



Optimisation par Évolution Artificielle et Application à la Robotique

Anne Auger, Nicolas Bredèche, Marc Schoenauer
Equipe-Projet TAO – INRIA Saclay et LRI

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

8 décembre 2008



Plan

- Contexte Optimisation
- L'algorithme le paradigme biologique
- Un exemple sans ordinateur
- Points-clé Diversité, exploitation vs exploration, représentation
- Les étapes
 - Evaluation, arrêt
 - Moteur d'évolution Darwinisme artificiel
 - Initialisation, opérateurs de variation Chaînes de bits, nombres réels et arbres
- Les instances historiques
- Conclusions et références



Optimisation

Données

- Ensemble de solutions possibles
- Critère de qualité

Espace de recherche

Fonction objectif
mais aussi performance, "fitness"

But

Trouver la meilleure solution (pour le critère donné)

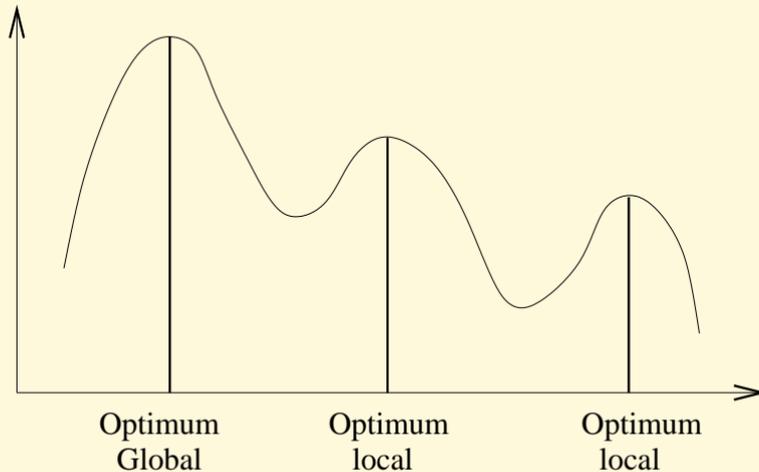
Formellement

Soit $\mathcal{F} : \Omega \mapsto \mathbb{R}$

Trouver $x^* \in \Omega$ tq $x^* = \text{ArgMax}(\mathcal{F})$



Optima (maximisation)



Optimum global : x^* t.q. $(\forall x \in \Omega) \mathcal{F}(x^*) \geq \mathcal{F}(x)$

Optimum local : x^* t.q. $(\exists \varepsilon > 0)$

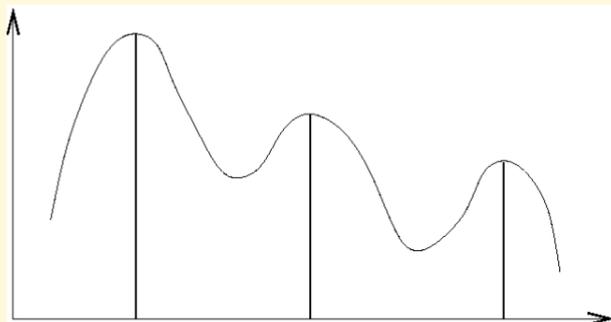
$$B(x, \varepsilon) \neq \{x\} \text{ et } (\forall x \in B(x^*, \varepsilon)) \neq \{x\}) \mathcal{F}(x^*) \geq \mathcal{F}(x)$$

Maxima **stricts** si inégalités strictes pour $x \neq x^*$

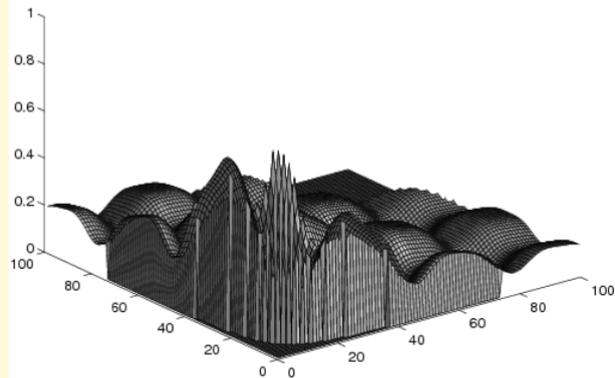


Paysages

"Graphe" de \mathcal{F} sur Ω



$\Omega = \mathbb{R}$



$\Omega = \mathbb{R}^2$



Sources de difficultés

- Espace de recherche trop grand (cas discret)

e.g. Problèmes NP-complets

- Fonction objectif "complexe"
irrégulière, non différentiable, non continue, ...
- Fonction objectif donnée via un calcul ou un expérience

Boîte noire



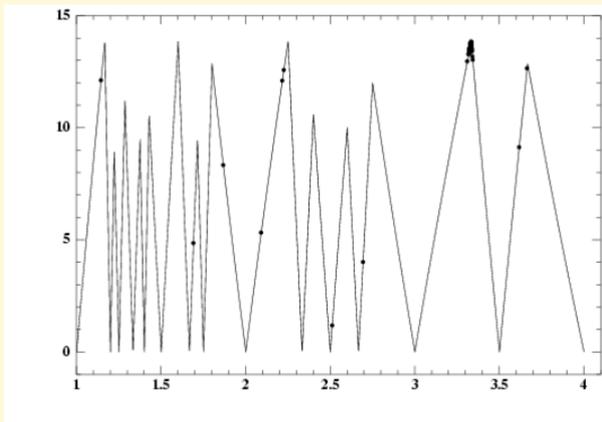
\mathcal{F} très chahutée

L. Taïeb, CMAP & Thomson

Espace de recherche : Interféromètres

Positionner des antennes

But : Maximiser la tolérance en conservant la précision.



Cas de 3 antennes, \mathcal{F} = Marge d'erreur (position 2ème antenne)

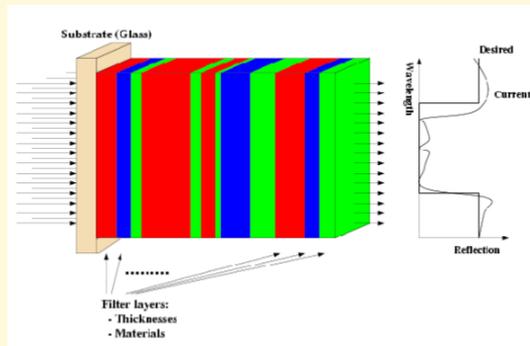


Ω mixte : discrets \times réels
 Schutz & Bäck, ICD, Dortmund.
 Martin, Rivory & Schoenauer,
 Optique des Solides Paris VI & CMAP.

Espace de recherche : Filtres optiques

(matériau, épaisseur)₁ ... (matériau, épaisseur)_N

But : Répondre au gabarit fixé.





Ω = Circuits analogiques

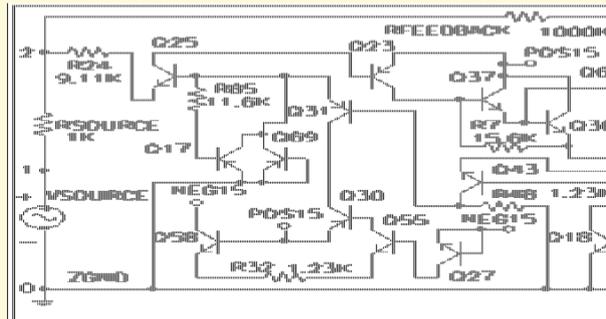
Koza et al., Stanford.

Espace de recherche : Circuits analogiques

Réseau de transistors, diodes, résistances

But : Fonctionnalités fixées

e.g. extraction de racine cubique





Ω espace de fonctions, \mathcal{F} mal connue

Robotique autonome On cherche la vitesse des moteurs en fonction des valeurs des capteurs
on veut que le robot avance en évitant les obstacles.



Espace de recherche : un espace de fonctions

But : Minimiser l'activation des capteurs

... tout en avançant !

