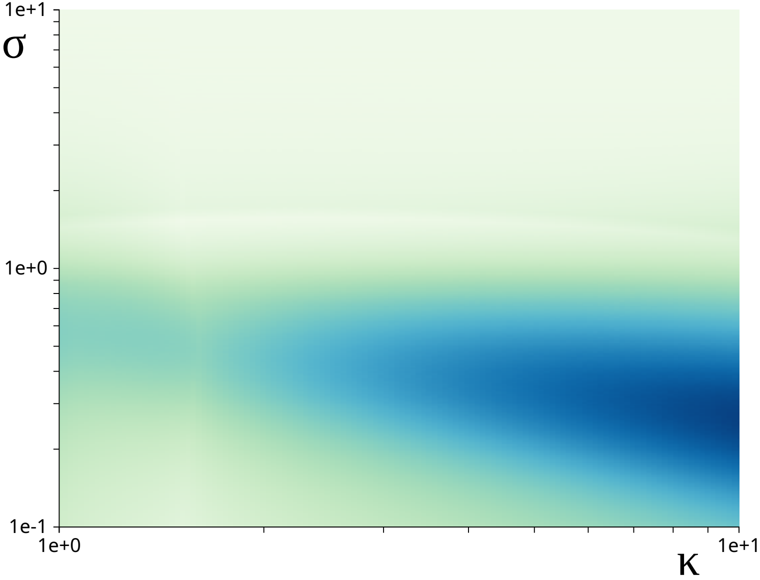
Grid Interpolation



Grid Interpolation



Grid Interpolation



Optimal (Canonical) Drift

- There is a very simple way of realizing optimal additive drift: **define** the potential to be the expected hitting time.

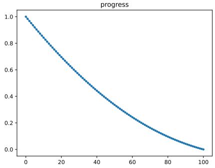
Optimal (Canonical) Drift

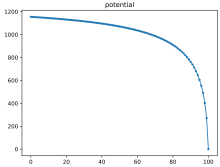
- There is a very simple way of realizing optimal additive drift: **define** the potential to be the expected hitting time.
- The construction yields an optimally tight (lower and upper) bound using only simple additive drift.

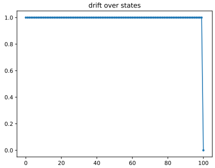
Optimal (Canonical) Drift

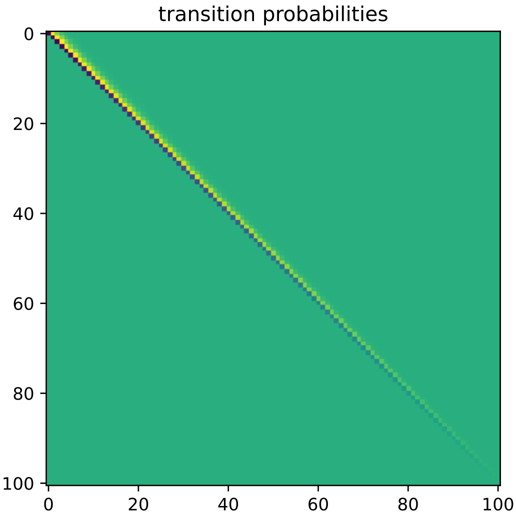
- There is a very simple way of realizing optimal additive drift: **define** the potential to be the expected hitting time.
- The construction yields an optimally tight (lower and upper) bound using only simple additive drift.
- Problem: the potential V is not an explicit function of the state (m, σ, C) — no closed form expression.

