EE496 : COMPUTATIONAL INTELLIGENCE
EA02-1 : FITNESS AND SELECTION

UGUR HALICI
METU: Department of Electrical and Electronics Engineering (EEE)
METU-Hacettepe U: Neuroscience and Neurotechnology (NSNT)

Fitness

topics

- Selective pressure
- Selection intensity
- Fitness-proportionate Selection
- Premature convergence
- Vanishing selective pressure
- Adapting of the fitness function

Principle of selection

- Better individuals (better fitness) should have better chances to create offspring (differential reproduction)
- **Selective pressure:** Strength of prefering good individuals
- Choice of selective pressure: Contrast of

  Exploration of the space:
  - deviation of the individuals over Ω as wide as possible
  - preferably big chances to find global optimum

  $\Rightarrow$ smaller selective pressure is desired

  Exploitation (of good individuals):
  - Strive (try) for (perhaps local) optimum in the vicinity of good individuals
  - Convergence to optimum

  $\Rightarrow$ higher selective pressure is preferred

Comparison of selection methods

metrics for comparison of methods created for selective pressure :

- **time to takeover:** # generations until population converges (population is called converged, if all individuals are identical)
- **selection intensity:** difference between average quality before and after the selection

Selection intensity

according to [Weicker, 2007]

Definition (Selection intensity)

Let $(\Omega, f, \succ)$ be a considered optimization problem and a selection operator $Sel^{\xi} : (\mathcal{G} \times \mathcal{Z} \times R)^r \rightarrow (\mathcal{G} \times \mathcal{Z} \times R)^s$ is applied on a population P with an average quality μ_f and standard deviation σ_f. Then, let μ^{sel}_f be the average quality of the population P_{sel} and the selection operator has the selection intensity

$$I_{\text{sel}} = \begin{cases} \frac{\mu_f^{\text{sel}}-\mu_f}{\sigma_f} & \cancel{\text{falls}}\ \text{if}\ \succ = >, \\ \frac{\mu_f-\mu_f^{\text{sel}}}{\sigma_f} & \cancel{\text{sonst.}}\ \text{otherwise} \end{cases}$$

if the problem is a maximization problem.

Selection intensity

- the higher I_{sel}, the higher the selective pressure

Example:

- 10 Indiv. with fitness: 2.0, 2.1, 3.0, 4.0, 4.3, 4.4, 4.5, 4.9, 5.5, 6.0
- selection leads to individuals with quality: 2.0, 3.0, 4.0, 4.4, 5.5

$$\mu_f = \frac{1}{|P|} \sum_{i=1}^{|P|} A^{(i)}.F \qquad \text{(Average of the fitness)}$$

$$\sigma_f = \sqrt{\frac{1}{|P|-1} \sum_{i=1}^{|P|} (A^{(i)}.F - \mu_f)^2} \qquad \text{(standard deviation)}$$

$$\Rightarrow \quad \mu_f = 4.07, \quad \sigma_f = 1.27, \quad \mu_f^{\text{sel}} = 3.78, \quad I_{\text{sel}} = \frac{4.07-3.78}{1.27} = 0.228$$

Criticism on selection intensity:

- metric requires a standard normal distribution of values
- rarely applicable on general optimization problems

Choice of the selective pressure

- **best strategy:** time-dependent selective pressure
  - low selective pressure in prior generations,
  - higher selective pressure in later generations

  ⇒ at first good exploration of the space,
  then exploitation of the promising region
- regulation of the selective pressure by adapting the fitness function or by the parameter of selection method
- important **selection methods**:
  - Roulette-wheel Selection,
  - Rank-based Selection,
  - Tournament Selection
- important **adaption methods**:
  - Adaption of the variation of the fitness,
  - linear dynamical scaling,
  - σ-scaling

