Algorithms Seminar 2001–2002,
F. Chyzak (ed.), INRIA, (2003), pp. 113–118.

Available online at the URL
http://algo.inria.fr/seminars/.

# Les algorithmes évolutionnaires : état de l'art et enjeux

*Marc Schoenauer*
Projet Fractales, Inria Rocquencourt (France)

October 15, 2001

*Summary by Philippe Dumas*

**Abstract**

Les algorithmes évolutionnaires sont des algorithmes d'optimisation stochastique fondés sur un parallèle grossier avec l'évolution darwinienne des populations biologiques. Ils fournissent une approche heuristique, à l'occasion performante et dans certains cas prouvée.

Un algorithme évolutionnaire a pour but d'optimiser une fonction réelle. Il repose sur une vision darwinienne relativement simpliste et une optimitisation stochastique résumées dans le diagramme de la Figure 1. La fonction $f$ à optimiser, appelée aussi performance, est définie sur un espace de recherche $\Omega$. L'algorithme fait évoluer une population, un sous-ensemble de l'espace de recherche. Cette évolution résulte d'une part d'un darwinisme artificiel, qui se manifeste par la sélection et le remplacement et ne dépend que de la performance $f$ ; d'autre part de l'effet du hasard, qui s'exprime dans l'initialisation et les opérateurs de variation et ne dépend que de la représentation de l'espace de recherche. L'idée fondamentale est que la sélection favorise les individus qui optimisent la performance et que les variations font apparaître dans la population sélectionnée des individus que l'ont peut espérer meilleurs au regard de la performance. Dans cette évolution, les générations successives de la population restent à taille constante et l'aspect stochastique ne dépend que de la génération précédente.

La mise en place d'un algorithme évolutionnaire est complexe et le coût de calcul est important. De tels algorithmes sont donc destinés à traiter des problèmes qui n'ont pas de solutions classiques. Si l'on veut bien négliger un discours pseudo-scientifique et des querelles de chapelles qui ont encombré le domaine dans ses premières décennies[1] il faut reconnaître à l'approche évolutionnaire des réussites frappantes.

[1]Ces aspects désagréables ne figuraient absolument pas dans l'exposé de M. Schoenauer.

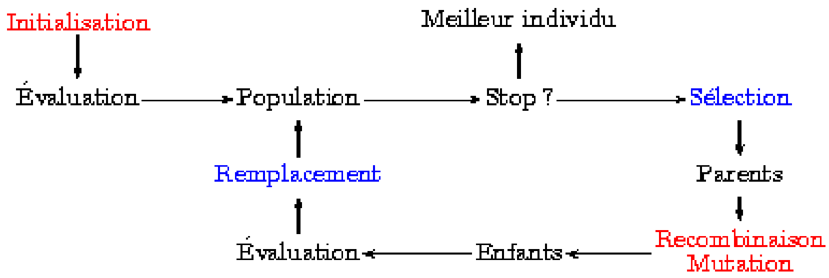


Figure 1. Variation (stochastique, en rouge) et darwinisme (déterministe ou stochastique, en bleu) sont les notions de base de l'algorithmique évolutionnaire.

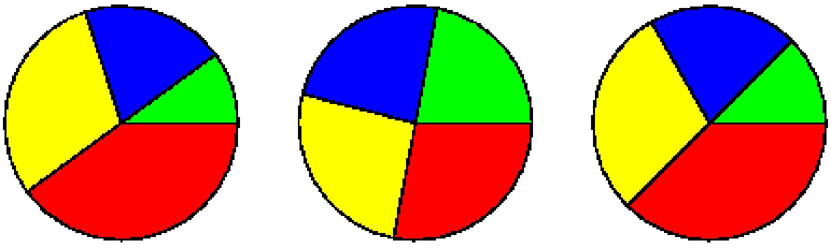

FIGURE 2. Une modification anodine de la fonction objectif peut avoir une effet marqué sur le processus de sélection par la roulette.

