Computational Intelligence
Winter Term 2011/12

Prof. Dr. Günter Rudolph
Lehrstuhl für Algorithm Engineering (LS 11)
Fakultät für Informatik
TU Dortmund
Three tasks:
2. Choice / design of variation operators acting in problem representation.
3. Choice of strategy parameters (includes initialization).

ad 1) different “schools“:

(a) operate on binary representation and define genotype/phenotype mapping
   + can use standard algorithm
   – mapping may induce unintentional bias in search

(b) no doctrine: use “most natural” representation
   – must design variation operators for specific representation
   + if design done properly then no bias in search
ad 1a) genotype-phenotype mapping

original problem  \( f: X \to \mathbb{R}^d \)

scenario: no standard algorithm for search space \( X \) available

- standard EA performs variation on binary strings \( b \in \mathbb{B}^n \)
- fitness evaluation of individual \( b \) via \( (f \circ g)(b) = f(g(b)) \) where \( g: \mathbb{B}^n \to X \) is genotype-phenotype mapping
- selection operation independent from representation
Genotype-Phenotype-Mapping $\mathbb{B}^n \rightarrow [L, R] \subset \mathbb{R}$

- Standard encoding for $b \in \mathbb{B}^n$

$$x = L + \frac{R - L}{2^n - 1} \sum_{i=0}^{n-1} b_{n-i} 2^i$$

→ Problem: *hamming cliffs*

<table>
<thead>
<tr>
<th>Genotype</th>
<th>000</th>
<th>001</th>
<th>010</th>
<th>011</th>
<th>100</th>
<th>101</th>
<th>110</th>
<th>111</th>
</tr>
</thead>
<tbody>
<tr>
<td>phenotype</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
</tbody>
</table>

L = 0, R = 7

$n = 3$
Genotype-Phenotype-Mapping $\mathbb{B}^n \rightarrow [L, R] \subset \mathbb{R}$

- Gray encoding for $b \in \mathbb{B}^n$

Let $a \in \mathbb{B}^n$ standard encoded. Then $b_i = \begin{cases} a_i, & \text{if } i = 1 \\ a_{i-1} \oplus a_i, & \text{if } i > 1 \end{cases}$  

<table>
<thead>
<tr>
<th>$b_0$</th>
<th>$b_1$</th>
<th>$b_2$</th>
<th>$b_3$</th>
<th>$b_4$</th>
<th>$b_5$</th>
<th>$b_6$</th>
<th>$b_7$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

OK, no hamming cliffs any longer …

⇒ small changes in phenotype „lead to“ small changes in genotype

since we consider evolution in terms of Darwin (not Lamarck):

⇒ small changes in genotype lead to small changes in phenotype!

**but:** 1-Bit-change: $000 \rightarrow 100 \Rightarrow \oplus$
Design of Evolutionary Algorithms

Genotype-Phenotype-Mapping $\mathbb{B}^n \rightarrow \mathbb{P}^{\log(n)}$ (example only)

- e.g. standard encoding for $b \in \mathbb{B}^n$

**individual:**

```
  010 101 111 000 110 001 101 100
```

consider index and associated genotype entry as unit / record / struct;
sort units with respect to genotype value, old indices yield permutation:

```
  000 001 010 100 101 101 110 111
```

= permutation
ad 1a) genotype-phenotype mapping

typically required: strong causality
→ small changes in individual leads to small changes in fitness
→ small changes in genotype should lead to small changes in phenotype

**but:** how to find a genotype-phenotype mapping with that property?

**necessary conditions:**

1. \( g: \mathbb{B}^n \rightarrow X \) can be computed efficiently (otherwise it is senseless)
2. \( g: \mathbb{B}^n \rightarrow X \) is surjective (otherwise we might miss the optimal solution)
3. \( g: \mathbb{B}^n \rightarrow X \) preserves closeness (otherwise strong causality endangered)

Let \( d(\cdot, \cdot) \) be a metric on \( \mathbb{B}^n \) and \( d_X(\cdot, \cdot) \) be a metric on \( X \).

\[ \forall x, y, z \in \mathbb{B}^n : d(x, y) \leq d(x, z) \Rightarrow d_X(g(x), g(y)) \leq d_X(g(x), g(z)) \]
ad 1b) use “most natural“ representation

typically required: strong causality
→ small changes in individual leads to small changes in fitness
→ need variation operators that obey that requirement