Roulette wheel selection

- best known selection method
- compute the relative fitness of the individuals A(i), $1 \leq i \leq |P|$

$$f_{\text{rel}}\left(A^{(i)}\right) = \frac{A^{(i)}.F}{\sum_{j=1}^{|P|} A^{(j)}.F}$$

and interprete $f_{rel}(A^{(i)})$ as a probability to be selected

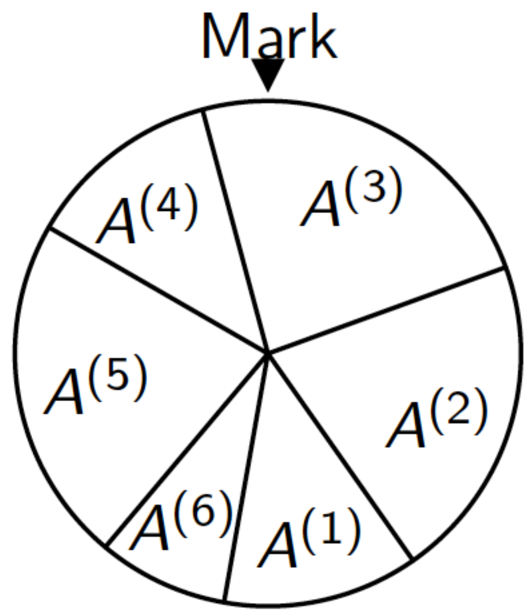
- (so called **fitness-proportionate Selection**)
- **note:** absolute fitness A.F may not be negative
- **Attention:** fitness has to be maximized (otherwise: selection of bad individuals with high probability)
- **Demonstration:** Roulette-wheel with 1 sector per individual $A^{(i)}$,

sector size = relative fitness values $f_{rel}(A^{(i)})$

Roulette-wheel selection: Demonstration

Selection of an individual:

1. set the roulette-wheel into motion
2. choose the indindividual of the
   corresponding sector



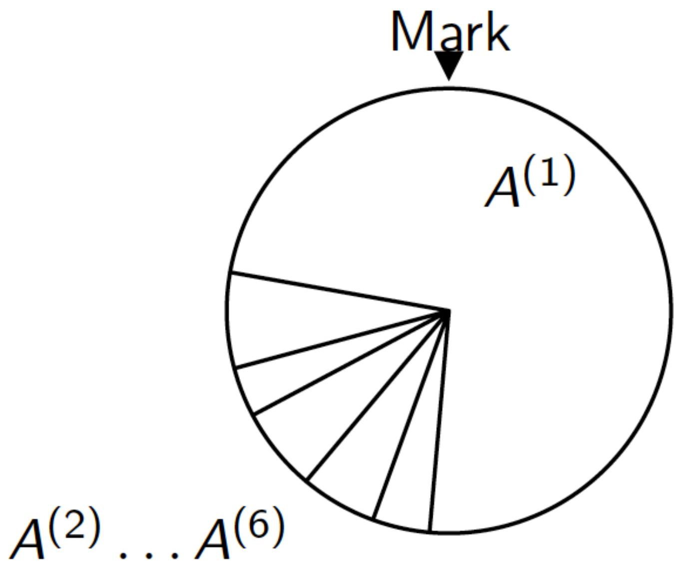
Selection of the next population:

- repeat selection # individuals-times

Disadvantage: Calculation of the relative fitness by summing up all fitness values (normalization factor)

- constant initial population during the selection
- aggravated (made worse) parallelization of the implementation

Roulette-wheel selection: Dominance Problem



- Individual with a very high fitness may dominate the selection
- Due to many copies/very similar individuals: dominance may become even stronger in subsequent generations

⇒ **Crowding:** population of very similar/identical individuals
  - results in a very fast find of the (local) optimum

- **Disadvantage:** diversity of the population vanishes
- Exploitation of worse (premature good) individuals
- No exploration of the space but local optimization (preferred in later generations, undesirable at the beginning)

Roulette-wheel selection: selection intensity

Theorem

When using a simple fitness-proportionate selection in a population with average quality μ_f and variation of the quality σ^2_f , the selection intensity is

$$I_{sel} = \frac{\sigma_f}{\mu_f}.$$

proof : tohose interested may look at the course book

Note: definition of I_{sel} was on page 3

Fitness function: Premature convergence

Dominance problem illustrates the strong influence of the fitness function on the effect of the fitness-proportionate selection

