UNIVERSITÉ CLAUDE BERNARD

LYON I

Contributions à l'étude probabiliste des algorithmes d'évolution

Thèse soutenue le 21 décembre 2001

par

Jean BÉRARD

en vue de l'obtention du

Diplôme de Doctorat

(arrêté du 30 mars 1992)

Spécialité : Mathématiques

Composition du jury :

Raphaël CERF	**Professeur (Université Paris-Sud), Rapporteur**
André GOLDMAN	**Professeur (Université Lyon 1)**
Alice GUIONNET	**Chargée de recherche CNRS (ENS Lyon)**
Christian MAZZA	**Professeur (Université Lyon 1), Directeur de thèse**
Didier PIAU	**Professeur (Université Lyon 1), Directeur de thèse**
Ofer ZEITOUNI	**Professeur (Technion Haifa, Israël), Rapporteur**

223-2001

Remerciements

Nombreuses sont les personnes qui ont contribué à la réalisation de ce travail, au cours des différentes étapes de son élaboration. Je n'en cite ici que quelques-unes, en espérant que tous ceux que j'oublie me pardonneront.

Je tiens tout d'abord à remercier Christian Mazza et Didier Piau, sous la direction collective desquels ce travail a été effectué. La grande ouverture d'esprit de Christian Mazza, et le bouillonnement d'idées qui le caractérise ont suscité mon enthousiasme pour l'étude mathématique de problèmes jusqu'alors peu ou mal formalisés, situés à l'interface entre plusieurs disciplines, et pour lesquels la théorie n'est pas encore écrite. L'exigence constante de Didier Piau, sa grande disponibilité, la confiance amicale et le soutien sans faille qu'il m'a toujours accordés ont fortement contribué à la réalisation de ma thèse, et je tiens à souligner combien j'ai apprécié son honnêteté intellectuelle et l'étendue de sa compétence mathématique. J'exprime également toute ma gratitude à André Goldman qui m'a accueilli au LaPCS il y a trois ans de cela. Son enthousiasme, sa détermination et son attention constante contribuent à créer au sein du laboratoire cette ambiance amicale dont j'ai tant bénéficié.

Raphaël Cerf et Ofer Zeitouni ont accepté d'être les rapporteurs de cette thèse, et je les remercie vivement du soin qu'ils ont apporté à la lecture du texte, des nombreuses remarques pertinentes qu'ils ont formulées, et de l'intérêt qu'ils ont ainsi manifesté pour mon travail.

Je remercie chaleureusement Alice Guionnet d'avoir bien voulu faire partie de ce jury.

Alexis Bienvenüe sait bien que sa contribution à ce travail ne se limite pas aux chapitres auxquels il a directement collaboré. Son amitié, sa bonne humeur et sa fantaisie de chaque jour sont pour beaucoup dans l'excellent souvenir que je garderai de ces trois années de thèse.

Je remercie également les collègues du groupe de travail sur les algorithmes génétiques constitué à Grenoble : Olivier François, Marc Joannidès et Éric Fontenas, avec la collaboration amicale desquels un chapitre de cette thèse a été écrit.

Enfin, *last, but not least,* je tiens à remercier Anne Perrut, qui, dans tous les rôles qu'elle a tenus - tour à tour professeur d'emacs, public pour exposé en cours d'élaboration, et collègue de bureau compatissante -, a toujours manifesté la gentillesse et l'attention pour lesquelles nous l'apprécions tous.

Table des matières

Introduction

Présentation générale

Motivations

Introduits formellement par J. Holland [39], les algorithmes génétiques forment une classe d'algorithmes stochastiques d'optimisation dont le principe repose sur une analogie avec les mécanismes biologiques de l'évolution (voir également D. Goldberg [34], et T. Bäck [3]). Dans le but de maximiser une *fonction d'objectif* $F : S \to \mathbb{R}$, définie sur l'espace d'états S, l'algorithme génétique fait évoluer une population de particules dans S, appelées *individus*, sous l'action répétée d'opérations de mutation et de sélection.

1. **Mutation :** chaque particule se déplace aléatoirement sur un site voisin, une structure de graphe sur S étant donnée.
2. **Sélection :** une nouvelle population de particules est créée à partir de la précédente. Les particules présentant de faibles valeurs relatives de F sont éliminées de la population, les particules présentant de grandes valeurs relatives de F sont plus fréquemment répliquées.

La mutation permet à l'algorithme d'explorer l'espace d'états, en autorisant la découverte de nouvelles solutions potentielles – éventuellement meilleures – au problème de la maximisation de F, et la sélection concentre l'exploration de S autour des "bonnes" solutions potentielles déjà découvertes. Un opérateur de mariage, ou *crossover*, qui crée de nouveaux individus par appariement d'individus existants, est parfois introduit, lorsque la structure de S permet de le définir de façon pertinente. Dans la suite, nous nous intéressons essentiellement aux algorithmes ne comportant que mutation et sélection.

Les algorithmes génétiques et leurs variantes (stratégies d'évolution, voir I. Rechenberg [64], H.-P. Schwefel [67], programmation évolutionnaire, voir L. Fogel, A. Owens et M. Walsh [31], programmation génétique, voir J. Koza [45]), que nous regrouperons sous le nom générique d'algorithmes d'évolution, sont couramment utilisés en pratique, et fournissent des résultats satisfaisants dans de nombreuses situations. Bien entendu, leurs performances ne peuvent se comparer à celles d'algorithmes spécifiques – lorsqu'il en existe –, qui exploitent les propriétés particulières de tel ou tel problème d'optimisation. L'intérêt des algorithmes d'évolution, outre leur facilité de mise en œuvre, est qu'ils continuent en général de fournir des résultats satisfaisants lorsque l'on ne dispose pas d'une connaissance suffisante du problème pour construire des algorithmes spécifiques efficaces. De nombreuses études expérimentales leur ont été consacrées, donnant lieu à un savoir-faire empirique concernant la manière d'ajuster les différents paramètres au problème considéré.

Parallèlement à cette motivation algorithmique, des modèles de ce type sont utilisés depuis longtemps en biologie, voir R. Fisher [30], M. Kimura [44], C. Neuhauser [55], en tant que description simplifiée des phénomènes d'évolution. La préoccupation principale est alors moins l'efficacité pratique dans un contexte d'optimisation, que la détermination précise des effets des opérateurs d'évolution (mutation, sélection et crossover) sur le comportement du modèle, notamment dans une perspective d'inférence statistique, voir S. Tavaré, R. Griffiths et P. Donnelly [71].

Enfin, les algorithmes génétiques présentent une analogie de forme avec les méthodes d'approximation particulaire des filtres optimaux dans le cadre du filtrage non-linéaire, dont l'utilisation s'est largement répandue au cours des dernières années, voir P. Del Moral [19].

Description formelle

Il existe de nombreuses manières de définir les opérateurs de mutation et de sélection, suivant que l'on autorise tout ou partie de la population à subir des mutations, ou que l'on rééchantillonne tout ou partie de la population par sélection, ou encore suivant le critère exact retenu pour conserver ou non les individus par sélection, voir D. Goldberg et K. Deb [36].

Nous nous restreignons dans toute cette thèse, à des mutations affectant indépendamment chaque individu de la population, et à une sélection de type proportionnel, définies ci-après. Ce choix n'est pas le plus général possible, mais il reprend les caractéristiques essentielles des algorithmes d'évolution, et présente suffisamment de souplesse pour englober de nombreux cas intéressants.

L'état de la population est représenté à tout instant par un élément X de S^p,

$$X = (X^{(1)}, \ldots, X^{(p)}),$$

p désignant la taille de la population à l'instant considéré.

- **Mutation :** $X \longrightarrow M(X)$ Par mutation, chaque particule $X^{(i)}$ évolue indépendamment des autres et effectue un pas d'une marche aléatoire sur S, une structure de graphe et un noyau markovien sur S étant fixés. Les particules mutées sont notées $M(X)^{(i)}$.

– **Sélection :** $X \longrightarrow S(X)$ Chaque $S(X)^{(i)}$, est choisi aléatoirement et indépendamment des autres dans l'ensemble $\{X^{(i)}\ 1 \leq i \leq p\}$, selon la loi

$$\frac{1}{Z}\sum_{i=1}^{p} G(X^{(i)})\delta_{X^{(i)}}, \quad \text{où} \quad Z = \sum_{i=1}^{p} G(X^{(i)}).$$

La fonction $G : S \to \mathbb{R}_+^*$ caractérise le poids accordé à un individu lors du rééchantillonage par sélection. Cette fonction, appelée *fitness* (adaptation), est naturellement reliée à F de telle façon que les éléments de S donnant de grandes valeurs à F donnent également de grandes valeurs à G. Dans les cas que nous considérons, ou bien $G(x) = F(x)$, ou bien $G(x) = \exp(\beta F(x))$, avec $\beta > 0$. Dans ce dernier cas, la constante positive β joue le rôle d'une température inverse, et contrôle l'intensité de la pression de sélection à laquelle est soumise la population. Plus β est grand, plus les individus donnant une grande valeur relative à F dans la population sont favorisés. Ce procédé de sélection est appelé *sélection de Boltzmann*, par analogie avec la mécanique statistique.

L'évolution temporelle des algorithmes que nous considérons est définie par l'itération d'étapes de mutation-sélection, c'est-à-dire $X_{n+1} = SM(X_n)$ ou $X_{n+1} = MS(X_n)$, suivant les cas.

Les particules $X_n^{(1)}, \ldots, X_n^{(p)}$ constituent la population à la date n, et les quantités auxquelles nous nous intéressons principalement sont, d'une part, la valeur maximale de la fonction d'objectif présente dans la population à la date n,

$$F^*(X_n) = \max_{i=1,\ldots,p} F(X_n^{(i)})$$

et, d'autre part, la moyenne empirique de ces valeurs dans la population :

$$F_n = \frac{1}{p}\sum_{i=1}^{p} F(X_n^{(i)}).$$

Résultats existants

Plusieurs approches heuristiques (voir J. Holland [39], H. Beyer [10], A. Prügel-Benett et J. Shapiro [61]) ont été proposées pour tenter d'analyser le comportement des algorithmes d'évolution dans un contexte d'optimisation, mais on ne dispose malheureusement que de peu de résultats mathématiques rigoureux dans cette direction. En particulier, l'interaction entre mutation, sélection et taille de la population demeure mal comprise. La raison principale tient naturellement à la difficulté même des problèmes d'optimisation auxquels ces algorithmes sont en général appliqués. Cependant, même des situations *a priori* plus simples se révèlent difficiles à analyser mathématiquement. En effet, les algorithmes d'évolution se présentent comme des chaînes de Markov en grande dimension, puisqu'ils décrivent l'évolution d'une population de particules en interaction, et non pas d'un seul individu, cette évolution étant généralement irréversible du fait de la sélection. En particulier, la forme des mesures invariantes est la plupart du temps inconnue. Par conséquent, les techniques usuelles d'analyse quantitative des chaînes de Markov (estimations spectrales ou inégalités de Sobolev logarithmiques, voir par exemple, P. Brémaud [14] chap. 6, ou P. Diaconis et L. Saloff-Coste [26]), ou de marches aléatoires (analyse harmonique et théorie du potentiel sur les graphes, voir P. Doyle et L. Snell [28], ou R. Lyons et Y. Peres [48]) sont en général difficiles à mettre en œuvre. Pour autant, un certain nombre de résultats rigoureux ont été obtenus dans divers contextes simplifiés, que nous décrivons à présent brièvement.

Théorie de Freidlin-Wentzell

Les premiers résultats de convergence asymptotique pour les algorithmes génétiques ont été obtenus par R. Cerf [15], [16], dans le cadre de la théorie de Freidlin-Wentzell des chaînes de Markov à transitions rares (voir également O. François [32]). Il s'agit d'une théorie perturbative, dans laquelle l'intensité de la sélection β tend vers l'infini, et le taux de mutation tend vers zéro. On obtient, lorsque β et le taux de mutation varient avec le nombre d'itérations, des résultats du type suivant :

$$\lim_{n \to +\infty} \mathrm{P}\left[F^*(X_n) = F_{max}\right] = 1,$$

c'est-à-dire la convergence en probabilité du maximum des valeurs de F présentes dans la population à la date n vers le maximum absolu de F sur S. Il existe également des versions de ces résultats pour les versions homogènes en temps de l'algorithme. Ces travaux mettent en évidence l'existence d'une taille de la population critique au dessus de laquelle le résultat ci-dessus est vérifié, et en dessous de laquelle l'algorithme ne se concentre plus asymptotiquement sur la valeur maximale de F. Il apparaît donc que la taille de la population doit être suffisament grande pour que l'algorithme fonctionne efficacement. Des estimations de cette taille de la population critique sont également disponibles.

Population infinie

Un autre cadre asymptotique dans lequel des résultats ont été obtenus est celui où la taille de la population est supposée infinie. Plus précisément, on peut montrer, sous des hypothèses très générales [20], que, lorsque la taille de la population p tend vers l'infini, la mesure empirique

$$\frac{1}{p} \sum_{i=1}^{p} \delta_{X_n^{(i)}}$$

converge en loi sur S vers une mesure de probabilité limite $\mu(n)$, pour tout instant n fixé. La suite des lois $(\mu(n))$ est appelée *algorithme génétique en population infinie*. L'intérêt de se placer dans ce contexte est que les fluctuations stochastiques liées à la taille finie de la population disparaissent. De fait, $\mu(n)$ satisfait une relation de récurrence explicite :

$$\mu(n+1) = KW(\mu(n)),$$

où $K : \mathcal{M}^1(S) \to \mathcal{M}^1(S)$ désigne l'action du noyau markovien de mutation sur les mesures de probabilités sur S, et où $W(\mu)$ est défini, pour toute mesure de probabilité μ sous laquelle G est intégrable, par

$$\frac{\mathrm{d}W(\mu)(x)}{\mathrm{d}\mu(x)} = \frac{G(x)}{\int G\mathrm{d}\mu}.$$

L'évolution temporelle de $\mu(n)$ est donc décrite par l'itération d'un *système dynamique à valeurs mesures* non-linéaire. Les propriétés asymptotiques de ce système dynamique lorsque n tend vers l'infini ont notamment été étudiées par M. Vose [73], voir également Y. Rabinovich et A. Wigderson [62], et C. Mazza et D. Piau [50]. Bien entendu, il reste à savoir si le comportement asymptotique de cet objet limite est relié à celui des algorithmes en population finie – qui sont les seuls réalisables en pratique –, au moins lorsque la taille de la population est suffisamment grande. Des résultats dans ce sens sont prouvés par P. Del Moral et A. Guionnet dans [21].

Complexité algorithmique

Nous nous intéressons particulièrement à la question de la complexité des algorithmes génétiques. Informellement : comment se comportent ces algorithmes lorsque le nombre d'itérations et l'espace mémoire disponibles sont limités ? La théorie de la complexité algorithmique fournit un cadre mathématique dans lequel ces questions, importantes en pratique, peuvent être formulées précisément, voir C. Papadimitriou [56]. On considère non plus un unique problème d'optimisation (S, F), mais une classe de problèmes d'optimisation de taille variable, le nombre d'itérations et l'espace mémoire utilisés par les algorithmes sur les problèmes de cette classe étant contraints à croître à une vitesse raisonnable avec la taille du problème. L'exemple le plus classique dans la théorie des algorithmes génétiques est celui de la classe de fonctions one-max définie par

$$S_q = \{0,1\}^q \ , \ F_q(x_1, \ldots, x_q) = x_1 + \cdots + x_q$$

où $q \geq 1$ paramètre la classe de fonctions considérée. La question typique que l'on se pose dans ce cadre est : quel est le comportement sur (S_q, F_q) de l'algorithme considéré si le nombre d'itérations et l'espace mémoire utilisés croissent au plus polynomialement avec q, lorsque q tend vers l'infini. Cette restriction est fondamentale, car, la taille de l'espace d'états S_q étant ici égale à 2^q, il est par exemple exclu pour l'algorithme d'explorer tout l'espace d'états pour trouver le maximum de F_q. La plupart des résultats de complexité algorithmique concernant les algorithmes d'évolution portent sur la fonction one-max ou des variantes (voir G. Rüdolph [66], et [62]), mais cette classe de problèmes ne présente, bien entendu, qu'un intérêt limité du point de vue de l'optimisation.

Il est à noter que la théorie de Freidlin-Wentzell se prête assez mal au traitement de ce genre de questions, car elle repose sur des approximations perturbatives effectuées à espace d'états fixé. D'autre part, les travaux de S. Arora [2] suggèrent que l'efficacité de l'approximation des algorithmes en population infinie par des algorithmes en population finie n'est pas garantie du point de vue de la complexité, même en ne considérant que des classes relativement restreintes de systèmes dynamiques à valeurs mesures. Ceci est confirmé par le fait que, dans certains cas concrets, les estimations fournies par [21] sur la taille de la population nécessaire pour obtenir une bonne approximation du comportement de l'algorithme en population infinie par des algorithmes en population finie, ont tendance à croître très rapidement avec la taille de l'espace d'états sous-jacent. Pour aborder les questions de complexité algorithmique, on doit donc se placer dans un contexte moins général que celui des travaux cités précédemment.

Objectifs

Notre objectif principal est d'étudier de façon précise le comportement d'algorithmes d'évolution simplifiés, afin de parvenir, sur des exemples spécifiques, à une compréhension détaillée des effets des étapes de mutation et de sélection, et du rôle joué par la taille de la population. La simplicité des modèles étudiés limite, certes, leur intérêt du point de vue de l'optimisation, mais ils présentent néanmoins une certaine pertinence, par exemple en biologie. D'autre part, leur traitement mathématique soulève de sérieuses difficultés, et nécessite le développement de techniques spécifiques. L'état d'esprit dans lequel les problèmes sont abordés est plus proche de l'étude des marches aléatoires (en interaction, en milieu aléatoire...) que des méthodes usuelles d'analyse d'algorithmes. En particulier, les résultats portant sur les problèmes de taille finie sont généralement obtenus comme corollaires de résultats asymptotiques plus généraux prouvés sur des espaces d'états infinis.

Les deux premiers modèles étudiés se rattachent à la catégorie des "fitness-space models", c'est-à-dire que l'effet de la mutation sur le fitness d'un individu donné ne dépend, en loi, que du fitness de cet individu. Le troisième modèle aborde le cas où le paysage de fitness comporte des irrégularités aléatoires.

Par ailleurs, nous présentons également des résultats expérimentaux sur l'utilisation de méthodes de *fitness sharing* dans le cadre du filtrage particulaire.

Plan

Les quatre parties de l'introduction qui suivent présentent de manière détaillée chacun des modèles étudiés, les motivations, les principaux résultats obtenus, ainsi que des esquisses de preuves. Elles se veulent une introduction à la lecture des chapitres suivants, plus techniques, qui reprennent pour l'essentiel quatre articles rédigés en anglais. En outre, diverses extensions et questions ouvertes y sont présentées.

Les quatre chapitres suivants contiennent l'exposé détaillé des résultats et des preuves, et peuvent être lus indépendamment les uns des autres, et indépendamment de l'intoduction :

- le chapitre 1 traite du modèle de mutation additive spatialement homogène, et correspond aux articles [7], [8] et [9],
- le chapitre 2 expose la généralisation des résultats du chapitre précédent au cas spatialement inhomogène, et correspond à l'article [6],
- le chapitre 3 présente les résultats expérimentaux obtenus sur les méthodes de *fitness sharing* en filtrage particulaire, et correspond à l'article [11],
- enfin, le chapitre 4 traite des modèles d'algorithmes génétiques en environnement aléatoire que nous avons étudiés, et correspond à l'article [5].

Mutation additive spatialement homogène

Résumé

Dans cette partie, nous nous intéressons aux algorithmes de mutation-sélection en population finie p définis sur l'ensemble $\mathbb{N}$ des entiers naturels de la façon suivante :

- $\mathbf{S}$ = sélection proportionnelle (fitness $F(x) = x$).
- $\mathbf{M}$ = mutation : $x \longmapsto x + \epsilon$, où ϵ est une variable de Bernoulli ± 1 symétrique.

Il s'agit donc d'un modèle dans lequel l'effet de la mutation sur le fitness est additif, et spatialement homogène, au sens où la mutation ajoute au fitness une quantité aléatoire dont la loi ne dépend pas du point où cette mutation est appliquée. Quoique très simple, ce modèle a néanmoins été utilisé en biologie pour étudier l'évolution de populations de virus. Des estimations récursives des moments associés à la population nous permettent de prouver un principe d'invariance pour la population normalisée, ainsi qu'un principe de grandes déviations. Les méthodes de preuve que nous employons se prêtent à une description heuristique simple en termes d'effets des opérateurs de mutation et de sélection sur les moments, ce qui permet de dégager les propriétés essentielles du modèle responsables du comportement observé. Le résultat principal est que, après n pas de temps, la position de la population est approximativement $\sqrt{n}$ fois la position au temps 1 d'un processus de Bessel de dimension $2p - 1$. Les résultats présentés dans cette partie ont été obtenus en collaboration avec Alexis Bienvenüe.

Motivations et résultats

Problématique et motivations

Par sa définition, ce modèle apparaît comme l'un des plus simples que l'on puisse imaginer. Pourtant, le choix de la fonction de fitness $F(x) = x$ n'est pas aussi restrictif qu'il y paraît. En effet, tout algorithme de mutation-sélection défini sur un espace d'états S avec fitness F tel que l'effet de la mutation sur le fitness d'un individu ne dépend, en loi, que de la valeur de ce fitness, peut se ramener à cette forme. Plus précisément, si K est le noyau markovien sur S qui définit la mutation, et si les transitions $F(x) \longmapsto F(M(x))$ sont données, en loi, par un noyau markovien K' sur $\mathbb{R}_+^*$, le processus $(Z_n) = (Z_n^{(i)})_{1\leq i\leq p}$ défini par $Z_n^{(i)} = F(X_n^{(i)})$ est lui-même un algorithme de mutation-sélection sur $\mathbb{R}_+^*$ avec mutation K' et fitness $G(x) = x$. Naturellement, les transitions $F(x) \longmapsto F(M(x))$ ne sont pas en général markoviennes, et considérer ce modèle revient donc à faire une hypothèse simplificatrice sur la façon dont les mutations affectent le fitness, en gommant la structure sous-jacente de l'espace d'états. Ce type de modèle est appelé modèle à espace de fitness, ou *fitness-space model.*

Il est clair que ce type de situation ne présente qu'un intérêt limité du point de vue de l'optimisation. Cependant, ce modèle a été utilisé en biologie pour étudier l'évolution de populations de virus, voir L. Tsimring, H. Levine, D. Kessler et D. Ridgway [72], [43] et D. Bonnaz [13].

Description du modèle

Les algorithmes auxquels nous nous intéressons sont définis comme des chaînes de Markov à temps discret (X_n) sur l'espace d'états $\mathbb{N}^p$, $p \geq 1$ désignant la taille de la population, de la façon suivante :

$$X_0 = (1, \ldots, 1).$$

1. **sélection :** $X_n \longrightarrow Y_n$ Si $X_n = (0, \ldots, 0)$, $Y_n = (1, \ldots, 1)$. Sinon, les $Y_n^{(i)}$, $1 \leq i \leq p$, sont choisis aléatoirement et indépendamment parmi les $\left\{X_n^{(i)},\ 1 \leq i \leq p\right\}$ selon la loi
$$\frac{1}{S_n}\sum_{i=1}^{p} X_n^{(i)}\delta_{X_n^{(i)}}, \quad \text{où} \quad S_n = \sum_{i=1}^{p} X_n^{(i)}.$$
2. **mutation :** $Y_n \longrightarrow X_{n+1}$
$$X_{n+1}^{(i)} = Y_n^{(i)} + \epsilon_i\ ,\ 1 \leq i \leq p,$$
les ϵ_i , $1 \leq i \leq p$ étant des variables aléatoires indépendantes et indépendantes du passé, de loi commune $\frac{1}{2}\delta_1 + \frac{1}{2}\delta_{-1}$.

F_n désigne le fitness moyen empirique :

$$F_n = \frac{1}{p} \sum_{i=1}^{p} X_n^{(i)},$$

et $d(X_n)$ le diamètre de la population, c'est-à-dire son extension spatiale :

$$d(X_n) = \max_{1 \leq i,j \leq p} |X_n^{(i)} - X_n^{(j)}|.$$

Remarque 1 *La réflexion en* 0 *dans la définition de la sélection est arbitraire, il s'agit simplement d'empêcher l'algorithme d'atteindre, par mutation, des valeurs négatives. Il serait également possible de modifier l'opérateur de mutation à l'aide d'une marche aléatoire réfléchie.*

Remarque 2 *L'algorithme étant initialisé en* 1*, et les pas de mutation pouvant faire croître les positions des particules d'au plus* 1 *à chaque pas, il est possible d'identifier jusqu'au temps* n*, par un couplage, les trajectoires de l'algorithme défini ci-dessus et celles de l'algorithme de mutation-sélection défini sur l'ensemble* $S(n) = \{0, 1, \ldots, n+1\}$ *et possédant les mêmes caractéristiques (fitness, mutation). En ce sens, les résultats asymptotiques sur la loi de* X_n *énoncés ci-dessous se traduisent directement en résultats asymptotiques sur le comportement au temps* n *de l'algorithme défini sur l'espace* $S(n)$*, lorsque* n *tend vers l'infini.*

Résultats

Nos deux résultats principaux sont les suivants :

Théorème 1 *Le processus* $\frac{1}{\sqrt{n}} \left(X_{\lfloor nt \rfloor} \right)_{0 \leq t \leq 1}$ *converge en loi dans l'espace de Skorokhod vers* $(1, \ldots, 1)R$*, où* R *est un processus de Bessel de dimension* $2p-1$ *issu de* 0*.*

Théorème 2 *Pour tout* i*, la suite* $\left(X_n^{(i)}/n \right)$ *satisfait un principe de grandes déviations sur* $[0,1]$*, relatif à la bonne fonction de taux* I*, où* I *est la fonction de taux associée à une marche aléatoire simple réfléchie sur* $\mathbb{N}$ *:*

$$\begin{cases} I(a) = \frac{1}{2} \left[(1-a)\log(1-a) + (1+a)\log(1+a) \right] & \text{si } a \in [0,1] \\ I(a) = +\infty & \text{sinon.} \end{cases}$$

La vitesse de croissance en $\sqrt{n}$ peut sembler surprenante, car on obtiendrait le même ordre de vitesse en l'absence de sélection, dans le cas d'une marche aléatoire simple réfléchie (cas $p = 1$). L'effet de la sélection se lit sur la loi limite, qui charge d'autant plus les grandes valeurs que p est grand. On notera que, sur l'échelle $\sqrt{n}$, le nuage de particules se comporte comme un seul point, ce qui signifie que le diamètre est petit devant le fitness moyen. D'autre part, le résultat de grandes déviations montre que, sur l'échelle n, on n'observe pas l'effet de la sélection.

Description informelle des preuves

La méthode que nous nous proposons d'employer, basée sur l'étude de l'effet d'un pas de mutation et de sélection sur les moments associés à la distribution empirique de la population, rejoint l'approche heuristique proposée par les physiciens théoriciens A. Prügel-Benett, M. Rattray et J. Shapiro [61], [63].

Pour simplifier, nous ne décrirons que la preuve de la convergence en loi de la population normalisée $X_n/\sqrt{n}$. Le plan est *grosso modo* le suivant :

- majoration du diamètre du nuage de particules
- preuve que le nuage de particules tend vers l'infini en probabilité
- étude des incréments du type $\mathrm{E}\left(F_{n+1}^{2q} - F_n^{2q} \,\middle|\, X_n\right)$ à l'aide des approximations permises par les deux étapes précédentes.

Estimation du diamètre

On vérifie facilement que la probabilité conditionnelle à X_n, pour que la sélection choisisse p fois le même individu est donnée, si $X_n \neq (0, \ldots, 0)$, par

$$\frac{1}{S_n^p} \sum_{i=1}^{p} \left(X_n^{(i)}\right)^p \geq p^{1-p}.$$

Par conséquent, la probabilité conditionnelle à X_n pour que la sélection contracte le nuage de particules sur un seul point est minorée par p^{1-p}. Ainsi, en remontant dans le passé du processus à partir de l'instant n, le premier instant avant n où le nuage de particules est réduit à un point, et n, ne sont pas très éloignés. Comme les étapes de mutation ne peuvent accroître le diamètre que d'au plus 2 à chaque pas, on obtient la majoration stochastique suivante :

$$d(X_n) \leq 2A_n,$$

où A_n est une variable aléatoire de loi géométrique de paramètre p^{1-p}. Cette majoration est une simple conséquence de l'amplitude bornée des mutations et du fait que l'on rééchantillonne la population en maintenant sa taille constante égale à p.

Fuite vers l'infini de X_n

Il semble intuitivement évident que la présence de la sélection contribue à faire croître le fitness vers de grandes valeurs, et que le nuage de particules doit croître au moins aussi rapidement qu'une marche aléatoire simple réfléchie. De fait, il est possible de minorer stochastiquement le nuage de particules par un processus de mutation-sélection avec une sélection uniforme et le même opérateur de mutation. Plus précisément, on peut construire un processus de mutation-sélection (V_n) sur $\mathbb{N}^p$, couplé à X, associé à une fonction de fitness constante, et dont les étapes de mutation sont données par une marche aléatoire simple réfléchie sur $\mathbb{N}$, tel que, pour tout $1 \leq i \leq p$, $V_n^{(i)} \leq X_n^{(i)}$. Un tel couplage est autorisé par la structure de la sélection par rééchantillonage avec des poids proportionnels au fitness, et le fait que, si $v \leq x$, on peut coupler les étapes de mutation de telle façon que $M(v) \leq M(x)$. L'intérêt de cette minoration est que, les probabilités de transition des étapes de sélection de V étant constantes, la loi de $V_n^{(i)}$ s'identifie à la loi au temps n d'une marche aléatoire simple réfléchie. En particulier, $V_n^{(i)}$, et donc $X_n^{(i)}$, tend vers l'infini en probabilité pour tout i.

Remarque 3 *En un sens, cette propriété n'est pas véritablement due à la présence de la sélection, puisque la mutation seule suffit à assurer la fuite de X_n vers l'infini en probabilité. Nous verrons, dans la partie suivante, des situations ou ce type de minoration ne peut plus être utilisé pour prouver la fuite vers l'infini.*

Étude des moments

Les deux résultats précédents : majoration du diamètre, fuite vers l'infini de (X_n), entraînent la convergence de $d(X_n)/F_n$ vers zéro dans tous les L^s ,$1 \leq s < +\infty$. Autrement dit, le nuage de particules peut toujours, au premier ordre, être traité comme un point, en ne commettant qu'une erreur contrôlée. Par exemple les probabilités de transition de l'étape de sélection peuvent être bien approchées par des probabilités de transition uniformes.

Étudions à présent l'effet d'un pas de mutation-sélection sur le fitness moyen empirique F_n. Soit $\Delta F_n = F_{n+1} - F_n$. Puisque la loi de la mutation est centrée, l'accroissement moyen de F_n n'est dû qu'à la sélection, et un calcul explicite montre que :

$$\mathrm{E}\left(\Delta F_n \mid X_n\right) = \mathbf{1}_{\{X_n \neq 0\}} \frac{1}{p^2 F_n} \sum_{1 \leq i < j \leq p} \left(X_n^{(j)} - X_n^{(i)}\right)^2 + \mathbf{1}_{\{X_n = 0\}}. \qquad (1)$$

Ce résultat peut sembler surprenant car, il montre, en vertu de la majoration du diamètre et de la fuite vers l'infini de F_n, que

$$\mathrm{E}\left(\Delta F_n \mid F_n\right)$$

tend vers zéro dans tous les L^s , $1 \leq s < +\infty$ lorsque n tend vers l'infini, alors que $|\Delta F_n| \propto d(X_n)$ est de l'ordre de l'unité. Un effet de compensation de signe se produit donc, lorsque l'on effectue une moyenne sur les pas de sélection.

Pour éliminer le F_n au dénominateur dans (1), de façon à se ramener à des quantités de l'ordre de l'unité, on est naturellement conduit à étudier les incréments moyens de F_n^2, puisque

$$\mathrm{E}\left(F_{n+1}^2 - F_n^2 \mid X_n\right) = 2F_n \mathrm{E}\left(\Delta F_n \mid X_n\right) + \mathrm{E}\left((\Delta F_n)^2 \mid X_n\right).$$

De la majoration $|\Delta F_n| \leq d(X_n)+2$, et de la majoration du diamètre, on déduit facilement que cette dernière expression est d'espérance majorée par une constante. On peut donc déja déduire que $\mathrm{E}(F_n^2) \leq Cn$. Cependant, nous souhaitons déterminer précisément le comportement asymptotique de cette expression. En introduisant le fitness après sélection $G_n = \frac{1}{p}\sum_{i=1}^p Y_n^{(i)}$, et le fait que les pas de mutation sont centrés et indépendants du passé, on obtient facilement que :

$$\mathrm{E}\left((\Delta F_n)^2 \;\middle|\; X_n\right) = p^{-1} + \mathrm{E}\left((G_n - F_n)^2 \;\middle|\; X_n\right)$$

L'expression explicite du dernier terme fait notamment intervenir des expressions du type

$$\frac{X_n^{(a)} X_n^{(b)}}{S_n^2}(X_n^{(a)} - X_n^{(i)})(X_n^{(b)} - X_n^{(j)}),$$

et il semble *a priori* difficile de parvenir à une forme exploitable, qui permettrait de déterminer le comportement des incréments de F_n^2. L'idée de la démonstration consiste alors à approcher, dans cette expression, les probabilités de transition

$$\frac{X_n^{(a)} X_n^{(b)}}{S_n^2}$$

de l'étape de sélection, par des probabilités de transition uniformes égales à $1/p^2$. Une simplification considérable en résulte, puisque l'on obtient que :

$$\mathrm{E}\left((G_n - F_n)^2 \;\middle|\; X_n\right) \approx p^{-3} \sum_{1\leq i<j\leq p} \left(X_n^{(j)} - X_n^{(i)}\right)^2.$$

En profitant de la symétrie de la loi des $(X_n^{(i)})_{1\leq i\leq p}$, et en prenant l'espérance, on obtient finalement que :

$$\mathrm{E}\big(F_{n+1}^2 - F_n^2\big) \approx \tfrac{1}{2}p(p-1)(2p^{-2}+p^{-3})\mathrm{E}\Big((X_n^{(1)} - X_n^{(2)})^2\Big) + p^{-1}. \tag{2}$$

L'incrément moyen de F_n^2 se présente donc comme la somme d'un terme dû à l'étape de sélection $X_n \to Y_n$, et d'un terme dû à l'étape de mutation $Y_n \to X_{n+1}$. Tout l'intérêt de cette dernière équation est qu'il est possible de déterminer précisément le comportement asymptotique de $\mathrm{E}\Big((X_n^{(1)} - X_n^{(2)})^2\Big)$. En effet, en approchant comme précédemment les probabilités de transition de l'étape de sélection par des probabilités de transition uniformes, on obtient que :

$$\mathrm{E}\Big((X_n^{(1)} - X_n^{(2)})^2\Big) \approx \frac{p-1}{p}\mathrm{E}\Big((X_{n-1}^{(1)} - X_{n-1}^{(2)})^2\Big) + 2.$$

Le facteur $\frac{p-1}{p} < 1$ est la probabilité, sous la sélection uniforme, que $Y_n^{(1)} = Y_n^{(2)}$, et le 2 est dû à la mutation. Cette équation traduit donc une propriété de *contraction approchée* de l'étape de sélection sur la variance du nuage de particules.

On déduit facilement de cette relation que

$$\lim_{n\to+\infty} \mathrm{E}\Big((X_n^{(1)} - X_n^{(2)})^2\Big) = 2p,$$

puis, grâce à (2), le comportement asymptotique exact :

$$\mathrm{E}F_n^2 \sim (2p-1)n.$$

Puisque le comportement asymptotique de $\mathrm{E}\Big((X_n^{(1)} - X_n^{(2)})^2\Big)$ est obtenu en remplaçant les étapes de sélection par un rééchantillonage uniforme des particules du nuage, on peut considérer que ce comportement est, en fait, uniquement dû à la mutation et au fait que la population est rééchantillonnée à chaque étape. Le coefficient de contraction par rééchantillonnage $\frac{p-1}{p}$ est d'autant plus proche de 1 que p est grand, et détermine la variance empirique asymptotique du nuage de particules en fonction de la variance des pas de mutation : plus la taille de population est grande, moins la contraction est forte et plus le nuage de particules peut exploiter la variance des pas de mutation pour accroître sa propre variance empirique. Lorsque p est grand, le terme dû à la mutation dans (2) devient négligeable, et l'incrément moyen de F_n^2 est essentiellement égal à $\mathrm{E}\Big((X_n^{(1)} - X_n^{(2)})^2\Big)$, ce qui traduit la façon dont la sélection exploite la variance empirique du nuage - essentiellement due à la mutation et au rééchantillonnage - pour accroître le fitness moyen.

La méthode peut en fait s'étendre, pour déterminer récursivement le comportement de tous les moments d'ordres pairs de F_n. En écrivant $F_{n+1}^{2q} = (F_n + \Delta F_n)^{2q}$ et en développant cette expression, on vérifie qu'il suffit de conserver les termes d'ordre 1 et 2 en ΔF_n. Autrement dit :

$$\begin{aligned}\mathrm{E}\left(F_{n+1}^{2q} - F_n^{2q} \;\middle|\; X_n\right) \approx{}& 2qF_n^{2q-1}\mathrm{E}\left(\Delta F_n \mid X_n\right)\\ &+q(2q-1)F_n^{2q-2}\mathrm{E}\left((\Delta F_n)^2 \mid X_n\right).\end{aligned}$$

En réutilisant les approximations précédentes, on parvient à exprimer

$$\mathrm{E}\Big(F_{n+1}^{2q} - F_n^{2q}\Big)$$

en termes d'expressions du type :

$$\mathrm{E}\big(F_n^{2q-2}\big) \text{ et } \mathrm{E}\Big(F_n^{2q-2}(X_n^{(1)} - X_n^{(2)})^2\Big).$$

La propriété de contraction approchée de la sélection s'applique à ces dernières expressions, sous la forme :

$$\mathrm{E}\Big(F_n^{2q-2}(X_n^{(1)} - X_n^{(2)})^2\Big) \approx \frac{p-1}{p}\mathrm{E}\Big(F_{n-1}^{2q-2}(X_{n-1}^{(1)} - X_{n-1}^{(2)})^2\Big) + 2\mathrm{E}\big(F_n^{2q-2}\big),$$

ce qui permet de clore la récurrence, en exprimant complètement le comportement asymptotique des incréments de F_n^{2q} en termes du comportement asymptotique de $\mathrm{E}F_n^{2q-2}$.

Une fois ces estimations de moments disponibles, il est aisé de prouver la convergence en loi de $F_n/\sqrt{n}$ par la méthode des moments et de déterminer la loi limite. De plus, comme $d(X_n)/F_n$ tend vers zéro en probabilité, on en déduit immédiatement la convergence en loi de $X_n/\sqrt{n}$.

Extensions et questions ouvertes

Convergence en loi du nuage non-normalisé

Au cours de la preuve, nous avons établi la convergence de $\mathrm{E}\Big((X_n^{(1)} - X_n^{(2)})^2\Big)$. Il est en fait possible de prouver la convergence en loi de $(X_n^{(1)} - X_n^{(2)})$, et plus généralement, de $\Theta(X_n)$, où, pour $x \in \mathbb{Z}^p$, $\Theta(x)$ désigne la classe d'équivalence de x dans $\mathbb{Z}^p / < (1, \ldots, 1) >$.

Taille de population

Comme nous l'avons vu, il est possible de déterminer précisément l'effet de la taille de population sur le comportement asymptotique de l'algorithme. Le modèle correspondant en population infinie a par ailleurs été étudié par C. Mazza et D. Piau [50], et fait apparaître une vitesse de croissance linéaire du fitness moyen avec le nombre d'itérations, à comparer à la vitesse en $\sqrt{n}$ de l'algorithme en population finie. Il serait intéressant de disposer d'informations sur des régimes intermédiaires dans lesquels la taille de population serait autorisée à croître avec le nombre d'itérations, par exemple $p_n = n$, afin de mieux comprendre cette transition population finie / population infinie. Ces régimes intermédiaires ne sont pour l'instant pas compris, et les simulations numériques que nous avons menées ne permettent pas de conjecturer un comportement précis, même s'il apparaît vraisemblable que le fitness moyen croît plus rapidement que $\sqrt{n}$.

Récurrence / Transience

La minoration stochastique de l'algorithme par un processus de mutation-sélection uniforme fournit la preuve de la fuite vers l'infini en probabilité de l'algorithme, c'est-à-dire, dans le langage des chaînes de Markov, que (X_n) est, soit transient, soit récurrent nul. Ce résultat suffit pour valider nos approximations, mais la détermination exacte du comportement de (X_n) est une question ouverte.

Couplages

Il serait intéressant de disposer, non pas seulement d'une comparaison stochastique de l'algorithme avec un processus de mutation-sélection dont la sélection est absente, mais de pouvoir comparer entre eux des processus de mutation-sélection plus généraux. Par exemple, est-il possible de coupler entre elles deux versions de (X_n) issues de $(1,\ldots,1)$ et de $(2,\ldots,2)$ de telle façon qu'une relation d'ordre entre les deux nuages de particules soit préservée ?

Par ailleurs, l'heuristique suivante a été suggérée par D. Piau [60] pour tenter d'expliquer la forme de la loi limite : on considère p particules situées au temps n sur un même sommet de $\mathbb{N}$, $\hat{x}_n$. Chacune mute en $\hat{x}_n \pm 1$ de façon i.i.d. et symétrique, donnant p positions z_k^n qui valent donc chacune $\hat{x}_n + 1$ ou $\hat{x}_n - 1$. Ensuite, on choisit $\hat{x}_{n+1}$, selon la loi conditionnelle

$$t_n^{-1} \sum_{k=1}^{p} \delta(z_k^n) \text{ avec } t_n = \sum_{k=1}^{p} z_k^n.$$

On voit que $\hat{x} = (\hat{x}_n)_n$ est une marche au plus proche voisin sur $\mathbb{N}$ dont les transitions vérifient :

$$p(x \to x+1) = \frac{1}{2}\left(1 + \frac{p-1}{x}(1+o(1))\right).$$

Par conséquent, $\hat{x}$ se comporte comme la version discrétisée d'un processus de Bessel de dimension d si d est solution $\frac{1}{2}(d-1) = p-1$, i.e. si $d = 2p-1$. Malheureusement, il ne nous a pas été possible de construire un couplage entre les processus (X_n) et $(\hat{x}_n)$ qui permettrait de déduire le résultat de convergence directement.

Stabilité

Les performances des algorithmes génétiques sont généralement considérées comme étant, en pratique, relativement indifférentes à la présence de bruit, voir E. Baum, D. Boneh et C. Garrett [4]. Dans notre contexte, il serait intéressant d'étudier le cas où F est bruitée, c'est-à-dire transformée, par exemple, en $F(x) = x + \epsilon_x$ où ϵ_x est une petite perturbation aléatoire.

Mutation additive spatialement inhomogène

Résumé

Dans cette partie, nous nous intéressons aux algorithmes génétiques de mutation-sélection en population finie sur l'intervalle $[1,+\infty[$, dont les étapes de mutation et de sélection sont définies de la façon suivante :

- **S** = sélection proportionnelle (fitness $F(x) = x$).

- $\mathbf{M}$ = mutation : $x \longmapsto \max(x + Lx^\alpha, 1)$, où $0 \leq \alpha < 1$ et où L est une variable aléatoire de loi μ donnée.

Il s'agit d'une généralisation du modèle précédent, qui correspond au cas où $\alpha = 0$ et $\mu = \frac{1}{2}\delta_1 + \frac{1}{2}\delta_{-1}$. Comme dans la partie précédente, la mutation agit additivement sur le fitness au sens où elle s'écrit, en omettant la barrière en 1, $X \longmapsto X + Lg(X)$, L étant indépendante de X. La différence essentielle réside dans le fait que l'amplitude de la loi de mutation dépend à présent du point où celle-ci est appliquée : il s'agit d'une mutation additive, mais spatialement inhomogène.

Afin d'étudier cette classe d'algorithmes, nous étendons les méthodes employées dans la partie précédente en systématisant l'utilisation des moments. Nous obtenons la généralisation du résultat principal de la partie précédente, à savoir la convergence en loi du nuage de particules normalisé, ainsi que des estimations sur le comportement asymptotique de la mutation itérée qui permettent la comparaison avec l'algorithme génétique. Plusieurs difficultés techniques absentes de la partie précédente doivent être surmontées, notamment le fait que la minoration de l'algorithme par un processus de mutation-sélection avec sélection uniforme ne fournit plus systématiquement une preuve de non-ergodicité.

