



# Représentations

Marc Schoenauer

Equipe TAO – INRIA Futurs et LRI

<http://tao.lri.fr/>

Cours

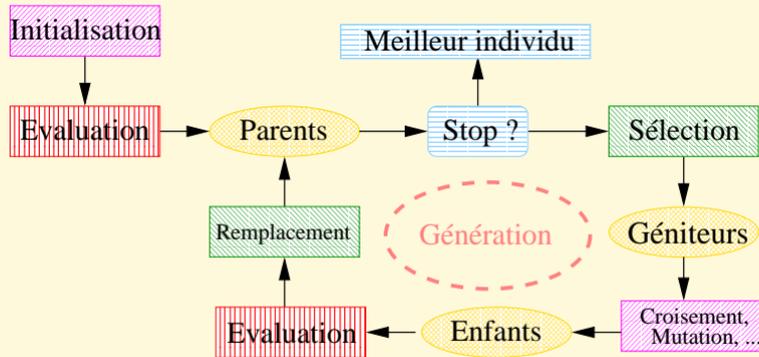
Optimisation par Evolution Artificielle  
et Application à la Robotique,  
le 14 décembre 2009

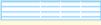


- Généralités Initialisation et opérateurs de variation
- Représentation binaire ... pour en finir avec la ...
  - Opérateurs usuels
  - Théorie des Schémas ... et sa discussion
  - Généralisation Opérateurs indépendants de la représentation
  - Etude de cas : les représentations des réels
- Représentation réelle Hors stratégies d'évolution
- Combinatorial Optimisation Generic vs specific operators
  - Direct representations hybridization
  - Memetic Algorithms
  - Indirect representations An example: train re-scheduling



# Algorithmes d'évolution : Le Squelette



-  Opérateurs stochastiques: Dépendent de la représentation
-  "Darwinisme" (stochastique ou déterministe)
-  Coût calcul
-  Critère d'arrêt, statistiques, ...



## Initialisation de la population

Choix de  $\Pi_0 = \{X_1, \dots, X_P\}$

- Par tirage uniforme dans  $\Omega$   $\Omega = \{0, 1\}^N$ ,  $X_i^j = 0$  ou  $1$  équiprobables  
 $\Omega = [0, 1]^N$ ,  $X_i^j = \text{random}()$
- ... mais attention au critère d'uniformité
- En tenant compte des connaissances a priori  
Ajoût de bonnes solutions manuelles  
Mais pas de biais vaut mieux qu'un mauvais biais
- Comme résultat d'une évolution précédente.  
Utilisation de plusieurs "milieux"  $\mathcal{F}_1, \mathcal{F}_2, \dots$

La diversité génétique est essentielle !

Opérateur d'exploration



## Opérateurs de variation

**Suivant l'arité** on définit

- opérateur unaire → **mutation**
- opérateur binaire → **crossover**    Peut modifier l'un des deux, ou les deux
- opérateur N-aire → **orgie**



## Le croisement

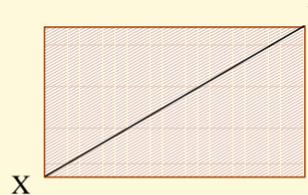
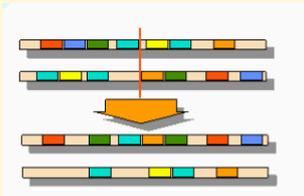
opérateur:  $\Omega \times \Omega \rightarrow \Omega$  (ou  $\Omega^2$ )

- **Analogie**: reproduction sexuée.
- **Intuition** :  
les enfants héritent des qualités de leurs deux parents...
- **Débat** : le croisement est  
L'opérateur majeur des AGs  
Un opérateur mineur (inutile) pour ES et EP
- **Recommandation** : **essayez !**  
l'intérêt dépend du problème  
notion de fragments de solution



# Opérateurs de variation : Le croisement

## Exemples classiques



Échange de **gènes** Croisement de paramètres réels

## Exemple orgiaque: Cinq parents

La foule subjuguée **boira** ses paroles enflammées  
 Ce plat **exquis** enchantait leurs papilles expertes  
 L'aube aux doigts de roses se leva sur un jour **nouveau**  
**Le cadavre** sanguinolent encombra la police nationale  
 Les coureurs assoiffés se jetèrent sur **le vin** pourtant mauvais  
**pour un enfant surréaliste**



## Croisement: discussion

### Propriétés:

- **Opérateur d'exploitation**
- Recombinaison des “bonnes” parties
- Effets destructeurs

### Choix du partenaire:

Aveugle en général, mais on peut introduire des préférences “sexuelles”

### Un exemple :

- Des parents sur des pics différents d'une fonction multi-modale donneront sans doute des enfants peu performants
- Croisement “restreint” : croiser  $X$  avec  $Y$  ssi  $d(\vec{X}, \vec{Y}) < \text{seuil}$
- Prise en compte des contraintes: your brain and my beauty



## La mutation

opérateur:  $\Omega \rightarrow \Omega$

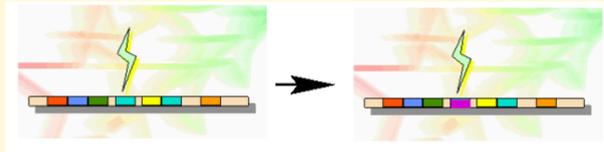
- **Analogie**: reproduction asexuée.
- **Intuition** : les enfants du croisement sont limités par  $\Pi_t$   
le seul contrepoids: la mutation
- **Grandes lignes** :  
L'enfant doit être en général **proche** du parent  
opérateur d'exploitation  
L'enfant doit pouvoir être **n'importe où**  
opérateur d'exploration – ergodicité
- **Quelle force?**  
Trop de mutations réussies (i.e. d'exploitation) mutations plus fortes  
sinon mutations plus faibles
- **Débat** : Un opérateur homéopathique des AGs  
L'opérateur majeur pour ES et EP



# Opérateurs de variation : La mutation

## Exemples classiques

- Mutation d'un gène



- Ajout de bruit Gaussien aux paramètres réels

## Un exemple sans queue ni tête

La terre est comme un orange **bleue**

La terre est **bleue** comme une orange



## Le débat croisement – mutation

### Le croisement:

- Effet dépend de la population  
Grandes modifications possibles si population diversifiée
- et donc de l'évolution  
Exploration au début, exploitation à la fin

Effet décroissant

### La mutation:

- Nécessaire
- Doit pouvoir faire de grands pas
- Mais effets destructeurs augmentent avec l'évolution

Ergodicité

opérateur d'exploration



## Le débat croisement – mutation (2)

### Efficacité empirique du croisement

- **Hypothèse constructive** :  
Assemblage de **briques de base**, i.e. pseudo-linéarité de la fonction fitness par rapport à des parties des individus
- **Hypothèse opportuniste** :  
Le croisement n'est qu'une **macro-mutation**

Headless-chicken crossover

### Besoin de mutation

- Ergodicité Résultats théoriques
- Relaxé en cas de **très grandes** population GP originel



## Reproduction

- Géniteurs (*mating pool*) → enfants ...
- Par application d'opérateurs de variation Généralement stochastiques
- Application des opérateurs
  - **Séquentielle**, avec une certaine **probabilité**  
AG historique
    - \* Croisement avec probabilité  $p_c$
    - \* Mutation avec probabilité  $p_m$  par individu
    - \* Exemple : avec  $p_c = 0.8$  et  $p_m = 0.3$ ,  
14% des individus sont inchangés
  - **Proportionnelle** selon des **poids** relatifs tirage de roulette  
GP historique: Croisement (6), Mutation(1), Copie(3)
  - ou toute combinaison



## Reproduction : le cas d'école

On dispose de

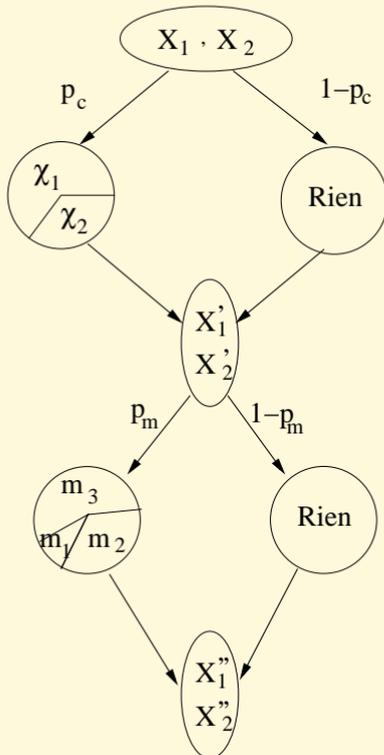
- 2 croisements  $\chi_1$  et  $\chi_2$ ,
- 3 mutations  $m_1, m_2, m_3$

On définit

- $p_c = 0.6$  et  $p_m = 0.3$ ,
- $pds_{\chi_1} = 2$  et  $pds_{\chi_2} = 1$
- $pds_{m_1} = 1, pds_{m_2} = 3$  et  $pds_{m_3} = 6$

**Exercice** : quel pourcentage

- subira  $\chi_2$  puis  $m_1$  ?
- subira seulement  $m_3$  ?
- ne sera pas modifié ?





