Local Performance of the $(\mu/\mu_I, \lambda)$-ES in a Noisy Environment

Dirk V. Arnold and Hans-Georg Beyer
presented by Andreas Beham

Contents of this presentation

- Noisy Optimization
- The $(\mu/\mu_I,\lambda)$-ES algorithm
- Introducing a performance measure
- Discussing the results
- Conclusions & Future Work

Noisy Optimization

- Noise is a constant companion in our life
- Specifically to optimization problems it can come from various sources
  - Measurement errors
  - Stochastic simulation
  - User interaction
- Consequences
  - Reduced convergence velocity
  - Inability to reach the global optimum

Noisy Optimization

- Nature itself is a noisy environment - Evolutionary Algorithms are nature inspired models and thus perceived to be robust with respect to noise
- Research on GA
  - J.M. Fitzpatrick and J. J. Grefenstette. 1988. „Genetic Algorithms in Noisy Environments”
  - B. L. Miller and D. E. Goldberg. 1997. „Genetic Algorithms, Selection Schemes, and the Varying Effects of Noise”
  - M. Rattray and J. Shapiro. 1997. „Noisy Fitness Evaluation in Genetic Algorithms and the Dynamics of Learning”
- Research on ES
  - I. Rechenberg. 1973. „Evolutionsstrategie: Optimierung Technischer Systeme nach den Prinzipien der biologischen Evolution“
  - H.-G. Beyer. 1993. „Toward a Theory of Evolution Strategies: Some Asymptotical Results from the (1,λ)/(1+λ)-Theory“
  - U. Hammel and T. Bäck. 1994. „Evolution Strategies on Noisy Functions. How to Improve Convergence Properties”

Noisy Optimization

- Empirical evidence shows that a population of solutions weakens the effect of noise
- Infinitely large population can remove the effect of noise entirely in some circumstances
  - great! though a little bit „impractical“
- Resampling – the „brute force“ approach
  - decrease in efficiency
- Balance between Population size and Resampling

Noisy Optimization

- Why this work?
  - Empirical evidence showed that larger populations helped mitigating the effect of noise, yet..
  - ...analysis of the ES was restricted to 1 parent versions!
  - Can the efficiency of the *(1+1)-ES* be topped in a noisy environment?
  - Is Genetic Repair still present when optimizing under noise?
  - Does the ES benefit from larger populations in the same way that the GA does?

The $(\mu/\mu_I,\lambda)$-ES algorithm

- μ parents, λ offsprings
- Comma notation
- Recombination
- Algorithm flow
  1. Initialize μ parents
  2. Compute the centroid $\langle x \rangle = \frac{1}{\mu}\sum_{i=1}^{\mu} x^{(i)}$
  3. Create λ offsprings $y^{(i)} = \langle x \rangle + z^{(i)},\ \ z^{(i)} = N(0,\sigma^2),\ \ i = 1,\dots,\lambda$
  4. Replace parents with μ best from y
  5. Continue with 2. until a stop criterion is satisfied

Fitness environment

- Quadratic Sphere

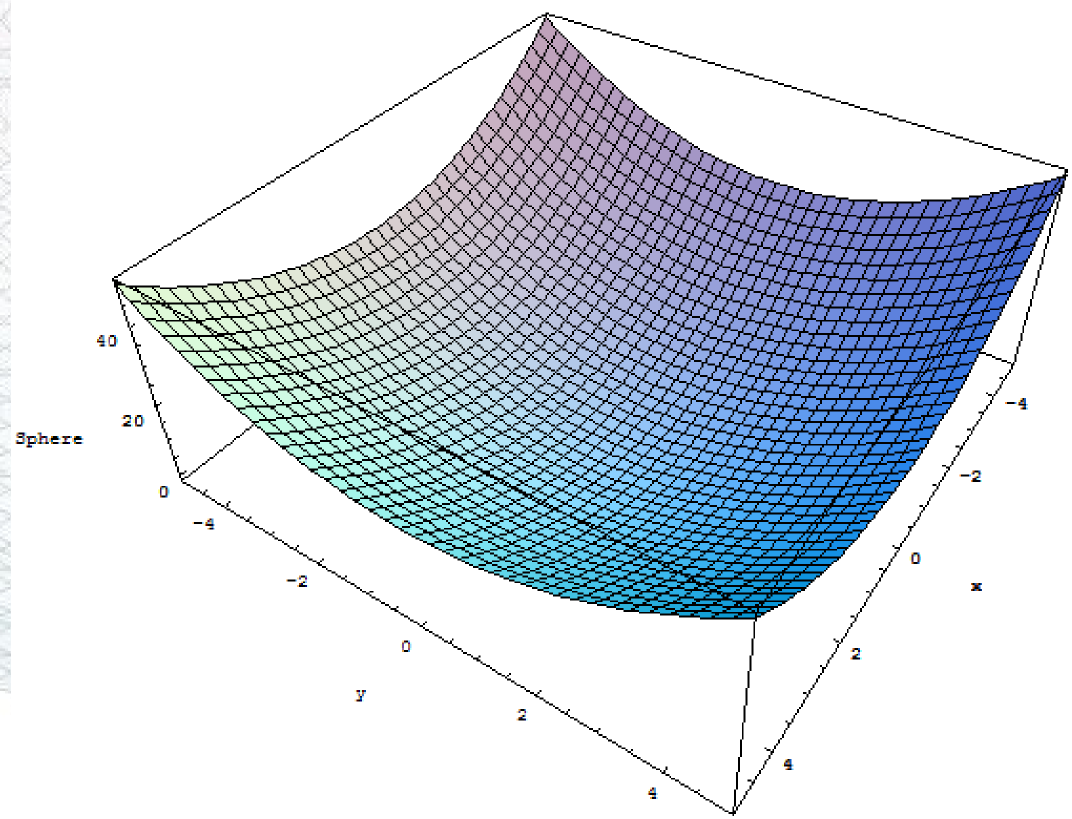
$$f(x) = \sum_{i=1}^{N} (\hat{x}_i - x_i)^2$$