Motivations et résultats

Problématique et motivations

Comme expliqué dans la partie précédente, l'étude des algorithmes génétiques définis sur $\mathbb{R}_+$ avec fitness $F(x) = x$ correspond en fait à l'étude des algorithmes génétiques définis sur un espace d'états arbitraire mais dont l'effet de la mutation sur le fitness ne dépend, en loi, que du fitness. Dans le modèle traité précédemment, la mutation agissait sur le fitness de façon additive et spatialement homogène, en ajoutant au fitness initial une variable aléatoire de Bernoulli ± 1. Dans cette partie, nous tentons de généraliser les résultats précédents en nous affranchissant de cette hypothèse d'homogénéité spatiale, tout en conservant deux propriétés essentielles du modèle précédent : le caractère additif de la mutation, ainsi que son caractère perturbatif. Autrement dit, nous nous intéressons au cas où l'effet de la mutation sur le fitness peut se mettre sous la forme $X \longmapsto X + Lg(X)$, où L est une variable aléatoire indépendante de X de loi μ fixée, et où $g(x)$ est petit devant x lorsque x tend vers l'infini, et dépend de x de manière régulière. Dès lors, le choix $g(x) = x^\alpha$ avec $0 \leq \alpha < 1$ apparaît naturellement comme un premier cas à envisager. Ce choix est également suggéré par les travaux de Y. Rabinovich et A. Wigderson, qui étudient dans [62] le comportement d'un algorithme génétique en population infinie défini sur $\mathbb{N}$ par sélection proportionnelle avec fitness $F(x) = x$, et mariage binomial, dont la version en population finie se ramène essentiellement au cas où $\mu = \mathcal{N}(0, 1)$ et $g(x) = \sqrt{x}$.

Une motivation supplémentaire à l'étude de cette classe d'algorithmes est fournie par la remarque suivante : dans le cas envisagé précédemment, la vitesse de croissance du fitness de l'algorithme était de l'ordre de $\sqrt{n}$, c'est-à-dire du même ordre que la vitesse de croissance d'une marche au hasard simple réfléchie. Autrement dit, la sélection n'agissait pas sur l'ordre de la vitesse, mais simplement sur la loi limite (loi normale réfléchie dans le cas d'une marche aléatoire, loi de Bessel pour l'algorithme génétique). La question se pose donc naturellement de savoir s'il est possible, en autorisant l'amplitude des mutations à croître avec le fitness, d'obtenir pour l'algorithme génétique un ordre de vitesse supérieur à celui obtenu en itérant la mutation seule.

Description du modèle

Les algorithmes auxquels nous nous intéressons sont définis comme des chaînes de Markov à temps discret (X_n) sur l'espace d'états $[1,+\infty[^p$, $p \geq 1$ désignant la taille de la population, de la façon suivante :

$$X_0 = (1, \ldots, 1).$$

1. **sélection :** $X_n \longrightarrow X'_n$
   Les $X'^{(i)}_n$, $1 \leq i \leq p$, sont choisis aléatoirement et indépendamment parmi les $\left\{X^{(i)}_n,\ 1 \leq i \leq p\right\}$ selon la loi
   $$\frac{1}{S_n}\sum_{i=1}^{p} X^{(i)}_n \delta_{X^{(i)}_n}, \quad \text{où} \quad S_n = \sum_{i=1}^{p} X^{(i)}_n.$$
2. **mutation :** $X'_n \longrightarrow X_{n+1}$
   $$X^{(i)}_{n+1} = \max\left(X'^{(i)}_n + L^{(i)}_n \left(X'^{(i)}_n\right)^{\alpha}\ ,\ 1\right),$$
   les $L^{(i)}_n$, $1 \leq i \leq p$ étant des variables aléatoires indépendantes et indépendantes du passé, de loi commune μ.

Remarque 4 *Les étapes de mutation ne sont pas définies exactement par $x \longmapsto x + Lx^{\alpha}$, car une barrière est placée en 1 de façon à éviter que X_n ne puisse prendre des valeurs négatives. Lorsque x est grand, cette troncature devient négligeable. En outre, cette formulation nous permet d'éviter le problème des petites valeurs de x (proches de zéro) qui ne sont pas pertinentes dans le problème.*

Remarque 5 *Les chaînes de Markov considérées ici sont définies sur des espaces d'états non-dénombrables. Cependant, avec les hypothèses sous lesquelles nous nous plaçons, l'essentiel de la théorie des chaînes de Markov à espace d'états dénombrable* *s'applique, notre référence concernant ces questions étant l'ouvrage de S. Meyn et R. Tweedie [51].*

Résultats

Dans tout ce qui suit, nous nous placerons dans le cas où la loi μ est centrée, c'est-à-dire le cas où, en moyenne, la mutation ne modifie pas le fitness. Les cas où μ n'est pas centrée sont brièvement abordés dans la section .

Nous supposerons que μ possède des moments de tous ordres, et, en outre, que μ possède une densité strictement positive par rapport à la mesure de Lebesgue. Cette dernière hypothèse n'est pas indispensable, mais simplifie significativement l'exposition des preuves.

Notre premier résultat porte sur le comportement de la chaîne de Markov Z_n obtenue en itérant les étapes de mutation seule :

$$Z_0 = 1 \ , \ Z_{n+1} = \max(Z_n + LZ_n^\alpha, 1).$$

Notons que $Z_n = X_n^{(1)}$ dans le cas où $p = 1$ (absence de sélection). Posons

$$\beta = \frac{1}{2(1-\alpha)} \ , \ r(p,\alpha) = (2p-3)(1-\alpha)^{-1} + 2.$$

Théorème 3 *Le comportement de Z_n dépend de α de la manière suivante :*
- *si $0 < \alpha < 1/2$, Z_n est récurrente nulle ou transiente et croît à vitesse n^β, au sens où $n^{-\beta}Z_n$ converge en loi vers*

$$\left((1-\alpha)\mathrm{E}(L^2)^{1/2}B(\alpha)\right)^{2\beta}$$

  où $B(\alpha)$ désigne la valeur au temps 1 d'un processus de Bessel de dimension $r(1,\alpha)$.
- *si $\alpha = 1/2$, Z_n est récurrente nulle ou transiente et croît à vitesse $o(n^\beta) = o(n)$, au sens où, $n^{-1}Z_n$ tend vers zéro dans tous les L^s, $1 \leq s < +\infty$,*
- *si $1/2 < \alpha < 1$, Z_n est récurrente positive.*

Le comportement asymptotique de Z_n présente donc une transition de phase en $\alpha = 1/2$. Pour $0 < \alpha < 1/2$, le comportement de Z_n prolonge celui d'une marche aléatoire simple réfléchie ($\alpha = 0$) et possède une vitesse au sens usuel. Pour $\alpha > 1/2$, l'amplitude des mutations est trop importante pour pemettre à Z_n de fuir vers l'infini en probabilité, et sa vitesse est nulle (dans tous les sens raisonnables). Dans le cas limite $\alpha = 1/2$, on obtient seulement une majoration de la vitesse de croissance de Z_n.

Le comportement de l'algorithme génétique est, quant à lui, donné par le théorème suivant, valable si $p \geq 2$:

Théorème 4 *Lorsque n tend vers l'infini, $n^{-\beta}X_n$ converge en loi vers*

$$\left((1-\alpha)\mathrm{E}(L^2)^{1/2}B(p,\alpha)\right)^{2\beta}(1,\ldots,1)$$

où $B(p,\alpha)$ désigne la valeur au temps 1 d'un processus de Bessel de dimension $r(p,\alpha)$.

Ainsi, la vitesse de croissance du fitness de l'algorithme est n^β pour toute valeur de α, et le comportement de l'algorithme ne présente donc pas de transition de phase. Dès que $\alpha \geq 1/2$, l'ordre de la vitesse de croissance de l'algorithme dépasse strictement celui de la mutation itérée. Contrairement au cas de la mutation itérée, l'ordre de la vitesse de l'algorithme croît continûment avec l'amplitude des mutations, qui est exploitée par la sélection dès que $p \geq 2$. On note que, comme dans la partie précédente, le diamètre de la population est petit devant le fitness moyen, la population se comportant au premier ordre comme un point sur l'échelle n^β.

Description informelle des preuves

La méthode que nous employons généralise celle de la partie précédente, en systématisant l'étude des moments. Le plan global de la preuve est identique :

- majoration *a priori* portant sur le diamètre du nuage de particules,
- preuve que le nuage de particules tend vers l'infini en probabilité,
- estimation récursive des moments à l'aide des approximations permises par les deux étapes précédentes.

Comme dans la partie précédente, la propriété de contraction de la sélection joue un rôle fondamental. L'inhomogénéité spatiale des mutations est traitée en ne considérant que des quantités dans lesquelles les mutations apparaissent normalisées par leur amplitude. D'autre part, la preuve de la fuite vers l'infini de (X_n), à la différence de la partie précédente, repose sur des estimations de moments. On note que les hypothèses d'intégrabilité sur μ entraînent l'existence de tous les moments considérés.

Comme précédemment, F_n désigne le fitness moyen empirique :

$$F_n = \frac{1}{p}\sum_{i=1}^{p} X_n^{(i)},$$

et $d(X_n)$ le diamètre de la population, c'est-à-dire son extension spatiale :

$$d(X_n) = \max_{1\leq i,j\leq p} |X_n^{(i)} - X_n^{(j)}|.$$

Définissons

$$L_n^* = \max_{1\leq i\leq p} |L_n^{(i)}|.$$

Estimations du diamètre

Du fait que la taille de la population est maintenue constante égale à p, la probabilité conditionnelle à X_n pour que l'étape de sélection choisisse p fois la même particule est minorée par p^{1-p}, ce qui entraîne une propriété de contraction de la sélection sur le diamètre de la population. Par conséquent, les effets de la mutation sur le diamètre ne s'accumulent pas et $d(X_n)$ est du même ordre que les pas de mutations, c'est-à-dire que $d(X_n) \propto F_n^\alpha$. Plus précisément, si l'on pose $s(x) = d(x)(\min x)^{-\alpha}$, la propriété de contraction s'écrit :

$$\mathrm{E}\left(s(X'_n) \mid \mathcal{F}_n\right) \leq \kappa s(X_n), \quad \text{où} \quad \kappa = 1 - p^{1-p} < 1.$$

D'autre part, à des termes d'ordre moins élevé en $d(X'_n)(\min_i X'^{(i)}_n)^{-\alpha}$ près,

$$s(X_{n+1}) \lessapprox d(X'_n)(\min_i X'^{(i)}_n)^{-\alpha} + 2L^*_n$$

On obtient donc :

$$\mathrm{E}s(X_{n+1}) \lessapprox \kappa \mathrm{E}s(X_n) + 2\mathrm{E}L^*_n.$$

La suite $v_n = \mathrm{E}(s(X_n))$ vérifie donc une relation de récurrence de la forme :

$$v_{n+1} \leq \gamma v_n + f(v_n)$$

où $0 < \gamma < 1$ et $f(x) = o(x)$, ce qui entraîne que (v_n) est bornée.

On montre plus généralement que, pour tout $q > 0$,

$$\mathrm{E}\left(d(X_n)^q(\min_i X^{(i)}_n)^{-\alpha q}\right) \text{ est borné par rapport à } n$$

Fuite vers l'infini de X_n

Comme dans la partie précédente, il est possible de construire un couplage entre (X_n) et un processus de mutation-sélection avec sélection uniforme qui le minore. Malheureusement, dès que $\alpha > 1/2$, ce processus est ergodique et la minoration obtenue ne nous est d'aucune utilité. Nous montrons donc directement que (X_n) tend vers l'infini en probabilité, grâce à l'action de la sélection, en étudiant les incréments moyens de F_n :

$$\mathrm{E}\left(F^\gamma_{n+1} - F^\gamma_n \mid X_n\right).$$

Les critères usuels d'incréments [51] ne semblant pas s'appliquer ici, nous utilisons une démonstration par l'absurde et un argument de type second moment. Nous établissons d'abord une majoration *a priori* de F_n en montrant que

$$\mathrm{E}(F_n^2) \leq C_1 n^{2\beta},$$

puis, sous l'hypothèse que (X_n) est ergodique, une minoration :

$$\mathrm{E}(F_n^{2(1-\alpha)}) \geq C_2 n.$$

Ces deux inégalités sont incompatibles avec l'ergodicité supposée de (X_n) car elles entraînent, grâce à une inégalité de second moment, que F_n prend des valeurs de l'ordre de n^β avec une probabilité minorée par une constante.

Le point crucial pour obtenir la majoration est que, μ étant centrée, la mutation n'a pas, en moyenne, d'effet sur le fitness. La variation moyenne de F_n est donc entièrement due à la sélection, pour laquelle l'effet de compensation de signe déjà observé dans la partie précédente se produit :

$$\mathrm{E}\left(\Delta F_n/F_n \mid X_n\right) \propto (d(X_n)/F_n)^2,$$

alors que $|\Delta F_n/F_n| \propto d(X_n)/F_n$.

Grâce à l'estimation du diamètre citée plus haut, on obtient que :

$$\mathrm{E}\left(\Delta F_n/F_n \mid X_n\right) \propto \mathrm{E}\left((\Delta F_n/F_n)^2 \;\middle|\; X_n\right) \propto F_n^{2(\alpha-1)} \tag{3}$$

On en déduit facilement que

$$\mathrm{E}\big(F_{n+1}^2 - F_n^2\big) \leq C\mathrm{E}F_n^{2\alpha},$$

ce qui entraîne la majoration annoncée.

Supposons à présent F_n grand, et cherchons γ tel que

$$\mathrm{E}\left(F_{n+1}^\gamma - F_n^\gamma \mid X_n\right)$$

soit de l'ordre de l'unité. L'estimation (3) montre que l'on doit choisir $\gamma = 2(1-\alpha)$ et conserver, au premier ordre, les deux premiers termes dans le développement de Taylor de $(1+\Delta F_n/F_n)^{2(1-\alpha)}$, d'où, à des termes d'ordre F_n^{-v}, $v>0$ près :

$$\begin{aligned}\mathrm{E}\left(F_{n+1}^\gamma - F_n^\gamma \mid X_n\right) &\approx 2(1-\alpha)\mathrm{E}\left(F_n^\beta(\Delta F_n/F_n) \;\middle|\; X_n\right)\\ &\quad +(1-\alpha)(1-2\alpha)\mathrm{E}\left(F_n^\beta(\Delta F_n/F_n)^2 \;\middle|\; X_n\right)\end{aligned}$$

En approchant au premier ordre le nuage de particules par un point, on obtient, à des termes d'ordre F_n^{-v} près, $(v>0)$, que

$$\begin{gathered}\mathrm{E}\left(F_{n+1}^{2(1-\alpha)} - F_n^{2(1-\alpha)} \;\middle|\; \mathcal{F}_n\right) \approx \\ q_1 \sum_{1\leq i<j\leq p} F_n^{-2\alpha}\left(X_n^{(i)} - X_n^{(j)}\right)^2 + q_2\left(\mathrm{E}L_1^2\right)\end{gathered} \tag{4}$$

L'incrément moyen est donc la somme d'un terme dû à la sélection et d'un terme dû à la mutation.

La difficulté est que, dès que $\alpha > 1/2$, q_2 est strictement négatif et la mutation a donc pour effet de diminuer l'incrément moyen. En particulier, si $p = 1$, le terme de sélection est nul, et ceci entraîne l'ergodicité de (X_n). Lorsque $p \geq 2$, l'idée de la preuve consiste à montrer que, en effectuant une moyenne sur plusieurs pas de la chaîne, la mutation donne suffisamment d'amplitude au nuage de particules pour que les termes de sélection $F_n^{-2\alpha}\left(X_n^{(i)} - X_n^{(j)}\right)^2$ compensent l'effet négatif de la mutation sur l'incrément moyen. À cette fin, on vérifie, toujours en supposant F_n grand et en approchant le nuage de particules par un point, que, à des termes d'ordre F_n^{-v} près,

$$\mathrm{E}\left(F_{n+1}^{-2\alpha}\left(X_{n+1}^{(1)} - X_{n+1}^{(2)}\right)^2 \;\middle|\; \mathcal{F}_n\right) \approx 2\left(\mathrm{E}L_1^2\right) + p^{-2} \sum_{1\leq i\neq j\leq p} F_n^{-2\alpha}\left(X_n^{(i)} - X_n^{(j)}\right)^2$$

En utilisant la symétrie de la loi des $(X_n^{(i)})_{1\leq i\leq p}$, il devient clair que le comportement moyen de $F_n^{-2\alpha}\left(X_n^{(i)} - X_n^{(j)}\right)^2$ se compare avec celui d'une suite définie récursivement par :

$$u_{n+1} = 2\mathrm{E}L_1^2 + \frac{p-1}{p}u_n,$$

qui tend vers $2p\mathrm{E}L_1^2$ lorsque n tend vers l'infini.

En substituant cette valeur dans (4), on obtient une valeur strictement positive pour l'incrément moyen. La difficulté pour traduire cet argument heuristique en preuve mathématique est que l'on doit supposer F_n grand pour que les approximations soient valables, et que celles-ci ne sont valables qu'en moyenne. Ceci nous conduit à raisonner indirectement, en supposant que (X_n) est ergodique de façon à pouvoir exploiter l'existence d'une mesure invariante.

Convergence en loi

Une fois que l'on a prouvé la fuite vers l'infini en probabilité de (X_n), les approximations effectuées dans la section ci-dessus sont automatiquement validées, et l'on peut étendre la méthode pour établir récursivement le comportement asymptotique de tous les incréments du type :

$$\mathrm{E}\Big(F_{n+1}^{2q(1-\alpha)} - F_n^{2q(1-\alpha)}\Big).$$

La preuve suit essentiellement la plan de la partie précédente, à ceci près que les quantités $\left(X_n^{(i)} - X_n^{(j)}\right)^2$ apparaissent normalisées par l'amplitude en F_n^{α} des mutations. Comme dans la partie précédente, on vérifie, en approchant le nuage de particules par un point, que ces incréments s'expriment, à des termes d'erreurs contrôlés près, en fonction d'espérances du type

$$\mathrm{E}\Big(F_n^{2(q-1)(1-\alpha)}\Big) \quad \text{et} \quad c_n = \mathrm{E}\bigg(F_n^{2(q-1)(1-\alpha)} F_n^{-2\alpha} \left(X_n^{(i)} - X_n^{(j)}\right)^2\bigg).$$

Ces derniers termes vérifient une propriété de contraction approchée du type précédent :

$$c_{n+1} \approx \frac{p-1}{p} c_n + 2\mathrm{E}L_1^2 \mathrm{E}F_n^{2(q-1)(1-\alpha)},$$

ce qui permet de clore la récurrence en exprimant complètement le comportement asymptotique de $\mathrm{E}F_n^{2q(1-\alpha)}$ en termes de celui de $\mathrm{E}F_n^{2(q-1)(1-\alpha)}$.

Extensions et questions ouvertes

Convergence en loi du processus

Par analogie avec le cas $\alpha = 0$, il paraît vraisemblable que le processus $(n^{-\beta} X_{\lfloor nt \rfloor})_{0 \leq t \leq 1}$ converge en loi dans l'espace de Skorokhod vers un processus de Bessel de paramètre $(2p-3)(1-\alpha)^{-1} + 2$. Ce résultat n'est pas encore démontré.

Cas où μ est non-centrée

On note que les estimations portant sur le diamètre ne font pas intervenir le fait que μ soit centrée, mais simplement l'existence de moments de tous ordres. En ce qui concerne les estimations d'incréments, si $\mathrm{E}(\mu) > 0$, on vérifie facilement que le terme dominant dans $\mathrm{E}\left(\Delta F_n / F_n \mid X_n\right)$ est dû à la mutation, et que

$$\mathrm{E}\left(\Delta F_n / F_n \mid X_n\right) \propto F_n^{(\alpha-1)}$$

On doit donc choisir $\gamma = (1-\alpha)$ pour obtenir des incréments de l'ordre de l'unité, et l'effet de la sélection sur ces incréments est négligeable. On montre ainsi, comme dans le cas où μ est centrée, que X_n tend vers l'infini en probabilité, et, plus précisément, on obtient le résultat suivant :

Théorème 5 *Si* $\mathrm{E}(\mu) > 0$, $n^{-2\beta} X_n$ *converge dans tous les* L^s *vers*

$$(1-\alpha)\mathrm{E}(\mu)(1, \ldots, 1).$$

La sélection n'a donc pas d'effet sur le comportement de l'algorithme.

Dans le cas où $\mathrm{E}(\mu) < 0$ un raisonnement analogue à celui qui précède, en supposant que X_n tende vers l'infini en probabilité, entraînerait que $n^{-2\beta}\mathrm{E}F_n$ converge vers $(1-\alpha)\mathrm{E}(\mu) < 0$, ce qui est manifestement impossible puisque $F_n \geq 1$. D'où :

Théorème 6 *Si* $\mathrm{E}(\mu) < 0$, X_n *est ergodique*

Une question ouverte est celle du comportement de l'algorithme si l'on réduit les hypothèses d'intégrabilité sur μ, par exemple, si μ est seulement supposée intégrable (voir par exemple F. Spitzer [69] pour le comportement des marches aléatoires standards dans cette situation).

Cas où $\alpha = 1$

Dans le cas limite $\alpha = 1$, les mutations ne sont plus des perturbations du fitness, et se mettent sous la forme :

$$X \longmapsto (1+L)X$$

(on suppose pour simplifier que $L_1 > -1$ presque sûrement pour éviter que X_n puisse prendre des valeurs négatives). L'effet des mutations est donc multiplicatif sur le fitness, et on vérifie facilement que

$$\Pi_n = \left((X_n^{(i)})(X_n^{(j)})^{-1} \right)_{1 \leq i < j \leq p}$$

est une chaîne de Markov, car les probabilités de transition de l'étape de sélection ne dépendent que de ces quantités, et les mutations multiplient simplement ces quantités par $(1+L_n^{(i)})(1+L_n^{(j)})^{-1}$. On vérifie que (Π_n) est positivement récurrente et apériodique grâce à la propriété de contraction de la sélection, et possède donc une mesure invariante π. De plus, $\log(X_n)$ s'identifie à un algorithme de mutation-sélection sur $\mathbb{R}$ avec fitness $F(x) = \exp x$, et mutation additive dont les pas suivent la loi de $\log(1+L)$. On obtient ainsi que

$$\frac{1}{p}\sum_{i=1}^{p} \log(X_{n+1}^{(i)}) - \frac{1}{p}\sum_{i=1}^{p} \log(X_n^{(i)}) = f(\Pi_n, U) + \frac{1}{p}\sum_{i=1}^{p} \log(1+L_n^{(i)}),$$

où U est une variable aléatoire indépendante de Π_n et de $L_n^{(i)}$ qui fournit l'alea de l'étape de sélection, et f une fonction mesurable qui vérifie

$$\mathrm{E}\left(f(\Pi_n, U) \mid \Pi_n = s \right) \geq 0,$$

l'inégalité étant stricte sauf si $\Pi_n = (1, \ldots, 1)$ (la sélection accroît le fitness en moyenne, strictement si le nuage n'est pas réduit à un point). Ainsi, lorsque

$$\mathrm{E}\log(1+L_1) \geq 0$$

(en particulier lorsque μ est centrée), le théorème ergodique entraîne que

$$n^{-1}\left(\frac{1}{p}\sum_{i=1}^{p} \log X_n^{(i)} \right)$$

converge presque sûrement vers

$$\int \mathrm{E}\left(\log f(\Pi, U) \mid \Pi = s\right) \mathrm{d}\pi(s) + \mathrm{E}(\log(1 + L_1)) > 0,$$

et X_n croît donc à vitesse exponentielle. La mesure π n'étant pas connue de manière explicite, il est difficile de déterminer le signe de la vitesse exponentielle lorsque $\mathrm{E}(\log(1 + L_1)) < 0$.

À titre d'exemple, lorsque $p = 2$ et $\mu = a\delta_{1/e-1} + (1 - a)\delta_{e-1}$, on peut vérifier que la loi de Π_n est invariante par le changement $a \longmapsto 1 - a$. Posons

$$v_a = \int \mathrm{E}\left(\log f(\Pi, U) \mid \Pi = s\right) \mathrm{d}\pi_a(s).$$

Comme la vitesse de croissance exponentielle est nécessairement inférieure à 1 car $X_n^{(i)} \leq e^n$ pour tout n, on a nécessairement $v_a + 1 - 2a \leq 1$. En utilisant le fait que $v_a = v_{1-a}$, on en déduit que la vitesse exponentielle, égale à $v_a + 1 - 2a$, est majorée par $2(1 - a) + 1 - 2a = 3 - 4a$. Ainsi, dès que $a > 3/4$, la vitesse exponentielle est négative. La question de savoir si, pour toute valeur de a, il existe p tel que la vitesse exponentielle soit positive, est ouverte. Nous verrons dans la partie suivante que le cas $\alpha = 1$ peut-être traité plus facilement lorsque p est autorisé à croître avec n.

Convergence du nuage normalisé par son amplitude

Au cours de la preuve du résultat principal, on montre la convergence des moments du type :

$$F_n^{-2\alpha} \left(X_n^{(i)} - X_n^{(j)} \right)^2$$

vers $2p\mathrm{E}L_1^2$. La méthode de preuve, qui repose sur la propriété de contraction approchée de la sélection, s'étend facilement pour montrer récursivement la convergence de tous les moments du type :

$$\prod_{1 \leq i < j \leq p} F_n^{-2\alpha q_{i,j}} \left(X_n^{(i)} - X_n^{(j)} \right)^{2q_{i,j}}$$

pour toute famille d'entiers $q_{i,j}$.

On obtient ainsi la convergence en loi de $(F_n^{-\alpha})\,\Theta(X_n)$ vers une loi limite, $\Theta(x)$ désignant la classe d'équivalence de x dans $\mathbb{R}^p/(1, \ldots, 1)$.

Grandes déviations

Dans la partie précédente, la vitesse de croissance était d'ordre $n^{1/2}$, et un principe de grandes déviations contrôlait les déviations d'ordre n. Le comportement des algorithmes étudiés dans cette partie sur d'autres échelles de normalisation que leur vitesse en n^{β} n'est pas encore déterminé.

Population infinie

L'étude par Rabinovich et Wigderson de l'algorithme génétique en population infinie comportant sélection proportionnelle avec fitness $F(x) = x$ et mariage binomial, présentée dans [62], repose sur l'utilisation de propriétés algébriques particulières du mariage binomial, et le comportement asymptotique des modèles en population infinie correspondant précisément aux algorithmes étudiés dans cette partie reste à déterminer. Cette question semble d'autant plus intéressante que le résultat de Rabinovich et Wigderson fait apparaître une vitesse linéaire de croissance du fitness moyen avec n, qui serait donc comparable à celle obtenue en population finie lorsque $\alpha = 1/2$, et même inférieure pour une taille de population suffisamment grande.

Contrôle des termes d'erreur

Toutes les estimations asymptotiques que nous avons employées reposent sur le fait que les termes d'erreur du type $\mathrm{E}F_n^{-v}$ où $v > 0$ tendent vers zéro lorsque n tend vers l'infini, car F_n tend vers l'infini en probabilité. Or la preuve du fait que F_n tend vers l'infini en probabilité ne fournit aucune estimation de la vitesse à laquelle ces termes d'erreur tendent vers zéro, contrairement à la partie précédente, où la minoration en loi du nuage de particules par une marche aléatoire simple réfléchie fournit une majoration des erreurs commises. La vitesse de convergence du nuage de particules normalisé vers sa loi limite reste donc à déterminer. Pour les mêmes raisons, le comportement asymptotique exact de Z_n lorsque $\alpha = 1/2$ est encore inconnu.

Lois limites

Plusieurs chaînes de Markov ergodiques ont été rencontrées : Z_n lorsque $\alpha > 1/2$, X_n lorsque $\mathrm{E}\mu < 0$. Ici encore, les méthodes employées pour prouver l'ergodicité ne fournissent que peu d'information sur les lois invariantes, excepté la non-intégrabilité dans le cas de Z_n lorsque $\alpha > 1/2$.

Fitness sharing pour le filtrage particulaire

Ce chapitre étant assez différent des autres par sa nature, puisqu'il présente exclusivement des résultats expérimentaux et non des preuves mathématiques, nous renvoyons au chapitre lui-même pour une présentation détaillée, et donnons juste un bref résumé.

Résumé

Les problèmes de filtrage non-linéaire multimodaux sont couramment traités à l'aide d'algorithmes de Monte-Carlo, qui utilisent des procédures d'échantillonnage similaires à la sélection proportionnelle dans les algorithmes génétiques, et se heurtent au problème de la convergence prématurée du fait de la taille finie de la population. Nous étudions, dans ce contexte, l'utilisation de méthodes de *fitness sharing* initialement introduites dans le cadre des algorithmes génétiques. Sur des exemples simplifiés, ces méthodes sont étudiées expérimentalement, et se révèlent d'une certaine efficacité. Le travail présenté dans cette partie est issu d'une collaboration avec Alexis Bienvenüe, Éric Fontenas, Olivier François et Marc Joannidès.

Paysage de fitness aléatoire

Résumé

Dans cette partie, nous étudions le comportement d'algorithmes génétiques en environnement aléatoire. L'espace d'états est un arbre régulier de hauteur finie q, $\mathbb{T}^q$, sur lequel la fonction d'objectif F est définie par une marche aléatoire indexée par l'arbre. Les étapes de mutation et de sélection sont définies de la façon suivante :

- $\mathbf{M}$ = mutation : marche aléatoire simple sur l'arbre (symétrique, aux plus proches voisins)
- $\mathbf{S}$ = sélection de type Boltzmann, proportionnelle à $\exp(\beta F(x))$.

La taille de la population est en outre autorisée à varier avec le nombre d'itérations. Notre résultat principal est que, lorsque q tend vers l'infini, la valeur maximale de F présente dans la population après q itérations de l'algorithme est asymptotiquement équivalente à la valeur maximale de F sur $\mathbb{T}^q$, avec une probabilité qui tend vers 1. Le résultat est prouvé en réalisant $\mathbb{T}^q$ comme sous-arbre d'un arbre infini pour lequel des techniques de renouvellement sur les arbres permettent d'obtenir des résultats presques sûrs, via un contrôle détaillé de la façon dont la population est répartie par l'algorithme. La preuve met en évidence le rôle fondamental de la taille de la population, dont la croissance entraîne un effet de moyenne sur les irrégularités du paysage de fitness.

Motivations et résultats

Problématique et motivations

Dans les deux parties précédentes, les paysages de fitness considérés possédaient de fortes propriétés de régularité vis-à-vis de la mutation. L'effet de la mutation sur le fitness d'un individu ne dépendait, en loi, que du fitness de l'individu en question, et en dépendait de manière régulière. Comme expliqué précédemment, ces modèles présentent une pertinence en tant que modèles simplifiés d'évolution biologique, mais ne peuvent s'appliquer à des situations complexes d'optimisation. Nous nous proposons d'étudier dans cette partie le comportement d'algorithmes génétiques dont le paysage de fitness présente des irrégularités aléatoires, et se rapproche donc davantage d'un problème d'optimisation réaliste. Plus précisément, nous supposerons que l'espace d'états est un arbre régulier sur lequel la fonction d'objectif est donnée par une marche aléatoire indexée par l'arbre. La géométrie de l'espace d'états vis-à-vis des mutations est donc fixée, et l'irrégularité provient du fait que les pas de mutation font aléatoirement croître ou décroître la fonction d'objectif. Cette situation d'optimisation a été suggérée par D. Aldous [1] pour analyser le comportement de l'algorithme de Metropolis en environnement aléatoire, et il semble donc naturel d'étudier le comportement d'algorithmes génétiques dans le même contexte, en mettant en évidence le rôle spécifique joué par la présence d'une population d'individus qui interagissent via la sélection. D'autre part, l'études de paysages de fitness irréguliers ("rugged fitness landscapes") présente un intérêt en biologie [58]. Une motivation supplémentaire est fournie par l'étude des algorithmes génétiques sur les arbres de Galton-Watson dont la fonction d'objectif est la distance à la racine, qui peut être abordée à l'aide des méthodes que nous présentons. Ce sujet rejoint celui des marches aléatoires biaisées sur les arbres de Galton-Watson, étudiées notamment par R. Lyons, R. Pemantle et Y. Peres dans [47], et par A. Dembo, N. Gantert, Y. Peres et O. Zeitouni dans [24].

Description du modèle

Le paysage de fitness L'espace d'états $\mathbb{T}^q$ est un arbre $m-$régulier dont on note r la racine. À chaque arête e de $\mathbb{T}^q$, on associe un coût $v(e) \in \{+1, -1\}$, et l'on définit la fonction d'objectif F sur les sommets de $\mathbb{T}^q$ par $F(r) = 0$ et :

$$F(x) = \sum_{e \in \{r \to x\}} v(e),$$

où $\{r \to x\}$ désigne l'unique chemin injectif qui relie r à x.

En choisissant les $v(e)$ indépendants et identiquement distribués de loi commune $\rho\delta_1 + (1-\rho)\delta_{-1}$, F s'identifie à une *marche aléatoire indexée par l'arbre* $\mathbb{T}^q$ (cf. [57] et [59]). Lorsque $\rho > 1/2$, le comportement de F le long d'un rayon de l'arbre est celui d'une marche simple avec biais positif : F croît à vitesse positive le long des rayons, et la recherche de grandes valeurs de F (lorsque q est grand) est facile. Au contraire, lorsque $\rho < 1/2$, le biais est négatif et le comportement typique de F le long d'un rayon est de tendre vers $-\infty$. Cependant, lorsque ρ est plus grand qu'une valeur critique $\rho_{cr} > 0$, la croissance exponentielle de l'arbre compense en un sens le biais négatif, et il existe des rayons dans l'arbre le long desquels F croît à vitesse positive. Bien entendu, de tels rayons sont rares dans l'arbre, et leur localisation est inconnue *a priori* du fait du caractère aléatoire des $v(e)$. Dans la suite, nous ferons l'hypothèse, plus restrictive, que $\rho > 1/m$. On pourra donc choisir $\rho < 1/2$ dès que $m \geq 3$. Dans ce cas particulier, il existe des rayons le long desquels F croît à vitesse 1. Autrement dit, le maximum de F sur $\mathbb{T}^q$ est de l'ordre de q lorsque q tend vers l'infini.

L'algorithme génétique Les algorithmes que nous étudions sont définis comme des chaînes de Markov à temps discret inhomogènes par :

$$X_0 = (r, \dots, r) \in (\mathbb{T}^q)^{p_0}$$

et par les transitions suivantes :

1. **Mutation :** $X_n \longrightarrow Y_n$
   Chaque particule $X_n^{(i)}$ évolue indépendamment des autres et effectue un pas d'une marche aléatoire simple sur $\mathbb{T}^q$, symétrique, aux plus proches voisins. Nous supposerons que la probabilité que la particule reste fixe est égale à $\mu > 0$. Les positions respectives des particules après mutation sont notées $Y_n^{(i)}$, $1 \leq i \leq p_n$.
2. **Sélection :** $Y_n \longrightarrow X_{n+1}$
   Chaque $X_{n+1}^{(i)}$, pour $1 \leq i \leq p_{n+1}$, est choisi aléatoirement et indépendamment des autres dans l'ensemble $\left\{Y_n^{(i)},\ 1 \leq i \leq p_n\right\}$, selon la loi

$$\frac{1}{S'_n} \sum_{i=1}^{p_n} \exp\left(\beta f(Y_n^{(i)})\right) \delta_{Y_n^{(i)}}, \quad \text{où} \quad S'_n = \sum_{i=1}^{p_n} \exp\left(\beta f(Y_n^{(i)})\right).$$

Les particules $X_n^{(1)}, \dots, X_n^{(p_n)}$ constituent la population à la date n. À la différence des parties précédentes, la taille de la population n'est pas supposée fixe, mais dépend de n de façon déterministe : $p = p_n$.

La quantité principale à laquelle nous nous intéresserons est la valeur maximale de F présente dans la population à la date n, $F^*(X_n)$, avec la notation : $F^*(x_1, \dots, x_p) = \max_{1 \leq i \leq p} F(x_i)$. Nous étudierons également la mesure de comptage des valeurs de F, $Z_F(X_n)$, avec la notation, pour $x \in \mathbb{T}^p$, $Z_F(x) = \sum_{1 \leq i \leq p} \delta_{F(x_i)}$

Résultats

Nous supposerons que la taille de la population croît sous-exponentiellement avec le nombre d'itérations, mais qu'elle croît strictement plus rapidement que $\log(n)^2$:

- $\lim_{n \to +\infty} p_n (\log n)^{-2} = +\infty$
- $\lim_{n \to +\infty} p_{n+1}/p_n = 1$

D'autre part, nous supposerons que l'intensité de la sélection est suffisament élevée :

- $\exp\beta > [(1-\mu)\,\rho\,]^{-1}(1+1/m)$
- $\exp\beta > 1/\mu$

Le résultat principal est le suivant :

Théorème 7 *Pour tout $\epsilon > 0$, lorsque q tend vers l'infini,*

$$\mathrm{P}\left[F^*(X_q) \geq (1-\epsilon)q\right] \to 1$$

Autrement dit, $q^{-1}F^*(X_q)$ tend vers 1 en probabilité lorsque q tend vers l'infini, ce qui signifie que, après q itérations, l'algorithme trouve une valeur de F de l'ordre de sa valeur maximale sur $\mathbb{T}^q$. On notera que la probabilité qui intervient dans le théorème se réfère aux deux niveaux d'aléatoire présents dans le problème : l'environnement aléatoire F, et la réalisation de l'algorithme, elle-même aléatoire. Ce résultat se rapproche par sa formulation des résultats de *rapid mixing* pour les chaînes de Markov finies, puisque l'on peut choisir (p_n) de telle façon que l'algorithme trouve une valeur de F proche du maximum tout en n'utilisant qu'une taille de population et un nombre d'itérations croissant au plus polynomialement avec q, alors que la taille de l'espace d'états croît exponentiellement vite avec q.

Description informelle des preuves

Plongement dans l'arbre infini

Une première idée de la preuve consiste à identifier $\mathbb{T}^q$ à un sous-arbre d'un arbre $m-$régulier enraciné infini, de même racine, muni d'une fonction $\tilde{F}$ qui prolonge F en loi. Ceci ne présente aucune difficulté : il suffit de définir, sur un arbre $m-$régulier enraciné infini $\mathbb{T}$, une fonction $\tilde{F}$ par : $\tilde{F}(r) = 0$ et :

$$\tilde{F}(x) = \sum_{e \in \{r \to x\}} w(e),$$

les $w(e)$ formant une famille i.i.d. de variables aléatoires de loi $\rho\delta_1 + (1-\rho)\delta_{-1}$. La restriction de $\tilde{F}$ à $\mathbb{T}^q$ possède trivialement la même loi que F. On s'intéresse ensuite au comportement de l'algorithme génétique $(\tilde{X}_n)$ défini sur $\mathbb{T}$ avec la fonction d'objectif $\tilde{F}$, avec les mêmes paramètres (population (p_n), intensité de sélection β) que précédemment. En notant que $(\tilde{X}_n)$ est initialisé à la racine, et que les pas de mutation sont aux plus proches voisins, on vérifie que, pour tout $1 \leq n \leq q$, et tout $1 \leq i \leq p_n$, $\tilde{X}_n^{(i)} \in \mathbb{T}^q$. Par conséquent, la réalisation de l'algorithme $(\tilde{X}_n)$ jusqu'au temps q ne fait pas intervenir les valeurs de $\tilde{F}$ hors de $\mathbb{T}^q$, et l'on a l'identité en loi suivante (que l'on peut aisément réaliser par un couplage) :

$$\left\{(\tilde{X}_n)_{1\leq n\leq q}\ ,\ \tilde{F}(\mathbb{T}^q)\right\} \stackrel{loi}{=} \left\{(X_n)_{1\leq n\leq q}\ ,\ F(\mathbb{T}^q)\right\}$$

Le résultat que nous nous proposons en fait de démontrer est le suivant :

Théorème 8 *Presque sûrement en la réalisation de $\tilde{F}$ et en la réalisation de $(\tilde{X}_n)$,*

$$\tilde{F}(\tilde{X}_n) = n - O(1).$$

L'identité en loi précédente montre que le théorème 7 est une conséquence du théorème ci-dessus, en utilisant le fait que la convergence presque sûre entraîne la convergence en probabilité. Comme $q^{-1}F^*(X_q)$ est borné, on a également convergence dans tous les L^s pour $1 \leq s < +\infty$.

Remarque 6 *Pour simplifier les notations, nous noterons à présent $(\tilde{X}_n)$ et $\tilde{F}$ par (X_n) et F respectivement.*

Renouvellement sur l'arbre infini

L'intérêt de se ramener à un problème sur un arbre infini est que les fortes propriétés d'indépendance du paysage de fitness peuvent être exploitées afin de prouver des résultats presque sûrs, comme l'article [47] en fournit plusieurs illustrations. La filtration associée naturellement à l'échantillonnage progressif du paysage de fitness par l'algorithme est définie de la façon suivante :

$$\mathcal{F}_n = \sigma\left(X_0, \ldots, X_n, v(e(X_0)), \ldots, v(e(X_{n-1}))\right)$$

où $e(x)$ désigne les arêtes de $\mathbb{T}$ issues de x. Dans ce contexte, les instants de renouvellement de l'algorithme, c'est-à-dire les instants où l'algorithme visite un sommet pour la première fois, jouent un rôle crucial. En effet, lorsqu'un sommet x se trouve représenté pour la première fois dans la population, les valeurs des $v(e)$, où e décrit l'ensemble des arêtes du sous-arbre de $\mathbb{T}$ issu de x, sont indépendantes du passé, et leur loi conditionnelle est donc connue. Nous aurons besoin d'une définition plus restrictive :

Définition *L'instant n est appelé temps de record de l'algorithme si, pour tout $0 \leq i \leq n-1$, $F^*(X_n) > F^*(X_i)$.*

En particulier, lorsque n est un temps de record, tous les individus de la population qui vérifient $F(X_n^{(i)}) = F^*(X_n)$ figurent pour la première fois dans la population. Les temps de record sont naturellement des temps d'arrêts de la filtration $(\mathcal{F}_n)$.

L'intérêt de cette définition réside dans le fait qu'il existe presque sûrement une infinité de temps de records, et que l'on dispose donc de bonnes propriétés de renouvellement pour l'algorithme.

En particulier, pour montrer qu'un événement A se produit presque sûrement, il suffit de montrer que, conditionnellement au fait que n soit un temps de record,

$$\mathrm{P}\left[A \mid \mathcal{F}_n\right] \geq \gamma > 0.$$

Nous choisirons naturellement

$$A = \{\exists n_0 \ , \ \forall n \geq n_0, F^*(X_{n+1}) = F^*(X_n) + 1\}.$$

Autrement dit, A signifie qu'après un certain temps, tout instant est un temps de record.

Expliquons brièvement comment prouver l'existence d'une infinité de temps de records. Il s'agit donc de montrer que la suite $F^*(X_n)$ n'est pas majorée. Montrons d'abord que $F^*(X_n)$ est presque sûrement minorée. Introduisons la suite $a_n(\lambda)$ définie par :

$$a_n(\lambda) = \max\left\{a : Z_F(X_n)([a, +\infty[) \geq \lambda p_n\right\}$$

La probabilité pour un individu de rester fixe par mutation étant positive, égale à μ, on a $Z_F(Y_n)([a_n(\lambda), +\infty[) \gtrapprox \mu\lambda p_n$. On vérifie alors que, si λ est choisi assez petit, du fait de la sélection, $Z_F(X_{n+1})([a_n, +\infty[) \gtrapprox \exp(\beta)\mu\lambda p_{n+1}$, d'où, d'après l'hypothèse $\exp(\beta)\mu > 1$, $a_{n+1}(\lambda) \geq a_n(\lambda)$. Les écarts à ce comportement moyen sont contrôlés en grandes déviations par des expressions du type $\exp(-Cp_n)$, et l'on vérifie donc, grâce au lemme de Borel-Cantelli, puisque l'on a supposé que $p_n \gg \log(n)^2$, que $a_{n+1}(\lambda) \geq a_n(\lambda)$ sauf pour un nombre fini d'indices n. Ceci qui entraîne que $F^*(X_n)$ est une suite presque sûrement minorée. La fin de la preuve consiste à vérifier que les valeurs de $F^*(X_n)$ ne peuvent pas s'accumuler.

Répartition de la population

Le résultat de la section précédente suggère d'introduire les événements :

$$A_n = \{n \text{ est un temps de record}\}$$

et de minorer

$$\mathrm{P}\left[\bigcap_{i=n+1}^{+\infty} A_i \;\middle|\; \mathcal{F}_n\right]$$

sur A_n.

Une première idée est d'essayer de minorer $\mathrm{P}\left[A_{n+1} \mid \mathcal{F}_n\right]$ sur A_n, puis d'effectuer le produit des minorations obtenues. Malheureusement, on ne peut obtenir de minoration utile si l'on ne fait pas d'hypothèse supplémentaire sur la répartition de la population. En effet, si par exemple, n étant un temps de record, tous les individus de niveau $F^*(X_n)$ sont localisés sur un seul sommet x de l'arbre, la probabilité conditionnelle à $\mathcal{F}_n$ pour que toutes les arêtes issues de x soient affectées d'un coût égal à -1 est $(1-\rho)^m$. Si ceci se produit, $n+1$ ne peut pas être un temps de record, et par conséquent, sur A_n,

$$\mathrm{P}\left[A_{n+1} \mid \mathcal{F}_n\right] \leq 1-(1-\rho)^m,$$

et la minoration de $\mathrm{P}\left[A \mid \mathcal{F}_n\right]$ que l'on obtient est triviale.