- Généralités
  - Initialisation et opérateurs de variation
- Représentation binaire
  - ... pour en finir avec la ...
  - Opérateurs usuels
  - Théorie des Schémas
    - ... et sa discussion
  - Généralisation
    - Opérateurs indépendants de la représentation
  - Etude de cas : les représentations des réels
- Représentation réelle
  - Hors stratégies d'évolution
- Combinatorial Optimisation
  - Direct representations
    - Generic vs specific operators
  - Memetic Algorithms
    - hybridization
  - Indirect representations
    - An example: train re-scheduling



# Représentation “binaire”

J. Holland

$$\Omega = \{0, 1\}^n$$

- Règles **Si ... Alors** en logique des prédicats      **Systèmes de classeurs**
- Nombreux problèmes combinatoires      **SAT, sac-à-dos, ...**
- Nombres entiers      **... et même réels**

$$X \in [a, b] \rightarrow i_X = 2^n \frac{X-a}{b-a} \in [0, 2^n - 1]$$

**Pros :** Simplicité

Analogie avec les **chromosomes** (Croisements)  
quelques arguments théoriques

**Cons :** **No Free Lunch Theorem**

d'autres arguments théoriques !



## Le croisement “binaire”

$$\Omega \times \Omega \rightarrow \Omega \times \Omega$$

$\Omega = \{0, 1\}^N$  : Echange de bits entre les parents.

- Croisement à 1 point

J. Holland

$$\left. \begin{array}{l} (b_1, \dots, b_N) \\ (c_1, \dots, c_N) \end{array} \right\} \xrightarrow{p_c} \left\{ \begin{array}{l} (b_1, \dots, b_l, c_{l+1}, \dots, c_N) \\ (c_1, \dots, c_l, b_{l+1}, \dots, b_N) \end{array} \right.$$

- Croisement à 2 points

DeJong

$$\left. \begin{array}{l} (b_1, \dots, b_N) \\ (c_1, \dots, c_N) \end{array} \right\} \xrightarrow{p_c} \left\{ \begin{array}{l} (b_1, \dots, b_l, c_{l+1}, \dots, c_m, b_{m+1}, \dots, b_N) \\ (c_1, \dots, c_l, b_{l+1}, \dots, b_m, c_{m+1}, \dots, c_N) \end{array} \right.$$

- Croisement uniforme

Syswerda

Echanger les bits avec probabilité 0.5

indépendamment pour chaque position

**Croisement destructeur :**

Attention à l'influence du codage génotypique



## Les mutations binaires

$$\{0, 1\}^N \rightarrow \{0, 1\}^N$$

Opérateurs unaires :-)

### La mutation “bit-flip”

Holland, Goldberg

- Pour chaque individu
- et pour chaque position  $l$ ,

$$(b_1, b_2, \dots, b_N) \xrightarrow{p_m} (b_1, b_2, \dots, \bar{b}_l, b_{l+1}, \dots, b_N)$$

- Pour que la mutation soit effective
- Typiquement :  $p_m = \frac{1}{N}$

$$p_m > \frac{1}{N \times P}$$

En moyenne, 1 bit muté  
mais proportion  $\approx \frac{1}{e}$  d'individus non modifiés



## La mutation “déterministe”

Toujours inverser le même nombre de bits par individu

- Avec probabilité  $p_m$  **par individu**
- Choisir aléatoirement  $k_m$  positions
- Inverser les bits correspondants

## Laquelle choisir ?

- Mutation standard : “bit-flip”
- Mais mutation déterministe parfois plus efficace

Temps d'atteinte de la solution

## Quelle probabilité $p_m$ ?

- Trop de mutations réussies
- sinon

augmenter  $p_m$

diminuer  $p_m$



- Généralités Initialisation et opérateurs de variation
- Représentation binaire ... pour en finir avec la ...
  - Opérateurs usuels
  - **Théorie des Schémas** ... et sa discussion
  - Généralisation Opérateurs indépendants de la représentation
  - Etude de cas : les représentations des réels
- Représentation réelle Hors stratégies d'évolution
- Combinatorial Optimisation Generic vs specific operators
  - Direct representations hybridization
  - Memetic Algorithms
  - Indirect representations An example: train re-scheduling



# Algorithme Génétique Canonique

J. Holland – 75

- $E = \{0, 1\}^N$ , maximisation
  - Initialisation uniforme des  $X_i^0, i = 1, \dots, P$
  - Boucle des générations : Population  $X_i^t, i = 1, \dots, P$ 
    - **Évaluation** des  $\mathcal{F}(X_i^t), i = 1, \dots, P$
    - **Sélection** par Roulette  $\longrightarrow X_i', i = 1, \dots, P$
    - **Croisement** 1-point, avec probabilité  $p_c$ , de paires aléatoirement choisies  $(X_i', X_j')$   $\longrightarrow X_i'', i = 1, \dots, P$
    - **Mutation** : pour chaque bit  $b_i$  de chaque  $X_j''$ ,  
 $b_i \rightarrow \bar{b}_i$ , probabilité  $p_m$   $\longrightarrow X_i^{t+1}, i = 1, \dots, P$
- Remplacement générationnel

Ultérieurement, sélection par le rang et par tournoi, Croisement 2-points et uniformes



## Théorie des Schémas

**Schéma** : ss-ensemble de  $\{0, 1\}^N$ .

$$H = (0 * 1 * 1)$$

$o(H)$  : nb de bits fixés

$$o(H) = 3$$

$\ell(H)$  : dist max entre bits fixés

$$\ell(H) = 4$$

$N(H, t)$ : nb d'individus dans  $H \cap \Pi_t$ .

$\bar{\mathcal{F}}(H, t)$ : moyenne  $\mathcal{F}(X)$ ,  $X \in H \cap \Pi_t$ .

**Théorème des schémas** : Holland

Sélection par roulette

Croisement à un point  $p_c$

Mutation bit-flip  $p_m$  (**par bit**)

$$\mathbb{E}[N(H, t + 1)] \geq N(H, t) \frac{\bar{\mathcal{F}}(H, t)}{\bar{\mathcal{F}}(\Omega, t)} \left[ 1 - p_c \frac{\ell(H)}{N - 1} \right] (1 - p_m o(H))$$



## Théorie des Schémas (2)

- Sélection proportionnelle :  $\mathbb{E}[N(H, t + 1)] = N(H, t) \frac{\bar{\mathcal{F}}(H, t)}{\bar{\mathcal{F}}(\Omega, t)}$
- P[H détruit par croisement] =  $p_c \frac{\ell(H)}{N-1}$  Mais il peut en apparaître
- P[H préservé par mutation] =  $(1 - p_m)^{o(H)} \approx (1 - p_m)^{o(H)}$   
 $p_m \ll 1$

### Conclusion abusive :

pour des schémas courts

$\ell(H)$  petit, et donc  $o(H)$  petit

et performants

$$\frac{\bar{\mathcal{F}}(H)}{\bar{\mathcal{F}}(\Omega)} > 1$$

Croissance exponentielle ( $\exists \epsilon > 0$  tq  $\mathbb{E}[N(H, t + 1)] \geq (1 + \epsilon)^t$  ???)



