L2 – Vie Artificielle

Alexandre Allauzen – Michèle Sebag
LIMSI – LRI

8 nov. 2013

Overview

Résumé

Optimisation

$$\text{Trouver } \mathit{ArgMax}\ (\mathcal{F}), \quad \mathcal{F} : \Omega \mapsto \mathbb{R}$$

- Ω: espace de recherche
- $\mathcal{F}$: fonction objectif

Algorithmes

- Classiques:

Résumé

Optimisation

$$\text{Trouver } \mathit{ArgMax}\ (\mathcal{F}), \qquad \mathcal{F} : \Omega \mapsto \mathbb{R}$$

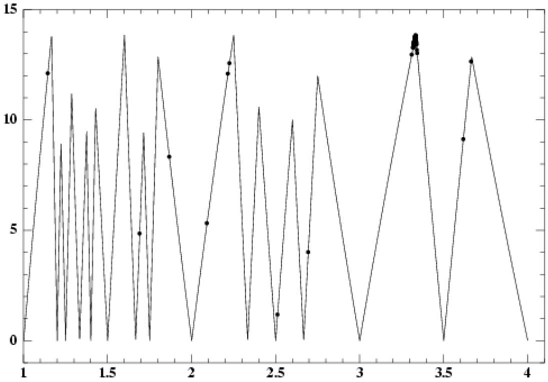
- Ω: espace de recherche
- $\mathcal{F}$: fonction objectif

Algorithmes

- Classiques: gradient
- Algorithmes d'évolution

Un problème d'ingénieur; $\mathcal{F}$ très chahutée

Espace de recherche : Interféromètres   Positionner des antennes
Objectif : Maximiser la tolérance en conservant la précision.



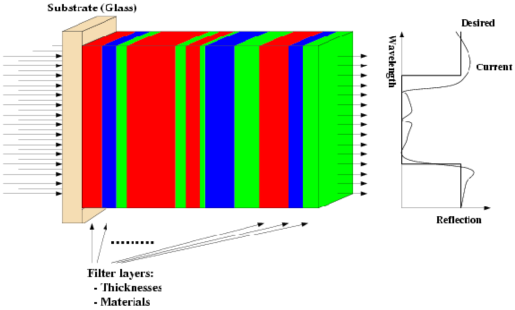
Cas de 3 antennes, $\mathcal{F} =$ Marge d'erreur (position 2ème antenne)

Ω mixte : discrets $\times$ réels

Espace de recherche : Filtres optiques

(matériau,épaisseur)$_1$ . . . (matériau, épaisseur)$_N$

Objectif : Répondre au gabarit fixé.

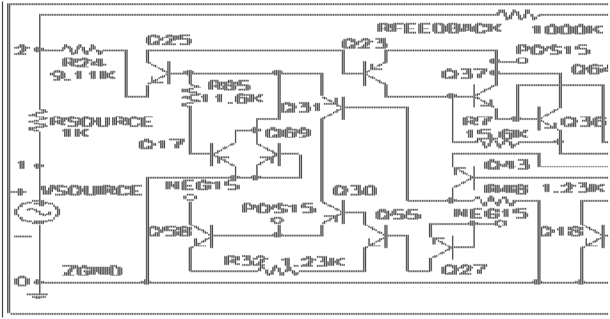


Ω = Circuits analogiques

Espace de recherche : Circuits analogiques

Réseau de transistors, diodes, résistances

Objectif : Fonctionalités fixées e.g. extraction de racine cubique

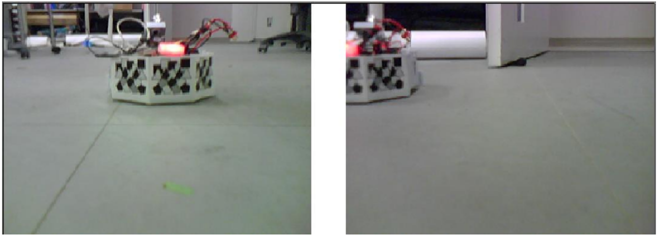


Ω = Réseaux neuronaux

Espace de recherche : $\mathbb{R}^d$

vecteur des poids du réseau

Objectif : un contrôleur de robot.