Nous chercherons donc à contrôler, outre le fait que n est un temps de record, la répartition de la population sur les sommets de niveau $F^*(X_n)$, via les deux quantités M_n et T_n^* définies par :

$$M_n = \#\{1 \leq i \leq p_n \ ; \ F(X_n^{(i)}) = F^*(X_n)\},$$

et,

$$T_n^* = \max\{T_n(x) \ ; \ F(x) = F^*(X_n)\},$$

où, pour $x \in \mathbb{T}$,

$$T_n(x) = \#\{1 \leq i \leq p_n \ ; \ X_n^{(i)} = x\}.$$

M_n est donc une quantité globale, à savoir le nombre total d'individus de la population localisés au niveau $F^*(X_n)$. T_n^* est une quantité locale, qui contrôle la répartition de ces individus sur chacun des sommets, à savoir le nombre maximum d'individus situés sur un sommet de niveau $F^*(X_n)$. La population est donc bien répartie sur le niveau $F^*(X_n)$ lorsque M_n est grand et T_n^* petit devant M_n.

Étudions l'effet moyen d'un pas de l'algorithme sur ces quantités, en supposant que nous disposons, à la date n, d'un contrôle sur ces paramètres du type :

$$B_n(m_n, t_n) = A_n \cap \{M_n \geq m_n \ , \ T_n^* \leq M_n/\ell_n\}$$

Conditionnellement au fait que n est un temps de record, les coûts $v(e)$ des arêtes e qui descendent des sommets x situés au niveau $F^*(X_n)$ sont des variables aléatoires indépendantes, de loi commune $\rho\delta_1+(1-\rho)\delta_{-1}$. D'autre part, la probabilité pour un individu de monter par mutation le long de l'une de ces arêtes est égale à $(1-\mu)m/(m+1)$. Par conséquent, il y a en moyenne $\rho(1-\mu)(m/(m+1))M_n$ individus situés après mutation au niveau $F^*(X_n)+1$. Les propriétés de la sélection de Boltzmann entraînent que, pour λ assez petit, et n assez grand,

$$M_{n+1} \gtrapprox \exp(\beta)\rho(1-\mu)(m/(m+1))M_n \text{ si } M_n \leq \lambda p_n,$$

$$M_{n+1} \gtrapprox \exp(\beta)\rho(1-\mu)(m/(m+1))\lambda p_{n+1} \text{ si } M_n > \lambda p_n$$

L'hypothèse $\exp\beta > [(1-\mu)\,\rho\,]^{-1}(1+1/m)$ montre donc que, en moyenne, $M_{n+1} > aM_n$ avec $a>1$, ou $M_{n+1} \geq \lambda p_{n+1}$.

D'autre part, si l'on considère un sommet y situé au niveau $F^*(X_n)+1$, le nombre d'individus situés sur y après mutation est en moyenne inférieur ou égal à $((1-\mu)/(m+1))T_n^*$ puisqu'il y a au plus T_n^* individus situés sur son prédecesseur ayant chacun une probabilité égale à $(1-\mu)/(m+1)$ d'atteindre y par mutation. Conditionnellement à M_{n+1}, les individus de niveau $F^*(X_n)+1$ ont tous la même probabilité d'être choisis lors de l'étape de sélection, d'où le fait que, en moyenne

$$T_{n+1}(y)M_{n+1}^{-1} \lessapprox ((1-\mu)/(m+1))T_n^*(\rho(1-\mu)(m/(m+1))M_n)^{-1},$$

d'où

$$(T_{n+1}^*) \lessapprox M_{n+1}(m\rho\ell_n)^{-1},$$

et nous avons supposé que $m\rho > 1$.

En admettant que le comportement réel s'écarte peu de ce comportement moyen, on obtient donc que $B_{n+1}(m_{n+1},\ell_{n+1})$ est réalisé avec $m_{n+1} \gtrapprox \inf(am_n, \lambda p_{n+1})$ et $\ell_{n+1} \gtrapprox m\rho\ell_n$.

De fait, les deux paramètres M_n et T_n^* suffisent à contrôler les grandes déviations par rapport à ce comportement moyen, si bien que, sur $B_n(m_n,\ell_n)$,

$$\mathrm{P}\left[B_{n+1}(m_{n+1},\ell_{n+1}) \mid \mathcal{F}_n\right] \geq 1-\exp\left(-C\ell_n\right)+\exp\left(-Cm_n\ell_n^{-1}\right),$$

On peut donc spécifier l'évolution temporelle de ces paramètres en deux phases :

1. un régime transitoire de croissance exponentielle de m_n, ℓ_n, et $m_n\ell_n^{-1}$ au cours de laquelle M_n finit par dépasser λp_n, du fait que nous avons supposé la croissance de la taille de la population sous-exponentielle,
2. un régime stable au cours duquel $m_n \geq \lambda p_n$ et $\ell_n \geq c(p_n)^{1/2}$, la population étant bien répartie sur le niveau $F^*(X_n)$.

de telle façon que le produit infini

$$\gamma = \prod_{i=1}^{+\infty}\left(1-\exp\left(-C\ell_n\right)+\exp\left(-Cm_n\ell_n^{-1}\right)\right)$$

soit strictement positif, puisque nous avons supposé que $p_n \gg \log(n)^2$.

On obtient ainsi le fait que, sur A_n, $\mathrm{P}\left[A \mid \mathcal{F}_n\right] \geq \gamma$, et donc le fait que $\mathrm{P}\left[A\right]=1$.

La croissance de l'arbre et la croissance de la taille de la population produisent donc un effet de moyenne sur les irrégularités du paysage de fitness (à comparer à l'effet de moyenne observé dans [24] du fait de l'incertitude sur la position de la marche) qui permet à la mutation et à la sélection de rechercher efficacement les grandes valeurs de F.

Extensions et questions ouvertes

Répartition empirique de F

On peut en fait montrer que, sous l'effet de la sélection, la répartition empirique des valeurs de F au sein de la population est concentrée autour de $F^*(X_n)$, au sens suivant : il existe $c > 0$ tel que, pour tout $a \geq 1$, et pour tout $n \geq k_a$, où k_a est aléatoire et fini presque sûrement,

$$p_n^{-1} Z_F(X_n)([-\infty, F^*(X_n) - a]) \leq c \exp(-\beta a).$$

Il est alors naturel de se demander si la suite des mesures définies par :

$$\mu_n(a) = Z_F(X_n)([-\infty, F^*(X_n) - a])$$

converge presque sûrement en loi lorsque $n \to +\infty$. Dans le cas où le paysage de fitness est régulier, on peut répondre par l'affirmative, mais, dans le cas présent, la preuve se heurte à des difficultés techniques qui ne sont pas encore surmontées.

Le cas Galton-Watson

Comme indiqué précédemment, on peut traiter, avec la méthode que nous venons de décrire, le cas où $\mathbb{T}$ est un arbre de Galton-Watson surcritique conditionné par la non-extinction, et $F(x) = d(r, x)$. L'irrégularité est donc présente dans l'opérateur de mutation lui-même, et la difficulté du problème tient à la présence de sommets sans descendants, qui sont des maxima locaux de F. On obtient le même type de résultat, à savoir que $F^*(X_n) = n - O(1)$ presque sûrement. L'intérêt est que, dans ce contexte, le comportement de l'algorithme de Metropolis, c'est-à-dire des marches aléatoires biaisées sur l'arbre de Galton-Watson, est relativement bien connu. L'existence d'une vitesse limite presque sûre est prouvée dans [47], et l'étude des grandes déviations par rapport à cette vitesse limite est menée dans [24]. On peut voir en particulier que, du fait de l'existence de grandes déviations, le supremum de p_n algorithmes de Metropolis indépendants a une probabilité exponentiellement petite d'être de l'ordre de n, car la vitesse asymptotique est strictement inférieure à 1. En ce sens, il est plus efficace, dans ce contexte d'optimisation, de faire interagir les individus par sélection plutôt que d'effectuer des recherches indépendantes.

Pression de sélection

Le résultat que nous démontrons suppose que β est suffisamment grand, mais n'impose pas de borne supérieure à la valeur que peut prendre β. Ceci peut sembler surprenant si l'on pense, par exemple, aux marches aléatoires sur les arbres de Galton-Watson. En effet, lorsque le biais vers les descendants est trop fort, la vitesse de croissance de la marche s'annule du fait que celle-ci est piégée dans les sous-arbres finis, dont les feuilles sont autant de maxima locaux de $d(r,x)$. Dans notre cas, la présence de maxima locaux n'affecte pas la vitesse asymptotique de l'algorithme lorsque β est grand, car la croissance de la taille de la population permet la découverte constante de nouvelles arêtes de l'arbre le long desquels la fonction d'objectif croît. Il est toutefois probable que la valeur de β influe sur la vitesse de convergence de $n^{-1}F^*(X_n)$ vers 1, - vitesse sur laquelle notre preuve ne fournit pas d'information. D'autre part, le comportement de l'algorithme dans le cas où β est petit n'est pas encore déterminé. Des simulations numériques sont en cours afin de préciser empiriquement l'influence de β.

Taille de la population

Le comportement de l'algorithme lorsque la taille de la population croît plus lentement que $\log(n)^2$, n'est pas non plus déterminé. Dans le cas d'une population constante, par exemple, les techniques usuelles pour étudier les comportements asymptotiques de marches aléatoires en environnement aléatoire, telles que l'ergodicité de l'environnement vu de la particule, semblent difficiles à mettre en œuvre. Même la transience de l'algorithme en tant que chaîne de Markov n'est pas établie. L'existence d'une limite presque sûre (positive ou négative) pour $n^{-1}F^*(X_n)$ reste donc conjecturale. On notera que, dans le cas de l'algorithme de Metropolis étudié par Aldous, où l'ergodicité de l'environnement vu de la particule est prouvée, le signe de la vitesse asymptotique demeure indéterminé car la mesure invariante n'est pas connue explicitement. L'intérêt de notre approche est justement de s'affranchir de ces difficultés techniques en autorisant la taille de la population à croître modérément avec le nombre d'itérations.

À l'inverse de l'algorithme en population constante finie, la version en population infinie de l'algorithme est facile à étudier. On raisonne à paysage de fitness fixé (quenched). Si ξ_n désigne la loi au temps n de l'algorithme en population infinie, ξ_n est déterminée par une formule du type Feynman-Kac [23], et l'on a

$$\xi_n(\{F \leq (1-\epsilon)n\}) = \frac{\mathrm{E}\big(\exp\left(\beta(F(U_1)+\ldots+F(U_n))\right)\mathbf{1}_{\{F(U_n)\leq(1-\epsilon)n\}}\big)}{\mathrm{E}(\exp\left(\beta(F(U_1)+\ldots+F(U_n))\right))}$$

où $(U_i)_{i\geq 0}$ est une marche aléatoire simple sur $\mathbb{T}$ de même loi que les étapes de mutation initialisée à la racine. Il existe presque sûrement en l'environnement un entier C et un rayon de l'arbre ne comportant que C arêtes de coût -1. En considérant la contribution d'un tel rayon, on obtient que :

$$\mathrm{E}(\exp\left(\beta(F(U_1)+\ldots+F(U_n))\right)) \geq a^n \exp\left(\tfrac{1}{2}\beta n(n+1) - nC\right),$$

où $a = (1-\mu)(m+1)$ est la probabilité de passer par mutation d'un sommet à l'un de ses voisins.

À présent, un chemin de l'arbre $r = u_0, u_1, \ldots, u_n$ tel que $F(u_n) \leq (1-\epsilon)n$ vérifie nécessairement, puisque $F(r) = 0$, et que F ne peut varier que d'au plus 1 à chaque pas,

$$F(u_1) + \ldots + F(u_n) \leq \tfrac{1}{2}n(n+1)C(\epsilon)$$

avec $C(\epsilon) < 1$.

On en déduit que

$$\xi_n(\{F \leq (1-\epsilon)n\}) \leq a^{-n} \exp\left(\tfrac{1}{2}\beta(C(\epsilon) - 1)n(n+1)\right),$$

et la loi de F sous ξ_n est donc asymptotiquement concentrée autour de n, au sens d'une inégalité exponentielle.

Notre résultat peut donc s'interpréter comme une propriété d'approximation de l'algorithme en population infinie par un algorithme dont la taille de la population est finie et croît avec le nombre d'itérations. Cependant, les résultats généraux d'approximation de [21] et [22] ne semblent pas s'appliquer ici. On notera par ailleurs que la preuve concernant le comportement de l'algorithme en population infinie n'utilise que l'existence dans l'arbre d'un rayon ne comportant qu'un nombre fini de coûts égaux à -1, et non la forme spécifique de la loi du paysage de fitness. Il est peu probable que les algorithmes en population finie se comportent de la même façon sur une classe aussi vaste de problèmes, et il serait intéressant de disposer d'un contre-exemple.

Valeur de ρ

Comme nous l'avons mentionné, notre résultat ne couvre qu'une partie des valeurs de ρ pour lesquelles la recherche de grandes valeurs de F sur l'arbre est pertinente. Dans le cas $\rho > 1/m$ que nous avons considéré, il existe dans l'arbre un sous-arbre de Galton-Watson surcritique infini le long des arêtes duquel tous les coûts sont égaux à 1. Notre résultat signifie que, pour découvrir les rayons d'un tel sous-arbre, il suffit d'explorer l'arbre total à l'aide d'un nombre de particules croissant lentement, par comparaison avec la croissance exponentielle caractéristique des processus de branchement surcritiques. Dans le cas où $\rho_{cr} < \rho \leq 1/m$, on peut encore construire, grâce à un processus de branchement surcritique, des rayons de l'arbre le long desquels F croît à vitesse positive (voir [59]). Cette construction pourrait permettre de généraliser notre résultat au cas où $\rho_{cr} < \rho < 1/m$. Par ailleurs, on peut facilement définir la marche aléatoire indexée par l'arbre à l'aide de coûts sur les arêtes plus généraux que des v.a. de Bernoulli ± 1, et il serait intéressant d'étudier le comportement d'algorithmes génétiques dans ce contexte.

Grandes déviations

L'étude des grandes déviations pour les marches aléatoires en environnement aléatoire est un sujet de recherche actif, (voir par exemple [17], [70], [74]) et il serait intéressant de disposer de résultats de ce type dans notre situation.

Chapitre 1

Sharp asymptotic results for mutation-selection algorithms with additive mutations

1.1 Abstract

We study the asymptotic behaviour of a mutation-selection genetic algorithm on the integers with finite population of size $p \geq 1$. The mutation is defined by the steps of a simple random walk and the fitness function is linear. We prove that the normalized population satisfies an invariance principle, that a large deviations principle holds, and that the relative positions converge in law. After n steps, the population is asymptotically around $\sqrt{n}$ times the position at time 1 of a Bessel process of dimension $2p-1$. The results presented in this chapter have been obtained in collaboration with Alexis Bienvenüe.

1.2 Introduction

1.2.1 Motivation

J. Holland [39] introduced genetic algorithms as an optimization method, inspired by a biological analogy with (what one assumes to be) the mechanisms of Darwinian evolution. They are now a popular tool for solving hard combinatorial optimization problems. In these algorithms, a finite population of particles evolves under the action of three operators:

- Selection randomly resamples the population. Particles with higher fitness are more likely to be selected, and there is a tendency to eliminate the particles with lower fitness.

- Mutation modifies randomly each particle.
- Mating creates a new population of "offsprings", from pairs of particles of the previous population.

In a combinatorial optimization context, particles are feasible solutions, and the fitness of a particle is the function to be maximized. Thus, the selection operator directs the evolution towards an increase of the fitness, while mutation and mating preserve the population diversity, and allow the algorithm to visit large parts of the space of solutions.

Despite the successes that genetic algorithms encounter in practical applications, and the numerous experimental studies of their properties, there exist few rigorous results on their behaviour. Among the exceptions is R. Cerf [15], who obtains asymptotic convergence results for genetic algorithms with rare transitions. See also Y. Rabinovich and A. Wigderson [62], who study the convergence speed of several genetic algorithms defined on binary strings, and the rate of escape of algorithms defined on the integers.

The algorithms we study belong to this latter category. We deliberately choose a very simplified context, to get a detailed understanding of the effects of selection through a fully rigorous mathematical treatment. The price to pay is that our results do not apply to complex optimization situations. However, this simple model has been used in biology to study the evolution of a population of viruses, see L. Tsimring, H. Levine, D. Ridgway, and D. Kessler [72], [43]. Moreover, D. Bonnaz [13] applied theoretical physics methods to get predictions about its behaviour (asymptotics of the mean and of the variance of the population) and supported them by numerical simulations. Our results confirm these predictions.

1.2.2 Description of the model and notations

A mutation-selection genetic algorithm on the integers with finite population of size $p \geq 1$ is specified by a simple random walk and by the fitness function $F(x) = x$, as follows.

Define a Markov chain $X_n = (X_n^{(i)})_{1 \leq i \leq p}$ with state space $\mathbb{N}^p$ and starting state $X_0 = (1, \ldots, 1)$, by the following transitions.

1. **Selection step:** $X_n \longrightarrow Y_n$

   If $X_n = (0, \ldots, 0)$, then $Y_n = (1, \ldots, 1)$.
   Else, each $Y_n^{(i)}$, $1 \leq i \leq p$, is chosen randomly and independently of the others in the set $\left\{X_n^{(i)},\ 1 \leq i \leq p\right\}$ according to the probability law

$$\frac{1}{S_n} \sum_{i=1}^{p} X_n^{(i)} \delta_{X_n^{(i)}}, \quad \text{where} \quad S_n = \sum_{i=1}^{p} X_n^{(i)}.$$

2. **Mutation step:** $Y_n \longrightarrow X_{n+1}$

   Each particle $Y_n^{(i)}$ evolves independently of the others, and performs one step of a simple random walk on $\mathbb{N}$ (symmetric, to the nearest neighbours). The new positions are $X_{n+1}^{(i)}, 1 \leq i \leq p$.

The particles $X_n^{(1)}, \ldots, X_n^{(p)}$ compose the population at time n. Thus, the number of particles does not vary during the time-evolution of the population. At a given time, several particles may assume the same value. The choice $X_n = (0, \ldots, 0) \longrightarrow Y_n = (1, \ldots, 1)$ is arbitrary. The point here is to prevent particles from taking negative values. One can also modify suitably the mutation operator.

We use the following notations. $\mathcal{F}_n$ and $\mathcal{G}_n$ are two $\sigma-$algebras:

$$\mathcal{F}_n = \sigma\left(X_0, Y_0, \ldots, X_{n-1}, Y_{n-1}, X_n\right),$$

and $\mathcal{G}_n$ is the σ-algebra generated by $\mathcal{F}_n \cup\ \sigma\left(Y_n\right)$. We set

$$S_n = \sum_{i=1}^{p} X_n^{(i)}, \quad T_n = \sum_{i=1}^{p} Y_n^{(i)},$$

$$F_n = S_n/p, \quad G_n = T_n/p, \quad \Delta F_n = F_{n+1} - F_n.$$

If $x \in \mathbb{N}^p$, E^x denotes the expectation under the law of the Markov chain X starting from $X_0 = x$. The barycenter $m(x)$ and the diameter $d(x)$ of x are

$$m(x) = (x_1 + \cdots + x_p)/p, \quad d(x) = \sup_{1 \leq i,j \leq p} |x_i - x_j|$$

Finally, $M_n(t) = X_{\lfloor nt \rfloor}/\sqrt{n}$ and $Z_n(t) = m(M_n(t)) = F_{\lfloor nt \rfloor}/\sqrt{n}$.

1.2.3 Statement of the results

For any fixed p, we describe the asymptotic behaviour of (X_n), when n goes to infinity, as follows.

Theorem 1 *The sequence of processes $(M_n(t))_{0 \leq t \leq 1}$ converges in law in the Skorohod space $\mathcal{D}([0,1], \mathbb{R}^p)$ to $(1, \ldots, 1)R$, where R is a Bessel process of dimension $(2p-1)$, starting from zero. Recall that a Bessel process in dimension $q \geq 1$ can be defined as the euclidean norm of a standard $q-$dimensional brownian motion starting from zero.*

Theorem 2 *For all i, The sequence $\left(X_n^{(i)}/n\right)$ satisfies a large deviations principle with good rate function I, where I is the rate function associated with a simple reflected random walk on $\mathbb{N}$, that is,*

$$\begin{cases} I(a) = \frac{1}{2}\left[(1-a)\log(1-a) + (1+a)\log(1+a)\right] & \text{if } a \in [0,1] \\ I(a) = +\infty & \text{otherwise.} \end{cases}$$

To state our last result, say that two elements $x, y \in \mathbb{N}^p$ are equivalent whenever $x - y$ is an integer multiple of $(1, \ldots, 1)$, and denote by $\Theta(x)$ the class of x for this equivalence relation.

Theorem 3 *$\Theta(X_n)$ converges in law to a law π_p, which is also the limiting law in the case of uniform selection (see the remarks about the description of π_p in Section 1.6).*

Theorem 1 has been announced in [8], and Theorems 2 and 3 have been announced in [7].

The $\sqrt{n}$ growth rate of the fitness is the same with or without selection, for any population size p. No selection would give a simple reflected random walk. However, as p increases, the limiting law is more and more different from the reflected normal law, that would hold without selection. It gives more and more weight to the large values. Moreover, normalizing by n, the large deviations principle is identical to the case of a simple reflected random walk. Thus, Theorems 2 and 3 imply that the selection has no effect, in the limit, upon X_n/n nor upon $\Theta(X_n)$.

The behaviour of the infinite population version of this model is strikingly different, see [50]. In the infinite population model, the fitness grows linearly, as opposed to our $\sqrt{n}$ scaling factor. A satisfactory explanation of this sharp transition phenomenon is still lacking.

1.2.4 Contents

In Section 1.3, we prove that the cloud of particles goes to infinity in probability, and we give a stochastic upper bound of the diameter of the cloud. Sections 1.4 and 1.5 are devoted to the detailed exposition of the invariance principle for the normalized cloud (Theorem 1) and of the large deviations principle (Theorem 2). In Section 1.6, we prove the convergence in law of the non-normalized cloud of particles (Theorem 3). Section 1.7 collects technical lemmas. In Section 1.8, we present numerical simulations of the algorithm.

1.3 Two preliminary results

From Lemma 1, the cloud of particles X_n is stochastically bounded from below by a mutation-selection process V_n associated with uniform selection. For fixed i, and n, each $V_n^{(i)}$ follows the same law as the position at time n of a simple reflected random walk. However, for fixed i, the process $(V_n^{(i)})_n$ itself is not a simple random walk. Lemma 2 provides a stochastic upper bound on the diameter of X_n.

Lemma 1 *Let $X_0 = x \in \mathbb{N}^p$. There is a process $V_n = (V_n^{(i)})_{1 \leq i \leq p}$ whose law does not depend on x, such that:*

- $V_n^{(i)} \leq X_n^{(i)}$;
- $V_n^{(i)}$ *follows the law of the position at time n of a simple symmetric random walk on $\mathbb{N}$ starting from zero with transition probabilities $p_{0\to 0} = p_{0\to 1} = 1/2$.*

Corollary 7 *For fixed $1 \leq i \leq p$, $X_n^{(i)}$ goes to infinity in probability, that is:*

$$\forall K > 0,\ \mathrm{P}\left[X_n^{(i)} \leq K\right] \xrightarrow[n\to+\infty]{} 0.$$

Proof of Lemma 1:

We couple X and V, and we introduce an intermediate step W_n between V_n and V_{n+1}. Let $V_0 = (0, \ldots, 0)$, and assume that X_t and V_t satisfy the desired properties for $t \leq n$.

Let $U_1, \ldots, U_p$ be independent random variables, independent of $\mathcal{F}_n$, of uniform law on $[0,1]$. Let $\epsilon_1, \ldots, \epsilon_p$ be independent random variables, independent of $\mathcal{F}_n$, of symmetric ± 1 Bernoulli law.

We denote by $x_1 = X_n^{(q_1)} \leq \cdots \leq x_p = X_n^{(q_p)}$ the ordered values of the p-tuple X_n. Let $s = \sum_{i=1}^p x_i$, and $v_1 = V_n^{(q_1)}, \ldots, v_p = V_n^{(q_p)}$. By the induction hypothesis, $v_i \leq x_i$ for all $i = 1, \ldots, p$. Define the selection step as follows:

- If $X_n = (0, \ldots, 0)$, $Y_n = (1, \ldots, 1)$ and $W_n = V_n$.
- Else:
$$Y_n^{(i)} = x_{k_i}, \qquad W_n^{(i)} = v_{l_i},$$
where k_i and l_i are uniquely defined by:
$$\sum_{j=1}^{k_i-1} x_j < sU_i \leq \sum_{j=1}^{k_i} x_j \qquad \forall i = 1, \ldots, p,$$
$$l_i - 1 < pU_i \leq l_i \qquad \forall i = 1, \ldots, p.$$
Since $(x_i)_i$ is a non-decreasing sequence, $l_i \leq k_i$ and
$$W_n^{(i)} = v_{l_i} \leq x_{l_i} \leq x_{k_i} = Y_n^{(i)}.$$

The mutation step is defined as follows:

$$\begin{aligned} X_{n+1}^{(i)} &= \left(Y_n^{(i)} + \epsilon_i\right), \\ V_{n+1}^{(i)} &= \sup\left\{W_n^{(i)} + \epsilon_i, 0\right\}. \end{aligned}$$

Then $V_{n+1}^{(i)} \leq X_{n+1}^{(i)}$, and, for all n and i, the law of $V_n^{(i)}$ is as desired. $\square$

Lemma 2 *Let $X_0 = x \in \mathbb{N}^p$. For all $n \in \mathbb{N}^*$, there is a random variable A_n, of geometric law on $\mathbb{N}^*$ with parameter p^{1-p}, such that:*

$$d(X_n) \leq d(x) + 2A_n$$

Proof of Lemma 2:

The proof relies on the following crucial tool :

Definition 1 *For $n \geq 0$, call n a crunch time if the population Y_n is concentrated on one point. That is, the selection step $X_n \longrightarrow Y_n$ amounts to choosing p times the same value.*

The idea of the proof is the following: the probability that n is a crunch time is strictly bounded away from zero. Hence, looking backwards in time, n and the last crunch time before n are not very distant apart. Since a mutation step does not increase the diameter by more than a constant, a bound on the diameter at time n follows.

Given $\mathcal{F}_n$, the conditional probability q_n that n is a crunch time satisfies

$$q_n \geq \frac{1}{S_n^p} \sum_{k=1}^{p} \left(X_n^{(k)} \right)^p .$$

Indeed, if $X_n = (0, \dots, 0)$, then $q_n = 1$. Else, the probability of selecting p times the index j is $(X_n^{(j)})^p / S_n^p$. Hence: $q_n \geq p^{1-p}$.

In order to couple $d(X_n)$ and A_n, we build the selection steps through a sequence of random parameters U_n, as follows.

- Let $(U_n)_n$ be i.i.d. random variables, of uniform law on $[0, 1]$, each U_n being independent of $\mathcal{F}_n$.
- If $U_n \leq q_n$, then $Y_n^{(1)} = \cdots = Y_n^{(p)}$.

Define B_n by:

$$B_n = n - \sup \left(\left\{ k < n, \ U_k \leq p^{1-p} \right\} \cup \{-1\} \right) .$$

The above construction yields

$$\{B_n \leq n\} \subset \{d(Y_{n-B_n}) = 0\}.$$

Moreover, the independence of the random variables U_n implies that B_n may be written as $B_n = \min\{A_n, (n + 1)\}$, where A_n follows a geometric law on $\mathbb{N}^*$ with parameter p^{1-p}. A selection step does not increase the diameter of the cloud of particles. A mutation step does not increase the diameter by more than 2. Hence,

- if $B_n \leq n$ (that is, if $A_n \leq n$), then $d(X_n) \leq 2B_n$.

- if $B_n = n+1$ (that is, if $A_n > n$), then $d(X_n) \le d(x) + 2n$.

$\square$

Corollary 8 *For all $x \in \mathbb{N}^p$ and $k \in \mathbb{N}$, there is a constant $\alpha_{p,k}$ such that, for all $n \ge 0$,*

$$\mathrm{E}^x d(X_n)^k \le d(x)^k + \alpha_{p,k}.$$

1.4 Invariance principle

The following proposition is the key result for our proof of the invariance principle. It gives asymptotic estimates of the even moments of F_n, that hold uniformly with respect to the initial value $X_0 = x$ of the chain. Theorem 1 follows, using the Markov property of X and the method of moments. We use the following definition in the statement of the proposition.

Definition 2 *For $h \in \mathbb{N}$, the function $(x,n) \mapsto \beta(x,n,h)$, defined on $n \in \mathbb{N}$ and $x \in \mathbb{N}^p$, satisfies the property (Q_h) if $\beta(x,n,h)$ is a sum of a finite number of terms of the form:*

$$\epsilon(x,n) m(x)^i n^j d(x)^k, \quad \textit{with } i,j,k \in \mathbb{N},\ 2j+i \le 2h,$$

where $\epsilon(x,n)$ is bounded and goes to zero as n goes to infinity, uniformly with respect to x.

Proposition 9 *For $h \in \mathbb{N}$, the following property holds, which we denote by (P_h). For $x \in \mathbb{N}^p$,*

$$\mathrm{E}^x F_n^{2h} = \sum_{i=0}^{h} f_i(h) m(x)^{2i} n^{h-i} + \alpha(x,n,h),$$

where $\alpha(x,n,h)$ satisfies (Q_h), and where the numbers $f_i(h) \in \mathbb{R}$ are uniquely determined by the following induction equations:

$$\begin{cases} f_h(h) = 1, \\ f_i(h) = \dfrac{2h(p-1) + h(2h-1)}{h-i} f_i(h-1), \quad 0 \le i \le h-1. \end{cases} \tag{1.1}$$

For notational convenience, we put :

$$f(x,n,h) = \sum_{i=0}^{h} f_i(h) m(x)^{2i} n^{h-i}.$$

Our proof of Proposition 9 is very technical, so we first sketch the idea. We use induction on h to prove that (P_h) holds for all h. The main step is to prove that, as $n \to +\infty$, the difference

$$\mathrm{E}F_{n+1}^{2h} - \mathrm{E}F_n^{2h} = \mathrm{E}(F_n + \Delta F_n)^{2h} - \mathrm{E}F_n^{2h} \tag{1.2}$$

involves only the expectations of the random variables F_r^{2q} with $r \leq n$ and $q \leq q-1$. To see this, we show that, expanding the $2h-$th power in (1.2), the leading terms when $n \to +\infty$ are those involving ΔF_n and $(\Delta F_n)^2$ only. Using the approximation of the transition probabilities of the selection step by uniform ones (see Lemma 7), we thus express (1.2) in terms of expectations of F_n^{2q} and of

$$\varphi_q(X_n) = (X_n^{(a)} - X_n^{(b)})^2 F_n^{2q},$$

for $q \leq h-1$. The asymptotic behavior of quantities of the last type stems from (P_h) and from an approximate contraction property of the selection operator which reads as follows:

$$\mathrm{E}\varphi_q(X_{n+1}) = \lambda \mathrm{E}\varphi_q(X_n) + \psi_h\left[\mathrm{E}(F_r^{2i})\,;\, r \leq n, i \leq h-1\right](1 + o(1)),$$

where $0 < \lambda < 1$ and ψ_q is explicit.

Proof of Proposition 9:

By induction on h. (P_0) holds trivially with $f(x, n, 0) = 1$. Suppose that (P_i) holds for all $i \leq h-1$, with $h \geq 1$. Then,

$$\mathrm{E}^x F_n^{2h} = m(x)^{2h} + \mathrm{E}^x \sum_{r=0}^{n-1} (F_{r+1}^{2h} - F_r^{2h}). \tag{1.3}$$

Expanding, we get:

$$\begin{aligned} \mathrm{E}^x(F_{r+1}^{2h} - F_r^{2h}) &= \mathrm{E}^x((F_r + \Delta F_r)^{2h} - F_r^{2h}) \\ &= \binom{2h}{1} \mathrm{E}^x\left(F_r^{2h-1}\mathrm{E}^x\left(\Delta F_r \mid \mathcal{F}_r\right)\right) \\ &\quad + \binom{2h}{2} \mathrm{E}^x\left(F_r^{2h-2}\mathrm{E}^x\left((\Delta F_r)^2 \mid \mathcal{F}_r\right)\right) \\ &\quad + \sum_{l=3}^{2h} \binom{2h}{l} \mathrm{E}^x\left(F_r^{2h-l}(\Delta F_r)^l\right). \end{aligned} \tag{1.4}$$

We start with the contribution of the rightmost sum in (1.4):
for $3 \leq l \leq 2h$,

$$\left|\mathrm{E}^x F_r^{2h-l}(\Delta F_r)^l\right| \leq \mathrm{E}^x(1 + F_r)^{2h-l}|\Delta F_r|^l. \tag{1.5}$$

Since $(1+F_r)^{2h-l} = (1+F_r)^{2h-l+1/2} \times (1+F_r)^{-1/2}$, Hölder's inequality yields:

$$\begin{aligned}\left|\mathrm{E}^x F_r^{2h-l}(\Delta F_r)^l\right| &\leq \left(\mathrm{E}^x\left((1+F_r)^{2h-l+1/2}\right)^\theta\right)^{1/\theta} \\ &\qquad \times \left(\mathrm{E}^x\left(\left((1+F_r)^{-1/2}\right)^{\theta'}(|\Delta F_r|)^{l\theta'}\right)\right)^{1/\theta'},\end{aligned}$$

where $\theta = (2h-2)/(2h-l+1/2) > 1$ and $\theta^{-1} + \theta'^{-1} = 1$. Since $l \geq 3$ and $|\Delta F_r| \leq d(F_r) + 2$:

$$\begin{aligned}\left|\mathrm{E}^x F_r^{2h-l}(\Delta F_r)^l\right| &\leq \left(\mathrm{E}^x(1+F_r)^{2h-2}\right)^{1/\theta} \\ &\qquad \times \left(\mathrm{E}^x\left(\left((1+F_r)^{-1/2}\right)^{\theta'}(d(X_r)+2)^{l\theta'}\right)\right)^{1/\theta'}. \quad (1.6)\end{aligned}$$

We first bound the rightmost factor in (1.6): from Schwarz's inequality,

$$\begin{aligned}&\mathrm{E}^x\left((1+F_r)^{-1/2}\right)^{\theta'}((d(X_r)+2))^{l\theta'} \\ &\qquad \leq \left[\mathrm{E}^x\left((1+F_r)^{-1/2}\right)^{2\theta'}\mathrm{E}^x\,(d(X_r)+2)^{2l\theta'}\right]^{1/2} \\ &\qquad \leq \left[\mathrm{E}^x(1+F_r)^{-\theta'}\right]^{1/2} c_{p,h}^{(1)}(1+d(x)^{8lh})^{1/2},\end{aligned}$$

according to Corollary 8, and since $2l\theta' \leq 8lh$). Furthermore:

$$\mathrm{E}^x(1+F_r)^{-l\theta'} \leq \mathrm{E}(1+m(Y_r))^{-l\theta'},$$

where (Y_n) is the process defined in Lemma 1. This last term goes to zero as n goes to infinity (and does not depend on x).

We now bound the leftmost factor in (1.6).

$$\left(\mathrm{E}^x(1+F_r)^{2h-2}\right)^{1/\theta} \leq \mathrm{E}^x(1+F_r)^{2h-2} = \sum_{i=0}^{2h-2}\binom{2h-2}{i}\mathrm{E}^x F_r^i.$$

Using the induction hypothesis $(P_0, \ldots, P_{h-1})$, we see that the last term of (1.4) satisfies (Q_{h-1}), hence its contribution in (1.3) — summing for $r = 0, \ldots, n-1$ — satisfies (Q_h).

Now we study of the first term in (1.4). From Lemma 6,

$$\binom{2h}{1}\mathrm{E}^x\left(F_r^{2h-1}\mathrm{E}^x\left(\Delta F_r \mid \mathcal{F}_r\right)\right) = 2hp^{-2}\mathrm{E}^x\sum_{1\leq i<j\leq p}F_r^{2h-2}(X_r^{(i)} - X_r^{(j)})^2 \quad (1.7)$$

For $0 \leq s \leq r-1$ and $i \neq j$,

$$\mathrm{E}^x F_{s+1}^{2h-2}(X_{s+1}^{(i)} - X_{s+1}^{(j)})^2 = \mathrm{E}^x(F_s + \Delta F_s)^{2h-2}(X_{s+1}^{(i)} - X_{s+1}^{(j)})^2,$$

and, expanding $(F_s+\Delta F_s)^{2h-2}$, we get:

$$
\begin{aligned}
\mathrm{E}^x F_{s+1}^{2h-2} & (X_{s+1}^{(i)}-X_{s+1}^{(j)})^2 \\
&= \sum_{l=0}^{2h-2} \binom{2h-2}{l} \mathrm{E}^x F_s^{2h-2-l} (\Delta F_s)^l (X_{s+1}^{(i)}-X_{s+1}^{(j)})^2 \\
&= \mathrm{E}^x F_s^{2h-2} \mathrm{E}^x \left((X_{s+1}^{(i)}-X_{s+1}^{(j)})^2 \;\middle|\; \mathcal{F}_s \right) \qquad (1.8)\\
&\quad \underbrace{+ \sum_{l=1}^{2h-2} \binom{2h-2}{l} \mathrm{E}^x F_s^{2h-2-l} (\Delta F_s)^l (X_{s+1}^{(i)}-X_{s+1}^{(j)})^2}_{(\star)} .
\end{aligned}
$$

Using the same method as for the last term in (1.4), we show that the sum $(\star)$ satisfies (Q_{h-1}).

Moreover, conditioning by $\mathcal{F}_s$ yields:

$$
\begin{aligned}
\mathrm{E}^x & \left((X_{s+1}^{(i)}-X_{s+1}^{(j)})^2 \;\middle|\; \mathcal{F}_s \right) \\
&= 2 + \mathrm{E}^x \left((Y_s^{(i)}-Y_s^{(j)})^2 \;\middle|\; \mathcal{F}_s \right) \\
&= 2 + \mathbf{1}_{\{X_s\neq 0\}} \sum_{1\le a,b\le p} \frac{X_s^{(a)} X_s^{(b)}}{S_s^2} (X_s^{(a)}-X_s^{(b)})^2. \qquad (1.9)
\end{aligned}
$$

If we replace in (1.9) the selection transition probabilities $X_s^{(a)}X_s^{(b)}/S_s^2$ by uniform selection transition probabilities p^{-2}, we get in (1.8) an error term that we bound using Lemma 7: for all $1\le a,b\le p$,

$$
\begin{aligned}
\mathbf{1}_{\{X_s\neq 0\}} & \left| F_s^{2h-2} \left(\frac{X_s^{(a)} X_s^{(b)}}{S_s^2} - \frac{1}{p^2} \right) (X_s^{(a)}-X_s^{(b)})^2 \right| \\
&\le 2p^{-1} d(X_s) F_s^{2h-3}.
\end{aligned}
$$

Hence, for $0\le s\le r-1$, equation (1.8) can be written

$$
\begin{aligned}
\mathrm{E}^x F_{s+1}^{2h-2} (X_{s+1}^{(i)}-X_{s+1}^{(j)})^2 &= 2\mathrm{E}^x F_s^{2h-2} + \frac{p-1}{p} \mathrm{E}^x F_s^{2h-2} (X_s^{(i)}-X_s^{(j)})^2 \\
&\quad + \beta_1(x,s,h-1),
\end{aligned}
$$

where $\beta_1(x,s,h-1)$ satisfies (Q_{h-1}). Summing these estimates for $s=0,\dots,r-1$, we get, for all $0\le r\le n-1$:

$$\begin{aligned}
\mathrm{E}^x F_r^{2h-2}(X_r^{(i)}-X_r^{(j)})^2 &= \left(\frac{p-1}{p}\right)^r (x_i-x_j)^2 m(x)^{2h-2}+\\
&\quad \sum_{k=0}^{r-1}\left(\frac{p-1}{p}\right)^{r-1-k}\Big(2\mathrm{E}^x F_k^{2h-2}+\beta_1(x,k,h-1)\Big)\\
&= \left(\frac{p-1}{p}\right)^r (x_i-x_j)^2 m(x)^{2h-2}+\\
&\quad \beta_2(x,r,h-1)+\sum_{k=0}^{r-1}\left(\frac{p-1}{p}\right)^{r-1-k} 2\mathrm{E}^x F_k^{2h-2},
\end{aligned}$$

where $\beta_2(x,r,h-1)$ satisfies (Q_{h-1}). Hence, according to (P_{h-1}):

$$\begin{aligned}
&\sum_{r=0}^{n-1}\mathrm{E}^x F_r^{2h-2}(X_r^{(i)}-X_r^{(j)})^2\\
&\quad = p\left(1-\left(\frac{p-1}{p}\right)^n\right)(x_i-x_j)^2 m(x)^{2h-2}\\
&\qquad +\sum_{r=0}^{n}\beta_2(x,r,h-1)\\
&\qquad +\sum_{r=0}^{n-1}\sum_{k=0}^{r-1}\left(\frac{p-1}{p}\right)^{r-1-k} 2\left(f(x,k,h-1)+\alpha(x,k,h-1)\right).
\end{aligned}$$

Here, the two first lines and the contribution of $\alpha(x,k,h-1)$ satisfy (Q_h). It remains

$$\begin{aligned}
\sum_{r=0}^{n-1}\mathrm{E}^x F_r^{2h-2}(X_r^{(i)}-X_r^{(j)})^2 &= 2\sum_{r=0}^{n-1}\sum_{k=0}^{r-1}\left(\frac{p-1}{p}\right)^{r-1-k} f(x,k,h-1)\\
&\quad +\beta_3(x,n,h),
\end{aligned}\tag{1.10}$$

where $\beta_3(x,r,h)$ satisfies (Q_h). We now consider the sum over k in (1.10). Using (P_{h-1}),

$$\begin{aligned}
&\sum_{k=0}^{r-1}\left(\frac{p-1}{p}\right)^{r-1-k} f(x,k,h-1)\\
&\quad = \sum_{k=0}^{r-1}\left(\frac{p-1}{p}\right)^{r-1-k}\sum_{i=0}^{h-1} f_i(h-1)m(x)^{2i}k^{h-i}\\
&\quad = \sum_{i=0}^{h-1} f_i(h-1)m(x)^{2i}\sum_{k=0}^{r-1}\left(\frac{p-1}{p}\right)^{r-1-k} k^{h-i}.
\end{aligned}$$

From Lemma 8, one has:

$$\sum_{k=0}^{r-1}\left(\frac{p-1}{p}\right)^{r-1-k} f(x,k,h-1)$$
$$= \quad p\sum_{i=0}^{h-1} f_i(h-1)m(x)^{2i}r^{h-i}\left[1+\epsilon_i(r)\right],$$

where the ϵ_i's are sequences such that $\epsilon_i(n)$ goes to zero as n goes to infinity. Using again (P_{h-1}), we get:

$$\sum_{k=0}^{r-1}\left(\frac{p-1}{p}\right)^{r-1-k} f(x,k,h-1) = pf(x,r,h-1)+\beta_4(x,r,h-1),$$

where $\beta_4(x,r,h-1)$ satisfies (Q_{h-1}). Plugging this estimate into (1.10) gives

$$\sum_{r=0}^{n-1}\mathrm{E}^x F_r^{2h-2}(X_r^{(i)}-X_r^{(j)})^2 = \beta_5(x,n,h)+2p\sum_{r=0}^{n-1} f(x,r,h-1), \tag{1.11}$$

where $\beta_5(x,r,h)$ satisfies (Q_h).

Plugging (1.11) into (1.7) gives:

$$\sum_{r=0}^{n}\binom{2h}{1}\mathrm{E}^x F_r^{2h-1}\mathrm{E}^x\left(\Delta F_r \mid \mathcal{F}_r\right)$$
$$= \quad \beta_6(x,n,h)+2h/p^2\sum_{1\le i<j\le p} 2p\sum_{r=0}^{n-1} f(x,r,h-1)$$
$$= \quad \beta_7(x,n,h)+2h(p-1)\sum_{r=0}^{n-1} f(x,r,h-1), \tag{1.12}$$

where $\beta_7(x,n,h)$ satisfies (Q_h).

We now study the second term in (1.4). From Lemma 6, using Lemma 7 to approximate the selection transition probabilities by uniform ones, as we did for the first term of (1.4), we get:

$$\begin{aligned}
&\mathrm{E}^x\ F_r^{2h-2}\mathrm{E}^x\left((\Delta F_r)^2 \;\middle|\; \mathcal{F}_r\right)\\
&\quad = \quad p^{-1}\mathrm{E}^x F_r^{2h-2}\\
&\qquad + p^{-4}\sum_{1\le i\ne j\le p}\ \sum_{1\le a,b\le p}\mathrm{E}^x F_r^{2h-2}(X_r^{(a)}-X_r^{(i)})(X_r^{(b)}-X_r^{(j)})\\
&\qquad + p^{-3}\sum_{1\le i\le p}\ \sum_{1\le a\le p}\mathrm{E}^x F_r^{2h-2}(X_r^{(a)}-X_r^{(i)})^2\\
&\qquad + \beta_8(x,r,h-1),
\end{aligned} \tag{1.13}$$

where $\beta_8(x,r,h-1)$ satisfies (Q_{h-1}).

