Mathematical study of stochastic models of evolution belonging to the ecological theory of adaptive dynamics

Nicolas Champagnat

To cite this version:

Nicolas Champagnat. Mathematical study of stochastic models of evolution belonging to the ecological theory of adaptive dynamics. Mathematics [math]. Université de Nanterre - Paris X, 2004. English. NNT : . tel-00091929

HAL Id: tel-00091929
https://theses.hal.science/tel-00091929

Submitted on 7 Sep 2006

UNIVERSITÉ PARIS X – NANTERRE
U.F.R. SEGMI – ÉQUIPE MODAL'X

N° attribué par la bibliothèque

|_|_|_|_|_|_|_|_|_|_|

THÈSE

POUR OBTENIR LE GRADE DE

**DOCTEUR DE
L'UNIVERSITÉ PARIS X**

Discipline : MATHÉMATIQUES

PRÉSENTÉE ET SOUTENUE PUBLIQUEMENT PAR

Nicolas CHAMPAGNAT

LE 6 DÉCEMBRE 2004

Étude mathématique de modèles stochastiques d'évolution issus de la théorie écologique des dynamiques adaptatives

M.	RÉGIS FERRIÈRE,	CO-DIRECTEUR, BIOLOGIE
M.	JEAN JACOD,	EXAMINATEUR
M.	CHRISTIAN LÉONARD,	EXAMINATEUR
MME.	SYLVIE MÉLÉARD,	DIRECTRICE DE THÈSE
M.	ÉTIENNE PARDOUX,	EXAMINATEUR
M.	ALAIN ROUAULT,	RAPPORTEUR

Remerciements

Je voudrais tout d'abord exprimer ma profonde gratitude à Sylvie Méléard qui a su me guider tout au long de cette thèse. Sa disponibilité exemplaire, ses conseils précieux, son soutien permanent et, souvent, sa patience m'ont permis de mener à bien ce travail. Je lui suis très reconnaissant de m'avoir fait bénéficier de son dynamisme et de son expérience pour mes premiers pas dans la recherche.

Je dois également énormément à Régis Ferrière. Par ses qualités humaines et par l'effort permanent qu'il investit dans le rapprochement des communautés biologiques et mathématiques, il fut pour moi un guide précieux dans les méandres de la littérature biologique, dont il a su me présenter les idées de façon accessible et stimulante pour un étudiant en mathématique. Son contact et son point de vue pluridisciplinaire sont pour moi une source de motivation continuelle.

Je souhaite également remercier Gérard Ben Arous pour m'avoir proposé ce sujet et avoir guidé mes premiers pas. Ses conseils et nos discussions à l'EPFL m'ont dévoilé la richesse et l'exigeance du monde de la recherche en mathématiques.

Je tiens à remercier vivement Amir Dembo et Alain Rouault de m'avoir fait l'honneur de rapporter ma thèse. Je leur témoigne toute ma reconnaissance d'avoir porté leur regard de mathématicien sur des problèmes biologiques dont ils n'étaient peut-être pas familiers. Je remercie également Alain Rouault de faire partie de mon jury.

Je voudrais remercier Jean Jacod, Christian Léonard et Etienne Pardoux de participer à mon jury de thèse. Je suis reconnaissant à chacun d'entre eux de l'intérêt qu'ils ont porté à mon travail. Je remercie également Jean Jacod pour ses précieux conseils, Christian Léonard pour nos discussions stimulantes, et Etienne Pardoux pour m'avoir aidé de ses conseils lorsque je cherchais un sujet de thèses à l'interface mathématiques-biologie.

J'ai eu beaucoup de plaisir à partager la vie de l'équipe Modal'X de l'université de Nanterre pendant ces trois années. Je remercie également les membres du laboratoire d'écologie de l'ENS Paris que j'ai cotoyé pendant cette thèse. Je n'oublie pas tous les thésards qui m'ont accompagné pendant ces années, notamment à Nanterre. Leur présence a constitué pour moi un soutien et un réconfort.

Je tiens à évoquer également le groupe de travail de l'ACI "populations structurées" coordonnée par Sylvie Méléard pour les discussions fructueuses du vendredi.

Mes derniers remerciements, et non les moindres, vont à mes amis et ma famille pour leur présence, leur soutien et leurs encouragements. Un grand merci notamment à Gaël pour nos parties de billard endiablées qui m'ont permi de m'aérer l'esprit, et à Mylène pour son aide et pour le stress de dernière minute partagé. Mes remerciements les plus profonds vont à Brigitte, qui a supporté avec patience mon humeur, mes découragement, mes questionnements perpétuels. Elle a su trouver les mots pour conserver ma motivation, et il est certain que, sans elle, cette thèse n'aurait jamais été menée à son terme.

Table des matières

Table des figures

Introduction

Cette thèse porte sur l'étude de processus stochastiques modélisant l'évolution adaptative des espèces, selon l'approche formelle de la “théorie des dynamiques adaptatives”. Ce nouveau domaine de la biologie théorique soulève de nombreuses questions mathématiques originales. Nous chercherons en particulier à justifier rigoureusement les modèles fondamentaux de cette théorie, à partir d'une description particulaire du phénomène de l'évolution darwinienne, puis nous étudierons en détail un modèle de diffusion dégénérée apparaissant dans le cadre de cette théorie.

Cette introduction s'organise en trois parties. La première est consacrée à la description du modèle particulaire sur lequel se fonde notre travail. Il s'agit d'un processus de sauts de Markov à valeurs dans un espace de mesures. La seconde décrit le contenu des quatre chapitres de cette thèse. Nous y énonçons les principaux résultats et donnons une idée de leur preuve. Nous terminons par une section de conclusions et perspectives.

Les quelques notions biologiques nécessaires à la compréhension de cette introduction seront définies au fur et à mesure. Pour plus de confort, nous les avons regroupées dans un glossaire couvrant également les notions biologiques utilisées dans le chapitre I, consacré à un développement détaillé de la biologie des dynamiques adaptatives.

1 Modélisation individuelle : système de particules en interaction

Le principe de la sélection naturelle est connu depuis Charles Darwin. L'évolution repose principalement sur trois ingrédients : l'*hérédité*, qui transmet les caractéristiques individuelles d'une génération à l'autre, une source de *variation* dans ces caractéristiques (comme les mutations ou le mécanisme de la reproduction sexuée) et enfin un mécanisme de *sélection*, pouvant résulter de la compétition entre les individus d'une même espèce (par exemple pour des ressources énergétiques ou pour la reproduction) ou de la compétition avec d'autres espèces (prédateurs, parasites, etc.).

Les dynamiques adaptatives proposent des modèles de l'évolution d'une population asexuée (reproduction clonale) *à l'échelle de la population* (le détail individuel de la dynamique temporelle de la population n'est pas décrit). Or, dans une population asexuée, seules les mutations sont source de variation parmi les caractéristiques individuelles, et la sélection résulte d'avantages individuels dans des populations mutantes de petit effectif, puisque constituées initialement d'un seul individu.

Afin de donner une justification mathématique rigoureuse de tels modèles, il est donc nécessaire de s'appuyer sur un modèle d'évolution *à l'échelle de l'individu*, simple mais réaliste.

Nous nous limiterons dans cette thèse à l'étude de l'évolution d'une seule espèce. Le lecteur pourra, pour l'étude biologique des dynamiques adaptatives de plusieurs espèces en interaction, se référer à Marrow *et al.* [57], Dieckmann et Law [18], Champagnat *et al.* [10], et Gragnani *et al.* [39].

Considérons un ou plusieurs traits quantitatifs caractérisant les individus d'une population (par exemple la taille, l'âge à maturité ou un taux d'acquisition de ressources nutritives) à valeurs dans un espace $\mathcal{X} \subset \mathbb{R}^d$ ($d \geq 1$), que nous supposerons dans cette introduction *fermé*. Nous allons construire un processus de Markov modélisant la naissance et la mort de chaque individu, avec hérédité clonale (un enfant a le même trait que son père), sauf lorsqu'une mutation a lieu, et avec une compétition modélisée par une densité-dépendance négative, c'est-à-dire une dépendance de la démographie (survie et fécondité) d'un individu par rapport à l'ensemble des individus présents dans la population.

La population est constituée à chaque instant $t \geq 0$ d'un nombre fini N_t d'individus, dont les traits sont notés $x_1,\dots,x_{N_t}$. Les paramètres du modèle sont les suivants :

- un individu de trait $x \in \mathcal{X}$ donne naissance à un nouvel individu avec taux $b(x) \in \mathbb{R}_+$;
- un individu de trait x dans une population de N_t individus de traits $x_1,\dots,x_{N_t}$ meurt avec taux $d(x) + \sum_{i=1}^{N_t} \alpha(x,x_i)$, où $d(x) \in \mathbb{R}_+$ est le taux de mort "naturelle" d'un individu de trait x, et $\alpha(x,y) \in \mathbb{R}_+$ est un noyau d'interaction modélisant l'effet d'un individu de trait y dans la population sur le taux de mort d'un individu de trait x ;
- chaque naissance donne lieu à une mutation avec probabilité $\mu(x) \in [0,1]$;
- le trait d'un individu mutant né d'un individu de trait x est donné par $y = x + h$, où h est dirigé par une variable aléatoire de loi $m(x,dh)$, de support inclu dans $\mathcal{X} - x = \{y - x : y \in \mathcal{X}\}$.

Ce type de modèle est aussi utilisé par les biologistes pour étudier la dynamique de la répartition spatiale d'une population, en remplaçant les mutations par des dispersions à la naissance (penser aux graines chez les plantes). De tels modèles ont notamment été utilisés, pour étudier l'évolution de la population, par Dieckmann [15], Metz *et al.* [59] et Ferrière *et al.* [32], pour étudier sa répartition spatiale, par Bolker et Pacala [5, 6], Law et Dieckmann [53], Law *et al.* [54], et pour étudier les deux phénomènes combinés, par Doebeli et Dieckmann [22].

La densité-dépendance du taux de mort peut être symétrique ($\alpha(x,y) = \beta(|x-y|)$, cf. Dieckmann et Doebeli [16] et Doebeli et Dieckmann [21]), ou asymétrique (par exemple $\alpha(x,y) = \beta(x-y)$ avec $\beta(\cdot)$ monotone, cf. Kisdi [49] et Kisdi et Geritz [50]).

1.1 Construction du modèle individuel

Etant donné que le nombre d'individus et de traits présents est susceptible de varier au cours du temps, il est naturel du point de vue mathématique de représenter une population de N_t individus de traits $x_1,\dots,x_{N_t}$ à l'instant $t \geq 0$, par la mesure de comptage

$$\nu_t = \sum_{i=1}^{N_t} \delta_{x_i}$$

où δ_x est la mesure de Dirac en x. Le taux de mort d'un individu de trait x dans une population ν_t peut alors s'écrire comme $d(x) + \int_{\mathcal{X}} \alpha(x,y)\nu_t(dy)$. Le nombre total d'individus et le nombre d'individus portant un trait x donné à l'instant t sont donnés par $\langle \nu_t, \mathbf{1} \rangle$ et $\langle \nu_t, \mathbf{1}_{\{x\}} \rangle$, où nous avons utilisé la notation $\langle \nu, f \rangle$ pour $\int f(x)\nu(dx)$.

Nous étudions donc un processus de saut de Markov à valeurs dans l'espace des mesures de comptage finies sur $\mathcal{X}$

$$\mathcal{M} = \left\{ \sum_{i=1}^{n} \delta_x : n \geq 0,\, x_1,\dots,x_n \in \mathcal{X} \right\}$$

dont le générateur infinitésimal s'écrit, pour toute fonction $\phi : \mathcal{M} \to \mathbb{R}$

$$\begin{aligned} L\phi(\nu) = & \int_{\mathcal{X}} [\phi(\nu + \delta_x) - \phi(\nu)](1 - \mu(x))b(x)\nu(dx) \\ & + \int_{\mathcal{X}} \int_{\mathbb{R}^d} [\phi(\nu + \delta_{x+h}) - \phi(\nu)]\mu(x)b(x)m(x,dh)\nu(dx) \\ & + \int_{\mathcal{X}} [\phi(\nu - \delta_x) - \phi(\nu)] \left(d(x) + \int_{\mathcal{X}} \alpha(x,y)\nu(dy) \right) \nu(dx) \end{aligned} \tag{1}$$

Le premiere terme (linéaire) décrit les naissances sans mutation, le second (linéaire) celles avec mutation, et le troisième (non linéaire) les morts par vieillesse ou compétition.

Fournier et Méléard [33] ont donné une construction d'un tel processus, sous l'hypothèse suivante :

(H) Il existe des constantes $\bar{b}$, $\bar{d}$ et $\bar{\alpha}$ telles que

$$0 \leq b(\cdot) \leq \bar{b}, \quad 0 \leq d(\cdot) \leq \bar{d} \quad \text{et} \quad 0 \leq \alpha(\cdot,\cdot) \leq \bar{\alpha}.$$

De plus, la mesure de probabilité $m(x,dh)$ a une densité $m(x,h)$ sur $\mathbb{R}^d$ par rapport à la mesure de Lebesgue, et il existe une fonction $m(h)$ mesurable telle que

$$\forall x \in \mathcal{X},\ \forall h \in \mathbb{R}^d,\ m(x,h) \leq m(h) \quad \text{et} \quad \int_{\mathbb{R}^d} m(h)dh < +\infty.$$

Rappelons cette construction. Soit $(\Omega,\mathcal{F},\mathbf{P})$ un espace de probabilité suffisamment riche pour contenir les cinq objets aléatoires indépendants suivants :

(i) une variable aléatoire ν_0 à valeurs dans $\mathcal{M}$ (la condition initiale) ;

(ii) une mesure de Poisson ponctuelle $N_1(ds,di,dv)$ sur $[0,\infty[\times\mathbb{N}\times[0,1]$ de mesure intensité $q_1(ds,di,dv) = \bar{b}ds\sum_{k\geq 1}\delta_k(di)dv$ (la mesure de Poisson ponctuelle des naissances sans mutation) ;

(iii) une mesure de Poisson ponctuelle $N_2(ds,di,dh,dv)$ sur $[0,\infty[\times\mathbb{N}\times\mathbb{R}^d\times[0,1]$ de mesure intensité $q_2(ds,di,dh,dv) = \bar{b}ds\sum_{k\geq 1}\delta_k(di)m(h)dhdv$ (la mesure de Poisson ponctuelle des naissances avec mutation) ;

(iv) une mesure de Poisson ponctuelle $N_3(ds,di,dv)$ sur $[0,\infty[\times\mathbb{N}\times[0,1]$ de mesure intensité $q_3(ds,di,dv) = \bar{d}ds\sum_{k\geq 1}\delta_k(di)dv$ (la mesure de Poisson ponctuelle des morts "naturelles") ;

(v) une mesure de Poisson ponctuelle $N_4(ds,di,dj,dv)$ sur $[0,\infty[\times\mathbb{N}\times\mathbb{N}\times[0,1]$ de mesure intensité $q_4(ds,di,dj,dv) = \bar{\alpha}ds\sum_{k\geq 1}\delta_k(di)\sum_{l\geq 1}\delta_l(dj)dv$ (la mesure de Poisson ponctuelle des morts par compétition).

Afin de résoudre le problème d'énumération des traits de la population, introduisons la fonction $H = (H^1,\dots,H^k,\dots) : \mathcal{M} \to (\mathbb{R}^d)^{\mathbb{N}^*}$ définie par

$$H\left(\sum_{i=1}^{n}\delta_{x_i}\right) = (x_{\sigma(1)},\dots,x_{\sigma(n)},0,\dots,0,\dots),$$

où $x_{\sigma(1)} \preccurlyeq \dots \preccurlyeq x_{\sigma(n)}$, pour l'ordre lexicographique $\preccurlyeq$ sur $\mathbb{R}^d$.

Introduisons le processus $(\nu_t,t \geq 0)$ à valeurs dans l'espace $\mathbb{D}(\mathbb{R}_+,\mathcal{M})$ des fonctions càdlàg de $\mathbb{R}_+$ dans $\mathcal{M}$, défini par

$$\begin{aligned}
\nu_t = \nu_0 &+ \int_0^T \int_{\mathbb{N}} \int_0^1 \mathbf{1}_{\{i \leq \langle \nu_{s-},\mathbf{1}\rangle\}} \delta_{H^i(\nu_{s-})} \\
&\qquad \mathbf{1}_{\left\{v \leq \frac{[1-\mu(H^i(\nu_{s-}))]b(H^i(\nu_{s-}))}{\bar{b}}\right\}} N_1(ds,di,dv) \\
&+ \int_0^T \int_{\mathbb{N}} \int_{\mathbb{R}^d} \int_0^1 \mathbf{1}_{\{i \leq \langle \nu_{s-},\mathbf{1}\rangle\}} \delta_{H^i(\nu_{s-})+h} \\
&\qquad \mathbf{1}_{\left\{v \leq \frac{\mu(H^i(\nu_{s-}))b(H^i(\nu_{s-}))}{\bar{b}} \frac{m(H^i(\nu_{s-}),h)}{m(h)}\right\}} N_2(ds,di,dh,dv) \\
&- \int_0^T \int_{\mathbb{N}} \int_0^1 \mathbf{1}_{\{i \leq \langle \nu_{s-},\mathbf{1}\rangle\}} \delta_{H^i(\nu_{s-})} \mathbf{1}_{\left\{v \leq \frac{d(H^i(\nu_{s-}))}{\bar{d}}\right\}} N_3(ds,di,dv) \\
&- \int_0^T \int_{\mathbb{N}} \int_{\mathbb{N}} \int_0^1 \mathbf{1}_{\{i \leq \langle \nu_{s-},\mathbf{1}\rangle\}} \mathbf{1}_{\{j \leq \langle \nu_{s-},\mathbf{1}\rangle\}} \delta_{H^i(\nu_{s-})} \\
&\qquad \mathbf{1}_{\left\{v \leq \frac{\alpha(H^i(\nu_{s-}),H^j(\nu_{s-}))}{\bar{\alpha}}\right\}} N_4(ds,di,dj,dv).
\end{aligned} \tag{2}$$

Le principe de cette écriture est simple, et décrit exactement comment simuler le processus : la mesure de Poisson associée à chaque type d'événement saute plus souvent que le processus ν ne le devrait. On décide alors si un saut a lieu en comparant v à une quantité liée au taux de l'événement correspondant. L'indicatrice faisant intervenir i et j permet de ne prendre en compte que les individus vivants à l'instant considéré.

La formule d'Itô pour processus de saut permet de montrer aisément, sous l'hypothèse (H), qu'un processus solution de (2) a bien (1) pour générateur infinitésimal ([33], proposition 2.6). De plus, le lemme de Gronwall permet de montrer que, si $\mathbf{E}[\langle \nu_0,\mathbf{1}\rangle] < +\infty$, un tel processus vérifie $\mathbf{E}[\sup_{t\in[0,T]} \langle \nu_t,\mathbf{1}\rangle] < +\infty$ pour tout $T > 0$ ([33], théorème 3.1). On peut montrer dans ce cas qu'il n'y a pas accumulation des sauts du processus, ce qui permet de déduire aisément l'existence et l'unicité de la solution ν de (2).

1.2 Limite des grandes populations

Nous allons nous intéresser dans cette thèse à une asymptotique particulière du processus précédent, faisant intervenir en particulier la limite de grande population ci-dessous, introduite dans [33].

Afin d'augmenter la taille de la population, il est biologiquement naturel de diminuer la force de la compétition en divisant le noyau $\alpha(\cdot,\cdot)$ par une constante $K > 0$ ayant vocation à tendre vers $+\infty$ (K représentera l'ordre de grandeur de la taille initiale de la population). Cette constante peut être biologiquement interprétée comme renormalisant l'espace ou les ressources disponibles.

Dans l'équation logistique classique ($\dot{n} = (r - kn)n$), on fait souvent apparaître une constante K similaire en écrivant l'EDO sous la forme

$$\dot{n} = r\left(1 - \frac{n}{K}\right)n.$$

K s'interprète comme la densité à l'équilibre de cette équation, et prend le nom de "capacité de charge" (*carrying capacity*).