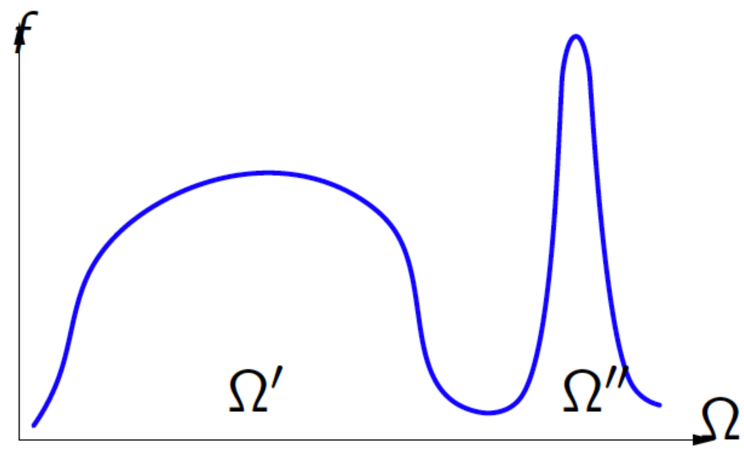
- Problem of **premature convergence:**

  If (value) range of the maximizing function is very huge

- Example: no chromosome at the beginning in the section Ω'' →

  population remains by selection in the vicinity of the local maximum in the section Ω'

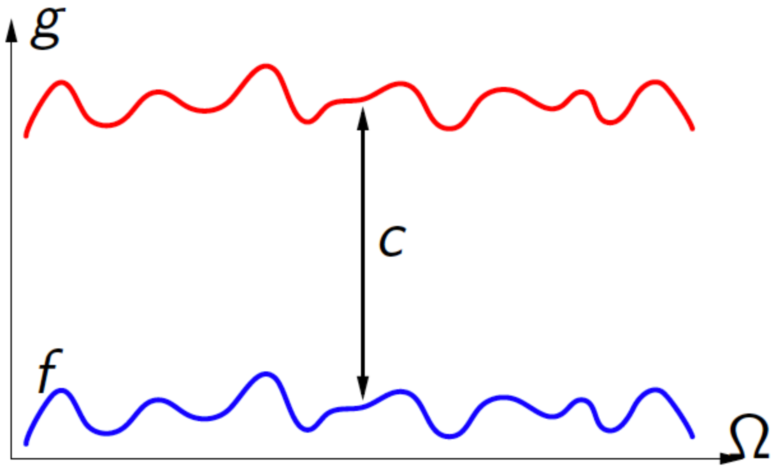
Individuals which converge to the section between Ω' and Ω'' have worse chances to create offspring



Vanishing selective pressure

Problem of the **absolute height** of the fitness values with respect to the **variation**

or: Problem of the **vanishing selective pressure**:



- Maximizing of f : Ω → R is equivalent to the maximization of g : Ω → R with g(x) ≡ f (x) + c, c ∈ R

  $c \gg \sup_{x\in\Omega} f(x) \Rightarrow x \in \Omega : g_{rel}(x) \approx 1 / |P|$ (|P| is pop.-size)

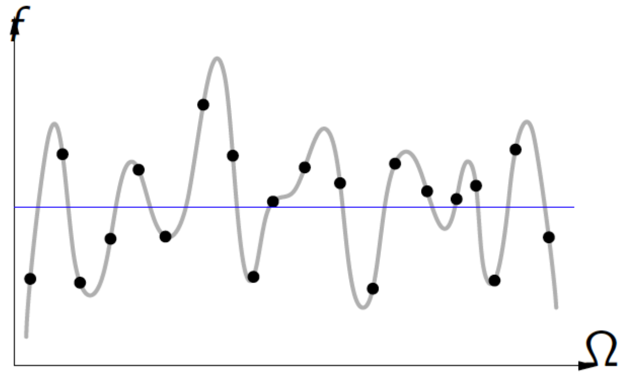
  ⇒ (too) small selective pressure

(Note: remember definition of f_{rel} on page 7:

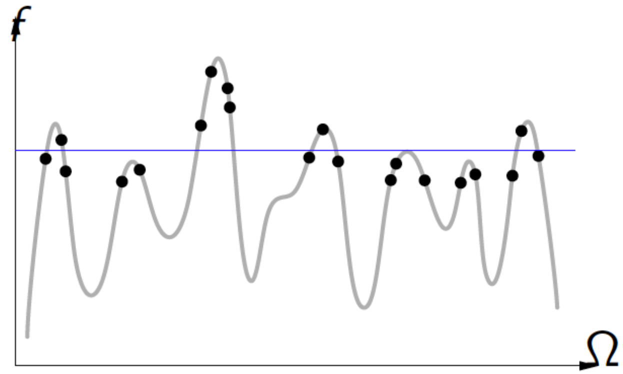
$$f_{\text{rel}}\left(A^{(i)}\right) = \frac{A^{(i)}.F}{\sum_{j=1}^{|P|} A^{(j)}.F}$$

Vanishing selective pressure

- Problem is perhaps grounded on the EA itself
- it increases tendentially the (average) fitness of the individuals
- Example: points illustrate individuals of the generation



early generation



later generation

- higher selective pressure at the beginning due to random fitness values
- later: smaller selective pressure (inverse way is preferred)

Adapting of the fitness function

Approach: **Scaling of the fitness**

linear dynamical scaling:

$$f_{lds}(A) = \alpha \cdot A.F - \min \{A^{(i)}.F \mid P(t) = \{A(1), \ldots, A(r)\}\}, \ \alpha > 0$$

- instead minimum of $P(t)$, minimum of the last k generations can be used
- usually $\alpha > 1$

σ-Scaling:

$$f_{\sigma}(A) = A.F - (\mu_f(t) - \beta \cdot \sigma_f(t)), \ \beta > 0$$

- Problem: Choice of the parameter α and β

Adapting of the fitness function

- consider **variation coefficient** of the fitness function

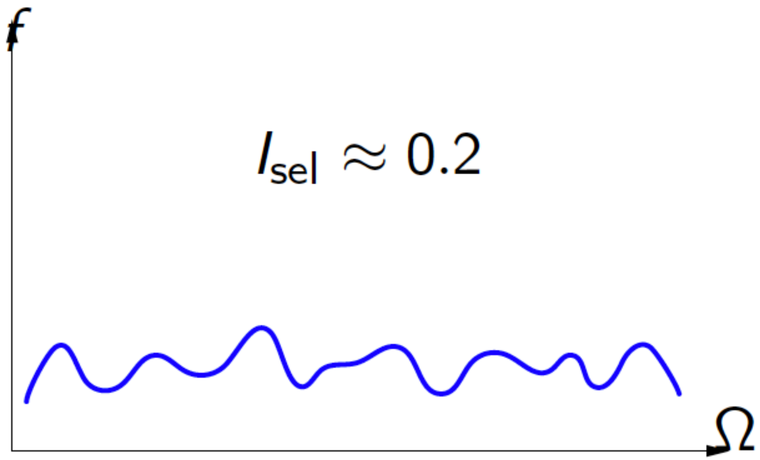
$$v = \frac{\sigma_f}{\mu_f} = \frac{\sqrt{\frac{1}{|\Omega|-1} \sum_{x' \in \Omega} \left(f(x') - \frac{1}{|\Omega|} \sum_{x \in \Omega} f(x) \right)^2}}{\frac{1}{|\Omega|} \sum_{x \in \Omega} f(x)}, \quad v(t) = \frac{\sigma_f(t)}{\mu_f(t)}$$