For an arbitrary sequence $(x_i)_{1\leq i\leq p}$, one has

$$\begin{aligned}\sum_{1\leq i\neq j\leq p}\ \sum_{1\leq a,b\leq p}(x_a-x_i)(x_b-x_j)\\ &= \sum_{1\leq a,b,i,j\leq p}(x_a-x_i)(x_b-x_j)-\sum_{i=j,a,b}(x_a-x_i)(x_b-x_j)\end{aligned}$$

By symmetry, the first sum in the right hand side is zero. As for the last sum, we separate cases $a=b$ and $a\neq b$:

$$\begin{aligned}\sum_{1\leq i\neq j\leq p}\ \sum_{1\leq a,b\leq p}(x_a-x_i)(x_b-x_j)\\ &= -\sum_{i,a}(x_a-x_i)^2-\sum_{i,a\neq b}(x_a-x_i)(x_b-x_i)\end{aligned}$$

In the last sum, let (u,v,w) be the ordered triple with elements i,a,b. With this notation, the last equation reads

$$\begin{aligned}\sum_{1\leq i\neq j\leq p}\ \sum_{1\leq a,b\leq p}(x_a-x_i)(x_b-x_j)\\ &= -\sum_{i,a}(x_a-x_i)^2-2\sum_{1\leq u<v<w\leq p}\begin{array}{l}(x_v-x_u)(x_w-x_u)\\ +(x_u-x_v)(x_w-x_v)\\ +(x_u-x_w)(x_v-x_w).\end{array}\end{aligned}$$

Note that , for all x,y,z,

$$2\left[(y-x)(z-x)+(x-y)(z-y)+(x-z)(y-z)\right]=(x-y)^2+(y-z)^2+(x-z)^2,$$

so that:

$$\begin{aligned}\sum_{1\leq i\neq j\leq p}\ \sum_{1\leq a,b\leq p}(x_a-x_i)(x_b-x_j)\\ &= -\sum_{i,a}(x_a-x_i)^2-\sum_{1\leq u<v<w\leq p}\begin{array}{l}(x_u-x_v)^2\\ +(x_v-x_w)^2\\ +(x_u-x_w)^2\end{array}\end{aligned}$$

Applying this to the sequence $(X_r^{(k)})_k$, and plugging it and equation (1.11) into (1.13), we get that:

$$
\begin{aligned}
\sum_{r=0}^{n-1} \mathrm{E}^x \; & F_r^{2h-2} \mathrm{E}^x \left((\Delta F_r)^2 \mid \mathcal{F}_r \right) \\
&= \Big[p^{-1} \\
&\qquad - p^{-4}(p(p-1) + 3\binom{p}{3})2p \\
&\qquad + p^{-3}p(p-1)2p \Big] \sum_{r=0}^{n-1} f(x,r,h-1) \\
&\quad + \beta_9(x,n,h) \\
&= \beta_9(x,n,h) + \sum_{r=0}^{n-1} f(x,r,h-1), \qquad (1.14)
\end{aligned}
$$

where $\beta_9(x,n,h)$ satisfies (Q_h).

Finally, thanks to (1.4), (1.12) and (1.14):

$$
\mathrm{E}^x F_n^{2h} = m(x)^{2h} + [2h(p-1) + h(2h-1)] \sum_{r=0}^{n-1} f(x,r,h-1) + \beta_{10}(x,n,h),
$$

where $\beta_{10}(x,n,h)$ satisfies (Q_h). Hence, (P_h) holds, with:

$$
\begin{cases} f_h(h) = 1, \\ f_i(h) = \dfrac{2h(p-1) + h(2h-1)}{h-i} f_i(h-1), \quad 0 \le i \le h-1. \end{cases} \qquad (1.15)
$$

$\square$

Remark 10 *The details of the proof of Proposition 9 yield that $\alpha(x,n,h)$ contains no term of the form $\epsilon(x,n)m(x)^{2h}d(x)^k$.*

The numbers $f_i(h)$ in Proposition 1.1 are related to the $(2p-1)-$dimensional Bessel process $(R(t))_{0\le t\le 1}$ starting from zero. The law of $R(t)$ is as follows.

Lemma 3 *For $u \in \mathbb{R}_+$,*

$$
\mathrm{E}^u R(t)^{2h} = \sum_{i=0}^{h} f_i(h) u^{2i} t^{h-i}.
$$

Proof of Lemma 3:

Let $d = 2p - 1$ and $\nu = d/2 - 1 = p - 3/2$. Using the explicit expression of the density of the Bessel semi-group density (see [65, p. 415]) and formulas 6.631 and 9.212 of [38], we get that:

$$\begin{aligned}
\mathrm{E}^u R(t)^{2h} &= \int_0^\infty t^{-1}\left(\frac{y}{u}\right)^\nu y^{2h+1}\exp\left(-\frac{u^2+y^2}{2t}\right) I_\nu(uy/t)\ dy \\
&= t^{-1}u^{-\nu} exp\left(-\frac{u^2}{2t}\right)\frac{(u/t)^\nu\,\Gamma(h+\nu+1)}{2^{\nu+1}(2t)^{-(h+\nu+1)}\Gamma(\nu+1)} \\
&\quad \times\Phi\left(h+\nu+1, \nu+1; \frac{u^2}{2t}\right) \\
&= (2t)^h\ \frac{\Gamma(h+\nu+1)}{\Gamma(\nu+1)}\ \Phi\left(-h, \nu+1; -\frac{u^2}{2t}\right) \\
&= \sum_{i=0}^h g_i(h)u^{2i}t^{h-i},
\end{aligned}$$

where, by the definition of $\Phi = {}_1F_1$,

$$g_0(h) = d\times(d+2)\times\ldots\times(d+2(h-1)),$$

and, for $1 \le i \le h$,

$$\begin{aligned}
g_i(h) &= g_0(h)\frac{h\times(h-1)\times\cdots\times(h-i+1)}{(\nu+1)\times(\nu+2)\times\cdots\times(\nu+1+i-1)}\frac{1}{2^i i!} \\
&= g_0(h)\frac{h\times(h-1)\times\cdots\times(h-i+1)}{d\times(d+2)\times\cdots\times(d+2(i-1))}(i!)^{-1} \\
&= \binom{h}{i}\prod_{i\le j\le h-1}(d+2j).
\end{aligned}$$

Since the $g_i(h)$ satisfy (1.1), we are done. □

The next two propositions study the finite-dimensional marginals of Z_n.

Proposition 11 *The finite dimensional distributions of the process Z_n converge to those of R.*

Proof of Proposition 11:

We prove by induction on m that, for $t = (t_i)_i$, t_i nonnegative real numbers with $t_i < t_{i+1}$ and for $h = (h_1, \ldots, h_m)$, h_i positive integers, as n goes to infinity,

$$\mathrm{E}(Z_n(t_1)^{2h_1}\times\cdots\times Z_n(t_m)^{2h_m}) \longrightarrow \mathrm{E}(R(t_1)^{2h_1}\times\cdots\times R(t_m)^{2h_m}).$$

For $m = 1$, $\mathrm{E}Z_n(t)^h = n^{-h}\mathrm{E}F^{2h}_{\lfloor nt\rfloor}$, that is

$$\mathrm{E}Z_n(t)^h = n^{-h}(f+\alpha)(0, \lfloor nt\rfloor, h) = t^h f_0(h) + o(1),$$

according to Proposition 9 and Corollary 8. According to Lemma 3, $\mathrm{E}Z_n(t)^h$ converges to $\mathrm{E}R(t)^h$ as n goes to infinity. Assume now that the result holds for $i \leq m$, where $m \geq 1$. For $x \in \mathbb{N}^p$, r and $i \in \mathbb{N}$, set $\phi(x,r,i) = \mathrm{E}^x(F_r^{2i})$. Fix $h = (h_1, \ldots, h_{m+1})$ and set

$$F(n,m) = F_{\lfloor nt_1 \rfloor}^{2h_1} \times \cdots \times F_{\lfloor nt_m \rfloor}^{2h_m} n^{-(h_1+\cdots+h_m)}.$$

Then,

$$\begin{aligned}
&\mathrm{E}(Z_n(t_1)^{2h_1} \times \cdots \times Z_n(t_{m+1})^{2h_{m+1}}) \\
&\quad = \mathrm{E}(F(n,m) \times F_{\lfloor nt_{m+1} \rfloor}^{2h_{m+1}}) n^{-h_{m+1}} \\
&\quad = \mathrm{E}\Big(F(n,m)\phi(X_{\lfloor nt_m \rfloor}, \lfloor nt_{m+1} \rfloor - \lfloor nt_m \rfloor, h_{m+1}) n^{-h_{m+1}}\Big) \\
&\quad = \mathrm{E}\Big(F(n,m)(f+\alpha)(X_{\lfloor nt_m \rfloor}, \lfloor nt_{m+1} \rfloor - \lfloor nt_m \rfloor, h_{m+1}) n^{-h_{m+1}}\Big) \\
&\quad = A_n + B_n,
\end{aligned}$$

say, where A_n and B_n are the contributions of f and α respectively. According to Proposition 9, the error term B_n is a sum of terms of the form

$$\begin{aligned}
b_{n,i,j} &= \mathrm{E}F(n,m)\epsilon(X_{\lfloor nt_m \rfloor}, \lfloor nt_{m+1} \rfloor - \lfloor nt_m \rfloor) \\
&\quad \times F_{\lfloor nt_m \rfloor}^i (\lfloor nt_{m+1} \rfloor - \lfloor nt_m \rfloor)^j d(X_{\lfloor nt_m \rfloor})^k n^{-h_{m+1}},
\end{aligned}$$

where $\epsilon(u,r)$ goes to zero as r goes to infinity uniformly with respect to u, and $2j + i \leq 2h_{m+1}$. According to Hölder's inequality,

$$\begin{aligned}
b_{n,i,j} &\leq \epsilon(\lfloor nt_{m+1} \rfloor - \lfloor nt_m \rfloor) \left(\mathrm{E}F(n,m)^2\right)^{1/2} \\
&\quad \times (\lfloor nt_{m+1} \rfloor - \lfloor nt_m \rfloor)^j n^{i/2 - h_{m+1}} \\
&\quad \times \left(n^{-2i}\mathrm{E}F_{\lfloor nt_m \rfloor}^{4i}\right)^{1/4} \left(\mathrm{E}d(X_{\lfloor nt_m \rfloor})^{4k}\right)^{1/4},
\end{aligned}$$

where $\epsilon(r)$ goes to zero as r goes to infinity.

By the induction hypothesis, $\mathrm{E}F(n,m)^2$ and $n^{-2i}\mathrm{E}F_{\lfloor nt_m \rfloor}^{4i}$ converge. Moreover, according to Corollary 8, $\mathrm{E}d(X_{\lfloor nt_m \rfloor})^{4k}$ is bounded. Hence, each $b_{n,i,j}$ goes to zero as n goes to infinity.

As regards the main term A_n,

$$\begin{aligned}
A_n &= \sum_{i=0}^{h_{m+1}} \mathrm{E}F(n,m) f_i(h_{m+1}) F_{\lfloor nt_m \rfloor}^{2i} (\lfloor nt_{m+1} \rfloor - \lfloor nt_m \rfloor)^{h_{m+1}-i} n^{-h_{m+1}} \\
&= \sum_{i=0}^{h_{m+1}} f_i(h_{m+1}) \mathrm{E}F_{\lfloor nt_1 \rfloor}^{2h_1} \times \cdots \times F_{\lfloor nt_{m-1} \rfloor}^{2h_{m-1}} \times F_{\lfloor nt_m \rfloor}^{2(h_m+i)} n^{-(h_1+\cdots+h_m+i)} \\
&\quad \times (t_{m+1} - t_m)^{h_{m+1}-i}(1 + o(1)).
\end{aligned}$$

Hence, according to the induction hypothesis, A_n converges to

$$\begin{aligned} A_\infty &= \sum_{i=0}^{h_{m+1}} f_i(h_{m+1}) \mathrm{E} R(t_1)^{2h_1} \times \cdots \times R(t_{m-1}^{2h_{m-1}}) \\ &\qquad \times R(t_m)^{2(h_m+i)} (t_{m+1} - t_m)^{h_{m+1}-i} \\ &= \mathrm{E}\Bigg(R(t_1)^{2h_1} \times \cdots \times R(t_m^{2h_m}) \\ &\qquad \sum_{i=0}^{h_{m+1}} f_i(h_{m+1}) R(t_m)^{2i} (t_{m+1} - t_m)^{h_{m+1}-i} \Bigg). \end{aligned}$$

According to Lemma 3,

$$\begin{aligned} A_\infty &= \mathrm{E} R(t_1)^{2h_1} \times \cdots \times R(t_m^{2h_m}) \mathrm{E}^{R(t_m)} R(t_{m+1} - t_m)^{2h_{m+1}} \\ &= \mathrm{E} R(t_1)^{2h_1} \times \cdots \times R(t_{m+1})^{2h_{m+1}}, \end{aligned}$$

and we are done with the proof of the induction.

Now, the method of moments entails that, for all $m \geq 1$, the following convergence in law holds:

$$(\epsilon_1 F_n(t_1), \ldots, \epsilon_m F_n(t_m)) \xrightarrow{\mathcal{L}} (\epsilon_1 R(t_1), \ldots, \epsilon_m R(t_m)),$$

as n goes to infinity, and where the $\epsilon_i \in \{-1, +1\}$ are i.i.d. random variables, independent of F_n and R, following a symmetric ± 1 Bernoulli law. The limiting law is uniquely determined by its moments since, for all $0 \leq t \leq 1$ and $\lambda \in \mathbb{R}_+$, the moment generating function $\mathrm{E}(e^{\lambda R_t})$ is finite-valued. The random variables $F_n(t_i)$ and $R(t_i)$ being nonnegative, this implies finally that:

$$(F_n(t_1), \ldots, F_n(t_m)) \xrightarrow{\mathcal{L}} (R(t_1), \ldots, R(t_m)).$$

□

Proposition 12 *There exists a constant C_p such that, for all $0 \leq t_1 \leq t \leq t_2 \leq 1$,*

$$\mathrm{E} \left(Z_n(t) - Z_n(t_1) \right)^2 \left(Z_n(t) - Z_n(t_2) \right)^2 \leq C_p (t_2 - t_1)^{3/2}.$$

Proof of Proposition 12:

For $x \in \mathbb{N}^p$ and $r, r_1, r_2 \in \mathbb{N}$, we define:

$$\begin{aligned} \psi_1(x, r) &= \mathrm{E}^x((F_r - F_0)^2) \\ \psi_2(x, r_1, r_2) &= \mathrm{E}^x((F_{r_1} - F_0)^2 \psi_1(X_{r_1}, r_2)) \end{aligned}$$

According to the Markov property, for all integers $a < b < c$, we have:

$$\mathrm{E}\left(F_b - F_a\right)^2\left(F_b - F_c\right)^2 = \mathrm{E}\left(\psi_2(X_a, b-a, c-b)\right). \tag{1.16}$$

Since $(s_1 - s_2)^2 \leq |s_1^2 - s_2^2|$ for nonnegative s_1 and s_2,

$$\psi_1(x,r) = \mathrm{E}^x(F_r - F_0)^2 \leq \mathrm{E}^x|F_r^2 - F_0^2|.$$

The estimates of Proposition 9 and Remark 10 lead to the inequality:

$$\mathrm{E}^x|F_r^2 - F_0^2| \leq c_p^{(2)}(1 + d(x)^2)r.$$

Hence:

$$\psi_2(x, r_1, r_2) \leq \mathrm{E}^x\left((F_{r_1} - F_0)^2 c_p^{(2)}(1 + d(X_{r_1})^2)r_2\right).$$

According to Hölder's inequality, we have:

$$\psi_2(x, r_1, r_2) \leq c_p^{(2)} r_2 \left(\mathrm{E}^x(F_{r_1} - F_0)^4\right)^{1/2}\left(\mathrm{E}^x(1 + d(X_{r_1})^2)^2\right)^{1/2} \tag{1.17}$$

From Corollary 8, $\mathrm{E}^x(1 + d(X_{r_1})^2)^2 \leq c_p^{(3)}(1 + d(x)^4)$. Using once again the estimates of Proposition 9 and Remark 10, we get that:

$$\mathrm{E}^x(F_r - F_0)^4 \leq c_p^{(4)}(m(x)^2 r + r^2)(1 + d(x)^4).$$

Using this bound in (1.17), we obtain:

$$\psi_2(x, r_1, r_2) \leq c_p^{(5)} r_2\left(m(x)^2 r_1 + r_1^2(1 + d(x)^4)\right)^{1/2}.$$

Hence, from (1.16) :

$$\mathrm{E}\left(F_b - F_a\right)^2\left(F_b - F_c\right)^2 \leq (c-b)\left[\left(F_a^2(b-a) + (b-a)^2\right)\left(1 + d(X_a)^4\right)\right]^{1/2}.$$

Setting $a = \lfloor t_1 n \rfloor$, $b = \lfloor tn \rfloor$, $c = \lfloor t_2 n \rfloor$, we get:

$$\begin{aligned}
&\mathrm{E}\left(Z_n(t) - Z_n(t_1)\right)^2\left(Z_n(t) - Z_n(t_2)\right)^2 \\
&\qquad \leq \quad c_p^{(5)} n^{-2}(\lfloor t_2 n \rfloor - \lfloor tn \rfloor) \\
&\qquad\qquad \Big(\mathrm{E}\left[m(X_{\lfloor t_1 n \rfloor})^2(\lfloor tn \rfloor - \lfloor t_1 n \rfloor) + (\lfloor tn \rfloor - \lfloor t_1 n \rfloor)^2\right] \\
&\qquad\qquad\qquad (1 + d(X_{\lfloor t_1 n \rfloor})^4)\Big)^{1/2}
\end{aligned}$$

Using Corollary 8 and Proposition 9, one can see that

$$\mathrm{E} n^{-1}(m(X_{\lfloor t_1 n \rfloor})^2(1 + d(X_{\lfloor t_1 n \rfloor})^4))$$

is bounded from above by a constant since $t_1 \in [0,1]$. We deduce that

$$\mathrm{E}\left(Z_n(t) - Z_n(t_1)\right)^2\left(Z_n(t) - Z_n(t_2)\right)^2 \leq c_p^{(5)}(t - t_1)\sqrt{t_2 - t}.$$

□

Proof of Theorem 1:

The convergence in law of $(Z_n)_{n\geq 0}$ to R is a consequence of the well-known Theorem 15.6 of [12], and of our Propositions 11 and 12.
Furthermore, according to Lemma 2, the p-dimensional process

$$n^{-1/2}\left(X^{(i)}_{\lfloor n\cdot\rfloor} - X^{(1)}_{\lfloor n\cdot\rfloor}\right)_{1\leq i\leq p}$$

goes to zero in probability as n goes to infinity for the uniform distance on $[0,1]$, hence for the Skorohod topology as well. □

1.5 Large deviations principle

We study the Laplace transform

$$\mathrm{E}\exp\left(tX^{(1)}_n\right)$$

Here again, the idea of the proof is to write recursion relations in which we approximate the transition probabilities of the selection steps by uniform ones. By a Gärtner-Ellis-type theorem (see [25]), the theorem is a consequence of the following result.

Proposition 13 *For $t \geq 0$,*

$$\frac{1}{n}\log \mathrm{E}\exp(tX^{(1)}_n) \xrightarrow[n\to+\infty]{} \log\cosh(t). \tag{1.18}$$

Let us sketch the idea of the proof. One has

$$\begin{aligned}\mathrm{E}\left(e^{t(X^{(1)}_{n+1})} \;\middle|\; \mathcal{F}_n\right) &= \mathbf{1}_{\{X_n\neq 0\}}\sum_{1\leq u\leq p}\frac{X^{(u)}_n}{S_n}e^{t(X^{(u)}_n)}\cosh(t)\\ &\quad + \mathbf{1}_{\{X_n=0\}}e^t\cosh(t).\end{aligned}$$

Hence:

$$\mathrm{E}(e^{t(X^{(1)}_{n+1})}) = \cosh(t)\mathrm{E}(e^{t(X^{(1)}_n)})(1+m_n(t)), \tag{1.19}$$

with

$$\begin{aligned}m_n(t) &= \frac{1}{\mathrm{E}(e^{t(X^{(1)}_n)})}\mathrm{E}\left(\mathbf{1}_{\{X_n\neq 0\}}\sum_{1\leq u\leq p}\frac{X^{(1)}_n - X^{(u)}_n}{S_n}e^{t(X^{(1)}_n)}\right)\\ &\quad + \frac{1}{\mathrm{E}(e^{t(X^{(1)}_n)})}\mathrm{P}\left[X_n=0\right](e^t-1).\end{aligned} \tag{1.20}$$

We want to prove that $m_n(t) \longrightarrow 0$ as $n \to \infty$. We first note that, in the definition of $m_n(t)$, the last term

$$\frac{1}{\mathrm{E}(e^{t(X_n^{(1)})})} \mathrm{P}\left[X_n = 0\right](e^t - 1)$$

goes to zero since $\mathrm{P}\left[X_n = 0\right](e^t - 1)$ is bounded (t is fixed here!), and since the denominator $\mathrm{E}(e^{t(X_n^{(1)})}) \to \infty$ as $n \to \infty$ (remember that $X_n^{(1)} \to \infty$ in probability). So we have to bound the first term. The inequality

$$\frac{\left|X_n^{(1)} - X_n^{(u)}\right|}{S_n} \leq \frac{d(X_n)}{pF_n},$$

yields

$$\begin{aligned} &\mathrm{E}\left(\mathbf{1}_{\{X_n \neq 0\}} \sum_{1 \leq u \leq p} \frac{X_n^{(1)} - X_n^{(u)}}{S_n} e^{t(X_n^{(1)})}\right) \\ &\qquad \leq \quad \mathrm{E}\left(\frac{d(X_n)}{pF_n} \mathbf{1}_{\{X_n \neq 0\}} \exp(t \max_i X_n^{(i)})\right). \end{aligned}$$

Hence Proposition 13 follows from the following result (stated below as Lemma 5): as $n \to \infty$,

$$\mathrm{E}\left(\frac{d(X_n)}{F_n} \mathbf{1}_{\{X_n \neq 0\}} \exp(t \max_i X_n^{(i)})\right) = o\left(\mathrm{E}\left(\exp(tX_n^{(1)})\right)\right).$$

We know that $d(X_n)$ is typically small when compared to F_n, however, since we are working on rare events here, we have to prove that, conditional on F_n taking values of order n, $d(X_n)/F_n$ is still small with large probability. The idea of the proof is the following: we condition the chain by its position at a time $n' \leq n$, where n' is such that the probability that there is a crunch time between n' and n, is high, and such that, on this event, the maximum possible diameter at time n is small, when compared to F_n. The key observation here, that we used already in the proof of Lemma 2, is that a non-negligible piece of this event can be made independent of the mutation process.

We first need a few technical results.

Lemma 4 below means that $\mathrm{E}\left(e^{t(X_n^{(1)})}\right)$ grows from step to step at least as fast as it should, according to Proposition 13 we want to prove. Corollary 14 states a similar result for $\mathrm{E}(e^{t \max_i X_n^{(i)}})$.

Lemma 4 *For all $n \geq 0$ and $t \geq 0$, $m_n(t) \geq 0$.*

Proof of Lemma 4:

Obviously, the last term in (1.20) is non-negative since $t \geq 0$. As regards the first term, by symmetry,

$$\begin{aligned} 2\mathrm{E}\mathbf{1}_{\{X_n \neq 0\}} S_n^{-1} \left(X_n^{(1)} - X_n^{(u)} \right) e^{t(X_n^{(1)})} \\ &= \mathrm{E}\mathbf{1}_{\{X_n \neq 0\}} S_n^{-1} \left(X_n^{(1)} - X_n^{(u)} \right) e^{t(X_n^{(1)})} \\ &\quad + \mathrm{E}\mathbf{1}_{\{X_n \neq 0\}} S_n^{-1} \left(X_n^{(u)} - X_n^{(1)} \right) e^{t(X_n^{(u)})} \\ &= \mathrm{E}\mathbf{1}_{\{X_n \neq 0\}} S_n^{-1} \left(X_n^{(1)} - X_n^{(u)} \right) (e^{t(X_n^{(1)})} - e^{t(X_n^{(u)})}), \end{aligned}$$

which is nonnegative since $y \mapsto e^y$ is an increasing function.

□

Corollary 14 *For all $t \geq 0$ and $n \geq q$,*

$$\mathrm{E}\left(\exp(t \max_i X_{n-q}^{(i)}) \right) \leq p(\cosh(t))^{-q} \mathrm{E}\left(\exp(t \max_i X_n^{(i)}) \right)$$

Proof of Corollary 14:

By Lemma 4,

$$\mathrm{E} \exp\left(t X_n^{(1)} \right) \geq \cosh^q(t) \mathrm{E} \exp\left(t X_{n-q}^{(1)} \right).$$

Now,

$$\begin{aligned} \mathrm{E}\left(\exp(t \max_i X_{n-q}^{(i)}) \right) &\leq \sum_{1 \leq i \leq p} \mathrm{E} \exp(t X_{n-q}^{(i)}) \\ &\leq p\mathrm{E}\left(\exp(t X_{n-q}^{(1)}) \right) \\ &\leq p(\cosh(t))^{-q} \mathrm{E}\left(\exp(t X_n^{(1)}) \right) \\ &\leq p(\cosh(t))^{-q} \mathrm{E}\left(\exp(t \max_i X_n^{(i)}) \right). \end{aligned}$$

□

We now prove the key result:

Lemma 5 *As n goes to infinity,*

$$\mathrm{E}\left(\frac{d(X_n)}{F_n} \mathbf{1}_{\{X_n \neq 0\}} \exp(t \max_i X_n^{(i)}) \right) = o\left(\mathrm{E}\left(\exp(t X_n^{(1)}) \right) \right) \tag{1.21}$$

Proof:

Let $(l_n)_n$ and $(u_n)_n$ be two increasing sequences of integers.

We first show that only large values of F_n contribute to the l.h.s. of (1.21).

Using the upper bound $d(X_n) \leq pF_n$,

$$\begin{aligned} \mathrm{E}\left(\frac{d(X_n)}{F_n}\mathbf{1}_{\{X_n\neq 0\}}\exp(t\max_i X_n^{(i)})\mathbf{1}_{\left\{\max_i X_n^{(i)}\leq u_n\right\}}\right) \\ \leq\ \mathrm{E}\left(p\exp(tu_n)\mathbf{1}_{\left\{\max_i X_n^{(i)}\leq u_n\right\}}\right) \\ \leq\ p\exp(tu_n) \end{aligned} \tag{1.22}$$

Assume that the selection steps are built through the sequence of random parameters U_m as defined in the proof of Lemma 2. Let

$$Q_n = \bigcup_{m=n-l_n}^{n-1}\{U_m \leq p^{1-p}\},$$

so that on Q_n, $d(X_n) \leq 2l_n$ since there is a crunch time between $n-l_n$ and n. Hence,

$$\begin{aligned} \mathrm{E}\left(\frac{d(X_n)}{F_n}\mathbf{1}_{\{X_n\neq 0\}}\exp(t\max_i X_n^{(i)})\mathbf{1}_{\left\{\max_i X_n^{(i)}\geq u_n\right\}}\mathbf{1}_{Q_n}\right) \\ \leq\ \frac{2l_n}{u_n/p}\mathrm{E}\left(\exp(t\max_i X_n^{(i)})\mathbf{1}_{Q_n}\right) \\ \leq\ \frac{2pl_n}{u_n}\sum_{i=1}^{p}\mathrm{E}\left(\exp(tX_n^{(i)})\mathbf{1}_{Q_n}\right) \end{aligned} \tag{1.23}$$

We introduce more notations concerning the one-step transitions of (X_n).

- **selection step:** $X_m \longrightarrow Y_m$

  $Y_m^{(i)} = X_m^{(f_m(i))}$, where the f_m's are random functions on $\{1,\ldots,p\}$ such that, knowing $\mathcal{F}_m$, the $(f_m(i))_{1\leq i\leq p}$ are i.i.d. with law

  $$\mathbf{1}_{\{X_n\neq 0\}}S_m^{-1}\sum_{j=1}^{p}X_m^{(j)}\delta_j + \mathbf{1}_{\{X_n=0\}}\delta_{(1,\ldots,1)}.$$

  Obviously, f_m and the random parameter U_m introduced above are not independent from each other.

- **mutation step:** $Y_m \longrightarrow X_{m+1}$

  $X_{m+1} = Y_m + \eta_m$, where $\eta_m = (\eta_m^{(1)},\ldots,\eta_m^{(p)})$ and the random variables $\eta_m^{(i)} \in \{-1,+1\}$, $1\leq i\leq p$ are independent from each other and independent from $\mathcal{G}_m$, with symmetric Bernoulli law.

With these notations, let, for $n-l_n \leq m \leq n$,

$$g_m = f_n \circ \cdots \circ f_m.$$

Thus, for all $j \in \{1,\ldots,p\}$:

$$\begin{aligned}
&\mathrm{E}\left(\exp(tX_n^{(j)})\mathbf{1}_{Q_n} \;\middle|\; \mathcal{F}_{n-l_n}\right) \\
&\quad = \mathrm{E}\left(\exp\left(t\left[X_{n-l_n}^{(g_{n-l_n}(j))} + \sum_{m=n-l_n}^{n-1} \eta_m^{(g_m(j))}\right]\right)\mathbf{1}_{Q_n} \;\middle|\; \mathcal{F}_{n-l_n}\right) \\
&\quad \leq \exp(t\max_i X_{n-l_n}^{(i)}) \\
&\quad\quad \times \mathrm{E}\left(\exp\left(t\sum_{m=n-l_n}^{n-1} \eta_m^{(g_m(j))}\right)\mathbf{1}_{Q_n} \;\middle|\; \mathcal{F}_{n-l_n}\right) \qquad (1.24)
\end{aligned}$$

Knowing $(U_{n-l_n},\ldots,U_{n-1})$, g_m depends only on η_r for $r<m$. Since the η_m's are i.i.d. Bernoulli random vectors independent of the U_m's, the conditional law knowing $(U_{n-l_n},\ldots,U_{n-1})$ of

$$\sum_{m=n-l_n}^{n-1} \eta_m^{(g_m(j))}$$

knowing $(U_{n-l_n},\ldots,U_{n-1})$ is the law of a sum of l_n symmetric Bernoulli ± 1 independent random variables. Hence,

$$\mathrm{E}\left(\exp\left(t\sum_{m=n-l_n}^{n-1} \eta_m^{(g_m(j))}\right)\mathbf{1}_{Q_n} \;\middle|\; \mathcal{F}_{n-l_n}\right) = (\cosh(t))^{l_n} \times \mathrm{P}\left(Q_n \mid \mathcal{F}_{n-l_n}\right).$$

Plugging this result into (1.24) gives

$$\mathrm{E}\left(\exp(tX_n^{(j)})\mathbf{1}_{Q_n}\right) \leq \mathrm{P}(Q_n)(\cosh(t))^{l_n}\mathrm{E}\exp(t\max_i X_{n-l_n}^{(i)})$$

Using Lemma 4, we get:

$$\mathrm{E}\left(\exp(tX_n^{(j)})\mathbf{1}_{Q_n}\right) \leq p\mathrm{E}\exp(t\max_i X_n^{(i)}). \qquad (1.25)$$

We now study Q_n^c:

$$\begin{aligned}
&\mathrm{E}\left(\frac{d(X_n)}{F_n}\mathbf{1}_{\{X_n\neq 0\}}\exp(t\max_i X_n^{(i)})\mathbf{1}_{\left\{\max_i X_n^{(i)}\geq u_n\right\}}\mathbf{1}_{Q_n^c}\right) \\
&\quad \leq p\mathrm{E}\left(\exp(t\max_i X_n^{(i)})\mathbf{1}_{Q_n^c}\right) \\
&\quad \leq p\sum_{j=1}^{p}\mathrm{E}\left(\exp(tX_n^{(j)})\mathbf{1}_{Q_n^c}\right). \qquad (1.26)
\end{aligned}$$

Using the same arguments as before, we get:

$$\mathrm{E}\left(\exp(tX_n^{(j)})\mathbf{1}_{Q_n^c}\right) \le \mathrm{P}(Q_n^c)(\cosh(t))^{l_n}\,\mathrm{E}\exp(t\max_i X_{n-l_n}^{(i)}).$$

Using Lemma 4 and the definition of Q_n, we get:

$$\begin{aligned}
\mathrm{E}\left(\exp(tX_n^{(j)})\mathbf{1}_{Q_n^c}\right) &\le (1-p^{1-p})^{l_n}(\cosh(t))^{l_n} \\
&\quad \times p(\cosh(t))^{-l_n}\,\mathrm{E}\exp(t\max_i X_n^{(i)}) \\
&\le p(1-p^{1-p})^{l_n}\,\mathrm{E}\exp(t\max_i X_n^{(i)}) \qquad (1.27)
\end{aligned}$$

Plugging (1.25) into (1.23) and (1.27) into (1.26), together with (1.22), we get:

$$\begin{aligned}
&\mathrm{E}\left(\frac{d(X_n)}{F_n}\mathbf{1}_{\{X_n\neq 0\}}\exp(t\max_i X_n^{(i)})\right) \qquad (1.28)\\
&\quad\le p\exp(tu_n) + \left(\frac{2p^3 l_n}{u_n} + p^3(1-p^{1-p})^{l_n}\right)\mathrm{E}\exp(t\max_i X_n^{(i)}).
\end{aligned}$$

We note that, thanks to Lemma 4,

$$\mathrm{E}\exp(tX_n^{(1)}) \ge (\cosh(t))^n.$$

Hence, taking $u_n = o(n)$ implies that

$$\exp(tu_n) = o\left(\mathrm{E}\exp(tX_n^{(1)})\right).$$

Moreover, using the inequality

$$\mathrm{E}\exp(t\max_i X_n^{(i)}) \le p\mathrm{E}\exp(tX_n^{(1)})$$

we see that it is enough to take $l_n = o(u_n)$ and $l_n \to +\infty$ to get the result thanks to the estimate (1.28).

□

1.6 Convergence of the cloud

For $p \in \mathbb{N}^*$, let

$$\mathbb{M}_p = \{(x_1,\ldots,x_p) \in \mathbb{Z}^p;\, \forall i,j,\ x_i - x_j \in 2\mathbb{Z}\},$$

and $\mathbb{H}_p$ be the quotient group $\mathbb{M}_p/\langle(1,\ldots,1)\rangle$: two elements of $\mathbb{M}_p$ are identified whenever they differ by an integer multiple of $(1,\ldots,1)$. Let $\Theta : \mathbb{M}_p \longrightarrow \mathbb{H}_p$ denote the canonical projection.

Let $(\Xi_n)_{n\geq 0}$ be the selection-mutation chain on $\mathbb{Z}^p$ with uniform selection and mutation ± 1, starting from zero. Then, $\xi = (\Theta(\Xi_n))_n$ is also a Markov chain on $\mathbb{H}_p$, irreducible, aperiodic and positively recurrent. Indeed,

$$\mathrm{P}(\xi_n = 0 \mid \xi_{n-1}) \geq 2^{-p} p^{1-p} \quad \text{a.s.}$$

Hence, ξ_n converges in law towards its invariant distribution π_p. We now prove Theorem 3, that is, that $\Theta(X_n)$ converges in law to π_p as well.

Proof of Theorem 3:

From Lévy's Theorem (cf. [29]), it is enough to prove that

$$\begin{aligned} D_n(t) &= \mathrm{E}\exp\left(i(t_1 X_n^{(1)} + \cdots + t_p X_n^{(p)})\right) \\ &\quad - \mathrm{E}\exp\left(i(t_1 \Xi_n^{(1)} + \cdots + t_p \Xi_n^{(p)})\right) \end{aligned}$$

goes to zero for every p-tuple $t = (t_1, \ldots, t_p)$ of real numbers summing to zero. We now consider the effects of mutation and selection over D_n. The mutation steps of X and Ξ multiply $D_n(t)$ by $\cos(t_1) \times \cdots \times \cos(t_p)$, hence, taking into account the selection steps as well, one gets

$$\begin{aligned} D_{n+1}(t) &= \mathrm{E}\mathbf{1}_{\{X_n=0\}} e^{i(t_1+\cdots+t_p)} \cos(t_1) \times \cdots \times \cos(t_p) \\ &\quad + \mathrm{E} \sum_{1\leq j_1,\ldots,j_p\leq p} \cos(t_1) \times \cdots \times \cos(t_p) \\ &\quad \times \left[\mathbf{1}_{\{X_n\neq 0\}} \left(\prod_{k=1}^{p} \frac{X_n^{(j_k)}}{S_n} \right) \exp\left(i \sum_{k=1}^{p} t_k X_n^{(j_k)} \right) \right. \\ &\quad \left. - \frac{1}{p^p} \exp\left(i \sum_{k=1}^{p} t_k \Xi_n^{(j_k)} \right) \right]. \end{aligned}$$

To use the approximation of the selection transition probabilities of X by those of a uniform selection, we add and substract p^{-p} to $\prod_k X_n^{(j_k)}/S_n$:

$$\begin{aligned} D_{n+1}(t) &= \mathrm{E}\mathbf{1}_{\{X_n=0\}} e^{i(t_1+\cdots+t_p)} \cos(t_1) \times \cdots \times \cos(t_p) \\ &\quad + \mathrm{E} \sum_{1\leq j_1,\ldots,j_p\leq p} \cos(t_1) \times \cdots \times \cos(t_p) \mathbf{1}_{\{X_n\neq 0\}} \\ &\quad \times \left(\left(\prod_{k=1}^{p} \frac{X_n^{(j_k)}}{S_n} \right) - \frac{1}{p^p} \right) \exp\left(i \sum_{k=1}^{p} t_k X_n^{(j_k)} \right) \\ &\quad + \mathrm{E} \sum_{1\leq j_1,\ldots,j_p\leq p} \cos(t_1) \times \cdots \times \cos(t_p) \\ &\quad \times \frac{1}{p^p} \left[\mathbf{1}_{\{X_n\neq 0\}} \exp\left(i \sum_{k=1}^{p} t_k X_n^{(j_k)} \right) - \exp\left(i \sum_{k=1}^{p} t_k \Xi_n^{(j_k)} \right) \right] \end{aligned}$$

Replacing $\mathbf{1}_{\{X_n\neq 0\}}$ by $1-\mathbf{1}_{\{X_n=0\}}$, we finally obtain:

$$\begin{aligned}
D_{n+1}(t) &= \mathrm{E}\mathbf{1}_{\{X_n=0\}}e^{i(t_1+\cdots+t_p)}\cos(t_1)\times\cdots\times\cos(t_p)\\
&\quad+\mathrm{E}\sum_{1\leq j_1,\ldots,j_p\leq p}\cos(t_1)\times\cdots\times\cos(t_p)\mathbf{1}_{\{X_n\neq 0\}}\\
&\qquad\times\left(\left(\prod_{k=1}^{p}\frac{X_n^{(j_k)}}{S_n}\right)-\frac{1}{p^p}\right)\exp\left(i\sum_{k=1}^{p}t_kX_n^{(j_k)}\right)\\
&\quad-\mathrm{E}\sum_{1\leq j_1,\ldots,j_p\leq p}\cos(t_1)\times\cdots\times\cos(t_p)\\
&\qquad\times\frac{1}{p^p}\mathbf{1}_{\{X_n=0\}}\exp\left(i\sum_{k=1}^{p}t_kX_n^{(j_k)}\right)\\
&\quad+\mathrm{E}\sum_{1\leq j_1,\ldots,j_p\leq p}\cos(t_1)\times\cdots\times\cos(t_p)\\
&\qquad\times\frac{1}{p^p}\left[\exp\left(i\sum_{k=1}^{p}t_kX_n^{(j_k)}\right)-\exp\left(i\sum_{k=1}^{p}t_k\Xi_n^{(j_k)}\right)\right].
\end{aligned}$$

Definition 3 *J_p is the set of applications $\{1,\ldots,p\}\longrightarrow\{1,\ldots,p\}$. For $f\in J_p$, we note $\#f=\#\{f(i),\ i\in\{1,\ldots,p\}\}$. Let $J_p^0=\{f\in J_p,\ \#f=1\}$. For $f\in J_p$ and $t\in\mathbb{R}^p$, we define $f(t)\in\mathbb{R}^p$ by:*

$$(f(t))_i=\sum_{f(j)=i}t_j\qquad\forall i\in\{1,\ldots,p\}.$$

We remark that $(f\circ g)(t)=f(g(t))$. With these notations, the last expression of $D_{n+1}(t)$ leads to:

$$D_{n+1}(t)=\varepsilon_n(t)+\frac{1}{p^p}\sum_{f\in J_p}\cos(t_1)\times\cdots\times\cos(t_p)\,D_n(f(t)),$$

where, thanks to Lemma 7,

$$\begin{aligned}
|\varepsilon_n(t)|\leq\varepsilon_n &= 2\mathrm{P}\left[X_n=0\right]+p^p\mathrm{E}\mathbf{1}_{\{X_n\neq 0\}}d(X_n)/F_n\\
&\leq 2\mathrm{P}\left[X_n=0\right]+p^p\left(\mathrm{E}d(X_n)^2\right)^{1/2}\left(\mathrm{E}\mathbf{1}_{\{X_n\neq 0\}}F_n^{-2}\right)^{1/2}.
\end{aligned}$$

Hence, thanks to Corollary 7 and Lemma 2, $\varepsilon_n\to 0$ as $n\to\infty$. Since the numbers t_i add to zero, $D_n(f(t))=0$ if $f\in J_p^0$. Hence:

$$D_{n+1}(t)=\varepsilon_n(t)+\frac{1}{p^p}\sum_{f\in J_p\setminus J_p^0}\cos(t_1)\times\cdots\times\cos(t_p)\,D_n(f(t)).\tag{1.29}$$

The set

$$\mathcal{J}_p(t) = \{f_k \circ \cdots \circ f_1(t),\ k \in \mathbb{N},\ f_1, \ldots, f_k \in J_p\}$$

is finite, and, for all $u \in \mathcal{J}_p(t)$, $\sum_i u_i = 0$. Introducing

$$\bar{D}_n(t) = \sup_{u \in \mathcal{J}_p(t)} |D_n(u)|,$$

equation (1.29) leads to:

$$\bar{D}_{n+1}(t) \leq \varepsilon_n + \frac{|J_p \setminus J_p^0|}{p^p} \bar{D}_n(t).$$

Since $|J_p \setminus J_p^0|/p^p < 1$, this proves that $D_n(t)$ goes to zero. □

Although we do not have a close formula for π_p, its Fourier transform can be algorithmically computed as follows. Let

$$\phi(t) = \mathrm{E} \exp\left(i \sum_{j=1}^{p} t_j Z(j)\right),$$

where $t = (t_1, \ldots, t_p)$ is a p-tuple of real numbers summing to zero, and $Z = (Z(1), \ldots, Z(p)) \in \mathbb{M}_p$ is a random vector, such that $\Theta(Z)$ has the law π_p (one can check that $\exp\left(i \sum_{j=1}^{p} t_j Z(j)\right)$ is really a function of $\Theta(Z)$). The invariance of π_p with respect to Ξ yields:

$$\begin{aligned}
\phi(t) &= \cos(t_1) \times \cdots \times \cos(t_p) p^{-p} \sum_{1 \leq j_1, \ldots, j_p \leq p} \mathrm{E} \exp\left(i \sum_{l=1}^{p} t_l Z(j_l)\right) \\
&= \cos(t_1) \times \cdots \times \cos(t_p) p^{-p} \sum_{f \in J_p} \phi(f(t)). \qquad (1.30)
\end{aligned}$$

Note $\#t = \#\{i,\ t_i \neq 0\}$. The set $\{f(t),\ f \in J_p\}$ contains p-tuples u such that either $\phi(u) = \phi(t)$ (when $\#u = \#t$, that is, when u is deduced from t by a permutation), or $\#u < \#t$. The relation (1.30) gives a method for computing $\phi(t)$ by induction on $\#t$, since $\phi(t) = 1$ when $t = 0$. As an example, we compute $\phi(t)$ when $\#t = 2$.