(1+1)-EA on OneMax

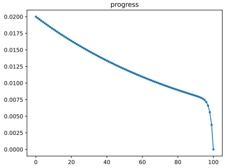


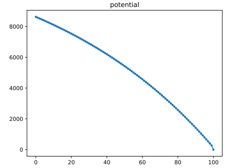


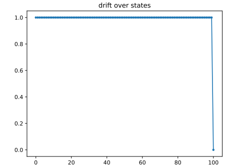


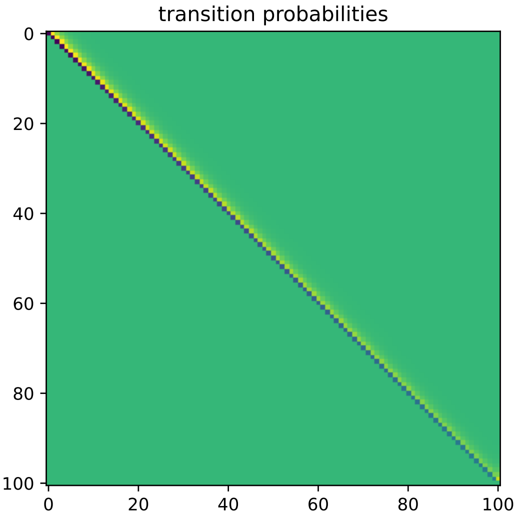


(1+1)-EA on LeadingOnes

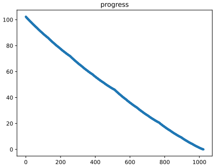


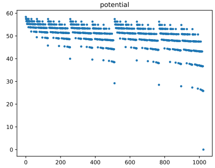


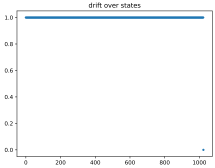


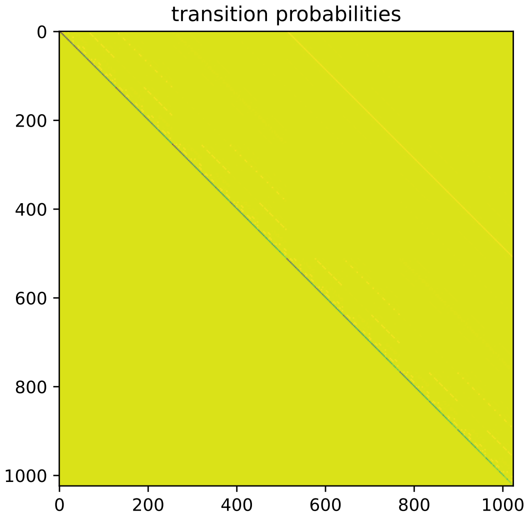


(1+1)-EA on BinaryValue

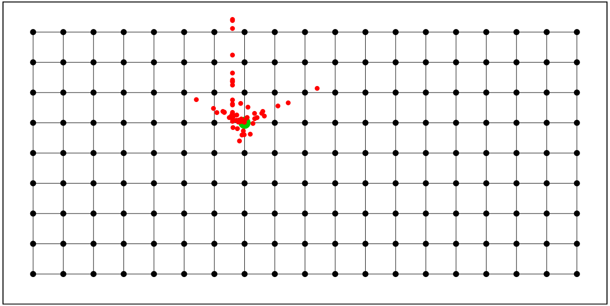






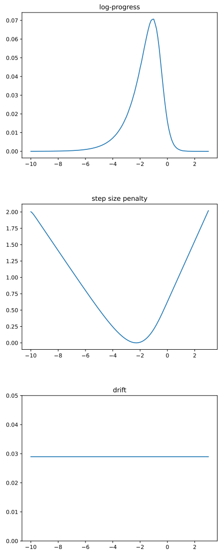


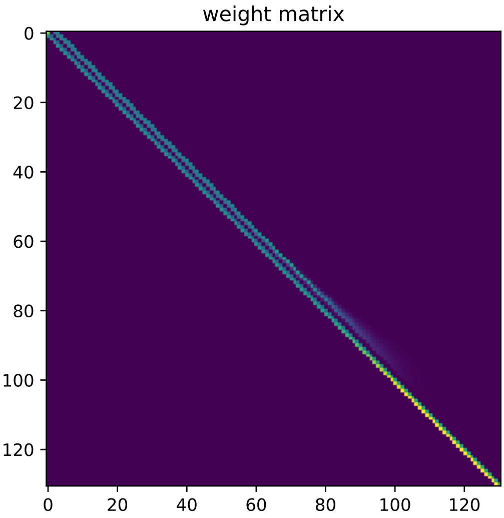
Canonical Drift on the Grid

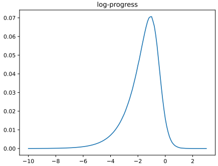


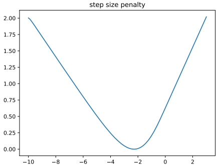
1. Transfer the canonical potential paradigm to the grid approach!
2. Ansatz: Ψ is parameterized by values on the grid + linear interpolation.
3. Using MC simulation data, we can solve the system.