Du point de vue mathématique, diviser le noyau de compétition $\alpha(\cdot,\cdot)$ par K correspond à une renormalisation naturelle des processus de sauts. En effet, dans le cas où la population initiale est *monomorphique* de trait x (c'est-à-dire composée d'individus ayant tous le même trait x) et où il n'y a pas de mutations ($\mu \equiv 0$), la population reste monomorphique en tout temps et, à l'instant $t \geq 0$, le taux de naissance global vaut $\langle \nu_t, \mathbf{1} \rangle b(x) = K(\langle \nu_t, \mathbf{1} \rangle / K) b(x)$, et le taux de mort $K(\langle \nu_t, \mathbf{1} \rangle / K)(d(x) + \alpha(x,x)(\langle \nu_t, \mathbf{1} \rangle / K))$. Si l'effectif initial de la population est proportionnel à K, on retrouve un cas particulier de la renormalisation classique des processus de sauts à valeurs dans $\mathbb{Z}^d$ étudiée dans le chapitre 11 de Ethier et Kurtz [28].

Guidés par ces observations, posons

$$\nu_t^K = \frac{1}{K}\nu_t,$$

et

$$\mathcal{M}^K = \left\{ \frac{1}{K}\nu : \nu \in \mathcal{M} \right\} \subset \mathcal{M}_F,$$

où $\mathcal{M}_F$ est l'ensemble des mesures positives finies sur $\mathcal{X}$.

Le taux de mort d'un individu de trait x dans une population $\nu = K\nu^K$ s'écrit alors $d(x) + \int (\alpha(x,y)/K)\nu(dy) = d(x) + \int \alpha(x,y)\nu^K(dy)$.

On peut donner une construction similaire à (2) du processus de Markov $(\nu_t^K, t \geq 0)$ à valeurs dans $\mathcal{M}^K$, dont le générateur infinitésimal s'écrit, pour toute fonction mesurable bornée $\phi : \mathcal{M}^K \to \mathbb{R}$,

$$\begin{aligned} L^K\phi(\nu) = & \int_{\mathcal{X}} \left[\phi\left(\nu + \frac{\delta_x}{K}\right) - \phi(\nu) \right] (1 - \mu(x)) b(x) (K\nu(dx)) \\ & + \int_{\mathcal{X}} \int_{\mathbb{R}^d} \left[\phi\left(\nu + \frac{\delta_{x+h}}{K}\right) - \phi(\nu) \right] \mu(x) b(x) m(x,dh) (K\nu(dx)) \\ & + \int_{\mathcal{X}} \left[\phi\left(\nu - \frac{\delta_x}{K}\right) - \phi(\nu) \right] \left(d(x) + \int_{\mathcal{X}} \alpha(x,y)\nu(dy) \right) (K\nu(dx)). \quad (3) \end{aligned}$$

Le résultat suivant montre que la renormalisation de $\alpha(\cdot,\cdot)$ en $1/K$ avec $K \to +\infty$ correspond effectivement à une asymptotique de grande population :

Théorème 1 (Fournier et Méléard [33]) *Sous (H), considérons une suite de processus* $(\nu^K)_{K\geq 1}$ *de générateurs infinitésimaux (3), telle que* ν_0^K *converge en loi pour la topologie vague sur* $\mathcal{M}_F$ *quand* $K \to +\infty$ *vers une mesure déterministe, positive et finie,* ξ_0 *sur* $\mathcal{X}$*, et que* $\sup_{K\geq 1} \mathbf{E}[\langle \nu_0^K, \mathbf{1}\rangle^3] < +\infty$*. Alors, pour tout* $T > 0$*, la suite* (ν^K) *converge en loi dans* $\mathbb{D}([0,T],\mathcal{M}_F)$ *pour la topologie de Skorohod, vers une fonction déterministe et continue* $(\xi_t)_{t\in[0,T]}$*, unique solution satisfaisant* $\sup_{t\in[0,T]}\langle \xi_t, \mathbf{1}\rangle < +\infty$ *de l'équation aux dérivées partielles faible, définie pour toute fonction* $f : \mathcal{X} \to \mathbb{R}$ *mesurable bornée, par*

$$\begin{aligned}\langle \xi_t, f\rangle = \langle \xi_0, f\rangle &+ \int_0^t \langle \xi_s(dx), (1-\mu(x))b(x)f(x)\rangle ds \\ &+ \int_0^t \left\langle \xi_s(dx), \mu(x)b(x)\int_{\mathbb{R}^d} f(x+h)m(x,h)dh \right\rangle ds \\ &- \int_0^t \left\langle \xi_s(dx), f(x)\left(d(x) + \int_{\mathcal{X}} \alpha(x,y)\xi_s(dy)\right)\right\rangle ds. \qquad (4)\end{aligned}$$

Ce théorème donne une interprétation microscopique à certains modèles classiques d'évolution de type EDP intégrales en population infinie (cf. Kimura [47], Bürger [7], Bürger et Bomze [8] et Diekmann *et al.* [20]).

Remarque 1 *Observons que, dans le cas où* $\mu \equiv 0$ *et* $\nu_0^K = n_0^K \delta_x$ *(une population* monomorphique*) avec* $n_0^K \to n_0$ *quand* $K \to +\infty$*, la fonction* ξ *du théorème s'écrit* $\xi_t = n(t)\delta_x$*, où* $n(\cdot)$ *est solution de l'équation logistique classique*

$$\dot n = n(b(x) - d(x) - \alpha(x,x)n) \quad \text{et} \quad n(0) = n_0. \qquad (5)$$

De même, si $\mu \equiv 0$ *et* $\nu_0^K = n_0^{K,x}\delta_x + n_0^{K,y}\delta_y$ *(une population* dimorphique*) avec* $n_0^{K,x} \to n_0^x$ *et* $n_0^{K,y} \to n_0^y$ *quand* $K \to +\infty$*, la fonction* ξ *du théorème s'écrit* $\xi_t = n_x(t)\delta_x + n_y(t)\delta_y$*, avec* $n_x(0) = n_0^x$*,* $n_y(0) = n_0^y$*, et*

$$\begin{cases} \dot n_x = n_x(b(x) - d(x) - \alpha(x,x)n_x - \alpha(x,y)n_y) \\ \dot n_y = n_y(b(y) - d(y) - \alpha(y,x)n_x - \alpha(y,y)n_y), \end{cases} \qquad (6)$$

Plus généralement, si la condition initiale est k-morphique (c'est-à-dire composée d'individus présentant k valeurs distinctes des traits) et converge vers un état déterministe k-morphique $\xi_0 = \sum_{i=1}^k n_{x_i}^0 \delta_{x_i}$*, alors, pour tout*

$t \geq 0$, $\xi_t = \sum_{i=1}^k n_{x_i}(t)\delta_{x_i}$, avec

$$\dot{n}_{x_i} = n_{x_i}\left(b(x_i) - d(x_i) - \sum_{j=1}^k \alpha(x_i,x_j)n_{x_j}\right), \quad \forall i \in \{1,\dots,k\}. \tag{7}$$

Ces résultats sont en fait des cas particuliers des résultats du chapitre 11 de [28].

Fournier et Méléard [33] démontrent ce théorème en prouvant la tension des loi des processus ν^K grâce aux critères d'Aldous [1] et de Rebolledo (cf. [44]), puis en identifiant la loi limite grâce au problème de martingale associé.

1.3 Simulations numériques

Nous avons réalisé des simulations numériques du processus ν^K pour deux modèles simples tirés de la littérature biologique. La simulation se fonde sur la méthode d'acceptation-rejet de la construction (2). Le source C du programme se trouve en annexe B.

Exemple 1 Il s'agit d'un modèle tiré de Kisdi [49], pour lequel les paramètres sont

$$\mathcal{X} = [0,4], \quad d(x) \equiv 0, \quad \mu(x) \equiv \mu,$$
$$b(x) = 4 - x, \quad \alpha(x,y) = 2\left(1 - \frac{1}{1 + 1.2\exp(-4(x-y))}\right),$$

et $m(x,dh)$ est une loi gaussienne de variance σ^2 conditionnée à ce que le trait mutant reste dans $\mathcal{X} = [0,4]$. La fonction $\alpha(x,y)$ ne dépend que de $x - y$ et tend vers 0 quand $x - y \to +\infty$ et vers 2 quand $x - y \to -\infty$. Cette fonction modélise donc une compétition asymétrique en faveur des traits grands (x ressent peu de compétition d'un trait y plus petit, mais une forte compétition d'un trait y plus grand). Puisque le taux de naissance est favorable aux traits petits, on s'attend à ce que l'évolution sélectionne un trait optimal.

Exemple 2 Ce modèle est tiré de Dieckmann et Doebeli [16]. Les paramètres sont

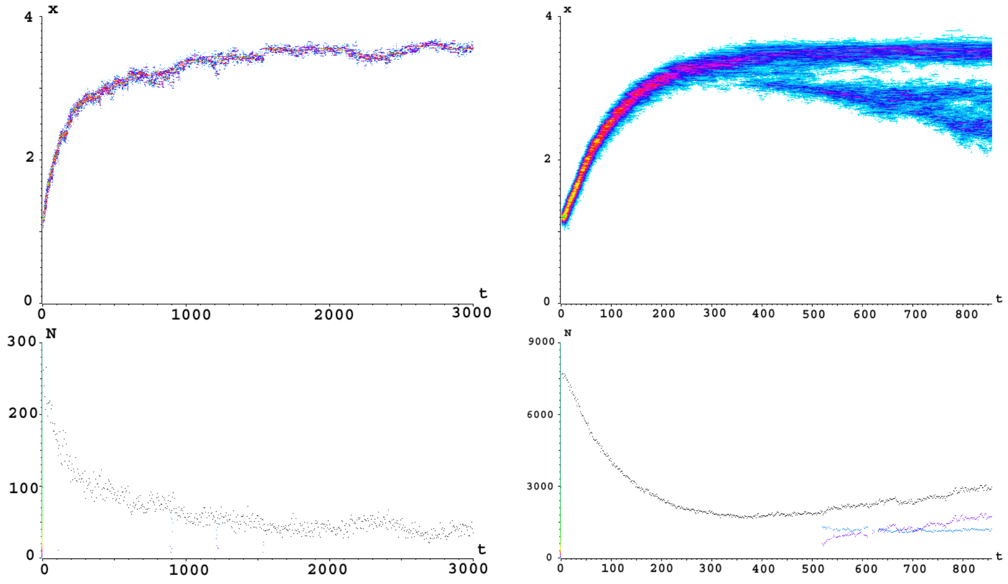
$$\mathcal{X} = [-2,2], \quad d(x) \equiv 0, \quad \mu(x) \equiv \mu,$$
$$b(x) = \exp(-x^2/2\sigma_k^2), \quad \alpha(x,y) = \exp(-(x-y)^2/2\sigma_c^2),$$

et $m(x,dh)$ est une loi gaussienne de variance σ^2 conditionnée à ce que le trait mutant reste dans $\mathcal{X} = [-2,2]$. Cette fois, la compétition est symétrique et les données du modèle se réduisent à deux paramètres seulement (outre μ, σ et K) : σ_k qui indique la taille de la zone où le taux de naissance est grand (autour de $x = 0$), et σ_c qui indique la portée de la compétition. On peut alors s'attendre à ce que la population se concentre au maximum de b, à moins que la compétition entre individus très semblables ne soit trop forte et "pousse" la population à se diversifier.

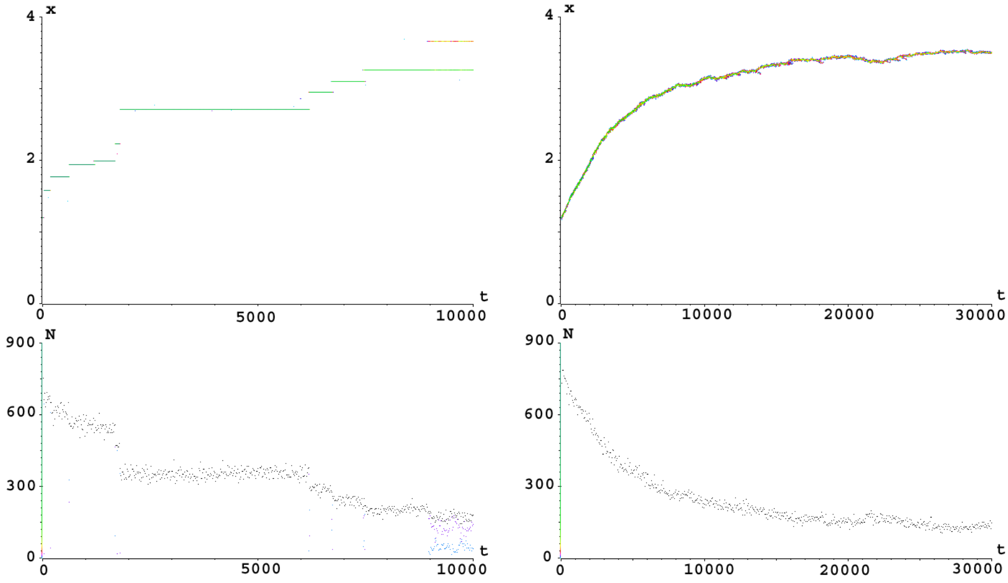
Nous avons réalisé des simulations de ces deux modèles, en partant d'une condition initiale monomorphique et en faisant varier les paramètres μ, K, σ, σ_k et σ_c, dont nous montrons un échantillon dans les figures suivantes. La partie supérieure de ces simulations utilise un code de couleurs pour représenter le nombre d'individus d'un trait x donné. La correspondance entre couleurs et effectifs est représentée sur l'axe des ordonnées de la partie inférieure des simulations, qui montre l'effectif total de la population ainsi que les effectifs de chaque "branche" (zone de l'espace des traits contenant une partie de la population, séparée d'une autre par une "large" zone vide), le cas échéant. Ces derniers sont représentés avec un code de couleurs afin de distinguer à quelle branche correspond l'effectif affiché : la branche avec la plus grande valeur du trait est représentée en violet, puis celle avec la valeur du trait immédiatement inférieure en bleu, puis rouge, orange et enfin vert (5 branches au maximum).

Les simulations montrent une très grande variété de comportements. Cependant, lorsque K et σ ne sont pas trop grands ($\sigma \lesssim 0.1$ et $K \lesssim 5000$), la population évolue selon un scenario général relativement stable (cf. figures 1 (a) (b) (d), et figures 2 (a) et (b)) : elle reste d'abord concentrée autour d'une valeur de trait qui se déplace progressivement vers un trait x^* proche de 3.5 dans l'exemple 1, et 0 dans l'exemple 2. Une fois ce point atteint, la population se stabilise, ou peut parfois se scinder en deux souspopulation concentrées autour de deux valeurs distinctes de traits qui coexistent et évoluent jusqu'à un nouvel équilibre, où la population peut se stabiliser ou de nouveau se scinder (ce phénomène est appelé *branchement évolutif*).

Afin de mieux comprendre ces phénomènes, les biologistes se sont intéressés au cas où les mutations sont rares (μ petit), sur lequel est fondée la théorie des dynamiques adaptatives. Les simulations montrent alors qu'à chaque instant, au plus deux ou trois traits sont présents dans la population. L'évolution procède d'abord par sauts (cf. figure 1 (c) et figure 2 (c)), avec apparitions puis invasions successives de traits mutants, qui remplacent le trait leur ayant donné naissance (appelé *trait résident*). Une fois la population

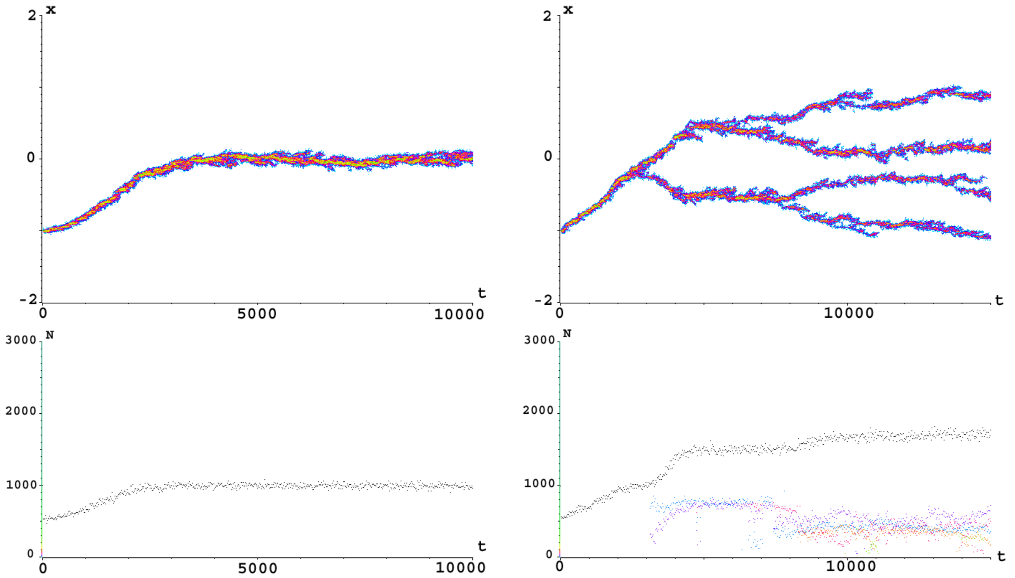


(a) $\mu = 0.1$, $K = 100$, $\sigma = 0.03$. (b) $\mu = 0.1$, $K = 3000$, $\sigma = 0.03$.



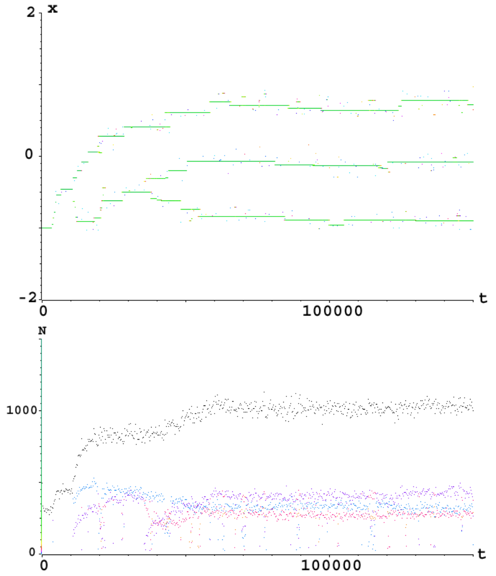
(c) $\mu = 0.00003$, $K = 300$, $\sigma = 0.2$. (d) $\mu = 1$, $K = 300$, $\sigma = 0.001$.

FIG. 1 – *Exemple 1. Simulations*



(a) $\mu = 0.05$, $K = 1000$, $\sigma = 0.01$, $\sigma_c = 0.8$, $\sigma_k = 0.9$.

(b) $\mu = 0.05$, $K = 1000$, $\sigma = 0.01$, $\sigma_c = 0.5$, $\sigma_k = 0.9$.



(c) $\mu = 0.0001$, $K = 500$, $\sigma = 0.08$, $\sigma_c = 0.4$, $\sigma_k = 1$.

FIG. 2 – *Exemple 2. Simulations*

arrivée au voisinage de x^*, il peut y avoir coexistence de deux traits ou plus (le trait ayant donné naissance au mutant ne disparaît pas).

Cette thèse porte principalement sur l'étude des conséquences mathématiques de cette hypothèse de mutations rares concernant la phase de convergence vers le point x^* (*avant* un éventuel branchement).

L'approche des dynamiques adaptatives fournit aussi, comme nous le verrons brièvement dans le chapitre I section 3.1, une explication biologique au phénomène de branchement évolutif et un critère y conduisant (cf. page 59). L'accord de ce critère avec les simulations numériques n'est cependant pas parfait. En effet, certaines simulations de l'exemple 1 (qui satisfait le critère de branchement) ne présentent pas le phénomène de branchement évolutif (cf. figures 1 (a) et (d)). De plus, pour les paramètres de l'exemple 2, le critère de branchement correspond à la condition $\sigma_c < \sigma_k$, alors qu'une étude numérique suggèrerait plutôt un critère de la forme $\sigma_c \gtrsim \sigma_k/2$ (comparer les figures 2 (a) et (b)). L'étude mathématique du phénomène de branchement n'est encore que partiellement comprise, et peut constituer une direction intéressante de prolongement du travail de cette thèse.

2 Résumé des chapitres I à IV

Chaque chapitre est indépendant des autres et utilise ses propres notations.