## 1. Représentation binaire

1.1. **Algorithme génétique classique.** Le modèle binaire [4] est le plus ancien et on parle à son sujet d'algorithme génétique classique. L'espace de recherche est l'ensemble des mots binaires de longueur donnée $\Omega = \{0,1\}^N$. Un élément de $\Omega$ est baptisé chromosome. La population est à chaque instant $t$ entier constituée de $\mu$ chromosomes $X_i^t$, qui forment un vecteur $X^t$. L'initialisation se fait suivant la loi uniforme. L'évaluation est simplement l'évaluation des $f\big(X_i^t\big)$. La sélection suit la méthode de la roulette : à chacun des chromosomes $X_i^t$ est associé sur un cercle un secteur d'angle proportionnel à la valeur $f\big(X_i^t\big)$ ; on effectue $\mu$ tirages suivant ce modèle probabiliste et à chaque fois que la roulette fournit le chromosome $X_i^t$, celui-ci est copié en un chromosome $X_j'$. On obtient ainsi une population intermédiaire $X'$, où les chromosomes les meilleurs au regard de la performance sont présents en plusieurs exemplaires, alors que les pires sont éliminés. Ensuite sont effectués des recombinaisons entre ces chromosomes : $\mu/2$ couples $\big(X_j', X_k'\big)$ sont tirés au hasard et avec une probabilité $p_c$ sont recombinés ; si la recombinaison a lieu, un point de croisement est tiré au hasard et les brins sont échangés. Après cela chaque chromosome subit une mutation : avec une probabilité $p_m$ chacun de ses bits est changé en son complément. Au terme de ce processus, on dispose de la génération suivante $X^{t+1}$. Dans ce modèle, il n'y a pas de remplacement.

Le réglage des paramètres comme $\mu$, $p_c$, $p_m$ ou le test d'arrêt est délicat. Indiquons un simple problème : l'optimisation de $f$ est équivalente à l'optimisation de $f' = af + b$ avec $a$ une constante positive et $b$ une constante. Cependant l'introduction de $a$ et $b$ peut avoir un effet drastique sur la sélection. On le voit sur la Figure 2, où l'on a supposé que la population comporte quatre individus. À gauche la performance prend les quatre valeurs 1, 2, 3, 4, d'où des secteurs de taille 0,1, 0,2, 0,3, 0,4. Au centre on a augmenté la performance de 10, ce qui donne les valeurs 11, 12, 13, 14 et des secteurs de taille 0,22, 0,24, 0,26, 0,28. Le tirage aléatoire est très fortement modifié et le meilleur individu est à peine avantagé. Ce problème peut être traité par une mise à l'échelle, qui est effectuée à chaque génération. Dans la mise à l'échelle linéaire, $a$ et $b$ sont choisis pour que la moyenne de la performance sur la population reste la même, $\overline{f}' = \overline{f}$, mais la meilleure valeur $f'_{\max}$ satisfait $f'_{\max} = \rho f_{\max}$ avec $\rho$ choisi entre 1 et 2. C'est ce qu'on a appliqué, avec $\rho = 1{,}5$ pour obtenir la version de droite, qui correspond à $a = 5/6$, $b = 5/12$ et aux quatre valeurs 1,25, 2,08, 2,9, 3,75 pour la performance.

1.2. **Algorithme évolutionnaire.** La représentation binaire, longtemps dominante, a été vivement critiquée car peu naturelle. Pour résoudre un problème numérique dont les solutions sont, par exemple, cherchées dans $[\,0,1\,]$ avec une précision de $10^{-10}$, elle amène à passer en représentation binaire avec des mots de trente-quatre bits (au moins), parce que 34 est le premier entier $k$ satisfaisant à $2^k \geq 10^{10}$. De plus les opérations peuvent ne pas être naturelles ou ne pas faire sens. Par exemple la recombinaison peut produire des chromosomes qui n'ont pas d'interprétation dans le modèle. Dans ce cas, on procède généralement à une pénalisation de ces chromosomes, qui vont ainsi être éliminés dans la sélection.