Robotique évolutionnaire



\mathcal{F} non calculable

Herdy & al., Berlin, PPSN96

Espace de recherche : Mélanges de café

But : Retrouver un arôme

\mathcal{F} = avis de l'expert

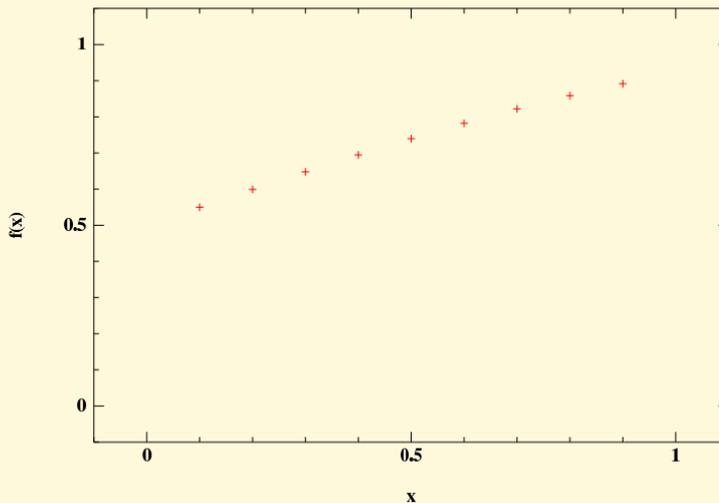




Scénario "boite-noire"

- Accès aux valeurs de \mathcal{F}
- Aucune hypothèse sur la fonction objectif

Aucune ? $\Omega = \mathbb{R}$, on a observé 10 points

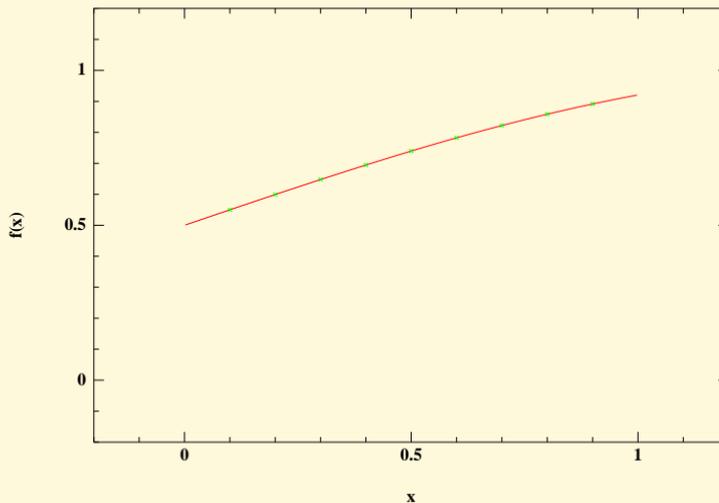




Scénario "boite-noire"

- Accès aux valeurs de \mathcal{F}
- Aucune hypothèse sur la fonction objectif

Aucune ?

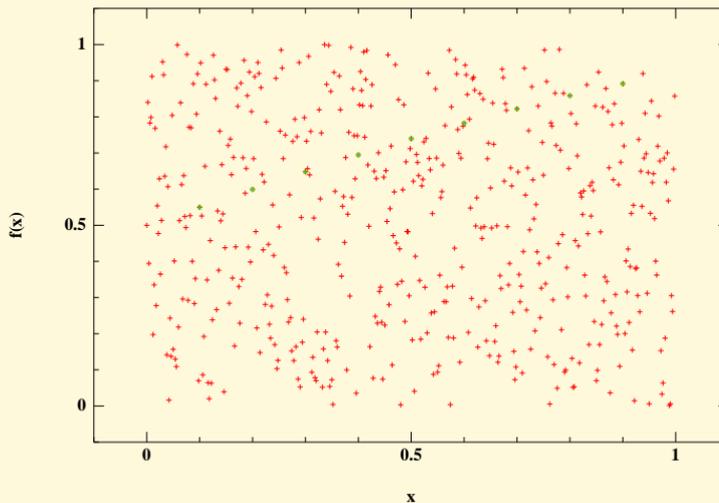




Scénario "boîte-noire"

- Accès aux valeurs de \mathcal{F}
- Aucune hypothèse sur la fonction objectif

Aucune ?





Contexte

- Optimisation globale

Recherche d'un optimum global
ou d'une bonne approximation

- Scénario boîte noire, mais ...

- Problème structuré : les voisinages de bonnes solutions peuvent contenir des solutions encore meilleures.



Algorithmes d'optimisation

Méthodes exactes

- Algorithmes de gradient
- Méthodes énumératives
- Hill-Climbing

cas continu

cas discret

Méthodes approchées

- Heuristiques
- Recherche locale
- Méthodes stochastiques (méta-heuristiques)



Méthodes de gradient

$$X_{i+1} = X_i + dX \times \nabla \mathcal{F}(X_i)$$

Contexte :

- Espace continu : $\Omega \subseteq \mathbb{R}^N$
- Fonction \mathcal{F} dérivable

dérivation numérique

Conditions nécessaires :

- Fonction régulière,
- ou Connaissance a priori

\mathcal{F} convexe .. optimum unique

X_0 bien choisi

Méthode locale, pour des problèmes continus et réguliers

pas boîte noire



Méthodes de type énumératif

Parcourir l'espace suivant un ordre déterministe

Contexte :

- Espace fini : $\Omega \equiv [1..N]$ **Attention** : $\text{opt.}_{discret} \neq \text{arrondi}(\text{opt.}_{continu})$
- Toute fonction \mathcal{F} , mais ...
- Ordre de parcours
 - fixé
 - dépend du problème Branch-and-Bound, A*, contraintes, ...

Conditions nécessaires :

- Taille de l'espace limitée
- Discrétisation bien choisie Méthode des intervalles

Méthode globale, coûteuse, pour des problèmes discrets

pas boîte noire



Hill-Climbing

$$X_{i+1} = \text{" Meilleur Voisin " de } X_i$$

Contexte :

- Toute fonction \mathcal{F}
- Tout espace Ω "naturel" ...
- pour lequel on sait définir des voisinages

Trouve l'optimum local (pour les voisinages choisis) le plus proche du point de départ

Conditions nécessaires :

- X_0 doit être bien choisi

Méthode locale, coûteuse



Exemples de voisinages

- $\Omega = \{0, 1\}^N$
- $d_H((x_i), (y_i)) = \sum |x_i - y_i|$
- $V_d(x) = \{y \in \{0, 1\}^N; d(x, y) \leq d\}$

distance de Hamming

Le choix de d détermine la "localité" de la recherche

- $d = N$: recherche globale (énumération !)
- $d = 1$: recherche très locale



Heuristiques

- Dépend **fortement** du problème
- Permet d'obtenir rapidement une solution
- pas très bonne ...

Exemple : le TSP

Heuristique gloutonne: choisir la ville la plus proche.

Plus de généralité : **méta-heuristiques**

Des heuristiques que l'on peut appliquer à plusieurs classes de problèmes



Méthodes stochastiques

Les points suivants sont choisis à l'aide de tirages aléatoires

- Monte-Carlo • X_i tiré avec une loi uniforme
• Marche aléatoire sans mémoire
- Recherche locale • Hill-climbing stochastique
- Métropolis, Recuit Simulé • Kirkpatrick, Gelatt and Vecchi, 1983
- Recherche Taboue • F. Glover – 1977 & 1989
- Algorithmes évolutionnaires ... • depuis 1965

Méthodes globales, mais TRÈS coûteuses.



Recherche locale (stochastique)

- Choisir x_0 uniformément dans Ω
- Calculer $\mathcal{F}(x_0)$
- Boucle e.g. Tant que améliorations ...
 1. Choisir y (uniformément) dans $V(x_t)$ voisinage
 2. Calculer $\mathcal{F}(y)$
 3. Si $\mathcal{F}(y) > \mathcal{F}(x_t)$, alors $x_{t+1} = y$ Acceptation
sinon $x_{t+1} = x_t$



Voisinages, exploration et exploitation

- Grands voisinages (e.g. $d = N$ dans $\{0, 1\}^N$)
≡ Monte-Carlo: **exploration** sans tenir compte du passé
- Petits voisinages (e.g. $d = 1$)
Recherche "très" locale: **exploitation** de la mémoire sans exploitation

Solutions

- Étendre la notion de voisinage
- Accepter aussi des points moins bons Item 3 de la boucle
Metropolis et Recuit simulé
- Interdire de revenir sur ses pas
Recherche taboue
- Utiliser plusieurs points simultanément et accepter des moins bons
Algorithmes évolutionnaires



Voisinages et opérateurs (stochastiques)

Voisinages : difficultés

- Définition d'un voisinage Espaces complexes: pas de distance (accessible)
- Échantillonnage des voisinages

Opérateurs stochastiques sur Ω

Soit $M : \Omega \times \Theta \mapsto \Omega$.

où Θ est un ensemble de paramètres

réels, variables aléatoires, ...

