# *Algorithmes Évolutionnaires (M2 MIAGE IA²)*

Andrea G. B. Tettamanzi
Laboratoire I3S – Équipe SPARKS
andrea.tettamanzi@univ-cotedazur.fr

univ-cotedazur.fr

# Séance 2
# Théorie

# Plan

- Les AE comme processus aléatoires
- Parallélisme implicite et théorème des schémas
- L’hypothèse des « blocs de construction »
- Quand la *fitness* est trompeuse
- Convergence

# Some notation

- Space of “genotypes” (individuals): $\Gamma$
- Space of “phenotypes” (candidate solutions): $\Phi$
- Decoding function: $M: \Gamma \to \Phi$
- Fitness function: $f: \Gamma \to [0, +\infty)$, $f(\gamma) = F[c(M(\gamma))]$
- For all $\gamma, \kappa \in \Gamma$, $f(\gamma) > f(\kappa)$ iff $c(M(\gamma)) < c(M(\kappa))$
- Space of populations: $\Gamma^*$
- The composition of a population is described by its share function $q: \Gamma \to [0, 1]$
- Share of individual $\gamma$ in the population: $q(\gamma)$

# EAs as Random Processes

$\left(\Gamma, 2^{\Gamma}, \mu\right)$ probability space

$x \in \Gamma^{(n)}$ a sample of size *n*

"random numbers"

$\left(\Omega, F, P\right)$

$\left(\Gamma, 2^{\Gamma}, \mu\right)$

trajectory

$\left\{ X_t(\omega) \right\}_{t=0,1,\ldots}$

evolutionary process

# The Schema Theorem

- Originally developed by Holland for simple GAs
  - Bit string representation
  - Uniform mutation
  - One-point crossover
  - Fitness-proportionate selection
- Historically the first justification of ***why*** EAs work
- Provides interesting insights into ***how*** EAs work
- Can be extended to other types of
  - Encoding
  - Genetic operators
  - Selection strategies

# Definition of Schema

- A *schema* is a subset S ⊆ Γ
- Represented by a template string
- Symbols in {0, 1, *}

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| * | * | * | 1 | 0 | * | 1 | * | * | * |

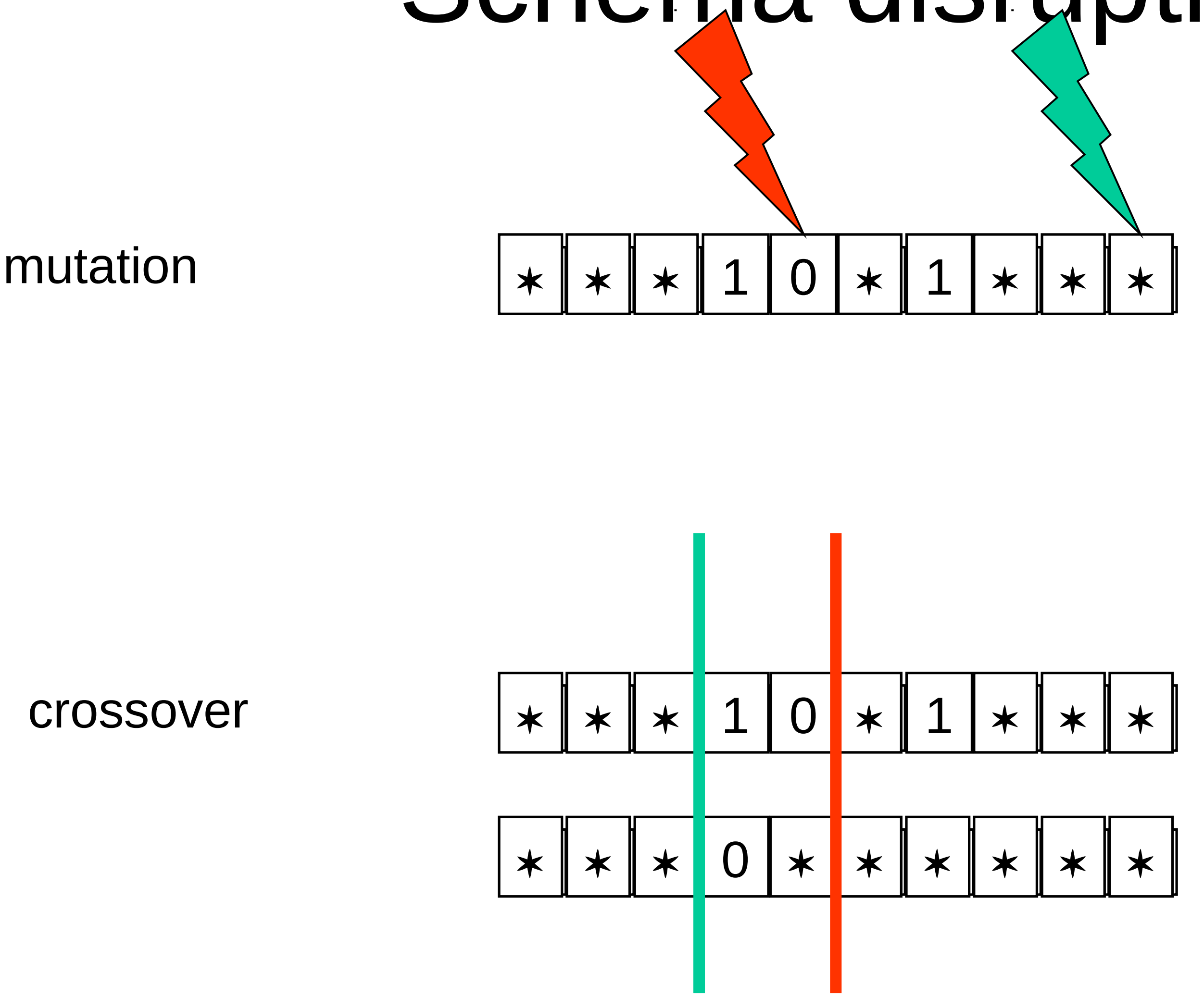
order of a schema: $o(S)$ = # fixed positions
defining length $\delta(S)$ = distance between first and last fixed position