Ces questions ont modifié le point de vue des spécialistes, qui parlent maintenant d'algorithmes évolutionnaires. On cherche une représentation naturelle des données et des opérateurs génétiques qui font sens dans le problème. Un bon algorithme évolutionnaire utilise une représentation qui permet à un spécialiste du domaine d'application d'interpréter les caractéristiques de la population et de proposer des choix de paramètres qui font converger l'algorithme vers une solution.

## 2. **Représentation réelle**

Dans une représentation réelle, l'espace de recherche est une partie $\Omega$ d'un $\mathbb{R}^N$. La similitude avec le darwinisme se fait par des stratégies évolutionnaires [1] dans lesquelles les mutations sont au premier plan, alors que les algorithmes génétiques misent plutôt sur la recombinaison.

Les mutations reposent essentiellement sur l'ajout d'un bruit gaussien centré d'écart-type $\sigma$ à chaque coordonnées de l'individu, tout l'art étant dans le choix du $\sigma$. On peut par exemple appliquer la règle du cinquième. Une mutation est réussie si elle fournit un individu meilleur que son parent[2] ; on note $\tau$ la proportion de mutations réussies sur les $T$ dernières générations. Si $\tau$ est plus grand que $1/5$, on change $\sigma$ en $1{,}22\,\sigma$ (on élargit la recherche) ; sinon on change $\sigma$ en $0{,}83\,\sigma$ (on restreint la recherche). S'il y a beaucoup de mutations réussies, les individus sont près d'un optimum local, mais l'optimum global risque d'être manqué ; s'il y a peu de mutations réussies, la région explorée est trop vaste et il convient de la restreindre.

On envisage aussi des mutations adaptatives. Chaque individu porte un paramètre $\sigma$, ce qui signifie que l'espace de recherche est maintenant une partie de $\mathbb{R}^N \times \mathbb{R}_+$. La mutation s'effectue en deux temps : on mute d'abord l'écart-type $\sigma$ (en le multipliant par l'exponentielle d'une variable gaussienne), puis l'individu lui-même en utilisant le nouvel écart-type. Si la valeur de $\sigma$ est aberrante, elle ne va pas fournir de mutation réussie et l'individu va être éliminé. Par contre les individus qui survivent ont une bonne valeur de $\sigma$. L'idée que l'individu porte les paramètres de la stratégie évolutionnaire a été développée en adjoignant à chaque individu une matrice de corrélation entre ses coordonnées, pour que les différentes coordonnées ne soient pas traitées indépendamment.

La recombinaison se fait par barycentre $z = (1-\alpha)x + \alpha y$, avec un poids $\alpha$ qui peut être une variable aléatoire.

À partir de la population de $\mu$ individus, les mutations et recombinaisons produisent $\lambda$ nouveaux individus. La sélection est déterministe et fournit les $\mu$ individus de la génération suivante. Dans la stratégie $(\mu, \lambda)$, les $\mu$ meilleurs individus parmi les $\lambda$ nouveaux individus ($\lambda > \mu$) sont conservés. Dans la stratégie $(\mu + \lambda)$, ce sont les $\mu$ meilleurs parmi les $\mu + \lambda$ individus disponibles ($\lambda > 1$) qui sont conservés. On ainsi appliqué le processus de remplacement de la Figure 1.

## 3. **Autres représentations**

3.1. **Représentation non structurée.** La recherche de structures matérielles optimales amène à une discrétisation de l'espace en cellules par découpages par des plans parallèles aux plans de coordonnées. Une cellule fait partie de la structure si elle est marquée 1 et sinon 0. On a ainsi un tableau de bits. L'inconvénient est qu'une représentation fine demande une énorme place en mémoire. Une représentation plus compacte repose sur la notion de diagramme de Voronoï. On fixe un domaine borné et un individu est une famille de $N$ points marqués dans ce domaine. À cette famille est associé son diagramme de Voronoï, dans lequel la cellule attachée à un point de la famille est constituée des points de l'espace qui sont plus près de ce point que des autres points de

[2] On voit qu'une évaluation a lieu dans la sélection. Le diagramme de la Figure 1 illustre seulement une idée.

la famille. Si le point de référence est marqué 1, alors tous les points de sa cellule font partie de la structure ; si le point est marqué 0, alors les points de la cellule ne sont pas dans la structure.