## Dogme génétique

- Les schémas courts et performants (**building blocks**) croissent et se multiplient.
- **Parallélisme implicite** : en évaluant 1 individu, on évalue tous les schémas auquel il appartient  
 *$N^3$  schémas utiles dans la population Dorigo 93*
- Un individu appartient à  $2^N$  schémas différents  
*1 schémas d'ordre 1,  $N$  schémas d'ordre  $N$ , ...*
- soit  $(2/3)^N$  des schémas possibles.
- Pour un alphabet de taille  $K$ , il est de  $(\frac{2}{K+1})^N$
- Ce nombre est maximal pour  $K = 2$

Le codage binaire est donc le plus performant. CQFD



## Critique de la théorie des schémas

- La performance moyenne de la population croit aussi !
- Performance observée  $\neq$  performance moyenne  
dépend de la **variance** de la performance dans le schéma  
Radcliffe-Surry 91-95
- Et si les schémas les plus performants ne contiennent pas l'optimum?  
Notion de déception (Goldberg 89, Das-Whitley 93)
- L'argument de l'alphabet minimal est FAUX (Antonisse 89).
- ... et tous les codages parmi les  $2^N!$  possibles ne peuvent pas être tous aussi performants.



## Déception

Compétition entre schémas courts : et si le vainqueur ne contient pas l'optimum?

**Définition** :  $f$  déceptive (totalement) à l'ordre  $k$  si un (tous les) vainqueur(s) des compétitions entre schéma d'ordre  $k$  ne contient(nent) pas l'optimum.

**Exemple** (Whitley et al, 1990):

Valeurs:  $000=28, 001=26, 010=22, 011=0, 100=14, 101=0, 110=0, 111=30$

$f(0**) > f(1**), f(*0*) > f(*1*)$  et  $f(**0) > f(**1)$

et de même à l'ordre 2.

**Attention:**

- $f \equiv 0$  sauf  $f(1\dots 1) = 1$  n'est **pas** déceptive!
- Totalemment non déceptif  $\equiv$  variables séparées
- GAs  $\neq$  optimiseurs (DeJong 92).

Needle in the haystack



# Critique de l'alphabet minimal

Antonisse 89

## Comparaison alphabet binaire – alphabet de taille $K$

- Comparer à taille totale d'espace égale :  
Trouver  $L$  tq  $2^N = K^L$ .       $2^{20} = 10^6$ , mais on a encore  $(2/3)^{20} > (2/11)^6$
- Il y a plus d'un symbole "joker":  
 $k = 3$  : il faut considérer  $\#_{0,1}$ ,  $\#_{0,2}$ ,  $\#_{1,2}$  et  $\#_{0,1,2}$ .
- Les schémas de  $K^L$  sont donc **beaucoup** plus nombreux que  $(K + 1)^L$ , soit  $(2^K - 1)^L$  parties non vides de  $\{0, \dots, k - 1\}$  par position.
- Un individu appartient à  $2^{(K-1) \times L}$  schémas  
 $2^{K-1}$  "jokers" possibles par position.
- et représente donc  $(\frac{2^{(K-1)}}{2^{K-1}})^L \approx (1/2)^L$  des schémas possibles.
- Et  $(1/2)^L > (2/3)^N = (4/9)^{N/2}$  (car  $N > 2 * L$  dès que  $N > 3$ ).



- Généralités Initialisation et opérateurs de variation
- Représentation binaire ... pour en finir avec la ...
  - Opérateurs usuels
  - Théorie des Schémas ... et sa discussion
  - **Généralisation** Opérateurs indépendants de la représentation
  - Etude de cas : les représentations des réels
- Représentation réelle Hors stratégies d'évolution
- Combinatorial Optimisation Generic vs specific operators
  - Direct representations hybridization
  - Memetic Algorithms
  - Indirect representations An example: train re-scheduling



# Généralisation

## Radcliffe 91-95

### Définitions: Représentations

- **génomiques** = suite ordonnées de valeurs (ou **allèles**)
- **alléliques** = suite d'allèles.
- **orthogonale** : toute combinaison d'alleles est valide.
- **dégénée**: plusieurs génotypes représentent le même phénotype.
- **redondante**: présence d'information inutile.

### Exemple : Les chaînes de bits

Génomique orthogonale sans redondance ni dégénérescence.



## Théorie des Formae

Question: quelles **caractéristiques** du problème sont importantes?  
→ **formae**, ou ensemble de relations d'équivalence sur  $\Omega$ .

**Exemple** : Couleur des yeux et couleur des cheveux

Relations:

$\mathcal{R}_y$  avoir les yeux de même couleur que ...

$\mathcal{R}_c$  avoir les cheveux de même couleur que ...

$\mathcal{R}_{yc}$  avoir yeux et cheveux de même couleur que ...

Formae:

Yeux bleus – Yeux verts – Blonds – Roux ...

Blonds aux yeux bleus – Roux aux yeux bleus ... (**ordre** plus grand)

**Codage binaire**:

Mêmes bits  $l$  et  $m$  → les schémas usuels.

Même différence entre les bits  $l$  et  $m$  → généralisation aux prédicats quelconques.



## Propriétés désirées Pas toujours compatibles

### des Formae

- **Calculabilité** en temps raisonnable
- **Redondance minimale**
- **Fermeture** : l'intersection de 2 formae doit être une formae.
- **Corrélation** : les formae de grande précision doivent être constituées d'individus de performance proches.

### des Opérateurs (mutation)

- **Ergodicité** :  
Tout point doit pouvoir être atteint par un nombre fini d'opérateurs de variation.



## Propriétés désirées (2)

### des Opérateurs (croisement)

- **Respect** :  
si  $X$  et  $Y$  appartiennent à  $\mathcal{H}$ ,  
 $\Xi(X, Y)$  appartient à  $\mathcal{H}$ .
- **Assortiment constructif** :  
si  $X$  est dans  $\mathcal{H}_1$  et  $Y$  dans  $\mathcal{H}_2$   
et  $\mathcal{H} = \mathcal{H}_1 \cap \mathcal{H}_2$  est non vide,  
il existe  $\Xi(X, Y)$  dans  $\mathcal{H}$ .
- **Transmission** : les enfants ne contiennent que des allèles présentes  
chez les parents.

Permet de définir des **opérateurs indépendants de la représentation**



## Croisements

Choix de la propriété à respecter:

- **$R^3$  (Random Respectful Recombination)**

Toutes les allèles communes + complément aléatoire.

2 parents blonds  $\Rightarrow$  enfant blond  
 $\equiv$  uniforme pour les bitstrings

- **RAR (Random Assorting Recombination)**

Choix parmi toutes les allèles des parents + complément aléatoire.

Blond  $\times$  Yeux-bleus peut donner blond aux yeux bleus  
 $\equiv$  uniforme pour les bitstrings

Voir aussi **GNX (Generalized N-point Xover)**

- **RTR (Random Transmitting Recombination)**

Choix parmi toutes les combinaisons valides d'allèles venant des parents

Un blond et un roux ne peuvent pas donner un brun  
 $\equiv$  uniforme pour les bitstrings



# Mutation binomiale minimale – BMM

## Hypothèses

- même nombre  $N$  d'allèle.  
 $d(x,y) = N - \#(\text{allèles communes})$ .
- on peut identifier des mutation minimales :  
si  $\mu(X) = Y$  il n'existe pas de  $Z$  avec  $\mu(X) = Z$  et  $d(Z, X) < d(Y, X)$
- une suite de mutation minimales permet de joindre deux points queconques.