**but:** how to find variation operators with that property?

⇒ need design guidelines ...
ad 2) design guidelines for variation operators

a) reachability
   every $x \in X$ should be reachable from arbitrary $x_0 \in X$
   after finite number of repeated variations with positive probability bounded from 0

b) unbiasedness
   unless having gathered knowledge about problem
   variation operator should not favor particular subsets of solutions
   $\Rightarrow$ formally: maximum entropy principle

c) control
   variation operator should have parameters affecting shape of distributions;
   known from theory: weaken variation strength when approaching optimum
Design of Evolutionary Algorithms

ad 2) **design guidelines for variation operators in practice**

**binary search space** $X = \mathbb{B}^n$

variation by k-point or uniform crossover and subsequent mutation

a) **reachability:**

regardless of the output of crossover
we can move from $x \in \mathbb{B}^n$ to $y \in \mathbb{B}^n$ in 1 step with probability

$$p(x, y) = p_m^H(x, y) \left(1 - p_m\right)^{n-H(x,y)} > 0$$

where $H(x,y)$ is Hamming distance between $x$ and $y$.

Since $\min\{p(x,y): x,y \in \mathbb{B}^n\} = \delta > 0$ we are done.
b) *unbiasedness*

don't prefer any direction or subset of points without reason

⇒ use maximum entropy distribution for sampling!

**properties:**
- distributes probability mass as uniform as possible
- additional knowledge can be included as constraints:
  → under given constraints sample as uniform as possible
Design of Evolutionary Algorithms

Formally:

Definition:
Let $X$ be discrete random variable (r.v.) with $p_k = P\{ X = x_k \}$ for some index set $K$. The quantity

$$H(X) = - \sum_{k \in K} p_k \log p_k$$

is called the entropy of the distribution of $X$. If $X$ is a continuous r.v. with p.d.f. $f_X(\cdot)$ then the entropy is given by

$$H(X) = - \int_{-\infty}^{\infty} f_X(x) \log f_X(x) \, dx$$

The distribution of a random variable $X$ for which $H(X)$ is maximal is termed a maximum entropy distribution.
Excursion: Maximum Entropy Distributions

Knowledge available:
Discrete distribution with support \{x_1, x_2, \ldots, x_n\} with \(x_1 < x_2 < \ldots < x_n < \infty\)

\[ p_k = P\{X = x_k\} \]

\[ \Rightarrow \text{leads to nonlinear constrained optimization problem:} \]

\[ - \sum_{k=1}^{n} p_k \log p_k \rightarrow \text{max!} \]

\[ \text{s.t.} \quad \sum_{k=1}^{n} p_k = 1 \]

solution: via Lagrange (find stationary point of Lagrangian function)

\[ L(p, a) = - \sum_{k=1}^{n} p_k \log p_k + a \left( \sum_{k=1}^{n} p_k - 1 \right) \]
Excursion: Maximum Entropy Distributions

\[ L(p, a) = - \sum_{k=1}^{n} p_k \log p_k + a \left( \sum_{k=1}^{n} p_k - 1 \right) \]

partial derivatives:

\[ \frac{\partial L(p, a)}{\partial p_k} = -1 - \log p_k + a \overset{!}{=} 0 \quad \Rightarrow \quad p_k \overset{!}{=} e^{a-1} \]

\[ \frac{\partial L(p, a)}{\partial a} = \sum_{k=1}^{n} p_k - 1 \overset{!}{=} 0 \]

\[ \Rightarrow \sum_{k=1}^{n} p_k = \sum_{k=1}^{n} e^{a-1} = n e^{a-1} \overset{!}{=} 1 \quad \Leftrightarrow \quad e^{a-1} = \frac{1}{n} \]

\[ p_k = \frac{1}{n} \quad \text{uniform distribution} \]
Excursion: Maximum Entropy Distributions

Knowledge available:

Discrete distribution with support \{ 1, 2, \ldots, n \} with \( p_k = P\{ X = k \} \) and \( E[X] = \nu \)

\[ - \sum_{k=1}^{n} p_k \log p_k \rightarrow \text{max!} \]

s.t.
\[ \sum_{k=1}^{n} p_k = 1 \quad \text{and} \quad \sum_{k=1}^{n} k p_k = \nu \]

solution: via Lagrange (find stationary point of Lagrangian function)

\[ L(p, a, b) = - \sum_{k=1}^{n} p_k \log p_k + a \left( \sum_{k=1}^{n} p_k - 1 \right) + b \left( \sum_{k=1}^{n} k \cdot p_k - \nu \right) \]
Excursion: Maximum Entropy Distributions

\[
L(p, a, b) = - \sum_{k=1}^{n} p_k \log p_k + a \left( \sum_{k=1}^{n} p_k - 1 \right) + b \left( \sum_{k=1}^{n} k \cdot p_k - \nu \right)
\]

partial derivatives:

\[
\frac{\partial L(p, a, b)}{\partial p_k} = -1 - \log p_k + a + b k \overset{!}{=} 0 \quad \Rightarrow \quad p_k = e^{a-1+b k}
\]

\[
\frac{\partial L(p, a, b)}{\partial a} = \sum_{k=1}^{n} p_k - 1 \overset{!}{=} 0
\]

\[
\frac{\partial L(p, a, b)}{\partial b} \overset{(*)}{=} \sum_{k=1}^{n} k p_k - \nu \overset{!}{=} 0 \quad \Rightarrow \quad \sum_{k=1}^{n} p_k = e^{a-1} \sum_{k=1}^{n} (e^b)^k \overset{!}{=} 1
\]