Alors $V_M(x) = \{y \in \Omega; \exists \theta \text{ réalisation de } \Theta; y = M(x, \theta)\}$
peut être vu comme un voisinage de x



Exemples sur $\{0, 1\}^N$

1. Opérateur **1-bit-flip**: pour $x \in \{0, 1\}^N$

- Tirer $i_0 = U[1, N]$ tirage uniforme dans $[1, N]$
- Inverser le bit i_0 de x

Ici, Θ est la VA entière de loi uniforme sur $[1, N]$, et tirer uniformément un élément de $V_M(x)$ est équivalent à appliquer M à x

2. Opérateur **bit-flip stochastique** de proba. $p_m \in (0, 1)$:
pour $x \in \{0, 1\}^N$

- Pour chaque bit x_i
- Inverser x_i avec probabilité p_m

Ici, Θ est la VA booléenne qui vaut *true* avec probabilité p_m

mais $V_M(x) = \{0, 1\}^N$ tout entier !



Recherche (locale?) stochastique (2)

- Choisir x_0 uniformément dans Ω
- Calculer $\mathcal{F}(x_0)$
- Boucle
 1. $y = \text{Move}(x_t)$
 2. Calculer $\mathcal{F}(y)$
 3. Si $f(y) > f(x_t)$, alors $x_{t+1} = y$
sinon $x_{t+1} = x_t$

e.g. Tant que améliorations ...

opérateur stochastique

Acceptation



Recherche stochastique (3)

- Choisir x_0 uniformément dans Ω
- Calculer $\mathcal{F}(x_0)$
- Boucle
 1. $y = \text{Move}(x_t)$
 2. Calculer $\mathcal{F}(y)$
 3. $x_{t+1} = \text{select}(y, x_t)$

e.g. Tant que améliorations ...

e.g. Boltzman \rightarrow Metropolis



Recherche stochastique (4)

- Choisir x_0 uniformément dans Ω
- Calculer $\mathcal{F}(x_0)$
- Boucle

e.g. Tant que améliorations ...

1. Faire k fois

- $y_i = \text{Move}(x_t)$
- Calculer $\mathcal{F}(y_i)$

2. $x_{t+1} = \text{select}(y_1, \dots, y_k, x_t)$

Avec une sélection déterministe, on a un $(1 + k) - ES$



Les algorithmes bio-inspirés

Des modèles mathématiques de phénomènes biologiques

- Réseaux de neurones
- Évolution artificielle
- Algorithmes de fourmis
- Optimisation par essaim de particules
- Systèmes immunitaires
- Membrane Computing
- DNA Computing

modèles rarement fidèles - peu de retours utiles



Algorithmes Evolutionnaires : les racines

- **Algorithmes Génétiques** J. Holland, 75, D.E. Goldberg 89
AI and biology – US, east coast
- **Strategies d'Évolution** I. Rechenberg 65, 73, H.P. Schwefel 65, 81
Engineers – Germany
- **Programmation Evolutionnaire** L.J. Fogel, 66, D.B. Fogel, 91, 95
Automata and time series – US, west coast
- **Programmation Génétique** J. Koza, 92 – a late root!
GAs on parse-trees

Aujourd'hui:

pragmatisme oecuménique ?

“Evolutionary Computation”



Plan

- Contexte Optimisation
- **L'algorithme** le paradigme biologique
- Un exemple sans ordinateur
- Points-clé Diversité, exploitation vs exploration, représentation
- Les étapes
 - Evaluation, arrêt
 - Moteur d'évolution Darwinisme artificiel
 - Initialisation, opérateurs de variation Chaînes de bits, nombres réels et arbres
- Les instances historiques
- Conclusions et références



Paradigme Darwinien

- Sélection naturelle avantage aux espèces adaptées à leur environnement
- + Variations aveugles
 parents → enfants par petites déviations apparemment non dirigées.
- = Adaptation résultante apparition d'espèces (e.g. bactéries résistantes).
- “Objectif” capacité de survivre et de se reproduire

Mais

- Source d'inspiration
- Aide à l'explication
- **Pas justification**



Algorithmes Evolutionnaires : La Métaphore

Modèle : L'évolution darwinienne des populations biologiques.

Les individus les plus adaptés survivent et se reproduisent

Vocabulaire :	Individu	Elément X de Ω
	Performance	Valeur de $\mathcal{F}(X)$
	Population	Ensemble de P éléments de Ω
	Génération	Passage de la population Π_i à Π_{i+1}

Processus :

- 1) Sous la pression du milieu,
- 2) Les individus se croisent, mutent et se reproduisent.
- 3) Au bout d'un nombre certain de générations, les individus les plus performants apparaissent dans la population.

≡ les **optima** de \mathcal{F} ...



Parallèle biologie/algorithmique

Différences

Environnement changeant mécanismes spécifiques ?	Généralement fixe
Performance inconnue finalité des plumes du paon* ?	C'est le point de depart
Le sélection élimine les nuls	Tentation eugénique Lamarckisme possible
Adaptation/sélection boîte noire	Etudes théoriques ?

* Richard Dawkins, *Le gène égoïste*



Parallèle biologie/algorithmique

Points communs

Diversité génétique essentielle

Maladies fatales
Multi-racial utile

Convergence prématurée
Solutions multiples utiles

Lenteur du processus

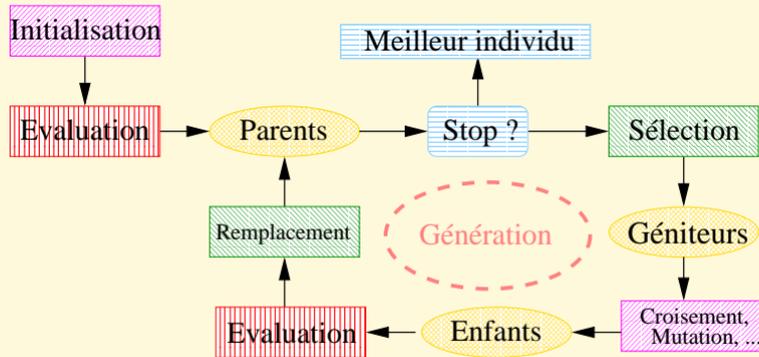
Néanderthal: -150 000 à -35 000
 \simeq 6000 générations
Cro-Magnon: -30 000 à nous
 \simeq 1200 générations

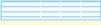
Vous verrez!

On est peut-être pas à l'optimum,
mais on a des solutions assez bonnes :-)



Algorithmes d'évolution : Le Squelette



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



Plan

- Contexte Optimisation
- L'algorithme le paradigme biologique
- **Un exemple** sans ordinateur
- Points-clé Diversité, exploitation vs exploration, représentation
- Les étapes
 - Evaluation, arrêt
 - Moteur d'évolution Darwinisme artificiel
 - Initialisation, opérateurs de variation Chaînes de bits, nombres réels et arbres
- Les instances historiques
- Conclusions et références



La feuille de papier qui tombe

P. Bentley

Le problème

- Trouver la forme d'un morceau de papier
- qui met le plus de temps possible à tomber

Difficultés

- Pas de simulation
- Pas d'a priori sur la forme de la solution

Quoique ...