- Introducing noise

$$\sigma_\varepsilon \delta, \quad \delta = \mathrm{N}(0,1)$$

- Perceived fitness

$$f(x) + \sigma_\varepsilon \delta$$



Two dimensional noiseless Sphere with optimum at (0,0)

Introducing a performance measure

- Fitness advantage

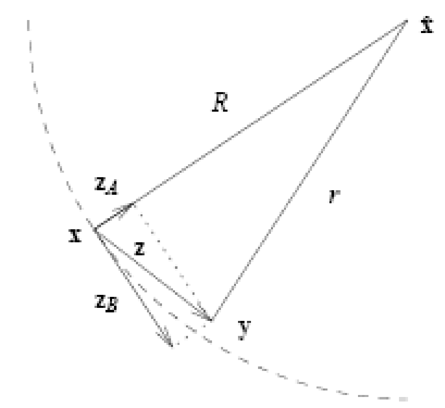
  $$q_x(z) = f(x) - f(x+z)$$

- Measurement of local performance
  - Progress rate φ
    - Residual distance between the centroids
  - Expected fitness gain q
    - Fitness distance between the centroids
  - For the Sphere φ and q coincide

Introducing a performance measure

- Seperating mutation vector z

$$z_A = \sigma(z_1, 0, \dots, 0)^T$$
$$z_B = \sigma(0, z_2, \dots, z_N)^T$$



- Introduce normalized quantities

$$\sigma^* = \sigma \frac{N}{R}, \quad \sigma^*_\varepsilon = \sigma_\varepsilon \frac{N}{2R^2}, \quad q^*_x(z) = q_x(z) \frac{N}{2R^2}$$

- The normalized fitness advantage becomes

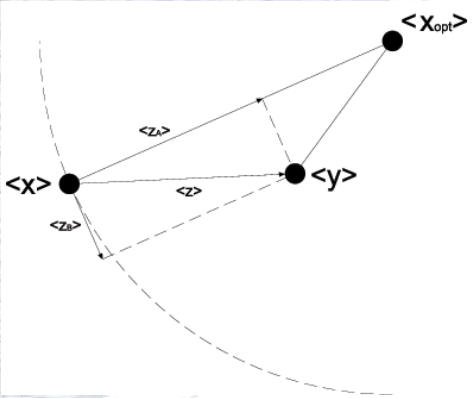
$$q^*(z) = \sigma^* z_1 - \frac{\sigma^{*2}}{2N} z_1^2 - \frac{\sigma^{*2}}{2N} \sum_{i=2}^{N} z_i^2 \approx \sigma^* z_1 - \frac{\sigma^{*2}}{2} \quad \text{für } N \rightarrow \infty$$

- The perceived normalized fitness advantage thus is

$$q^*(z) + \sigma_\varepsilon^* \delta = \sigma^*(z_1 + \vartheta \delta) - \frac{\sigma^{*2}}{2}, \quad \vartheta = \frac{\sigma_\varepsilon^*}{\sigma^*} \text{ noise-to-signal ratio}$$

