A closer look on self-adaptation

- Review of basics
- Experiments
- Conclusions
- Self-adaptive individuals as agents

Self-adaptation principles

- Biological model: repair enzymes, mutator genes
- No deterministic control: strategy parameters *evolve*
- *Indirect* link between fitness and useful strategy parameter settings
- Strategy parameters are conceivable as an *internal model* of the local topology
- Individual space:

  I = M x S

  – *M*: Search space
  – *S*: Strategy parameters space

The crucial claim (Schwefel 1987, 1992)

Self-adaptation of strategy parameters works

- Without exogenous control
- By recombining/mutating the strategy parameters
- By exploiting the implicit link between fitness and useful internal model

The crucial claim (Schwefel 1987, 1992)

Necessary conditions (found by experiments):

- Generation of an offspring surplus, $\lambda > \mu$
- $\mu > 1$ is necessary
- (μ, λ)-selection (to guarantee extinction of misadapted individuals
- A not too strong selective pressure, heuristic: $\lambda / \mu \approx 7$ e.g., (15,100)
- Recombination also on strategy parameters (especially: intermediate recombination)

Empirical test design

- On simple functions (with predictable optimal σ values)
- To check whether it works
- To investigate impact of various setups
- To compare observed and theoretically optimal behavior (if known)

Test functions for experiments

- One common step size ($n_\sigma = 1$): Sphere model

$$f_1(\overline{x}) = \sum_{i=1}^{n} x_i^2$$

- Appropriate scaling of variables ($n_\sigma = n$):

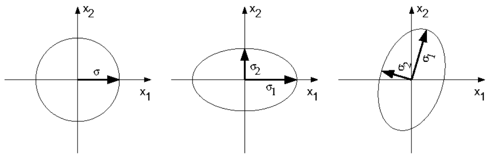
$$f_2(\overline{x}) = \sum_{i=1}^{n} i \cdot x_i^2$$

- A metric ($n_\sigma = n$, $n_\alpha = n \cdot (n - 1) / 2$)

$$f_3(\overline{x}) = \sum_{i=1}^{n} \left(\sum_{j=1}^{i} x_j \right)^2$$

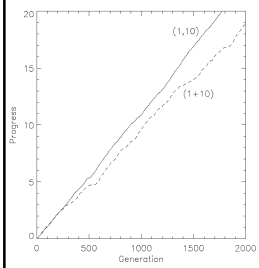
Refresh: effects of self-adaptation varaints

Sketch of the lines of equal probability density



Standard mutation $n_\sigma = 1$ | Standard mutations $n_\sigma = 2$ | Correlated mutations $n_\sigma = 2$, $n_\alpha = 1$

Experiment 1: (λ,μ) vs. (λ+μ)



Convergence velocities: (1, 10) vs. (1 + 10) ES (sphere model f_1 with n = 30 and $n_\sigma = 1$)

Progress measure:

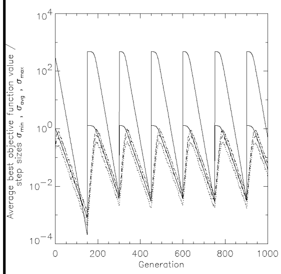
$$P_t = \log \sqrt{\frac{f_{\min}(0)}{f_{\min}(t)}}$$

- Time measured by number of generations
- Counterintuitive: elitist strategy is a bad choice
- Reason: misadapted σ might survive in an elitist strategy
- Forgetting is necessary to prevent stagnation periods

Time-varying sphere model

- Sphere model $f_1(\bar{x}) = || \bar{x} - \bar{x}^*||^2$
- Optimum location $\bar{x}^*$ is shifted every 150 generations
- (15,100)-ES, $n_\sigma = 1$, n = 30, no recombination.
- Simple model of a dynamic environment (with "catastrophes")

Experiment 2: learning σ



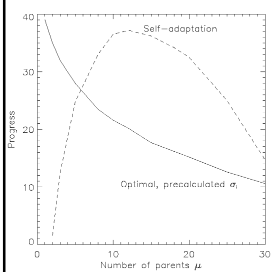
Best objective function value and minimum, average, maximum and optimal standard deviation.

- f is the time varying sphere model
- Optimal σ value known:

$$\sigma_{opt} = c_{\mu,\lambda} \frac{R}{n} = c_{\mu,\lambda} \frac{\sqrt{f(\bar{x})}}{n}$$

- The algorithm learns optimal σ's ("internal strategy") without exogenous control

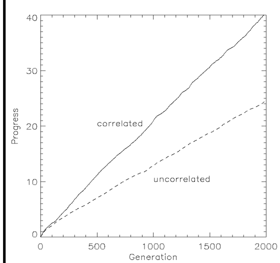
Experiment 3: effect of population size



Average convergence velocity on f_2

- (μ, 100)-ES with μ ∈ {1, … , 30}
- n_σ = n = 30, and the optimum $\sigma_i \propto 1/\sqrt{i}$ is known
- Optimum setting of σ_i: μ = 1 performs best
- Self-adaptation: for μ = 12 diverse parents are approx. as good as the optimal strategy
- Individuals exchange information about their "internal models" by recombination
- Self-adaptation is collective learning

Experiment 4: self-adaptation of covariances



Convergence velocities: ES with and w/o correlated mutations on f_3.

- (15, 100)-ES, n = n_σ = 10, n_α = 45
- Recombination:
  - Intermediary on x_i
  - Global intermediary on σ_i
  - None on α_j (covariances)
- Covariances increase effectiveness in case of rotated coordinate systems

Other variants for continuous search spaces

- Original Evolutionary Programming:
  $$\sigma' = \sigma \cdot (1 + \alpha \cdot N(0,1))$$
  Equivalent to log-normal with $n_\sigma = 1$, $\tau_0 = \alpha$ (Beyer 1995).
- Two-point distribution:
  $$\sigma' = \begin{cases} \sigma \cdot \alpha, & \text{if } u \sim U(0,1) \leq 1/2 \\ \sigma / \alpha, & \text{if } u \sim U(0,1) > 1/2 \end{cases}$$
  (Mutational step size control after Rechenberg, $\alpha = 1.3$).
- Substitution of N(0, 1) by other distributions (e.g., one-dimensional Cauchy, Yao and Liu 1996).

Self-adaptation: conclusions

- Powerful & robust parameter control scheme
- Optimal conditions concerning selection, population size, etc: ?
- Optimal learning rate settings (i.e., speed of self-adaptation): ?
- Few theoretical results

Self-adaptation: individuals as agents

- Individuals are *autonomous*; internal control of their behavior (mutation)
- Individuals *communicate* by exchanging partial information (recombination)
- Individuals are *reactive* to their environment (objective function)
- Further possibilities:
  - Spatial communication structure (graph)
  - Parallel implementation
  - More complex internal strategies: including symbolic representation