# Properties of schemata

- A schema S matches $2^{l - o(S)}$ distinct strings
- A string of length $l$ is matched by $2^{l}$ schemata

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| * | * | * | 1 | 0 | * | 1 | * | * | * |

# Schema disruption



# “Implicit Parallelism”

In a population of $n$ individuals of length $l$

$2^l \leq$ # schemata processed $\leq n2^l$

$n^3$ of which are processed usefully (Holland 1989)

(i.e. they are not disrupted by crossover and mutation)

# Absolute fitness of a schema

- Expected fitness for any of the matched strings
- Cannot be directly observed by the GA

$$f(S) = \frac{1}{\|S\|} \sum_{\gamma \in S} f(\gamma)$$

# Relative fitness of a schema

- Expected fitness of an individual in a population *x*...
- ... Given that it belongs to a schema *S*

$$f_x(S) = \frac{1}{q_x(S)} \sum_{\gamma \in S} q_x(\gamma) f(\gamma)$$

# The Schema Theorem

$\{X_t\}_{t=0,1,...}$ populations at times *t*

suppose that $$\frac{f_{X_t}(S) - f(X_t)}{f(X_t)} = c$$ is constant

$$E[q_{X_t}(S) | X_0] \geq q_{X_0}(S)(1+c)^t \left(1 - p_{cross}\frac{\delta(S)}{l-1} - o(S)\, p_{mut}\right)^t$$

i.e. short, low-order above-average schemata increase exponentially!

# The Schema Theorem (proof)

$$E[q_{X_t}(S) | X_{t-1}] \geq q_{X_{t-1}}(S) \frac{f_{X_{t-1}}(S)}{f(X_{t-1})} P_{surv}[S] = q_{X_{t-1}}(S)(1+c) P_{surv}[S]$$

$$P_{surv}[S] = 1 - p_{cross} \frac{\delta(S)}{1-l} - p_{mut} o(S)$$

# Discussion of the Schema Theorem

- Hypothesis: ratio between schema fitness and average fitness of population is constant.
- This is only true “locally”, i.e., within few generations.
- Therefore, the exponential growth of above-average schemata happens only locally.
- However, the growth of high-fitness schemata has an effect on the average fitness in the population.
- Soon this effect compensates and balances the growth.

