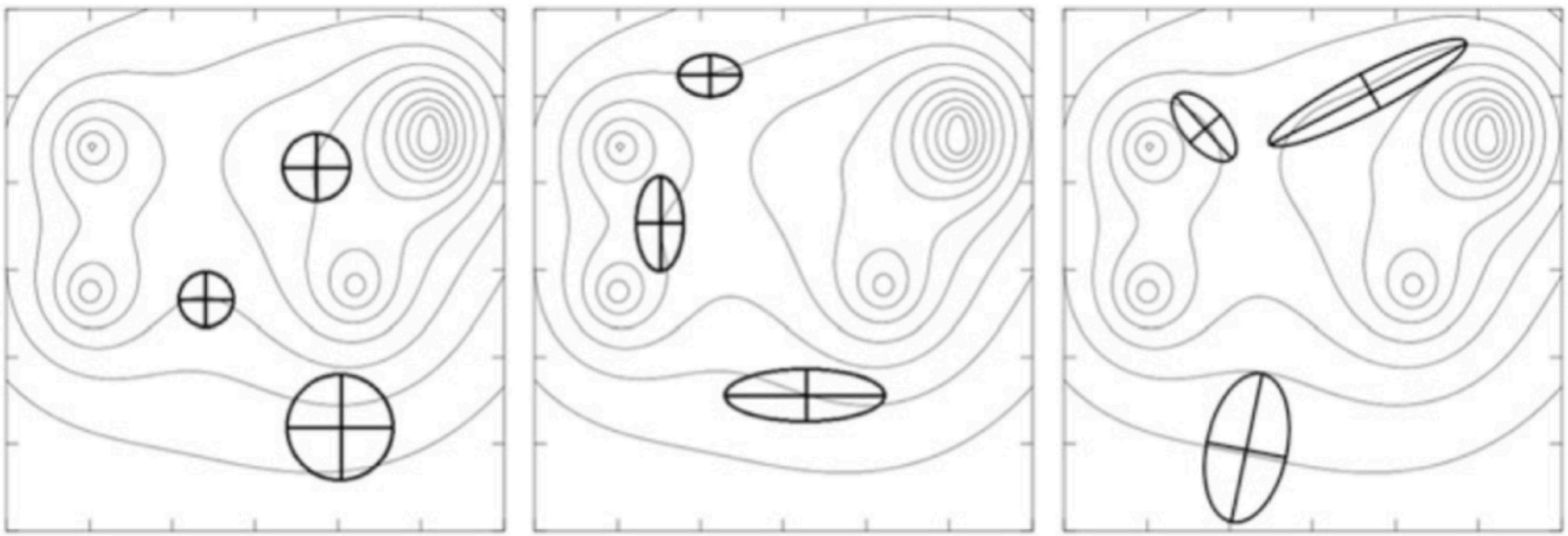
11: Evolution Strategies 2

- ES recombination
- Parent selection
- Survivor selection
- Example application
- Textbook Chapter 6.2, 4.4.2

Prof. Ting Hu, School of Computing, Queen’s University

Self-adaption in ES



constant variance for all dimensions

different variance for each dimension

covariance matrix adaptation (CMA)

n_σ	n_α	Structure of individuals	Remark
1	0	$\langle x_1, \ldots, x_n, \sigma \rangle$	Standard mutation
n	0	$\langle x_1, \ldots, x_n, \sigma_1, \ldots, \sigma_n \rangle$	Standard mutations
n	$n \cdot (n-1)/2$	$\langle x_1, \ldots, x_n, \sigma_1, \ldots, \sigma_n, \alpha_1, \ldots, \alpha_{n \cdot (n-1)/2} \rangle$	Correlated mutations

ES technical sketch

Representation	Real-valued vectors
Recombination	Discrete or intermediary
Mutation	Gaussian perturbation
Parent selection	Uniform random
Survivor selection	Deterministic elitist replacement by (μ, λ) or $(\mu + \lambda)$
Speciality	Self-adaptation of mutation step sizes