On définit sur cette représentation des opérateurs d'évolutions. Par exemple la recombinaison de deux individus consiste à les couper tous les deux par un même plan de l'espace et à échanger les points marqués des deux individus qui sont dans l'un des demi-espaces limité par le plan.

3.2. **Représentation en arbre.** Un thème éculé du domaine est celui de l'ajustement. On dispose de données numériques et on cherche une fonction qui les modélise. Un espace de recherche possible est celui des arbres qui représentent des expressions de fonction. La recombinaison se fait par échange de sous-arbres. La mutation consiste en le remplacement d'un sous-arbre par un arbre aléatoire et en la mutation gaussienne des constantes. La performance tient compte à la fois de l'erreur quadratique dans l'ajustement et de la complexité de l'expression comptée à l'aide du nombre de nœuds de l'arbre.

3.3. **Programmation génétique.** Un pas supplémentaire dans la généralisation amène à accepter une population constituée de programmes informatiques. On parle alors de programmation génétique. Une éventuelle solution du problème posé est obtenu en appliquant un programme de la population à un embryon.

L'espace de recherche est constitué de programmes c'est-à-dire d'arbres étiquetés par des symboles de fonctions, qui représentent des fonctions mathématiques comme les fonctions et opérations usuelles (constantes, noms de variables, fonctions exp, ln, $\ldots$, opérations $+$, $\times$, $\ldots$) ou des fonctions informatiques (valeurs de vérité, conditionnement, boucle). La performance d'un programme se mesure sur la solution qu'il produit à partir de l'embryon. La population initiale peut être constituée de tous les arbres dont la profondeur est bornée par un certain entier. Les opérations génétiques sont celles que nous avons vues au sujet des arbres.

Cette idée a par exemple été utilisée dans la conception de circuits analogiques. L'embryon est un circuit simple et les fonctions qui apparaissent dans les programmes sont des modificateurs de circuits. La performance se mesure en testant le circuit pour un échantillon de fréquences. Il faut noter que pour la conception d'un filtre passe-bas 60 dB exposée dans [5], la population est de taille $\mu = 640\,000$, ce qui limite la programmation génétique à des exercices d'école.

## 4. Domaines d'application

L'approche évolutionnaire est valablement appliquée quand une technique classique n'est pas disponible ; quand le coût de calcul des méthodes standard est trop élevé ; quand la performance n'a pas les propriétés de régularité que requièrent les méthodes standard. Une démarche naïve sur des exemples d'école est donc sans intérêt. Citons trois exemples dans lesquels l'approche évolutionnaire montre son intérêt [3].

Le calcul du profil d'une aile d'avion est extrêmement coûteux. Une approche évolutionnaire, dans laquelle la performance est la différence entre la pression calculée et la pression désirée, a permis une amélioration de 14% de cette performance par rapport aux méthodes classiques.

Un algorithme génétique a permis de sélectionner quatre-cents chaînes peptidiques potentiellllement actives comme bactéricide. Cinq d'entre elles ont été synthétisées, par exemple pour concevoir de nouveaux additifs alimentaires anti-bactériens.

La radiothérapie utilise un faisceau radioactif qui détruit les tumeurs mais aussi des tissus sains. On savait prévoir la forme de la zone lésée suivant celle du faisceau. Un algorithme génétique a permis de déterminer la forme du faisceau à produire pour atteindre une zone de forme donnée.

| chromosome | état | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $s_1$ | $s_2$ | $s_3$ | $s_4$ | $s_5$ | $s_6$ | $s_7$ | $s_8$ | $s_9$ | $s_{10}$ |
| 00 | 2 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 01 | 0 | 1 | 0 | 0 | 2 | 1 | 1 | 0 | 0 | 0 |
| 10 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 2 | 1 | 0 |
| 11 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 2 |

FIGURE 3. La matrice $Z$ dans le cas où les populations sont constituées de deux chromosomes de longueur 2, les chromosomes et les états étant rangés dans un ordre lexicographique.