Proposition 15 *For all $1 \leq i \neq j \leq p$, the sequence $X_n^{(i)} - X_n^{(j)}$ converges in law to a random variable Δ of law*

$$(1 - \alpha)\delta_0 + \sum_{n \geq 1} (1 - a)^{n-1} a (\delta_{2n} + \delta_{-2n}) \alpha/2,$$

where $\alpha = \sqrt{p}/(\sqrt{p} + 1)$ and $a = 2/(\sqrt{p} + 1)$. The variable Δ may be written as $\Delta = 2ZG$, where the law of the random variable Z is

$$(\delta_1 + \delta_{-1})\alpha/2 + (1 - \alpha)\delta_0,$$

and G is a geometric random variable independent of Z: for $n \geq 1$,

$$\mathrm{P}\left[G = n\right] = (1-a)^{n-1}a.$$

Proof of Proposition 15:

For $t \in \mathbb{R}$, equation (1.30) gives:

$$\begin{aligned}
\phi(t,-t,0,\ldots,0) &= \cos^2 t \times p^{-p}\left(\sum_{f\in J_p, f(1)\neq f(2)} \phi(t,-t,0,\ldots,0)\right. \\
&\qquad \left. + \sum_{f\in J_p, f(1)=f(2)} 1\right) \\
&= \cos^2 t\left(\frac{p-1}{p}\phi(t,-t,0,\ldots,0) + \frac{1}{p}\right),
\end{aligned}$$

leading to:

$$\phi(t,-t,0,\ldots,0) = \frac{\cos^2 t}{p-(p-1)\cos^2 t}.$$

The result follows by a simple computation. □

1.7 Technical lemmas

Lemma 6 computes $\mathrm{E}^x\left(\Delta F_n \mid \mathcal{F}_n\right)$ and $\mathrm{E}^x\left((\Delta F_n)^2 \mid \mathcal{F}_n\right)$. A useful consequence of Lemma 7 is that the transition probabilities of the selection step can be approximated by those corresponding to a uniform selection. Indeed, by Lemma 2, $d(X_n)$ is essentially a finite number, and, by Corollary 7, F_n goes to infinity. We omit the proof of Lemma 8.

Lemma 6 *For all $x \in \mathbb{N}^p$,*

$$\mathrm{E}^x\left(\Delta F_n \mid \mathcal{F}_n\right) = \mathbf{1}_{\{X_n\neq 0\}}\frac{1}{p^2F_n}\sum_{1\leq i<j\leq p}\left(X_n^{(j)} - X_n^{(i)}\right)^2 + \mathbf{1}_{\{X_n=0\}},$$

and

$$\begin{aligned}
\mathrm{E}^x\left((\Delta F_n)^2 \mid \mathcal{F}_n\right) &= p^{-1} + p^{-2}\sum_{1\leq i\neq j\leq p}\left[\mathbf{1}_{\{X_n\neq 0\}}H_{i,j}(X_n) + \mathbf{1}_{\{X_n=0\}}\right] \\
&\quad + p^{-2}\sum_{1\leq i\leq p}\left[\mathbf{1}_{\{X_n\neq 0\}}K_i(X_n) + \mathbf{1}_{\{X_n=0\}}\right],
\end{aligned}$$

where we used the following shorthands:

$$\begin{aligned} H_{i,j}(X_n) &= \sum_{1\le a,b\le p} \frac{X_n^{(a)}X_n^{(b)}}{S_n^2}(X_n^{(a)}-X_n^{(i)})(X_n^{(b)}-X_n^{(j)}), \\ K_i(X_n) &= \sum_{1\le a\le p} \frac{X_n^{(a)}}{S_n}(X_n^{(a)}-X_n^{(i)})^2. \end{aligned}$$

Proof of Lemma 6:

We prove the formula for $(\Delta F_n)^2$ and omit the easier proof of the formula for ΔF_n. Note that

$$\mathrm{E}^x\left((\Delta F_n)^2 \;\middle|\; \mathcal{G}_n\right) = \mathrm{E}\left((G_n - F_n + (\epsilon_1+\cdots+\epsilon_p)/p)^2 \;\middle|\; \mathcal{G}_n\right),$$

where $(\epsilon_i)_i$ is i.i.d., of symmetric ± 1 Bernoulli law, and independent of $\mathcal{G}_n$. Hence:

$$\mathrm{E}^x\left((\Delta F_n)^2 \;\middle|\; \mathcal{G}_n\right) = (G_n - F_n)^2 + p^{-1}.$$

We get the result of the lemma by noting that:

$$\mathrm{E}^x\left((G_n - F_n)^2 \;\middle|\; \mathcal{F}_n\right) = p^{-2}\sum_{1\le i,j\le p}\mathrm{E}^x\left((Y_n^{(i)}-X_n^{(i)})(Y_n^{(j)}-X_n^{(j)}) \;\middle|\; \mathcal{F}_n\right).$$

□

The following lemma gives a precise formulation of the basic fact that, if $d(X_n)$ is small when compared to F_n, the transition probabilities of the selection step at time n can be efficiently approximated by uniform transition probabilities.

Lemma 7 *Let $x = (x(1), \ldots, x(p))$ be any p-tuple of nonnegative real numbers, with $x \neq 0$. Set $S = x(1) + \cdots + x(p)$. For k and $i_1, \ldots, i_k$ in $\{1, \ldots, p\}$,*

$$\left|\left(\prod_{j=1}^k \frac{x(i_j)}{S}\right) - \frac{1}{p^k}\right| \le kd(x)/S.$$

Proof of Lemma 7:

By induction on k. If $k = 1$,

$$\left|\frac{x(i)}{S} - \frac{1}{p}\right| = \left|\frac{\sum_{j=1}^p (x(i)-x(j))}{pS}\right| \le d(x)/S.$$

If the result holds for all $l \le k-1$ with $k \ge 2$, one has, reducing to the same denominator,

$$\left|\prod_{l=1}^k \frac{x(i_l)}{S} - \frac{1}{p^k}\right| = \frac{1}{(pS)^k}\left|\sum_{1\le j_1,\ldots,j_k\le p}\left(\prod_{l=1}^k x(i_l) - \prod_{l=1}^k x(j_l)\right)\right|.$$

Adding and substracting $x(i_1) \times \cdots \times x(i_{k-1}) \times x(j_k)$ to each term of this sum, we get:

$$\begin{aligned}\left|\prod_{l=1}^{k} \frac{x(i_l)}{S} - \frac{1}{p^k}\right| &= \frac{1}{(pS)^k}\left|\sum_{1\leq j_1,\ldots,j_k\leq p}\left[(x(i_k)-x(j_k))\prod_{l=1}^{k-1}x(i_l)\right.\right.\\ &\qquad\left.\left.+\,x(j_k)\left(\prod_{l=1}^{k-1}x(i_l)-\prod_{l=1}^{k-1}x(j_l)\right)\right]\right|,\end{aligned}$$

which leads to the bound

$$\begin{aligned}\left|\prod_{l=1}^{k}\frac{x(i_l)}{S}-\frac{1}{p^k}\right| &\leq \left(\prod_{l=1}^{k-1}\frac{x(i_l)}{S}\right)\left|\sum_{j_k=1}^{p}\frac{x(i_k)-x(j_k)}{pS}\right|\\ &\quad+\frac{x(j_k)}{S}\left|\sum_{1\leq j_1,\ldots,j_{k-1}\leq p}\left(\prod_{l=1}^{k-1}x(i_l)-\prod_{l=1}^{k-1}x(j_l)\right)\right|\\ &\leq \left|\frac{x(i_k)}{S}-\frac{1}{p}\right|+\left|\left(\prod_{l=1}^{k-1}\frac{x(i_l)}{S}\right)-\frac{1}{p^{k-1}}\right|\\ &\leq (1+(k-1))d(x)/S\end{aligned}$$

by the induction hypothesis. $\square$

Lemma 8 *Let* $\alpha \in]-1,1[$ *and* $q\in\mathbb{N}$. *As* r *goes to infinity,*

$$\sum_{k=0}^{r-1}\alpha^{r-1-k}k^q = \frac{1}{1-\alpha}r^q + o(r^q).$$

1.8 Simulations

We have performed numerical simulations of the algorithms studied in this paper. These simulations first yielded empirical insights of the behaviour of the chain. Afterwards, they showed that the asymptotic results of the previous sections were indeed observed on real finite-size systems.

As regards the technical side of the simulations, we designed a `C` computer program using a Mersenne-Twister random generator, see M. Matsumoto and T. Nishimura [49]. We used an `egcs` `C` compiler on a generic Pentium II ® computer.

Figure 1.1 shows the time-evolution of the average fitness F_n for a single realization of the chain. Figure 1.2 shows the average time-evolution of F_n performed over 10^4 independent realizations, for several values of the population size p. Figure 1.3 displays the empirical distribution of $X_n^{(1)}/\sqrt{n}$ for $n=2\cdot 10^4$ and $p=2$. Figure 1.4

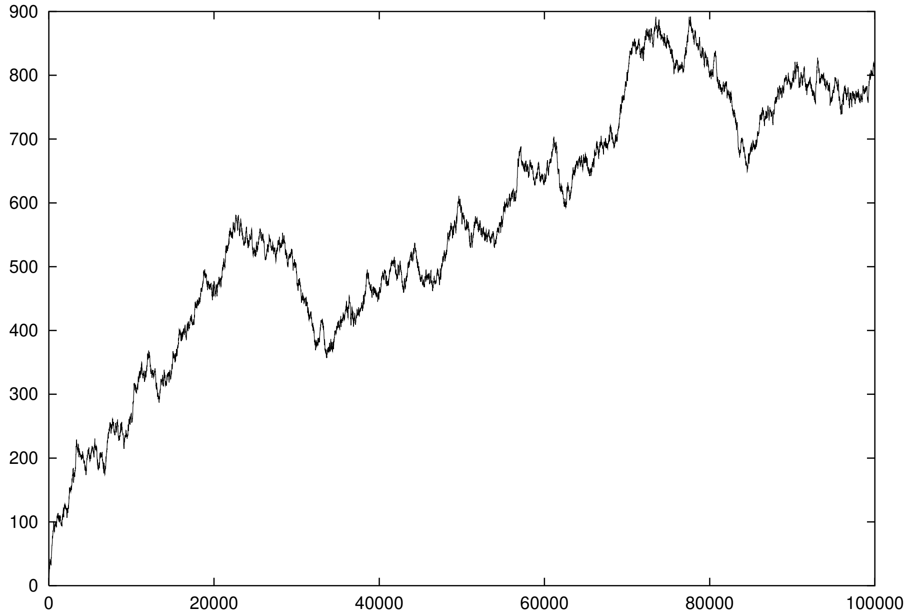

Figure 1.1: Single trajectory

Single realization of F_n *up to time* 10^5, 4 *particles.*

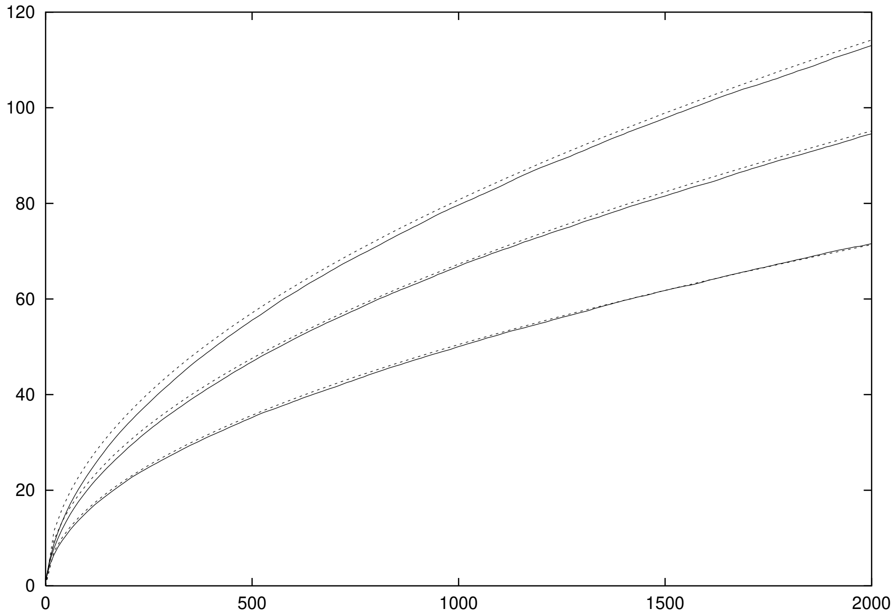

Figure 1.2: Average evolution

Average of F_n *over* 10^4 *realizations up to time* 10^5,
$p = 2, 3, 4$ *particles.*

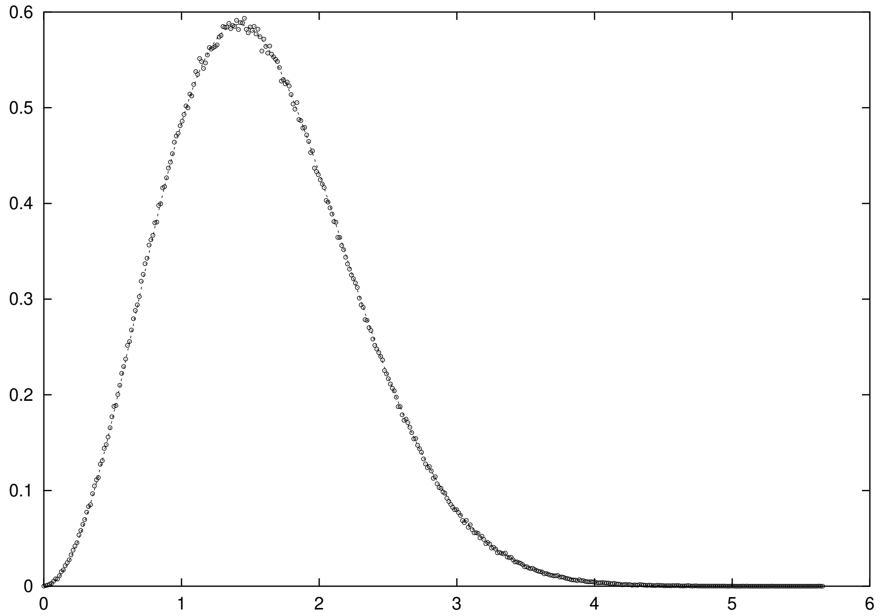

Figure 1.3: Empirical distribution
Empirical distribution of $X_n^{(1)}/\sqrt{n}$, *2 particles, time* 2.10^4, 2.10^6 *realizations.*

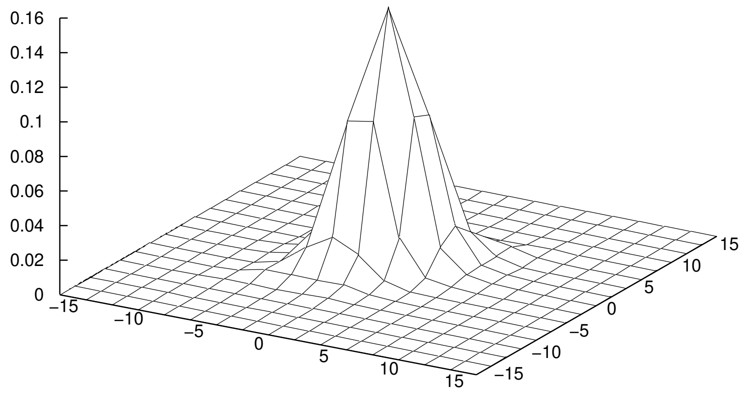

Figure 1.4: Cloud distribution

Frequency of the relative positions of the three particles, time 5.10^4*,*
10^6 *realizations.*

displays the empirical joint distribution of $(X_n^{(2)} - X_n^{(1)}, X_n^{(3)} - X_n^{(1)})$ for $n = 5 \cdot 10^4$ and $p = 3$. Dashed lines display the theoretical asymptotical values.

Chapitre 2

Asymptotics of a mutation-selection algorithm with space inhomogenous additive mutations

2.1 Abstract

We study the asymptotic behavior of finite population mutation-selection algorithms defined on $[1, +\infty)$ with fitness $F(x) = x$ and space-inhomogenous additive mutations $X \mapsto X + LX^{\alpha}$ where L is a centered random variable and $0 < \alpha < 1$. The asymptotic speed of growth of the average fitness is determined.

2.2 Introduction

2.2.1 Motivations

Genetic algorithms are popular stochastic optimization methods, formally introduced by J. Holland [39] and based on a biological analogy. One seeks to maximize the *objective function* $F : S \to \mathbb{R}$, defined on the *state space* S. The mutation-selection genetic algorithms we consider evolve a finite population of particles in S under the following iterated two-stages procedure:

1. **Mutation:** each particle moves at random to a neighboring element of S, a convenient graph structure on S being given.

2. **Selection:** a new population of particles is created by resampling. Particles with low relative values of F tend to be eliminated, particles with large relative

values of F are more often replicated.

Mutation allows for the discovery of new, possibly better, feasible solutions to the maximization of F, and selection concentrates the exploration of S around good, already discovered solutions. Despite the successes of this approach in the applications and numerous experimental studies, few rigorous mathematical results are available, that would explain how genetic algorithms work. In particular, the interaction between mutation, selection and population size remains poorly understood, even for simple examples. Among the exceptions is the paper by R. Cerf [15], which gives general asymptotic convergence results in the context of Markov chains with rare transitions. See the papers by Y. Rabinovich and A. Wigderson [62], G. Rudolph [66] and J. Bérard and A. Bienvenüe [7], [8] where simple examples are worked out, and C. Mazza and D. Piau [50] for results in the infinite-population case. In [8] and [7], we studied finite population genetic algorithms on the integers where mutation works by adding a ± 1 Bernoulli random variable to the fitness. In this paper, we extend our study to additive space-inhomogenous perturbative mutations, that is, mutations whose effect on the fitness is of the form $X \longmapsto X + Lg(X)$, where L is a random variable independent from X with given law μ, and where $g(x)$ is small compared to x when x is large. More precisely, we make the specific choice $g(x) = x^{\alpha}$ with $0 < \alpha < 1$, that appears as a natural first case to study. Another motivation lies in the work of Y. Rabinovich et A. Wigderson, who study in [62] the behavior of an infinite population genetic algorithm on the integers with binomial mating. The finite population version is essentially equivalent to the choice $\mu = \mathcal{N}(0,1)$ et $g(x) = \sqrt{x}$. Another motivation for the study of these algorithms lies in the following remark: in the case investigated in [8] and [7], the growth speed of the average fitness was shown to be of order $\sqrt{n}$, that is, of the same order as the growth speed of a simple reflected random walk. In other words, the selection had no effect on the order of the speed, but only on the limiting law. It is thus a natural question to ask whether one can observe a strictly larger order for the growth speed of the genetic algorithm than for the iterated mutation, if the amplitude of mutation is allowed to increase with fitness.

2.2.2 Description of the model and notations

The algorithms we study are defined as discrete time Markov chains on the state space $[1, +\infty[^p$, $p \geq 1$ standing for population size, as follows:

$$X_0 = (1, \dots, 1).$$

1. **selection:** $X_n \longrightarrow X'_n$

The $X_n'^{(i)}$'s, $1 \le i \le p$, are chosen randomly and independently among the $\left\{X_n^{(i)},\ 1 \le i \le p\right\}$'s according to the probability law

$$\frac{1}{S_n}\sum_{i=1}^{p} X_n^{(i)}\delta_{X_n^{(i)}}, \quad \text{where} \quad S_n = \sum_{i=1}^{p} X_n^{(i)}.$$

2. **mutation:** $X_n' \longrightarrow X_{n+1}$

$$X_{n+1}^{(i)} = \max\left(X_n'^{(i)} + L_n^{(i)}\left(X_n'^{(i)}\right)^{\alpha}\ ,\ 1\right),$$

the $L_n^{(i)}$'s , $1 \le i \le p$ being independent identically distibuted random variables, independent from the past, with common law μ.

(X_n) is called the *population at time* n, or the *cloud of particles.*
F_n and F_n' are the average fitness values in X_n and X_n' respectively:

$$F_n = \frac{1}{p}\sum_{i=1}^{p} X_n^{(i)}\ ,\ F_n' = \frac{1}{p}\sum_{i=1}^{p} X_n'^{(i)}.$$

For $x \in \mathbb{R}^p$, the diameter of x is defined as:

$$d(x) = \max_{1 \le i,j \le p} |x_i - x_j|$$

$$\mathcal{F}_n = \sigma\left(X_0, X_0', \ldots, X_{n-1}, X_{n-1}', X_n\right),$$

and $\mathcal{F}_n'$ the σ-algebra generated by $\mathcal{F}_n \cup\ \sigma\left(X_n'\right)$.

We define

$$L_n^* = \max_{1 \le i \le p} |L_n^{(i)}|,$$

and

$$\beta = \frac{1}{2(1-\alpha)}\ ,\ r(p,\alpha) = (2p-3)(1-\alpha)^{-1} + 2.$$

Remark 16 *The mutation steps are not defined exactly as $x \longmapsto x + Lx^{\alpha}$ since a barrier is put in 1 to prevent X_n from taking negative values. When x is large, this truncation becomes negligible. Moreover, this formulation allows us not to deal with small values of x (close to zero) that would not be relevant in the problem.*

2.2.3 Statement of the results

We assume that μ has moments of all orders and has zero mean. This implies that all the moments we consider are finite. Moreover, we assume that μ has a density with respect to the Lebesgue measure, that is everywhere positive. This assumption simplifies the proofs significantly.

Our first results concerns the behavior of the Markov chain Z_n that evolves through mutation steps only.

$$Z_0 = 1 \ , \ Z_{n+1} = \max(Z_n + LZ_n^\alpha, 1)$$

Note that $Z_n = X_n^{(1)}$ when $p = 1$ (no selection).

Theorem 4 *The behavior of Z_n depends on α as follows:*

- *if $0 < \alpha < 1/2$, Z_n is null recurrent or transient and grows at speed n^β in the sense that $n^{-\beta} Z_n$ converges in law to*

$$\left((1-\alpha)\mathrm{E}(L^2)^{1/2} B(\alpha)\right)^{2\beta}$$

  where $B(\alpha)$ is the value at time 1 of a Bessel process of dimension $r(1,\alpha)$,

- *if $\alpha = 1/2$, Z_n is null recurrent or transient and grows at speed $o(n^\beta) = o(n)$, in the sense that $n^{-1} Z_n$ goes to zero in L^s for all $1 \leq s < +\infty$,*

- *if $1/2 < \alpha < 1$, Z_n is positive recurrent.*

The asymptotic behavior of Z_n has a phase transition at $\alpha = 1/2$. For $0 < \alpha < 1/2$, the behavior of Z_n extends that of a simple reflected random walk ($\alpha = 0$) and has a speed in the usual sense. For $\alpha > 1/2$, the amplitude of the mutations is so large that Z_n cannot escape to infinity in probability, and it has zero speed. In the limiting case $\alpha = 1/2$, we only have an upper bound on the growth speed of Z_n.

The behavior of the genetic algorithm is stated in the following theorem, valid for $p \geq 2$,

Theorem 5 *As n goes to infinity, $n^{-\beta} X_n$ converges in law to*

$$\left((1-\alpha)\mathrm{E}(L^2)^{1/2} B(p,\alpha)\right)^{2\beta} (1, \dots, 1)$$

where $B(p,\alpha)$ is the value at time 1 of a Bessel process of dimension $r(p,\alpha)$.

Thus, the speed of growth of the algorithm is n^β for all $0 < \alpha < 1$, and the behavior of the algorithm does not have a phase transition. As soon as $\alpha > 1/2$, the order of the growth speed is strictly larger than the one for iterated mutation. Unlike the iterated mutation case, when $\alpha > 1/2$, the algorithm uses the amplitude of the mutations to grow. We note that, as in the $\alpha = 0$ case investigated in [8] and [7], the diameter of the population is small compared to the average fitness, and the whole population behaves as a single point on the n^β scale.

2.2.4 Contents

In Section 2.3, we sketch the arguments that allow us to use the standard results of countable state space Markov chain theory in our setting. In Section 2.4, we study the behavior of the Markov chains Z_n that evolve through mutation steps only. In Section 2.5, we prove our main estimate on the diameter of the cloud. In Section 2.6, we prove the non-positivity of the Markov chain (X_n). In Section 2.7, we prove the convergence in law of the normalized cloud of particles. In Section 2.8, we prove the main technical lemmas needed for our proofs.

2.3 A few words on Markov chains

The Markov chains we study here are defined on uncountable spaces. However, most of the results of standard countable space Markov chain theory apply here, as we briefly explain. Our main reference concerning general state space Markov chains is the book by S. Meyn and R. Tweedie [51] from which we use the terminology.

With our assumptions, it is routine to check that both X_n and Z_n are $\phi-$irreducible, that $(1, \dots, 1)$ is an accessible atom, that every compact set is a small set, and that X_n and Z_n are strongly aperiodic.

As a consequence, if X_n (resp. Z_n) is positive recurrent, there exists a unique invariant probability measure π such that X_n (resp. Z_n) converges in law to π, and if the Markov chain (X_n) (resp. (Z_n)) is null, for every compact set K, $\mathrm{P}\left[X_n \in K\right]$ (resp. $\mathrm{P}\left[Z_n \in K\right]$) goes to zero as n goes to infinity.

2.4 Asymptotics for mutation-only chains

Proof of Theorem 4:

Step 1 - Upper bound on $\mathrm{E}(Z_n^2)$

We prove that

$$\mathrm{E} Z_n^2 \leq C n^{2\beta}$$

Let

$$B_n = \{Z_n + L_n Z_n^{\alpha} < 1\}$$

Remark 17 *For* $x \geq 1$, $x - Lx^{\alpha} \leq 1 \Rightarrow x \leq (1 + |L|)^{2\beta}$.

As a consequence, for all $q \geq 0$,

$$\mathbf{1}_{B_n} \leq Z_n^{-q} (1 + |L_n|)^{2\beta q}$$

$$\begin{aligned}
\mathrm{E}Z_{n+1}^2 &= \mathrm{E}\mathbf{1}_{B_n^c}(Z_n + L_n Z_n^\alpha)^2 + \mathrm{E}\mathbf{1}_{B_n} \\
&= \mathrm{E}\mathbf{1}_{B_n^c}(Z_n^2 + 2L_n Z_n^{1+\alpha} + L_n^2 Z_n^{2\alpha}) + \mathrm{P}(B_n).
\end{aligned}$$

By the above inequality with $q = 2$, and using the fact that L_n has moments of all orders, Hölder's inequality implies that

$$\mathrm{E}\big(Z_{n+1}^2 \mathbf{1}_{B_n}\big) \leq C$$

Using the fact that L_n is centered and independent from Z_n, we finally have

$$\mathrm{E}Z_{n+1}^2 \leq \mathrm{E}Z_n^2 + \left(\mathrm{E}L_n^2\right)\left(\mathrm{E}Z_n^{2\alpha}\right) + \mathrm{cte}$$

Since $\alpha < 1$, by Hölder's inequality,

$$\mathrm{E}\big(Z_n^{2\alpha}\big) \leq \big(\mathrm{E}\big(Z_n^2\big)\big)^\alpha$$

Thus, the sequence $u_n = \mathrm{E}\big(Z_n^2\big)$ satisfies:

$$u_{n+1} \leq u_n + C u_n^\alpha + C$$

By Lemma 19, $u_n = O(n^{2\beta})$.

Step 2 - Drift analysis of $\mathrm{E}(Z_n)$

$$\begin{aligned}
\mathrm{E}Z_{n+1} - \mathrm{E}Z_n &= \mathrm{E}(Z_n + L_n Z_n^\alpha) + \mathrm{E}\mathbf{1}_{B_n}(1 - (Z_n + L_n Z_n^\alpha)) - \mathrm{E}Z_n \\
&= \mathrm{E}L_n Z_n^\alpha + \mathrm{E}\mathbf{1}_{B_n}(1 - (Z_n + L_n Z_n^\alpha)) \\
&= \mathrm{E}\mathbf{1}_{B_n}(1 - (Z_n + L_n Z_n^\alpha)),
\end{aligned}$$

since L_n is centered and independent from Z_n.

Define φ on $[1, +\infty)$ by

$$\varphi(x) = \mathrm{E}\mathbf{1}_{\{x + Lx^\alpha < 1\}}(1 - (x + Lx^\alpha)),$$

so that

$$\mathrm{E}Z_{n+1} - \mathrm{E}Z_n = \mathrm{E}\varphi(Z_n).$$

By Remark 17,

$$\begin{aligned}
\varphi(x) &\leq \mathrm{E}\mathbf{1}_{B_n}(1 + |x + xL^\alpha|) \\
&\leq \mathrm{E}\left[1 + (1 + |L|)^{2\beta} + (1 + |L|)^{2\beta+\alpha}\right] \\
&\leq 3\mathrm{E}(1 + |L|)^{2\beta+\alpha}.
\end{aligned}$$

One easily checks that φ is continuous since μ is absolutely continuous with respect to the Lebesgue measure. Note that $\varphi(x) \geq 0$ for all x. Moreover, since $L \not\equiv 0$, φ is strictly positive on some interval $[1, 1+\epsilon[$, with $\epsilon > 0$.

Now, if we assume Z_n to be ergodic for some $\alpha \leq 1/2$, Z_n converges in law to its invariant measure μ. Since φ is a bounded continuous function,

$$\lim_{n\to+\infty} \mathrm{E}(\varphi(Z_n)) = \int \varphi \mathrm{d}\mu = e_1$$

It is easy to check that μ must charge 1. Hence $e_1 > 0$. Thus,

$$\mathrm{E}(Z_n) \sim e_1 n.$$

If $\alpha < 1/2$, this contradicts the upper bound $\mathrm{E}Z_n^2 = O(n^{2\beta})$. If $\alpha = 1/2$, the inequality

$$\mathrm{P}\left[Z_n \geq \frac{1}{2}\mathrm{E}Z_n\right] \geq \frac{1}{8}\frac{(\mathrm{E}Z_n)^2}{\mathrm{E}(Z_n^2)}$$

implies that the sequence of laws of Z_n is not tight, contradicting the assumed ergodicity.

Hence, for $\alpha \leq 1/2$, the Markov chain (Z_n) is null. The usual drift criteria [51] may be applied here, using only the fact that ϕ is positive and bounded. However, the structure of the argument will be the same for the genetic algorithm, for which these criteria do not seem to work.

Step 3 - Drift analysis of $\mathrm{E}\left(Z_n^{2(1-\alpha)}\right)$

On B_n^c,

$$Z_{n+1}^{2(1-\alpha)} - Z_n^{2(1-\alpha)} = Z_n^{2(1-\alpha)}\left[(1 + L_n Z_n^{\alpha-1})^{2(1-\alpha)} - 1\right]$$

A Taylor expansion shows that:

$$\left|(1+x)^{2(1-\alpha)} - 1 + 2(1-\alpha)x + (1-\alpha)(1-2\alpha)x^2\right| \leq Cx^3$$

for all $|x| \leq 1/2$.

Define

$$D_n = \left\{\left|L_n Z_n^{\alpha-1}\right| \leq 1/2\right\}$$

From the above inequality, on $B_n^c \cap D_n$, if we replace

$$Z_{n+1}^{2(1-\alpha)} - Z_n^{2(1-\alpha)}$$

by

$$2(1-\alpha)L_n Z_n^{(1-\alpha)} + (1-\alpha)(1-2\alpha)(L_n)^2$$

the error committed is less than $C|L_n|^3 Z_n^{(\alpha-1)}$.

Assume troughout that the Markov chain (Z_n) is null. Then, using the fact that μ has moments of all orders and Hölder's inequality, $\mathrm{E}\big(|L_n|^3 Z_n^{\alpha-1}\big)$ goes to zero as n goes to infinity since $\alpha < 1$. Thus, the difference between

$$\mathrm{E}\left[\left(Z_{n+1}^{2(1-\alpha)} - Z_n^{2(1-\alpha)}\right) \mathbf{1}_{D_n} \mathbf{1}_{B_n^c}\right]$$

and

$$\mathrm{E}\left[\left(2(1-\alpha)L_n Z_n^{1-\alpha} + (1-\alpha)(1-2\alpha)L_n^2\right) \mathbf{1}_{D_n} \mathbf{1}_{B_n^c}\right]$$

goes to zero as n goes to infinity. Remark 17 and the nullity of the Markov chain (Z_n) imply that any expectation of the form:

$$\mathrm{E}\left(|L_n|^{q_1} Z_n^{q_2} \mathbf{1}_{D_n^c}\right) \ , \ \mathrm{E}\left(|L_n|^{q_1} Z_n^{q_2} \mathbf{1}_{B_n}\right)$$

goes to zero as n goes to infinity. Thus, the difference between

$$\mathrm{E}\left(Z_{n+1}^{2(1-\alpha)} - Z_n^{2(1-\alpha)}\right)$$

and

$$\mathrm{E}\left(2(1-\alpha)L_n Z_n^{1-\alpha} + (1-\alpha)(1-2\alpha)L_n^2\right)$$

goes to zero as n goes to infinity. Since L_n is centered and independent from Z_n, we finally have:

$$\lim_{n \to +\infty} \mathrm{E}\left(Z_{n+1}^{2(1-\alpha)} - Z_n^{2(1-\alpha)}\right) = (1-\alpha)(1-2\alpha)\mathrm{E}\left(L^2\right)$$

If $\alpha > 1/2$, $(1-\alpha)(1-2\alpha) < 0$ and we deduce that $\mathrm{E}\left(Z_n^{2(1-\alpha)}\right)$ goes to $-\infty$ as $n \to +\infty$, a contradiction since $Z_n \geq 1$ by definition. Hence, for $\alpha > 1/2$, Z_n is ergodic. For $\alpha < 1/2$, this proves that n^β is the right growth speed for Z_n, and for $\alpha = 1/2$, that $\mathrm{E}Z_n = o(n)$. □

2.5 Moment estimates on the diameter

Proposition 18 *For all $q > 0$, there exists a constant $A_{\alpha,q}$ such that, for all n,*

$$\mathrm{E}\left(\frac{d(X_n)^q}{\left(\min_i X_n^{(i)}\right)^{\alpha q}}\right) \leq A_{\alpha,q}.$$

Proof:

Let

$$U_n = \frac{d(X_n)^q}{(\min_i X_n^{(i)})^{\alpha q}}, m = \min_i X_n'^{(i)}$$

and define $\mathcal{H}_n$ the $\sigma-$algebra generated by $\mathcal{F}_n$ and $(L_n^{(i)})_{1\le i\le p}$.

Let

$$h(M) = \max\left(\left(\frac{M}{\theta_{\alpha,q,\eta}}\right)^{2\beta}, (1+M)^{2\beta}\right),$$

where $\theta_{\alpha,q,\eta}$ is defined in Lemma 18,

and

$$A_n = \left\{\min_i X_n'^{(i)} \ge h(L_n^*)\right\}.$$

By Lemma 17,

$$\left(\min_i X_{n+1}^{(i)}\right)^{-\alpha q} \le \left(\max\left(m - L_n^* m^\alpha, 1\right)\right)^{-\alpha q}.$$

By Remark 17, on A_n, $m - L_n^* m^\alpha \ge 1$ and

$$(m - L_n^* m^\alpha)^{-\alpha q} \le (1+\eta) m^{-\alpha q}.$$

On the other hand, on A_n^c,

$$h(L_n^*)^{\alpha q} m^{-\alpha q} \ge 1.$$

Thus, with

$$g(M) = \max(1+\eta, h(M)^{\alpha q}),$$

$$\left(\min_i X_{n+1}^{(i)}\right)^{-\alpha q} \le g(L_n^*) m^{-\alpha q} \tag{2.1}$$

Recall that, for $x \in [1, +\infty)^p$, $\max x = \min x + d(x)$. Thus, Lemma 15 implies that:

$$d(X_{n+1})^q \le \left(d(X_n') + 2L_n^* \left(\min_i X_n'^{(i)} + d(X_n')\right)^\alpha\right)^q$$

Expanding, we get:

$$U_{n+1} \le \sum_{k=0}^{q} B_k$$

where

$$B_k = (\min_i X_{n+1}^{(i)})^{-\alpha q} \binom{q}{k} d(X_n')^{q-k} (2L_n^*)^k (\min_i X_n'^{(i)} + d(X_n'))^{\alpha k}$$

Using the fact that $\min_i X_n'^{(i)} \ge \min_i X_n^{(i)}$, $\max_i X_n'^{(i)} \le \max_i X_n^{(i)}$, and $d(X_n') \le d(X_n)$, and inequality (2.1), we obtain that $B_k \le C_k$, where

$$C_k = g(L_n^*)(\min_i X_n^{(i)})^{-\alpha q} \binom{q}{k} d(X_n)^{q-k} (2L_n^*)^k (\min_i X_n^{(i)} + d(X_n))^{\alpha k}$$

Upper bound for $C_k, k \geq 1$

For $k \geq 1$,

$$(\min_i X_n^{(i)} + d(X_n))^{\alpha k} \leq \gamma_{\alpha k} \left((\min_i X_n^{(i)})^{\alpha k} + d(X_n)^{\alpha k} \right)$$

where, for $s > 0$, γ_s is a positive constant such that for all $x, y > 0$, $(x+y)^s \leq \gamma_\beta (x^s + y^s)$.

Returning to the definition of C_k, we have

$$C_k \leq g(L_n^*) \binom{q}{k} 2^k \gamma_{\alpha k} (L_n^*)^k \left(\frac{d(X_n)^{q-k}}{\left(\min_i X_n^{(i)}\right)^{\alpha(q-k)}} + \frac{d(X_n)^{q-k+\alpha k}}{\left(\min_i X_n^{(i)}\right)^{\alpha q}} \right).$$

Since $\min_i X_n^{(i)} \geq 1$, we deduce that:

$$C_k \leq g(L_n^*) \binom{q}{k} 2^k \gamma_{\alpha k} (L_n^*)^k \left((U_n)^{(q-k)/q} + (U_n)^{(q-k+\alpha k)/q} \right).$$

Thanks to Hölder's inequality and to the fact that L_n^* has moments of all orders,

$$\mathrm{E} C_k \leq \text{constante}(\mu, p, q, k, \alpha, \eta) \times (\mathrm{E}(U_n))^{v(q,k,\alpha)}$$

where $v(q,k,\alpha) < 1$.

Upper bound for $\mathrm{E}B_0$

We study separately $\mathrm{E}B_0 \mathbf{1}_{A_n}$ and $\mathrm{E}B_0 \mathbf{1}_{A_n^c}$.

We first show that:

$$\mathrm{P}\left[A_n^c \mid \mathcal{H}_n \right] \leq p h(L_n^*) (F_n)^{-1} \tag{2.2}$$

Let $J_n^{(j)}$ denote the index chosen by selection to define $X_n'^{(j)}$, that is,

$$X_n'^{(j)} = X_n^{J_n^{(j)}},$$

and define

$$A_{i,j} = \{ X_n^{(i)} \leq h(L_n^*), J_n^{(j)} = i \}.$$

Noting that $\min_i X_n'^{(i)} \geq \min_i X_n^{(i)}$,

$$A_n^c \subset \bigcup_{1 \leq i,j \leq p} A_{i,j}$$

Since $\mathrm{P}\left[A_{i,j} \mid \mathcal{H}_n \right] \leq h(L_n^*)(pF_n)^{-1}$,

$$\mathrm{P}\left[A_n^c \mid \mathcal{H}_n \right] \leq p h(L_n^*) (F_n)^{-1}$$

Noting that $C_0 = g(L_n^*)U_n$ is $\mathcal{H}_n-$measurable, and thanks to (2.2),

$$\mathrm{E}\left(C_0\mathbf{1}_{A_n^c} \mid \mathcal{H}_n\right) \leq C_0\mathrm{P}\left[A_n^c \mid \mathcal{H}_n\right] \leq C_0 p h(L_n^*)(F_n)^{-1}$$

Turning back to the definition of C_0, and using the fact that $d(X_n) \leq pF_n$,

$$\mathrm{E}\left(C_0\mathbf{1}_{A_n^c} \mid \mathcal{H}_n\right) \leq p^2 h(L_n^*)g(L_n^*)d(X_n)^{q-1}\left(\min_i X_n^{(i)}\right)^{-\alpha q}$$

Since $\min_i X_n^{(i)} \geq 1$,

$$d(X_n)^{q-1}\left(\min_i X_n^{(i)}\right)^{-\alpha q} \leq (U_n)^{(q-1)/q}$$

Thanks to Hölder's inequality and to the fact that L_n^* has moments of all orders,

$$\mathrm{E}\big(C_0\mathbf{1}_{A_n^c}\big) \leq \mathrm{constant}(\mu, p, q, k, \alpha, \eta) \times (\mathrm{E}U_n)^{(q-1)/q}$$

On the other hand, by definition, on A_n,

$$B_0 = (\min_i X_{n+1}^{(i)})^{-\alpha q} d(X_n')^q \leq (1+\eta)m^{-\alpha q}d(X_n')^q$$

By positivity,

$$\mathrm{E}\left(B_0\mathbf{1}_{A_n} \mid \mathcal{H}_n\right) \leq (1+\eta)\mathrm{E}\left(m^{-\alpha q}d(X_n')^q \mid \mathcal{H}_n\right)$$

The contraction property of the selection step, stated in Lemma 16, implies that:

$$\mathrm{E}\left(m^{-\alpha q}d(X_n')^q \mid \mathcal{H}_n\right) \leq \kappa U_n$$

Hence:

$$\mathrm{E}(B_0\mathbf{1}_{A_n}) \leq \kappa(1+\eta)\mathrm{E}U_n$$

Putting together the previous estimates, we have:

$$\mathrm{E}U_{n+1} \leq \kappa(1+\eta)\mathrm{E}(U_n) + c(\mu, p, q, \alpha, \eta)\sum_{l=1}^{c(q)}(\mathrm{E}(U_n))^{v(q,l,\alpha)}$$

where $\max_{l=1,\ldots,c(q)} v(q, l, \alpha) < 1$.

Since $\kappa < 1$, we can find η such that $\kappa(1+\eta) < 1$, and the non-negative sequence $u_n = \mathrm{E}(U_n)$ satisfies an induction relation of the following form:

$$u_{n+1} \leq \kappa(1+\eta)u_n + f(u_n),$$

where f is a function satisfying $f(x) = o(x)$ as x goes to infinity. As a consequence, u_n is a bounded sequence. $\square$

2.6 Non-positivity of the chain

Remark 19 *In all the proofs that follow, we write $V(n, param)$ for a random variable with moments of all orders bounded with respect to n when param is fixed. Such a property is very useful in our proofs since it allows for an extensive use of Hölder's inequality to control error terms.*

A typical example of such a random variable is $(L_n^*)^{q_1} d(X_n)^{q_2} F_n^{-\alpha q_2}$. Recall that, for every $q \geq 1$ and random variable Y with $q-$th order moment, and every $\sigma-$algebra $\mathcal{G}$, $|\mathrm{E}(Y \mid \mathcal{G})|^q \leq \mathrm{E}(|Y|^q \mid \mathcal{G})$ a.s., so that, $\mathrm{E}(V(n, param) \mid \mathcal{G})$ still has moments of all orders bounded with respect to n when param is fixed.

The usual drift criteria for non-positivity (see [51]) do not seem to work in our case. Indeed, it is easy to show that $\mathrm{E}(\Delta F_n \mid \mathcal{F}_n)$ is almost surely positive, but it is certainly not bounded, as is usually required. Instead, we use a proof by contradiction by assuming that (X_n) is an ergodic Markov chain.

Proposition 20 *There exists a constant $C_1 > 0$ such that:*

$$\mathrm{E}(F_n^2) \leq C_1 n^{2\beta}$$

Proposition 21 *Under the assumption that (X_n) is positive, there exists a constant $C_2 > 0$ such that:*

$$\mathrm{E}(F_n^{2(1-\alpha)}) \geq C_2 n$$

According to the two propositions above, if we assume (X_n) to be positive, the following second moment inequality:

$$\mathrm{P}\left[Z \geq \frac{1}{2}\mathrm{E}Z\right] \geq \left[\frac{1}{2}\frac{\mathrm{E}Z}{(\mathrm{E}Z^u)^{\frac{1}{u}}}\right]^{\frac{u}{u-1}},$$

valid for all $u > 1$, with $Z = F_n^{2(1-\alpha)}$ and $u = 2\beta$, implies that the sequence of the laws of F_n is not tight, thus contradicting positivity.

Proof of proposition 20:

One has

$$F_{n+1}^2 - F_n^2 = (F_{n+1} - F_n)^2 + 2F_n(F_{n+1} - F_n).$$

According to the inequality

$$|F_{n+1} - F_n| \leq d(X_n) + L_n^*(pF_n)^\alpha,$$

stated in Lemma 10,

$$(F_{n+1} - F_n)^2 \leq 2\left(d(X_n)^2 + (L_n^*)^2(pF_n)^{2\alpha}\right)$$

Thus, by Proposition 18,

$$(F_{n+1} - F_n)^2 \le V(n,\alpha) F_n^{2\alpha}.$$

On the other hand,

$$\mathrm{E}\left(F'_n - F_n \;\middle|\; \mathcal{H}_n\right) = (p^2 F_n)^{-1} \sum_{1 \le i < j \le p} \left(X_n^{(i)} - X_n^{(j)}\right)^2$$

so

$$\mathrm{E}\left(F_n(F'_n - F_n) \;\middle|\; \mathcal{H}_n\right) \le C(p) d(X_n)^2$$

As before, by Proposition 18,

$$\mathrm{E}\left(F_n(F'_n - F_n) \;\middle|\; \mathcal{H}_n\right) \le V(n,\alpha) F_n^{2\alpha}$$

Then, according to Lemma 11, and by Hölder's inequality,

$$\mathrm{E}\big((F_{n+1} - F'_n) F_n \mathbf{1}_{D(n)^c}\big)$$

is bounded with respect to n.

On $D(n)$,

$$F_{n+1} - F'_n = p^{-1} \sum_{i=1}^{p} L_n^{(i)} (X_n'^{(i)})^\alpha,$$

and, as above,

$$\mathrm{E}\left(F_n \left(\sum_{i=1}^{p} L_n^{(i)} \left(X_n'^{(i)}\right)^\alpha\right) \mathbf{1}_{D(n)^c}\right)$$

is bounded.