## Mutation BMM:

- paramètre  $p_m \in [0, 1]$
- nombre de mutations minimales tiré selon la loi binomiale  
 $B(n, p_m)$   $P[B(N, p_m) = k] = C_N^k p_m^k (1 - p_m)^{N-k}$
- Bitstring : mutation standard (modulo les retours en arrière).



- Généralités Initialisation et opérateurs de variation
- Représentation binaire ... pour en finir avec la ...
  - Opérateurs usuels
  - Théorie des Schémas ... et sa discussion
  - Généralisation Opérateurs indépendants de la représentation
  - **Etude de cas : les représentations des réels**
- Représentation réelle Hors stratégies d'évolution
- Combinatorial Optimisation Generic vs specific operators
  - Direct representations hybridization
  - Memetic Algorithms
  - Indirect representations An example: train re-scheduling



## Représentation des réels

### Représentation binaire

$$\text{Réal} \rightarrow \text{Entier} \rightarrow \text{Chaîne de bits}$$
$$X \in [a, b] \rightarrow i_X = 2^n \frac{X-a}{b-a} \in [0, 2^n - 1]$$

### Problèmes

- Discrétisation
- Création d'optima locaux
- Manque de continuité, Falaises de Hamming

Passer de 7 (0111) à 8(1000) par mutation est "impossible"

- Croisement binaire viole le respect et l'assortiment constructif:

$$8(1000) \times 7(0111) \rightarrow [0,15]$$

$$8(1000) \times 0(0000) \rightarrow \{8,0\}$$



## Codage Gray

- Deux entiers consécutifs ne diffèrent que d'un bit.  
Si  $x \in \{0, 1\}^N$  est le codage binaire de  $i$ ,  
 $x^T \hat{G}$  est **un** code de Gray pour  $i$ , avec

$$O^+ = \begin{pmatrix} 1 & 1 & 0 & \dots & 0 \\ 0 & 1 & 1 & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix}$$

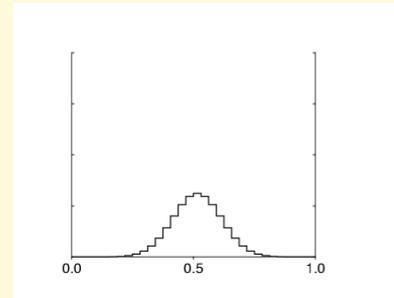
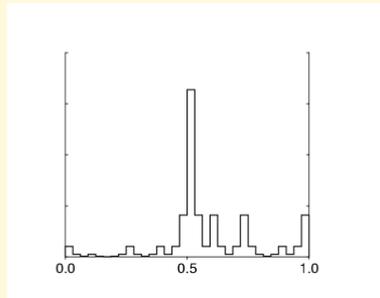
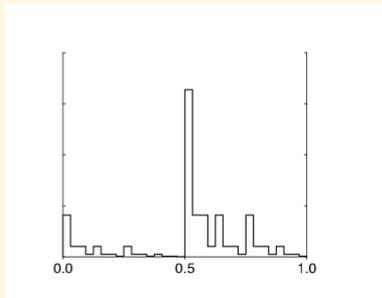
Vrai également avec toute permutation de colonnes de  $G$  (de l'ordre de  $N!2^N$  codes différents).

- Moins d'optima locaux (Whitley 97).
- Une mutation d'un bit reste "incontrôlée".



# Mutations pour des codages binaires des réels

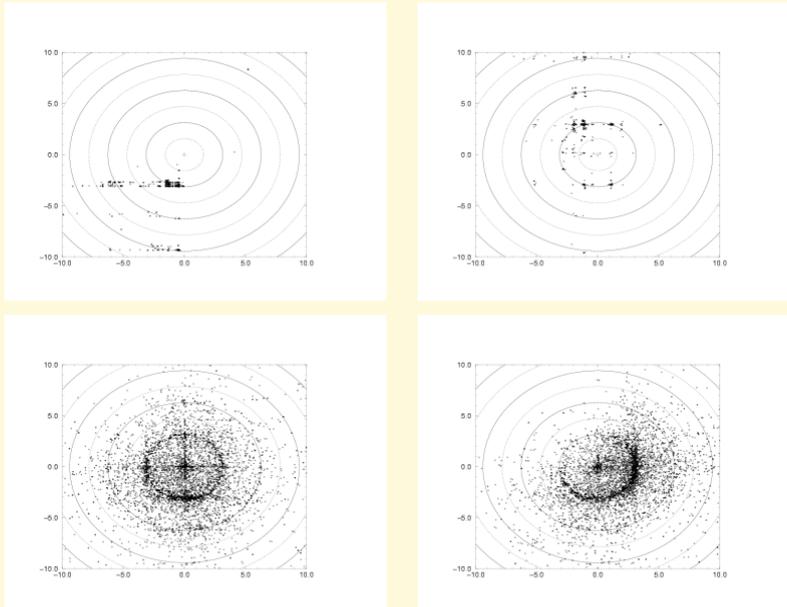
## Surry & Radcliffe 96



[0,1] discrétisé sur 5 bits – loi de distribution des enfants du point 0.5+



# Un run (9000 points) – Fonction “F6” de Schaffer Surry & Radcliffe 96



(a) : Représentation binaire – (b) : Représentation Gray

(c) : Représentation Dedekind – (d) Représentation IsoDedekind

Optima locaux sur chacun des cercles concentriques, optimum global en (0,0).



## Tentatives d'amélioration (2)

### **Mutation variable**

Faire dépendre la probabilité de mutation de la position des bits

Fogarty 89

### **Dynamic Parameter Encoding**

Zoom autour de la meilleure solution trouvée

Whitley 91



## Vers la représentation réelle

### Important

- Localité (ou continuité, ou causalité forte des ES)
- Interval-schemata (Eshelman & Schaffer 93)
- Forma possible:  $I_{x,\tau} = ]x - \tau, x + \tau[$

### Etude de la variance

$f$  Lipschitz

$$\begin{aligned}
 \sigma^2(t, \tau) &= \frac{1}{\tau} \int_{t-\tau/2}^{t+\tau/2} (f(x) - \bar{f}(t, \tau))^2 dx \\
 &\leq \frac{1}{\tau} \int_{t-\tau/2}^{t+\tau/2} (f(x) - f(x_\tau))^2 dx \quad \text{pour un } x_\tau \in [t - \tau/2, t + \tau/2] \\
 &\leq \frac{1}{\tau} \int_{t-\tau/2}^{t+\tau/2} \alpha^2 |x - x_\tau|^2 dx \quad \leq \frac{\alpha^2 \tau^2}{3}
 \end{aligned}$$



## Approche par Forma

- **Relations d'équivalence** de base sur  $[0, 2^N - 1]$   
 $xR_iy$  si  $(x, y) < i$  ou  $\geq i$
- Les classes d'équivalence de base sont les coupures  
 $\xi_i^0 = \{0, \dots, i - 1\}$   
 $\xi_i^1 = \{i, \dots, N - 1\}$   
et leurs intersections (les intervalles).
- **Représentation de Dedekind**  
sur  $2^N$  bits (extrêmement redondante).

$$i \rightarrow 00 \dots 0 \overset{i}{1} \dots 11$$

**Ne pas tenter d'implémenter!**



## Passage à la limite Mutation

- La mutation minimale est un pas de longueur  $\Delta = \frac{b-a}{2^N}$
- On en tire  $S = B(\ell_N, p_m)$ ,  $\ell_N$  facteur d'échelle à déterminer

Etude asymptotique de la V.A.  $X$  (déplacement total),  
pour  $p_m$  fixé et  $N$  (et  $\ell_N$ ) tendant vers  $+\infty$ ,

- $S \approx N(\ell_N p_m, \sqrt{\ell_N p_m (1 - p_m)})$   $\ell_N p_m \rightarrow +\infty$
- Pour  $S$  fixé, marche aléatoire 1D de pas  $\Delta$   
 $X|S \approx N(0, \sqrt{s}\Delta)$
- et la fonction génératrice de  $X|S$  est  
 $\mathcal{M}_{X|S}(t) = \mathbb{E}[e^{t(X|S)}] = \exp(s\Delta^2 t^2 / 2)$