Recombination

- Create one child per recombination
- How to create an allele for one child
  - discrete recombination
  - intermediary recombination
- How to choose parents
  - global recombination: select two parents for each position

Recombination

- ES typically uses global recombination
- Discrete recombination for object variables
  - preserves diversity within the phenotype (solution) space
  - allows the trial of very different combinations of variables
- Intermediary recombination for strategy parameters (mutation step size)
  - assures a more cautious adaptation of strategy parameters

Parent selection

- Parents are selected by uniform random distribution whenever an operator needs one/some
- Thus: ES parent selection is unbiased - every individual has the same probability to be selected
- Note that in ES, "parent" means a population member (in GA, a population member selected to undergo variation)

Survivor selection

- Applied after creating λ children from the μ parents by recombination and mutation
- Deterministically picks the best ones
- Basis of selection is either:
  - the set of children only: (μ, λ)
  - the set of parents and children: $(\mu + \lambda)$

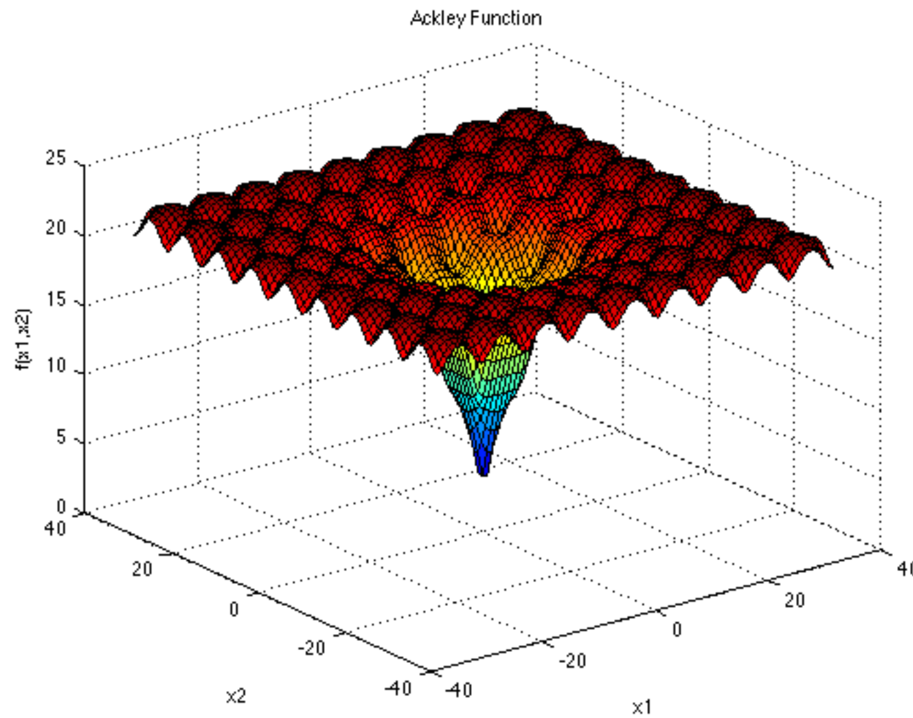
Survivor selection

- $(\mu + \lambda)$ - selection is an elitist strategy, monotonic progression
- (μ, λ) - selection can "forget" and more fair, punctuated
- Often (μ, λ) - selection is preferred by most users:
    - better in leaving local optima, especially working with a small μ
    - better in following moving optima
    - using the + strategy bad σ values can survive in $< x, \sigma >$ too long if their host *x* is very fit
- Selective pressure in ES is high ($\lambda \approx 7 \times \mu$ is the common setting)

Prerequisites for self-adaptation

- $\mu > 1$ to carry different strategies (individuals)
- $\lambda > \mu$ to generate offspring surplus
- Strong selection, e.g. $\lambda \approx 7 \times \mu$
- (μ, λ) - selection to get rid of mis-adapted σ
- Intermediate recombination for strategy parameters while discrete recombination for object variables