Le premier chapitre de cette thèse présente l'approche heuristique, suivie par les biologistes, de la dynamique adaptative d'une population. Les deux chapitre suivants portent sur la justification mathématique de l'heuristique biologique présentée dans le premier chapitre, en appliquant une asymptotique originale au modèle particulaire présenté dans la section précédente. Nous obtenons ensuite dans le chapitre III une famille de modèles d'évolution par diffusion indexée par un paramètre $\varepsilon > 0$, avec dérive discontinue et diffusion dégénérée. Ces diffusions sont étudiées en détail dans le chapitre IV, où l'on prouve notamment un principe de grandes déviations quand $\varepsilon \to 0$, et où l'on applique ce résultat à l'étude du temps et du lieu de sortie d'un domaine attracteur par ces diffusions.

2.1 Chapitre I : la biologie des dynamiques adaptatives

Ce chapitre présente les prérequis, les idées et les principaux résultats biologiques sur le dynamiques adaptatives, afin de familiariser le lecteur mathématicien avec cette théorie récente. Cette présentation met en lumière, de manière originale et aussi rigoureuse que possible, l'enchaînement des idées biologiques, leurs limites et les difficultés qui leur sont associées. Nous avons en particulier détaillé les hypothèses biologiques nécessaires à cette étude, rarement explicitées dans la littérature biologique. Nos sources principales sont Metz *et al.* [61, 59], Dieckmann et Law [18] et Diekmann [19].

Un tel chapitre nous a paru nécessaire pour plusieurs raisons. En premier lieu, le travail de défrichage de la littérature biologique conduisant à cette présentation a constitué l'indispensable première étape de cette thèse. En outre, ce chapitre permet de comprendre le cheminement et les méthodes mathématiques des chapitres suivants, ainsi que l'intérêt biologique de notre travail. D'autre part, il fournit une justification relativement convaincante des modèles biologiques, et introduit un modèle nouveau, objet de l'étude du chapitre III. Enfin, cette présentation éclaircit les notions d'équilibres évolutifs et de points de branchement, évoquées au cours de l'étude numérique de la section 1.3, et permet d'envisager diverses perspectives à notre travail.

Les modèles obtenus dans le chapitre I décrivent l'évolution globale des traits présents dans la population, mais s'expriment en fonction d'un certain nombre de paramètres individuels. Nous allons les écrire ici en fonction des paramètres b, d, α, μ, m et K du processus $(\nu^K_t,t \geq 0)$ de la section précédente.

Modèle TSS monomorphique

Le premier modèle s'obtient sous les hypothèses :

Hypothèses biologiques
(HB1) *les mutations sont rares (μ petit) ;*
(HB2) *l'effectif de la population est grand (K grand) ;*

et sous une hypothèse supplémentaire interdisant la coexistence à long terme de deux traits distincts (cf. chapitre I section 2.4). L'évolution procède alors par sauts dans l'espace des traits $\mathcal{X}$ de manière markovienne. La population est en tout temps monomorphique, et des mutants envahissent et remplacent le trait résident à une échelle de temps infinitésimale par rapport à l'échelle de temps des mutations. Ce modèle est appelé *trait substitution sequence* (TSS) monomorphique (cf. [59] et [18]), et son générateur infinitésimal s'écrit, pour toute fonction φ mesurable bornée sur $\mathcal{X}$,

$$A\varphi(x) = \int_{\mathbb{R}^d} (\varphi(x+h) - \varphi(x))\mu(x)b(x)\bar{n}_x \frac{[f(x+h,x)]_+}{b(x+h)} m(x,dh), \qquad (8)$$

où $[\cdot]_+$ désigne la partie positive, où la *fonction de fitness*

$$f(y,x) = b(y) - d(y) - \alpha(y,x)\bar{n}_x \qquad (9)$$

mesure l'avantage sélectif du trait mutant y par rapport au trait résident x, et où

$$\bar{n}_x := \frac{b(x) - d(x)}{\alpha(x,x)}$$

est l'unique équilibre stable de la dynamique déterministe (5) d'une population monomorphique de trait x.

Nous ne détaillerons pas ici le raisonnement biologique conduisant à ce modèle. Retenons cependant que (HB1) induit une séparation entre l'échelle de temps des mutations et celle des naissances et des morts (ce point sera précisé dans (14)), et que (HB2) permet de prévoir l'issue de la compétition entre un trait résident et un trait mutant en supposant que leur dynamique est déterministe et régie par (6).

L'équation canonique des dynamiques adaptatives

En supposant de plus :

Hypothèse biologique
(HB3) *les mutations sont de petite amplitude,*

Dieckmann et Law [18] ont proposé une approximation du modèle TSS monomorphique, dans le cas où $\mathcal{X} \subset \mathbb{R}$ et où la mesure $m(x,\cdot)$ est symétrique pour tout $x \in \mathcal{X}$ (c'est-à-dire que $m(x,A) = m(x,A^0)$ pour tout $A \subset \mathbb{R}$ mesurable, où A^0 désigne le symétrique de A par rapport à 0), sous forme de l'EDO suivante :

$$\frac{dx}{dt} = \frac{\sigma^2(x)}{2}\mu(x)\bar{n}_x\partial_1 f(x,x), \tag{10}$$

où $\sigma^2(x)$ désigne la variance de la mesure de probabilité $m(x,\cdot)$.

Modèle TSS k-morphique

Si l'on adjoint à (HB1) et (HB2) une hypothèse imposant que la population reste k-morphique en tout temps (c'est-à-dire que la compétition entre $k+1$ traits distincts conduit à long terme à la disparition d'un seul de ces traits), on obtient un processus de saut à valeurs dans $\mathcal{X}^k$, que nous appellerons modèle TSS k-morphique. Il s'agit d'un modèle original en dynamiques adaptatives, généralisant le modèle TSS monomorphique. Son générateur infinitésimal est donné, pour toute fonction ϕ mesurable bornée sur $\mathcal{X}^k$, par

$$A\phi(\mathbf{x}) = \sum_{i=1}^{k} \int_{\mathbb{R}^d} (\phi(\mathbf{x} + (h)_i) - \phi(\mathbf{x}))[g_i(x_i + h; \mathbf{x})]_+ m(x_i,dh), \tag{11}$$

où $\mathbf{x} = (x_1, \dots, x_k) \in \mathcal{X}^k$, et où $(h)_i$ est le vecteur nul, sauf à la $i^{\text{ème}}$coordonnée qui vaut h, et où

$$g_i(y; x_1, \dots, x_k) = \mu(x_i) b(x_i) \bar{n}^i_{(x_1,\dots,x_k)} \frac{f(y; x_1, \dots, x_k)}{b(y)}, \tag{12}$$

avec $\bar{n}_{(x_1,\dots,x_k)} = (\bar{n}^1_{(x_1,\dots,x_k)}, \dots, \bar{n}^k_{(x_1,\dots,x_k)})$ l'unique équilibre stable (cf. chapitre I section 2.5) de (7), et avec

$$f(y; x_1, \dots, x_k) = b(y) - d(y) - \sum_{i=1}^{k} \alpha(y, x_i) \bar{n}^i_{(x_1,\dots,x_k)}.$$

2.2 Chapitre II : justification du processus de saut monomorphique des dynamiques adaptatives à partir du modèle individu-centré

Nous établissons dans ce chapitre la convergence du modèle individuel vers le processus TSS monomorphique ci-dessus, lorsque la probabilité de mutation tend vers 0 en même temps que $K \to +\infty$. Il s'agit d'un résultat de séparation d'échelles de temps entre les mutations, et les événements de naissance et de mort. Pour ce, introduisons un paramètre $u_K \in [0,1]$ tendant vers 0 quand K tend vers l'infini, et multiplions la probabilité de mutation $\mu(\cdot)$ par u_K. Le générateur du processus $(\nu^K_t, t \geq 0)$ de la section 1.2 s'écrit alors, pour toute fonction mesurable bornée $\phi : \mathcal{M}^K \to \mathbb{R}$,

$$\begin{aligned} L^K \phi(\nu) = & K \int_{\mathcal{X}} \left[\phi\left(\nu + \frac{\delta_x}{K}\right) - \phi(\nu) \right] (1 - u_K \mu(x)) b(x) \nu(dx) \\ & + K \int_{\mathcal{X}} \int_{\mathbb{R}^d} \left[\phi\left(\nu + \frac{\delta_{x+h}}{K}\right) - \phi(\nu) \right] u_K \mu(x) b(x) m(x, dh) \nu(dx) \\ & + K \int_{\mathcal{X}} \left[\phi\left(\nu - \frac{\delta_x}{K}\right) - \phi(\nu) \right] \left(d(x) + \int_{\mathcal{X}} \alpha(x, y) \nu(dy) \right) \nu(dx), \end{aligned} \tag{13}$$

et ν^K peut être défini par une équation analogue à (2).

Nous aurons besoin, en plus de l'hypothèse (H) introduite à la section 1.1, des hypothèses :

(H1) Pour tout $x \in \mathcal{X}$, $\mu(x) > 0$ et $b(x) - d(x) > 0$, et il existe $\underline{\alpha} > 0$ tel que

$$\alpha(\cdot, \cdot) \geq \underline{\alpha}.$$

(H2) Pour tout $x \in \mathcal{X}$ fixé, pour presque tout $y \in \mathcal{X}$ (au sens de la mesure de Lebesgue),

$$\text{soit} \quad (b(y)-d(y))\alpha(x,x)-(b(x)-d(x))\alpha(y,x)<0,$$
$$\text{soit} \quad \begin{cases} (b(y)-d(y))\alpha(x,x)-(b(x)-d(x))\alpha(y,x)>0, \\ (b(x)-d(x))\alpha(y,y)-(b(y)-d(y))\alpha(x,y)<0. \end{cases}$$

(H2) assure que, soit l'équilibre $(\bar{n}_x,0)$ du système (6) est stable, soit $(0,\bar{n}_y)$ l'est, et dans ce dernier cas, toute solution de (6) avec condition initiale dans $(\mathbb{R}_+^*)^2$ converge vers cet équilibre quand $t \to +\infty$ (cf. l'hypothèse (HB5') de la section 2.4 du chapitre I, et l'annexe A).

Le principal résultat de ce chapitre s'écrit :

Théorème 2 (II.1.1) *Supposons (H), (H1), (H2),*

$$\forall C>0, \quad \exp(-CK) \ll u_K \ll \frac{1}{K\log K} \tag{14}$$

et $\nu_0^K=(\gamma_K/K)\delta_x$ *pour un* $x \in \mathcal{X}$ *fixé et pour une suite* $(\gamma_K) \in \mathbb{N}^{\mathbb{N}^*}$ *telle que* $\gamma_K/K \to \gamma > 0$*. Alors, le processus* $(\nu^K_{t/Ku_K},t \geq 0)$ *défini ci-dessus converge au sens des lois fini-dimensionnelles pour la topologie de* $\mathcal{M}_F$ *induite par les applications* $\nu \mapsto \langle \nu,f \rangle$ *où* f *parcourt l'ensemble des fonctions mesurables bornées sur* $\mathcal{X}$*, vers le processus*

$$Y_t = \begin{cases} \gamma\delta_x & si\ t=0 \\ \bar{n}_{X_t}\delta_{X_t} & si\ t>0, \end{cases}$$

où le processus de Markov $(X_t,t \geq 0)$ *a pour générateur infinitésimal (8), et pour état initial* x*.*

Quelques remarques s'imposent sur ce résultat :

1. La renormalisation du temps en $1/Ku_K$ revient à se placer à l'échelle de temps des mutations. En effet, la taille de la population est proportionnelle à K, et chaque individu donne naissance à un mutant à un taux proportionnel à u_K. Ainsi, le taux global d'apparition d'un mutant est proportionnel à Ku_K. Le théorème 2 signifie qu'à l'échelle de temps des mutations, la population est en tout temps monomorphique, et que la phase de remplacement d'un trait résident par un trait mutant a lieu sur une échelle de temps infinitésimale par rapport aux mutations. C'est la condition (14) qui conduit à cette séparation d'échelles de temps, et qui fournit la formulation mathématique convenable de l'hypothèse (HB1).

2. Il est indispensable de prendre simultanément les deux limites pour obtenir le modèle TSS monomorphique. D'une part, effectuer d'abord la limite $K \to +\infty$ puis $\mu \to 0$ ne pourrait donner un processus stochastique, puisque la première limite donne un processus déterministe (cf. théorème 1). D'autre part, effectuer la limite $\mu \to 0$ sans faire tendre K vers l'infini ne donnerait rien. En effet, si l'on construit le processus ν à l'aide de la formule (2) en remplaçant $\mu(\cdot)$ par $u\mu(\cdot)$ afin de faire tendre u vers 0, un calcul élémentaire (cf. theorem 2.1 chapitre II) montre que, indépendamment de u, le nombre total d'individus dans la population est dominé stochastiquement par un processus $(Z_t, t \geq 0)^*$ de naissance et mort "logistique" de transitions de i à $i+1$ avec taux $2\bar{b}i$, et de i à $i-1$ avec taux $\underline{\alpha}i^2$. Or, à cause de l'hypothèse $\underline{\alpha} > 0$, on montre facilement (cf. remarque 2.1 chapitre II) que le processus Z s'éteint presque sûrement en temps fini. placer à l'échelle de temps des mutations (en t/u) conduirait le processus à l'extinction presque pour tout $t > 0$ lorsque $u \to 0$.
3. Il est impossible d'obtenir un résultat de convergence *fonctionnelle*, à cause de la discontinuité à droite en $t = 0$ du processus Y, qui n'est donc pas càdlàg. Plus généralement, la phase de remplacement d'un trait résident par un trait mutant conduit à une discontinuité du processus qu'il est difficile de contrôler.
4. Ce résultat est différent des résultats de séparation d'échelles de temps du chapitre 7 de Freidlin et Wentzell [34] ou du chapitre 3 de Skorohod *et al.* [71] (principe de moyennisation), d'une part à cause de l'absence d'hypothèse d'ergodicité (cf. [71]), d'autre part parce qu'il combine deux limites simultanément.

Idée de la preuve

Elle se fonde sur deux ingrédients : la convergence d'une population monomorphique vers sa densité d'équilibre et l'étude du phénomène d'invasion d'un mutant dans cette population.

Le premier de ces problèmes résulte des *grandes déviations* pour la convergence de ν_t^K vers $n(t)\delta_x$, où $n(t)$ est solution de (5), lorsque $\mu \equiv 0$ et ν_0^K est monomorphique (cf. remarque 1 page 7). Toute solution de l'équation logistique (5) de condition initiale strictement positive converge vers $\bar{n}_x$. Les grandes déviations pour de tels processus de naissance et de mort sont connues (cf. Dupuis et Ellis [24]), et permettent de montrer que le temps pendant lequel le processus stochastique reste dans un voisinage de sa limite (problème de sortie de domaine, cf. Freidlin et Wentzell [34]) est de l'ordre de $\exp(KC)$ pour une certaine constante $C > 0$.

Or, lorsque u_K est petit, le processus ν^K issu d'une condition initiale monomorphique est proche du même processus avec $\mu \equiv 0$, tant qu'aucune mutation n'a lieu. Ainsi, l'inégalité de gauche dans l'hypothèse (14) permet d'assurer qu'avec une grande probabilité, le processus ν^K issu de $(\gamma_K/K)\delta_x$ est proche de $\bar{n}_x\delta_x$ à l'instant de la première mutation. La preuve de cette assertion (lemme 3.1 page 83) utilise des résultats de *domination stochastique* entre $\langle\nu^K_t,\mathbf{1}\rangle$ et des processus de naissances et de morts de type logistique.

L'étude du second problème peut être résumée par la figure 3 ci-après. Nous distinguerons trois phases dans l'invasion d'un mutant y dans une population monomorphique de trait x, et la compétition qui s'ensuit :

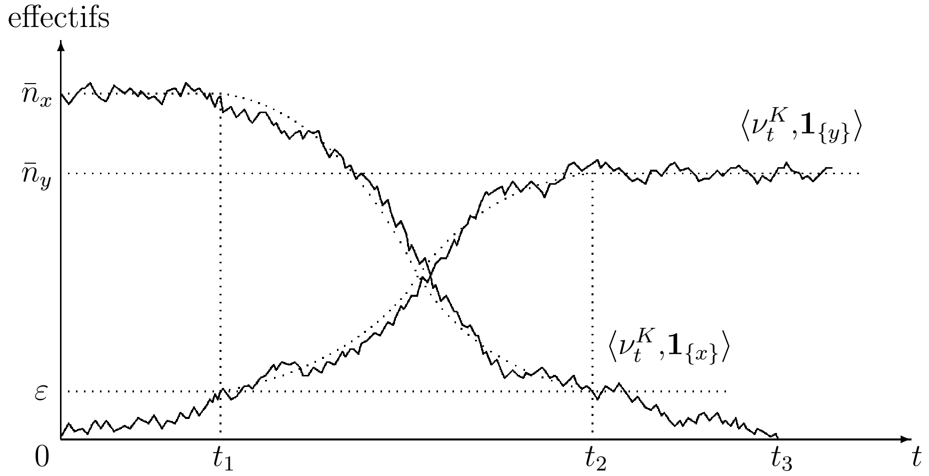


FIG. 3 – *Les trois phases de l'invasion et de la fixation d'un mutant y dans une population monomorphique de trait x.*

- La phase d'invasion (entre 0 et t_1) que nous interpréterons comme le passage de la densité de mutants $\langle\nu^K_t,\mathbf{1}_{\{y\}}\rangle$ de $1/K$ (un seul individu) à un niveau $\varepsilon > 0$ fixé ayant vocation à tendre vers 0. Le temps d'arrêt t_1 correspond donc au temps d'atteinte du premier entier supérieur ou égal à εK par le nombre de mutants $K\langle\nu^K_t,\mathbf{1}_{\{y\}}\rangle$. Tant que la densité de mutants est faible, la dynamique de la population résidente est proche de ce qu'elle serait en l'absence de mutants, donc $\langle\nu^K_t,\mathbf{1}_{\{x\}}\rangle$ reste proche de $\bar{n}_x$. Le taux de mort d'un individu de trait y est alors approximativement égal à la constante $d(y)+\alpha(y,x)\bar{n}_x$, et l'effectif des mutants suit approximativement un processus de branchement binaire, pour lequel la probabilité d'atteindre εK tend, quand $K \to +\infty$, vers sa probabilité de non extinction, qui vaut $[f(y,x)]_+/b(y)$, où f est définie dans (9).

- Une fois l'invasion réussie, d'après la remarque 1 page 7, la dynamique des deux traits x et y est, avec grande probabilité, proche de la solution du système d'EDO (6) avec condition initiale $(\bar{n}_x,\varepsilon)$, représentée en pointillés dans la figure 3 entre t_1 et t_2. Or, (H2) garantit que, lorsque $f(y,x) > 0$, cette solution converge vers $(0,\bar{n}_y)$ quand $t \to +\infty$. Ainsi, avec grande probabilité, la densité de mutants atteint tout voisinage de son équilibre $\bar{n}_y$ donné à l'avance, et la densité de résidents atteint le niveau ε petit, au bout d'un temps fini t_2.
- Enfin, une heuristique similaire à celle de la première phase montre que la population résidente suit approximativement un processus de branchement binaire de taux de naissance $b(x)$ et de taux de mort $d(x) + \alpha(x,y)\bar{n}_y$. (H2) garantit que ce processus de branchement est sous-critique, et donc que la population résidente s'éteint en un temps fini t_3 avec grande probabilité.

Cette heuristique est rendue rigoureuse dans le lemme 3.2 du chapitre II, en utilisant diverses *dominations stochastiques* des processus $\langle \nu_t^K,\mathbf{1}_{\{x\}}\rangle$ et $\langle \nu_t^K,\mathbf{1}_{\{y\}}\rangle$. La preuve nécessite en outre des résultats sur les *processus de branchement*, qui permettent de montrer que les temps t_1 et $t_3 - t_2$ sont de l'ordre de $\log K$, tandis que $t_2 - t_1$ est borné, avec grande probabilité. Ainsi, l'inégalité de droite dans l'hypothèse (14) garantit que l'invasion et la fixation d'un mutant ont lieu à une échelle de temps infinitésimale par rapport à l'échelle de temps des mutations.