(continued on next slide)
Excursion: Maximum Entropy Distributions

\[ e^{a-1} = \frac{1}{\sum_{k=1}^{n} (ce^b)^k} \quad \Rightarrow \quad p_k = e^{a-1+b}k = \frac{(e^b)^k}{\sum_{i=1}^{n} (e^b)^i} \]

\[ \Rightarrow \quad \text{discrete Boltzmann distribution} \quad p_k = \frac{q^k}{\sum_{i=1}^{n} q^i} \quad (q = e^b) \]

\[ \Rightarrow \quad \text{value of } q \text{ depends on } \nu \text{ via third condition: } (*) \]

\[ \sum_{k=1}^{n} k p_k = \frac{\sum_{k=1}^{n} k q^k}{\sum_{i=1}^{n} q^i} = \frac{1 - (n + 1)q^n + nq^{n+1}}{(1-q)(1-q^n)} \quad \Rightarrow \quad \nu \]
Excursion: Maximum Entropy Distributions

Boltzmann distribution
(n = 9)

\( \nu = 2 \)

\( \nu = 3 \)

specializes to uniform distribution if \( \nu = 5 \)
(as expected)

\( \nu = 4 \)

\( \nu = 5 \)

\( \nu = 6 \)

\( \nu = 7 \)

\( \nu = 8 \)
Excursion: Maximum Entropy Distributions

Knowledge available:

Discrete distribution with support \( \{1, 2, \ldots, n\} \) with \( \mathbb{E}[X] = \nu \) and \( \mathbb{V}[X] = \eta^2 \)

\( \Rightarrow \) leads to nonlinear constrained optimization problem:

\[
- \sum_{k=1}^{n} p_k \log p_k \rightarrow \max!
\]

s.t. \( \sum_{k=1}^{n} p_k = 1 \) and \( \sum_{k=1}^{n} k p_k = \nu \) and \( \sum_{k=1}^{n} (k - \nu)^2 p_k = \eta^2 \)

Solution: in principle, via Lagrange (find stationary point of Lagrangian function)

but very complicated analytically, if possible at all

\( \Rightarrow \) consider special cases only

Note: constraints are linear equations in \( p_k \)
Excursion: Maximum Entropy Distributions

Special case: $n = 3$ and $E[X] = 2$ and $V[X] = \eta^2$

Linear constraints uniquely determine distribution:

1. $p_1 + p_2 + p_3 = 1$
2. $p_1 + 2p_2 + 3p_3 = 2$
3. $p_1 + 0 + p_3 = \eta^2$

II – I:

$\begin{align*}
p_2 + 2p_3 &= 1 \\
p_2 &= 1 - \eta^2
\end{align*}$

I – III:

$\begin{align*}
p_1 &= \frac{\eta^2}{2} \\
p_3 &= \frac{\eta^2}{2}
\end{align*}$

$\Rightarrow p = \left( \frac{\eta^2}{2}, 1 - \eta^2, \frac{\eta^2}{2} \right)$
Excursion: Maximum Entropy Distributions

Knowledge available:

Discrete distribution with unbounded support \{ 0, 1, 2, \ldots \} and \( E[X] = \nu \)

\[ - \sum_{k=0}^{\infty} p_k \log p_k \rightarrow \max! \]

s.t. \[ \sum_{k=0}^{\infty} p_k = 1 \quad \text{and} \quad \sum_{k=0}^{\infty} kp_k = \nu \]

solution: via Lagrange (find stationary point of Lagrangian function)

\[ L(p, a, b) = - \sum_{k=0}^{\infty} p_k \log p_k + a \left( \sum_{k=0}^{\infty} p_k - 1 \right) + b \left( \sum_{k=0}^{\infty} kp_k - \nu \right) \]
Excursion: Maximum Entropy Distributions

\[ L(p, a, b) = - \sum_{k=0}^{\infty} p_k \log p_k + a \left( \sum_{k=0}^{\infty} p_k - 1 \right) + b \left( \sum_{k=0}^{\infty} k \cdot p_k - \nu \right) \]

partial derivatives:

\[ \frac{\partial L(p, a, b)}{\partial p_k} = -1 - \log p_k + a + b k \overset{\parallel}{=} 0 \quad \Rightarrow \quad p_k = e^{a-1+b k} \]

\[ \frac{\partial L(p, a, b)}{\partial a} = \sum_{k=0}^{\infty} p_k - 1 \overset{\parallel}{=} 0 \]

\[ \frac{\partial L(p, a, b)}{\partial b} \overset{(*)}{=} \sum_{k=0}^{\infty} k p_k - \nu \overset{\parallel}{=} 0 \quad \Rightarrow \quad \sum_{k=0}^{\infty} p_k = e^{a-1} \sum_{k=0}^{\infty} \left(e^b\right)^k \overset{\parallel}{=} 1 \]