## Passage à la limite Mutation (2)

- Loi de  $X$  :  $p(x) = \int_{-\infty}^{+\infty} p(x|s)p(s)ds$

- Fonction génératrice de  $X$  :

$$\begin{aligned}\mathcal{M}_X(t) = \mathbb{E}[e^{tX}] &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p(x|s)p(s)e^{tx} ds dx \\ &= \int_{-\infty}^{+\infty} p(s) \int_{-\infty}^{+\infty} p(x|s)e^{tx} dx ds \\ &= \int_{-\infty}^{+\infty} p(s)\mathcal{M}_{X|S}(t) ds \\ &= \int_{-\infty}^{+\infty} p(s)\exp(s\Delta^2 t^2/2) ds \\ &= \mathcal{M}_S(\Delta^2 t^2/2)\end{aligned}$$



## Passage à la limite Mutation (3)

Or  $\mathcal{M}_S(t) \approx \exp(\ell_N p_m t + \ell_N p_m (1 - p_m) t^2 / 2)$

Et donc

$$\mathcal{M}_X(t) = \exp\left(\frac{\ell_N p_m (b - a)^2}{2 \cdot 2^{2N}} t^2 + \frac{\ell_N p_m (1 - p_m) (b - a)^4}{8 \cdot 2^{4N}} t^4\right)$$

En prenant maintenant  $\ell_N = 2^{2N}$  il vient

$$\lim_{N \rightarrow \infty} \mathcal{M}_X(t) = \exp(p_m (b - a)^2 t^2 / 2)$$

D'où finalement

$$BMM \rightarrow N(0, \sqrt{p_m} (b - a))$$



# Passage à la limite

## Croisements

### En dimension 1

Choisir les enfants uniformément dans l'intersection de toutes les formes auxquelles appartiennent les 2 parents – i.e. le segment les joignant:

Croisement barycentrique (ou arithmétique)

$$(\vec{X}, \vec{Y}) \rightarrow \alpha \vec{X} + (1 - \alpha) \vec{Y}$$

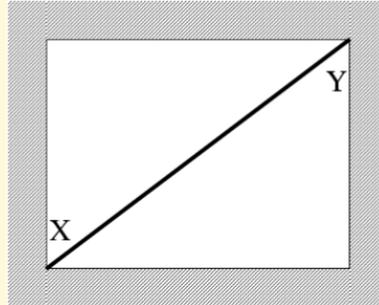


## Passage à la limite

### Croisements (2)

En dimension  $n$

Dedekind sur chaque coordonnée



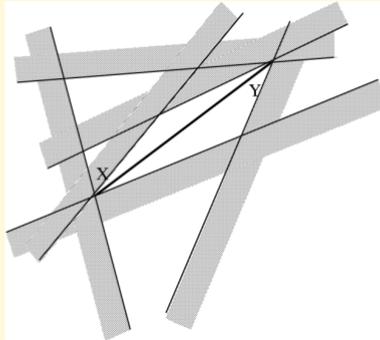
Croisement barycentrique (ou arithmétique) généralisé :

$$(\vec{X}, \vec{Y}) \rightarrow \alpha_i X_i + (1 - \alpha_i) Y_i$$



## Passage à la limite Croisements (3)

IsoDedekind, les forma de base sont les demi-espaces



Croisement barycentrique (ou arithmétique) :

$$(\vec{X}, \vec{Y}) \rightarrow \alpha \vec{X} + (1 - \alpha) \vec{Y}$$



- Généralités Initialisation et opérateurs de variation
- Représentation binaire ... pour en finir avec la ...
  - Opérateurs usuels
  - Théorie des Schémas ... et sa discussion
  - Généralisation Opérateurs indépendants de la représentation
  - Etude de cas : les représentations des réels
- Représentation réelle Hors stratégies d'évolution
- Combinatorial Optimisation Generic vs specific operators
  - Direct representations hybridization
  - Memetic Algorithms
  - Indirect representations An example: train re-scheduling



## Représentation “réelle

Rechenberg & Schwefel, Michalewicz, Radcliffe, Fogel, ...

$$\Omega \subset \mathbb{R}^n$$

- Historiquement, **Real Coded Genetic Algorithms**
- Aujourd’hui, utiliser CMA-ES, mais
- Ce n’est pas toujours possible
  - Espace de recherche mixte (ou pire !)      *Adaptation souvent peu efficace*
  - Multi-objectif      *mais un CMA-ES multi-objectif est récemment apparu*



## Croisement réel

$$\Omega = \mathbb{R}^N \text{ or } \prod [a_i, b_i]$$

### Croisement vectoriel

Le croisement “binaire” peut être rendu **générique** pour tout type de “vecteur”, échangeant les variables correspondantes

**Exemple:** croisement ES “discret”  $\equiv$  croisement uniforme  
 $\equiv$  échange aléatoire de variables

### Croisement global

ES historiques

- Pour chaque variable, on choisit un “partenaire” différent
- Intermédiaire ou discret

### Invariance par rotation

Ces croisements ne sont **pas** invariants par rotation  
→ peu performants sur fonctions non séparables



## Croisement réel (2)

### Croisement arithmétique, barycentrique, intermédiaire, ...

Eshelman-Schaffer (BLX- $\alpha$ ), Michalewicz, Radcliffe ...

**Idée de base** : combinaison linéaire des parents

- Coordonnée par coordonnée

Rotations ...

$$(\vec{X}, \vec{Y}) \rightarrow (\alpha_i \vec{X}_i + (1 - \alpha_i) \vec{Y}_i)$$

avec  $\alpha_i = U([0, 1])$  VA indépendantes.

- Globalement

$$(\vec{X}, \vec{Y}) \rightarrow \alpha \vec{X} + (1 - \alpha) \vec{Y}$$

avec  $\alpha = U([0, 1])$

- Application contractante (perte de diversité)

Extension :  $\alpha_i = U([-0.5, 1.5])$ .

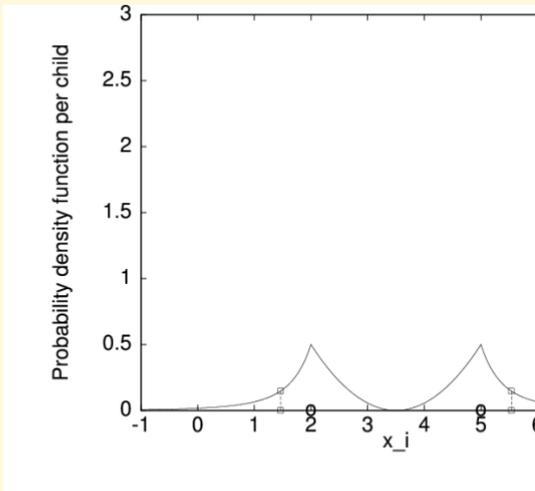
Attention aux bornes des variables



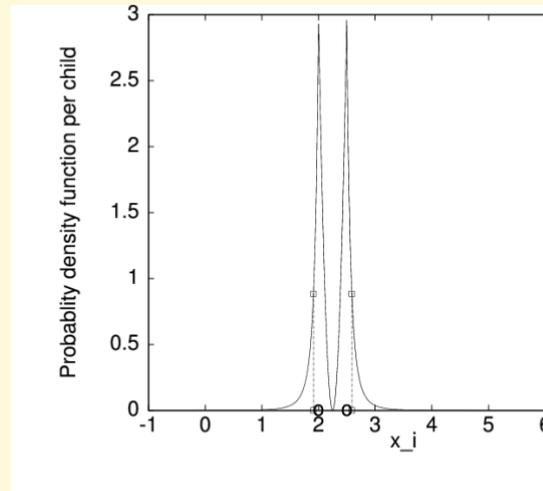
## Croisement réel (3)