Since the $L_n^{(i)}$ are centered and independent from X_n and X'_n,

$$\mathrm{E}\left(F_n \sum_{i=1}^{p} L_n^{(i)} (X_n'^{(i)})^\alpha\right) = 0$$

and finally

$$\mathrm{E}(F_n(F_{n+1} - F'_n))$$

is bounded.

Putting these estimates together,

$$\mathrm{E}\big(F_{n+1}^2 - F_n^2\big) \le C_1 + \mathrm{E}\big(V(n,\alpha) F_n^{2\alpha}\big).$$

Since $\alpha < 1$, by Hölder's inequality,

$$\mathrm{E}\big(F_{n+1}^2 - F_n^2\big) \le C_1 + C_2 \left(\mathrm{E} F_n^2\right)^\alpha$$

Thus, the sequence $u_n = \mathrm{E}F_n^2$ satisfies an induction relation of the form:

$$u_{n+1} \leq u_n + C_1 + C_2 u_n^{\alpha}.$$

The result follows by Lemma 19. □

Let

$$\delta_n(\lambda, r) = \mathrm{E}\left[\left(F_{n+r}^{2(1-\alpha)} - F_n^{2(1-\alpha)}\right) \mathbf{1}_{\{F_n \geq \lambda\}}\right]$$

Proposition 21 above is a consequence of the following lemma:

Lemma 9 *Under the assumption that (X_n) is positive, there exists $\lambda > 0$ and $r \geq 1$ such that:*

$$\liminf_{n \to +\infty} \delta_n(\lambda, r) > 0$$

Proof of implication: Lemma 9 $\Rightarrow$ Proposition 21:

Assume that (X_n) is positive, and let $\theta_n = \max\left(F_n^{2(1-\alpha)}, \lambda^{2(1-\alpha)}\right)$.

- on $\{F_n < \lambda\}$, $\theta_n = \lambda^{2(1-\alpha)}$ and, since by definition $\theta_{n+r} \geq \lambda^{2(1-\alpha)}$, we have $\theta_{n+r} - \theta_n \geq 0$
- on $\{F_n \geq \lambda\}$, $\theta_n = F_n^{2(1-\alpha)}$, and, since by definition $\theta_{n+r} \geq F_{n+r}^{2(1-\alpha)}$, we have $\theta_{n+r} - \theta_n \geq F_{n+r}^{2(1-\alpha)} - F_n^{2(1-\alpha)}$.

As a consequence,

$$\mathrm{E}(\theta_{n+r} - \theta_n) \geq \delta_n(\lambda, r)$$

□

Proof of Lemma 9:

Assume troughout that (X_n) is positive. Then, (X_n) has an invariant measure ν_p such that X_n converges in law to ν_p.

Step 1 - G is not integrable w.r.t. ν_p

Let us prove that $G(x_1, \ldots, x_p) = (1/p)(x_1 + \ldots + x_p)$ is not integrable w.r.t. ν_p. If it were the case, by invariance of ν_p,

$$\int G \mathrm{d}\nu_p = \mathrm{E}_{\nu_p} F_0 = \mathrm{E}_{\nu_p} F_1.$$

This would contradict the fact that, for all $x \in [1, +\infty)^p$,

$$\mathrm{E}^x\left(\Delta F_0 \mid \mathcal{F}_0\right) > 0.$$

Let ν be the law of G under ν_p. Since μ is absolutely continuous with respect to the Lebesgue measure, ν has no atoms except perhaps 1. By non-integrability

$$\sum_{k=1}^{+\infty} \nu([k, +\infty)) = +\infty.$$

Thus, there exists a sequence of integers Λ_ℓ going to infinity and satisfying:

$$\nu([\Lambda_l, +\infty)) \geq \Lambda_l^{-2}$$

Step 2 - Lower bound for $\delta_n(\lambda, r)$
Writing

$$F_{n+r}^{2(1-\alpha)} - F_n^{2(1-\alpha)} = \sum_{h=0}^{r-1} F_{n+h+1}^{2(1-\alpha)} - F_{n+h}^{2(1-\alpha)}$$

and using Lemma 12, we get:

$$\begin{aligned} \mathrm{E}\left(F_{n+r}^{2(1-\alpha)} - F_n^{2(1-\alpha)} \;\middle|\; \mathcal{F}_n \right) = \\ Q_1(1,\alpha,p)\mathrm{E}\left(\sum_{h=0}^{r-1} \sum_{1\leq i<j\leq p} F_{n+h}^{-2\alpha} \left(X_{n+h}^{(i)} - X_{n+h}^{(j)} \right)^2 \;\middle|\; \mathcal{F}_n \right) \\ +rQ_2(1,\alpha,p)\mathrm{E}L_1^2 + \mathrm{E}\left(\sum_{h=0}^{r-1} V(n+h,\alpha) F_{n+h}^{-v} \;\middle|\; \mathcal{F}_n \right) \end{aligned} \tag{2.3}$$

Now, by Lemma 14, $F_{n+h}^{-v} \leq V(n,r,\alpha)F_n^{-v}$ since $\mathbf{1}_{\{F_{n+h}\leq(1/2)F_n\}} \leq V(n,h,\alpha)F_n^{-v}$. Thus, we can rewrite the error term:

$$\mathrm{E}\left(\sum_{h=0}^{r-1} V(n+h,\alpha) F_{n+h}^{-v} \;\middle|\; \mathcal{F}_n \right) = V(r,n,\alpha)F_n^{-v}$$

Adding $\mathbf{1}_{\{F_n\geq\lambda\}}$ in (2.3), and taking expectations, we get, by the symmetry of the law of the $(X_m^{(i)})_{1\leq i\leq p}$,

$$\begin{aligned} \delta_n(\lambda,r) = Q_1(1,\alpha,p)(1/2)p(p-1)\sum_{h=0}^{r-1} \mathrm{E}\left[F_{n+h}^{-2\alpha}\left(X_{n+h}^{(1)} - X_{n+h}^{(2)} \right)^2 \mathbf{1}_{\{F_n\geq\lambda\}} \right] \\ +rQ_2(1,\alpha,p)\mathrm{E}L_1^2\mathrm{P}\left[F_n \geq \lambda\right] + \mathrm{E}\big(V(r,n,\alpha)F_n^{-v}\mathbf{1}_{\{F_n\geq\lambda\}}\big) \end{aligned} \tag{2.4}$$

Let

$$a_n(\lambda,h) = \mathrm{E}\left[\frac{(X_{n+h}^{(1)} - X_{n+h}^{(2)})^2}{F_{n+h}^{2\alpha}} \mathbf{1}_{\{F_n\geq\lambda\}} \right]$$

Using Lemma 13 and 14 , we have, for $0 \leq h \leq r-1$,

$$\begin{aligned} \mathrm{E}\left(F_{n+h+1}^{-2\alpha}\left(X_{n+h+1}^{(1)} - X_{n+h+1}^{(2)} \right)^2 \;\middle|\; \mathcal{F}_n \right) = 2\left(\mathrm{E}L_1^2\right) \\ +p^{-2}\mathrm{E}\left(\sum_{1\leq i\neq j\leq p} F_{n+h}^{-2\alpha}\left(X_{n+h}^{(i)} - X_{n+h}^{(j)} \right)^2 \;\middle|\; \mathcal{F}_n \right) + V(r,n,\alpha)F_n^{-v} \end{aligned}$$

Adding $\mathbf{1}_{\{F_n \geq \lambda\}}$ in the above equation, and taking expectations, we get, using the symmetry of the law of the $(X_m^{(i)})_{1 \leq i \leq p}$, that for $0 \leq h \leq r-1$:

$$a_n(\lambda, h+1) = \frac{p-1}{p} a_n(\lambda, h) + 2(\mathrm{E}L_1^2)\mathrm{P}\left[F_n \geq \lambda\right] + \mathrm{E}\big(V(r,n,\alpha)F_n^{-v}\mathbf{1}_{\{F_n \geq \lambda\}}\big)$$

Iterating, and using the fact that a_n is non-negative, we get that, for $1 \leq h \leq r-1$:

$$a_n(\lambda, h) \geq \sum_{k=0}^{h-1} \left(\frac{p-1}{p}\right)^k 2\mathrm{E}L_1^2\mathrm{P}\left[F_n \geq \lambda\right] + \mathrm{E}\big(V(r,n,\alpha)F_n^{-v}\mathbf{1}_{\{F_n \geq \lambda\}}\big)$$

Using the fact that $\sum_{k=0}^{+\infty} \left(\frac{p-1}{p}\right)^k = p$, and

$$Q_1(1,\alpha,p)p^2(p-1) + Q_2(1,\alpha,p) = (2p-3+2q(1-\alpha)) > 0$$

as soon as $p \geq 2$, we see that the inequality (2.4) implies that, for large enough r,

$$\mathrm{E}\delta_n(\lambda, r) \geq \epsilon\mathrm{P}\left[F_n \geq \lambda\right] + \mathrm{E}\big(V(r,n,\alpha)F_n^{-v}\mathbf{1}_{\{F_n \geq \lambda\}}\big)$$

Choose $s > 1$ close enough to 1 so that $-2(1/s-1) - v < 0$. By hölder's inequality,

$$\mathrm{E}\big(V(r,n,\alpha)F_n^{-v}\mathbf{1}_{\{F_n \geq \lambda\}}\big) \leq C(r,\alpha)(\mathrm{P}\left[F_n \geq \lambda\right])^{1/s}\lambda^{-v}$$

Letting n go to infinity, we have

$$\liminf_{n \to +\infty} \mathrm{E}\delta_n(\lambda, r) \geq \epsilon\nu([\lambda, +\infty)) + C(r,\alpha)(\nu([\lambda, +\infty)))^{1/s}\lambda^{-v}$$

Now, taking $\lambda = \Lambda_\ell$, where Λ_ℓ has been defined in step 1, one checks that

$$(\nu([\Lambda_\ell, +\infty)))^{1/s}\Lambda_\ell^{-v} = o(\nu([\Lambda_\ell, +\infty))$$

as ℓ goes to infinity, since

$$(\nu([\Lambda_\ell, +\infty)))^{1/s-1} \leq \Lambda_\ell^{-2(1/s-1)}$$

and $-2(1/s-1) - v < 0$, with λ_ℓ going to infinity as ℓ goes to infinity. Thus, taking ℓ large enough,

$$\liminf_{n \to +\infty} \mathrm{E}\delta_n(\Lambda_\ell, r) > 0.$$

$\square$

2.7 Convergence in law

Proof of Theorem 5:

We prove that, for all $q \geq 0$,

$$\mathrm{E}F_n^{2q(1-\alpha)} = \mu_q n^q + o(n^q),$$

where μ_q is the $2q-$th moment of a Bessel process of dimension $r(p,\alpha)$.

The result is proved by induction on q. For $q = 0$, it is trivial. Let us assume that this holds for all $0 \leq k \leq q-1$.

By Lemma 12,

$$\mathrm{E}\left(F_{n+1}^{2q(1-\alpha)} - F_n^{2q(1-\alpha)} \;\middle|\; \mathcal{F}_n\right) = Q_1(q,\alpha,p)F_n^{2(q-1)(1-\alpha)} \sum_{1\leq i<j\leq p} F_n^{-2\alpha}\left(X_n^{(i)} - X_n^{(j)}\right)^2 + \\ Q_2(q,\alpha,p)\left(\mathrm{E}L_1^2\right)F_n^{2(q-1)(1-\alpha)} + V(n,q,\alpha)F_n^{2(q-1)(1-\alpha)}F_n^{-v}$$

Taking expectations, and using the symmetry of the law of the $(X_n^{(i)})_{1\leq i\leq p}$,

$$\begin{aligned} \mathrm{E}\Big(F_{n+1}^{2q(1-\alpha)} - F_n^{2q(1-\alpha)}\Big) &= \\ (1/2)p(p-1)Q_1(q,\alpha,p)\mathrm{E}\Big(F_n^{2(q-1)(1-\alpha)}F_n^{-2\alpha}\left(X_n^{(1)} - X_n^{(2)}\right)^2\Big) \\ +Q_2(q,\alpha,p)(\mathrm{E}L_1^2)\mathrm{E}F_n^{2(q-1)(1-\alpha)} + \mathrm{E}\Big(V(n,q,\alpha)F_n^{2(q-1)(1-\alpha)}\Big) \end{aligned} \tag{2.5}$$

For $q = 1$, Hölder's inequality implies that:

$$\mathrm{E}\Big(V(n,q,\alpha)F_n^{2(q-1)(1-\alpha)}F_n^{-v}\Big) \leq C\left(\mathrm{E}\big(F_n^{-2v}\big)\right)^{1/2},$$

and, since $F_n \geq 1$ and goes to infinity in probability, this implies that

$$\mathrm{E}\Big(V(n,q,\alpha)F_n^{2(q-1)(1-\alpha)}F_n^{-v}\Big)$$

goes to zero as n goes to infinity. Now assume $q \geq 2$. By the induction hypothesis

$$\mathrm{E}F_n^{(2q-2)(1-\alpha)} \sim \mu_{q-1}n^{q-1}$$

Hölder's inequality implies that:

$$\mathrm{E}\Big(V(n,q,\alpha)F_n^{2(q-1)(1-\alpha)}F_n^{-v}\Big) \leq C\left(\mathrm{E}F_n^{(2q-2)(1-\alpha)}\right)^b,$$

where $b < 1$. Hence it is $o(n^{q-1})$ as n goes to infinity.

Define

$$c_n(q) = \mathrm{E}\left[F_n^{2(q-1)(1-\alpha)}F_n^{-2\alpha}\left(X_n^{(1)} - X_n^{(2)}\right)^2\right]$$

By Lemma 13, one has

$$\mathrm{E}\left(F_{n+1}^{2(q-1)(1-\alpha)}F_{n+1}^{-2\alpha}\left(X_{n+1}^{(1)}-X_{n+1}^{(2)}\right)^2 \;\middle|\; \mathcal{F}_n\right) = 2\left(\mathrm{E}L_1^2\right)F_n^{2(q-1)(1-\alpha)}$$
$$+p^{-2}F_n^{2(q-1)(1-\alpha)}\sum_{1\le i\ne j\le p}F_n^{-2\alpha}\left(X_n^{(i)}-X_n^{(j)}\right)^2+V(q,n,\alpha)F_n^{2(q-1)(1-\alpha)}F_n^{-v}$$

Taking expectations, and using the symmetry of the law of the $(X_n^{(i)})_{1\le i\le p}$,

$$c_{n+1}(q)=\frac{p-1}{p}c_n(q)+2\left(\mathrm{E}L_1^2\right)\mathrm{E}\Big(F_n^{2(q-1)(1-\alpha)}\Big)+\mathrm{E}\Big(V(n,q,\alpha)F_n^{2(q-1)(1-\alpha)}F_n^{-v}\Big)$$

As before, the last term in this expression is $o(n^{q-1})$.
Using the induction hypothesis, we find that:

$$c_{n+1}(q)=\frac{p-1}{p}c_n(q)+2\left(\mathrm{E}L_1^2\right)\mu_{q-1}n^{q-1}+o(n^{q-1})$$

As a consequence,

$$c_n=2p\left(\mathrm{E}L_1^2\right)\mu_{q-1}n^{q-1}+o(n^{q-1})$$

This, in turn, implies, with equation 2.5, that:

$$\mathrm{E}\Big(F_{n+1}^{2q(1-\alpha)}-F_n^{2q(1-\alpha)}\Big)=q(1-\alpha)\mu_{q-1}\left(\mathrm{E}L_1^2\right)(2p-3+2q(1-\alpha))n^{q-1}+o(n^{q-1})$$

We deduce that:

$$\mathrm{E}F_n^{2q(1-\alpha)}=(1-\alpha)\mu_{q-1}\left(\mathrm{E}L_1^2\right)(2p-3+2q(1-\alpha))n^{q}+o(n^{q})$$

Thus, we know explicitly the asymptotic behavior of the even moments of the positive random variable $n^{-1/2}F_n^{(1-\alpha)}$. They are easily identified as those of $(1-\alpha)(\mathrm{E}L^2)^{1/2}B(p,\alpha)$, where $B(p,\alpha)$ is a random variable whose law is the law at time 1 of a Bessel process of dimension $r(p,\alpha)$. Carleman's condition (see [29]) is easily checked and we conclude, by the method of moments, that $n^{-1/2}F_n^{(1-\alpha)}$ converges in law to $(1-\alpha)(\mathrm{E}L^2)^{1/2}B(p,\alpha)$. By Proposition 18, $d(X_n)/F_n$ goes to zero in probability, and this gives the convergence of $n^{-\beta}X_n$.

□

2.8 Technical lemmas

Lemma 10 *For all $n \geq 0$,*

$$|\Delta F_n| \leq d(X_n) + L_n^*(pF_n)^\alpha,$$

Proof:

One has:

$$|\Delta F_n| \leq |F_{n+1} - F'_n| + |F'_n - F_n|$$

For all $1 \leq i \leq p$,

$$\left|X_{n+1}^{(i)} - X_n'^{(i)}\right| \leq \left|L_n^{(i)}\right| \left(X_n'^{(i)}\right)^\alpha \leq L_n^*(pF_n)^\alpha$$

On the other hand, for all $1 \leq i \leq p$,

$$\left|X_n^{(i)} - X_n'^{(i)}\right| \leq d(X_n)$$

$\square$

Definition 4 *Let us define the event $D(n)$ by:*

- $d(X_n) \leq (1/4)F_n$
- $L_n^*(pF_n)^\alpha \leq (1/4)F_n$
- $F_n \geq 2(1 + L_n^*)^{2\beta}$

According to the inequality

$$|\Delta F_n| \leq d(X_n) + L_n^*(pF_n)^\alpha,$$

$D(n)$ implies that $|\Delta F_n| \leq (1/2)F_n$.

According to the inequality $\min_i X_n^{(i)} \geq F_n - d(X_n)$, $D(n)$ implies that

$$\min_i X_n^{(i)} > (1 + L_n^*)^{2\beta}$$

so, on $D(n)$,

$$X_{n+1}^i = X_n'^{(i)} + L_n^{(i)} \left(X_n'^{(i)}\right)^\alpha.$$

Lemma 11 *For all $q > 0$,*

$$\mathbf{1}_{D(n)^c} \leq F_n^{-q} V(n, q, \alpha)$$

Proof:

One has

$$\mathbf{1}_{\{d(X_n)\geq(1/4)F_n\}} \leq 4^q (d(X_n)/F_n)^q$$

and

$$\mathbf{1}_{\{L_n^*(pF_n)^\alpha\geq(1/4)F_n\}} \leq 4^q (pL_n^*)^q (F_n)^{q(\alpha-1)}$$

Likewise,

$$\mathbf{1}_{\left\{F_n\leq 2(1+L_n^*)^{2\beta}\right\}} \leq 2^q F_n^{-q}(1+L_n^*)^{2q\beta}$$

Thanks to Proposition 18 and to the fact that L_n^* has moments of all orders, for all q, we get the stated inequality. □

Lemma 12

$$\mathrm{E}\left(F_{n+1}^{2q(1-\alpha)} - F_n^{2q(1-\alpha)} \;\middle|\; \mathcal{F}_n\right) = Q_1(q,\alpha,p)F_n^{2(q-1)(1-\alpha)} \sum_{1\leq i<j\leq p} F_n^{-2\alpha}\left(X_n^{(i)} - X_n^{(j)}\right)^2 + Q_2(q,\alpha,p)\left(\mathrm{E}L_1^2\right)F_n^{2(q-1)(1-\alpha)} + V(n,q,\alpha)F_n^{2(q-1)(1-\alpha)}F_n^{-v}$$

where

$$Q_1(q,\alpha,p) = 2q(1-\alpha)p^{-2} + q(1-\alpha)(2q(1-\alpha)-1)p^{-3}$$

and

$$Q_2(q,\alpha,p) = q(1-\alpha)(2q(1-\alpha)-1)p^{-1}$$

Proof:

One has

$$F_{n+1}^{2q(1-\alpha)} - F_n^{2q(1-\alpha)} = F_n^{2q(1-\alpha)}\left((1+\Delta F_n/F_n)^{2q(1-\alpha)} - 1\right)$$

From Lemma 11 above,

$$F_{n+1}^{2q(1-\alpha)}\mathbf{1}_{D(n)^c} \leq V(q,n,\alpha)F_n^{-1}\ ,\ F_n^{2q(1-\alpha)}\mathbf{1}_{D(n)^c} \leq V(q,n,\alpha)F_n^{-1}$$

For $|x|\leq 1/2$, a Taylor expansion shows that

$$\left|(1+x)^{2q(1-\alpha)} - 1 - 2q(1-\alpha)x - q(1-\alpha)(2q(1-\alpha)-1)x^2\right| \leq C|x|^3 \tag{2.6}$$

We have, by Lemma 10,

$$|\Delta F_n|^3 \leq Cd(X_n)^3 + CL_n^{*3}(F_n)^{3\alpha}$$

Thus, using Proposition 18 and the fact that L_n^* has moments of all orders, we deduce that:

$$F_n^{2q(1-\alpha)}|\Delta F_n/F_n|^3 \leq V(q,n,\alpha)F_n^{(2q-3)(1-\alpha)}$$

so

$$\mathrm{E}\left(F_{n+1}^{2q(1-\alpha)} - F_n^{2q(1-\alpha)} \;\middle|\; \mathcal{F}_n\right) = 2q(1-\alpha)\gamma^{(1)} + q(1-\alpha)(2q(1-\alpha)-1)\gamma^{(2)} \\ + V(q,n,\alpha)F_n^{2(q-1)(1-\alpha)}F_n^{-(1-\alpha)}$$

where

$$\gamma^{(1)} = \mathrm{E}\left(F_n^{2q(1-\alpha)}\left(\frac{\Delta F_n}{F_n}\right)\mathbf{1}_{D(n)} \;\middle|\; \mathcal{F}_n\right)$$

$$\gamma^{(2)} = \mathrm{E}\left(F_n^{2q(1-\alpha)}\left(\frac{\Delta F_n}{F_n}\right)^2\mathbf{1}_{D(n)} \;\middle|\; \mathcal{F}_n\right)$$

Let us first study $\gamma^{(1)}$.
Write $\Delta F_n = F_{n+1} - F'_n + F'_n - F_n$.
On $D(n)$,

$$F_{n+1} - F'_n = p^{-1}\sum_{i=1}^{p} L_n^{(i)}(X_n'^{(i)})^{\alpha}$$

Thus,

$$\gamma^{(1)} = \mathrm{E}\left(F_n^{2q(1-\alpha)}\left(p^{-1}\sum_{i=1}^{p} L_n^{(i)}(X_n'^{(i)})^{\alpha} + F'_n - F_n\right)F_n^{-1}\mathbf{1}_{D(n)} \;\middle|\; \mathcal{F}_n\right)$$

Moreover, the error commited when removing $\mathbf{1}_{D(n)}$ in the above conditional expectation is less than $V(q,n,\alpha)F_n^{-1}$ by Lemma 11.

On the other hand, a simple computation shows that

$$\mathrm{E}\left(F'_n - F_n \;\middle|\; \mathcal{F}_n\right) = (p^2F_n)^{-1}\sum_{1\le i<j\le p}\left(X_n^{(i)} - X_n^{(j)}\right)^2$$

Conditioning by $\mathcal{F}'_n$ and using the fact that the $L_n^{(i)}$ are centered random variables independent from $\mathcal{F}'_n$, we see that

$$\mathrm{E}\left(F_n^{2q(1-\alpha)}\left(p^{-1}\sum_{i=1}^{p} L_n^{(i)}(X_n'^{(i)})^{\alpha}\right)F_n^{-1} \;\middle|\; \mathcal{F}_n\right) = 0 \; a.s.$$

Hence

$$\gamma^{(1)} = p^{-2}F_n^{2(q-1)(1-\alpha)}\sum_{1\le i<j\le p} F_n^{-2\alpha}\left(X_n^{(i)} - X_n^{(j)}\right)^2 + V(q,n,\alpha)F_n^{-1}$$

Let us now study $\gamma^{(2)}$.
Write $(\Delta F_n)^2 = (F_{n+1} - F'_n)^2 + 2(F_{n+1} - F'_n)(F'_n - F_n) + (F'_n - F_n)^2$

Let us first study the contribution of the first two terms in $\gamma^{(2)}$.

On $D(n)$,

$$F_{n+1} - F'_n = p^{-1} \sum_{i=1}^{p} L_n^{(i)} (X_n'^{(i)})^{\alpha}$$

From Lemma 11, we have, as before,

$$\mathrm{E}\left(F_n^{2q(1-\alpha)-2} (F_{n+1} - F'_n)(F'_n - F_n) \mathbf{1}_{D(n)} \;\middle|\; \mathcal{F}_n \right) =$$

$$\sum_{i=1}^{p} \mathrm{E}\left(F_n^{2q(1-\alpha)-2} (p^{-1} L_n^{(i)} (X_n'^{(i)})^{\alpha})(F'_n - F_n) \;\middle|\; \mathcal{F}_n \right) + V(q,n,\alpha) F_n^{-1}$$

Using the fact that the $L_n^{(i)}$ are centered random variables independent from $\mathcal{F}'_n$, we see that the conditional expectations in the right-hand side of the above equation are zero.

Expanding:

$$\mathrm{E}\left(F_n^{2q(1-\alpha)-2} (F_{n+1} - F'_n)^2 \mathbf{1}_{D(n)} \;\middle|\; \mathcal{F}_n \right) =$$

$$p^{-2} \sum_{1 \le i,j \le p} \mathrm{E}\left(F_n^{2q(1-\alpha)-2} L_n^{(i)} L_n^{(j)} (X_n'^{(i)})^{\alpha} (X_n'^{(j)})^{\alpha} \mathbf{1}_{D(n)} \;\middle|\; \mathcal{F}_n \right)$$

As before, since, for $i \neq j$, $L_n^{(i)} L_n^{(j)}$ is centered an d independent from $\mathcal{F}_n$,

$$\sum_{1 \le i \neq j \le p} \mathrm{E}\left(F_n^{2q(1-\alpha)-2} L_n^{(i)} L_n^{(j)} (X_n'^{(i)})^{\alpha} (X_n'^{(j)})^{\alpha} \mathbf{1}_{D(n)} \;\middle|\; \mathcal{F}_n \right) = V(q,n,\alpha) F_n^{-1}$$

On $D(n)$, for all $1 \le i \le p$,

$$\left| (X_n'^{(i)})^{2\alpha} - F_n^{2\alpha} \right| \le C F_n^{2\alpha} d(X_n)/F_n = V(n) F_n^{\alpha}.$$

Thus the error commited when replacing $X_n'^{(i)}$ by F_n in

$$\mathrm{E}\left(F_n^{2q(1-\alpha)-2} \left(L_n^{(i)} (X_n'^{(i)})^{\alpha} \right)^2 \mathbf{1}_{D(n)} \;\middle|\; \mathcal{F}_n \right)$$

is less than:

$$V(n) F_n^{2(q-1)(1-\alpha)-\alpha}$$

Moreover, the error commited when removing $\mathbf{1}_{D(n)}$ in the above conditional expectations is less than $V(q,n,\alpha)F_n^{-1}$.

Hence:

$$\mathrm{E}\left(F_n^{2(q-1)\alpha-2} (F_{n+1} - F'_n)^2 \mathbf{1}_{D(n)} \;\middle|\; \mathcal{F}_n \right) = p^{-1} (\mathrm{E} L_1^2) \left(F_n^{2(q-1)\alpha} \right) + F_n^{2(q-1)(1-\alpha)} F_n^{-v}$$

Let us now study the last term (as before, we remove the indicator $\mathbf{1}_{D(n)}$ and commit an error less than $V(q,n,\alpha)F_n^{-1}$):

$$\begin{aligned}\mathrm{E}\left((F_n' - F_n)^2 \;\middle|\; \mathcal{F}_n\right) &= p^{-2} \sum_{1\le i\ne j\le p} H_{i,j}(X_n) \\ &\quad + p^{-2} \sum_{1\le i\le p} K_i(X_n),\end{aligned}$$

with the following notations:

$$\begin{aligned}H_{i,j}(X_n) &= \sum_{1\le a,b\le p} \frac{X_n^{(a)}X_n^{(b)}}{S_n^2}(X_n^{(a)} - X_n^{(i)})(X_n^{(b)} - X_n^{(j)}), \\ K_i(X_n) &= \sum_{1\le a\le p} \frac{X_n^{(a)}}{S_n}(X_n^{(a)} - X_n^{(i)})^2.\end{aligned}$$

Let us study the error committed when replacing the transition probabilities of the selection step by uniform ones in $H_{i,j}$ and K_i. We have:

$$\left|\frac{X_n^{(a)}}{S_n} - 1\right| \le \frac{d(X_n)}{pF_n}$$

$$\left|\frac{X_n^{(a)}X_n^{(b)}}{S_n^2} - 1\right| \le 2\frac{d(X_n)}{pF_n}$$

Thus, replacing in

$$\mathrm{E}\left((F_n' - F_n)^2 \;\middle|\; \mathcal{F}_n\right) F_n^{2(q-1)(1-\alpha)} F_n^{-2\alpha}\mathbf{1}_{D(n)}$$

the error terms are bounded by:

$$d(X_n)^3 F_n^{-3\alpha} F_n^{2(q-1)(1-\alpha)} F_n^{-(1-\alpha)} = V(q,n,\alpha) F_n^{2(q-1)(1-\alpha)} F_n^{-(1-\alpha)}$$

For an arbitrary sequence $(x_i)_{1\le i\le p}$,

$$\begin{aligned}&\sum_{1\le i\ne j\le p}\;\sum_{1\le a,b\le p} (x_a - x_i)(x_b - x_j) \\ &= \sum_{1\le a,b,i,j\le p} (x_a - x_i)(x_b - x_j) - \sum_{i=j,a,b} (x_a - x_i)(x_b - x_j)\end{aligned}$$

By symmetry, the first sum in the right-hand side is zero. As for the last sum, we separate cases $a = b$ and $a \ne b$:

$$\begin{aligned}&\sum_{1\le i\ne j\le p}\;\sum_{1\le a,b\le p} (x_a - x_i)(x_b - x_j) \\ &= -\sum_{i,a} (x_a - x_i)^2 - \sum_{i,a\ne b} (x_a - x_i)(x_b - x_i)\end{aligned}$$

In the last sum, let (u,v,w) be the ordered triple i,a,b. With this notation, the last equation reads:

$$\begin{aligned}\sum_{1\le i\ne j\le p}\ \sum_{1\le a,b\le p}(x_a-x_i)(x_b-x_j)\\ = -\sum_{i,a}(x_a-x_i)^2-2\sum_{1\le u<v<w\le p} & (x_v-x_u)(x_w-x_u)\\ & +(x_u-x_v)(x_w-x_v)\\ & +(x_u-x_w)(x_v-x_w).\end{aligned}$$

Note that, for all x,y,z,

$$2\left[(y-x)(z-x)+(x-y)(z-y)+(x-z)(y-z)\right]=(x-y)^2+(y-z)^2+(x-z)^2,$$

whence:

$$\begin{aligned}\sum_{1\le i\ne j\le p}\ \sum_{1\le a,b\le p}(x_a-x_i)(x_b-x_j)\\ = -\sum_{i,a}(x_a-x_i)^2-\sum_{1\le u<v<w\le p} & (x_u-x_v)^2\\ & +(x_v-x_w)^2\\ & +(x_u-x_w)^2\end{aligned}$$

Thus, replacing the transition probabilities of the selection step by uniform ones yields the stated result.

□

Lemma 13

$$\mathrm{E}\left(F_{n+1}^{2(q-1)(1-\alpha)}F_{n+1}^{-2\alpha}\left(X_{n+1}^{(1)}-X_{n+1}^{(2)}\right)^2\ \middle|\ \mathcal{F}_n\right)=2\left(\mathrm{E}L_1^2\right)F_n^{2(q-1)(1-\alpha)}$$
$$+p^{-2}F_n^{2(q-1)(1-\alpha)}\sum_{1\le i\ne j\le p}F_n^{-2\alpha}\left(X_n^{(i)}-X_n^{(j)}\right)^2+V(q,n,\alpha)F_n^{2(q-1)(1-\alpha)}F_n^{-v}$$

Proof:

As before,

$$\mathrm{E}\left(F_{n+1}^{2(q-1)(1-\alpha)}F_{n+1}^{-2\alpha}\left(X_{n+1}^{(1)}-X_{n+1}^{(2)}\right)^2\ \middle|\ \mathcal{F}_n\right)=$$
$$\mathrm{E}\left(F_{n+1}^{2(q-1)(1-\alpha)}F_{n+1}^{-2\alpha}\left(X_{n+1}^{(1)}-X_{n+1}^{(2)}\right)^2\mathbf{1}_{D(n)}\ \middle|\ \mathcal{F}_n\right)+V(n,q,\alpha)F_n^{2q(1-\alpha)}F_n^{-v}$$

On $D(n)$,

$$X_{n+1}^{(1)}-X_{n+1}^{(2)}=X_n'^{(1)}+L_n^{(1)}\left(X_n'^{(1)}\right)^\alpha-X_n'^{(2)}+L_n^{(2)}\left(X_n'^{(2)}\right)^\alpha$$

Whence

$$\left(X^{(1)}_{n+1}-X^{(2)}_{n+1}\right)^2=\left(X'^{(1)}_n-X'^{(2)}_n\right)^2+\left(L^{(1)}_n\left(X'^{(1)}_n\right)^\alpha-L^{(2)}_n\left(X'^{(2)}_n\right)^\alpha\right)^2+$$
$$2\left(L^{(1)}_n(X'^{(1)}_n)^\alpha-L^{(2)}_n(X'^{(2)}_n)^\alpha\right)\left(X'^{(1)}_n-X'^{(2)}_n\right)$$

Let us first study the contribution of the last term. As before, we commit an error less than $V(q,n,\alpha)F_n^{-1}$ when removing $\mathbf{1}_{D(n)}$ from the above conditional expectations. Conditioning by $\mathcal{F}'_n$ and using the fact that the $L^{(i)}_n$ are centered random variables independent from $\mathcal{F}'_n$, we deduce that the contribution of this last term is less than $V(q,n,\alpha)F_n^{-1}$.

As for the contribution of the previous term, we write:

$$\left(L^{(1)}_n\left(X'^{(1)}_n\right)^\alpha-L^{(2)}_n\left(X'^{(2)}_n\right)^\alpha\right)^2=\left(L^{(1)}_n\left(X'^{(1)}_n\right)^\alpha\right)^2+$$
$$\left(L^{(2)}_n\left(X'^{(2)}_n\right)^\alpha\right)^2+2L^{(1)}_n\left(X'^{(1)}_n\right)^\alpha L^{(2)}_n\left(X'^{(2)}_n\right)^\alpha$$

Just as before, the contribution of the last term is $V(q,n,\alpha)F_n^{-1}$. Using the fact that, on $D(n)$, for all $1\le i\le p$,

$$F_{n+1}^{-2\alpha}\left|\left(X'^{(i)}_n\right)^{2\alpha}-F_{n+1}^{2\alpha}\right|\le C(d(X_n)/F_n+F_n^{-\alpha}),$$

and since the $L^{(i)}_n$ are independent from $\mathcal{F}'_n$, we see that the contribution of the two previous terms is

$$\left(F_n^{2(q-1)(1-\alpha)}\left(2\mathrm{E}L_1^2\right)\right)+V(q,n,\alpha)F_n^{2(q-1)(1-\alpha)}F_n^{-v}$$

Now,

$$\mathrm{E}\left((X'^{(1)}_n-X'^{(2)}_n)^2\;\middle|\;\mathcal{F}_n\right)=\sum_{1\le i\ne j\le p}\frac{X^{(i)}_nX^{(j)}_n}{S_n^2}\left(X^{(i)}_n-X^{(j)}_n\right)^2$$

Replacing the transition probabilities of the selection step by uniform ones, as in the proof of the previous lemma, we get the stated result.

□

Lemma 14 *For all $h\ge 1$, and all $q\ge 1$,*

$$\mathbf{1}_{\{F_{n+h}\le(1/2)F_n\}}\le V(q,n,h,\alpha)F_n^{-q}$$

Proof:

Note that, as in the proof of Lemma 11, for all $l\ge 1$,

$$\mathbf{1}_{\left\{|\Delta F_l|\geq(1-(1/2)^{1/h})F_l\right\}} \leq V(l,h,\alpha)F_l^{-q},$$

so that, on

$$\bigcap_{l=n}^{n+k-1}\left\{|\Delta F_l| < (1-(1/2)^{1/h})F_l\right\},$$

$$\mathbf{1}_{\left\{|\Delta F_{n+k}|\geq(1-(1/2)^{1/h})F_{n+k}\right\}} \leq V(l,n,h,\alpha)((1/2)^{k/h}F_n)^{-q}.$$

□

Lemma 15

$$d(X_{n+1}) \leq d(X'_n) + 2L^*_n\left(\max_{1\leq i\leq p} X'^{(i)}_n\right)^{\alpha},$$

Proof:

$$X^{(i)}_{n+1} = \max\left(\left(X'^{(i)}_n + L_i\left(X'^{(i)}_n\right)^{\alpha}\right) 1\right),$$

hence:

$$\begin{aligned} d(X_{n+1}) &\leq d(X'_n) + d\left(\left(L_i\left(X'^{(i)}_n\right)^{\alpha}\right)_{1\leq i\leq p}\right) \\ &\leq d(X'_n) + 2\max_{1\leq i\leq p}\left|L_i\left(X'^{(i)}_n\right)^{\alpha}\right| \\ &\leq d(X'_n) + 2\left(\max_{1\leq i\leq p}|L_i|\right)\left(\max_{1\leq i\leq p} X'^{(i)}_n\right)^{\alpha}. \end{aligned}$$

□

Definition 5 *We say that n is a crunch time if $d(X'_n) = 0$*

Definition 6 *We say that a function $s : [1,+\infty)^p \to \mathbb{R}_+$ is contracting under selection if:*

- *$s(x) = 0$ whenever $d(x) = 0$*
- *$s(y) \leq F(x)$ whenever $\{y_i\} \subset \{x_i\}$*

Two useful examples of such functions are $s(x) = d(x)^q$ and $s(x) = d(x)^q(\min x)^{-\alpha q}$ with $q > 0$. The following lemma justifies the term ¨contraction under selection¨

Lemma 16 *If s is contracting under selection,*

$$\mathrm{E}\left(s(X'_n)\;\middle|\;\mathcal{F}_n\right) \leq \kappa s(X_n), \quad \textit{where} \quad \kappa = 1 - p^{1-p} < 1.$$

Proof:

Let A be the event ¨n is a crunch time¨. One has $\mathrm{P}\left[A \mid \mathcal{F}_n\right] \geq p^{1-p}$. On A, $s(X'_n) = 0$. On A^c, $s(X'_n) \leq s(X_n)$. □

Lemma 17

$$\min_i X_{n+1}^{(i)} \geq \max\left(\left(\min_i X_n'^{(i)} - L_n^* \left(\min_i X_n'^{(i)}\right)^\alpha\right), 1\right).$$

Proof:

Recall that

$$X_{n+1}^{(i)} = \max\left(\left(X_n'^{(i)} + L_i \left(X_n'^{(i)}\right)^\alpha\right), 1\right).$$

Let $i \in \arg\min_j X_n'^{(j)}$.

$$X_{n+1}^{(i)} = X_n'^{(i)} + L_i \left(X_n'^{(i)}\right)^\alpha \geq X_n'^{(i)} - L_n^* \left(X_n'^{(i)}\right)^\alpha.$$

Set $x = X_n'^{(i)}$, and $y = X_n'^{(j)}$ for $j \neq i$. We have $1 \leq x \leq y$. Let us consider the function $f : u \mapsto u - L_n^* u^\alpha$. We note that, for $u > 0$,

$$f'(u) = 1 - \alpha L_n^* u^{\alpha-1} \geq 1 - L_n^* u^{\alpha-1}$$

Thus, if $x - L_n^* x^\alpha \geq 1$, f is non-decreasing on $[x, +\infty)$, so:

$$X_{n+1}^{(j)} \geq f(y) \geq f(x) = \left(X_n'^{(i)} - L_n^* \left(X_n'^{(i)}\right)^\alpha\right) \geq 1$$

If $x - L_n^* x^\alpha \leq 1$, we have by definition $X_{n+1}^{(j)} \geq 1$.

□

Lemma 18 *For all positive real numbers M, η, m such that $m \geq 1$, $m - Mm^\alpha \geq 1$ and all positive integer q, there exists $\theta_{\alpha,q,\eta}$ such that*

$$m \geq \left(\frac{M}{\theta_{\alpha,q,\eta}}\right)^{2\beta} \quad \Longrightarrow \quad \left|\frac{1}{m^{\alpha q}} - \frac{1}{(m - Mm^\alpha)^{\alpha q}}\right| \leq \frac{\eta}{m^{\alpha q}}$$

Lemma 19 *If (u_n) is a sequence of non-negative real numbers satisfying:*

$$u_{n+1} \leq u_n + C_1 u_n^\alpha + C_2$$

then $u_n = O\left(n^{2\beta}\right)$

Proof:

Let (u_n) satisfy the inequality above. As n goes to infinity,

$$(n+1)^{2\beta} = n^{2\beta} + (2\beta + o(1))n^{2\alpha\beta}$$

Thus, if $u_n \leq An^{2\beta}$

$$u_{n+1} \leq An^{2\beta} + C_1 A^{\alpha} n^{2\alpha\beta} + C_2$$

Thus, for all large A and n, $u_n \leq An^{2\beta}$ implies that

$$u_{n+1} \leq A(n+1)^{2\beta}$$

$\square$

Chapitre 3

Niching in Monte Carlo Filtering Algorithms

3.1 Abstract

Nonlinear multimodal filtering problems are usually addressed via Monte Carlo algorithms. These algorithms involve sampling procedures that are similar to proportional selection in genetic algorithms, and that are prone to failure due to genetic drift. We investigate the feasibility and the relevance of niching strategies in this context. Sharing methods are evaluated experimentally, and prove to be efficient in such issues. This work is a collaboration with Alexis Bienvenüe, Éric Fontenas, Olivier François and Marc Joannidès.

3.2 Introduction

In evolutionary computation, genetic drift is often considered as a source of premature convergence. Given a problem with multiple solutions, a genetic algorithm (GA) will at best ultimately converge to a population containing only one of these solutions. This phenomenon has been observed in natural as well as in artificial evolution, and is undesirable in many applications (e.g. multi-objective optimization). To overcome the above problem, several methods have been proposed that take their inspiration from mathematical ecology [35]. GA were developed that are capable of forming and maintaining stable subpopulations, or *niches*. GAs which employ *niching* mechanisms become capable of finding multiple solutions to a problem, within a single population [18], [33], [40]. Among these methods, the most popular is *fitness sharing*, that works by modifying the objective function according to the presence of nearby individuals.

Beyond the field of evolutionary computation, similar phenomena have been observed in Monte Carlo strategies such as iterated bootstrap or particle filtering [53], [54]. Such algorithms are based on selection procedures as well, and are closely connected to the traditional GA framework (see Section 3.3). These strategies have proved their efficiency in high-dimensional nonlinear problems. In terrain navigation for instance, an aircraft measures its relative elevation sequentially, and the goal of filtering is to estimate the position and velocity of the aircraft. This problem is multimodal as several positions might correspond to a single relative elevation. While Monte Carlo strategies are theoretically able to simulate the true distribution of the aircraft position, selection often concentrates the solutions on a single mode leading to wrong decisions.

While the benefit of sharing methods has been intensively studied by the EC community, few efforts have been devoted to the other contexts. This work evaluates the feasibility of sharing methods in Monte Carlo nonlinear filtering algorithms. Section 3.3 presents an account on the filtering problem and Monte Carlo methods. Section 3.4 introduces niching strategies in sampling procedures for particle filters and discusses the choice of a sharing bandwidth. In Section 3.5, the algorithm is evaluated on a set of one-dimensional test problems similar to those encountered in terrain navigation. For these simple models, the solution to the filtering problems can be computed exactly. On this basis, the niching method is compared to the standard algorithm, and proves to be beneficial in this context.

3.3 Monte Carlo Filtering

Filtering addresses the issue of predicting an unknown signal (X_t) given noisy observations of this signal. Mathematically, the signal is modelled as a Markov process taking values in some measurable space $\mathcal{X}$

$$X_t = F(X_{t-1}) + V_t, \quad t \geq 1 \tag{3.1}$$

where F is a deterministic function and (V_t) is a sequence of independent identically distributed centered random variables. More generally, such dynamics can be specified according to some Markov kernel $Q(x, dx)$ that describes the transition probabilities between successive states. The distribution of X_t is often called the *prior distribution*.

In the filtering problem, the signal cannot be observed directly. Instead, data are (indirectly) gathered from the observation of a second signal

$$Y_t = H(X_t) + W_t, \quad t \geq 0, \tag{3.2}$$

where H is usually a nonlinear function and W_t a noisy variable independent from X_t.

The filtering problem consists of making predictions about the original signal X_t given the observations Y_0, Y_1, ..., Y_t. This amounts to estimating (or computing) the conditional distribution of X_t given these observations. This distribution is called the *posterior distribution*.