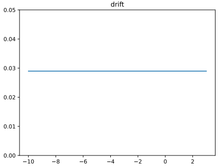
(1+1)-ES on Sphere

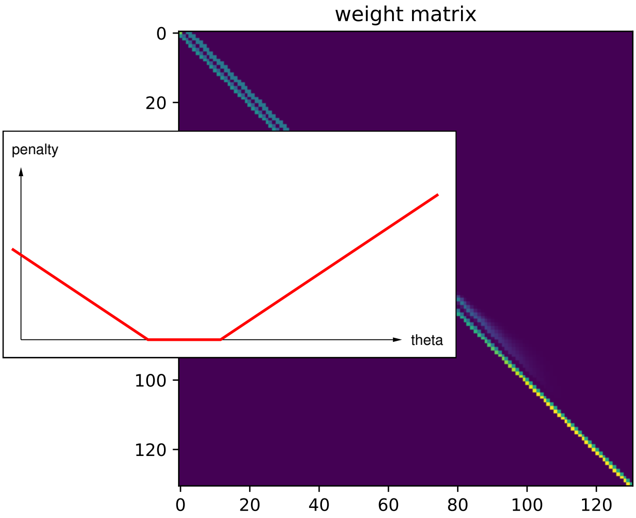





log-progress
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
−10
−8
−6
−4
−2
0
2


step size penalty
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
−10
−8
−6
−4
−2
0
2


drift
0.05
0.04
0.03
0.02
0.01
0.00
−10
−8
−6
−4
−2
0
2


weight matrix
0
20
100
120
0
20
40
60
80
100
120
penalty
theta

Technicalities

- Choice of the grid (range and density) turn out to be uncritical.

Technicalities

- Choice of the grid (range and density) turn out to be uncritical.
- Replaced MC integration with a "proper" numerical scheme with a-priori controlled error.

Technicalities

- Choice of the grid (range and density) turn out to be uncritical.
- Replaced MC integration with a "proper" numerical scheme with a-priori controlled error.
- Floating point errors are currently ignored.

Technicalities

- Choice of the grid (range and density) turn out to be uncritical.
- Replaced MC integration with a "proper" numerical scheme with a-priori controlled error.
- Floating point errors are currently ignored.
- Controlled effect of integration error on the penalty term.

Towards Provable Drift

- We aim to turn the technique into a computer-assisted proof.

Towards Provable Drift

- We aim to turn the technique into a computer-assisted proof.
- Interpolation: Lipschitz constant of the drift controls gaps in between grid points.

Towards Provable Drift

- We aim to turn the technique into a computer-assisted proof.
- Interpolation: Lipschitz constant of the drift controls gaps in between grid points.
- Extrapolation: control asymptotic effects beyond the grid.

Towards Provable Drift

- We aim to turn the technique into a computer-assisted proof.
- Interpolation: Lipschitz constant of the drift controls gaps in between grid points.
- Extrapolation: control asymptotic effects beyond the grid.
- Only possible in low dimensions (likely only $n = 2$).

Achievements

- It is possible to analyze CMA this way.
  A first study (Bachelor thesis) is on its way.

Achievements

- It is possible to analyze CMA this way.
  A first study (Bachelor thesis) is on its way.
- Convergence rate and duration of the adaptation phase can be bounded.

Achievements

- It is possible to analyze CMA this way.
  A first study (Bachelor thesis) is on its way.
- Convergence rate and duration of the adaptation phase can be bounded.
- No (analytic) dependency on parameters (dimension, problem instance).

Achievements

- It is possible to analyze CMA this way.
  A first study (Bachelor thesis) is on its way.
- Convergence rate and duration of the adaptation phase can be bounded.
- No (analytic) dependency on parameters (dimension, problem instance).
- Computationally feasible only in dimension 2.

Conclusions

- The analysis of CMA-ES is progressing significantly.

Conclusions

- The analysis of CMA-ES is progressing significantly.
- Convergence is solved!

Conclusions

- The analysis of CMA-ES is progressing significantly.
- Convergence is solved!
- Drift for CMA-ES would be a valuable addition...

Conclusions

- The analysis of CMA-ES is progressing significantly.
- Convergence is solved!
- Drift for CMA-ES would be a valuable addition...
- ...but it is still an open problem.

Thanks

Stephan Frank

Alexander Jungeilges

Thank you!

Thank you!

Questions?