La preuve du théorème se conclut en examinant, pour $t > 0$ fixé, la loi de l'état de la population ν_{t/Ku_K}^K en fonction du nombre aléatoire de mutations ayant eu lieu entre 0 et t/Ku_K. Pour chacune de ces mutations, on connaît la probabilité d'invasion et de fixation du mutant et on peut borner la durée de ces événements. Une délicate utilisation en cascade de la *propriété de Markov forte* permet de conclure.

2.3 Chapitre III : l'équation canonique et le modèle de diffusion

Certains des résultats de ce chapitre ont fait l'objet d'une publication écrite en collaboration avec Régis Ferrière et Gérard Ben Arous, parue dans la revue Selection [10].

L'objet de ce chapitre est d'obtenir rigoureusement une équation canonique des dynamiques adaptatives plus générale que (10) à partir du processus TSS k-morphique de générateur infinitésimal (11), puis d'en déduire un modèle de diffusion approchant ce processus.

Equation canonique k-morphique

Afin d'effectuer une asymptotique des petits sauts sur le processus TSS k-morphique, introduisons un paramètre $\varepsilon > 0$, et multiplions l'amplitude des sauts par ε, ce qui revient à remplacer la mesure de sauts $m(x,dh)$ par $m(x,dh/\varepsilon)$. Il est alors nécessaire d'accélérer le temps pour obtenir une limite non constante. La renormalisation convenable, en $1/\varepsilon^2$, n'est pas standard, et conduit (après changement de variable) à introduire le générateur infinitésimal suivant :

$$A^\varepsilon\phi(\mathbf{x}) = \frac{1}{\varepsilon^2}\sum_{i=1}^{k}\int_{\mathbb{R}^d}(\phi(\mathbf{x}+(\varepsilon h)_i)-\phi(\mathbf{x}))[g_i(x_i+\varepsilon h;\mathbf{x})]_+ m(x_i,dh), \tag{15}$$

où g_i est définie dans (12). Notons qu'un processus engendré par cet opérateur doit rester dans $\mathcal{X}$ en tout temps pour pouvoir être correctement défini. C'est pourquoi il est nécessaire d'ajouter une hypothèse sur $\mathcal{X}$ ou sur $m(x,dh)$ le garantissant. Nous avons choisi de supposer $\mathcal{X}$ convexe et de prendre $\varepsilon < 1$.

Nous obtenons alors le résultat suivant :

Théorème 3 (III.3.1) *Sous des hypothèses de régularité et de bornitude sur g_i et $m(x,dh)$, considérons, pour tout $\varepsilon \in]0,1[$, un processus de Markov X^ε de générateur infinitisimal (15) et de condition initiale $X_0^\varepsilon \in \mathcal{X}^k$ bornée dans $\mathbb{L}^1$ et convergeant en loi vers une variable aléatoire X_0 quand $\varepsilon \to 0$. Alors, $(X^\varepsilon)_{\varepsilon>0}$ converge en loi pour la topologie de Skorohod dans $\mathbb{D}(\mathbb{R}_+,\mathcal{X}^k)$ vers le processus $(X(t) = (x_1(t),\dots,x_k(t)),t \geq 0)$ d'état initial X_0 et de trajectoires $\mathcal{C}^1$ déterministes, solutions du système d'EDO*

$$\frac{dx_i}{dt} = \int_{\mathbb{R}^d} h[h\cdot\nabla_1 g_i(x_i;X)]_+ m(x_i,dh) \qquad \textit{pour } 1 \leq i \leq k, \tag{16}$$

où $\nabla_1 g_i$ désigne le gradient de $g_i(y;X)$ par rapport à la première variable y.

On retrouve l'équation (22) de la section 3.2 dans le cas particulier où $k = d = 1$ et $m(x,\cdot)$ est symétrique pour tout $x \in \mathcal{X}$. Grâce au théorème 2, on obtient donc une justification mathématique rigoureuse de l'équation canonique des dynamiques adaptatives à partir d'un modèle particulaire individuel.

Remarquons que, pour qu'une solution de (16) soit définie sur $\mathbb{R}_+$, il faut qu'elle ne sorte pas de $\mathcal{X}^k$. Nous avons choisi de résoudre ce problème de modélisation en supposant non viables les traits à la frontière[1] de $\mathcal{X}^n$. Ceci

---

[1] Dans le chapitre III, nous supposons en fait $\mathcal{X}$ ouvert, ce qui se justifie biologiquement par le fait que tout mutant né d'un trait x viable et suffisamment proche de x, est aussi viable.

s'exprime biologiquement par la relation $g_i(\mathbf{x}) = 0$ lorsque $x_i \in \partial\mathcal{X}$. On prolonge aussi l'application $x \mapsto m(x,dh)$ par δ_0 à la frontière de $\mathcal{X}$.

Le seconde difficulté tient à la mauvaise régularité du générateur A^ε et des coefficients du système (16), du fait de la présence de parties positives $[\cdot]_+$. Ceci conduit à des hypothèses de régularité pour $m(\cdot,dh)$ faisant intervenir des *métriques de Kantorovich* sur des espaces de mesures (cf. Rachev [65]). Ces hypothèses permettent en particulier de montrer l'existence globale en temps des solutions de (16).

La preuve en elle-même est fondée sur la propriété de *tension*, et la *propriété de martingale*. La tension ne pose pas de problème et découle du critère d'Aldous [1]. La principale difficulté pour établir la propriété de martingale consiste à montrer la *convergence des générateurs*, du fait de la présence des parties positives, ici encore. Un développement limité de l'expression (15) conduit, au premier ordre, à l'opérateur

$$A^0\phi(\mathbf{x}) = \sum_{i=1}^{k} \int_{\mathbb{R}^d} (h \cdot \nabla_i\phi(\mathbf{x}))[h \cdot \nabla_1 g_i(x_i;\mathbf{x})]_+ m(x_i,dh) = b(\mathbf{x}) \cdot \nabla\phi(\mathbf{x})$$

engendrant le processus X du théorème 3, où $\nabla_i\phi$ désigne le gradient de $\phi(x_1,\dots,x_k)$ par rapport à la $i^{\text{ème}}$ variable x_i, où $\cdot$ désigne le produit scalaire usuel, et où $b(\mathbf{x}) = (b^1(\mathbf{x}),\dots,b^k(\mathbf{x}))$ avec

$$b^i(\mathbf{x}) = \int_{\mathbb{R}^d} h[h \cdot \nabla_1 g_i(x_i;\mathbf{x})]_+ m(x_i,dh).$$

Approximation diffusion

Nous cherchons ensuite à dériver du processus de saut TSS k-morphique un processus de diffusion. Il y a plusieurs motivations biologiques à cette recherche.

L'un des inconvénients des modèles TSS tient au fait que l'ensemble des traits accessibles par mutation à partir d'une population donnée, est limité par le signe de la fonction de *fitness*. Par exemple, dans le cas monomorphique, pour f générique, puisque $f(x,x) = 0$, l'ensemble des traits mutants pouvant envahir une population de trait x est délimité par la courbe définie implicitement par $\{y : f(y,x) = 0\}$, et est donc, au voisinage de x, limité à un demi-espace. Il s'agit d'un comportement peu réaliste du point de vue biologique, parce que les populations naturelles sont soumises à des fluctuations aléatoires (dérive génétique, cf. Kimura [46, 48]), et que l'invasion de la population par un mutant légèrement désavantagé est toujours possible. De ce point de vue, (HB2) est une hypothèse excessive. Un processus de diffusion peut rendre compte de toutes les possibilités d'invasions.

Par ailleurs, une diffusion sur $\mathbb{R}^d$ avec $d \geq 2$ sort en général en temps fini avec une probabilité positive de tout domaine borné, alors que ce n'est pas forcément le cas pour le processus TSS. Il est en particulier facile de construire des fonctions de *fitness* dont la configuration de signe empêche tout éloignement d'une singularité évolutive. Ce comportement est irréaliste à longue échelle de temps, où l'intuition biologique suggère la possibilité de changements, graduels ou brutaux, de l'état global de la population (cf. Rand et Wilson [66]). Dans l'optique de l'étude de l'évolution en temps long, question biologique fondamentale, un processus de diffusion est donc biologiquement plus réaliste.

La méthode mathématique (cf. Ethier et Kurtz [28] chapitre 11) consiste, partant d'une famille de processus de générateurs L^ε ($\varepsilon > 0$) convergeant vers un opérateur L^0 lorsque $\varepsilon \to 0$, à effectuer un développement limité au premier ordre de L^ε autour de cette limite, et à ne garder que les termes d'ordres 0 et 1. On obtient une famille d'opérateurs de diffusion $(\tilde{L}^\varepsilon)_{\varepsilon>0}$. Ethier et Kurtz appellent *approximation diffusion* un processus généré par $\tilde{L}^\varepsilon$.

De telles diffusions ne correspondent cependant pas à un processus limite de la suite de processus considérée, comme c'est le cas pour les approximations diffusions du type du principe d'invariance de Donsker (cf. [28] chapitre 7). Ethier et Kurtz montrent cependant, dans le cas d'une famille de processus de naissances et morts, que l'approximation par les diffusions obtenues avec la méthode précédente, est du même ordre que celle fournie par le théorème de la limite centrale ([28] chapitre 11).

Un développement limité délicat de la formule (15) donne :

$$A^\varepsilon \phi(\mathbf{x}) = \tilde{A}^\varepsilon \phi(\mathbf{x}) + o(\varepsilon)$$

avec

$$\tilde{A}^\varepsilon \phi(\mathbf{x}) = (b(\mathbf{x}) + \varepsilon \tilde{b}(\mathbf{x})) \cdot \nabla \phi(\mathbf{x}) + \frac{\varepsilon}{2} \sum_{i=1}^{k} \sum_{1 \leq l,m \leq d} a^i_{lm}(\mathbf{x}) \frac{\partial_i^2 \phi}{\partial_i x_l \partial_i x_m}(\mathbf{x}), \qquad (17)$$

où $\partial_i \phi / \partial_i x_l$ désigne la dérivée partielle de $\phi(x_1, \ldots, x_n)$ ($x_i \in \mathcal{X} \subset \mathbb{R}^d$) par rapport à la $l^{\text{ème}}$ coordonnée de la $i^{\text{ème}}$ variable x_i, et avec les paramètres $\tilde{b}(\mathbf{x}) = (\tilde{b}^1(\mathbf{x}), \ldots, \tilde{b}^k(\mathbf{x}))$,

$$\tilde{b}^i(\mathbf{x}) = \begin{cases} \dfrac{1}{2} \displaystyle\int_{\{h \cdot \nabla_1 g_i(x_i;\mathbf{x}) > 0\}} h(h^* H_1 g_i(x_i;\mathbf{x}) h) m(x_i, dh) & \text{si } \nabla_1 g_i(x_i;\mathbf{x}) \neq 0, \\ \dfrac{1}{2} \displaystyle\int_{\mathbb{R}^d} h[h^* H_1 g_i(x_i;\mathbf{x}) h]_+ m(x_i, dh) & \text{si } \nabla_1 g_i(x_i;\mathbf{x}) = 0, \end{cases}$$

$$\text{et}\quad a^i_{lm}(\mathbf{x})=\int_{\mathbb{R}^d}h_lh_m[h\cdot\nabla_1 g_i(x_i;\mathbf{x})]_+m(x_i,dh),$$

où H_1g_i désigne la matrice hessienne de $g_i(y;\mathbf{x})$ par rapport à la première variable y, et où $h=(h_1,\dots,h_k)$. Nous noterons a^i la matrice $(a^i_{lm})_{1\leq l,m\leq d}$, et a la matrice diagonale par blocs, de blocs $a^i,\dots,a^k$.

Ces coefficients ont la régularité suivante :

Proposition 1 (III.4.2) *Sous des hypothèses de régularité et bornitude sur* g_i *et* $m(x,dh)$*, les fonctions* a*,* b *et* $\tilde{b}$ *sont bornées. De plus,* a *et* b *sont continues sur* $\mathcal{X}^k$*, et la fonction* $\tilde{b}^i$ *est continue sur* $\mathcal{X}^k$*, sauf éventuellement aux points de l'ensemble* $\Gamma_i=\{X=(x_1,\dots,x_k)\in\mathcal{X}^k:\nabla_1g_i(x_i;\mathbf{x})=0\}$*. Enfin, la matrice* $a(\mathbf{x})$ *est symétrique positive, non dégénérée si* $\mathbf{x}\in\mathcal{X}^k\setminus(\partial(\mathcal{X}^k)\cup\Gamma)$*, où* $\Gamma=\Gamma_1\cup\Gamma_k$*, et* $a^i(\mathbf{x})=0$ *si* $\nabla_1g_i(x_i;\mathbf{x})=0$*.*

Ici encore, ce résultat technique nécessite des hypothèses de régularité sur $m(x,dh)$, formulées en termes de métriques de Kantorovich.

Il est aisé de voir que $\tilde{b}$ est en général *discontinue* aux points de Γ. Par exemple, si $d=1$, $k=1$ et si $m(x,\cdot)$ est symétrique pour tout $x\in\mathcal{X}$,

$$\tilde{b}(x)=\frac{1}{2}\text{signe}[\partial_1g(x;x)]\partial^2_{11}g(x;x)\int_{\mathbb{R}}|h|^3m(x,dh),\tag{18}$$

avec la convention que signe$(0)=0$.

Puisque la matrice a n'est en général pas uniformément elliptique et que $\tilde{b}$ n'est pas continue, ce processus n'entre pas dans le cadre des résultats standards du calcul stochastique. En particulier, le problème de l'existence faible d'un processus ayant pour générateur infinitésimal $\tilde{A}^\varepsilon$ est difficile, et nous n'avons pu le résoudre que dans le cas où $k=1$ (et d quelconque), auquel est consacré le chapitre suivant.

2.4 Chapitre IV : étude du modèle de diffusion, grandes déviations et problème de sortie de domaine

Ce chapitre est consacré à l'étude des diffusions X^ε solution de

$$dX^\varepsilon_t=(b(X^\varepsilon_t)+\varepsilon\tilde{b}(X^\varepsilon_t))dt+\sqrt{\varepsilon}\sigma(X^\varepsilon_t)dW_t,\tag{19}$$

où $\sigma(x)\sigma(x)^*=a(x)$ dans le cas monomorphique ($k=1$) et sous l'hypothèse que $m(x,\cdot)$ est une mesure symétrique sur $\mathbb{R}^d$ pour tout $x\in\mathcal{X}$. Les paramètres du processus dépendent alors d'une seule fonction g définie sur $\mathcal{X}^2$ et des mesures $m(x,dh)$, et l'ensemble des singularités évolutives est $\Gamma:=\{x\in\mathcal{X}:\nabla_1g(x,x)=0\}$.

Notre objectif biologique est l'étude en temps long du processus X^ε, afin de comprendre et décrire les changements brutaux de l'état de la population aucours de l'évolution ("évolution par équilibres ponctués", cf. Rand et Wilson [66]). Ce problème se rapporte, mathématiquement, à l'étude du temps et du lieu de sortie d'un domaine attracteur par la diffusion X^ε, reposant sur les *grandes déviations* de X^ε quand $\varepsilon \to 0$, et sur la propriété de Markov forte (cf. Freidlin et Wentzell [34]). L'étude de ces deux questions constitue le cœur du chapitre.

En renforçant les hypothèses de la proposition 1, la régularité des coefficients peut être améliorée comme suit :

Proposition 2 (IV.2.1) *Sous des hypothèses de régularité et bornitude sur* g_i *et* $m(x,dh)$, *les fonctions* a, b *et* $\tilde{b}$ *sont bornées,* a *et* b, *prolongées par 0 hors de* $\mathcal{X}$, *sont globalement lipschitziennes sur* $\mathbb{R}^d$, *et la fonction* $\tilde{b}$ *prolongée par 0 hors de* $\mathcal{X}$ *est continue, sauf éventuellement aux points de* Γ. *De plus, la matrice* a *est symétrique, définie positive, si et seulement si* $x \in \mathcal{X} \setminus (\partial\mathcal{X} \cup \Gamma)$, *et nulle sinon. Si l'on définit* $\sigma(x)$ *comme la racine carrée (au sens des matrices symétriques positives) de* $a(x)$, *alors* σ *est bornée, globalement höldérienne d'exposant* $1/2$ *et localement lipschitzienne dans* $\mathcal{X} \setminus (\partial\mathcal{X} \cup \Gamma)$. *Enfin,* $a(x) = b(x) = \tilde{b}(x) = 0$ *lorsque* $x \in \partial\mathcal{X} \cup \Gamma$.

Existence faible, unicité en loi, propriété de Markov forte et problème d'atteinte des singularités évolutives

Puisque les points de $\partial\mathcal{X} \cup \Gamma$ sont des points où une solution de (19) peut s'arrêter, on peut construire explicitement une solution faible X^ε de (19) jusqu'au temps

$$\tau = \inf\{t \geq 0, X_t^\varepsilon \in \partial\mathcal{X} \cup \Gamma\}$$

par approximation par des diffusions non dégénérées dont l'existence faible est connue. En posant le processus constant après τ, on obtient une solution explicite de (19).

La mise en œuvre de cette méthode donne :

Théorème 4 (IV. 3.1) *Sous des hypothèses de régularité et de bornitude sur* g *et* $m(x,dh)$, *pour tout* $x \in \mathcal{X}$, *il existe une mesure de probabilité* $\mathbf{P}_x$ *sur* $\mathcal{C}(\mathbb{R}_+,\mathbb{R}^d)$ *solution du problème de martingale associé à (19) avec état initial* x, *sous laquelle le processus canonique est constant après* τ.

Nous étudierons dans ce chapitre les solutions particulières de (19) obtenues à partir de la solution $\mathbf{P}_x$ du problème de martingales ci-dessus par la construction canonique sur une extension de $\mathcal{C}(\mathbb{R}_+,\mathbb{R}^d)$ (cf. Karatzas et Shreve [45] page 315).

Lorsque $d = 1$, les coefficients de la diffusion prennent la forme

$$b(x) = \frac{1}{2}\partial_1 g(x;x) \int_{\mathbb{R}} |h|^2 m(x,dh), \quad a(x) = \frac{1}{2}|\partial_1 g(x;x)| \int_{\mathbb{R}} |h|^3 m(x,dh),$$

et (18) pour $\tilde{b}$. Au voisinage d'un point x_0 de Γ, $b(x)$ se comporte comme $C_1(x - x_0)$, $\tilde{b}(x)$ comme $C_2\text{signe}(x - x_0)$, et $a(x)$ comme $C_3|x - x_0|$, c'est-à-dire que $\sigma(x)$ se comporte comme $\sqrt{C_3|x - x_0|}$. Cerny et Engelbert [11] ont montré qu'il y a unicité en loi, pour des diffusions sur $\mathbb{R}$ très semblables, si et seulement si elles n'atteignent pas x_0 en temps fini. Il semble alors raisonnable d'étudier dans notre cas la relation entre l'unicité (et, ce qui est souvent lié, la propriété de Markov forte) et le temps d'arrêt $\tau = \inf\{t \geq 0 : X^\varepsilon_t \in \Gamma \cup \partial\mathcal{X}\}$.

Théorème 5 (IV.3.2) *Sous des hypothèses de régularité et de bornitude sur g et $m(x,dh)$, considérons $\mathbf{P}_x$ une solution du problème de martingales associé à (19) avec état initial x, et X^ε le processus canonique associé. Si $\mathbf{P}_x(\tau = \infty) = 1$, alors, il y a unicité en loi pour les solutions de (19) avec état initial x et X^ε vérifie la propriété de Markov forte.*

Ce résultat s'appuie sur des approximations du processus canonique par des diffusions pour lesquelles l'unicité en loi et la propriété de Markov forte sont connues.

La question de savoir sous quelles conditions $\mathbf{P}_x(\tau = \infty) = 1$ est vraie se pose alors. Nous obtenons une condition necessaire et suffisante pour cette égalité lorsque $d = 1$ (théorèmes 3.3 et 3.4 pages 138 et 142), puis nous donnons des conditions l'impliquant, et d'autres l'infirmant, pour $d \geq 2$ (théorème 3.5 page 142). En particulier, il existe pour tout $d \geq 2$ des conditions sous lesquelles il est possible d'atteindre les singularités évolutives isolées. Ceci tient au fait que σ est höldérienne, et non lipschitzienne, à leur voisinage.