Simulated Binary Crossover (Deb et al., 97)

$$\left\{ \begin{array}{l} \beta = (2u)^{\frac{1}{\eta+1}} \text{ si } u = U([0, 1]) < 0.5, \left(\frac{1}{2(1-u)}\right)^{\frac{1}{\eta+1}} \text{ sinon} \\ x_1 \leftarrow 0.5[(1 + \beta)x_1 + (1 - \beta)x_2] \\ x_2 \leftarrow 0.5[(1 - \beta)x_1 + (1 + \beta)x_2] \end{array} \right.$$



Parents distants

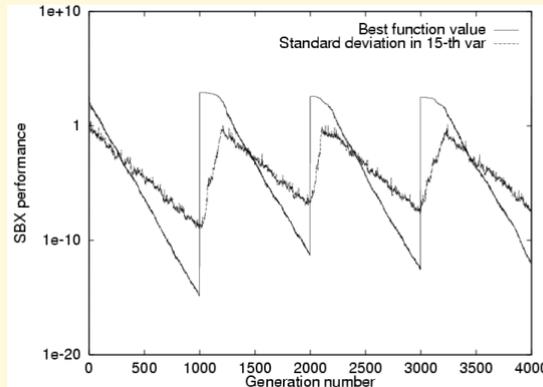


Parents proches



## Variance des variables dans la population

Fonction "ellipsoïde" dynamique (le minimum se déplace toutes les  $K$  générations)



Fitness et  $var(x_{15})$  pour SBX

- Mais peu performant sur fonctions non séparables
- Voir aussi PCX (Parent Centric Crossover)



## Mutations réelles

$$\mathbb{R}^N \rightarrow \mathbb{R}^N$$

**Idée:** ajout de bruit Gaussien

Tout l'art est dans le choix de  $\sigma$ !

$$X_i := X_i + N(0, \sigma)$$

- Basé sur la fitness (type EP)
- Basé sur l'historique
- Mutations adaptatives (ES modernes)
- Covariance Matrix Adaptation

**Rechenberg, règle des 1/5**

$\sigma$  ajustée par l'évolution !

State-of-the-art



## Autres mutations réelles

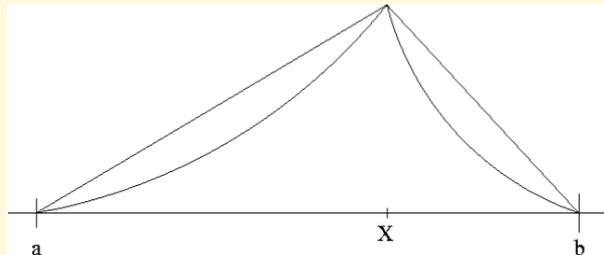
- Mutation Gaussienne, avec  $\sigma$  décroissant

Difficile à régler

- Mutation uniforme (Michalewicz)

$$x_i = U([x_i - \varepsilon, x_i + \varepsilon])$$

- Mutation “non-uniforme” (Michalewicz)



$$\begin{cases} x + \Delta(t, b - x) & \text{if } (U([0, 1]) > 0.5) \\ x - \Delta(t, x - a) & \text{otherwise} \end{cases} \quad \text{where } \Delta(t, y) = y(1 - r^{1 - \frac{t}{T}})$$



- Généralités Initialisation et opérateurs de variation
- Représentation binaire ... pour en finir avec la ...
  - Opérateurs usuels
  - Théorie des Schémas ... et sa discussion
  - Généralisation Opérateurs indépendants de la représentation
  - Etude de cas : les représentations des réels
- Représentation réelle Hors stratégies d'évolution
- **Combinatorial Optimisation**
  - Direct representations Generic vs specific operators
  - Memetic Algorithms hybridization
  - Indirect representations An example: train re-scheduling



# Combinatorial Optimisation

- NP-hard problems      Search space of exponential size w.r.t. some parameter
- Typically, search for a permutation of  $[1, N]$
- **Examples:** Travelling Salesman (TSP), Graph Coloring, Constraint Satisfaction (SAT), . . .

and many real-world problems

Standard Evolutionary Algorithms are **not efficient**

Handbook of Evolutionary Computation 94-97

A visionary paper:

J. J. Grefenstette. Incorporating problem specific knowledge in genetic algorithms. In Davis L., editor, *Genetic Algorithms and Simulated Annealing*, pages 42-60. Morgan Kaufmann, (1987).



## Historical of Evolutionary TSP

- Specific Crossover  
Grefenstette, 87 (20%), Suh & van Gucht, 89 (1.7, then 1.3 %, 100 cities)
- *Genetic Edge Recombination*  
Matthias & Whitley, 92 (3.8%), Tang & Leung, 94 (0.95%, 532 cities)
- Changing the representation  
Radcliffe, 91-05
- Lamarckism through OR techniques  
Mühlenbein, 91, modified by Ulder, 91 (0.17%, 532 cities)
- Genetic Local Search / Memetic Algorithms  
Merz & Freisleben, 96 (0%, 15 000 cities)
- Inver-Over operator  
and EAX crossover  
Tao & Michalewicz, 98  
Tagata, 97-06 (0%, 24 000 cities)



- Généralités Initialisation et opérateurs de variation
- Représentation binaire ... pour en finir avec la ...
  - Opérateurs usuels
  - Théorie des Schémas ... et sa discussion
  - Généralisation Opérateurs indépendants de la représentation
  - Etude de cas : les représentations des réels
- Représentation réelle Hors stratégies d'évolution
- Combinatorial Optimisation Generic vs specific operators
  - **Direct representations** hybridization
  - Memetic Algorithms
  - Indirect representations An example: train re-scheduling



## Evolving permutations

### Crossovers

- **UOX** Uniform Order-based Crossover Davis, 85  
uniformly  
Copy some positions from parent 1  
Complete using order from parent 2
- **PMX** Partially Matched Crossover Goldberg & Lingle, 85  
Choose a substring  $S$  from parent 1  
Perform minimal modifications on parent 2 so  $S$  appears

### Mutations

- **Transposition** : exchange 2 values. Possibly within a limited radius  
repeat  $N$  times Radcliffe's BMM
- **Inversion** : choose 2 positions, inverse the substring
- **Scramble** : choose 2 positions, shuffle the substring



## Different representations for the TSP

Radcliffe, 91-95

- **Permutation** : (1234)

Degeneracy : starting point (easy to fix),  
and direction (neglected).

Redundancy (amount of useless information): last city.

- **Undirected edges** : (12,23,34,14).

No degeneracy, redundancy : last edge.

- **Directed edges** : ( $\vec{12}$ ,  $\vec{23}$ ,  $\vec{34}$ ,  $\vec{41}$ )

Degeneracy : direction, redundancy : last edge.

- **Coins** : (1(24),2(13),3(24),4(13))

No degeneracy, large redundancy.



## Representation-independent operators

### Crossovers

- Exchange of “genes”
- Random complement (constraint-wise)
- ... or repair

1-pt, uniform, ...

### Binomial Minimal Mutation

- Repeat  $B(n, p)$  times a minimal mutation

e.g. transposition

### Correlation

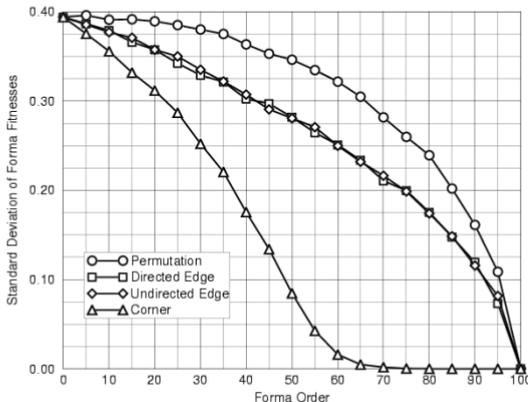
- Schema  $\equiv$  share some common genes
- Fitness variance w.r.t. precision  
Taking redundancy into account (corner)

number of individuals in schema

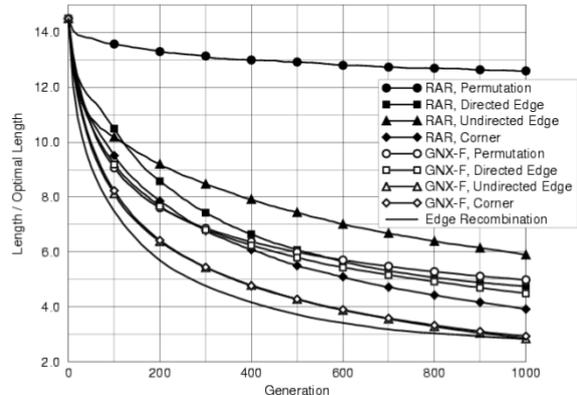


# Correlation

## Experimental results



Variance of fitness vs precision



Performance of corresponding EAs  
(representation-independent operators)