Un algorithme sans ordinateur

Matériel nécessaire

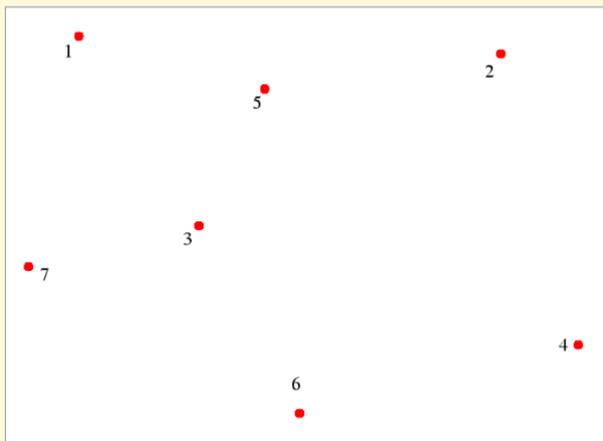
- Une ramette de feuilles A4
- Des ciseaux
- Un bac à sable format A4
- Une dizaine de petites cailloux numérotés
- Un chronomètre
- Une fléchette
- ... de la patience



Initialisation

- Lancer les petits cailloux en l'air au dessus du bac à sable
- Reporter les positions des cailloux tombés dans le bac sur une feuille A4

Géotype : Liste ordonnée des points





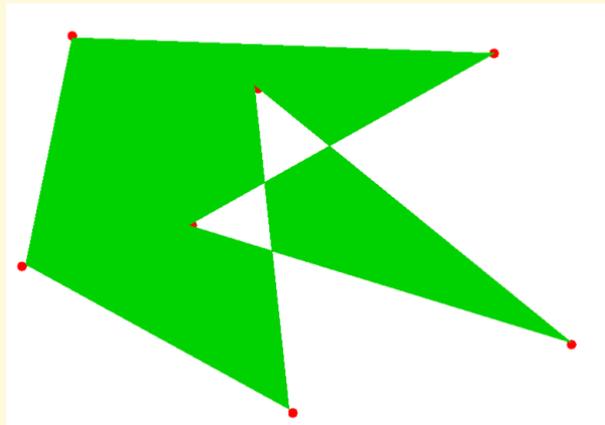
Morphogénèse

passage du génotype au phénotype

- Tracer le polygone rempli correspondant
- Découper la forme obtenue

à la MacPaint

Phénotype : La forme de papier





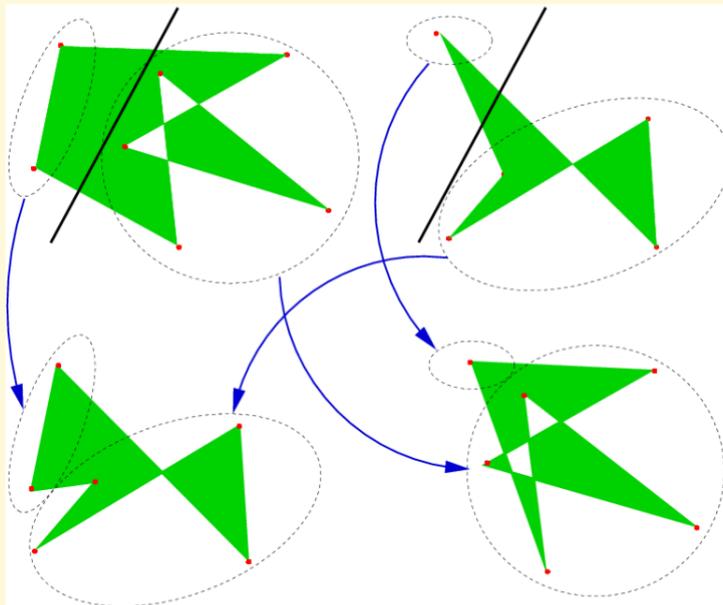
Évaluation

Pour chaque forme

- Faire 5 fois
- Lâcher la forme de 2m de haut
- Chronométrer le temps d'atteinte du sol
- Faire la moyenne des 5 temps



Croisement

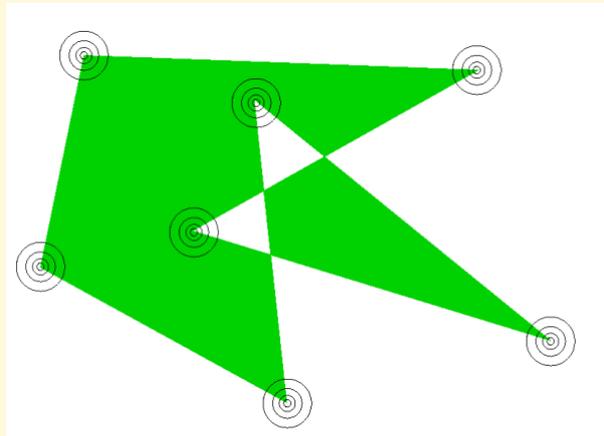


Un croisement possible, pour lequel les ciseaux et le scotch ne suffisent pas.



- Viser avec la fléchette chaque point tour à tour

Distance à ajuster selon votre habileté



- Tracer sur une nouvelle feuille les impacts

Si la fléchette sort du cadre, enlever le point

Si votre portable sonne, ajouter un point



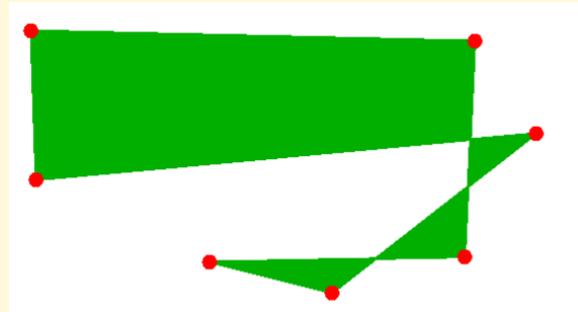
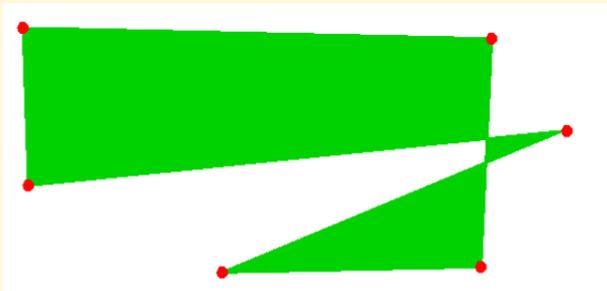
Darwinisme (remplacement déterministe)

Déchirer la moitié des formes – les plus rapides



Résultats

Avec 5 formes et 10 générations



Deux des meilleures formes obtenues

Meilleure forme aléatoire : 0.8s

Toutes les formes après 10 générations de 10 formes : $> 2s$

Comportement “hélicoptère”





Plan

- Contexte Optimisation
- L'algorithme le paradigme biologique
- Un exemple sans ordinateur
- **Points-clé** Diversité, exploitation vs exploration, représentation
- Les étapes
 - Evaluation, arrêt
 - Moteur d'évolution Darwinisme artificiel
 - Initialisation, opérateurs de variation Chaînes de bits, nombres réels et arbres
- Les instances historiques
- Conclusions et références



Points clé

- Espace de recherche quelconque

Choix de la représentation

- Une population, pas un individu

Attention à la perte de diversité génétique

- Exploration de l'espace / Optimisation locale

Le dilemme EVE

- Indépendance objectif / moteur d'optimisation

Boîte noire ou connaissances du contexte ?

No Free Lunch Theorem



Perte de la Diversité Génétique

Si les individus d'une population se ressemblent trop,

1. Les populations suivantes deviennent de plus en plus homogènes
fragilité au changement
2. évolution d'une population → évolution d'un individu
3. Découverte du plus proche optimum local et enlèvement de la recherche

Dans la pratique, **la population ne se rediversifie pas.**

⇒ Convergence Prématuration



Le Dilemme Exploitation vs Exploration

- Exploitation des bons individus.
recherche locale: chercher dans le voisinage des meilleurs individus de la population.
- Exploration des zones inconnues de Ω .
recherche globale: il faut pouvoir aller partout.

Excès d'exploitation \implies

Convergence prématurée
enlèvement dans un optimum local

Excès d'exploration \implies

Pas de convergence
 \approx marche aléatoire



La représentation

- Le **Darwinisme** ne dépend que de la **performance**
- **L'initialisation et les opérateurs de variation** ne dépendent que de la **représentation**.

Trois exemples de base:

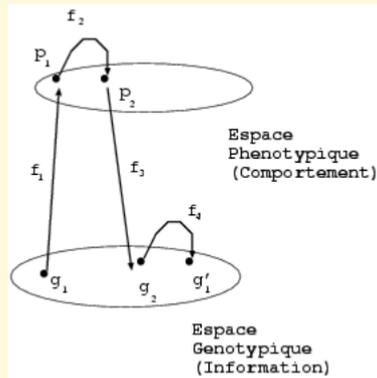
- Représentation “binaire” $\Omega = \{0, 1\}^N$
- Représentation “réelle” $\Omega = [0, 1]^N$ or $\mathbb{R}^N \dots$
- Représentation par arbres GP

Le choix de la représentation est **crucial** sera mis en avant dans tout le cours.