Nous n'expliciterons pas ici ces résultats, mais nous bornerons à donner une idée de leur preuve. En dimension $d = 1$, nous utilisons les résultats d'élimination de la dérive d'Engelbert et Schmidt, ainsi que le critère d'explosion de Feller (cf. [45] pages 345–348). En dimension $d \geq 2$, nous cherchons à nous ramener à la dimension 1 en étudiant le processus $d(X^\varepsilon_t,\Gamma)$. Si, par exemple, $0 \in \Gamma$, la formule d'Itô permet d'exprimer $\|X^\varepsilon_t\|$ sous forme d'une partie à variation bornée et d'une martingale. Le théorème de Dubins-Schwartz exhibe un changement de temps transformant cette martingale en un mouvement brownien standard. En majorant et minorant la partie dérive de ce nouveau processus, nous sommes capables de l'encadrer entre deux diffusions dans $\mathbb{R}_+$, apparentées aux processus de Bessel. Il est aisé de savoir si elles atteignent ou non 0, en temps fini ou infini. La dernière partie de la preuve, délicate, consiste à examiner si ces propriétés sont toujours valides à l'échelle de temps initiale.

Principe de grandes déviations

Nous montrons ensuite un principe de grandes déviations essentiellement complet pour X^ε, sans avoir recours à l'hypothèse que $\tau = \infty$ p.s. La méthode utilisée s'inspire d'un article de Doss et Priouret [23], lui-même inspiré des travaux d'Azencott [4]. Il s'agit d'un principe de transfert du principe de grandes déviations pour la famille $(\sqrt{\varepsilon}W)_{\varepsilon>0}$, où W est un mouvement brownien d-dimensionnel (théorème de Shilder, cf. Dembo et Zeitouni [13] page 185) aux diffusions X^ε.

Théorème 6 (IV.4.1) *Sous des hypothèses de régularité et de bornitude sur g et $m(x,dh)$, et si les points de Γ sont isolés, considérons une solution X^ε de (19) avec état initial $x \in \mathcal{X}$, et constante après τ. Alors, pour tout $T > 0$, pour tout ouvert O et tout fermé F de $\mathcal{C}([0,T],\mathcal{X})$, tel que $\mathcal{C}^1([0,T],\mathcal{X} \setminus (\partial\mathcal{X} \cup \Gamma))$ est dense dans F,*

$$\liminf_{\varepsilon\to 0} \varepsilon \ln \mathbf{P}(X^\varepsilon \in O) \geq - \inf_{\psi\in O} I_{T,x}(\psi)$$
$$\limsup_{\varepsilon\to 0} \varepsilon \ln \mathbf{P}(X^\varepsilon \in F) \leq - \inf_{\psi\in F} I_{T,x}(\psi),$$

où

$$I_{T,x}(\psi) = \begin{cases} \frac{1}{2}\int_0^{t_\psi\wedge T} [\dot\psi(t) - b(\psi(t))]^* a^{-1}(\psi(t))[\dot\psi(t) - b(\psi(t))]dt \\ \qquad \text{si } \psi(0) = x,\ \psi \text{ est a.c. sur } [0,T] \text{ et constante} \\ \qquad\quad \text{après l'instant } t_\psi \text{ d'atteinte de } \partial\mathcal{X} \cup \Gamma, \\ +\infty \qquad \text{sinon.} \end{cases} \tag{20}$$

Rappelons brièvement la méthode utilisée dans [23] : si X^ε est solution de l'EDS

$$dX_t^\varepsilon = b^\varepsilon(X_t^\varepsilon)dt + \sqrt{\varepsilon}\sigma(X^\varepsilon)dW_t,$$

avec b^ε convergeant uniformément vers une fonction b, et avec b, b^ε et σ lipschitziennes, introduisons la fonction $S : \mathcal{C}^{ac}([0,T],\mathbb{R}^d) \to \mathcal{C}^{ac}([0,T],\mathbb{R}^d)$, où $\mathcal{C}^{ac}$ est l'espace des fonctions absolument continues, définie par

$$S(\phi)_t = x + \int_0^t b(S(\phi)_s)ds + \int_0^t \sigma(S(\phi)_s)\dot\phi_s ds, \tag{21}$$

et posons

$$I(\psi) = \inf\{J(\phi) : S(\phi) = \psi\}, \tag{22}$$

où J est la fonction de taux du principe de grandes déviations pour $(\sqrt{\varepsilon}W)_{\varepsilon>0}$. Les hypothèses de [23] permettent de montrer que I est une *bonne* fonction de taux.

L'image par S de $\sqrt{\varepsilon}W$ est proche de X^ε dans le sens suivant : pour toute fonction $\phi \in \mathcal{C}^{ac}([0,T],\mathbb{R}^d)$ telle que $\phi(0) = 0$ et $J(\phi) < \infty$, et pour tout $\eta,R > 0$, il existe $\delta > 0$ tel que

$$\limsup_{\varepsilon\to 0} \varepsilon \ln \mathbf{P}(\|X^\varepsilon - S(\phi)\| \geq \eta, \|\sqrt{\varepsilon}W - \phi\| \leq \delta) \leq -R, \tag{23}$$

où $\|\cdot\|$ désigne la norme uniforme sur $[0,T]$.

Grâce à ce résultat, il est facile d'obtenir la borne inférieure des grandes déviations en montrant que, pour toute fonction ϕ absolument continue telle que $\phi(0) = 0$ et pour tout $\eta > 0$, il existe $\delta > 0$ tel que $\mathbf{P}(\|X^\varepsilon - S(\phi)\| \leq \eta)$ est exponentiellement minorée par $\mathbf{P}(\|\sqrt{\varepsilon}W - \phi\| \leq \delta)$, lui-même exponentiellement minoré par $-J(\phi)$.

Concernant la borne supérieure des grandes déviations, si F est un fermé de $\mathcal{C}([0,T],\mathbb{R}^d)$ tel que $\inf_{\psi\in F} I(\psi) > 0$, et si $0 < u < \inf_{\psi\in F} I(\psi)$, alors $\Psi(u) := \{\psi : I(\psi) \leq u\}$ est compact et $\Psi(u) \cap F = \emptyset$, donc $d(F,\Psi(u)) = \eta > 0$. En appliquant (23) à toutes les fonctions $\phi \in \Phi(u) := \{\phi : J(\phi) \leq u\}$ avec $R > u$, on montre alors que $\mathbf{P}(X^\varepsilon \in F)$ est exponentiellement majoré par $\mathbf{P}(\sqrt{\varepsilon}W \notin U)$, où U est un voisinage ouvert de $\Phi(u)$, et cette probabilité est donc exponentiellement majorée par $-u$.

Dans notre cas, plusieurs difficultés vont nous obliger à modifier substantiellement ce programme. Premièrement, S n'est pas correctement définie par (21), puisque σ n'est pas lipschitzienne au voisinage des points de $\Gamma\cup\partial\mathcal{X}$. Il y a existence d'une solution, puisqu'il suffit de prendre $S(\phi)$ constant dès qu'il atteint $\Gamma \cup \partial\mathcal{X}$ (rappelons que $b = \tilde{b} = \sigma = 0$ sur cet ensemble), mais il n'y a pas nécessairement unicité. Nous choisissons de prendre pour $S(\phi)$ cette solution particulière et, de manière similaire, de considérer la solution particulière X^ε de (19) constante après τ. Cette définition de S conduit à la formule (20) pour $I_{T,x}$ défini par (22).

La seconde difficulté vient du fait que la fonction S ainsi obtenue n'est pas continue au sens de la norme uniforme et, par suite, la fonction de taux $I_{T,x}$ n'est pas semi-continue inférieurement (proposition 4.1 page 153). Ceci interdit d'utiliser la méthode ci-dessus pour établir la borne supérieure.

Enfin, du fait de la mauvaise régularité des coefficients, il nous est impossible de prouver (23) pour toutes les fonctions ϕ telles que $J(\phi) < +\infty$: nous devons supposer ϕ constante après le premier instant où $S(\phi)$ atteint $\Gamma \cup \partial\mathcal{X}$.

Malgré cela, il est possible d'obtenir la borne inférieure des grandes déviations par une méthode très semblable à celle présentée ci-dessus, en choisissant correctement la fonction ϕ à laquelle appliquer (23).

Concernant la borne supérieure, la méthode de [23] ne fonctionne pas. La solution que nous proposons est d'établir (23) en inversant les rôles de X^ε et $\sqrt{\varepsilon}W$ afin d'employer une méthode similaire à celle de la borne inférieure : pour tout $\delta,R > 0$, il existe $\eta > 0$ tel que

$$\limsup_{\varepsilon\to 0} \varepsilon \ln \mathbf{P}(\|X^\varepsilon - S(\phi)\| \leq \eta, \|\sqrt{\varepsilon}W - \phi\| \geq \delta) \leq -R. \tag{24}$$

Si F est compact, il suffit alors de majorer $\mathbf{P}(\|X^\varepsilon - \psi\| \leq \eta)$ pour toute fonction $\psi \in F$ afin de conclure, et cette majoration peut être faite en comparant cette probabilité avec $\mathbf{P}(\|\sqrt{\varepsilon}W - \phi\| \leq \delta)$ si $\psi = S(\phi)$ grâce à l'inégalité précédente.

Cependant, toutes les fonctions $\psi \in F$ ne peuvent pas s'écrire sous la forme $\psi = S(\phi)$ et, de surcroît, nous ne sommes parvenus à montrer (24) que pour les fonctions ϕ telles que $J(\phi) < \infty$ et $S(\phi)$ n'atteigne pas Γ sur $[0,T]$. La première de ces difficultés nous oblige à faire l'hypothèse que $\mathcal{C}^1([0,T],\mathcal{X} \setminus (\partial\mathcal{X} \cup \Gamma))$ est dense dans $F \subset \mathcal{C}([0,T],\mathcal{X})$, ce qui n'est pas une hypothèse très restrictive si $d \geq 2$. La deuxième de ces difficultés conduit à une preuve de la borne supérieure assez technique, nécessitant en particulier de restreindre l'intervalle de temps sur lequel appliquer (24), pour chaque fonction $\psi \in F$ atteignant Γ.

La borne supérieure se généralise à tout fermé F satisfaisant les hypothèses du théorème 6 en utilisant des bornes de tension exponentielle (lemme 4.4 page 157) et en agrandissant de manière technique les compacts intervenant dans le calcul afin qu'ils satisfassent l'hypothèse de densité des fonctions $\mathcal{C}^1$.

La preuve des inégalités (23) et (24) dans les cas mentionnés plus haut (lemme 4.2 page 156) constitue la plus grande partie de la démonstration du principe de grandes déviations. Elle requiert l'usage de la formule de Girsanov, de majorations exponentielles d'intégrales stochastiques (lemme 4.7 page 167), et de discrétisations du processus X^ε.

Problème de sortie de domaine

Le théorème 6 a pour conséquence la convergence en probabilité de la diffusion X^ε issue de x vers la solution de

$$\dot{\phi} = b(\phi) \tag{25}$$

de même état initial. Si l'on considère une domaine $G \subset \mathcal{X}$ attracteur pour cette EDO, c'est-à-dire tel qu'aucune solution de (25) ne puisse en sortir, combien de temps X^ε met-il pour sortir de G ? En quel point de ∂G cette sortie a-t-elle lieu ? Le théorème suivant répond partiellement à ces questions.

Théorème 7 (IV.5.1) *Sous des hypothèses de régularité et de bornitude sur* g *et* $m(x,dh)$*, considérons un sous-ensemble* G *ouvert borné de* $\mathcal{X}$ *tel que* $G \cap \Gamma = \{0\}$ *et tel que toute solution de (25) avec état initial dans* $\overline{G}$ *ne sorte pas de* $\overline{G}$ *et converge vers 0 quand* $t \to +\infty$*. Considérons pour tout* $x \in \mathcal{X}$ *une solution* $\mathbf{P}_x$ *au problème de martingale associé à (19) avec état initial* x*, et* X^ε *le processus canonique associé (constant après* τ*). Supposons que, pour tout* $x \in G \setminus \{0\}$*,* $\mathbf{P}_x(\tau = \infty) = 1$ *et* $\mathbf{P}_x(\lim_{t\to+\infty} X^\varepsilon_t = 0) = 0$*. Définissons pour tout* $y,z \in \mathcal{X}$

$$V(y,z) = \inf_{t>0} \inf_{\{\psi\in\mathcal{C}([0,t],\mathcal{X}):\psi(0)=y,\psi(t)=z\}} I_{t,y}(\psi),$$

et

$$\bar{V}(0,z) := \lim_{\rho\to 0} \inf_{\|y\|=\rho} V(y,z), \tag{26}$$

et supposons que $\bar{V} := \inf_{z\in\partial G} \bar{V}(0,z) < \infty$*. Définissons enfin*

$$\tau^\varepsilon = \inf\{t > 0 : X^\varepsilon_t \in \partial G\}.$$

Alors, pour tout $x \in G \setminus \{0\}$ *et tout* $\delta > 0$*,*

$$\lim_{\varepsilon\to 0} \mathbf{P}_x(\tau^\varepsilon > e^{(\bar{V}-\delta)/\varepsilon}) = 1. \tag{27}$$

De plus, pour tout sous-ensemble fermé N *de* ∂G *tel que* $\inf_{z\in N} \bar{V}(0,z) > \bar{V}$*, et pour tout* $x \in G \setminus \{0\}$*,*

$$\lim_{\varepsilon\to 0} \mathbf{P}_x(X^\varepsilon_{\tau^\varepsilon} \in N) = 0. \tag{28}$$

En particulier, s'il existe $z^* \in \partial G$ *tel que* $\bar{V}(0,z^*) < \bar{V}(0,z)$ *pour tout* $z \in \partial G \setminus \{z^*\}$*, alors*

$$\forall \delta > 0, \quad \forall x \in G \setminus \{0\}, \quad \lim_{\varepsilon\to 0} \mathbf{P}_x(\|X^\varepsilon_{\tau^\varepsilon} - z^*\| < \delta) = 1.$$

Ce résultat constitue une première étape dans l'étude de la “chaîne de visite” par X^ε des attracteurs de (25) (cf. Freidlin et Wentzell [34] chapitre 6).

Quelques commentaires sur ce théorème s'imposent :

- Premièrement, l'hypothèse $\mathbf{P}_x(\tau = \infty) = 1$ permet d'utiliser la propriété de Markov forte (théorème 5), et on montre que l'hypothèse $\mathbf{P}_x(\lim_{t\to+\infty} X^\varepsilon_t = 0) = 0$ garantit que $\tau^\varepsilon < +\infty$ presque sûrement.

- Deuxièmement, la forme classique du potentiel de Freidlin et Wentzell, $V(0,z)$, est inadaptée à notre cas. En effet, puisque X^ε reste constant et égal à 0 dès qu'il atteint ce point, $V(0,z) = +\infty$ dès que $z \neq 0$. C'est pourquoi nous introduisons le potentiel $\bar{V}(0,z)$.
- Enfin, (27) ne fournit qu'une borne inférieure pour τ^ε, alors que le résultat classique donne un encadrement. La borne supérieure n'a pu être obtenue, du fait de la dégénérescence de σ en 0, qui peut piéger X^ε au voisinage de 0 pour des durées difficiles à majorer.

La méthode que nous utilisons est largement inspirée de la preuve du théorème 5.7.11, chapitre 5, section 7 du livre de Dembo et Zeitouni [13]. La plupart des lemmes impliqués dans cette méthode ne nécessitent que des modifications mineures, et la preuve de (27) est très semblable à la preuve de [13]. La preuve de (28), en revanche, a du être modifiée, à cause de l'absence de borne supérieure sur τ^ε.

3 Conclusions et perspectives

Du point de vue biologique, plusieurs prolongements du travail de cette thèse sont envisageables. Le premier consiste à considérer plusieurs espèces en interaction. L'ensemble des chapitres III et IV peut se généraliser aisément à ce cas. Les résultats du chapitre II peuvent aussi l'être, à condition que la dynamique déterministe (correspondant à la limite $K \to +\infty$) d'une population où chaque espèce est monomorphique, possède un unique équilibre globalement asymptotiquement stable (cf. annexe A), et que l'on puisse donner une condition analytique similaire à (H2) garantissant l'impossibilité de la coexistence de plusieurs traits.

La seconde direction naturelle pour prolonger ce travail consiste à considérer le cas d'espèces à reproduction sexuée. La reproduction sexuée est une source de variabilité génétique différente de la mutagénèse, mais aussi d'une homogénéisation des phénotypes. Quelques études empiriques du phénomène de branchement évolutif ont déjà été menées dans des populations sexuées (cf. Dieckmann et Doebeli [16]), mais l'étude de la dynamique adaptative d'une telle population reste un problème difficile.

On peut aussi rajouter des structurations à la population, comme par exemple une structuration spatiale. De multiples modèles spatialement structurés ont déjà été étudiés par diverses méthodes. Citons Durrett [25], Durrett et Levin [27], Renshaw [67], Metz et Gyllenberg [60] et Gyllenberg *et al.* [40], Bolker et Pacala [5], Law et Dieckmann [53] et Law *et al.* [54] et Fournier et Méléard [33]. Les méthodes des dynamiques adaptatives ont déjà été appliquées à certains modèles “à patchs” (cf. [60] et [40]). L'étude des dynamiques adaptatives dans le cas des autres modélisations spatiales est une question biologique fondamentale, mais il s'agit là aussi de problèmes difficiles.

Citons un dernier problème important en dynamiques adaptatives, probablement plus accessible : comment généraliser la classification des singularités évolutives lorsque l'espace des traits est multi-dimensionnel, en tenant compte des fortes corrélations possibles entre les différentes composantes du trait lors des mutations (penser à la taille et au poids des individus) ? Quelle est, notamment, leur influence sur le branchement (cf. Leimar [56]), et quels sont le nombre, le poids et la direction d'évolution des branches ?

Venons-en maintenant à une discussion mathématique des chapitres II à IV de cette thèse.

Le chapitre II introduit des techniques susceptibles d'être appliquées dans des cas plus généraux, où il est possible d'avoir coexistence de deux traits ou plus. Il faut alors remplacer (H2) par une hypothèse garantissant que, si k traits $x_1,\dots,x_k$ sont à un équilibre asymptotiquement stable $(\bar{n}_1,\dots,\bar{n}_k)$ de leur dynamique déterministe, pour presque tout mutant y né de l'un de ces traits, il existe un voisinage $\mathcal{V}$ de $(0,\bar{n}_1,\dots,\bar{n}_k)$ dans $\mathcal{X}^{k+1}$ tel que la dynamique déterministe des $k+1$ traits $y,x_1,\dots,x_k$ converge vers un unique équilibre asymptotiquement stable pour toute condition initiale dans $\mathcal{V}$. Trouver de telles conditions est un problème difficile et peu étudié. Nous obtiendrions alors une justification mathématique d'un modèle TSS “polymorphique” où le nombre de traits peut varier.

Dans l'espoir d'éliminer l'inconvénient biologique de l'hypothèse de grande population, on peut envisager de travailler en population finie (K constant). Nous envisageons deux solutions permettant de contourner l'extinction presque sûre de la population, interdisant de prendre l'asymptotique $\mu(\cdot)\to 0$:

- La plus simple et la moins réaliste de ces solutions consiste à partir d'un modèle individuel semblable à ν dans lequel l'extinction totale de la population est impossible. La notion d'équilibre d'un trait résident correspond alors à la mesure stationnaire d'une chaîne de Markov récurrente, et la notion d'invasion d'un mutant correspond à la survie du trait mutant plus longtemps que le trait résident. La probabilité d'invasion est en général impossible à calculer et, par suite, l'éventuel processus TSS obtenu lorsque $\mu(\cdot)\to 0$ ne peut être explicité. Cependant, il est envisageable de calculer un développement limité de cette probabilité quand le trait mutant tend vers le trait résident. Ceci permettrait, comme dans le chapitre III, de calculer la limite de ce processus TSS lorsque l'amplitude des mutations tend vers 0. On s'attend à trouver une diffusion qu'il serait intéressant de comparer avec l'approximation diffusion du chapitre III.