$\mathcal{F}$ non calculable

Espace de recherche : Mélanges de café

Objectif : Retrouver un arôme

$\mathcal{F}$ = note de l'expert

Programming by Optimization

http://www.prog-by-opt.net/

Overview

Optimisation

Le problème

$$\text{Trouver } x^* \ / \ \mathcal{F}(x^*) = Sup\{\mathcal{F}(y), y \in \Omega\}$$

E espace mesuré, $\Omega \subset E$, $\mathcal{F} : \Omega \mapsto \mathbb{R}$
$\mathcal{F}$ est la **Fonction objectif**.

Maximum global :

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

Maximum local :

$$x^* \text{ t.q. } (\exists \varepsilon > 0)(\forall x \in B(x^*, \varepsilon) \cap \Omega)\mathcal{F}(x^*) \geq \mathcal{F}(x)$$

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

Algorithmes d'optimisation

- Algorithmes de gradient
- Hill-Climbing
- Méthodes énumératives
- Méthodes stochastiques (méta-heuristiques)

Critères de comparaison

- Type d'espace de recherche
- Régularité de la fonction objectif (contraintes)
- Recherche locale – recherche globale

Méthodes de gradient

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

Contexte :

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

Conditions nécessaires :

- Fonction régulière, — $\mathcal{F}$ convexe .. optimum unique
- ou Connaissance a priori — X_0 bien choisi

Méthode locale, non applicable sur un espace discret ou mixte

Hill-Climbing

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

Contexte :

- Tout espace Ω "naturel"
- Toute fonction $\mathcal{F}$

Conditions nécessaires :

- X_0 doit être bien choisi

Méthode locale, coûteuse, détermine le plus proche optimum local.

Méthodes de type énumératif

Parcourir l'espace suivant un ordre déterministe

Contexte :

- Espace fini : $\Omega \equiv [1..N]$ *Attention* : opt.$_{discret}$ $\neq$ arrondi(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, non fiable pour des problèmes continus.

Méthodes stochastiques

Sélectionner X_i dans l'espace Ω

D'après une distribution de probabilité

- Monte-Carlo — X_i tiré avec une loi uniforme
- Recuit Simulé — Kirkpatrick, Gelatt and Vecchi, 1983
- Recherche Taboue — F. Glover – 1977 & 1989
- Algorithmes évolutionnaires — depuis 1965

Méthodes globales, mais TRÈS coûteuses.

Overview

Algorithmes d'Evolution : 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 cost

- **Programmation Génétique**

  J. Koza, 92 – a late root!

  GAs on parse-trees

Aujourd'hui:

"Evolutionary Computation"

Paradigme Darwinien

- Sélection naturelle

  avantage aux espèces adaptées à leur environnement
- Variabilité

  parents $\rightarrow$ enfants par petites déviations apparament non dirigées.
- "Objectif"

  capacité de survivre et de se reproduire
- Adaptation résultante

  apparition d'espèces (e.g. bacteries resistantes).

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 :

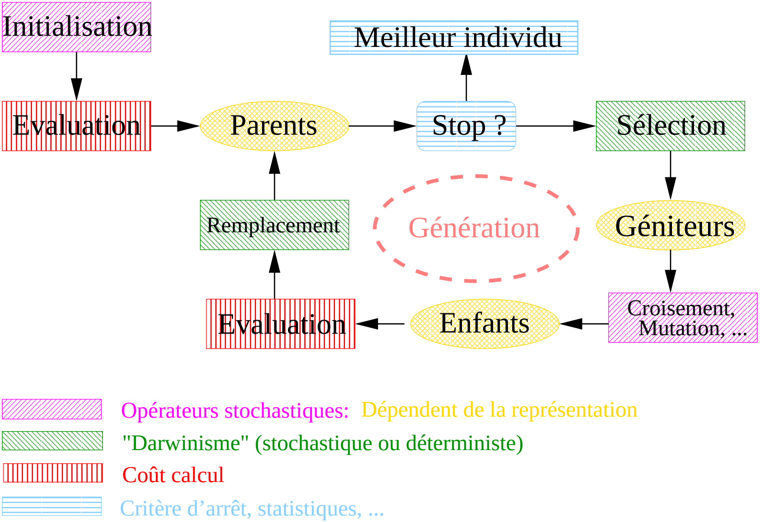
Individu	Elément X de Ω
Performance	Valeur de $\mathcal{F}(X)$
Population	Ensemble de P éléments de Ω
Génération	Passage de la population Π_i à Π_{i+1}

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.

$\equiv$ les **optima** de $\mathcal{F}$...

Algorithmes d'évolution : Le Squelette



Overview

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 enlisement de la recherche

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

$$\Longrightarrow \textit{Convergence Prématurée}$$

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 explorer tout l'espace.