Filtering has an old tradition that goes back to the seminal paper by Kalman [42]. The standard approach assumes that F and H are linear and that V_t and W_t are Gaussian random variables of known covariance matrices. In contrast, solving nonlinear filtering problems turns out to be particularly difficult, and the difficulty is even increased when the signal becomes unidentifiable (e.g., H not invertible). In such a case, the posterior distribution may be multimodal. Kalman filters would therefore lead to erroneous predictions since these methods always predict a single mode.

Kunita and Stettner [46] developed general recursion schemes that compute the exact solution of the filtering problem based on Bayes' formula. Despite their closed form, these equations are hardly of practical interest since numerical computations of high-dimensional integrals are involved.

Monte Carlo filtering is an algorithmic alternative to Kunita-Stettner recursion. It consists of a computer intensive technique, and can be useful where linear filtering fails. This method is based on a *particle system approach* [37], [19], [27] in which the posterior distribution is computed empirically. In this approach, a population of n particles is evolved in the signal space according to a two-stages procedure. More precisely, let $x_t = (x_1, \ldots, x_n)$ be the population at time t (x_0 is randomly initialized). The two steps are iterated as follows.

1) **Prediction.** Create n new particles x'_1, ..., x'_n by sampling from the transition kernel $Q(x, dx)$

$$x'_i \sim Q(x_i, dx), \quad i = 1, \ldots, n.$$

Conditional to $x_1, \ldots, x_n$, the new particles are independent.

2) **Correction.** Resample the particles according to a proportional scheme taking the observation y_t (at time t) into account

$$x''_i \sim \frac{L_t(y_t - H(x'_i))}{\sum_j L_t(y_t - H(x'_j))}$$

where L_t is the likelihood function of the observation noise W_t. Define the population at time $t+1$ as being $x_{t+1} = x''$.

The convergence of this algorithm to the optimal solution of the filtering problem has been proved in [19], when the population size goes to infinity.

Turning to an evolutionary computation perspective, there is a close connection between Monte Carlo filters and the simple GA without recombination. This connection has been emphasized in previous works by [19], [22]. In the above algorithm, particles can be identified as individuals in a population, where the set of phenotypes corresponds to the possible states of the signal. The first step, called *prediction*, is similar to the mutation step in GAs. Each individual generates an offspring by mutation from the kernel $Q(x, dx)$ (and the offspring replaces its parent). In the second step, called *correction*, a random selection of the offspring is performed. The selection strategy is similar to the proportional selection scheme used in GAs. However, the fitness function is time-dependent as it must account for the information contained in the data at each instant. Mathematically, the fitness of offspring x_i can actually be defined as

$$f_t(x_i) = L_t(y_t - H(x_i)). \tag{3.3}$$

3.4 Niching in a filtering context

3.4.1 Niching algorithms

The reason why Monte Carlo filtering methods should work is that their infinite-population models correspond to Kunita's recursion scheme precisely. However, the shortcomings of infinite-population models are well-known. By their very nature, they may not truly reflect the finite-population properties that are of major interest to a practitioner. For instance, the effect of stochastic fluctuations during the correction step are neglected in this approach.

The same kind of remark also holds for the traditional GA. To overcome the above problem, Goldberg and Richardson proposed a method based on sharing [35]. These methods require that the objective *fitness* function be shared as a single resource among similar individuals in a population. Niching is achieved by degrading the objective function (i.e., the unshared fitness) of an individual according to the presence of nearby individuals. This type of niching requires a distance metric on the phenotype of the individuals. The objective fitness $f(i)$ of an individual i is degraded by first summing all of the shared values of individuals within a fixed bandwidth σ_{sh} of that individual and then dividing $f(i)$ by this sum, which is known as the *niche count*. More specifically, if two individuals are separated by distance $d(i, j)$, then a shared value is added

$$\text{sh}(d(i,j)) = \begin{cases} 1 - (\frac{d(i,j)}{\sigma_{\text{sh}}})^\alpha & \text{if } d(i,j) < \sigma_{\text{sh}} \\ 0 & \text{otherwise.} \end{cases} \tag{3.4}$$

to the niche count

$$m(i) = \sum_{j=1}^{n} \text{sh}(d(i,j)). \tag{3.5}$$

The parameters σ_{sh} and α are chosen by the user of the niched GA based on some a priori knowledge of the fitness landscape. The parameter α is often set to one, yielding the triangular sharing function. Finally, the shared fitness is defined as

$$f'(i) = \frac{f(i)}{m(i)}. \tag{3.6}$$

The actual fitness of each individual is modulated according to the density of the population around it: the fitness of isolated individuals is increased, while that of individuals in well represented regions of the search space is decreased.

3.4.2 Sharing methods in Monte Carlo filters

In this paper, we investigate the maintenance of stable subpopulations in Monte Carlo filtering algorithms using the method of sharing function. Since the available computational resources do not allow the number of particles to be arbitrarily large, standard Monte Carlo filters often suffer from a loss of diversity due to the stochastic nature of resampling. In real-world applications (e.g. real-time target tracking algorithms), this premature loss of diversity implies loosing the signal for some time by concentrating all individuals in a possibly wrong region of the search space. Maintaining stable niches in Monte Carlo filter is therefore a crucial point, since these niches actually correspond to existing modes of the posterior distribution. A niching procedure can be included in Monte Carlo filters as follows.

1) **Prediction (unchanged).** Create n new particles $x'_1, \ldots, x'_n$ by sampling from the transition kernel $Q(x, dx)$

$$x'_i \sim Q(x_i, dx), \quad i = 1, \ldots, n.$$

Conditional to $x_1, \ldots, x_n$, the new particles are independent.

2) **Correction.** Resample the particles according to a proportional scheme taking the observations into account

$$x''_i \sim \frac{f'_t(x'_i)}{\sum_j f'_t(x'_j)}$$

where $f'_t(x'_i)$ is the shared fitness of $f_t(x'_i) = L_t(y_t - H(x'_i))$.

In implementing this algorithm, choosing the bandwidth σ_{sh} is a critical step. Deb's rule sets this parameter by taking into account distances between peaks, and relative fitnesses. Specifically,

$$\sigma_{\mathrm{sh}} = \min_{i,j} \left\{ \frac{d(x'_i, x'_j)}{1 - r_{ij}} \right\} \tag{3.7}$$

where

$$r_{ij} = \min\left(\frac{f_t(x'_j)}{f_t(x'_i)}, \frac{f_t(x'_i)}{f_t(x'_j)}\right) \tag{3.8}$$

and the metric d corresponds to the Euclidean Distance. In one-dimensional filtering problems, a better-supported rule is given by Silverman [68] inspired from density estimation

$$\sigma_{\text{sh}} = 0.9 \min\left(\text{sd}(x'), \text{iqr}(x')/1.34\right) \, n^{-0.2}.$$

This rule is based on the minimum of the standard deviation of x' and its interquartile range divided by 1.34. Using this rule is quite natural in a niching context. Sharing indeed degrades the fitness function by dividing by the density of nearby particles and actually involves an estimation of this density. Note that a proper use of this rule requires that the sharing method be build upon a Gaussian kernel instead the triangular function. Similar rules also exist in higher dimensions.

3.5 Experiments

3.5.1 Test problems

Evaluating the impact of the sharing method in Monte Carlo filters is difficult in general. A number of test problems have been chosen to assess the performances empirically. The selection of six test problems is inspired from target tracking issues. Three noisy dynamical systems describe the motion of a target in one dimension. The first motion is the classical $AR(1)$ dynamics [42]

$$X_t = 0.9X_{t-1} + V_t, \quad V_t \sim \mathcal{N}(0,1) \tag{3.9}$$

where the (random) initial condition X_0 is sampled according to the Gaussian distribution $\mathcal{N}(0, 5.26315)$ (so that the signal is stationary). The second motion is called the *piecewise linear dynamics*, and can be described as

$$X_t = X_{t-1} - 0.1\,\text{sign}(X_{t-1}) + V_t, \quad V_t \sim \mathcal{N}(0,1) \tag{3.10}$$

and $X_0 = 0$. The third motion is called the *double well dynamics* [52]

$$X_t = X_{t-1} - 0.04X_{t-1}(X_{t-1}^2 - 1) + V_t, \quad V_t \sim \mathcal{N}(0, 0.01 \times q) \tag{3.11}$$

where q is a parameter set to 0.24 in [52], and $X_0 = 0$. In addition, these three motions are observed through different functions. The first observation function is a symmetric one

$$H(x) = |x|, \tag{3.12}$$

and the second is non-symmetric

$$H(x) = \begin{cases} 2x & \text{if } x \geq 0\,; \\ -x/2 & \text{if } x \leq 0\,. \end{cases} \tag{3.13}$$

The observation noise is standard Gaussian $W_t = \mathcal{N}(0, \sigma^2)$ (σ is often set to 1). The length of simulation runs is equal to $T = 100$. Regarding the symmetric observation function, the posterior distribution is bimodal while this is not always the case for the non-symmetric function. For the six problems, the posterior distribution can be computed exactly using Kunita's recursions. Knowing the exact solution will be useful in assessing the accuracy of the filtering procedures during the experiments.

Figure 3.1 displays a typical trajectory from the double well dynamics (a) and the exact posterior distributions computed via Kunita's recursions under a symmetric observation function (b). A population of 20 individuals is evolved using the classical Monte Carlo procedure (c) and the niching algorithm (d).

3.5.2 Experimental design

Simulation runs contain simulated trajectories of the signal motion and the corresponding observations. For each of the six models, simulations are repeated 100 times so that the performances can be evaluated statistically. In the experimental design, the following parameter settings are experimented.

NS	No Sharing
Sil	Silverman's rule for the sharing bandwidth
Deb	Deb's rule for the sharing bandwidth
σ_{sh}	constant sharing bandwidth (values = 0.1, 1, 10)

The variable NS means that no sharing is used. The algorithm corresponds to the classical Monte Carlo filtering method. The next levels indicate how the sharing bandwidth has been set up: Silverman's rule, Deb's rule or constant. Except for Silverman's rule, the triangular function is chosen (parameter sh = 1 in Figure 3.2). Other choices were tested but the results did not change significantly.

3.5.3 Performance measures

As shown by Figure 3.1, sharing can be helpful in combating genetic drift in Monte Carlo filtering algorithms. Indeed, the distribution of individuals seems closer to the true distribution than the population corresponding to the classical procedure. To quantify these observations, several performance measures can be introduced. Such measures assess the distance of the empirical distribution of the population from the true distribution. The ratio of good decisions taken by the algorithms can be compared.

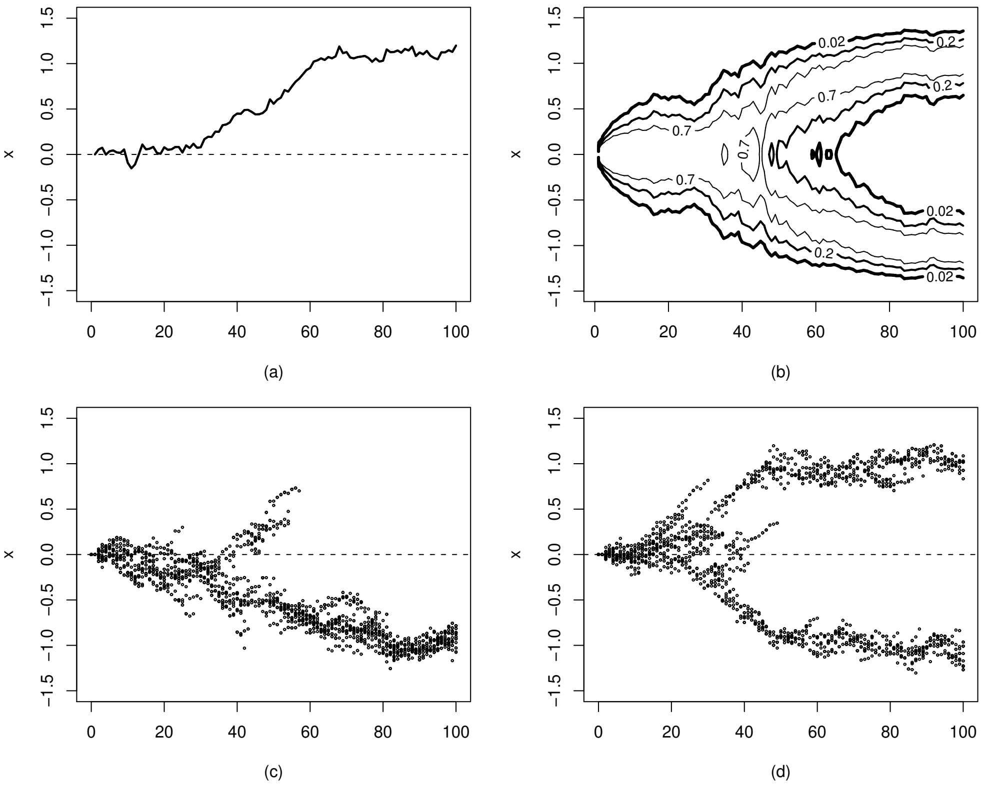

Figure 3.1: (a) A simulated trajectory from the double well dynamics. (b) Contour plot of the posterior distribution densities. The observation function is symmetric $H(x) = |x|$. (c) Classical Monte Carlo filtering plot (population size = 20). (d) Monte Carlo + Niching simulation plot.

KS distance. The Kolmogorov-Smirnov metric is a standard measure of the distance between two probability distributions μ_1 and μ_2. It is defined as

$$D_{KS}(\mu_1, \mu_2) = \sup_t |F_1(t) - F_2(t)| \in [0, 1]$$

where F is the cumulative distribution function of μ. Here, μ_1 denotes the empirical distribution of the population computed from Monte Carlo algorithms and μ_2 denotes the true distribution.

Measure of symmetry: MS. In the case where H is symmetric (half of the simulation runs), the posterior distribution of X_t is bimodal (and symmetric). MS is the fraction of time during which two niches (modes) subsist. (A niche is considered active when more than 10 percent of the population are present.)

Ratio of good decisions: RGD. In the case where H is non symmetric, we say that an algorithm takes a decision when all individuals are located at the same side of zero during 5 time steps. Good decisions correspond to all individuals evolving in the same region as the signal.

3.5.4 Results

This Section presents the most significant results obtained after the series of experiments. Symmetric observation functions are discussed first. Figure 3.2 reports the values of MS and the KS distance for the double well dynamics (DW). Similar results have been obtained for AR(1) and the piecewise linear dynamics. These results are summarized in Table 3.1.

Figure 3.2 shows that the fraction of time (MS) during which two niches subsist averages to 0.346 in the Monte Carlo filtering algorithm. This fraction increases to 0.922 when sharing is used together with Deb's rule (the best bandwidth is however $\sigma_{\rm sh} = 1$). Simulation runs have also been performed with a population size of 100 individuals. The improvement due to sharing is more significant when the population size is small. Similar remarks can be made regarding the KS distance. For large population sizes, the goodness-of-fit seems better when Silverman's rule is used.

Table 3.1 reports the relative gain in using Deb's rule computed for each measure (MS and KS). This gain represents the difference between measures with sharing / without sharing averaged over 100 runs. The results are given as percentages (MS and KS are floating point numbers between 0 and 1). Numbers in brackets represent the best ratio obtained from constant bandwidth rules (when this ratio is significantly better than Deb's rule). The star means that MS reached the maximal value (100%). The following set of comments can be made about these results.

1) For this set of bimodal problems, sharing always improves Monte Carlo filters (the improvement may sometimes be a minor one).

2) Deb's rule is competitive (and has the advantage to be adaptive). This explains why this rule is chosen as a reference rule in Table 3.1.

3) Silverman's rule outperforms the other rules for large population sizes. (In some sense, this rule is optimal in density estimation when n grows to infinity.)

4) Sharing is less efficient when the observation noise is small.

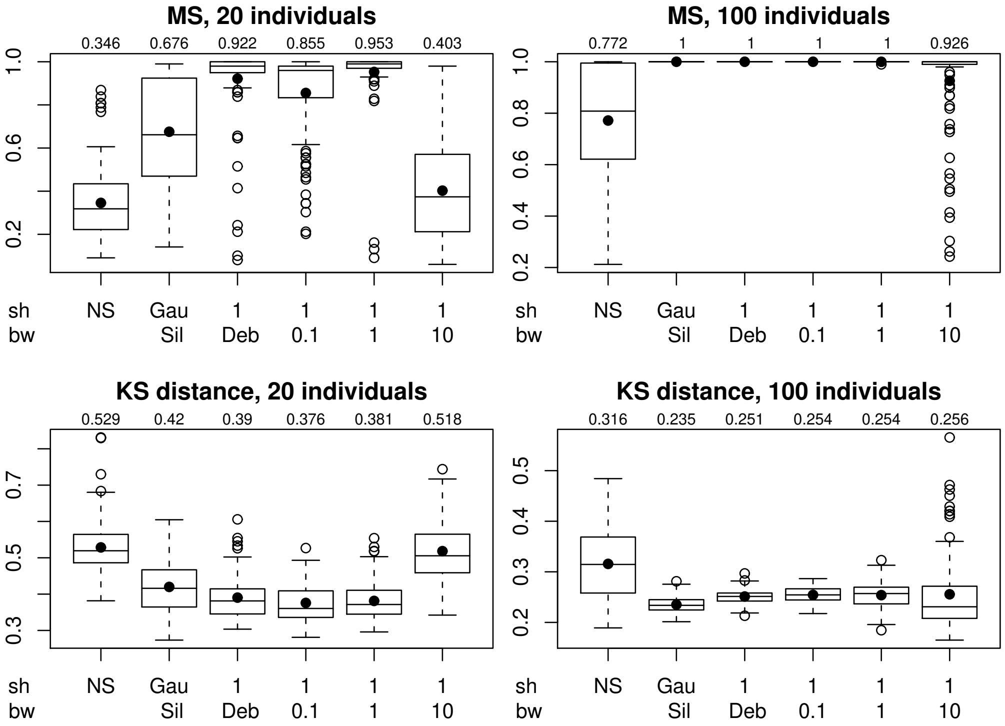

Figure 3.2: Performance measures corresponding to the double well dynamics ($q = 0.24$, $\sigma = 1$) and the symmetric observation function. sh refers to the sharing function (Gaussian or triangular) and bw to the bandwidth parameter.

Turn now to the experimental results in the non-symmetric observation context. These results are summarized in Table 3.2 using the same notations as before. In the non-symmetric context, the measure of symmetry has been replaced by the ratio of good decisions (RGD). The posterior distribution is indeed multimodal only during a short interval of time, after that it concentrates on a single mode.

	gains (%)			
	20 individuals		100 individuals	
dynamics	MS	KS	MS	KS
$AR(1)$, $\sigma = 1$	24	5	4*	3
$AR(1)$, $\sigma = 0.1$	0	0	*(15)* 2	*(7)* 1
piecewise linear, $\sigma = 1$	17	6	13*	8
piecewise linear, $\sigma = 0.2$	1	1	*(20)* 7	4
DW, $q = 0.05$, $\sigma = 1$	*(49)* 35	*(15)* 8	25*	6
DW, $q = 0.24$, $\sigma = 1$	57	14	23*	6
DW, $q = 0.1$, $\sigma = 1$	62	17	21*	6
DW, $q = 0.24$, $\sigma = 0.05$	*(21)* 12	4	40	15

Table 3.1: Improvement obtained from sharing with Deb's rule (symmetric observation function).

In contrast to bimodal problems, high gains can hardly be expected. The aim of sharing is maintaining individuals in niches when such niches truly exist. Note that this method do not create artificial niches. Significant gains can nevertheless be observed when the posterior distribution remains multimodal within a long period before concentrating on a single mode. In the period during which the filtering problem is multimodal, maintaining subpopulations in all niches is crucial, as the algorithm should be capable of tracking the mode that will subsist.

3.6 Discussion

This paper presented a new paradigm in Monte Carlo filtering algorithms: the method of likelihood sharing. While niching methods are widely accepted in evolutionary computation, the benefit of these techniques remains unexplored in several neighboring domains.

Our results give evidences that sharing methods can improve Monte Carlo filtering algorithms significantly. These methods are dedicated to problems for which posterior distributions are multimodal and standard algorithms are not efficient. Adding a sharing method allows population sizes to be reduced by a large factor, and contributes to the global efficiency of the algorithm.

The method can be beneficial as well for problems that are not purely multimodal. This occurs in tracking a specific mode among several others that would be prominent after a while.

The empirical results presented in this paper have been obtained for one-dimensional problems, for which the solution can be computed by standard numerical methods.

	gains (%)			
	20 individuals		100 individuals	
dynamics	RGD	KS	RGD	KS
$AR(1)$, $\sigma = 1$	2	-1	2*	-3
$AR(1)$, $\sigma = 0.1$	*(1)* -5	0	-2	0
piecewise linear, $\sigma = 1$	*(2)* -4	1	0	-5
piecewise linear, $\sigma = 0.2$	-2	2	4	3
DW, $q = 0.05$, $\sigma = 1$	11	9	17*	*(1)* -5
DW, $q = 0.24$, $\sigma = 1$	*(19)* 6	6	10*	*(1)* -12
DW, $q = 0.24$, $\sigma = 0.05$	-2	-1	5	-3

Table 3.2: Performances of sharing for the non-symmetric observation function.

Further work is needed to extend this contribution to real-world problems (such as those arising in terrain navigation) and higher dimensional issues.

Chapitre 4

Genetic algorithms in random fitness landscapes

4.1 Abstract

We study the asymptotic behavior of two mutation-selection genetic algorithms with finite population in random environments. First, the state space is a supercritical Galton-Watson tree and the objective function is the distance from the root. Second, the state space is a regular tree and the objective function is a sample of a tree-indexed random walk. In both cases, we prove that, after n steps, the algorithm finds the maximum possible value of the objective function up to a finite random constant.

4.2 Introduction

4.2.1 Motivation

Genetic algorithms are popular stochastic optimization methods, formally introduced by J. Holland [39] and based on a biological analogy. One seeks to maximize the *objective function* $F : S \to \mathbb{R}$, defined on the *state space* S. The mutation-selection genetic algorithms we consider evolve a finite population of particles in S under the following iterated two-stages procedure:

1. **Mutation:** each particle moves at random to a neighboring element of S, a convenient graph structure on S being given.

2. **Selection:** a new population of particles is created by resampling. Particles with low relative values of F tend to be eliminated, particles with large relative values of F are more often replicated.

Mutation allows for the discovery of new, possibly better, feasible solutions to the maximization of F, and selection concentrates the exploration of S around good, already discovered solutions. Note that the algorithms we study do not include the mating operator, since it is difficult to define it in our context. Moreover, genetic algorithms that include mating are in general harder to analyze than mutation-selection ones.

Despite the successes of this approach in the applications and numerous experimental studies, few rigorous mathematical results are available, that would explain how genetic algorithms work. In particular, the interaction between mutation, selection and population size remains poorly understood, even for simple examples. Among the exceptions is the paper by R. Cerf [15], which gives general asymptotic convergence results in the context of Markov chains with rare transitions. See the papers by Y. Rabinovich and A. Wigderson [62], G. Rudolph [66] and J. Bérard and A. Bienvenüe [7], [8] where simple examples are worked out, and C. Mazza and D. Piau [50] for results in the infinite-population case.

In this paper, we study the asymptotic behavior of two mutation-selection genetic algorithms defined on trees in random contexts. First, the state space is a random supercritical Galton-Watson tree conditioned upon non-extinction, see for instance the paper [47] by R. Lyons, R. Pemantle and Y. Peres. The objective function $F(x)$ at a vertex x of the tree is the distance between x and the root of the tree. In the second case, the state space is a regular tree, and the objective function is a sample of a tree-indexed simple random walk, see [57].

Both situations display randomness, since either the graph structure of the state space or the objective function itself is random. In our opinion, this makes interesting the study of these simple optimization algorithms.

In both cases, the objective function is unbounded. We are looking for the asymptotic growth speed of the objective function values that the algorithm discovers, rather than for convergence rates to a finite maximum value. As explained by D. Aldous in [1], the study of the transient, in the Markov chain sense, behavior of infinite-state algorithms is intended as a toy model for the pre-equilibrium behavior of randomized optimization algorithms on large finite sets. Our second case is precisely the situation that Aldous analyses for the Metropolis algorithm.

A second motivation comes from the study of diffusion phenomena in random environments, see for instance the book by B. Hughes [41]. The biased random walk on Galton-Watson trees, studied in [47], corresponds to the Metropolis algorithm on these trees with our objective function $F(x)$. Thus, it is natural to ask for the behavior of genetic algorithms in the same context.

Our results show that, at least under some restrictions on the parameters, part of which are believed to be purely technical, the algorithms perform very well for the two examples considered. See below for a detailed formulation. Yet another motivation is the study of the approximation of infinite-population genetic algorithms by finite-population ones, that are the only computationally realistic algorithms. This is an important theoretical question since infinite-population genetic algorithms are often used to model the behavior of large population ones, see the paper by P. Del Moral, L. Kallel and J. Rowe [22], and the book by M. Vose [73]. From the results below, our two specific examples display such an approximation property, in a strong sense.

4.2.2 Description of the model and notations

We use proportional selection with Boltzmann weights, see [15], and mutation steps given by simple random walks on the state space. Instead of being kept constant as is usually the case, see [39], the population size p_n at time n, goes to infinity as n goes to infinity, at a possibly slow rate.

We use the same notations for the two cases, the case under consideration being clear from the context.

Galton-Watson case

Let $\mathbb{T}$ denote a random Galton-Watson tree, with root r. For $x \in \mathbb{T}$, $|x|$ is the tree-distance between r and x, and the objective function F is $F(x) = |x|$. The reproduction law of the Galton-Watson tree is q, and Q is a random variable of law q. We assume that q has bounded support, $Q \leq q^*$ say, and we set

$$\theta = \mathrm{E}(Q/(Q+1)).$$

Random objective function case

Fix $m \geq 2$, and let $\mathbb{T}$ denote a rooted regular $m-$ary tree, with root r. For every edge e of $\mathbb{T}$, choose $v(e) \in \{+1, -1\}$. Then, let $F(r) = 0$, and, for every vertex $x \neq r$ of $\mathbb{T}$,

$$F(x) = \sum_{e \in \{r \to x\}} v(e),$$

where $\{r \to x\}$ denotes the set of edges on the unique injective path from r to x. We assume that $(v(e))_e$ is a collection of i.i.d. random variables of law

$$\rho\, \delta_1 + (1-\rho)\, \delta_{-1}.$$

In both cases, we fix $\beta > 0$ and a deterministic sequence $(p_n)_n$ of positive integers. The genetic algorithm is a time-inhomogeneous Markov chain

$$X_n = (X_n^{(i)})_{1 \leq i \leq p_n},$$

such that $X_n \in \mathbb{T}^{p_n}$, with initial state $X_0 = (r, \dots, r) \in \mathbb{T}^{p_0}$, defined by the following transitions.

1. **Mutation step:** $X_n \longrightarrow Y_n$

   Each particle $X_n^{(i)}$ evolves independently of the others, and performs one step of a simple random walk on $\mathbb{T}$, symmetric and to the nearest neighbors.

   In the random objective function case, the probability of staying at the same vertex is $\mu > 0$. In the Galton-Watson case, we assume that $\mu = 0$. The new positions are $Y_n^{(i)}$, $1 \leq i \leq p_n$.

2. **Selection step:** $Y_n \longrightarrow X_{n+1}$

   Each $X_{n+1}^{(i)}$ for $1 \leq i \leq p_{n+1}$ is chosen randomly and independently of the others in the set $\left\{Y_n^{(i)},\ 1 \leq i \leq p_n\right\}$, according to the probability law

$$\frac{1}{S'_n}\sum_{i=1}^{p_n} \exp\left(\beta F(Y_n^{(i)})\right)\delta_{Y_n^{(i)}}, \quad \text{where} \quad S'_n = \sum_{i=1}^{p_n} \exp\left(\beta F(Y_n^{(i)})\right).$$

The particles $X_n^{(1)}, \ldots, X_n^{(p_n)}$ compose the population at time n. Note that, at a given time, several particles may be located at the same vertex. Let N_n denote the counting measure of X_n, and $Z_F(X_n)$ the counting measure of the objective function values:

$$N_n = \sum_{i=1}^{p_n} \delta_{X_n^{(i)}}, \quad Z_F(X_n) = \sum_{i=1}^{p_n} \delta_{F(X_n^{(i)})}.$$

For $x_1, \ldots, x_p \in \mathbb{T}$, let

$$F^*(x_1, \ldots, x_p) = \sup\{F(x_i)\,;\, 1 \leq i \leq p\}.$$

4.2.3 Statement of the results

We use the following assumptions.

- **A1**: $\lim_{n\to+\infty} p_n (\log n)^{-2} = +\infty$
- **A2**: $\lim_{n\to+\infty} p_{n+1}/p_n = 1$

Galton-Watson case

- **A3**: $\theta > 1/2$
- **A4**: $\exp \beta > 1/\theta$

Random objective function case

- **A5**: $m\rho \geq 1$

- **A6**: $\exp\beta > [(1-\mu)\,\rho\,]^{-1}(1+1/m)$
- **A7**: $\exp\beta > 1/\mu$

Theorem 6 *Assume that A1, A2, A3 and A4 hold. Then, almost surely on the non-extinction of the underlying Galton-Watson tree $\mathbb{T}$, and almost surely on the realization of the algorithm,*

$$F^*(X_n) = n - O(1).$$

Moreover, the empirical law of $F(X_n)$ concentrates around its maximum value $F^*(X_n)$, in the following sense. There exists $c > 0$ such that, for all $a \geq 1$, and for all $n \geq k_a$, where k_a is random and almost surely finite,

$$p_n^{-1} Z_F(X_n)([0, F^*(X_n) - a]) \leq c \exp(-\beta a).$$

Theorem 7 *Assume that A1, A2, A5, A6 and A7 hold. Then, almost surely on the random objective function F and on the realization of the algorithm,*

$$F^*(X_n) = n - O(1).$$

Moreover, the empirical law of $F(X_n)$ concentrates around its maximum value $F^*(X_n)$, in the sense of Theorem 6:

$$p_n^{-1} Z_F(X_n)((-\infty, F^*(X_n) - a]) \leq c \exp(-\beta a).$$

We now comment upon these results and the hypotheses. Note that trivially $F^*(X_n) \leq n$. Thus, in both cases, up to a finite random constant, the cloud of particles X_n contains the largest possible value of F after n steps, and the empirical distribution of the particles is concentrated, in the sense of an exponentially decreasing tail, around this maximum value. Thus, in the Galton-Watson case, there is no slowdown phenomenon, despite the many dead-ends in the tree. Recall that the Metropolis algorithm is slowed down at low temperature, see [47]. Moreover, the large deviations results of A. Dembo, N. Gantert, Y. Peres and O. Zeitouni [24] show that the supremum of p_n independent Metropolis algorithms has an exponentially small probability of being of order n. Hence, the genetic algorithm is more efficient in this context. In the random objective function case, the algorithm finds the exponentially small part of the state space where F takes values of order n. In both cases, our proof of the efficiency of the interaction between mutation and selection relies upon the averaging effect produced by the growth of p_n.

Remark 22 *The behavior stated in Theorems 6 and 7 is easily seen to be shared by the infinite-population versions of our algorithms. Since assumption A1 allows for a slow growth rate of the population size with time, our results may be interpreted as a strong approximation of the infinite-population behavior by finite-population algorithms. See [22] and [73] for an account on these questions.*

Assumption A3 implies that the average number $\mathrm{E}(Q)$ of children under q, is strictly greater than one, so that the non-extinction of $\mathbb{T}$ occurs with positive probability and the conditioning in the theorem is trivially defined. However, A3 is in general strictly stronger than the condition $\mathrm{E}(Q) > 1$, so our result does not include every supercritical Galton-Watson tree. However, the conclusion of Theorem 6 may hold as soon as $\mathrm{E}(Q) > 1$.

When A3 is satisfied, A4 is satisfied as soon as $\beta > \log(2)$, so β has not to be very large for the theorem to be valid. Whether the behavior for small values of β is different from the one stated in Theorem 6 is an open question.

Assumption A5 says that, given a vertex x, the average number of children y of x such that $F(y) = F(x) + 1$ is strictly greater than 1, so this condition is analogous to A3. In fact, there is a critical value $\rho_{\mathrm{cr}} \in (0, 1/m)$, see [57], such that, if $\rho < \rho_{\mathrm{cr}}$, there is almost surely only a finite number of vertices x such that $F(x) \geq -c$, for any c. When $\rho > \rho_{\mathrm{cr}}$, there exist infinite rays R in $\mathbb{T}$ and $c > 0$ such that $F(R(i)) \geq ci$ for all i. From A5, ρ is far away from ρ_{cr}, and indeed we have the stronger property that there exist infinite rays R in $\mathbb{T}$ and $c > 0$ such that $F(R(i)) \geq i - c$ for all i. However, when $m \geq 3$, this includes non-trivial optimization situations since, as soon as $\rho < 1/2$, the typical behavior of F along a ray is to go to $-\infty$ at a positive speed.

The two cases considered have different structures, but the methods that we use for proving the results are so close that we give the full proof only in the Galton-Watson case. We sketch the slight modifications that are needed in the random function case.

4.3 Preliminaries on the Galton-Watson case

From now on, we assume that A1 to A4 are satisfied.

4.3.1 Additional notations and definitions

Let $M_n = Z_F(X_n)(F^*(X_n))$ be the number of particles maintained by X_n at the top level. Define N'_n, $Z_F(X'_n)$ and M'_n in the same way as N_n, $Z_F(X_n)$ and M_n, where Y_n replaces X_n.

We use the standard construction for random diffusion processes in random environments. Thus, the Galton-Watson tree is progressively sampled as the algorithm explores it, rather than being fixed once for all. A standard result is that, if a property of the algorithm holds almost surely with respect to this construction, then it holds for almost every realization of the environment, and for almost every realization of the algorithm conditional on this environment. In this context, $\mathbb{T}_n$ denotes the subtree uncovered by the algorithm up to time n.

Remark 23 *In all the proofs, we work on unconditioned Galton-Watson trees, and condition upon non-extinction only at the end.*

The σ-algebra $\mathcal{F}_n$ is generated by X_k, Y_k and $\mathbb{T}_{k+1}$, for $0 \leq k \leq n-1$, and X_n. The σ-algebra $\mathcal{G}_n$ is generated by $\mathcal{F}_n$, $\mathbb{T}_{n+1}$ and Y_n.

Let x_1, x_2, ..., x_{L_n}, with $L_n \geq 1$, denote the distinct vertices of the tree maintained by X_n at the level $F^*(X_n)$. Note that $L_n \leq M_n$. Each x_i has a random, possibly zero, number z_i of children and, if $z_i \geq 1$, we denote these children by $x_i(1), \dots, x_i(z_i)$. Let $T_{n,i} = N_n(x_i)$ denote the number of particles maintained by the algorithm at x_i. Thus, $M_n = \sum_{i=1}^{L_n} T_{n,i}$. Let

$$T_n^* = \sup\{T_{n,i}\,;\, 1 \leq i \leq L_n\}.$$

Definition 7 *Say that $H_n(\ell)$ holds if $T_n^* \leq M_n/\ell$.*

Remark 24 *$H_n(1)$ is the certain event.*

Let $\psi \in (1/\theta, \exp(\beta))$ (from A4, $\theta \exp(\beta) > 1$). Let $\nu > 1$ such that $\nu^2 < \psi\theta$ and $\nu < 2\theta$ (from A3, $\theta > 1/2$).

Definition 8 *For $\lambda > 0$, define the event $D_1(n)$ by:*

- $F^*(X_{n+1}) = F^*(X_n) + 1$
- $M_{n+1} \geq \lambda p_{n+1}$ *if* $M_n \geq \lambda p_n$ *and* $M_{n+1} \geq \nu^2 M_n$ *if* $M_n \leq \lambda p_n$

For $\ell \geq 1$, set

$$D_2(n, \ell) = D_1(n) \cap H_{n+1}(\nu\, \ell).$$

4.3.2 Record epochs

Definition 9 *Call n a record epoch if $F^*(X_n) > F^*(X_p)$ for all $0 \leq p \leq n-1$.*

Record epochs are a fundamental tool. Indeed, in the construction above, fresh random decisions, that is independent from the past, are made concerning the number of children of $x_1, \dots, x_{L_n}$, at each record epoch n.

Proposition 25 *Conditional upon non-extinction of the tree, there exists almost surely an infinite number of record epochs.*

In other words, the sequence $F^*(X_n)$ is unbounded from above almost surely conditional on the tree being infinite.

Proof of Proposition 25:

Assume that the mutation steps of the algorithm are defined as follows. Set $\Lambda = (q^* + 1)!$. First, for every $d \leq q^* + 1$, fix a partition $\mathcal{P}_d$ of the set $\{1, \dots, \Lambda\}$

into d equal parts. Then, let d_i be the number of neighbors of $X_n^{(i)}$, $d_i \leq q^* + 1$. We draw a random integer $u_i(n)$ according to the uniform law on $\{1, \dots, \Lambda\}$, and choose the next position $Y_n^{(i)}$ among the neighbors of $X_n^{(i)}$ according to the element of $\mathcal{P}_{d_i}$ to which $u_i(n)$ belongs.

Consider an arbitrary vertex $x \in \mathbb{T}$, a positive integer r, and a finite sequence of integers $\xi(0), \dots, \xi(r-1)$ in $\{1, \dots, \Lambda\}$. Introduce the event

$$\Delta_n(x) = \{x \in X_n \text{ and } F(x) = F^*(X_n)\}.$$

On $\Delta_n(x)$, let i denote the smallest index $j \leq p_n$ such that $X_n^{(j)} = x$. Say that the event $\Gamma_n(x)$ occurs if there exists a sequence $(i_k)_{0 \leq k \leq r}$ such that $i_0 = i$ and, for all $0 \leq k \leq r-1$,

$$u_{i_k}(n+k) = \xi(k) \text{ and } X_{n+k+1}^{(i_{k+1})} = Y_{n+k}^{(i_k)}.$$

Finally, define

$$\Delta(x) = \{\Delta_n(x) \text{ occurs infinitely often}\}$$

and

$$\Gamma(x) = \{\Gamma_n(x) \text{ occurs infinitely often}\}.$$

A particle $Y_n^{(j)}$ such that $F(Y_n^{(j)}) \geq F^*(Y_n) - b$ has a probability greater than

$$1 - \left(1 - \exp(-\beta b) p_n^{-1}\right)^{p_{n+1}}$$

of being kept by the selection step, conditional on $\mathcal{G}_n$. After r mutation steps, the distance between the current maximum and a particle initially located on the maximum cannot exceed $2r$ (in the worst case, the particle performs r downward steps, and the maximum performs r upward steps). Moreover, for a given mutation step, $\xi(k)$ is drawn with probability Λ^{-1}. Hence, on $\Delta_n(x)$,

$$\mathrm{P}\left[\Gamma_n(x) \mid \mathcal{F}_n\right] \geq \Lambda^{-r} \left(1 - \left(1 - \exp(-\beta 2r) p_n^{-1}\right)^{p_{n+1}}\right)^r.$$

This expression being bounded away from zero independently from n, by Lemma 25,

$$\mathrm{P}\left[\Delta(x) \setminus \Gamma(x)\right] = 0. \tag{4.1}$$

Call B the event that $F^*(X_n)$ is bounded from above. If B occurs, there is at least one vertex $x \in \mathbb{T}$ such that $\Delta(x)$ occurs. Hence, according to (4.1) and Lemma 25, for every finite sequence $\xi(0), \dots, \xi(r-1)$, $\Gamma(x)$ occurs almost surely on B. But the non-extinction of $\mathbb{T}$ implies that, for all $C > 0$, there is a positive integer r and a path of length r connecting x to another vertex at distance greater than C from the root. This path is associated with at least one sequence of mutation decisions $\xi(0), \dots, \xi(r-1)$. Since this holds for all C, we get that, conditional on non-extinction, B has zero probability.

□

4.4 Proof in the Galton-Watson case

The theorem is a simple consequence of Propositions 26 and 27. The technical lemmas needed are stated in section 4.6. Lemma 20 below contains our basic estimates on the one-step transitions of the algorithm.

Lemma 20 *There exist $\lambda > 0$, $K > 0$, $n_0 \geq 1$ and $g > 0$, such that, for all $n \geq n_0$,*

- *if n is a record epoch,*
$$\mathrm{P}\left[D_1(n) \mid \mathcal{F}_n\right] \geq g,$$
- *if n is a record epoch, if $M_n \geq K$, and if $H_n(\ell_n)$ occurs with $\ell_n \leq (M_n)^{1/2}$, then*
$$\mathrm{P}\left[D_2(n, \ell_n) \mid \mathcal{F}_n\right] \geq 1 - C_1 \exp\left(-C_2 (M_n)^{1/2}\right) - \exp\left(-C_3 \ell_n\right).$$

(recall that the definitions of $D_1(n)$ and $D_2(n, \ell)$ involve a parameter λ.)

Proof of Lemma 20:

Set $B_n = \{n \text{ is a record epoch}\} \cap H_n(\ell_n)$. We put $\ell_n = 1$ for the proof of the first assertion of the lemma. Let $t_n = M_n/\ell_n$. Let $\epsilon \in (0, 1)$, whose value will be fixed later. Denote by $C_r(expression)$, any positive constant that depends only on $expression$ and on the law q.