- La deuxième solution, plus réaliste mais nettement plus complexe, est d'étudier le modèle individu-centré de la section 1 conditionnellement à la non extinction. La notion d'équilibre d'un trait résident correspond alors à une mesure quasi-stationnaire (cf. Seneta et Vere-Jones [69]). On peut alors s'attendre, en faisant tendre $\mu(\cdot)$ vers 0, à obtenir un modèle de type TSS monomorphique, dans lequel le conditionnement aurait pour effet de favoriser l'apparition des meilleurs mutants lors des mutations et, lors de la phase de compétition, de sélectionner systématiquement le trait ayant la meilleure capacité de survie.

Comme le suggèrent les simulations numériques (figure 1 (d)), il existe une alternative à l'approche du chapitre II, consistant à fixer μ et faire tendre directement la taille des sauts vers 0. Le support de ν_t^K semble converger vers un singleton, semblant évoluer selon une dynamique déterministe. Pour la même raison que dans le chapitre II, il est nécessaire de combiner cette limite avec la limite $K \to +\infty$. Signalons que cette même limite a récemment été étudiée dans des modèles *déterministes* très simples semblables à (4) par Diekmann *et al.* [20].

Les méthodes mathématiques du chapitre III sont plus classiques, mais l'étude du modèle de diffusion soulève des problèmes intéressants. En effet, la question de l'existence faible de cette diffusion dans le cas où $k \geq 1$ ou $m(x,dh)$ n'est pas symétrique n'est pas résolue. Dans le cas monomorphique, si cette diffusion atteint avec une probabilité non nulle la singularité évolutive x^* en temps fini, avec $\tilde{b}_{x^*} > 0$ (cette constante est définie au chapitre IV, remarque 3.6), il est peut-être possible de montrer que la diffusion ne peut s'échapper de x^*, et qu'il n'y a donc pas existence lorsque $\tilde{b}(x^*) \neq 0$, ce qui peut être le cas lorsque $m(x^*,\cdot)$ n'est pas symétrique.

Un autre problème intéressant consiste à démontrer la convergence du processus TSS polymorphique vers l'équation canonique lorsque les mutations tendent vers 0, sans supposer fixe le nombre de traits présents à tout instant. Les travaux de Geritz *et al.* [36] et Geritz [35] vont dans ce sens (cf. aussi chapitre I, sections 2.4 et 2.5).

Enfin, l'étude du modèle de diffusion du chapitre IV a abouti, malgré les difficultés, à un résultat satisfaisant, même s'il reste encore quelques questions ouvertes. Est-il possible d'utiliser des méthodes de minoration de $d(X^\varepsilon,\Gamma)$ par des processus de type Bessel (cf. chapitre IV, théorème 3.5), pour majorer le temps passé par la diffusion au voisinage de Γ ? Ces majorations permettent-elles d'obtenir la borne supérieure de τ^ε dans le théorème 7 de sortie de domaine ?

Chapter I

An introduction to adaptive dynamics

Abstract: This chapter presents the main results of the theory of adaptive dynamics. We detail the biological ideas and assumptions needed in this framework, and we introduce, as rigourously as possible, several "trait substitution sequence" models (in the polymorphic, monomorphic and k-morphic cases), and the canonical equation of adaptive dynamics. We also present the classification of evolutionary singularities in dimension 1.

We introduce hereafter the main results of the theory of adaptive dynamics, so that the reader becomes more familiar with this recent theory and its biological scope. This presentation may seem unsatisfactory for the mathematician reader. The aim of chapters II and III is to make this heuristics mathematically precise. We will try, in this biological presentation, to be as rigorous as possible, and to discuss the limits of the biological ideas and methods. In particular, we will precisely state the biological hypotheses underlying this theory, that are rarely explicitely described in the literature. This will of preciseness leads to a presentation quite different from the usual biological references, among which we mention Metz *et al.* [61, 59], Dieckmann and Law [18] and Diekmann [19].

This chapter is divided in four sections. In the first one, we place adaptive dynamics back in the context of the general phenomenon of natural selection. In the second one, we get interested in the main consequences of the biological assumptions of *rare mutations* and *large population*, in particular concerning the *invasion* of a population by a mutant. We recall the classical notion of fitness and we describe three models of evolution by jumps. The first one is similar to the general *trait substitution sequence* (TSS) model of [59], even if trying to be rigorous involves some modifications of this model. The second one is the classical monomorphic TSS model of [18]. The third model is an original one, the so-called k-morphic TSS model, which generalizes the monomorphic TSS model. In the third section, we examine the consequences of the additional biological assumption of *small mutations*, allowing to classify the evolutionary singularities in dimension 1, and to obtain an ordinary differential equation (ODE) approaching the evolutionary dynamics of the population. This chapter ends with a short conclusion.

The main biological notions used in this chapter are defined in a glossary at the end of this document.

1 Biological approaches of natural selection

1.1 The mechanisms of evolution

The principle of natural selection is known since Charles Darwin: among several individuals of a species, the ones that are better adapted to their natural environment transmit their characteristics to a larger number of descendants than the ones that are less adapted. This implies, on large time scale, the disappearance of the disadvantaged individual characteristics and the persistence of the adapted characteristics.

Since then, our understanding of this phenomenon has progressed mainly in the details of the molecular and physiological mechanisms. DNA was identified as the chemical support of heredity. The genetic code, translating genes into proteins, has been deciphered, and we know that proteins are responsible for almost every chemical reactions in a living organism (syntheses, enzymes, regulations, cycles, transport, etc.). But we still barely understand the other main mecanisms: how the *genotype* of an individual translates into its *phenotype* (*i.e.* the set of its global characteristics, such as reproductive efficiency, rate of food intake, height, etc.) which determines its adaptation to its environment? how to quantify this adaptation from the individual phenotypes? how do the differences between individual lead to the selection of the best adapted ones, in large time scales? Out of these three problems, it is probably the last one for which mathematics can be the most useful.

Evolution, as understood nowadays, can be represented as in figure I.1 below.

Each individual is characterized by a *genotype* that expresses into an individual *phenotype* during the developpement and through the metabolism of this individual. The set of phenotypes in a given population enters a competition process for resources (energy, space, time, information) or with other species (predators, preys, parasites, etc.), which affects each individual's sur-

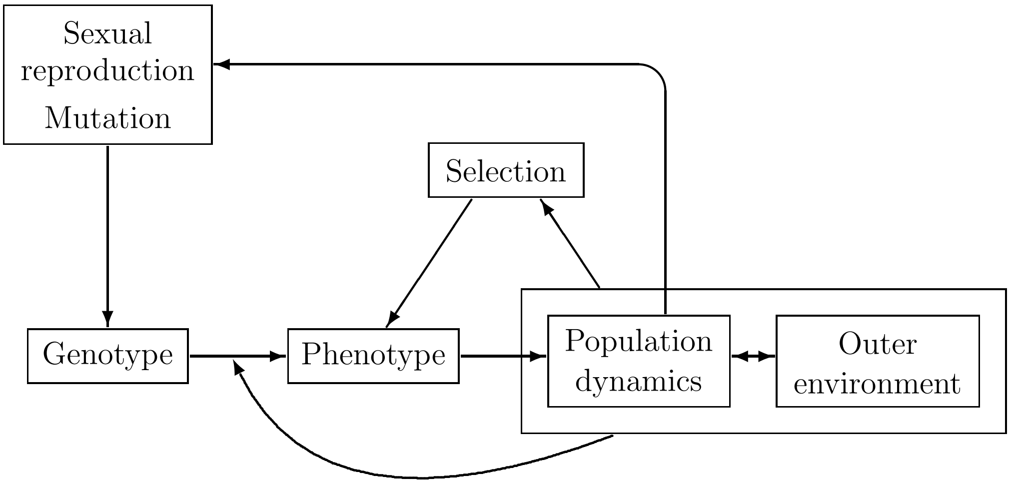


Fig. I.1 – *Diagram of evolution.*

vival and reproduction ability and thus induces a *population dynamics*. From this competition follows a *selection* process on phenotypes (and not on genotypes), which is a consequence of the differences between each individual's reproductive efficiency. Let us insist on the fact that the dynamics of the population depends on the outer environment (through natural resources or other species), which itself depends of the population we consider which exploits these resources or compete with the other species. In particular, the adaptation of a phenotype depends on its environment, but also, though this environment, on the composition of the rest of the population of the same species. The population dynamics has also an effect on the set of *genotypes* in the population through *sexual reproduction* (meiose, recombination) and *mutations* in the genome of a newborn individual. Finally, the environment of an individual (and thus the population dynamics) may also have an effect on the translation of its genotype to its phenotype, during its development, or as a consequence of the amount of energetic resources available, for example.

In the diagram above, most of the mecanisms are highly involved and very hard to model :

- the random map associating the genotype of a new individual to the ones of its parents, in the case of sexual reproduction (very hard to describe because of recombination) ;
- the map associating the phenotype of an individual to its genotype ;
- the application mapping the set of phenotypes in a population to the macroscopic characteristics of the population and their dynamics ;
- the selection phenomenon resulting, on a large time scale, from the population dynamics ;

– the influence of the environment on all these mecanisms.

In view of this, simplifications and approximations are needed in order to study evolution. A first approach, the so-called “quantitative genetics”, insists on the mecanisms of sexual reproduction. The size of the population is often supposed constant (as in the Wright-Fisher model) and the selective advantage, usually called *fitness*, of an individual is given *a priori*, and depends directly on its genotype. This allows to ignore the effects of the environment and not to consider explicitely phenotypes in the models. In this approach, the *fitness* describes a **fixed** *adaptive landscape* on which the population tries to reach a maximum (“*hill climbing*”, cf. Wright [72]). We refer to Durrett [26] and Ethier and Kurtz [29].

1.2 The approach of “adaptive dynamics”

The recent alternative approach of the theory of adaptive dynamics consists in simplifying the reproductive scheme by considering an *asexual* population (clonal reproduction, as for bacteria and some plants) and in paying more attention to the *ecology* of the system. The phenotypic variability then only comes from mutations. This allows to skip the notion of genotype, considering that the mutations directly act at the level of phenotypes. Figure I.2 sums up the main outlines of the modelling of evolution in the framework of adaptive dynamics.

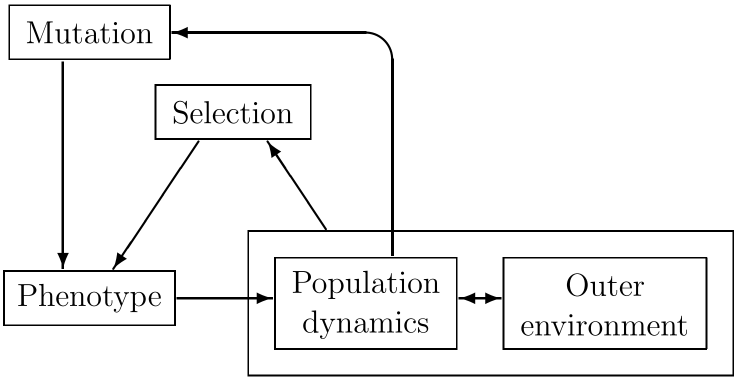


Fig. I.2 – *Modelling principle of adaptive dynamics.*

The main difference with the approach of quantitative genetics is that we now take into account the dependence between the population dynamics and the environment. The adaptive landscape on which evolution takes place now depends on the composition of the population. It is no longer **fixed** but evolves in the same time that the population. Moreover, an approach based on an explicit description of the ecology of the system allows, as we will see further, to define a quite precise notion of *fitness*. Finally, in the models of adaptive dynamics, the population size can fluctuate, what Wright-Fisher model and its variants did not allow.

Biological assumptions

In the framework of adaptive dynamics, biologists usually make the three following simplifying assumptions :

Biological hypotheses
(HB1) *Mutations are rare.*
(HB2) *Population size is large.*
(HB3) *Mutations have a small phenotypic effect.*

We recall that the mutations are considered as acting directly on the phenotype, *i.e.* any other mutation can be ignored. Since mutations having a phenotypic effect have to be non-synonymous[1], affecting a coding part of the genome (about 5 % of the human genome) and altering the chemical properties of the encoded protein (cf. Anfinsen [2]), it is biologically reasonable to consider that the mutations are rare.

The main purpose of (HB1) is to induce a time scale separation, between the ecological one (corresponding to the population dynamics) and the evolutionary one (corresponding to the mutations). Indeed, as mutations are rare, selection has enough time to discriminate between phenotypes in the population before a mutation produces a new phenotype. Therefore, the number of phenotypes that can coexist at the same time in the population is reduced as much as possible. The population can be *monomorphic*, *i.e.* made of individuals holding the same phenotype, or *dimorphic* (resp. n-*morphic*), *i.e.* composed of individuals holding exactly 2 (resp. n) different phenotypes. We will call *polymorphic* a population which is not monomorphic.

Assumption (HB2) allows to deal with densities of population instead of discrete population sizes and to suppose that the dynamics of these densities are deterministic. This last assumption is very classical in ecology, even if it may seem questionable to probabilists. We will see in chapter II what precise signification we can give to this approximation.

Finally, (HB3) is a classical assumption in evolution, and implies that the evolution proceeds gradually, by small steps.

---

[1] That is, involving a change in the amino-acid chain of the protein encoded in the corresponding gene.

2 Evolution by successive invasions of mutants: the adaptive dynamics jump processes

The approach that we present in this section refers to the first works in the theory of adaptives dynamics. The three fundamental papers are Hofbauer and Sigmund [41], Marrow *et al.* [57] and Metz *et al.* [61], but the most detailled presentation of the basis and main consequences of this approach can be found in Metz *et al.* [59]. From the mathematical point of view, no rigourous and systematic work had been yet engaged to support the theoretical framework sketched in these articles. The main purpose of chapters II and III of this thesis is to provide a mathematical basis for some adaptive dynamics models.

Let us consider one or several phenotypic traits characterizing individuals in a single asexual population (we restict ourselves to a single species in order to simplify notations). Let $\mathcal{X}$ be the space of possible trait values (*i.e.* a finite number of phenotypic characteristics). For example, $\mathcal{X}$ can be a subset of $\mathbb{R}^d$. Let us begin our study with the case of a monomorphic population with trait x (*i.e.* every individual of the population has the same trait value $x \in \mathcal{X}$).

Under (HB1), the only possible evolution for this monomorphic population is the appearence of a unique mutant trait $y \in \mathcal{X}$, followed by the death or the survival of traits x or y (this mutant trait cannot be helped by another mutant trait, since the mutations are rare). x is called the *resident trait*, as opposed to the *mutant trait* y. We will sometimes call "resident (resp. mutant) population x" the resident (resp. mutant) population with trait x. When the mutant trait y do not get immediately extinct, we will say that the *invasion* of the resident population x by the mutant trait y has occured. This invasion can be defined as the increase of the mutant population's size from a single individual to a "significative" size with respect to the resident population's size (this notion will be described in more details below). Finally, once the invasion of y has occured, the competition between x and y may drive both, one or none of them to extinction. We will say that *fixation* has occured when only y survives.

2.1 Invasion and fitness

The invasion of the resident population x by the mutant trait y depends on the dynamics of the mutant sub-population in the whole population. As long as the mutant sub-population is small, competition between individuals with trait y can be omitted, which corresponds classically in ecology to a

"Malthusian growth", *i.e.* an exponential increase (or decrease) of the mutant population size. Hence, the invasion of the mutant depends on an appropriate notion of exponential growth rate of the corresponding population. If this rate is negative, the mutant trait will disappear with a very high probability, and if it is positive, invasion is possible (but does not necessarily take place, since the random fluctuations of the mutant population cannot be neglected, as long as the size of this sub-population remains small). This growth rate can therefore be regarded as a quantitative measure of the adaptive value (or *fitness*) of the mutant trait y.

As can be seen in figure I.2, the dynamics of the mutant trait depends on its environment, depending itself on the rest of the population, which is made of individuals carrying the trait x. Therefore the growth rate of the trait y depends not only on y but also on x and on the number (or density) of residents when the mutant appears. Here again, (HB1) allows to simplify this description: since mutations are supposed to be rare, the resident population had enough time to reach an equilibrium (to be defined properly) of its dynamics before y has appeared. Suppose moreover that this equilibrium exists and is uniquely determined by x. Then, the growth rate of y depends on the resident population only through its trait x (its size being at equilibrium). Obviously, we have omitted several problems: what is the right definition of this equilibrium? how to ensure its existence and uniqueness? We will see in chapter II how to solve these questions in the case of a particular dynamics.

This leads to the following biological definition (Metz *et al.* [61]), that we will illustrate ans discuss in a few forthcoming examples:

Definition 1 *We call* fitness *of the trait* $y \in \mathcal{X}$ *with respect to the trait* $x \in \mathcal{X}$*, and we denote* $f(y,x)$*, the exponential growth rate of an initially small population with trait* y*, in a monomorphic resident population with trait* x*, at equilibrium.*

The biological originality of this notion of fitness lies in the fact that it does not depend only on the mutant trait value, as in most of the models in quantitative genetics, but also on the resident trait value.

In the first three examples, the resident population has an influence on the dynamics of the mutant trait y only through the value x of its trait.

Example 1 The dynamics of the mutant trait y, with density $n_y(t)$, is governed by the following logistic equation:

$$\dot{n}_y = n_y(r(y,x) - \alpha(y,x)n_y). \tag{1}$$

The quantity $r(y,x)$ is the natural growth rate of the mutant population and $\alpha(y,x)$ represents the competition inside the mutant population.

The mutant population being initially rare means that $n_y(0)$ is small. In this simple case, the fitness of y with respect to x is defined by

$$f(y,x) := \lim_{n_y(0)\to 0} \lim_{t\to 0} \frac{1}{t} \frac{n_y(t) - n_y(0)}{n_y(0)}. \tag{2}$$

We get $f(y,x) = r(y,x)$, which is coherent with the notion of invasion in this particular case: the mutant trait y can invade the resident population x if the steady state 0 of equation (1) is repulsive[2], that is if $f(y,x) = r(y,x) > 0$. If this steady state is asymptotically stable ($f(y,x) = r(y,x) \leq 0$), invasion is impossible.

Example 2 The dynamics of the mutant population y, with size $N_y(t)$, is governed by a continuous time binary branching process, with individual birth rate $b(y,x)$ (there is only one new individual per birth) and individual death rate $d(y,x)$. The mutant population being initially rare means here that $N_y(0) = 1$. In this case, the fitness of y with respect to x can be defined by

$$\begin{aligned} f(y,x) &:= \lim_{t\to 0} \frac{1}{t} \mathbf{E}\left[\frac{N_y(t) - N_y(0)}{N_y(0)} \;\middle|\; N_y(0) = 1 \right] \\ &= \lim_{t\to 0} \frac{\mathbf{E}[N_y(t) \mid N_y(0) = 1] - 1}{t}. \end{aligned} \tag{3}$$

Using the fact that $(e^{-(b(y,x)-d(y,x))t} N_y(t), t \geq 0)$ is a martingale (cf. Athreya and Ney [3] p. 111), we get $f(y,x) = b(y,x) - d(y,x)$ (which is the intuitive notion of growth rate of the population).

In this example, the invasion of the mutant trait can be interpreted as the survival of the mutant population. The probability of this event is given by (cf. [3] p. 109) :

$$\mathbf{P}(\forall t \geq 0, N_y(t) \geq 1) = \begin{cases} 1 - \dfrac{d(y,x)}{b(y,x)} = \dfrac{f(y,x)}{b(y,x)} & \text{if } b(y,x) > d(y,x) \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

Here again, we see that the possibility of invasion is determined by the sign of the fitness.

---

[2] The notions of repulsive, asymptotically stable, globally asymptotically stable and unstable steady states are properly defined in Annex A.

Example 3 The dynamics of the mutant population y, with size $N_y(t)$, is governed by a logistic birth and death Markov process, with birth rate (transition from i to $i+1$) $ib(y,x)$ for $i \geq 0$ and death rate (transition from i to $i-1$) $id(y,x) + i(i-1)\alpha(y,x)$ for $i \geq 1$. Such a process is natural from the biological point of view: in a sense, it is the stochastic analog of the logistic equation (1). It contains linear "natural" birth and death rates, corresponding to a branching process, and a death by competition rate which is proportional to the number of pairs of individuals, modeling the interactions in the population (cf. Lambert [51]).