# Punctuated Equilibria

- Evolution proceeds in leaps and bounds
- Long periods of equilibrium
- Punctual events cause revolutionary changes
- Innovation
- Good match with observations of natural evolution

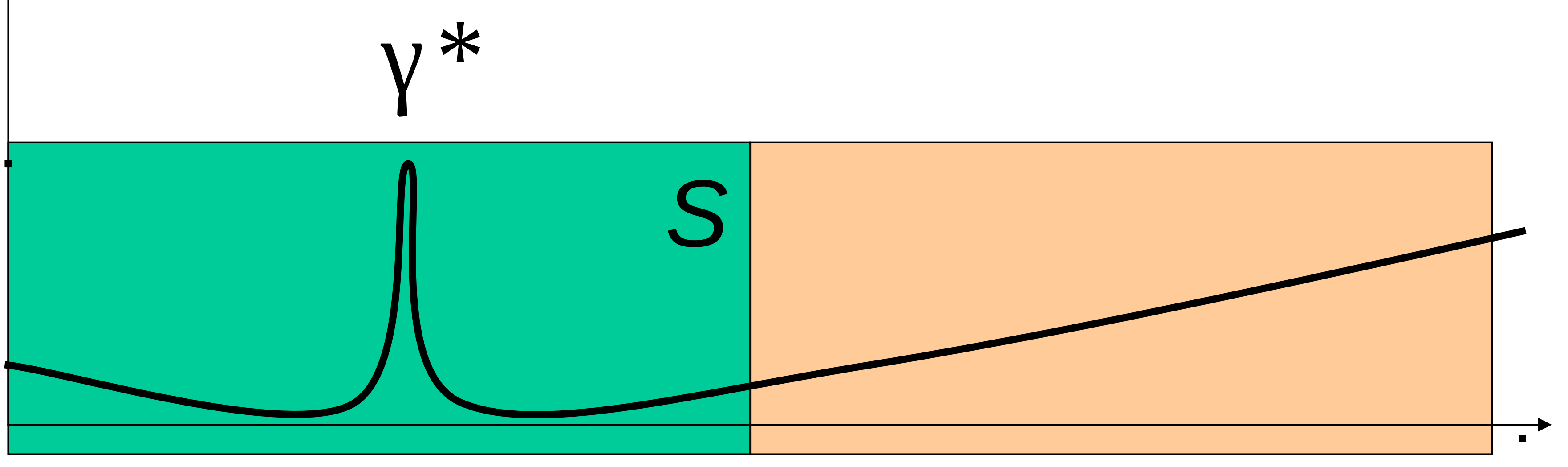
# The Building Blocks Hypothesis

‘‘An evolutionary algorithm seeks near-optimal performance through the juxtaposition of short, low-order, high-performance schemata — the building blocks’’

# Deception

i.e. when the building block hypothesis does not hold:

for some schema *S*, $\gamma^* \in S$ but $f(S) < f(\bar{S})$



# Deception: Example

$S_1 = 111{*}{*}{*}{*}{*}{*}{*}$

$S_2 = {*}{*}{*}{*}{*}{*}{*}{*}11$

$\gamma^* = 1111111111$

$S = 111{*}{*}{*}{*}{*}11$

$\bar{S} = 000{*}{*}{*}{*}{*}00$

$f(S_1) = 3$

$f(S_2) = 2$

$f(S) = 5$

$f(\bar{S}) = 10$

# Critical success factors

- **Solution encoding**
- **Fitness function**
- Prior knowledge of the problem (if available)
- Trial and error (otherwise)
- Rules of thumb
- Analogy with other problems

# Convergence properties of EAs

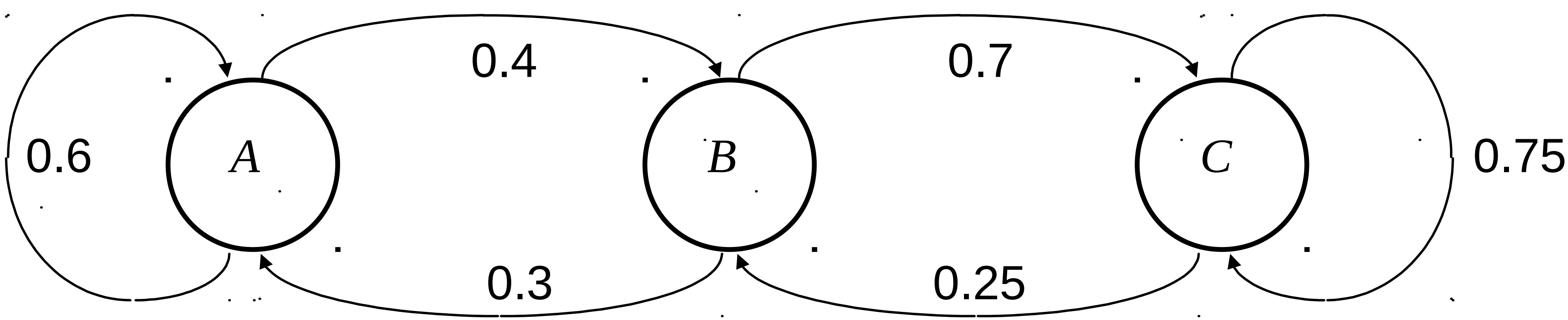
- Does my evolutionary algorithm converge to the optimal solution?
- How fast does my evolutionary algorithm find the optimal solution?

