Nature-Inspired Computing

Evolutionary Strategies

Dr. Şima Uyar
September 2006

Overview

- developed in Germany in the 1970's
- typically applied to numerical optimization problems
- features:
  - fast
  - good optimizer for real-valued optimisation
  - relatively much theory behind it
  - self-adaptation of (mutation) parameters is standard

Overview

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

Introductory Example

- Task: minimimise $f : R^n \rightarrow R$
- Algorithm: "two-membered ES" using
  - vectors from R^n as chromosomes
  - population size 1
  - only mutation creating one child
  - greedy selection

Introductory Example: Pseudocode

```
set t = 0
create initial point xᵗ = 〈 x1ᵗ,…,xnᵗ 〉
repeat until (termin criteria)
do
  draw zi from a normal distr. for all i = 1,…,n
   yiᵗ = xiᵗ + zi
   if f(xᵗ) < f(yᵗ) then
     xᵗ⁺¹ = xᵗ
   else
     xᵗ⁺¹ = yᵗ
   fi
    set t = t+1
od
```

Introductory Example: Mutation Mechanism

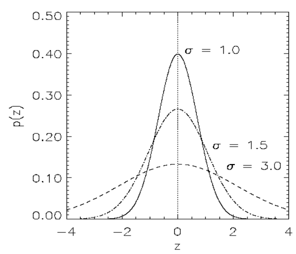
- z values drawn from normal distribution $N(\xi,\sigma)$
  - mean ξ is set to 0
  - variation σ is called mutation step size
- σ is varied on the fly by the "1/5 success rule"

Introductory Example: Mutation Mechanism

- this rule resets σ after every k iterations by
  - $\sigma = \sigma / c$ if $p_s > 1/5$
  - $\sigma = \sigma \cdot c$ if $p_s < 1/5$
  - $\sigma = \sigma$ if $p_s = 1/5$

where p_s is the % of successful mutations, $0.8 \le c \le 1$

Illustration of Normal Distribution



Representation

- chromosomes consist of three parts:
  - object variables: $x_1,\dots,x_n$
  - strategy parameters:
    - mutation step sizes: $\sigma_1,\dots,\sigma_{n_\sigma}$
    - rotation angles: $\alpha_1,\dots, \alpha_{n_\alpha}$

Representation

- not every component is always present
- full size: $\langle x_1,\dots,x_n, \sigma_1,\dots,\sigma_n ,\alpha_1,\dots, \alpha_k\rangle$

where k = n(n-1)/2 (no. of i,j pairs)

Mutation

- main mechanism: changing value by adding random noise drawn from normal distribution

$x'_i = x_i + N(0,\sigma)$

Mutation

- key idea:
  - σ is part of the chromosome $\langle x_1,\dots,x_n, \sigma \rangle$
  - σ is also mutated into σ' (see later how)
- thus: mutation step size σ is co-evolving with the solution x

Mutation

- net mutation effect:

$\langle x, \sigma \rangle \rightarrow \langle x', \sigma' \rangle$

- order is important:
  - first $\sigma \rightarrow \sigma'$ (see later how)
  - then $x \rightarrow x' = x + N(0,\sigma')$

Mutation

- rationale: new $\langle x', \sigma' \rangle$ is evaluated twice
  - primary: x' is good if f(x') is good
  - secondary: σ' is good if the x' created from it is good
- not holds if mutation order is reversed

Mutation Case 1: Uncorrelated Mutation with One σ

- chromosomes: $\langle x_1,\dots,x_n, \sigma \rangle$
  $\sigma' = \sigma \cdot \exp(\tau \cdot N(0,1))$
  $x'_i = x_i + \sigma' \cdot N(0,1)$
- typically the "learning rate" $\tau \propto 1/ n^{½}$

Mutation Case 2: Uncorrelated Mutation with n σ's

- chromosomes: $\langle x_1,\dots,x_n, \sigma_1,\dots, \sigma_n \rangle$
  $\sigma'_i = \sigma_i \cdot \exp(\tau' \cdot N(0,1) + \tau \cdot N_i (0,1))$
  $x'_i = x_i + \sigma'_i \cdot N_i (0,1)$
- two learning rate parameters:
  - τ' overall learning rate
  - τ coordinate-wise learning rate
  $\tau \propto 1/(2 n)^{½}$ and $\tau \propto 1/(2 n^{½})^{½}$

Recombination

- creates one child
- acts per variable / position by either
  - averaging parental values, or
  - selecting one of the parental values
- from two or more parents by either:
  - using two parents to make a child
  - selecting two parents for each position anew

Names of Recombinations

	Two fixed parents	Two parents selected for each i
$z_i = (x_i + y_i)/2$	Local intermediary	Global intermediary
z_i is x_i or y_i chosen randomly	Local discrete	Global discrete

Parent Selection

- selected by uniform random distribution whenever an operator needs one/some
- ES parent selection is unbiased - every individual has the same probability to be selected

Parent Selection

- in ES "parent" means a population member
- in GA's: a population member selected to undergo variation

Survivor Selection

- applied after creating λ children from the μ parents by mutation and recombination
- deterministically chops off the "bad stuff"
- basis of selection is either:
  - set of children only: (μ,λ)-selection
  - set of parents and children: $(\mu+\lambda)$-selection

Survivor Selection cont'd

- $(\mu+\lambda)$-selection is an elitist strategy
- (μ,λ)-selection can "forget"
- selective pressure in ES is very high ($\lambda \approx 7 \cdot \mu$ is the common setting)