If we apply formula (3) to this process, writing

$$\begin{gathered}
\mathbf{P}(N_y(t) = 2 \mid N_y(0) = 1) = b(y,x)t + o(t), \\
\mathbf{P}(N_y(t) = 0 \mid N_y(0) = 1) = d(y,x)t + o(t), \\
\mathbf{P}(N_y(t) = 1 \mid N_y(0) = 1) = 1 - (b(y,x) + d(y,x))t + o(t) \\
\text{and} \quad \mathbf{P}(N_y(t) > 2 \mid N_y(0) = 1) = o(t)
\end{gathered}$$

(the terms involving $\alpha(y,x)$ cancel due to the fact that $N_y(0) = 1$), we obtain the relation $f(y,x) = b(y,x) - d(y,x)$.

However, in this case, the notion of invasion and its interpretation in term of fitness is far from being clear, since, when $\alpha(y,x) > 0$, this birth and death process gets almosy surely extinct in finite time (cf. the introduction of this thesis, section 2.2).

Example 4 The three previous examples are biologically somewhat unrealistic because the dynamics of the mutant population is influenced by the resident population only *via* its trait x. In this situation, the notion of equilibrium state of the resident trait was useless to determine the fitness. It is biologically much more realistic to include a *density-dependance* between the populations with traits x and y. The most classical example is the logistic competition model:

$$\begin{cases} \dot{n}_x = n_x(r(x) - \alpha(x,x)n_x - \alpha(x,y)n_y) \\ \dot{n}_y = n_y(r(y) - \alpha(y,x)n_x - \alpha(y,y)n_y), \end{cases} \tag{5}$$

where n_x is the density of the resident population x, n_y is the density of the mutant population y, and where the "natural" growth rate $r(\cdot)$ and the competition coefficient $\alpha(\cdot,\cdot)$ are positive. In this case, the fitness, as defined in (2), is given by $f(y,x) = r(y) - \alpha(y,x)n_x(0)$. Therefore, the initial density of the resident population has to be known in order to determine $f(y,x)$.

The dynamics of the resident population alone follows

$$\dot{n}_x = n_x(r(x) - \alpha(x,x)n_x). \tag{6}$$

This ODE has two steady states: 0, which is repulsive (as $r(x) > 0$) and

$$\bar{n}_x := \frac{r(x)}{\alpha(x,x)},$$

which is globally asymptotically stable. Therefore, Definition 1 leads to the formula ($n_x(0) = \bar{n}_x$):

$$f(y,x) = r(y) - \frac{\alpha(y,x)}{\alpha(x,x)} r(x). \tag{7}$$

Now, the stability matrix of the steady state $(\bar{n}_x,0)$ is

$$\begin{pmatrix} -r(x) & -r(x)\alpha(x,y)/\alpha(x,x) \\ 0 & r(y) - r(x)\alpha(y,x)/\alpha(x,x) \end{pmatrix}.$$

Therefore, this steady state is locally asymptotically stable if $f(y,x) < 0$ and unstable if $f(y,x) > 0$. Here again, the relation between fitness and invasion is satisfied.

Example 5 The stochastic analog of the previous example is the birth, death and competition $\mathbb{N}^2$-valued Markov process $((N_x(t),N_y(t)),\ t \geq 0)$ with transitions

- from (i,j) to $(i+1,j)$ with rate $ib(x)$, for $i,j \geq 0$;
- from (i,j) to $(i,j+1)$ with rate $jb(y)$, for $i,j \geq 0$;
- from (i,j) to $(i-1,j)$ with rate $id(x) + i(i-1)\alpha(x,x) + ij\alpha(x,y)$, for $i \geq 1$ and $j \geq 0$;
- from (i,j) to $(i,j-1)$ with rate $jd(y) + ji\alpha(y,x) + j(j-1)\alpha(y,y)$, for $i \geq 0$ and $j \geq 1$.

The competition rates are proportional to the number of pairs of competitors, as in Example 3.

In this example, formula (3) gives $f(y,x) = b(y) - d(y) - \alpha(y,x)N_x(0)$, in the case where $N_x(0)$ is deterministic. Here again, the fitness depends on the initial resident population, but the notion of equilibrium for this population is not clear, for the same reasons as in Example 3. Therefore, the fitness is not well defined, and the notion of invasion is far from being clear in this case.

We could also imagine deterministic models for which the fitness is not well defined, by assuming a dynamics for the monomorphic resident population admitting several stable equilibria. Additional problems can appear when the resident dynamics is cyclic ot even chaotic (cf. De Feo and

Ferrière [12] and Dercole *et al.* [14]). The problem of defining an appropriate notion of fitness for these kinds of dynamics has been also discussed by Metz *et al.* [59].

Finally, let us mention that it is also possible to define the notion of fitness in the case of discrete-time dynamics, even if the computations may get much more involved (cf. Ferrière and Gatto [31] and section 8 of Diekmann [19]). In this thesis, we are only interested in continuous-time models.

The large population assumption

Thanks to assumption (HB2), we can get rid of some of the difficulties described above: as the population is supposed to be large, we can approximate the resident population's dynamics by a density-dependent ODE, or a system of density-dependent ODEs when this dynamics is coupled with environmental variables such as external resources or other species' density. For the fitness to be well-defined by Definition 1, we need the following additional assumption:

Biological hypothesis
(HB4) *Under (HB2), the monomorphic population dynamics is governed by a density-dependent ODE (or a system of ODEs), admitting* a unique *non null globally asymptotically stable steady state.*

In particular, this assumption is true for (6), and, more generally, for all the logistic-like ODEs of the form

$$\dot{n} = n(r - \alpha(n)),$$

where $r > 0$, $\alpha : \mathbb{R}_+ \to \mathbb{R}_+$ is increasing, $\lim_{u\to+\infty} \alpha(u) > r$ and $\alpha(0) = 0$. r is the "natural" growth rate of the population, and α represents the intra-specific competition. For this ODE, 0 is a repulsive steady state ($r > 0$), and the unique solution of $\alpha(u) = r$ is a globally asymptotically stable steady state.

We can now state the biological principle linking fitness and invasion:

Biological principle 1 *Under (HB4), the fitness is properly defined by Definition 1. Moreover, if $x \in \mathcal{X}$ is the trait of an equilibrium monomorphic population in which a mutant trait $y \in \mathcal{X}$ has just appeared, then, under (HB1), (HB2) and (HB4),*

- *if $f(y,x) < 0$, invasion by the mutant is impossible;*
- *if $f(y,x) > 0$, invasion by the mutant is possible (but not sure).*

The fitness, as defined above, satisfies the following fundamental biological property :

Fundamental property 1 *For any* $x \in \mathcal{X}$,

$$f(x,x) = 0.$$

Actually, since the resident population with trait x is supposed to be at equilibrium, the growth rate of a single individual with trait x (a mutant which is identical to the resident) has to be null. In particular, the fitness (7) of Example 4 satisfies this property.

Invasion probability

Metz *et al.* [59] and Dieckmann and Law [18] have proposed the following solution for computing the invasion probability of an advantaged mutant (*i.e.* with positive fitness). Under (HB2), invasion means that the mutant population size increases from a unique individual (*i.e.* an “infinitesimal” density) to a “significative” size with respect to the whole population (*i.e.* a positive density). During this phase, the competition between two mutant individuals can be neglected, as the density of the mutant population remains small. Its growth is therefore Malthusian and can be modeled by a binary branching process (cf. Example 2 page 40), which survival probability (4) is known. Thus, we obtain:

Biological principle 2 *Under (HB1) and (HB2), the invasion probability of a mutant trait* y *in a monomorphic resident population with trait* x *is*

$$\frac{[b(y,x) - d(y,x)]_+}{b(y,x)},$$

where $[a]_+$ *is the positive part of* a*, and where* $b(y,x)$ *and* $d(y,x)$ *are respectively the birth and death rates of a unique mutant individual with trait* y *in the equilibrium resident population with trait* x*.*

If we assume moreover (HB4), then the resident population is at its equilibrium and the fitness $f(y,x)$ *depends on the rates* $b(y,x)$ *and* $d(y,x)$ *through the relation (cf. Example 2)*

$$f(y,x) = b(y,x) - d(y,x).$$

A mathematical definition of the parameters $b(y,x)$ and $d(y,x)$ will be obtained for a particular model in chapter II.

However, there are several cases to which this principle, as the previous one, does not apply. For example, if the spatial component of the environment plays an important role, or if the resident population is not homogeneous in space, then a branching process is no longer a realistic model for the dynamics of the small mutant population. In some cases, it may also be unrealistic to consider that the birth and death rates do not depend on time (because of seasonal or day-night cycles, for example).

2.2 Fixation or coexistence?

Once the invasion of the mutant trait is achieved, the question of its fixation or of the coexistence of the two traits arises. Under (HB2), this is determined by the system of ODEs governing the dynamics of the two traits. The trajectory after invasion may converge towards an equilibrium where the density of the resident n_x is null (fixation), where the density of mutants n_y is null (*the resident strikes back*, cf. Mylius and Diekmann [62]), where the two densities are both positive (coexistence), or both null (*evolutionary suicide*, cf. Ferrière [30] and Dieckmann and Ferrière [17]), or may converge to a limiting cycle or a chaotic path (cf. De Feo and Ferrière [12]).

Here again, Example 4 is enlightening. The following proposition is a corollary of Theorem 1 of Annex A.

Proposition 1 *Assume that the resident density* n_x *and the mutant density* n_y *are governed by the logistic system (5) and that* $f(y,x) > 0$*. Then, there is fixation if* $f(x,y) < 0$ *and coexistence if* $f(x,y) > 0$*. More precisely, there exists a neighborhood* $\mathcal{V}$ *of the steady state* $(\bar{n}_x,0)$ *such that any solution of (5) with initial condition in* $\mathcal{V} \cap (\mathbb{R}_+^*)^2$ *converges to* $(0,\bar{n}_y)$ *when* $t \to +\infty$ *if* $f(x,y) < 0$*, and to the unique steady-state of (5) belonging to* $(\mathbb{R}_+^*)^2$ *if* $f(x,y) > 0$*.*

Metz *et al.* [59] have generalized this criterion by the following heuristics: according to the Biological Principle 1, the invasion of the population x by the mutant trait y requires that $f(y,x) > 0$ (we exclude the null case), which implies that the equilibrium $(\bar{n}_x,0)$ is unstable for the dynamics of the two traits x and y. But (HB4) implies that there is no other (locally) stable attractor in $\mathbb{R}_+ \times \{0\}$, so that $n_y(t)$ cannot converge to 0 as $t \to +\infty$. Now, if $f(x,y) > 0$, the equilibrium $(0,\bar{n}_y)$ is also unstable and the same arguments show that $n_x(t)$ cannot converge to 0 as $t \to +\infty$. Therefore, there is coexistence of x and y. Conversely, if $f(x,y) < 0$, the equilibrium $(0,\bar{n}_y)$ is locally stable, but we cannot deduce in general the convergence of $(n_x(t),n_y(t))$ towards this equilibrium for any initial condition next to $(\bar{n}_x,0)$.

Therefore, we have to introduce an additional assumption :

Biological hypothesis
(HB5) *Under (HB4), for any x, $y \in \mathcal{X}$, if $f(y,x) > 0$ and $f(x,y) < 0$, then the dynamics $(n_x(t),n_y(t))$ of the dimorphic population with traits x and y, after invasion of the resident equilibrium population x by the mutant y, converges as $t \to +\infty$ towards $(0,\bar{n}_y)$.*

As we will see in section 2.4, this assumption is reasonable under (HB3). The works of Geritz [35] and Geritz *et al.* [36] establish under which conditions it is satisfied for various models.

We can now state:

Biological principle 3 *Under (HB1), (HB2), (HB4) and (HB5), after the invasion of the resident trait $x \in \mathcal{X}$ by the mutant trait $y \in \mathcal{X}$, there is* fixation *of the mutant if $f(x,y) < 0$, and* coexistence *of the two traits if $f(x,y) > 0$.*

2.3 Adaptive dynamics jump process: general case

This section is inspired from [59], but its contents are mainly original.

We can now predict the fate (invasion, fixation, coexistence) of a mutant trait appearing in a monomorphic population. Under (HB1), the next mutant trait will appear only once invasion, fixation or coexistence of the first one has been decided, and its fate can be analyzed in the same way. The global evolution of the population in a large time scale results from the succession of these invasions. This evolution may be a gradual process (cf. Rand and Wilson [66]) when each mutant gets fixed in the population, or a diversification process if coexistence of many traits takes place.

To go further in our study, we have to generalize the results of the previous sections to the case when the initial population is polymorphic. Assumption (HB4) has to be modified as follows:

Biological hypothesis
(HB4') *Under (HB2), the dynamics of a k-morphic population with traits $x_1,\dots,x_k \in \mathcal{X}$ is governed by a system of density-dependent ODEs, with a* unique *non-null globally asymptotically stable steady state.*

Note that this is a very strong assumption and there is no general method known to verify it for particular dynamics. However, it is a very convenient tool to construct the biological model. It is possible to weaken this assumption and to describe a more general biological model by allowing the existence of several stable steady states (see section 3 of the introduction of this thesis).

Under (HB4'), we denote by

$$\bar{n}_{(x_1,\dots,x_k)} := (\bar{n}^1_{(x_1,\dots,x_k)},\dots,\bar{n}^k_{(x_1,\dots,x_k)}) \tag{8}$$

the unique stable equilibrium for the dynamics $(n_{x_1}(t),\dots,n_{x_k}(t))$ of the k-morphic population with traits $x_1,\dots,x_k$. By (HB4'),

Property 2 *For any $k \geq 2$, $x_1,\dots,x_k \in \mathcal{X}$ and $i \in \{1,\dots,k\}$ such that $\bar{n}^i_{(x_1,\dots,x_k)} = 0$,*

$$\bar{n}_{(x_1,\dots,x_{i-1},x_{i+1},\dots,x_k)} = [\bar{n}_{(x_1,\dots,x_k)}]_i,$$

where $[(u_1,\dots,u_k)]_i$ denotes the vector $(u_1,\dots,u_{i-1},u_{i+1},\dots,u_k)$.

The definition of the fitness generalizes as follows:

Definition 2 *For any $k \geq 1$, we call* fitness *of the trait $y \in \mathcal{X}$ with respect to the traits $x_1,\dots,x_k \in \mathcal{X}$ and we denote by $f(y;x_1,\dots,x_k)$, the exponential growth rate of an initially small population with trait y, in a k-morphic resident population with traits $x_1,\dots,x_k$, at equilibrium.*

The fitness function is therefore defined on $\mathcal{X} \times (\cup_{k\geq 1}\mathcal{X}^k)$. By the same arguments as above, it is easy to see that this definition is biologically correct under assumptions (HB2) and (HB4') and that the fitness function satisfies the following properties:

Fundamental properties 3

- *For any $k \geq 1$, $x_1,\dots,x_k \in \mathcal{X}$ and $i \in \{1,\dots,k\}$ such that $\bar{n}^i_{(x_1,\dots,x_n)} > 0$,*

$$f(x_i;x_1,\dots,x_k) = 0. \tag{9}$$

- *For any permutation σ of $\{1,2,\dots,k\}$ and for any $y \in \mathcal{X}$,*

$$f(y;x_{\sigma(1)},\dots,x_{\sigma(k)}) = f(y;x_1,\dots,x_k).$$

- *The k and $k+1$-morphic fitnesses satisfy the relation*

$$f(y;x_1,x_1,x_2,\dots,x_k) = f(y;x_1,\dots,x_k). \tag{10}$$

- *If $k \geq 2$ and if $i \in \{1,\dots,k\}$ satisfies $\bar{n}^i_{(x_1,\dots,x_k)} = 0$, then*

$$f(y;x_1,\dots,x_k) = f(y;x_1,\dots,x_{i-1},x_{i+1},\dots,x_k).$$

Moreover, the Biological Principles 1 and 2 generalise as follows:

Biological principle 4 *Under (HB1), (HB2) and (HB4'), the invasion probability of a trait y in a k-morphic resident population with traits $x_1,\dots,x_k$ is given by*

$$\frac{[b(y;x_1,\dots,x_k)-d(y;x_1,\dots,x_k)]_+}{b(y;x_1,\dots,x_k)},$$

where $b(y;x_1,\dots,x_k)$ and $d(y;x_1,\dots,x_k)$ are respectively the birth and death rates of a unique mutant individual with trait y in the equilibrium resident population with traits $x_1,\dots,x_k$. Moreover, these demographic rates are related with the fitness function via

$$f(y;x_1,\dots,x_k)=b(y;x_1,\dots,x_k)-d(y;x_1,\dots,x_k).$$

The the question of survival of the traits living in the population after an invasion can be easily answered thanks to (HB4'): after the mutant trait y has invaded the k-morphic population with traits $x_1,\dots,x_k$, the dynamics of the $k+1$ traits $y,x_1,\dots,x_k$ has a unique globally asymptotically stable steady state, towards which it converges. The positive coordinates of this equilibrium correspond to the traits that will survive the intraspecific competition.

Is it possible to determine these surviving traits from the fitness function? Using similar arguments as in section 2.2,

- If $\forall i\in\{1,\dots,k\}$, $f(x_i,y;x_1,\dots,x_{i-1},x_{i+1},\dots,x_k)>0$, the unique stable equilibrium of the dynamics of the $k+1$ traits $y,x_1,\dots,x_k$ has no null coordinate and there is coexistence.
- If $\exists i\in\{1,\dots,k\}$ such that $f(x_i;X)<0$ et $\bar n_X\in(\mathbb{R}_+^*)^k$, where $X=(y,x_1,\dots,x_{i-1},x_{i+1},\dots,x_k)$, the equilibrium $(\bar n_X^1,\dots,\bar n_X^i,0,\bar n_X^{i+2},\bar n_X^k)$ is locally asymptotically stable for the $k+1$-morphic dynamics and therefore, by (HB4'), is also globally asymptotically stable.
- If, conversely, $\exists i\in\{1,\dots,k\}$ such that $f(x_i;X)<0$ and $\bar n_X\notin(\mathbb{R}_+^*)^k$, it can happen that the dynamics converges to an equilibrium having two (or more) null coordinates, but also towards an equilibrium for which all the coordinates are non-null. It is actuallty possible to construct two dynamics of three traits x, y and z such that $f(y,x)<0$ and $f(z,x)<0$ (the two traits y and z are unable to invade the trait x alone, *i.e.* $(\bar n_x,0,0)$ is a stable steady state in $\mathbb{R}_+\times\{0\}\times\mathbb{R}_+$ and $\mathbb{R}_+\times\mathbb{R}_+\times\{0\}$) and therefore $f(y;x,z)=f(y,x)<0$ and $f(z;x,y)=f(z,x)<0$, where in one dynamics, the steady state $(\bar n_x,0,0)$ is globally asymptotically stable in $\mathbb{R}_+^3$, and in the other, it is unstable in $\mathbb{R}_+^3$, and the unique globally asymptotically stable steady state has non-null coordinates.

Therefore, there is no simple criterion based on the fitness for fixation and coexistence.

Let us introduce the following biological parameters: for any $k \geq 1$ and $x_1, \dots, x_k \in \mathcal{X}$ with $\mathcal{X} \subset \mathbb{R}^d$ $(d \geq 1)$, we consider

- $\bar{n}_{(x_1,\dots,x_k)} = (\bar{n}^1_{(x_1,\dots,x_k)}, \dots, \bar{n}^k_{(x_1,\dots,x_k)})$, defined in (8);
- $b^i(x_1, \dots, x_k)$, the birth rate of an individual with trait x_i in the k-morphic population of traits $x_1, \dots, x_k$ at equilibrium $i \in \{1, \dots, k\}$);
- $\mu(x)$, the probability that a birth from an individual with trait $x \in \mathcal{X}$ may cause a mutation;
- $m(x,dh)$, the law of $y - x$, where y is a mutant trait born from an individual with trait x; since $y \in \mathcal{X}$, the support of this probability measure is a subset of $\overline{\mathcal{X}} - x := \{z - x : z \in \overline{\mathcal{X}}\}$;
- $b(y; x_1, \dots, x_k)$ and $d(y; x_1, \dots, x_k)$, defined in the Biological Principle 4 ;
- $f(y; x_1, \dots, x_k) = b(y; x_1, \dots, x_k) - d(y; x_1, \dots, x_k)$.