# Markov Chains

A stochastic process $\left\{ X_t(\omega) \right\}_{t=0,1,\ldots}$

is a Markov chain if and only if its state depends only on the previous state, i.e., for all t,

$$P[X_t = x | X_0, X_1, \ldots, X_{t-1}] = P[X_t = x | X_{t-1}]$$



# EAs as Random Processes

$(\Gamma, 2^{\Gamma}, \mu)$ probability space

$x \in \Gamma^{(n)}$ a sample of size *n*

"random numbers"

$(\Omega, \mathrm{F}, P)$

$(\Gamma, 2^{\Gamma}, \mu)$

trajectory

$\{X_t(\omega)\}_{t=0,1,\ldots}$

evolutionary process

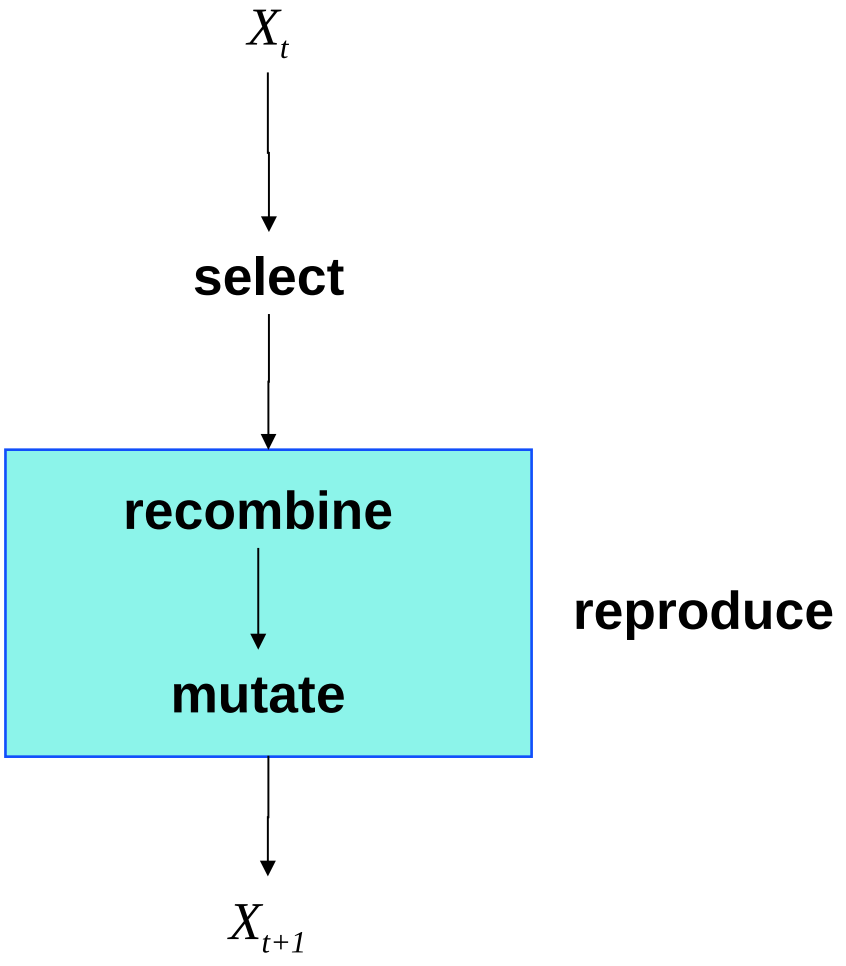
# Abstract Evolutionary Algorithm

Stochastic functions:

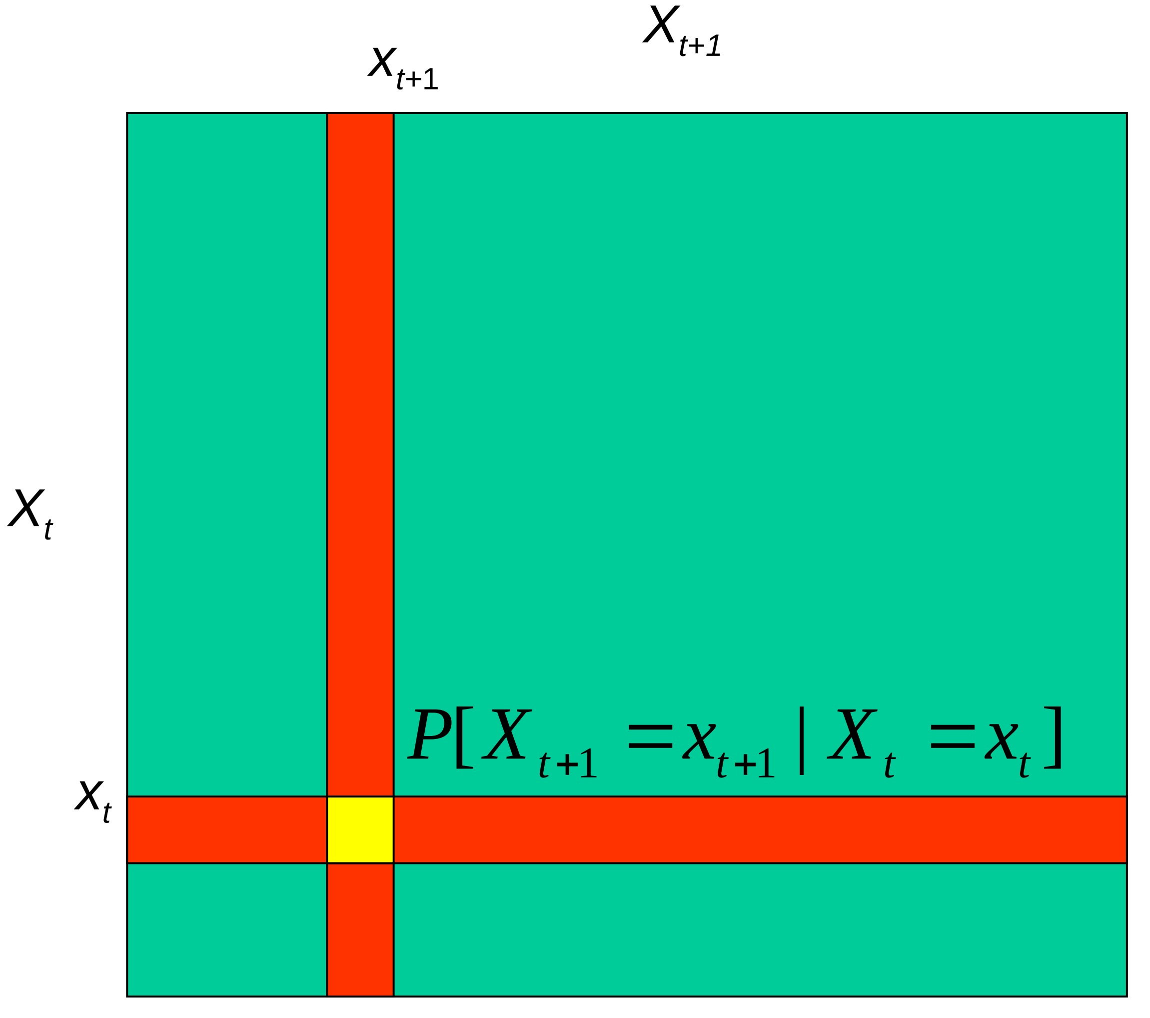
**select**: $\Gamma^{(n)} \times \Omega \rightarrow \Gamma$
**recombine**: $\Gamma \times \Gamma \times \Omega \rightarrow \Gamma$
**mutate**: $\Gamma \times \Omega \rightarrow \Gamma$
**reproduce**: $\Gamma \times \Omega \rightarrow \Gamma$

Transition function:

$$X_{t+1}(\omega) = T_t(\omega)\left(X_t(\omega)\right)$$



# Transition Matrix



# Convergence to Optimum

**Theorem:** if $\{X_t(\omega)\}_{t = 0, 1, \ldots}$ is *monotone*, *homogeneous*, $x_0$ is given, $\forall y$ in **reach**($x_0$) $\exists \gamma \in \Gamma^{(n)}_O$ reachable, then

$$\lim_{t \to \infty} P[X_t \in \Gamma_O^{(n)} \mid X_0 = x_0] = 1.$$

**Theorem:** if **select**, **mutate** are *generous*, the neighborhood structure is *connective*, transition functions $T_t(\omega)$, $t = 0, 1, \ldots$ are i.i.d. and *elitist*, then

$$\lim_{t \to \infty} P[X_t \in \Gamma_O^{(n)}] = 1.$$

# Speed of Convergence

- Still an open problem
- Very difficult
- Partial attempts
- Complexity classes