## 5. Théorie

**5.1. Schémas.** Un schéma $H$ est un mot sur l'alphabet $\{0, 1, \#\}$. Un mot binaire est une réalisation du schéma $H$, s'il coïncide avec le schéma pour les lettres différentes du joker $\#$. L'ordre d'un schéma est le nombre $\omega(H)$ de caractères 0 ou 1 qu'il contient et sa longueur utile $\lambda(H)$ est la distance maximale entre deux lettres de $H$ autres que le joker $\#$. Holland a prouvé l'énoncé suivant.

**Théorème 1** (dit des schémas). *La suite des $m(H,t)$, nombre de chromosomes qui réalisent le schéma $H$ à la génération $t$, satisfait à l'inégalité ($\mathbf{E}$ désigne l'espérance)*

$$\mathbf{E}\ m(H, t+1) \geq \mathbf{E}\ m(H,t)\frac{\overline{f}(H,t)}{\overline{f}(X^t)}\left(1 - \frac{\lambda(H)}{N-1}p_c\right)(1-p_m)^{\omega(H)}.$$

Dans cet énoncé $\overline{f}(H,t)$ est la valeur moyenne de la performance sur les réalisations du schéma dans la population, alors que $\overline{f}(X^t)$ est la moyenne de la performance sur la population. L'inégalité vient du fait que le schéma peut apparaître par mutation. Ce résultat est interprété de la manière suivante : un schéma de faible longueur utile, de faible ordre, dont la performance est supérieure à la moyenne, a un nombre de chromosomes qui augmente exponentiellement dans la population. Il fournit une explication à la convergence vers un optimum.

**5.2. Chaînes de Markov.** La théorie des schémas ne permet pas d'expliquer la composition de la population au cours des générations. Une approche par chaînes de Markov a été développée dans [6, 7]. Le nombre de chromosomes possibles est $2^N$. Chaque tirage avec remise de $\mu$ individus dans ces $2^N$ chromosomes fournit une population. Le nombre d'états de la chaîne de Markov est donc

$$\nu = \binom{\mu + 2^N - 1}{2^N - 1}.$$

Chaque état peut être vu comme une ligne d'une matrice $Z = (z_{c,s})$ (Figure 3), dans laquelle $z_{c,s}$ est le nombre d'occurrences du chromosome $c$ dans la population $s$.

**Théorème 2.** *Si la probabilité de mutation $p_m$ est non nulle, la chaîne de Markov associée à un algorithme génétique classique est ergodique. En particulier elle possède une distribution limite.*

Le théorème de Perron–Frobenius permet de préciser le comportement de la chaîne.

Cerf [2] a utilisé la notion de chaîne de Markov, mais avec une approche différente basée sur la théorie des perturbations stochastiques des systèmes dynamiques. Un algorithme génétique simpliste qui ne comporte pas d'opérateurs de variations et effectue une sélection dégénérée est perturbé selon un paramètre $\ell$ qui gouverne la probabilité de mutation ($p_m = \ell^{-a}/N$), la probabilité de recombinaison ($p_c = \ell^{-b}$) et aussi le mode de sélection. Cette approche fournit un seuil pour la taille de la population, dans le cas inhomogène où $\ell$ dépend du temps.

**Théorème 3.** *On suppose que la suite* $(\ell_t)$ *tend vers* $+\infty$ *et que la taille* $\mu$ *de la population est supérieure à une valeur critique* $\mu^*$ *qui s'exprime à l'aide de la performance et des paramètres du modèle. Il y a alors équivalence entre les deux assertions :*

– *il existe des exposants* $\alpha$ *et* $\beta$ $(0 < \alpha < \beta)$ *satisfaisant à* $\displaystyle\sum_{t=0}^{+\infty} \frac{1}{\ell_t^\alpha} = +\infty$ *et* $\displaystyle\sum_{t=0}^{+\infty} \frac{1}{\ell_t^\beta} < +\infty$ *;*
– *pour toute population originelle dans l'espace de recherche, en un temps fini, la population est presque sûrement toute entière dans l'ensemble des individus qui fournissent l'optimum global de la performance.*

Ce résultat n'a pas de conséquence pratique mais il a comme corollaire que la taille critique de la population est de l'ordre de $N$, ce que Goldberg avait obtenu expérimentalement.