## Conclusions

- Variance is a good predictor of EA performance, but ...
- outperformed by Genetic Edge Recombination



# Genetic Edge Recombination

Whitley et al., 89

- Build the map of links

For each city, list of towns linked to it in at least one of both parents

- Choose a starting city

Random, or as in 3

1. Remove current city from link map

2. If current city has no more city in its list,

- If unvisited cities, choose one with least links

Goto 1

- Else, STOP

3. Choose, in cities linked to current city, one with least links

Goto 1

Tour is complete, and has many of both parents' edges



## Inver-over operator

Tao & Michalewicz, 98

For each  $T = (v_0, \dots, v_n)$  in the population

- Choose  $v$  uniformly in  $T$
- While  $v' \neq Succ_T(v)$ 
  - With proba.  $p$ 
    - Choose  $v' \neq v$  uniformly in  $T$
    - Inverse  $v - v'$
  - With proba  $1 - p$ 
    - Choose  $T'$  uniformly in the population
    - $v' = Succ_{T'}(v)$
    - Inverse  $v - v'$
- Replace  $T$  by its offspring if it has better fitness

typically 0.02

in  $T'$

in  $T$  !



## Inver-over operator Performances

### Advantages

- Only 3 parameters
  - Population size 100
  - Inversion probability 0.02
  - Stopping criterion 10 loops without improvement
- One "evaluation" = 4 operations as only 2 edges are modified

### Resultats

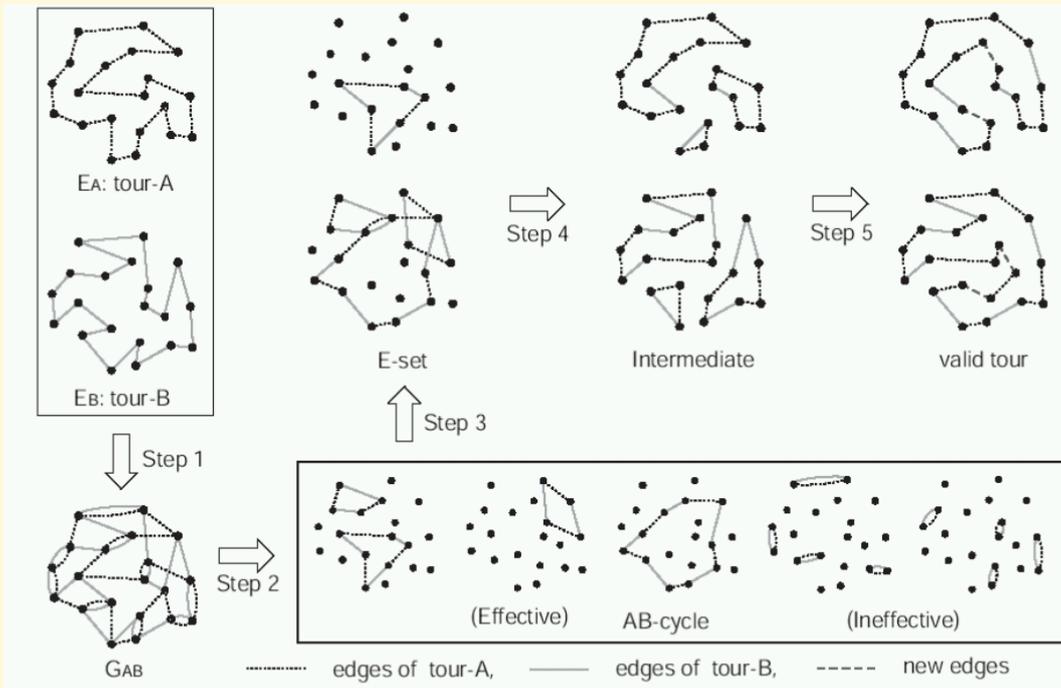
- Incredibly fast and efficient up to 500 cities
- Fast and efficient up to 10000 cities
- Only 1000 times slower than Lin-Kernighan

But didn't scale up...



# Edge Assembly Crossover

Y. Nagata, 97-06



Y. Nagata, PPSN 2006



## Edge Assembly Crossover (2)

**Step 2** : Random starting points, alternate edges from A and B

**Step 4** : Remove from parent its own edges in the sub-tour

**Step 5** : Greedy algorithm – increasing size of sub-tours

### Strategies for step 3

- **Rand** : uniform choice of AB-cycles
- **1AB** : a unique tour is used
- **Block** :
  - Choose largest AB-cycle  $C_0$
  - Build intermediate solution
  - For each sub-tour  $k > 1$ ,  
choose an AB-cycle  $C_k$  that touches it
  - Do step 3 with all  $C_i$

2006

step 4



## Edge Assembly Crossover : selection

### Diversity preserving selection

- $N$  parents are replaced by  $N$  offspring
- Repeat  $N$  times
  - For each pair of parents
  - Create  $N_e$  offspring using EAX
  - Best of  $N_e$  offspring and 2 parents survives



## Edge Assembly Crossover : results

- Using 1AB, then Rand if stagnation Nagata & Kobayashi, 1997
  - $N = 300$ ,  $N_e = 30$
  - 0% on 3038 cities
  - in 2h30 (Intel 200MHz)
  - but scales up poorly for larger problems
- Using 1AB, then Block if stagnation Nagata, 2006
  - $N = 300$ ,  $N_e = 30$
  - 0% on 24978 cities, in 45mn (Itanium 1.3 GHz)
  - Best known results for  
8079 (2mn), 14773 (6mn) and 33708 (1h25) cities!



- Généralités Initialisation et opérateurs de variation
- Représentation binaire ... pour en finir avec la ...
  - Opérateurs usuels
  - Théorie des Schémas ... et sa discussion
  - Généralisation Opérateurs indépendants de la représentation
  - Etude de cas : les représentations des réels
- Représentation réelle Hors stratégies d'évolution
- Combinatorial Optimisation Generic vs specific operators
  - Direct representations hybridization
  - **Memetic Algorithms**
  - Indirect representations An example: train re-scheduling



## Memetic Algorithm for the TSP

### Genetic Local Search – Freisleben & Merz - 96

- Greedy initialisation Random start, then closest city
- Local Optimisation (Lin-Kernighan, 3-opt) Population of local optima
- *Distance Preserving Crossover*  

$$d(\text{offspring}, \text{parent1}) \approx d(\text{offspring}, \text{parent2}) \approx d(\text{parent1}, \text{parent2})$$
- “4-edges” mutation In order to leave the local optimum

State-of-the-art results (96) up to 10000-15000 cities

<b>But</b> , e.g. on f11400:	Nearest neighbour	20%	12s
	Lin-Kernighan	2-3%	68s
	<i>Genetic Local Search</i>	0%	57h



# Memetic algorithms

## Other successes

- **Graph colouring** Dorne & Hao 97-98
  - Specific crossover
  - Tabu-search-based mutation
  - Best known results (500 to 1000 nodes)
- **Quadratic assignment** Merz & Freisleben 1997
  - *Distance Preserving Crossover*
  - Mutation by inversion
  - Memetic algorithm – using degraded 2-opt
  - Best known results (30-256 locations).

idem ER, EAX, ...