Phénotypes – Génotypes

Lewontin – 1974



- Espace phénotypique :
- Espace génotypique :

Evaluation, sélection
Croisement, mutation

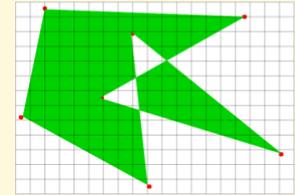
Préhistoire : Génotypes binaires universels

Aujourd'hui : Importance des connaissances
dans la représentation
dans les opérateurs



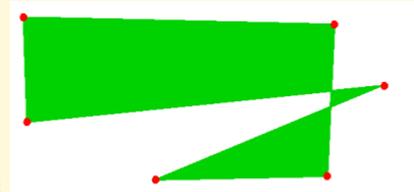
Représentations pour les petits papiers

- Paramétrique bestiale
- Compacte structurée limitée
- Compacte non structurée
- Par “programme” e.g. $\text{quadrangle}(\dots) +_4 \text{triangle}(\dots) +_3 \text{triangle}(\dots)$



Les épaisseurs pour un bi-profil

Les petits cailloux :-)



+ Modularité, réutilisabilité

– Morphogénèse complexe, causalité diffuse



Plan

- Contexte Optimisation
- L'algorithme le paradigme biologique
- Un exemple sans ordinateur
- Points-clé Diversité, exploitation vs exploration, représentation
- **Les étapes**
 - **Evaluation, arrêt**
 - Moteur d'évolution Darwinisme artificiel
 - Initialisation, opérateurs de variation Chaînes de bits, nombres réels et arbres
- Les instances historiques
- Conclusions et références



Évaluation de la population courante

DE TRES LOIN l'étape la plus coûteuse

- Ne pas recalculer $\mathcal{F}(X)$ inutilement
- utiliser une estimation de \mathcal{F}
- ... mais pas trop longtemps

voire la construire à la volée

Optimum approché \neq optimum réel



Critères d'arrêt

Pas si simple !

- Quand on a trouvé l'optimum ...:-)
- Quand on n'espère pas trouver mieux

Perte de diversité

- Quand le ratio $\frac{\text{gain espéré}}{\text{surcoût de calcul requis}}$
est trop élevé

Rationalité limitée

- Quand on a épuisé ses ressources

Nombre fixé d'évaluations

Règle heuristique: après un nombre donné d'évaluations sans amélioration.



Evaluation des résultats

Point de vue informatique

Ne **jamais** tirer de conclusions d'**un seul** essai !

Utiliser des mesures statistiques

Moyennes et écarts-types, médianes, ...

Pour **comparer** des algorithmes, utiliser des tests statistiques

Paramétriques (e.g. Student) ou non-paramétriques (e.g. Wilcoxon)

Point de vue de l'application

Contexte de **conception**

Trouver au moins une fois une très bonne solution

Contexte de **production**

Trouver en moyenne une solution assez bonne



Plan

- Contexte Optimisation
- L'algorithme le paradigme biologique
- Un exemple sans ordinateur
- Points-clé Diversité, exploitation vs exploration, représentation
- Les étapes
 - Evaluation, arrêt
 - **Moteur d'évolution** Darwinisme artificiel
 - Initialisation, opérateurs de variation Chaînes de bits, nombres réels et arbres
- Les instances historiques
- Conclusions et références



Sélection

Objet : Choisir ceux qui se reproduisent

- Darwinisme: Biais en faveur des plus adaptés
- Biais trop important: Perte de diversité
- Biais trop petit : pas de convergence

Différentes modalités

- Déterministe, comparaisons de fitness ES historique
- Proportionnelle roulette, GAs historiques
- Stochastique, basée sur des comparaisons de fitness GAs historiques, EP historique

Opérateur d'exploitation



Remplacement

Objet : Choisir ceux qui survivent

- L'autre étape Darwinienne
- Peut être déterministe ou stochastique
- Choix parmi les enfants seulement, ou conflit de générations

Opérateur d'exploitation



Moteur d'évolution

Sélection et remplacement

- sont deux étapes de “sélection”
- Sélection : un individu peut être sélectionné plusieurs fois
- Remplacement : sélectionne au plus 1 fois chaque individu

Darwinisme artificiel = sélection + remplacement
≡ **Moteur d'évolution**



Elitisme

Garantir la survie du meilleur ?

- **Théorie** : Pas si clair

Il faut accepter de perdre pour gagner
cf théorie des $(\mu, \lambda) - ES$

- **Application** : psychologiquement trop dur!

conserver le meilleur en mémoire

Elitisme fort: quelques meilleurs survivent de toute façon

Elitisme faible:

Si $\text{Max}(\mathcal{F}(\Pi_{t+1})) < \text{Max}(\mathcal{F}(\Pi_t))$

Remplacer le pire de Π_{t+1}
par le meilleur de Π_t

Heuristique d'exploitation



Sélection Déterministe

Historiquement : Les moteurs d'évolution $(\mu ; \lambda)$ -ES

- Sélection uniforme (\equiv pas de sélection)
- μ parents donnent λ enfants
- Remplacement :
 (μ, λ) -ES : prochains μ parents = meilleurs parmi les λ enfants
Pour: meilleurs résultats de convergence
Contre: on peut perdre les meilleurs

$(\mu + \lambda)$ -ES : prochains μ parents =
meilleurs parmi les μ parents + les λ enfants
Pour: robustesse pratique
Contre: on peut converger vers un opt. local

Paramètres : μ, λ



Sélection Stochastique

(I) La roulette

Objet : Tirer P parents parmi $\Pi_t = \{X_1, \dots, X_P\}$

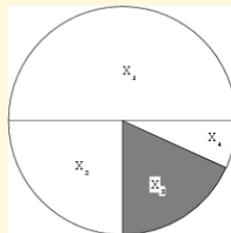
On tire P fois un individu avec remplacement

Proba[select. X_i] $\propto \mathcal{F}(X_i)$

Roulette :

Largeur de la case i : $\frac{\mathcal{F}(X_i)}{\sum_j \mathcal{F}(X_j)}$

Exemple, pour $P = 4$ et
 $\mathcal{F}(X_i) = \{50, 25, 15, 10\}$



Tirage d'un parent :

Lancer la boule;

si elle tombe dans la case i , sélectionner X_i



Algorithme génétique d'Epinal

Historiquement: Le "Simple GA"

- Sélectionner P individus par tirage de roulette avec remplacement
- Générer P enfants à l'aide des opérateurs de variation
- Remplacer les parents par les enfants remplacement générationnel

Paramètres: pas de paramètres

[voir plus loin](#)



Tirage à la roulette (2)

- Pour :** favorise les meilleurs
mais les mauvais ont des chances
- Contre :** coût élevé
variance élevée
perte de diversité possible

Pression sélective ps : proba. de sélection du meilleur / proba. moyenne
= nb de copies du meilleur sur P tirages

$$ps = P \times \frac{\mathcal{F}(X_{max})}{\sum_j \mathcal{F}(X_j)}$$

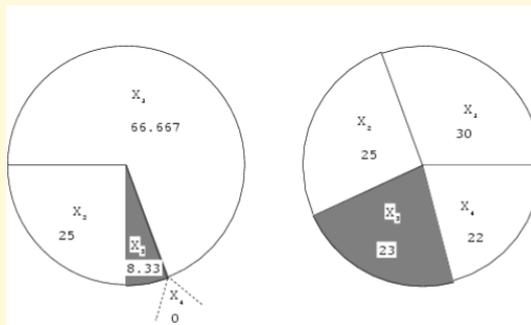
- **Trop forte** : le meilleur prend le dessus rapidement
- **Trop faible** : pas de sélection (\sim marche aléatoire)

Tirage à la roulette (3): Mise à l'échelle



Optimiser $\mathcal{F} \equiv$ Optimiser $\alpha \mathcal{F} + \beta$

$$\frac{5}{3}(\mathcal{F} - 10)$$



$$0.2\mathcal{F} + 20$$

Heuristique de Mise à l'échelle

fitness scaling

- Ajuster α et β pour une pression sélective donnée ($\bar{\mathcal{F}}$ fixé)

tronquer les valeurs négatives

- Tirage à la roulette sur $\mathcal{F}' = \alpha\mathcal{F} + \beta$.