Example application (ES_1)



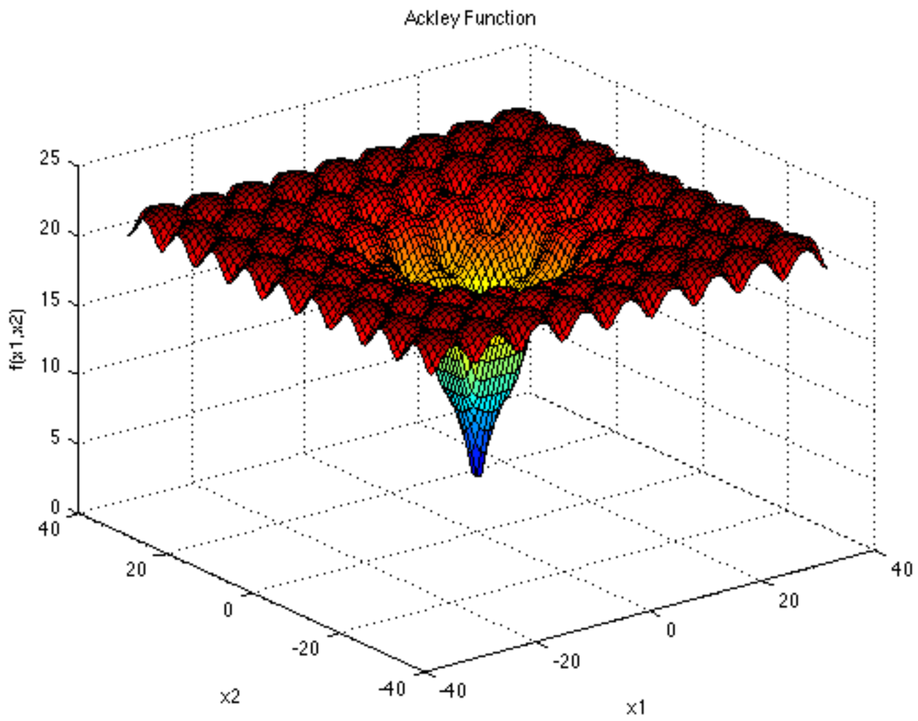
- Task is to minimize the Ackley function with $n = 30$ (optimal 0 at $x = (0,0,\ldots,0)$)

$$f(x) = -20 \cdot \exp\left(-0.2\sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} x_i^2}\right) - \exp\left(\frac{1}{n}\sum_{i=1}^{n}\cos(2\pi x_i)\right) + 20 + e$$

- Representation: $< x_1, x_2, \ldots, x_{30}, \sigma >$, $-30 < x_i < 30$
- Selection: (30, 200)
- Termination: after 100k fitness evaluations
- Initial standard deviations: $\sigma = 3.0$

Back and Schwefel: An overview of evolutionary algorithms for parameter optimization, 1993

Example application (ES_{30})