Introducing a performance measure

- The progress vector for a population of selected offspring solutions, is the difference between the centroids of the parent population and the selected offsprings
- The expected fitness gain is the expected value of the normalized fitness advantage between the centroids



$$q^*(\langle z\rangle)=\sigma^*\langle z_1\rangle-\frac{\sigma^{*2}}{2\text{N}}\langle z_1\rangle^2-\frac{\sigma^{*2}}{2\text{N}}\sum_{i=2}^{N}\langle z_i\rangle^2$$

- The function can be simplified similar to the normalized fitness advantage

$$q^*(\langle z\rangle) \approx \sigma^*\langle z_1\rangle-\frac{\sigma^{*2}}{2\mu}$$

- The factor μ reflects the presence of genetic repair

Introducing a performance measure

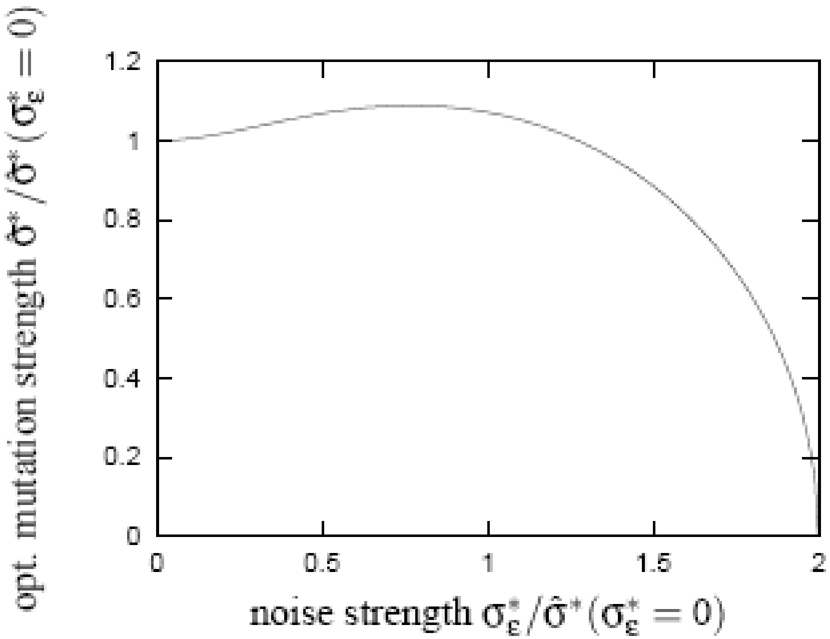
- Calculating the expected value of the normalized fitness advantage is quite a challenging task
- The final progress rate law is

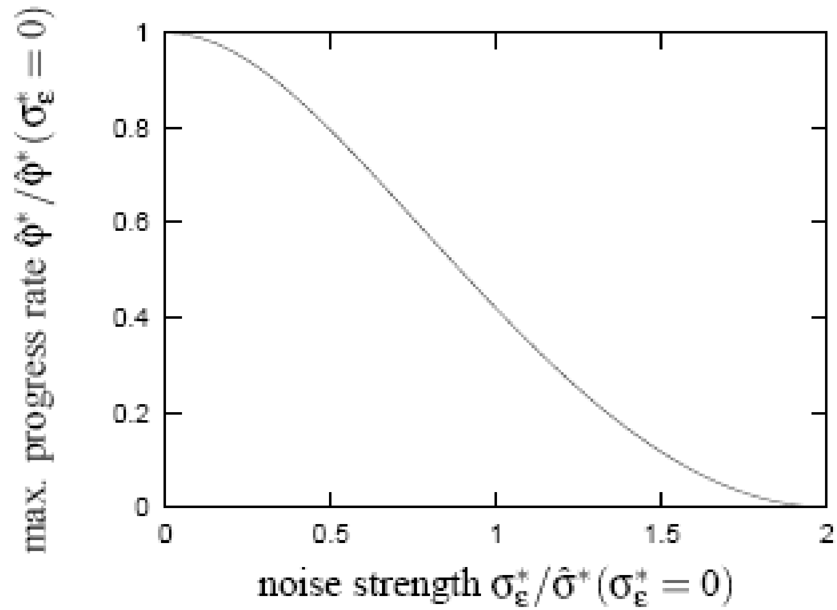
$$\varphi^* = \frac{c_{\mu/\mu,\lambda}\sigma^*}{\sqrt{(1+\vartheta^2)}} - \frac{\sigma^{*2}}{2\mu}, \qquad c_{\mu/\mu,\lambda} = \frac{\lambda-\mu}{2\pi}\binom{\lambda}{\mu}\int_{-\infty}^{\infty} e^{-x^2}[\Phi(x)]^{\lambda-\mu-1}[1-\Phi(x)]^{\mu-1}\,\mathrm{d}x$$ [1]