Exemples et problème du super-individu



Tirage à la roulette (4)

Variantes principales

- Reste stochastique** Sélection déterministe $E\left(\frac{\mathcal{F}(X_i)}{\sum_j \mathcal{F}(X_j)}\right)$
 Puis roulette sur le reste décimal
 diminue la variance
 P sélections simultanées
- Ranking**
 Largeur de la case i : $\left(\frac{i}{P \times (P-1)/2}\right)^\alpha$
 = roulette sur le rang
 meilleur : case P
 en général, $\alpha=1$ (ranking linéaire)
 lisse la sélection
 Les valeurs exactes sont sans importance
- Troncation** :
 Sélection à la roulette parmi les μ meilleurs
 augmente la pression sélective

Sélection Stochastique

(II) Le tournoi



- **Tournoi (déterministe)**

$ps \approx T$

- Tournoi de taille $T \in \mathbb{N}$
- Choix uniforme de T individus
Rendre le meilleur

avec ou sans remplacement?

- **Tournoi stochastique (binaire)**

$ps \approx 2t$

- Taux $t \in [0.5, 1]$
- Choix uniforme de 2 individus
Rendre le meilleur avec probabilité t

- **Pour** : Robustesse par rapport aux erreurs sur \mathcal{F}
Facile à paramétrer T or t

- **Contre** : forte variance



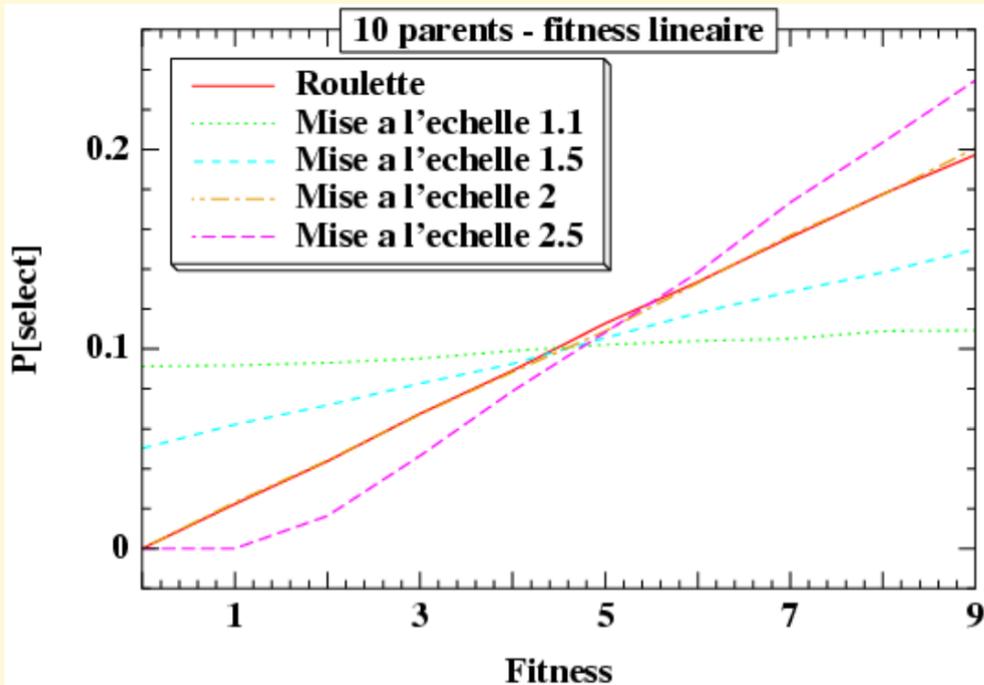
Sélection : illustration

- 10 parents de fitness toutes différentes
- étagée linéairement, exponentiellement ou logarithmiquement
- Sélection de 100000 géniteurs Variance
- Plusieurs valeurs de pression sélective ... Mise à l'échelle, rang
- de taille (ou de probabilité) pour les tournois



Sélection : illustration (2)

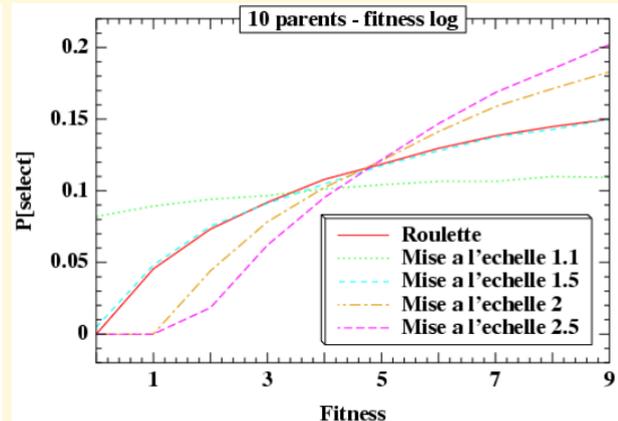
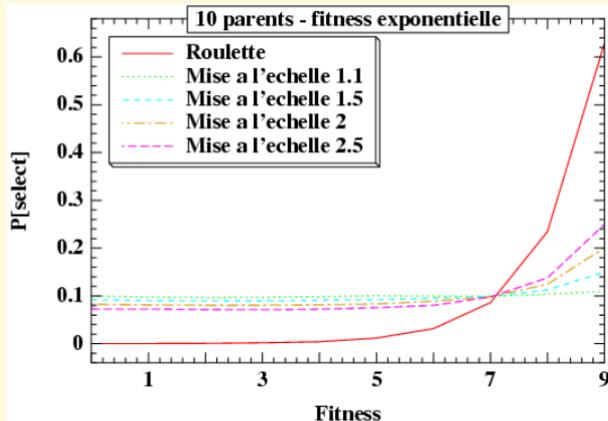
Fitness linéaire – Mise à l'échelle (linéaire)





Sélection : illustration (3)

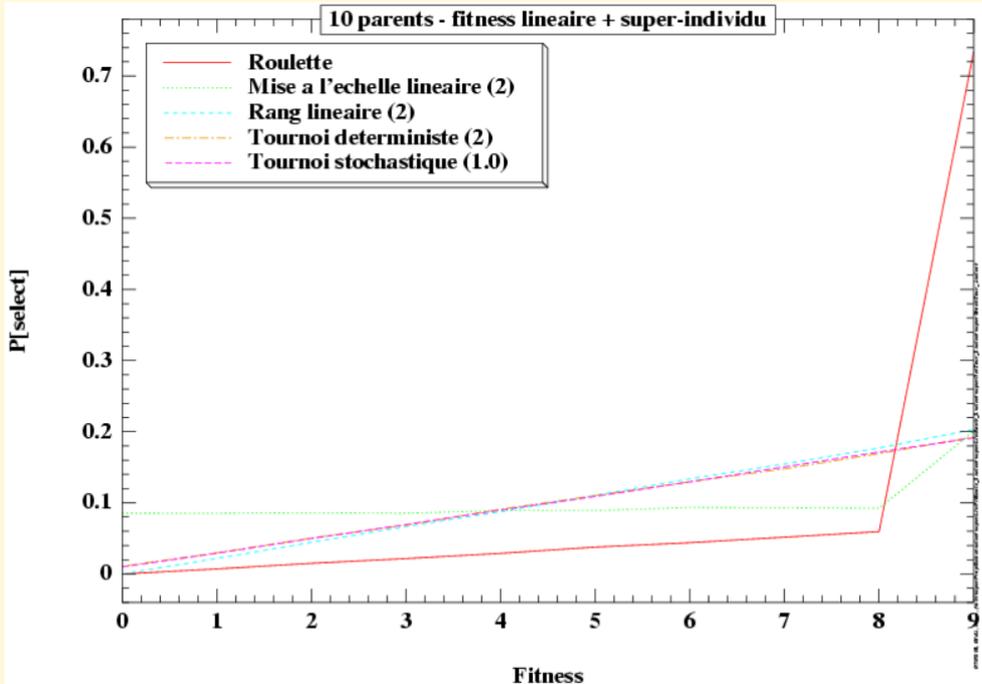
Fitness exp. et log. – Mise à l'échelle (linéaire)