(continued on next slide)
Excursion: Maximum Entropy Distributions

\[ e^{a-1} = \frac{1}{\sum_{k=0}^{\infty} (e^b)^k} \quad \Rightarrow \quad p_k = e^{a-1+b} k = \frac{(e^b)^k}{\sum_{i=0}^{\infty} (e^b)^i} \]

set \( q = e^b \) and insists that \( q < 1 \) \( \Rightarrow \) \( \sum_{k=0}^{\infty} q^k = \frac{1}{1-q} \)

\( \Rightarrow \) \( p_k = (1-q) q^k \) for \( k = 0, 1, 2, \ldots \) geometrical distribution

it remains to specify \( q \); to proceed recall that \( \sum_{k=0}^{\infty} k q^k = \frac{q}{(1-q)^2} \)
Excursion: Maximum Entropy Distributions

⇒ value of $q$ depends on $\nu$ via third condition: $(\star)$

$$
\sum_{k=0}^{\infty} k p_k = \frac{\sum_{k=0}^{\infty} k q^k}{\sum_{i=0}^{\infty} q^i} = \frac{q}{1 - q} \equiv \nu
$$

⇒ $q = \frac{\nu}{\nu + 1} = 1 - \frac{1}{\nu + 1}$

⇒ $p_k = \frac{1}{\nu + 1} \left(1 - \frac{1}{\nu + 1}\right)^k$
Excursion: Maximum Entropy Distributions

geometrical distribution

with $E[x] = \nu$

$p_k$ only shown for $k = 0, 1, \ldots, 8$
Excursion: Maximum Entropy Distributions

Overview:

- support \{ 1, 2, \ldots, n \} \Rightarrow \textit{discrete uniform} distribution
- and require \( E[X] = \theta \) \Rightarrow \textit{Boltzmann} distribution
- and require \( V[X] = \eta^2 \) \Rightarrow \text{N.N. (not Binomial distribution)}

- support \( \mathbb{N} \) \Rightarrow \text{not defined!}
- and require \( E[X] = \theta \) \Rightarrow \textit{geometrical} distribution
- and require \( V[X] = \eta^2 \) \Rightarrow \text{?}

- support \( \mathbb{Z} \) \Rightarrow \text{not defined!}
- and require \( E[|X|] = \theta \) \Rightarrow \textit{bi-geometrical} distribution (\textit{discrete Laplace distr.})
- and require \( E[|X|^2] = \eta^2 \) \Rightarrow \text{N.N. (discrete Gaussian distr.)}
Excursion: Maximum Entropy Distributions

support $[a,b] \subseteq \mathbb{R}$ \ \Rightarrow \ uniform\ distribution

support $\mathbb{R}^+$ with $E[X] = \theta$ \ \Rightarrow \ Exponential\ distribution

support $\mathbb{R}$
with $E[X] = \theta$, $V[X] = \eta^2$ \ \Rightarrow \ normal/\ Gaussian\ distribution\ $N(\theta, \eta^2)$

support $\mathbb{R}^n$
with $E[X] = \theta$
and $Cov[X] = C$ \ \Rightarrow \ multinormal\ distribution\ $N(\theta, C)$

expectation\ vector $\in \mathbb{R}^n$

covariance\ matrix $\in \mathbb{R}^{n,n}$

positive\ definite: $\forall x \neq 0 : x'Cx > 0$
Excursion: Maximum Entropy Distributions

for permutation distributions?

→ uniform distribution on all possible permutations

\[
\text{set } v[j] = j \text{ for } j = 1, 2, \ldots, n \\
\text{for } i = n \text{ to } 1 \text{ step } -1 \\
\quad \text{draw } k \text{ uniformly at random from } \{1, 2, \ldots, i\} \\
\quad \text{swap } v[i] \text{ and } v[k] \\
\text{endfor}
\]

Guideline:

Only if you know something about the problem \textit{a priori} or

if you have learnt something about the problem \textit{during the search}

⇒ include that knowledge in search / mutation distribution (via constraints!)
ad 2) design guidelines for variation operators in practice

continuous search space $X = \mathbb{R}^n$

a) reachability
b) unbiasedness
c) control

leads to CMA-ES