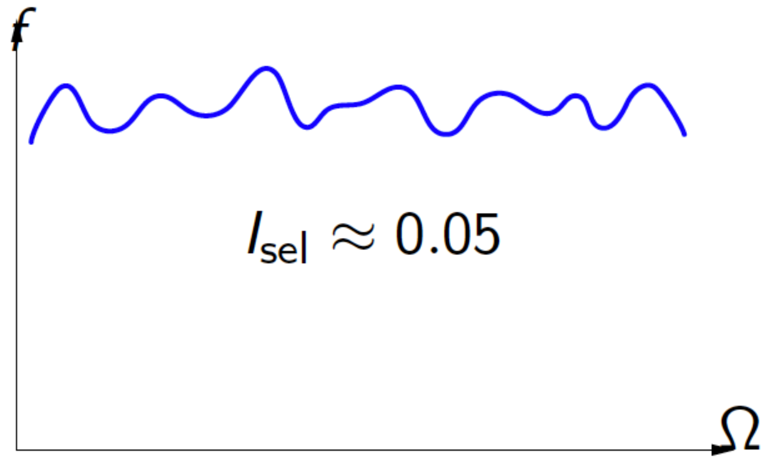
- empirical discovery: $v \approx 0.1$ yields good ratio of exploration and exploitation
- if $v \neq 0.1$, then adapting of f (e.g. by scaling)
- v can be estimated, not be calculated,
- but practical calculations of v: Replace of Ω by P(t)
- hence: approximation of v by selection intensity I_{sel} (t)

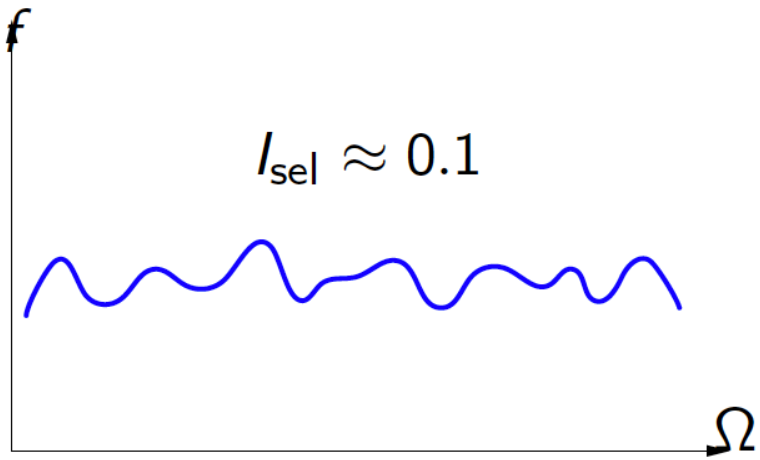
⇒ in each generation: calculate I_{sel} (t) and adapt f accordingly

(σ-scaling with $\beta = 1/(I^*_{sel})$, $I^*_{sel} = 0.1$)

Illustration of the selection intensity







- too high I_{sel}: premature convergence
- too small I_{sel}: vanishing selective pressure
- appropriate: $I_{sel} \approx 0.1$

Adaption of the fitness function:

- determine relative fitness not directly from f(x) but from

$$g(x) \equiv \exp\left(f(x)/(kT)\right)$$

- time-dependent **temperature** T controls selective pressure
- k is normalizing constant
- Temperature decreases e.g. linearly considering the predefined maximum number of generations

Selection

- Roulette-wheel Selection
- Expected value model
- Rank-based Selection
- Tournament selection
- Elitism
- Characterization

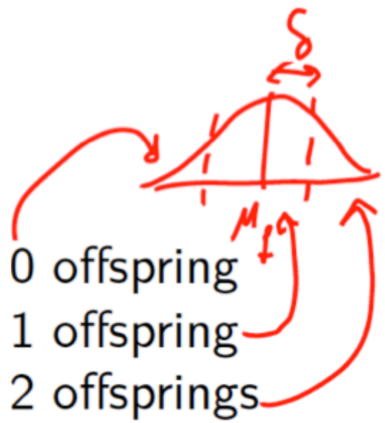
Roulette-wheel Selection: variance problem