Sélection : illustration (4)

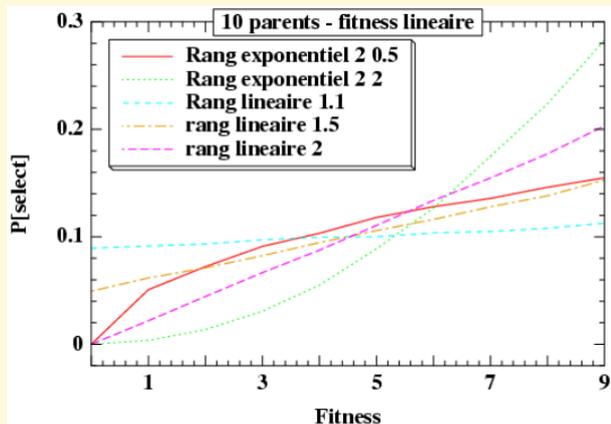
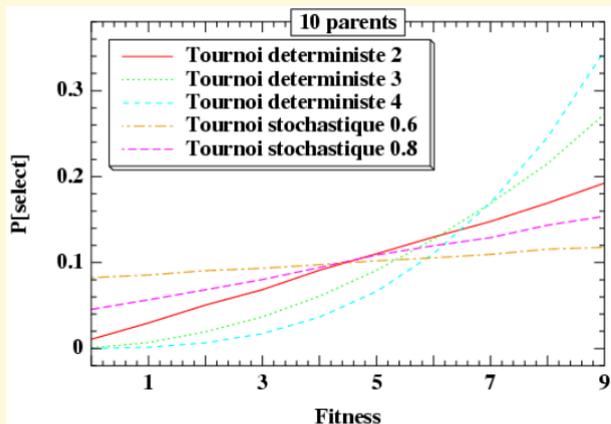
Présence d'un super-individu





Sélection : illustration (5)

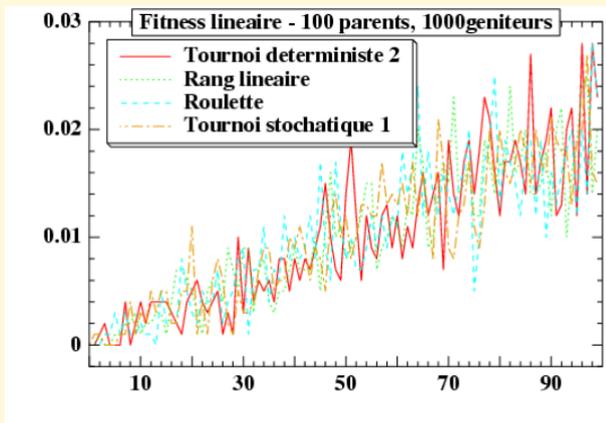
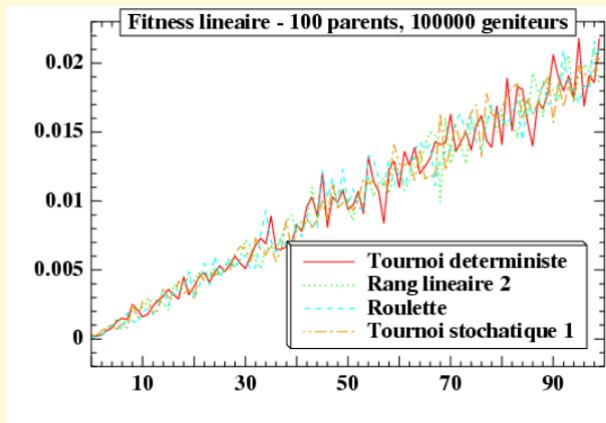
Rang et tournois





Sélection : illustration (5)

Variance





Moteur AG standard

Historiquement :

- Sélectionner P individus par sélection stochastique

avec remplacement

- Générer P enfants à l'aide des **opérateurs de variation**

- Remplacer les parents par les enfants

remplacement générationnel

Paramètres: La sélection – et ses paramètres

- Proportionnelle (roulette):
 P pression sélective et mise à l'échelle
- Tournoi : T ou t



Moteur “Steady State GA”

Historiquement (SSGA)

- Sélectionner 1 individu par tournoi
- Générer 1 enfants à l’aide des **opérateurs de variation**
- Remplacer 1 parent par l’enfant par “anti-tournoi”

Paramètres: Tailles des tournois sélection et remplacement

Note: Autres stratégies de remplacement – le pire, le plus vieux, ...



Sélection Stochastique

(III) Le tournoi EP

Utilisé comme remplacement

Chaque individu est sélectionné au plus une fois ...

- Taille du tournoi $T \geq 2$
- Pour chaque individu I
 - Choix uniforme de $T - 1$ individuals
- Score(I) = # I “battus”
- Choisir (déterministiquement) les meilleurs scores

avec ou sans remplacement?



Le moteur EP

Historiquement : L'algorithme EP (Evolutionary Programming)

- Sélection uniforme (\equiv pas de sélection)
- Chaque parent donne naissance à **un** enfant par mutation seulement
- Remplacement:
Sélection de P individus parmi les P parents + P enfants par tournoi EP

Paramètres: Taille du tournoi T

Note: Ressemble à un moteur $(P + P) - ES$



Plan

- Contexte Optimisation
- L'algorithme le paradigme biologique
- Un exemple sans ordinateur
- Points-clé Diversité, exploitation vs exploration, représentation
- Les étapes
 - Evaluation, arrêt
 - Moteur d'évolution Darwinisme artificiel
 - **Initialisation, opérateurs de variation**
Chaînes de bits, nombres réels et arbres
- Les instances historiques
- Conclusions et références



La partie **opérateurs de variation** est transférée dans les cours suivants.



Plan

- Contexte Optimisation
- L'algorithme le paradigme biologique
- Un exemple sans ordinateur
- Points-clé Diversité, exploitation vs exploration, représentation
- Les étapes
 - Evaluation, arrêt
 - Moteur d'évolution Darwinisme artificiel
 - Initialisation, opérateurs de variation Chaînes de bits, nombres réels et arbres
- **Les instances historiques**
- Conclusions et références



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, rang, tournoi} $\longrightarrow X_i', i = 1, \dots, P$
 - **Croisement** {1,2,N}-points, 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'' ,
 $b_i \longrightarrow \bar{b}_i$, probabilité p_m $\longrightarrow X_i^{t+1}, i = 1, \dots, P$

Remplacement générationnel

Originellement, roulette et 1-point



Evolutionary Programming

L. Fogel – 63

- Tout espace de recherche Ω
... où l'on peut définir un opérateur de mutation
- Initialisation uniforme (?)
- Boucle des générations : Population $X_i^t, i = 1, \dots, P$
 - Appliquer la **mutation** une fois à chaque parent
 $\longrightarrow Y_i, i = 1, \dots, P$
 - **Évaluation** des $\mathcal{F}(Y_i^t), i = 1, \dots, P$
 - **Tournoi EP** (stochastique) parmi les parents + enfants
 $\{X_i^t, i = 1, \dots, P\} \wedge \{Y_i^t, i = 1, \dots, P\}$

Note: Originellement, automates à états finis, et remplacement déterministe



Stratégies d'évolution

Rechenberg and Schwefel – 1965

- $E = \mathbb{R}^N$ or $\prod [a_i, b_i]$, minimization
- Initialisation uniforme $X_i^0, i = 1, \dots, \mu$
- Boucle des générations : Population $X_i^t, i = 1, \dots, \mu$

– **Évaluation** des $\mathcal{F}(X_i^t), i = 1, \dots, \mu$

– Faire λ fois

- Générer un enfant X_j' par **croisement**

Discret ou intermédiaire, Global ou “local”

Sur les variables et les “ σ ”

- Mutation adaptative de X_j' $\longrightarrow Y_j, j = 1, \dots, \lambda$

– **Sélection** déterministe de $X_i^{t+1}, i = 1, \dots, \mu$

Parmi les λY_j $(\mu, \lambda) - ES$

Parmi les μX_i^t et les λY_j $(\mu + \lambda) - ES$

Note: Originellement, 1+1, pas de croisement(!), règle des 1/5



Genetic Programming

J. Koza – 92

- Espaces d'arbres représentant des programmes Pseudo-LISP
- Initialisation non triviale “ramp half-and-half”
- Tout moteur d'évolution type GA SSGA