Random sampling of the tree

Conditionally to $\mathcal{F}_n$ and B_n, n is a record epoch and the numbers of children $z_i, 1 \leq i \leq L_n$ of the vertices $x_i, 1 \leq i \leq L_n$ are i.i.d. random variables with law q. Let
$$J_n = \sum_{i=1}^{L_n} T_{n,i}\, z_i/(z_i + 1), \quad E_1 = \{J_n \geq (1 - \epsilon)\theta M_n\}.$$
From Lemma 22, on B_n,
$$\mathrm{P}\left[E_1^c \mid \mathcal{F}_n\right] \leq \exp\left(-C_1(\epsilon) M_n / T_n^*\right).$$
Hence,
$$\mathrm{P}\left[E_1^c \mid \mathcal{F}_n\right] \leq \exp\left(-C_1(\epsilon) \ell_n\right) \triangleq e_1.$$

Mutation

Conditionally to $\mathcal{F}_n$ and $\mathbb{T}_{n+1}$, the number of particles $I_1, \ldots, I_{L_n}$ that move by mutation from each vertex x_i to the children of x_i are independent random variables, with respective binomial laws $\mathcal{B}(T_{n,i}, z_i/(z_i + 1))$. Let
$$E_2 = \{M'_n \geq (1 - \epsilon)^2 \theta M_n, F^*(Y_n) = F^*(X_n) + 1\}.$$

The fact that $I_1+\cdots+I_{L_n} \geq (1-\epsilon)^2\theta M_n$ implies that $F^*(Y_n) = F^*(X_n)+1$ and that $M'_n = I_1 + \cdots + I_{L_n}$ since $(1-\epsilon)^2\theta M_n > 0$. Thus, E_1 and $I_1 + \ldots + I_{L_n} \geq (1-\epsilon)J_n$ imply E_2. By Lemma 23, on E_1,

$$\mathrm{P}\left[E_2^c \mid \mathcal{F}_n, \mathbb{T}_{n+1}\right] \leq \exp\left(-C_2(\epsilon)J_n\right)$$

so that on E_1,

$$\mathrm{P}\left[E_2^c \mid \mathcal{F}_n, \mathbb{T}_{n+1}\right] \leq \exp\left(-C_3(\epsilon)M_n\right) \triangleq e_2$$

Let us define the set of indices $\mathcal{A}_n$ by:

$$\mathcal{A}_n = \{1 \leq i \leq L_n \mid T_{n,i} \geq (1/2)t_n, z_i \geq 1\}.$$

Fix $i \in \mathcal{A}_n$ and $1 \leq j \leq z_i$. The number $N'_n(x_j)$ of particles that move by mutation from x_i to $x_i(j)$, follows, conditionally to $\mathcal{F}_n$ and $\mathbb{T}_{n+1}$, a binomial law

$$\mathcal{B}(T_{n,i}, (z_i+1)^{-1}).$$

Let

$$E_3(i,j) = \{N'_n(x_i(j)) \leq (1+\epsilon)T_{n,i}(z_i+1)^{-1}\}.$$

We have by Lemma 24

$$\mathrm{P}\left[(E_3(i,j))^c \mid \mathcal{F}_n, \mathbb{T}_{n+1}\right] \leq \exp\left(-C_4(z_i,\epsilon)T_{n,i}\right).$$

Using the fact that $z_i \leq q^*$, and that $T_{n,i} \geq (1/2)t_n$ since $i \in \mathcal{A}_n$, we get

$$\mathrm{P}\left[E_3(i,j)^c \mid \mathcal{F}_n, \mathbb{T}_{n+1}\right] \leq \exp\left(-C_5(\epsilon)t_n\right).$$

Let

$$E_3 = \bigcap_{i \in \mathcal{A}_n} \bigcap_{j=1}^{z_i} E_3(i,j).$$

The above intersection runs over at most $q^* M_n$ events. Hence

$$\mathrm{P}\left[E_3^c \mid \mathcal{F}_n, \mathbb{T}_{n+1}\right] \leq q^* M_n \exp\left(-C_5(\epsilon)t_n\right) \triangleq e_3.$$

Selection

Define E_4 by:

- on $\{M_n > \lambda p_n\}$,

$$M_{n+1} \geq \psi\theta(1-\epsilon)^4 \lambda p_{n+1}, \quad F^*(X_{n+1}) = F^*(Y_n),$$

- on $\{M_n \leq \lambda p_n\}$,

$$M_{n+1} \geq \psi\theta(1-\epsilon)^4 M_n, \quad F^*(X_{n+1}) = F^*(Y_n).$$

Recall that ψ is defined after Remark 24. From now on, we assume that ϵ is small enough so that: $\psi\theta(1-\epsilon)^4 \geq \nu^2$, and we choose λ as in Lemma 21. By Lemma 21, conditionally to $\mathcal{G}_n$, the number $Z_F(X_{n+1})(F^*(Y_n))$ of particles at level $F^*(Y_n)$ contained in X_{n+1} follows a binomial law $\mathcal{B}(p_{n+1}, \mu_{n+1})$, with

$$\mu_{n+1} \geq \psi \inf(\lambda, M'_n p_n^{-1}) = h.$$

Let

$$G = \{Z_F(X_{n+1})(F^*(Y_n)) \geq (1-\epsilon) h p_{n+1}\}.$$

Since $h > 0$, G implies that $M_{n+1} = Z_F(X_{n+1})(F^*(Y_n))$. Moreover, one can easily check that for n sufficiently large (that is, such that $p_{n+1}/p_n \geq 1-\epsilon)$), $G \cap E_2 \subset E_4$. On the other hand, we have $h \leq \psi\lambda$, and by Lemma 24, on E_2,

$$\mathrm{P}\left[E_4^c \mid \mathcal{G}_n\right] \leq \exp\left(-C_6(\lambda,\epsilon) M_n\right) \triangleq e_4.$$

Conditionally to $\mathcal{G}_n$, M_{n+1} and $F^*(X_{n+1})$, the number $N_{n+1}(y)$ of particles put by selection on a vertex y at level $F^*(X_{n+1})$ follows a binomial law

$$\mathcal{B}(M_{n+1}, N'_n(y)(Z_F(X'_n)(F^*(X_{n+1})))^{-1}).$$

Assume that E_4 occurs, so that $F^*(X_{n+1}) = F^*(Y_n)$. Hence $Z_F(X'_n)(F^*(X_{n+1})) = M'_n$. For $i \in \mathcal{A}_n$ and $1 \leq j \leq z_i$, E_3 implies that $N'_n(x_i(j)) \leq (1+\epsilon)(1/2)T^*_n$, and E_2 implies that $M'_n \geq (1-\epsilon)^2\theta M_n$. Thus, on $B_n \cap E_2 \cap E_3$:

$$N'_n(y)(M'_n)^{-1} \leq (2\theta)^{-1}(1+\epsilon)(1-\epsilon)^{-2}(\ell_n)^{-1}.$$

Let

$$E_5(i,j) = \{N_{n+1}(x_i(j))(M_{n+1})^{-1} \leq (2\theta)^{-1}(1+\epsilon)^2(1-\epsilon)^{-2}(\ell_n)^{-1}\}.$$

From now on, we assume that ϵ is small enough so that $(2\theta)^{-1}(1+\epsilon)^2(1-\epsilon)^{-2} \leq \nu^{-1}$. By lemma 24, on $B_n \cap E_2 \cap E_3 \cap E_4$,

$$\mathrm{P}\left[E_5(i,j)^c \mid \mathcal{G}_n, M_{n+1}, F^*(X_{n+1})\right] \leq \exp\left(-C_8(\epsilon) M_{n+1}\ell_n^{-1}\right).$$

Note that, on E_4, we have (easily) $M_{n+1} \geq \lambda M_n$ since $\lambda < 1$. Let

$$E_5 = \bigcap_{i\in\mathcal{A}_n} \bigcap_{j=1}^{z_i} E_5(i,j).$$

On $B_n \cap E_2 \cap E_3 \cap E_4$,

$$\mathrm{P}\left[E_5^c \mid \mathcal{G}_n, M_{n+1}, F^*(X_{n+1})\right] \leq q^* M_n \exp\left(-C_9(\epsilon,\lambda) M_n/\ell_n\right) \triangleq e_5.$$

When y's parent is one of the $x_i, i \notin \mathcal{A}_n$, we have, $N'_n(y) \leq (1/2)t_n$. Let

$$E_6(i,j) = \{N_{n+1}(x_i(j))(M_{n+1})^{-1} \leq (1/2)\theta^{-1}(1+\epsilon)(1-\epsilon)^{-2}(\ell_n)^{-1}\}$$

and

$$E_6 = \bigcap_{i \notin \mathcal{A}_n} \bigcap_{j=1}^{z_i} E_6(i,j).$$

Using the same argument as for E_5, we get that, on $E_2 \cap E_4$,

$$\mathrm{P}\left[E_6^c \mid \mathcal{G}_n, M_{n+1}, F^*(X_{n+1})\right] \leq q^* M_n \exp\left(-C_{11}(\epsilon,\lambda)M_n/\ell_n\right) \triangleq e_6.$$

Set $U_n = E_1 \cap E_2 \cap E_3 \cap E_4 \cap E_5 \cap E_6$ and $W_n = E_1 \cap E_2 \cap E_4$.

Let us summarize the properties of X_{n+1} contained in $B_n \cap U_n$.

- $F^*(X_{n+1}) = F^*(X_n) + 1$ and so $n+1$ is a record epoch.
- $M_{n+1} \geq \lambda p_n$ if $M_n \geq \lambda p_n$, and $M_{n+1} \geq \nu^2 M_n$ if $M_n \leq \lambda p_n$
- $T^*_{n+1} M^{-1}_{n+1} \leq \nu^{-1}\ell_n^{-1}$ and so $H_{n+1}(\nu\ell_n)$ occurs

The two first properties are also consequences of $B_n \cap W_n$. Note that the e_i are measurable with respect to $\mathcal{F}_n$, and that $e_1, e_2, e_4 < 1$. Moreover, assuming $\ell_n \leq M_n^{1/2}$ and $M_n \geq K$ for K large enough, we get that: $e_7 \triangleq e_1 + e_2 + e_3 + e_4 + e_5 + e_6 < 1$. Indeed, $e_1 \leq \exp\left(-C_1(\epsilon)\ell_n\right)$, and all the other terms are bounded from above by an expression of the form $C_{12} M_n \exp\left(-C_{13}(\epsilon,\lambda)M_n/\ell_n\right)$. The inequality $\ell_n \leq (M_n)^{1/2}$ entails that this expression is itself bounded from above by $C_{12} M_n \exp\left(-C_{13}(\epsilon,\lambda)M_n^{1/2}\right)$, hence by $C_{14}\exp\left(-C_{15}(\epsilon,\lambda)M_n^{1/2}\right)$, and this last expression goes to zero as $M_n \to +\infty$.

Conditioning $\mathrm{P}\left[U_n \mid \mathcal{G}_n, M_{n+1}, F^*(X_{n+1})\right]$, we get that, if $\ell_n \leq (M_n)^{1/2}$, on $B_n \cap \{M_n \geq K\}$,

$$\mathrm{P}\left[U_n \mid \mathcal{F}_n\right] \geq (1-e_6-e_5)(1-e_4)(1-e_3-e_2)(1-e_1) \geq 1-e_7,$$

and, on B_n, without restrictions on ℓ_n or M_n,

$$\mathrm{P}\left[W_n \mid \mathcal{F}_n\right] \geq (1-e_4)(1-e_2)(1-e_1).$$

□

Proposition 26 below is the key result in our proof of the main theorem. The idea of its proof is the following: starting at a record epoch, M_n has a positive (conditional) probability of growing exponentially fast until $M_n \geq \lambda p_n$ and then to remain greater than λp_n forever. Such a regular growth is made possible by maintaining many different vertices at level $F^*(X_n)$, no vertex containing too many particles, so that an averaging phenomenon occurs.

Proposition 26 *Conditional on the non-extinction of the tree, there exists $\lambda > 0$ such that there exists almost surely a time k such that, for all $n \geq k$, $M_n \geq \lambda p_n$ and $F^*(X_n) = F^*(X_{n-1}) + 1$.*

Proof of Proposition 26:

We take λ, $K \geq 1$, n_0 as they were defined in the proof of Lemma 20. Let $n_1 \geq n_0$ be such that, for all $n \geq n_1$, $\lambda p_n \geq K$, $p_{n+1}/p_n \leq \nu^2$. (Such an n_1 exists thanks to assumptions A1 and A2.) Assume throughout that r is a record epoch greater than n_1. Let r_1 be such that $\nu^{2r_1} \geq K$ (remember that $\nu > 1$). Let

$$R_1 = \{r + r_1 \text{ is a record epoch}, M_{r+r_1} \geq K\}.$$

Note that

$$\bigcap_{i=r}^{r+r_1-1} D_1(i) \subset R_1.$$

By Lemma 20,

$$\mathrm{P}\left[R_1 \mid \mathcal{F}_r\right] \geq g^{r_1}.$$

Let $r_2 = \inf\{k \geq 0 \mid \nu^{2k} K \geq \lambda p_{r+r_1+k}\}$ (r_2 is well-defined thanks to A2.) Let

$$R_2 = \{r + r_1 + r_2 \text{ is a record epoch}\} \cap \{M_{r+r_1+r_2} \geq \lambda p_{r+r_1+r_2}\} \cap H_{r+r_1+r_2}(\nu^{r_2}).$$

We set $D_3(r, r_1, -1) = \Omega$ and, for $k \geq 0$,

$$D_3(r, r_1, k) = \bigcap_{i=r+r_1}^{r+r_1+k} D_2(i, \nu^{i-r-r_1}).$$

According to the definition of r_2:

$$D_3(r, r_1, r_2 - 1) \cap R_1 \subset R_2.$$

For $0 \leq k \leq r_2 - 1$,

$$\mathrm{P}\left[D_3(r, r_1, k) \mid \mathcal{F}_{r+r_1+k}\right] = \mathrm{P}\left[D_2(r + r_1 + k, \nu^k) \,\middle|\, \mathcal{F}_{r+r_1+k}\right] \mathbf{1}_{D_3(r,r_1,k-1)}.$$

On $D_3(r, r_1, k-1) \cap R_1$, $M_{r+r_1+k} \geq \nu^{2k} K$ thanks to the definition of r_2, $H_{r+r_1+k}(\nu^k)$ occurs, and, in particular, $M_{r+r_1+k} \geq K$, and $\nu^k \leq M_{r+r_1+k}^{1/2}$. By Lemma 20, on R_1,

$$\mathrm{P}\left[D_3(r, r_1, k) \mid \mathcal{F}_{r+r_1+k}\right] \geq v(k) \mathbf{1}_{D_3(r,r_1,k-1)},$$

where we put:

$$v(k) = 1 - C_1 \exp\left(-C_2 (\nu^{2k} K)^{1/2}\right) - \exp\left(-C_3 \nu^k\right).$$

We conclude that, on R_1,

$$\mathrm{P}\left[R_2 \mid \mathcal{F}_{r+r_1}\right] \geq \prod_{k=0}^{r_2-1} v(k).$$

Let

$$R_3 = \{\forall n \geq r+r_1+r_2, F^*(X_{n+1}) = F^*(X_n)+1, M_n \geq \lambda p_n\}.$$

For $n < r+r_1+r_2$, we set $D_4(r,r_1,r_2,n) = \Omega$. For $n \geq r+r_1+r_2$,

$$D_4(r,r_1,r_2,n) = \bigcap_{i=r+r_1+r_2}^{n} D_2(i,(\lambda K^{-1}p_i)^{1/2}).$$

We have

$$D_4(r,r_1,r_2,+\infty) \cap R_2 \subset R_3.$$

$$\mathrm{P}\left[D_4(r,r_1,r_2,n) \mid \mathcal{F}_n\right] = \mathrm{P}\left[D_2(n,(\lambda K^{-1}p_n)^{1/2}) \;\middle|\; \mathcal{F}_n\right] \mathbf{1}_{D_4(r,r_1,r_2,n-1)}.$$

On $D_4(r,r_1,r_2,n-1) \cap R_2$, $M_n \geq \lambda p_n$, $H_n\left((\lambda K^{-1}p_n)^{1/2}\right)$ occurs, and in particular $M_n \geq K$, and $(\lambda K^{-1}p_n)^{1/2}) \leq M_n^{1/2}$, so that, by Lemma 20, on R_2,

$$\mathrm{P}\left[D_4(r,r_1,r_2,n) \mid \mathcal{F}_n\right] \geq w(n)\mathbf{1}_{D_4(r,r_1,r_2,n-1)},$$

where we put:

$$w(n) = 1 - C_1 \exp\left(-C_2(\lambda p_n)^{1/2}\right) - \exp\left(-C_3(\lambda K^{-1}p_n)^{1/2})\right).$$

We conclude that, on R_2,

$$\mathrm{P}\left[R_3 \mid \mathcal{F}_{r+r_1+r_2}\right] \geq \prod_{i \geq r+r_1+r_2} w(i),$$

the above infinite product being well-defined and positive according to A1. Setting

$$\Xi_1 = \prod_{k \geq 0} v(k) \ , \ \Xi_2 = \prod_{i \geq n_1} w(i),$$

we get that, if r is a record epoch,

$$\mathrm{P}\left[R_1 \cap R_2 \cap R_3 \mid \mathcal{F}_r\right] \geq g^{r_1} \times \Xi_1 \times \Xi_2 > 0.$$

Let

$$R = \{\exists\, r \ , \ \forall n \geq r \quad F^*(X_{n+1}) = F^*(X_n)+1\}$$

We have:

$$R_1 \cap R_2 \cap R_3 \subset R.$$

By Proposition 25, there is, conditional on non-extinction, an infinite number of record epochs. We conclude, thanks to Lemma 25, that conditional on non-extinction, R occurs almost surely. $\square$

Proposition 27 *Conditional on non-extinction, there exists $\lambda > 0$ such that, for all $a \geq 1$, and all $n \geq k_a$ (k_a is random),*

$$p_n^{-1} Z_F(X_n)([0, F^*(X_n) - a]) \leq 4(\theta\lambda)^{-1} \exp(-\beta a).$$

Proof of Proposition 27:

We use Proposition 26 and its proof. Conditional on non-extinction, there exists almost surely k such that, for all $n \geq k$, $M_n \geq \lambda p_n$ and $H_n((\lambda K^{-1} p_n)^{1/2})$ occurs. For such an n, $Z_F(X_{n+1})([0, F^*(X_{n+1}) - a])$ follows, conditionally to $\mathcal{G}_n$, a binomial law $\mathcal{B}(p_{n+1}, \phi_n)$, where

$$\phi_n \leq \exp(-\beta a) p_n (M_n')^{-1}.$$

One can easily check that

$$\{M_n' \geq (1/2)\theta\lambda p_n, Z_F(X_{n+1})([0, F^*(X_{n+1}) - a]) \leq 2\lambda\phi_n\}$$

occurs with conditional probability greater than $1 - C_{14} p_n \exp(-C_{15}(p_n)^{1/2})$. The second Borel-Cantelli lemma implies the result, according to A1.

$\square$

4.5 The random objective function case

We keep the same notations as before. Record epochs are defined exactly as in the Galton-Watson case. Here, $\mathcal{F}_n$ is the σ-algebra generated by X_k, Y_k, $v(e(X_k))$, for $0 \leq k \leq n-1$, and X_n, where, for $x \in \mathbb{T}$, $e(x)$ stands for the set of edges incident to x. The $\sigma-$algebra $\mathcal{G}_n$ is generated by $\mathcal{F}_n$, $e(X_n)$ and Y_n. The analog of Proposition 25 is

Proposition 28 *Almost surely on the random objective function F and on the realization of the algorithm, there is an infinite number of record epochs.*

Proof of Proposition 28 (sketch):

We have to prove that $F^*(X_n)$ is not bounded from above. The difference from the Galton-Watson case is that F^* may take arbitrarily large negative values, so we have to work a bit more. For a fixed $\lambda > 0$, we define a_n as the largest integer a such that:

$$Z_F(X_n)([a, F^*(X_n)]) \geq \lambda p_n.$$

We first prove that a_n cannot decrease more than an almost surely finite number of times. Conditionally to $\mathcal{F}_n$, the number of particles that are fixed by mutation is greater than $(1 - \epsilon)\mu\lambda p_n$ with probability greater than $1 - \exp(-C(\epsilon)\lambda p_n)$. A result analogous to Lemma 21 shows that, conditionally to $\mathcal{G}_n$ and to the inequality above, $Z_F(X_{n+1})([a_n, F^*(X_{n+1})])$ follows a binomial law $\mathcal{B}(p_{n+1}, \phi_n)$ with $\phi_n \geq \zeta\lambda$, where $\zeta > 1$ for ϵ and λ small enough, thanks to A7. Thus, the probability that $a_{n+1} < a_n$ is less than $\exp(-C_1 p_n)$. The Borel-Cantelli lemma proves the stated result. As a consequence, $F^*(X_n)$ is almost surely bounded from below. Thus, if $F^*(X_n)$ is bounded from above with positive probability, two situations may occur. Either, with positive probability there exists $x \in \mathbb{T}$ such that $F(x) = \limsup_n F^*(X_n)$, and such that $x \in X_n$ and $F^*(X_n) = F(x)$ an infinite number of times. In which case an argument similar to the proof in the Galton-Watson case, using the fact that there are almost surely arbitrarily large values of F somewhere in $\mathbb{T}$, works to get a contradiction. Or, with positive probability, there is an infinite number of times n such that X_n contains a vertex y in $\mathbb{T}$ for the first time, satisfying $F^*(X_n) = F(y) = \limsup_n F^*(X_n)$. For all $L \in \mathbb{Z}$, call $\Xi_n(L)$ the event that X_n contains a vertex y in $\mathbb{T}$ for the first time, satisfying $F^*(X_n) = F(y) = L$. Conditional on $\mathcal{F}_n$ and on $\Xi_n(L)$, the probability that $F(y(1)) = L + 1$ and that $y(1) \in X_{n+1}$, is bounded from below by a constant. By Lemma 25, if $\Xi_n(L)$ occurs an infinite number of times, we must have $\limsup F^*(X_n) \geq L + 1$. Thus, we get a contradiction in the second case as well, by taking $L = \limsup F^*(X_n)$.

□

Lemma 20 holds as in the Galton-Watson case. A sketch of the proof follows. Conditionally to $\mathcal{F}_n$ and B_n, n is a record epoch and the increments $F(x_i(j)) - F(x_i)$ form a collection of i.i.d. ± 1 random variables. Let

$$J_n = \sum_{i=1}^{L_n} \sum_{j=1}^{m} \mathbf{1}_{\{F(x_i(j)) = F^*(X_n)+1\}} N_n(x_i)$$

and let

$$E_1 = \{J_n \geq (1-\epsilon) m\rho M_n\}.$$

Lemma 22 shows that, on B_n,

$$\mathrm{P}\,[E_1^c \mid \mathcal{F}_n] \leq \exp(-C_1(\epsilon) M_n / T_n^*).$$

Define now

$$E_2 = \{M_n' \geq (1-\epsilon)^2 (1-\mu)\rho m (m+1)^{-1} M_n, F^*(Y_n) = F^*(X_n) + 1\}.$$

Lemma 23 shows that, on $B_n \cap E_1$,

$$\mathrm{P}\,[E_2^c \mid \mathcal{F}_n, F(\mathbb{T}_{n+1})] \leq \exp(-C_2(\epsilon) J_n).$$

The end of the proof is similar to the Galton-Watson case. With the two above results, that is, the existence of an infinite number of record epochs, and Lemma 20, we can prove propositions 26 and 27 exactly as in the Galton-Watson case.

4.6 Technical lemmas

Lemma 21 *The number of particles at level $F^*(Y_n)$ put by selection in X_{n+1} follows, conditionally to Y_n, a binomial law $\mathcal{B}(p_{n+1}, \mu_{n+1})$. For all $\psi < \exp\beta$, there exists $\lambda_0 > 0$ such that, for all $0 < \lambda < \lambda_0$,*

$$\mu_{n+1} \geq \psi \inf\left(\lambda, M'_n p_n^{-1}\right).$$

Proof:

Denote by $\mathcal{H}_n$ the set of indices $1 \leq i \leq p_n$ such that $F\left(Y_n^{(i)}\right) = F^*(Y_n)$. For all $i \notin \mathcal{H}_n$, we have, by definition, $F\left(Y_n^{(i)}\right) \leq F^*(Y_n) - 1$. By definition of the proportional selection,

$$\mu_{n+1} = M'_n \exp\left(\beta F^*(Y_n)\right) \left(\sum_{i=1}^{p_n} \exp\beta f\left(Y_n^{(i)}\right)\right)^{-1}.$$

Thanks to the preceding remark,

$$\sum_{i \notin \mathcal{H}_n} \exp\left(\beta f\left(Y_n^{(i)}\right)\right) \leq \exp(-\beta) \exp\left(\beta F^*(Y_n)\right)(p_n - M'_n).$$

Simplifying by $\exp(\beta F^*(Y_n))$ in the expression of μ_{n+1}, we conclude that:

$$\mu_{n+1} \geq M'_n \left(M'_n + (p_n - M'_n)\exp(-\beta)\right)^{-1}.$$

The two assertions stated in the lemma are easy consequences of this inequality. □

Lemma 22 *Let $\chi_1, \ldots, \chi_k$ be k nonnegative real numbers, χ^* be the maximum of the χ_i. Let $s_1, \ldots, s_k$ be i.i.d. random variables with common law supported by $[0, a]$. Let $\theta = \mathrm{E}s_1$. Then, for all $\epsilon > 0$, there exists $C > 0$ such that:*

$$\mathrm{P}\left[\sum_{i=1}^{k} s_i\chi_i \leq (1-\epsilon)\theta \sum_{i=1}^{k} \chi_i\right] \leq \exp\left(-C\epsilon\theta(\chi^*)^{-1}\sum_{i=0}^{k}\chi_i\right).$$

Proof:

By the exponential Chebyshev bound, this probability is bounded from above, for all $t \geq 0$, by:

$$\mathrm{E}\left(\exp\left(-t\sum_{i=1}^{k} s_i\chi_i\right)\right)\exp\left(t(1-\epsilon)\theta\sum_{i=1}^{k}\chi_i\right).$$

By independence,

$$\mathrm{E}\left(\exp\left(-t\sum_{i=1}^{k} s_i\chi_i\right)\right) = \exp\left(\sum_{i=1}^{k} \log \mathrm{E}(\exp\left(-ts_i\chi_i\right))\right).$$

For u near zero, $\mathrm{E}(\exp(-us_1)) = 1 - u\theta + o(u)$. In particular, for all $\gamma > 0$, there exists $\eta > 0$ such that, for all $t \in (0, \eta/\chi^*)$,

$$\log \mathrm{E}(\exp\left(-ts_i\chi_i\right)) \leq -(1-\gamma)t\theta\chi_i.$$

Hence, taking $\gamma = \epsilon/2$, an upper bound for the original probability is:

$$\exp\left(-t\theta\epsilon/2\sum_{i=1}^{k}\chi_i\right).$$

□

Lemma 23 *Let $\chi_1, \ldots, \chi_k$ be k nonnegative real numbers, and $s_1, \ldots, s_k$ in $[0,1]$. Let $V = \sum_{i=1}^{k} s_i\chi_i$. Consider now $Z_1, \ldots, Z_k$ independent binomial random variables with respective binomial laws $\mathcal{B}(\chi_i, s_i)$. Then, for all $\epsilon \in (0,1)$, there exists $C(\epsilon) > 0$ such that:*

$$\mathrm{P}\left[\sum_{i=1}^{k} Z_i \leq (1-\epsilon)V\right] \leq \exp\left(-C(\epsilon)V\right).$$

Proof:

The Laplace transform is given by:

$$\mathrm{E}\left(\exp\left(-t\sum_{i=1}^{k} Z_i\right)\right) = \exp\left(\sum_{i=1}^{k} \chi_i \log\left(1 + s_i(e^{-t} - 1)\right)\right).$$

For all $0 < \gamma' < 1$, there exists $\phi > 0$ such that, for all $0 \leq t \leq \phi$,

$$\exp\left(\sum_{i=1}^{k} \chi_i \log\left(1 + s_i(e^{-t} - 1)\right)\right) \leq \exp\left(-t(1-\gamma')\sum_{i=1}^{k} s_i\chi_i\right).$$

We conclude by an argument similar to the one used in the previous lemma. □

Lemma 24 *Let S be a random variable with binomial law $\mathcal{B}(n,p)$. For all $0 < p < 1$ and all $0 < \epsilon < 1$, we have:*

$$\mathrm{P}\left[S \leq n(1-\epsilon)p\right] \leq \exp\left(-(1/2)\epsilon^2 np\right).$$

$$\mathrm{P}\left[S \geq n(1+\epsilon)p\right] \leq \exp\left(-(1/2)\epsilon^2 np\right).$$

Proof:

We use the exponential Chebyshev bound, valid for every $t > 0$:

$$\mathrm{P}\left[S \le n(1-\epsilon)p\right] \le \exp\left(n(1-\epsilon)pt\right) \times \mathrm{E}\exp\left(-tS\right),$$

with $t = \epsilon$. The inequality:

$$\exp\left(-\epsilon\right) - 1 \le \epsilon^2/2 - \epsilon$$

yields the first stated inequality. We proceed likewise to get the second inequality. □

Lemma 25 *Let* $(\Omega, \mathcal{F}, \mathrm{P})$ *be a probability space with a filtration* $(\mathcal{F}_n)$*, an event* $A \in \sigma(\mathcal{F}_n, n \ge 0)$*, and* $(\tau_i)_{i\ge 1}$ *a family of not necessarily finite stopping times. We assume that for all* i *and* k

$$\mathrm{P}\left[A \mid \mathcal{F}_k\right] \mathbf{1}_{\{\tau_i = k\}} \ge \gamma \mathbf{1}_{\{\tau_i = k\}},$$

where γ *is a positive constant. Let*

$$B = \{\forall n, \exists i\ ,\ n \le \tau_i < +\infty\}.$$

Then $\mathrm{P}\left[B \setminus A\right] = 0$.

Proof:

Define, for $n \ge 0$, $s_n = \inf\{\tau_i\ :\ \tau_i \ge n\}$, with the convention $\inf \emptyset = +\infty$. Note that s_n is a $(\mathcal{F}_n)-$ stopping time. On B, each s_n is finite, hence

$$\mathbf{1}_B \le \sum_{u=n}^{+\infty} \mathbf{1}_{\{s_n = u\}} \le 1. \tag{4.2}$$

For all $u \ge n$, we have:

$$\mathrm{E}\left(\mathbf{1}_{\{s_n = u\}} \mathbf{1}_A \;\middle|\; \mathcal{F}_u\right) \ge \gamma \mathbf{1}_{\{s_n = u\}}.$$

Conditioning by $\mathcal{F}_n$ and summing, (4.2) implies that:

$$\mathrm{E}\left(\mathbf{1}_A \mid \mathcal{F}_n\right) \ge \gamma \mathrm{E}\left(\mathbf{1}_B \mid \mathcal{F}_n\right).$$

Thanks to the Lévy theorem, see [29] p.263, taking the limit $n \to +\infty$ in the above inequality yields:

$$\mathbf{1}_A \ge \gamma \mathbf{1}_B \quad \mathrm{P} - a.s.$$

□

4.7 Appendix: the regular case

Our techniques for dealing with random fitness landscapes work indeed in the case where the fitness landscape is regular. For instance, take $\mathbb{Z}$ as the state space, and $F(x) = x$, as the objective function. The fitness is $G(x) = \exp \beta F(x)$, and the mutation is a nearest neighbor random walk $X \longmapsto X + \epsilon$, where ϵ is independent from X with law $b\delta_1 + (1-b)\delta_{-1}$. The algorithm starts at 0.

As before:

Theorem 8 *If the following assumptions are met:*

- $\lim_{n\to+\infty} p_n (\log n)^{-1} = +\infty$,
- $\lim_{n\to+\infty} p_{n+1}/p_n = 1$,
- $\exp \beta > 1/b$,

almost surely with respect to the algorithm

$$F^*(X_n) = n - O(1).$$

Moreover, there exists $c > 0$ such that, for all $a \geq 1$, and for all $n \geq k_a$, where k_a is random and almost surely finite,

$$p_n^{-1} Z_F(X_n)((-\infty, F^*(X_n) - a]) \leq c \exp(-\beta a).$$

The assumption on the population growth is slightly weaker than in the random fitness landscape case since we do not have to control the repartition of the particles at level $F^*(X_n)$, but only the total number of particles at that level.

Moreover, we can show that the empirical distibution of the fitness relative to $F^*(X_n)$ converges in law toward the limiting law that we would get in the infinite population case.

Theorem 9 *Under the assumptions of Theorem 8, the sequence of measures μ_n defined on $2\mathbb{N}$ by*

$$\mu_n(a) = p_n^{-1} Z_F(X_n)(F^*(X_n) - a)$$

converges almost surely in law toward the probability measure $\nu_{b,\beta}$, defined, for all $u \in 2\mathbb{Z}$, by:

$$\nu_{b,\beta}(u) = \left(\frac{1-b}{b}\right)^{u/2} \left(\prod_{i=1}^{u/2} \frac{1}{\exp(2\beta i) - 1}\right) c_{b,\beta}$$

where $c_{b,\beta}$ is the normalizing constant.

The measures we consider are concentrated on $2\mathbb{N}$ since the population is started at a single point, all mutation steps being ± 1.

Proof (sketch):

We first study the infinite population case: let ζ_n be the law at time n of the infinite population model started at δ_0, and, for all $a \in 2\mathbb{N}$, $\nu_n(a) = \zeta_n(n-a)$. One first checks, using the proof of Lemma 21, that $\liminf \nu_n(0) > 0$. As an easy consequence, for all $u \in 2\mathbb{N}$, $\liminf \nu_n(u) > 0$. As in the proof of Proposition 27, one has : $\limsup \nu_n(u) \leq c \exp(-\beta u)$, and the sequence ν_n is tight. Then, putting, for large n, $r_n(u) = \nu_{n+1}(u+2)\nu_{n+1}(u)^{-1}$, one has the following induction equations:

$$r_{n+1}(0) = \exp(-2\beta)\left(r_n(0) + \frac{1-b}{b}\right),$$

and, for all even integer $u \geq 2$,

$$r_{n+1}(u) = \exp(-2\beta)\left(\frac{br_n(u)+1-b}{b+(1-b)r_n(u-2)^{-1}}\right).$$

Since $\exp(-2\beta) < 1$, these induction equations have a contraction property, and allow us to prove by induction on u that $r_n(u)$ converges, as n goes to infinity, to an explicit value. This result, together with the tightness of the sequence ν_n, implies that ν_n converges in law as n goes to infinity.

The proof for the empirical laws μ_n follows the same scheme. The fact that $\liminf \mu_n(0) > 0$ almost surely, and the almost sure tightness of the sequence (μ_n) stems from the previous results. Now, it is easy to check, using large deviations estimates on binomial random variables, and the second Borel-Cantelli lemma, as in the proof of Proposition 27, that almost surely $\liminf \mu_n(u) > 0$ for all $u \in 2\mathbb{N}$. Define the sequences $s_n(u)$ for μ_n in the same way as the r_n's are defined for ν_n. Using again the second Borel-Cantelli lemma and large deviations estimates around the average behavior, one gets that, for large n :

$$s_{n+1}(0) = \exp(-2\beta)\left(s_n(0) + \frac{1-b}{b}\right) + \eta_n(0),$$

and, for $u \geq 2$, $u \in 2\mathbb{N}$,

$$s_{n+1}(u) = \exp(-2\beta)\left(\frac{bs_n(u)+1-b}{b+(1-b)s_n(u-2)^{-1}}\right) + \eta_n(u),$$

where, for all u, $\eta_n(u)$ goes to zero almost surely. By the contracting character of these approximate induction equations, the η_n terms do not affect the limiting behavior, and $s_n(u)$ converges almost surely to the same value as $r_n(u)$ for all u.

$\square$

Liste de publications

[1] J. Bérard. Genetic algorithms in random environment : two examples, soumis.

[2] J. Bérard. Asymptotics of a genetic algorithm with additive space-inhomogenous mutations, Preprint.

[3] J. Bérard et A. Bienvenüe. Convergence of a Genetic Algorithm with finite population, Mathematics and Computer Science, Algorithms, Trees, Combinatorics and Probabilities, 155-163. Birkhäuser Verlag

[4] J. Bérard et A. Bienvenüe. Un principe d'invariance pour un algorithme génétique en population finie, C. R. Acad. Sci. Paris Sér. I Math. 331 469-474.

[5] J. Bérard Avec A. Bienvenüe. Sharp asymptotic results for simplified mutation-selection algorithms, 2001, soumis.

[6] A. Bienvenüe, M. Joannidès, J. Bérard, E. Fontenas et O. François. Niching in Monte-Carlo filtering algorithms, Proceedings of the EA 2001 conference on Artificial Evolution. À paraître dans Springer Lecture Notes in Computer Science.

Bibliography

[1] D. Aldous. A Metropolis-type optimization algorithm on the infinite tree, Algorithmica, 22: 388–412, 1998.

[2] S. Arora. Simulating Quadratic Dynamical Systems is PSPACE-complete. *ACM symposium on Theory of Computing*, 1994.

[3] T. Bäck. Evolutionary Algorithms in theory and practice. Oxford University Press, 1995.

[4] E. Baum, D. Boneh, and C. Garrett. On Genetic Algorithms. *COLT* , 1995.

[5] J. Bérard. Genetic Algorithms in Random Environment: two examples. Submitted, 2001.

[6] J. Bérard. Asymptotics of a genetic algorithm with additive space-inhomogenous mutation. Preprint.

[7] J. Bérard and A. Bienvenüe. Convergence of a genetic algorithm with finite population. In D. Gardy and A. Mokkadem, editors, *Mathematics and Computer Science. Algorithms, Trees, Combinatorics and Probabilities*, Trends in Mathematics, pages 155–163, Basel, 2000. Birkhäuser Verlag.

[8] J. Bérard and A. Bienvenüe. Un principe d'invariance pour un algorithme génétique en population finie. *C. R. Acad. Sci. Paris Sér. I Math.*, 331:469–474, October 2000.

[9] J. Bérard and A. Bienvenüe. Sharp asymptotic results for simplified mutation-selection algorithms. Submitted, 2001.

[10] H.-G. Beyer. An Alternative Explanation for the Manner in which Genetic Algorithms Operate. *Biosystems*, 41:1-15,1997.

[11] A. Bienvenüe, M. Joannidès, J. Bérard, O. Fontenas et O. François. Niching in Monte-Carlo filtering algorithms. Proceedings of the 5th International Conference on Artificial Evolution. To appear in Springer Lecture Notes in Computer Science.

[12] P. Billingsley. *Convergence of Probability Measures.* John Wiley & Sons, New York, 1968.

[13] D. Bonnaz. *Modèles stochastiques de dynamique des populations.* PhD Thesis, Université de Lausanne, 1998.

[14] P. Brémaud. *Markov chains, Gibbs Fields, Monte Carlo simulation and Queues.* Texts in Applied Mathematics, 31, Springer-Verlag, New-York, 1999.

[15] R. Cerf. Asymptotic convergence of genetic algorithms. *Adv. in Appl. Probab.*, 30(2):521–550, 1998.

[16] R. Cerf. The dynamics of mutation-selection algorithms with large population sizes. *Ann. Inst. H. Poincaré Probab. Statist* 32(4):455–508, 1996.

[17] F. Comets, N. Gantert, O. Zeitouni. Quenched, annealed and functional large deviations for one-dimensional random walk in random environment. *Probab. Theory Related Fields* 118 (2000), no. 1, 65–114.

[18] K. Deb and D.E. Goldberg. *An investigation of niche and species formation in genetic function optimization.* In J.D. Schaffer, editor, Proceedings of the Third International Conference on Genetic Algorithms, pp. 42–50. Morgan Kaufmann, Fairfax VA, 1989.

[19] P. Del Moral. Nonlinear Filtering : Interacting particle resolution. *Markov Processes and Related Fields*, Vol. 2, 1996.

[20] P. Del Moral. Measure Valued Processes and Interacting Particle Systems. Application to nonlinear Filtering Problems. *Ann. Appl. Probab.* 8(2), pages 438-495, 1998.

[21] P. Del Moral and A. Guionnet. On the stability of interacting processes with applications to filtering and genetic algorithms. *Ann. Inst. H. Poincaré Probab. Statist.* 37(2), pages 155–194, 2001.

[22] P. Del Moral, L. Kallel and J. Rowe. Modelling Genetic Algorithms with Interacting Particle Systems. In L. Kallel, B. Naudts, A. Rogers, editors, *Natural Computing Series: Theoretical Aspects of Evolutionary Computing, 10–67*, Springer Verlag, Berlin, 2001.

[23] P. Del Moral and L. Miclo. Branching and interacting particle systems approximations of Feynman-Kac formulae with applications to non-linear filtering. *Séminaire de probabilités*, Lecture notes in Math. 1729, Springer, 2000.

[24] A. Dembo, N. Gantert, Y. Peres, O. Zeitouni. Large Deviations for Random Walks on Galton-Watson Trees: Averaging and Uncertainty. To appear in *Probab. Theory Related Fields.*

[25] A. Dembo and O. Zeitouni. *Large deviations techniques and applications.* Springer-Verlag, New York, second edition, 1998.

[26] P. Diaconis, L. Saloff-Coste. Logarithmic Sobolev inequalities for finite Markov chains. *Ann. Appl. Probab.* 6, no. 3, 695–750, 1996.

[27] A. Doucet, J.F.D. Freitas and N.J. Gordon, Eds. *Sequential Monte Carlo methods in practice*, Springer-Verlag, New-York, 2001.

[28] P. Doyle and L. Snell. *Random walks and electric networks.* Carus Mathematical Monographs, 22. Mathematical Association of America, Washington D.C., 1984.

[29] R. Durrett. *Probability: theory and examples.* Duxbury Press, Belmont, CA, second edition, 1996.

[30] R. A. Fisher. *The Genetical Theory of Natural Selection.* Oxford University Press, 1930.

[31] L. J. Fogel, A. J. Owens, and M. J. Walsh. *Artificial Intelligence through Simulated Evolution.* John Wiley, New York, 1966.

[32] O. François. Global optimization with exploration/selection algorithms and simulated annealing. To appear in *The Annals of Applied Probability.*

[33] C.M. Fonseca and P. J. Fleming. *An overview of evolutionary algorithms in multiobjective optimization*. Evolutionary Computation, **3**:1 (1995), pp. 1–16.

[34] D. E. Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley, Reading, MA, 1989.

[35] D. E. Goldberg and J. Richardson. *Genetic Algorithms with Sharing for Multimodal Function Optimization*. In J. J. Grefenstette, editor, Genetic Algorithms and Their Applications: Proc. of the 2nd international Conference on Genetic Algorithms, July 28-31, Cambridge MIT, L. Erlbaum Associates, Hillsdale, MA, pp. 41–49, 1987.

[36] D. E. Goldberg and K. Deb. A comparative analysis of selection schemes used in genetic algorithms. In G .J. E. Rawlins, editor, *Foundations of Genetic Algorithms*, pages 69–93, Morgan Kaufmann, 1991. Birkhäuser Verlag.

[37] N. Gordon, D. Salmond and A. Smith. *Novel approach to nonlinear/non-Gaussian Bayesian state estimation*. IEEE Proceedings, Part F, **140** (1993), pp. 107–113.

[38] I. S. Gradshteyn and I. M. Ryzhik. *Table of integrals, series, and products.* Academic Press Inc., San Diego, CA, fifth edition, 1996.

[39] J. H. Holland. *Adaptation in natural and artificial systems.* University of Michigan Press, Ann Arbor, Mich., 1975. An introductory analysis with applications to biology, control, and artificial intelligence.

[40] J. Horn. The nature of niching: Genetic algorithms and the evolution of optimal, cooperative populations. PhD Thesis, Univ. Illinois at Urbana-Champaign, 1997.

[41] B. Hughes. *Random walks and Random environments. Vol. 2. Random environments.* The Clarendon Press, Oxford University Press, New York, 1996.

[42] R. E. Kalman. *A New Approach to Linear Filtering and Prediction Problems*, Transaction of the ASME–Journal of Basic Engineering (1960), pp. 35-45.

[43] D. Kessler, H. Levine, D. Ridgway and L. Tsimring (1997). Evolution on a smooth landscape. *J. of Statist. Physics*, 87(3/4): 519-544.

[44] M. Kimura. *The Neutral Theory of Molecular Evolution.* Cambridge University Press, 1983.

[45] J. R. Koza. *Genetic Programming: On the Programming of Computers by means of Natural Evolution.* MIT Press, 1994.

[46] H. Kunita. *Asymptotic behavior of non-linear filtering errors of Markov processes*, J. Multivariate Analysis, **1**:4 (1971), pp. 365–393.

[47] R. Lyons, R. Pemantle, Y. Peres. Biased random walks on Galton-Watson trees, Probab. Theory Related Fields, 106: 249–264, 1996

[48] R. Lyons, Y. Peres. *Probability on Trees and Networks.* Preliminary draft available at http://www.math.gatech.edu/~rdlyons

[49] M. Matsumoto and T. Nishimura. Mersenne twister a 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Transactions on Modeling and Computer Simulation*, 8(1):3–30, January 1998. Special issue on uniform random number generation.

[50] C. Mazza and D. Piau. On the effect of selection in genetic algorithms. *Random Structures and Algorithms*, 18:185–200, 2001.

[51] S. P. Meyn and R. L. Tweedie. *Markov Chains and Stochastic Stability.* Springer-Verlag, Berlin, 1996.

[52] R.N. Miller, E.F. Carter, S.T. Blue. *Data Assimilation into nonlinear stochastic models*, Tellus, **51A** (1999), pp. 167–194.

[53] C. Musso and N. Oudjane. *Particle methods for multimodal filtering.* In Proc. of the second international conference on Information Fusion, Silicon Valley, CA, July 6-8, IEEE Press, pp. 785–792, 1999.

[54] C. Musso, N. Oudjane, F. Legland. *Improving regularized particle filters.* In A. Doucet, J.F.D. Freitas and N.J. Gordon, eds: Sequential Monte Carlo methods in practice, Springer-Verlag, New-York, 2001.

[55] C. Neuhauser. Mathematical models in population genetics. In D. Balding, M. Bishop and C. Cannings, editors, *Handbook of Statistical Genetics*, pages 179–208, New York, 2001. John Wiley and Sons.

[56] C. Papadimitriou. *Computational Complexity.* Addison Wesley, MA, 1994.

[57] R. Pemantle. Tree-indexed processes. *Statist. Sci.*, 10: 200–213, 1995.

[58] A. S. Perelson and S. A. Kauffman, Editors. *Molecular Evolution on rugged landscapes: Proteins, RNA, and the immune system,* Santa Fe Institue, Addison Wesley, CA, 1991.

[59] Y. Peres. Probability on Trees: an Introductory Climb. *Springer Lecture Notes in Mathematics,* 1717, pages 193-280, 1999.

[60] D. Piau. *Marches au hasard, arbres aléatoires, biologie moléculaire et algorithmes génétiques.* Habilitation à diriger les recherches, Université Lyon 1, 2000.

[61] A. Prugel-Benett and J. L. Shapiro. An analysis of genetic algorithms using statistical mechanics. *Phys. Rev. Lett.*, 72(9):1305-1309, 1994.

[62] Y. Rabinovich and A. Wigderson. Techniques for bounding the convergence rate of genetic algorithms. *Random Structures and Algorithms*, 14(2):111–138, 1999.

[63] M. Rattray. *Modelling the dynamics of Genetic Algorithms using Statistical Mechanics.* PhD thesis, Manchester University, 1996.

[64] I. Rechenberg. *Evolutionstrategie: Optimierung Technischer System nach Prinzipien des Biologischen Evolution.* Fromman-Holzboog Verlag, Stuttgart, 1973.

[65] D. Revuz and M. Yor. *Continuous Martingales and Brownian Motion.* Springer-Verlag, Berlin, 1991.

[66] G. Rüdolph. *Convergence Properties of Evolutionary Algorithms.* Hamburg: Kovac, 1997.

[67] H. -P. Schwefel. *Numerical Optimization of Computer Models.* John Wiley & Sons, New-York, second edition, 1995.

[68] B. W. Silverman. Density estimation. Chapman and Hall, New-York, 1986.

[69] F. Spitzer. *Principles of random walks.* Graduate Texts in Mathematics, Vol. 34. Springer-Verlag, New York-Heidelberg, second edition, 1976.

[70] A.-S. Sznitman. Slowdown and neutral pockets for a random walk in random environment. *Probab. Theory Related Fields* 115 (3), 287–323, 1999.

[71] S. Tavaré, R. C. Griffiths and P. Donnelly (1997). Inferring coalescence times for molecular sequence data. *Genetics* 145:505-518.

[72] L. Tsimring, H. Levine, and D. A. Kessler. RNA virus evolution via a fitness-space model. *Phys. Rev. Lett.*, 76:4440, 1996.

[73] M. Vose. *The Simple Genetic Algorithm. Foundations and theory.* MIT Press, Cambridge MA, 1999.

[74] M. Zerner. Lyapounov exponents and quenched large deviations for multidimensional random walk in random environment. *Ann. Probab.* 26 (1998), no. 4, 1446–1476.