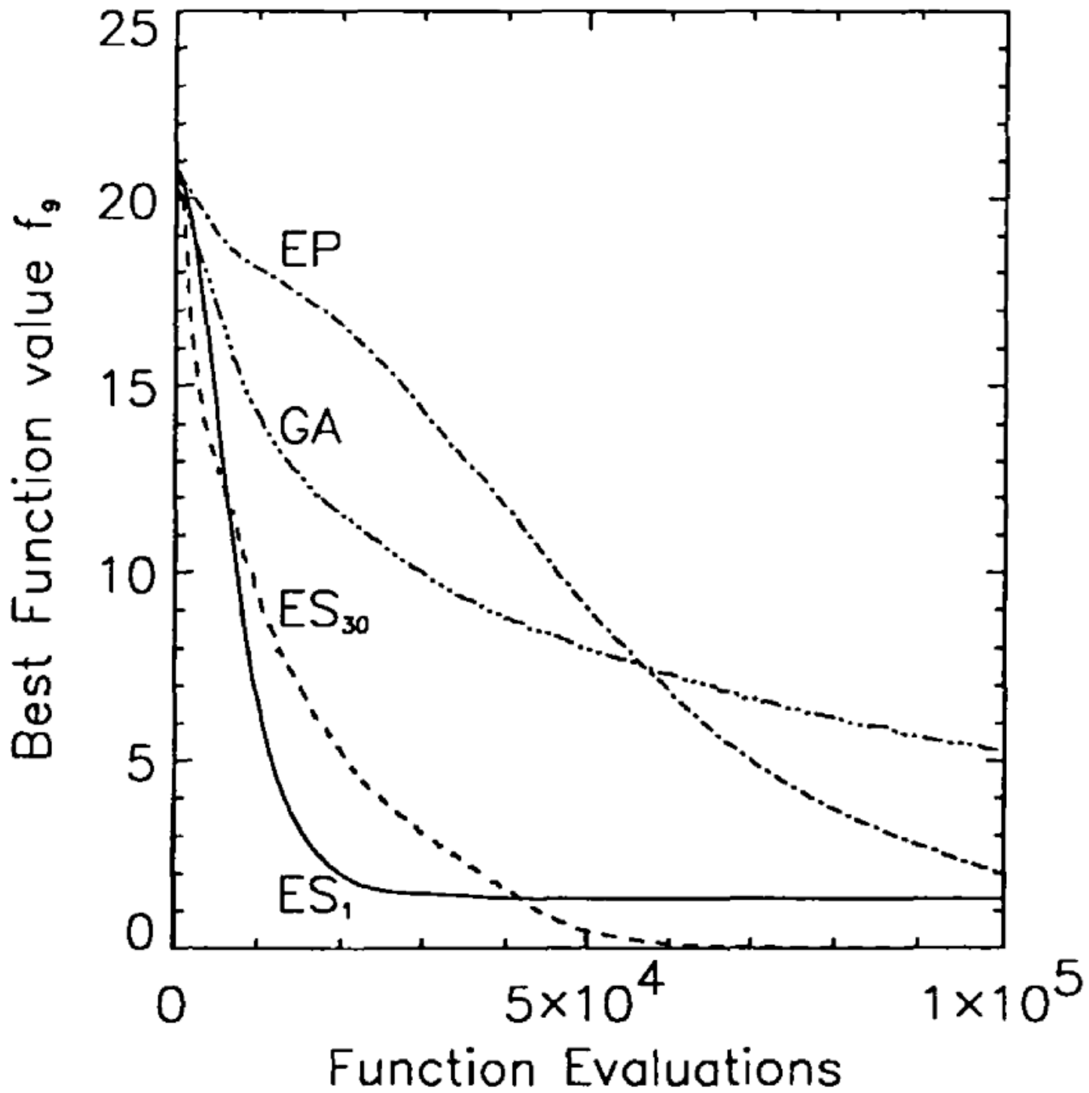
- Task is to minimize the Ackley function with $n = 30$ (optimal 0 at $x = (0,0,...,0)$)

$$f(x) = -20 \cdot \exp\left(-0.2\sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} x_i^2}\right) - \exp\left(\frac{1}{n}\sum_{i=1}^{n}\cos(2\pi x_i)\right) + 20 + e$$

- Representation: $< x_1, x_2, \ldots, x_{30}, \sigma_1, \sigma_2, \ldots, \sigma_{30} >$, $-30 < x_i < 30$
- Selection: (30, 200)
- Termination: after 100k fitness evaluations
- Initial standard deviations: $\sigma_i = 3.0$

Back and Schwefel: An overview of evolutionary algorithms for parameter optimization, 1993

Result comparison



20 runs	Mean best	SD
ES_1	1.326	1.039
ES_{30}	1.618×10^{-3}	9.290×10^{-4}

Back and Schwefel: An overview of evolutionary algorithms for parameter optimization, 1993