- This is in accordance to the analytical results obtained from the $(\mu/\mu_I,\lambda)$-ES on the noiseless Sphere
- On the noisy Sphere for μ=1 results of the $(1,\lambda)$-ES appear as special cases to the equation obtained here

1 H.-G. Beyer, „Toward a Theory of Evolution Strategies: On the Benefit of Sex – the (μ/μ,λ)-Theory“, Evolutionary Computation, 3(1), pages 81-111, (1995)

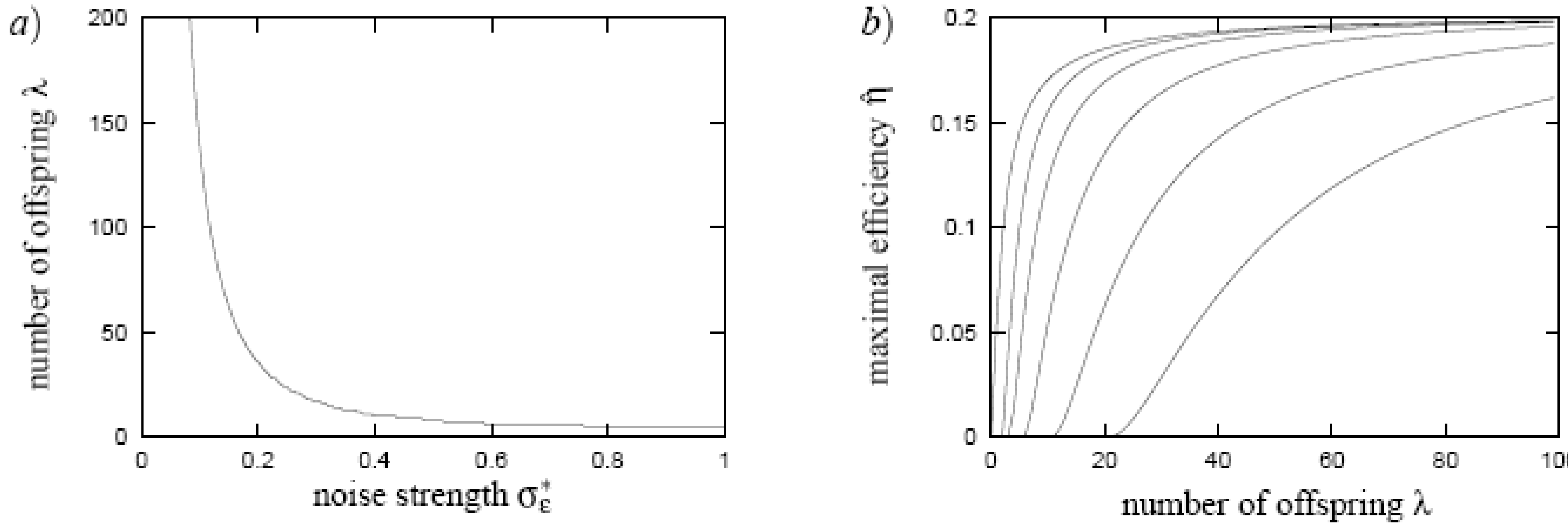
Discussing the results





- The $(\mu/\mu_I,\lambda)$-ES is able to achieve a positive progress rate up to a normalized noise strength of 2μc and thus by scaling the parameter μ we can deal with any noise strength
- Genetic repair is present and allows for higher mutation strengths, which in turn decreases the noise-to-signal ratio

Discussing the results



- From a) we can see that in the presence of noise the $(2/2_I,6)$-*ES* is more efficient compared to the $(1+1)$-*ES* for noise strength > 0.59 already
- In b) we can see the efficiency of the $(\mu/\mu_I,\lambda)$-*ES* for various normalized noise strengths (0,1,2,4,8,16).
- By increasing the number of offspring we asymptotically close the gap to the max efficiency of 0.202 in the absence of noise

Conclusions

- We could show that by increasing the population size we can remove the effects of noise entirely
- Genetic Repair is present, this makes it possible to choose higher mutation strengths and thus decrease the noise to signal ratio
- The efficiency of the *(1+1)-ES* could be exceeded by a multi-parent strategy with recombination for small population sizes already

Future Work

- Analyze the performance of the $(\mu/\mu_I,\lambda)\text{-}ES$ in other fitness environments and finite-dimensional parameter spaces
- Extend the analysis to dominant rather than intermediate recombination
- Investigate the effects of noise on the behavior of mutation strength adaption schemes

Thank you for your time