Excès d'exploitation $\Longrightarrow$ **Convergence prématurée**
enlisement dans un optimum local

Excès d'exploration $\Longrightarrow$ **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 $\mathbb{R}^N$
- Représentation par arbres Genetic Programming

Le choix de la représentation est **crucial**

sera mis en avant dans tout le cours.

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

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

Pour : Simplicité
Analogie avec les *chromosomes* (Croisements)
quelques arguments théoriques

Contre : *No Free Luch Theorem*
d'autres arguments théoriques !

Représentation réelle

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

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

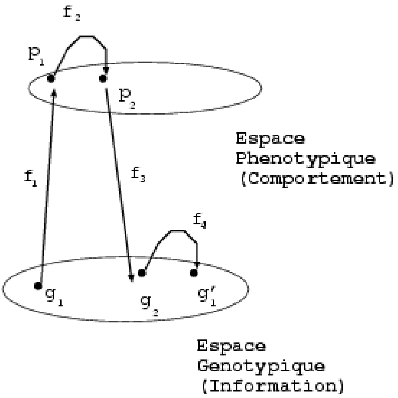
- Tout problème paramétrique — i.e. numérique
- . . . et bien d'autres ! — Comparer avec les méthodes déterministes

Pros : "Naturel"
quelques arguments théoriques

Cons : *No Free Luch Theorem :-)*
cache la forêt

Phénotypes – Génotypes

Lewontin – 1974



- Espace phénotypique : Evaluation, sélection
- Espace génotypique : Croisement, mutation

Préhistoire : Génotypes binaires universels

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

Overview

Évaluation de la population courante

DE TRES LOIN l'étape la plus coûteuse

- Ne pas recalculer $\mathcal{F}(X)$ inutilement
- utiliser une estimation de $\mathcal{F}$ voire la construire à la volée
- ...mais pas trop longtemps 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 resssources — 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 ecarts-types, médianes, ...

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

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
- Proportionelle — 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
- Choix parmi les enfants seulement, ou conflit de générations
- Peut également être déterministe ou stochastique

Sélection et remplacement

- sont deux étapes de "sélection"
- le remplacement sélectionne 0 ou 1 fois chaque individu

Attention:

Darwinisme artificiel $\equiv$ sélection + remplacement

Moteur d'évolution

Elitisme

Garantir la survie du meilleur ?

- *Théorie* : Non — Il faut accepter de perdre pour gagner
  cf théorie des $(\mu, \lambda) - ES$
- *Application* : Oui — conserver le meilleur en mémoire

Elitisme fort: quelques meilleurs survivent de toute façon

Elitisme faible:

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

Remplacer le pire de Π_{t+1}

par le meilleur de Π_t

Heuristique d'exploitation

Sélection Déterministe

Historiquement : Les moteurs d'évolution $(\mu \overset{+}{,} \lambda)$-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

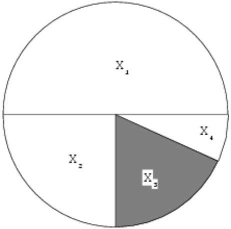
Objet : Tirer P parents parmi $\Pi_t = \{X_1, \ldots, X_P\}$

On tire P fois un individu avec replacement

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

Sélection Stochastique

(II) Le tournoi

- **Tournoi (déterministe)**
  - Tournoi de taille $T \in \mathbb{N}$
  - Choix uniforme de T individus   avec ou sans replacement?
    Rendre le meilleur
- **Tournoi stochastique (binaire)**
  - 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

Initialisation de la population

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

- Par tirage uniforme dans Ω $\quad \Omega = \{0,1\}^N$, $X_i^j = 0$ 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" $\mathcal{F}_1, \mathcal{F}_2, ..$

La diversité génétique est essentielle !

Opérateur d'exploration

Reproduction

- parents sélectionnés $\longrightarrow$ enfants . . .
- par application d'opérateurs de variation Généralement stochastiques
- Choix parmi les opérateurs
  - Séquentiel : Les opérateurs sont appliqués successivement avec une certaine **probabilité**
  - Proportionel : sélection à la roulette d'un opérateur Selon des **priorités**
  - ou toute combinaison

Opérateurs de variation

Suivant l'arité on définit

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

Le croisement

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

- *Analogue*: 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

Le croisement binaire

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

- Croisement à 1 point — J. Holland

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

- Croisement à 2 points — DeJong

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

- Croisement uniforme — Syswerda
  Echanger les bits avec probabilité 0.5 *indépendamment pour chaque position*

Croisement destructeur :
Attention à l'influence de la représentation

Croisement réel

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

Croisement arithmetique, barycentrique, intermediaire, . . .