5.3. **Convergence des stratégies évolutionnaires.** En s'appuyant sur la théorie des surmartingales, Rudolph [8] a prouvé un énoncé qui donne en particulier le résultat suivant.

**Théorème 4.** *Une stratégie évolutionnaire, de type* $(1, \lambda)$ *avec* $\lambda \geq 2$ *et des mutations sphériques (les mutations font passer d'un point* $x$ *de* $\mathbb{R}^N$ *à un point* $x + \varepsilon$ *où* $\varepsilon$ *est une variable aléatoire uniforme sur une sphère de centre* 0 *de rayon adapté à la performance) et une performance strictement convexe au voisinage du point* $X^*$ *qui fournit l'optimum, converge presque sûrement et en moyenne vers* $X^*$*. De plus la convergence est géométrique.*

5.4. **Problèmes.** Les énoncés qui viennent d'être cités ne donnent qu'une faible idée de ce qui a été produit. Cependant il faut conclure que l'approche évolutionnaire manque de fondement théorique. Ceci est dû au fait que dans chaque cas les algorithmes sont adaptés à la situation par des choix de paramètres et des variantes que ne couvrent pas la théorie. Il en résulte par exemple que les dynamiques associées à des représentations différentes d'un même problème sont différentes et ne sont pas comparables faute d'un cadre adapté.

## Bibliography


[1] Bäck (Thomas), Hoffmeister (Frank), and Schwefel (Hans-Paul). – A survey of evolution strategies. In Belew (Richard K.) and Booker (Lashon B.) (editors), *Proceedings of the Fourth International Conference on Genetic Algorithms, San Diego, CA, USA, July 1991.* pp. 2–9. – Morgan Kaufmann, 1991.

[2] Cerf (Raphaël). – *Une théorie asymptotique des algorithmes génétiques.* – Thèse de doctorat, Université de Montpelier II, 1994.

[3] Evolution@work. – Disponible en ligne à `http://www.evonet.polytechnique.fr/evoweb/resources/evolution_work/`.

[4] Holland (John H.). – *Adaptation in natural and artificial systems.* – University of Michigan Press, Ann Arbor, Mich., 1975, ix+183p. An introductory analysis with applications to biology, control, and artificial intelligence.

[5] Koza (John R.), Andre (David), Bennett, III (Forrest H.), and Keane (Martin A.). – Evolution of a low-distortion, low-bias 60 decibel op amp with good frequency generalization using genetic programming. In Koza (John R.) (editor), *Late breaking papers at the Genetic Programming 1996 conference, Stanford University, July 28-31, 1996*, pp. 94–100. – 1996. Disponible en ligne à `http://www.genetic-programming.com/jkpubs96.html`.

[6] Liepins (Gunar E.) and Vose (Michael D.). – Representational issues in genetic optimization. *Journal of Experimental and Theoretical Artificial Intelligence*, vol. 2, n° 2, 1990, pp. 4–30.

[7] Nix (Allen E.) and Vose (Michael D.). – Modelling genetic algorithms wit Markov chains. *Annals of Mathematics and Artificial Intelligence*, vol. 5, n° 1, 1992, pp. 79–88.

[8] Rudolph (Günter). – Convergence of non-elitist strategies. In *Proceedings of the First IEEE Conference on Evolutionary Computation, IEEE World Congress on Computational Intelligence, June 27-29, 1994, Orlando, Florida, USA*, pp. 63–66. – 1994.

[9] Schoenauer (Marc) and Michalewicz (Zbigniew). – Evolutionary computation. *Control and Cybernetics*, vol. 26, n° 3, 1997, pp. 307–338. – Disponible en ligne à `http://www.eeaax.polytechnique.fr/papers.html`.