Note:

- Une communauté en pleine expansion
- Approche morphogénétique systématique
- Résultats fascinants Synthèse de circuits analogiques



Les racines - résumé

	GA	ES	EP	GP
Représentation	$\{0, 1\}^N$	\mathbb{R}^N	FSA	Arbres
Sél. reproduction	Oui	Non	Non	Oui
Natalité	1	1-10	1	SSGA-like
Croisement	Oui	Pourquoi pas?	Jamais !	OUI
Mutation	Mineure	Adaptative	Oui	NON
Sél. environment	Generationnel	Déterministe	Tournoi	SSGA

Tendance Moderne : une boîte à outils générale



Recuit Simulé

Minimisation

- Individu $x_k \in E$ (Pas de sélection!)
- Choix de x_{k+1} dans un voisinage de x .
- Loi de Boltzmann:
 $\Delta = F(x_{k+1}) - F(x)$
Probabilité d'accepter $x_{k+1} : \exp(-\frac{\Delta}{T})$,
 T "température" $\rightarrow 0$ qd $k \rightarrow \infty$
 - Toujours accepté si $\Delta < 0$
 - De moins en moins d'augmentations acceptées qd $k \rightarrow \infty$.

Choix du voisinage ?

Comment faire varier T ?



Schémas de recuit

- Metropolis:

$$T_k = \frac{T_0}{\ln(k)}$$

Convergence (Geman & Geman 1984)

- Schéma exponentiel (trés utilisé)

$$T_{k+1} = cT_k$$

- RS adaptatif (ASA, Ingber 1989)

$E = \mathbb{R}^D$. Une température par coordonnée.

$$T_i(k) = T_{0i} \exp(-c_i k^{1/D})$$

+ techniques adaptatives.



Recherche Taboue

Glover 89-95

Espaces discrets

- Petites populations (1 historiquement).
- Sélection du(es) meilleur(s).
- Création des enfants dans un voisinage.
- Utilisation de l'**histoire de la recherche** pour éliminer des candidats-enfants
- Remplacement des parents par les meilleurs enfants.
- On garde trace du (des) meilleurs individus rencontrés.

Choix du voisinage ?

Gestion de l'histoire ?



Histoire

- Histoire récente:
Liste taboue des derniers essais.
Décourage les mouvements vers des attributs (variables) présents dans les derniers points examinés.
- Histoire fréquente:
Décourage les mouvements vers des attributs (variables) souvent rencontrés.

Stratégies d'intensification

Après un nombre d'essais sans amélioration, on repart de "vieux" optima rencontrés.

Stratégies de diversification

On encourage les mouvements vers des attributs (variables) rarement vus.



Plan

- Contexte Optimisation
- L'algorithme le paradigme biologique
- Un exemple sans ordinateur
- Points-clé Diversité, exploitation vs exploration, représentation
- Les étapes
 - Evaluation, arrêt
 - Moteur d'évolution Darwinisme artificiel
 - Initialisation, opérateurs de variation Chaînes de bits, nombres réels et arbres
- Les instances historiques
- Conclusions et références



Suite du cours

Hybridisation

- Représentations et opérateurs (MS)
Couplage avec des méthodes locales
- Optimisation continue (AA)
Les stratégies d'évolution, Covariance Matrix Adaptation
- Quelques résultats théoriques (AA)
Temps d'atteinte (cas discret), estimations d'erreur (cas continu).
- Techniques avancées (MS) Optimisation multi-modale, prise en compte des contraintes, Optimisation multi-critères, Parallélisation
- Robotique évolutionnaire (NB)
 - Évolution de contrôleurs
 - Évolution et embryogénèse



Conclusions provisoires

Echecs:

- J'ai essayé en boîte noire, ça ne marche pas...
- J'ai essayé sur un pb résolu, c'était ridiculement lent, comparé à ...

Contextes recommandés:

- Problèmes non résolus fonctions chahutées, contraintes chahutées
- Plusieurs optima critères implicites, multi-critères
- Problèmes (très) mal posés validation de l'utilisateur
- A coupler avec des méthodes locales avec mesure



Faire attention à

- La représentation du problème le coeur du problème
- Les opérateurs d'évolution le croisement a-t-il un sens ?
Principe de causalité forte
préférez le sur-mesure
- Choix des paramètres Force de sélection
Taille population
probabilité des opérateurs
- Le calcul de la fonction \mathcal{F} S'il y a un bug, l'évolution le trouvera...

Travaillez, prenez de la peine !

Bibliography



LIVRES: Si vous n'en lisez qu'un :-)

- A.E. Eiben and J.E. Smith *Introduction to Evolutionary Computing*, Springer Verlag, 2003
- W. Banzhaf, P. Nordin, R.E. Keller, F.D. Francone, *Genetic Programming — An Introduction On the Automatic Evolution of Computer Programs and Its Applications*, Morgan Kaufmann, 1997.

LIVRES: Les ancêtres

- L.J. Fogel and A.J. Owens and M.J. Walsh, *Artificial Intelligence Through Simulated Evolution*, John Wiley & Sons, 1966
- J. Holland, *Adaptation in natural and artificial systems*, University of Michigan Press, Ann Arbor, 1975
- H.-P. Schwefel, *Numerical Optimization of Computer Models*, John Wiley & Sons, 1981 (Nelle édition 1995).
- D. E. Goldberg, *Genetic algorithms in search, optimization and machine learning*, Addison Wesley, 1989 Classique, mais dépassé
- J. Koza, *Genetic Programming*, I, II (& III), MIT Press, 1992, 1994 (& 1999) ... + videos!



Les référence

- Z. Michalewicz, *Genetic Algorithms+Data Structures = Evolution Programs*, Springer Verlag, 1992–1996. Toujours jeune !
- D. B. Fogel, *Evolutionary Computation, Toward a New Philosophy of Machine Intelligence*, IEEE Press, 1995.
- Melanie Mitchell, *An Introduction to Genetic Algorithms*, MIT Press, 1996.
- T. Bäck, *Evolutionary Algorithms in theory and practice*, New-York:Oxford University Press, 1995.
- Th. Bäck and D.B. Fogel and Z. Michalewicz, *Handbook of Evolutionary Computation*, Oxford University Press, 1997.
- Michael D. Vose, *The Simple Genetic Algorithm: foundations and theory*, MIT Press, 1999.
- D. E. Goldberg, *The Design of Innovation: Lessons from and for Competent Genetic Algorithms*, Kluwer Academic Publishers, 2001.



Journaux

- Evolutionary Computation MIT Press
- Transactions on Evolutionary Computation IEEE
- Genetic Programming and Evolvable Hardware Journal Kluwer
- Theoretical Computer Science – TCS-C – Theory of Natural Computing Elsevier
- Natural Computing Kluwer
- Applied Soft Computing Elsevier
- BioSystems Elsevier
- Journal of Heuristics Kluwer
- Journal of Global Optimization (electronic)



Conférences

Historiquement

- **ICGA** Int. Conf. on Genetic Algorithms
Depuis 1987, USA
- **PPSN** Parallel Problems Solving from Nature
Depuis 1990, Europe
Paris 2000!!!
- **EP** Annual Conf. on Evolutionary Programming
Depuis 1992, USA
- **ICEC** IEEE Int. Conf. on Evolutionary Computation
Depuis 1994, World-wide
- **GP** Genetic Programming Int. Conf.
Depuis 1996, USA



Aujourd'hui

- GP + ICGA = **GECCO** (annuelle depuis 1999)
Genetic and Evolutionary Computation Conference
Conférence **ACM** depuis ... 2005
- ICEC + EP + Galesia = **CEC** (annuelle depuis 1999)
Congress on Evolutionary Computation
- **EuroGP, EvoCOP et EvoNet Workshops** (annuels depuis 1998)
European Conf. on Genetic Programming
Evolutionary Combinatorial Optimization
Application of Evolutionary Computation Workshops
- **EA** (tous les 2 ans depuis 94) Evolution Artificielle
Toulouse 94, Brest 95, Nîmes 97, Calais 99, Dijon 01, Marseille 03, Lille 05,
Tours 2007
- ... et toujours **PPSN** !