Eshelman-Schaffer (BLX-α), Michalewicz, Radcliffe . . .

Idée de base : combinaison linéaire des parents

- Globalement
$$(\vec{X}, \vec{Y}) \rightarrow \alpha\vec{X} + (1-\alpha)\vec{Y}$$
avec $\alpha = U([0,1])$
- Coordonnée par coordonnée
$$(\vec{X}, \vec{Y}) \rightarrow (\vec{\alpha}_i X_i + (1-\alpha_i)Y_i)$$
avec $\alpha_i = U([0,1])$ VA indépendantes.
- Application contractante (perte de diversité)
Extension : $\alpha_i = U([-0.5, 1.5])$. Attention aux bornes des variables

Croisement: discussion

Propriétés:

- **Opérateur d'exploitation**
- Recombinaison des "bonnes" parties
- Effets desctructeurs

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: $\Omega \rightarrow \Omega$

- *Analogue*: reproduction asexuée.
- *Intuition* : les enfants du croisement sont limités par Π_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?*
  De nombreuses mutations heureuses mutation plus faible
  sinon mutations plus forte
- *Débat* : Un opérateur homéopathique des AGs
  L'opérateur majeur pour ES et EP

Les mutations binaires

$\{0,1\}^N \rightarrow \{0,1\}^N$
Opérateurs unaires :-)

La mutation "bit-flip" Holland, Goldberg; since 80s

- Pour chaque individu
- et pour chaque position l,

$$(b_1, b_2, \ldots, b_N) \xrightarrow{p_m} (b_1, b_2, \ldots, \overline{b_l}, b_{l+1}, \ldots, b_N)$$

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

Mutations binaires (2)

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 “bit-flip” $=$ mutation standard
- Mutation déterministe parfois plus efficace

Quelle probabilité p_m ?

- De nombreuses mutations heureuses augmenter p_m
- sinon diminuer p_m

Mutations réelles

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

Idée: ajout de bruit Gaussien $X_i := X_i + N(0, \sigma)$

Tout l'art est dans le choix de σ!

- Basé sur la fitness (type EP)
  - Le meilleur mute peu Exploitation
  - Le pire mute beaucoup Exploration
- Basé sur l'historique Rechenberg, **règle des 1/5**
  - Une mutation est *réussie* si l'enfant est meilleur que le parent
  - $\tau = \#$ mutations réussies dans les dernières T générations
  - Si $\tau > 0.2$ $\sigma = 1.22\sigma$
    sinon $\sigma = 0.83\sigma$
- Mutations adaptatives (ES modernes) σ ajustée par l'évolution !

Mutations adaptatives

Les paramètres de mutation sont portés par l'individu, e.g.

$$\vec{X} = (X_1, \ldots, X_N,\ \sigma)$$

Une mutation en deux temps:

- Muter σ et tous les paramètres de mutation
- Muter X_i en utilisant la nouvelle valeur de σ

Les σ sont ajustées gratuitement !

Pourquoi ?

- Des mutations successives avec un σ aberrant ne peuvent pas être constament réussies
- Les individus qui survivent longtemps résultent de nombreuses mutations successives réussies
- Les individus qui survivent longtemps doivent donc avoir des "bons" σ

Le débat croisement – mutation

Le croisement:

- permet de grandes modifications
- les modifications dépendent de la population
- plutôt opérateur d'exploitation
- effets décroissants avec l'évolution

La mutation:

- nécessaire
- nécessité de pouvoir faire de grands pas
- plutôt opérateur d'exploration
- effets destructeurs augmentent avec l'évolution

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

Autres Aspects

liste non exhaustive

- Découverte de plusieurs optima Heuristiques de partage (sharing)
- Programmation génétique Evolution dans des espaces d'arbres
- Couplage avec des méthodes locales
- Optimisation sous contrainte
- Optimisation multi-critères
- ...

Conclusions provisoires

Echecs:

- J'ai essayé en boite 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 TRES mal posés — validation de l'utilisateur
- Préférer le sur-mesure — Couplage avec des méthodes locales

Faire attention à

- La Représentation du problème — on ne saurait y passer trop de temps
- 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 !