- Selection of individuals is indeed proportional to the fitness, but random
- no guarantee that “fitter“ individuals are taken to the next generation, not even for the best individual
- high deviation **(high variance)** of the offspring of an individual
- very simple and recommendable solution is

  Discretization of the fitness range

  compute μ_f (t) and σ_f (t) of P

  – if $\mu_f (t) - \sigma_f (t) > f (x)$ then generate 0 offspring
  – if $\mu_f (t) - \sigma_f (t) \leq f (x) \leq \mu_f (t) + \sigma_f (t)$ then generate 1 offspring
  – if $f (x) > \mu_f (t) + \sigma_f (t)$ then generate 2 offsprings



Expected value model: Sol. of the variance

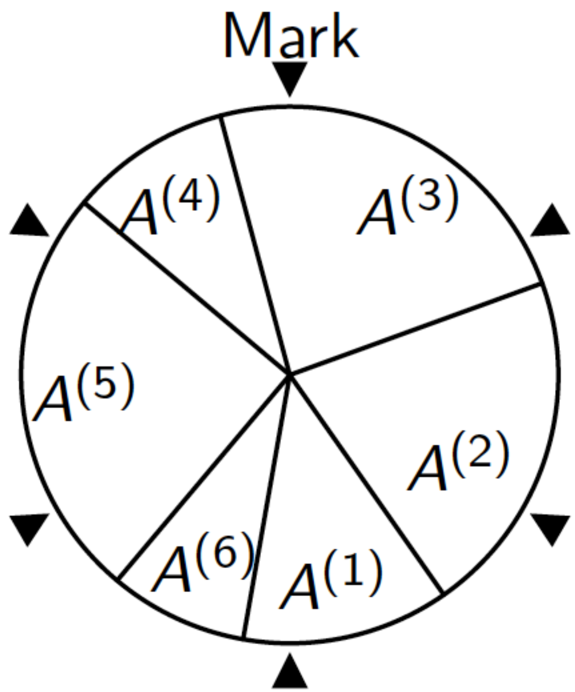
- generate $\lfloor f_{rel}(s) \cdot |P| \rfloor$ individuals for each solution candidate
- fill the (intermediary) population by Roulette-wheel selection

Alternative:

Stochastic Universal Sampling

Selection of the next population:

- Rotate Roulette-wheel once
- Choose one chromosome per mark
- Here:
  1 x $A^{(1)}$, 1 x $A^{(2)}$, 2 x $A^{(3)}$, 2 x $A^{(5)}$.
- Better-than-average individuals are taken into the next population definitely



Rank-based Selection

1. Sort individuals decendingly according to their fitness:
   Rank is assigned to each individual in population
2. Define prob. distribution over Rank scale:
   the lower the rank the lower the probability
3. Roulette-wheel selection based on the distribution

Adavantage:

- Avoidance of dominance problem: decoupling of fitness value and selection probability
- regulation of the selective pressure by prob. distribution on rank scale

Disadvantage: Sort of individuals (complexity: $|P| \cdot \log |P|$)

Tournament selection

1. Draw k individuals ($2 \leq k < |P|$) randomly from P(t) (selection without regarding the fitness, let k be the **tournament size**).
2. Individuals carry out the tournament and best indivual wins: Tournament winner receives a descendant in the next population
3. All participants (even the winner) of the tournament are returned to P(t)

Advantage:

- Avoidance of the dominance problem: decoupling of fitness value and selection probability
- regulation of the selective pressure by tournament size with limitations

Modification: f_{rel} of the participants determine winning probability (Roulette-wheel selection of an individual in tournament)

Elitism

best individual in next population:

- in general it is not ensured that the best individual enters the next generation
- no protection from modifications by genetic operators

⇒ fitness of the best individual can decrease from one generation to the next (= undesired)

Solution: Elitism

- unchanged transfer of the best individual (or the k, $1 \leq k < |P$ best individuals) into the next generation
- elite of a population never gets lost, hence elitism

Attention: elite is not exclued from normal selection: genetic operator can improve them