

How to Assess Step-Size Adaptation Mechanisms in Randomised Search

Nikolaus Hansen*, Asma Atamna*, and Anne Auger*

Motivations

- Assessment of step-size adaptation mechanisms often done on too restrictive scenarios (the sphere function): misleading results
- Introduce a methodology to assess step-size adaptation mechanisms more thoroughly

Conclusion / Summary

- Testing only the sphere function is misleading, a methodology with more test functions is advisable
- All methods assume too large step-sizes on the sphere function in small dimension
- Self-adaptation (SA) often assumes (far) too small step-sizes, despite its bias

Proposed Methodology

- Evaluate on functions with known optimal step-size
- What to analyse: convergence rate, step-size change, and/or normalised step-size, for each function

Random fitness

- Unbiased random walk of $\ln \sigma^{(t)}$ is desired

Linear function $f: x \mapsto x_1$

- Test how quickly the step-size is increased
- Linear increase of $\ln \sigma^{(t)}$ is demanded: at least a factor of **1.1 (2)** within n evaluations (iterations)
- Step-size increment should be comparable to step-size decrement on the sphere

Sphere function $f: x \mapsto \|x\|^2$

- Test how quickly the step-size is decreased
- Limited loss in convergence rate: at most **3** times slower

- Optimal step-size: $\sigma^{(t)} = \frac{\sigma_{\text{opt}}^*(n)}{n} \times f(x^{(t)})^{1/2}$

Ellipsoid function $f: x \mapsto \sum_{i=1}^n \alpha^{(i-1)/(n-1)} x_i^2$

- Test the ability to estimate the optimal step-size
- With two moderate values of the condition number $\alpha = 10$, 100

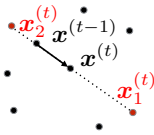
Stationary sphere

- Test the ability to estimate the optimal step-size without approaching the optimum
- Artificial model with stationary optimal step-size: at each iteration, the mean vector is projected on the unit sphere

Tested Algorithms

Two-Point Step-Size Adaptation (TPA)

- Performs a coarse line-search along the direction of the last mean shift from $x^{(t-1)}$ to $x^{(t)}$
- Sample the first two offspring along this line



$$x_{1/2}^{(t)} = x^{(t)} \pm \sigma^{(t)} \times \|\mathcal{N}_t(\mathbf{0}, \mathbf{I})\| \frac{x^{(t)} - x^{(t-1)}}{\|x^{(t)} - x^{(t-1)}\|}$$

- If $x_1^{(t)}$ is better than $x_2^{(t)}$, increase $\sigma^{(t+1)}$, otherwise decrease it

$$\sigma^{(t+1)} = \sigma^{(t)} \times \exp\left(\frac{s^{(t)}}{d_\sigma}\right)$$

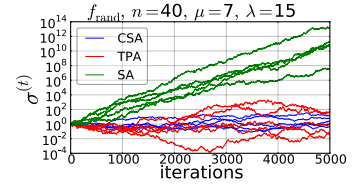
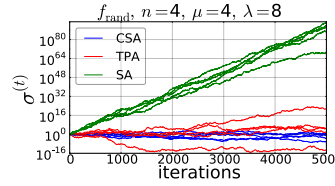
$$s^{(t)} = (1 - c_\sigma)s^{(t-1)} + c_\sigma \frac{\text{rank}(x_2^{(t)}) - \text{rank}(x_1^{(t)})}{\lambda - 1}, \quad d_\sigma = \sqrt{n}, \quad c_\sigma = 0.3$$

Cumulative Step-Size Adaptation (CSA) [1]

Self-Adaptation (SA) [2]

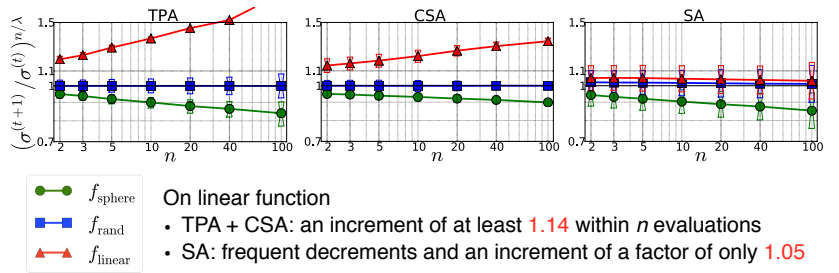
Results

Random fitness



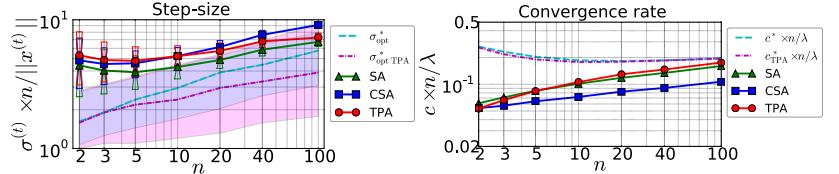
- Biased random walk for SA

Step-size change on random fitness, linear, and sphere functions



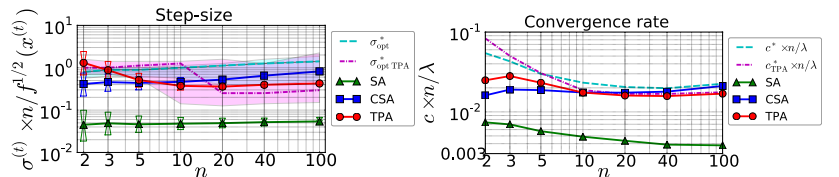
- On linear function
 - TPA + CSA: an increment of at least **1.14** within n evaluations
 - SA: frequent decrements and an increment of a factor of only **1.05**

Sphere function



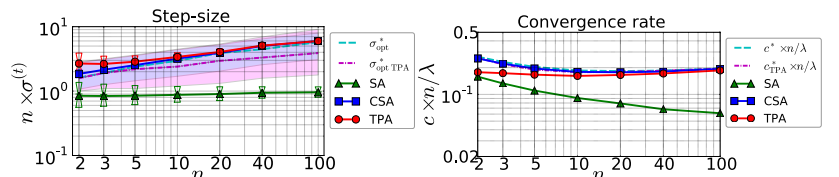
- Shaded areas: values with at most 20% performance loss
- Small dimension: all algorithms are up to **5** times slower than optimal
- The gap to the optimal step-size decreases with increasing n

Ellipsoid function with $\alpha = 100$



- TPA + CSA: close-to-optimal convergence rate in $n \geq 10$
- SA: performance deteriorates drastically compared to the sphere, step-size **20** times smaller than the optimal, convergence rate up to **6** times slower than the optimal

Stationary sphere



- TPA + CSA: close-to-optimal convergence rate in large dimensions, a moderate loss in small ones for TPA
- SA: step-size **2** to **6** times smaller than optimal, convergence rate **3** times slower than optimal only in $n = 100$

References

* Inria, TAO Team, Univ. Paris-Sud, Rue Noetzlin, Bat. 660, 91405 Orsay Cedex France

1. N. Hansen and A. Ostermeier. Completely derandomized self-adaptation in evolution strategies. *Evolutionary Computation*, 9(2): 159–195, 2001
 2. H.-P. Schwefel. *Evolution and Optimum Seeking*. Sixth-Generation Computer Technology. Wiley Interscience, New York, 1995