# Memetic algorithms

## Initialisation

**Constructive heuristics:** e.g. greedy

**Inoculation** of good solutions

- Obtained using other methods
- Should they be perturbed?
- Introduces a bias
  - Radcliffe and Surry, 94 Design of a gas network  
Best average results using inoculation  
Best overall solution using random initial population
  - Sourd and MS, 98 Mesh renumbering  
Semet and MS, 2005 Train re-scheduling  
Best results (average and best) from inoculated population



# Memetic algorithms

## Local optimisation

### When, what, and how?

- As a mutation operator
  - To a few individuals only
  - only partial optimisation?
  - Depends on the number and positioning of the local optima
- To all newborn offspring?
  - Genetic Local Search
  - Search space = local optima

### Lamarck vs Baldwin

- **Lamarck** : result of local optimisation replaces original individual  
Inheritance of acquired characteristics
- **Baldwin** : individual is unchanged, but its fitness is that of the result of local optimisation  
Hinton & Nowlan, 87

Problem-dependent issues :- (



# Memetic algorithms

## Variation operators

- Local operators
  - Must be different from operators of local optimiser
  - Must escape local optima
- Specific operators e.g. ER, EAX, ...
  - Require expert knowledge



# Memetic algorithms

## Conclusion

- The best of several worlds
- Improvement by several orders of magnitude **both in quality and cost**

State of the art on several benchmarks, **but** ...

- Real-world problems **scheduling, vehicle routing, load balancing, ...**  
are **heavily constrained**
- writing variation operators amounts to ...  
solve an optimisation problem!



- Généralités Initialisation et opérateurs de variation
- Représentation binaire ... pour en finir avec la ...
  - Opérateurs usuels
  - Théorie des Schémas ... et sa discussion
  - Généralisation Opérateurs indépendants de la représentation
  - Etude de cas : les représentations des réels
- Représentation réelle Hors stratégies d'évolution
- Combinatorial Optimisation Generic vs specific operators
  - Direct representations hybridization
  - Memetic Algorithms
  - **Indirect representations** An example: train re-scheduling



## Hybridisation ... within the representation

**Direct representation** : genotype  $\equiv$  phenotype

- For each {room, driver, machine}
- list of {courses, clients, tasks} with schedule

### Constraint handling

- Within variation operators
- A posteriori, using repair
  - Can be difficult too
  - Can bias the search

Is an optimisation problem per se!



## Hybridisation ... within the representation (2)

### Indirect representation

- **Genotype** = permutation of {course, clients, tasks}
- **Morphogenesis** = a specific program that **meets the constraints**  
e.g. *scheduler*
- **Phenotype** = apply *scheduler*, items ordered by the genotype

### Pros and Cons

- + Use standard variation operators for permutations
- + Any domain knowledge can be put in the *scheduler*
- Morphogenesis can be costly
- Weak causality                      *Small mutations can result in large fitness variations*



## Train (Re)scheduling

Y. Semet & MS, 04-06, coll. SNCF

- Re-schedule trains after some incident
- Strict constraints
  - Security Time and space spacing
  - Connections
  - ...
- Degrees of freedom: schedule of train in each station

**State-of-the-art** : CPLEX (c) Ilog

Mixed Linear Integer Programming  
Millions of variables for realistic cases



## Indirect representation

**Genotype** : a permutation of all trains to be re-scheduled

But it is not a TSP!

**Morphogenesis** : a scheduler

- Greedy Each train is scheduled as early as possible in all stations
- as allowed by the constraints Expert knowledge

**Question**: Does this representation cover the whole space of schedules?

**Answer**: Certainly not :-)



## Variation operators

### Choice among usual operators for permutations

By trial-and-error

- No crossover was found useful

Uniform Order-based Crossover, Partially Matched Crossover, ...

- Basic mutation OK

simple inversion, ...

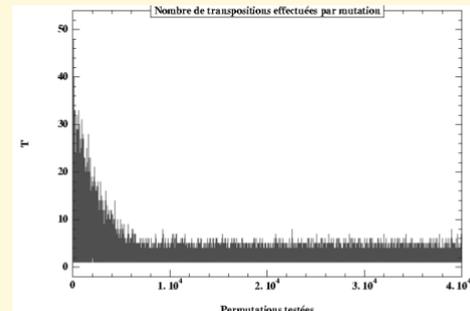
- MM = inversion of 2 trains

- with a limited radius to avoid too large changes

### How to apply mutation

- Following a binomial law
- of mean  $T$
- decreasing along evolution

e.g. from 30 to 3





## Initialisation Inoculation

- Random permutations have a very poor fitness
- Was expected, but . . . a not-too-bad solution is known:  
the initial schedule before the incident happened

### Procedure

- Reverse-engineer the genotype of the initial schedule Using an EA!  
Only its phenotype is know
- Initialise population with perturbed initial schedule
  - Mass Mutation: mutate initial schedule  $k = 3$  times
  - Layered Mutation: mutate initial schedule 0, 10 or 500 times



## Results

On real instances

541 trains around Paris

### Experimental conditions

- (10+70)-ES
- Initialisation: Mass Mutation or Layered
- $T$  (average number of application of inversion mutation) fixed or decreasing
- Comparisons through Wilcoxon unsigned test with 99% confidence

### Tendencies

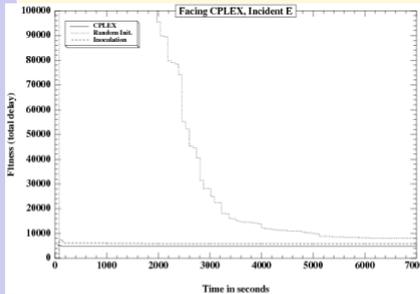
- For fixed  $T$ , MM is the best choice
- For decreasing  $T$ , MM and Layered perform equally well



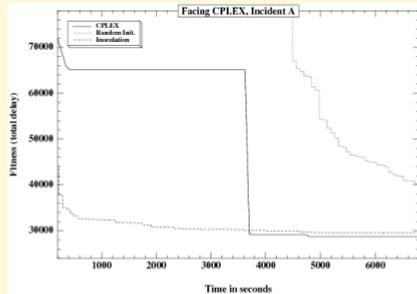
## EA vs CPLEX

Deterministic state-of-the-art: CPLEX (c) Ilog

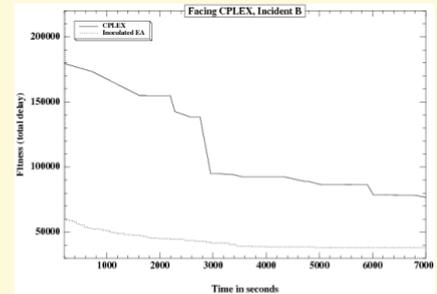
541 trains  $\equiv$  1M variables, 3M constraints



Easy incident  
CPLEX unbeatable



Difficult incident  
CPLEX catches up  
in 2 hours



Very difficult incident  
CPLEX catches up  
in 6 hours

On the very very very difficult incidents, no algorithm gives any useful result ... whereas human do!



## Conclusion on train re-scheduling

- All expert knowledge is captured in the scheduler

Tedious programming

- Variation operator: the simplest of all!
- ... but cleverly applied
  - (10+70)-ES
  - Repeatedly applied,
  - Binomial law with decreasing mean
- Initialisation : key to success
- CPLEX is beaten

but it only lost a battle :-)



## A few other applications

Among many!

- Crew scheduling Levine 97
  - University scheduling Paechter & al. 98
  - Job-shop scheduling Hart & Ross 98, 2000
  - Frequency assignment Crisan & al. 97, Dorne 98-99
  - Air Traffic Control Alliot, Delahaye & al. 95-00
  - Mobile phone cell definition ARNO project, 00
  - Routing in complex networks Knowles & Corne, 98-00
  - Vehicle routing Chronopost International, 2002
- To be updated . . . .



# Evolutionary Combinatorial Optimisation

## Conclusions

- No Free Lunch Generic doesn't work
- Hybridisation at all steps
  - Representation Indirect representations
  - Operators Can be an optimisation problem on their own
  - Initialisation Inoculation
  - Lamarckism (or Baldwinism) Through OR or other metaheuristics

Hard work mandatory, but will pay off :-)