In a k-morphic population with traits $x_1, \dots, x_k$ at equilibrium, each birth from an individual with trait x_i causes a mutation with probability $\mu(x_i)$. These births take place at individual rate $b^i(x_1, \dots, x_k)$ and the density of individuals with traits x_i is $\bar{n}^i_{(x_1,\dots,x_k)}$. Intuitively, the number of individuals with trait x_i is proportional to the corresponding density, so the appearance rate of a mutant born from an individual with trait x_i is proportional to $\mu(x_i) b^i(x_1, \dots, x_k) \bar{n}^i_{(x_1,\dots,x_k)}$. Therefore, if we renormalize properly the time scale, mutations from individuals with trait x_i in the population $x_1, \dots, x_k$ occur with rate $\mu(x_i) b^i(x_1, \dots, x_k) \bar{n}^i_{(x_1,\dots,x_k)}$. Here again, this heuristics will be made clear in chapter II for a particular model.

Finally, under (HB1), (HB2) and (HB4'), the long-time evolution of the population follows a Markovian process governed by the following algorithm:

1. At time $t = 0$, the population is made of a finite number of traits belonging to $\mathcal{X}$.
2. If, at time $t \geq 0$, the population is k-morphic, with traits $x_1, \dots, x_k$, thanks to (HB1), (HB2) et (HB4'), we can suppose that it is at the equilibrium $\bar{n}_{(x_1,\dots,x_k)}$.
3. Consider k independent exponential random variables, with respective parameters $\mu(x_i) b^i(x_1, \dots, x_k) \bar{n}^i_{(x_1,\dots,x_k)}$ for $1 \leq i \leq k$, and assume that the j^{th} one is the smallest. At this exponential time, a mutant born from an individual with trait x_j appears. Its trait is given by $y = x_j + h$, where h has law $m(x_j,dh)$ and is independent of the previous random variables.
4. Let U be a random variable uniformly distributed on [0,1], independent of the previous ones. If $U \leq [f(y; x_1, \dots, x_k)]_+ / b(y; x_1, \dots, x_k)$, we go to

the next step, otherwise, we go back to step 2.

5. The mutant y managed to invade the population. We obtain as follows the traits that will survive the competition : we define $J \subset \{1,2,\dots,j+1\}$ by $j \in J \Leftrightarrow \bar{n}^j_{(y,x_1,\dots,x_k)} > 0$, then the process goes back to step 2, with a card(J)-morphic population, with traits x_{j-1} for $j \in J$ and $j \geq 2$, and y if $1 \in J$.

We obtain a Markov jump process taking values in $\cup_{k\geq 0}\mathcal{X}^k$, sometimes called *polymorphic trait substitution sequence*, that we will call in the sequel "TSS polymorphic model". It is not easy to write its infinitesimal generator, because the dimension of the population state space is not fixed and because there is no simple criterion to decide which traits will survive the competition at step 5.

2.4 Adaptive dynamics jump process: monomorphic case

The monomorphic case is the only one in which we have a criterion to determine which traits will survive to the intra-specific competition after an invasion (Biological Principle 3). Therefore, it is natural to prevent stable polymorphic population from appearing, by modifying (HB5) so that the invasion of a mutant implies its fixation:

Biological hypothesis

(HB5') "Invasion-implies-fixation-principle": *under (HB4), for any* $x \in \mathcal{X}$ *and almost every mutant trait* y *born from* x*, either* $f(y,x) < 0$*, or* $f(y,x) > 0$ *and* $f(x,y) < 0$*, and in this last case the dynamics* $(n_x(t),n_y(t))$ *of the dimorphic population with traits* x *and* y*, after invasion of the resident* x *at equilibrium by the mutant* y*, converges as* $t \to +\infty$ *towards* $(0,\bar{n}_y)$.

We assumed that this assumption was satisfied for *almost every* mutant trait in order to exclude the degenerate case of traits y satisfying $f(y,x) = 0$ (for example when $y = x$).

Under (HB1), (HB2), (HB4) et (HB5'), we can describe as in the previous section a model for the evolution of the population, depending only on the following global parameters: for $x \in \mathcal{X} \subset \mathbb{R}^d$, we consider

– $\bar{n}_x$, the equilibrium density of a monomorphic population with trait x (see assumption (HB4), page 43);

– $b(x)$, the birth rate of an individual with trait x in a monomorphic population with trait x at equilibrium;

– $\mu(x)$ and $m(x,dh)$, defined above (section 2.3);

- $b(y,x)$ et $d(y,x)$, defined in the Biological Principle 2 page 44;
- $f(y,x) = b(y,x) - d(y,x)$.

The evolution follows an algorithm similar to the one of the previous section, except that the initial condition is monomorphic, that there is only one exponential time to be considered in step 3, and that there is always a single trait (the mutant one) surviving the competition in step 5.

We obtain a Markov jump process, taking values in $\mathcal{X}$, sometimes called *monomorphic trait substitution sequence* and that we will call in the sequel “monomorphic TSS model”. Its infinitesimal generator A is defined, for any bounded measurable function $\varphi : \mathcal{X} \to \mathbb{R}$, by

$$A\varphi(x) = \int_{\mathbb{R}^d} (\varphi(x+h) - \varphi(x))\mu(x)b(x)\bar{n}_x \frac{[f(x+h,x)]_+}{b(x+h,x)} m(x,dh). \tag{11}$$

This model was first described by Dieckmann and Law [18]. This is the simplest stochastic model in adaptive dynamics, for which we will obtain a microscopic justification in chapter II.

Assumption (HB5’) and small jumps

Let us consider two traits x and y in $\mathcal{X}$ and assume that the fitness function is differentiable on $\mathcal{X}^2$. By Property 1, $f((x+y)/2,(x+y)/2) = 0$, from what follows that

$$\begin{aligned} f(y,x) &= f\left(\frac{x+y}{2} - \frac{x-y}{2}, \frac{x+y}{2} + \frac{x-y}{2}\right) \\ &= -\frac{x-y}{2}\left[\partial_1 f\left(\frac{x+y}{2},\frac{x+y}{2}\right) - \partial_2 f\left(\frac{x+y}{2},\frac{x+y}{2}\right)\right] + o(|x-y|) \end{aligned} \tag{12}$$

and

$$\begin{aligned} f(x,y) &= f\left(\frac{x+y}{2} + \frac{x-y}{2}, \frac{x+y}{2} - \frac{x-y}{2}\right) \\ &= \frac{x-y}{2}\left[\partial_1 f\left(\frac{x+y}{2},\frac{x+y}{2}\right) - \partial_2 f\left(\frac{x+y}{2},\frac{x+y}{2}\right)\right] + o(|x-y|), \end{aligned} \tag{13}$$

where $\partial_i f$ is the partial derivative of f with respect to the i^{th} variable ($i = 1,2$).

From these two equations follows that $f(y,x)$ and $f(x,y)$ are of opposite signs for x and y close enough and satisfying $\partial_1 f((x+y)/2,(x+y)/2) - \partial_2 f((x+y)/2,(x+y)/2) \neq 0$. Therefore, (HB3) implies (HB5’) as long as

x is far enough from the zeroes of $z \mapsto \partial_1 f(z,z) - \partial_2 f(z,z)$. Geritz [35] has mathematically justified this heuristics for particular models.

Differentiating the relation $f(x,x) = 0$ (Property 1), we get that $\partial_1 f(x,x) - \partial_2 f(x,x) = 0$ if and only if $\partial_1 f(x,x) = \partial_2 f(x,x) = 0$. Such values of the trait x, for which (HB5') may not be satisfied, are called *evolutionary singularities.*

Pairwise invasibility plots

This paragraph will be devoted to explain how to get a visual representation of the evolution of a population characterized by a unique one-dimensional trait ($\mathcal{X} \subset \mathbb{R}$). The possibility of invasion only depends on the sign of the fitness function, so let us decompose $\mathcal{X}^2$ according to the sign of $f(y,x)$. The figure we obtain is called *pairwise invasibility plot*, abbreviated into *PIP* and was introduced in Metz *et al.* [61].

If we assume that the fitness function is $\mathcal{C}^1$, since $f(x,x) = 0$ for any $x \in \mathcal{X}$, we can write

$$f(y,x) = (x - y)\phi(x,y), \tag{14}$$

where the function ϕ is continuous on $\mathcal{X}^2$ and $\mathcal{C}^1$ on $\mathcal{X}^2 \setminus \{(x,x), x \in \mathcal{X}\}$. The set of pairs (x,y) such that $f(y,x) = 0$ is therefore the straight line $y = x$ and the set of the zeroes of ϕ, which is, generically speaking, an implicitly defined curve.

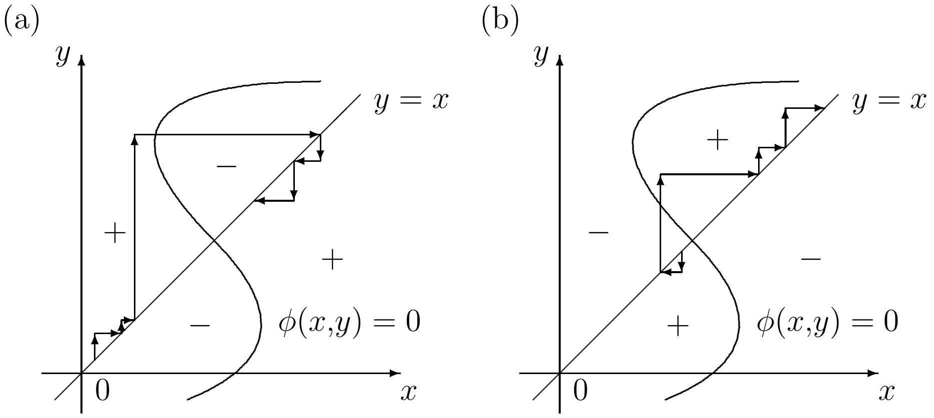


Fig. I.3 – *Two examples of PIP and possible paths of the TSS monomorphic model.*

We give two examples of PIP in Figures I.3 (a) and (b). The + and − signs correspond to the sign of $f(y,x)$. The arrows represent two possible sequences of invasions by mutants: for the invasion of a resident population

x by a mutant y to be possible, the point (x,y) has to belong to an area where $f(y,x) > 0$. That is the meaning of the vertical arrows. The horizontal arrows indicate that, once an invasion occured, there is fixation of the mutant.

2.5 Adaptive dynamics jump process: k-morphic case

We obtain in this section an original TSS model, generalizing the previous one, that will be studied in chapter III.

In view of the criteria of page 48, allowing to predict which trait will survive the competition after the invasion of a mutant, it is natural to generalise the monomorphic TSS model to the k-morphic case, by modifying assumption (HB5) as follows:

Biological hypothesis

(HB5'') *Under (HB4') (page 46), for any* $x_1,\dots,x_k \in \mathcal{X}$ *such that* $\bar{n}_{(x_1,\dots,x_k)} \in (\mathbb{R}_+^*)^k$ *and for almost every mutant trait* y *born from* x_i *where* $i \in \{1,\dots,k\}$*, we have* $f(y;x_1,\dots,x_k) < 0$*, or* $f(y;x_1,\dots,x_k) > 0$ *and in this second case*

$$\bar{n}_{(y,x_1,\dots,x_{i-1},x_{i+1},\dots,x_k)} \in (\mathbb{R}_+^*)^k, \tag{15}$$

$$f(x_i;y,x_1,\dots,x_{i-1},x_{i+1},\dots,x_k) < 0 \tag{16}$$

$$\textit{et} \quad f(x_j;y,x_1,\dots,x_{j-1},x_{j+1},\dots,x_k) > 0 \quad \textit{for} \quad j \neq i. \tag{17}$$

As we saw in Section 2.3, this assumption simply means that, when the invasion by the mutant y born from x_i occurs, the dynamics of the $k+1$ traits $y,x_1,\dots,x_k$ converges to the equilibrium $\bar{n}_{(y,x_1,\dots,x_k)}$, whose coordinates are all different from zero, except the one corresponding to the trait x_i. Therefore, the invasion of a mutant trait implies its fixation, in the sense that it replaces the trait from which it was born, whereas the other traits survive. In particular, if the initial population is k-morphic, it will remain so for ever.

Under the biological assumptions (HB1), (HB2), (HB4') and (HB5''), the evolution of the population can be described by the algorithm of section 2.3, where, in step 5, all the traits survive except the one from which is born the mutant trait. We obtain a $\mathcal{X}^k$-valued Markov process with infinitesimal generator given, for any bounded measurable function $\phi : \mathcal{X}^k \to \mathbb{R}$ and $\mathbf{x} = (x_1,\dots,x_k) \in \mathcal{X}^k$, by

$$A\phi(\mathbf{x}) = \sum_{i=1}^{k} \int_{\mathbb{R}^d} (\phi(\mathbf{x} + (h)_i) - \phi(\mathbf{x}))[g_i(x_i + h;\mathbf{x})]_+ m(x_i,dh), \tag{18}$$

where $(h)_i$ is the vector whose coordinates are all zero except the i^{th} one which is h, and where

$$g_i(y;x_1,\dots,x_k)=\mu(x_i)b^i(x_1,\dots,x_k)\bar{n}^i_{(x_1,\dots,x_k)}\frac{f(y;x_1,\dots,x_k)}{b(y;x_1,\dots,x_k)},\tag{19}$$

with the notations of section 2.3.

Assumption (HB5") and small jumps

We can (heuristically) prove that, under (HB3) and if f is $\mathcal{C}^1$, (HB5") is a reasonable assumption, by writing equations similar to (12) and (13):

- On one hand, since $\bar{n}_{(x_1,\dots,x_k)}\in(\mathbb{R}^*_+)^k$, for any $j\in\{1,\dots,k\}$, excluding the case when the fitness is null,

  $$f(x_j;x_1,\dots,x_{j-1},x_{j+1},\dots,x_k)>0.$$

  Hence, for y close enough from x_i, the same inequalities hold replacing x_j by y, which implies (15).
  Moreover, for any $j\neq i$, by (10) (Properties 3 page 47),

  $$f(x_j;x_1,\dots,x_{j-1},x_{j+1},\dots,x_k)=f(x_j;x_i,x_1,\dots,x_{j-1},x_{j+1},\dots,x_k),$$

  and thus $f(x_j;y,x_1,\dots,x_{j-1},x_{j+1},\dots,x_k)>0$ if y and x_i are close enough.
- On the other hand,

  $$f(y;x_1,\dots,x_k)=-\frac{x_i-y}{2}\times$$
  $$(\partial_1 f-\partial_{i+1}f)\left(\frac{x_i+y}{2};x_1,\dots,x_{i-1},\frac{x_i+y}{2},x_{i+1},\dots,x_k\right)+o(|x_i-y|)$$

  and

  $$f(x_i;x_1,\dots,x_{i-1},y,x_{i+1},\dots,x_k)=\frac{x_i-y}{2}\times$$
  $$(\partial_1 f-\partial_{i+1}f)\left(\frac{x_i+y}{2};x_1,\dots,x_{i-1},\frac{x_i+y}{2},x_{i+1},\dots,x_k\right)+o(|x_i-y|),$$

  where $\partial_j f$ denotes the partial derivative of f, as a function defined on $\mathcal{X}^{k+1}$, with respect to the j^{th} variable. As in Section 2.4, this implies that $f(y;x_1,\dots,x_k)$ and $f(x_i;x_1,\dots,x_{i-1},y,x_{i+1},\dots,x_k)$ have opposite signs if x_i-y is small enough and if $\partial_1 f-\partial_{i+1}f$ is different from zero at $((x_i+y)/2;x_1,\dots,x_{i-1},(x_i+y)/2,x_{i+1},\dots,x_k)$.

Therefore, (HB3) implies (HB5”) as long as the population is far enough from the non-trivial zeros of f and from the *evolutionary singularities*, defined as the points $(x_1,\dots,x_k)$ satisfying $\partial_i f(x_i;x_1,\dots,x_k)=0$ for some $i\in\{1,\dots,k\}$ (according to (9), this is equivalent to the fact that $(\partial_1 f-\partial_{i+1}f)(x_i;x_1,\dots,x_k)=0$).

3 Small jumps assumption: biological consequences

This section deals with two consequences of assumption (HB3) forming important biological advances in the theory of adaptive dynamics. We will first present the classification of evolutionary singularities in dimension 1 ($\mathcal{X}\subset\mathbb{R}$) according to the behavior of the TSS models in their neighborhood, which provides in particular a biological interpretation of the phenomenon of evolutionary branching observed in the simulations of section 1.3 of the introduction of this thesis. Let us insist on the fact that our mathematical study of adaptive dynamics will not deal with this particular phenomenon. We include these biological results for sake of completeness about the major issues of adaptive dynamics. The second part of this section is devoted to a brief presentation of the canonical equation of adaptive dynamics — an ODE approximating the evolution of trait in the population — that is the subject of the mathematical study of chapter III.

These results generalize of the game-theoretical approach of evolution (cf. Maynard Smith [58]), since they extend the notion of ESS (*evolutionary stable strategy*) and allow to describe the evolutionary transition leading to these equilibria, so that one can predict in which ESS the population is going to settle.

3.1 Classification of the evolutionary singularities in dimension 1

This section follows closely section 7 of Diekmann [19], except for the justification of the branching criterion. Let us also refer to Metz *et al.* [59] and Geritz *et al.* [37].

We will assume in this section that $\mathcal{X}\subset\mathbb{R}$ and that the fitness function is $\mathcal{C}^2$ on $\mathcal{X}^2$. It follows immediately from the formula $f(x,x)=0$ (Property 1) that, for any $x\in\mathcal{X}$,

$$\begin{aligned}&\partial_1 f(x,x)+\partial_2 f(x,x)=0\\ \text{and}\quad &\partial^2_{11}f(x,x)+2\partial^2_{12}f(x,x)+\partial^2_{22}f(x,x)=0.\end{aligned}\tag{20}$$

Let us consider a monomorphic population with trait x, and assume first that x is not an evolutionary singularity ($\partial_1 f(x,x) \neq 0$). Under (HB3), the mutant traits born from x for which the invasion is possible are determined by the sign of $\partial_1 f(x,x)$: if $\partial_1 f(x,x) > 0$, then $f(y,x) > 0$ for $y > x$ in a neighborhood of x, and the evolution can only proceed towards greater traits. Similarly, if $\partial_1 f(x,x) < 0$, the evolution can only proceed towards lesser traits. Therefore, the "fitness gradient" $\partial_1 f(x,x)$ determines locally the direction of evolution, except when x is an evolutionary singularity

Now, let us consider a monomorphic population with trait x^* where x^* is an evolutionary singularity, *i.e.* a solution to $\phi(x^*,x^*) = 0$, where ϕ has been introduced in (14). In other words, (x^*,x^*) is an intersection point between the curve $\phi(x,y) = 0$ and the line $y = x$ (cf. Figure I.3).

We will be interested in the following biologically motivated properties of evolutionary singularities:

Definition 3

- *The evolutionary singularity x^* is an ESS (*evolutionary stable strategy*) if no other trait y in a neighborhood of x^* can invade the equilibrium resident population with trait x^*.*
- *The evolutionary singularity x^* is a CSS (*convergence stable strategy*) if x^* is an* attractor *for the monomorphic TSS model, in the sense that the invasion by a mutant trait $y < x$ (resp. $y > x$) of a resident population with trait $x < x^*$ (resp. $x > x^*$) is impossible in a neighborhood of x^*.*
- *We will say that there is* mutual invasibility *at the evolutionary singularity x^* when there exists two traits x and y in any neighborhood of x^* such that $f(y,x) > 0$ and $f(x,y) > 0$ (*i.e.* there is coexistence of x and y, by the Biological Principle 3).*

An ESS corresponds to an *absorbing state* for the monomorphic TSS model, whereas a CSS corresponds to a *stable steady state.* In this sense, the expression *evolutionary stable strategy* is somewhat incorrect.

The notion of mutual invasibility takes into account the difference between the monomorphic TSS model and the general one: it allows to identify the evolutionary singularities where the monomorphic TSS model do not describe correctly the evolution, and where a stable dimorphism may emerge from a monomorphic population.

We will give criterions for these three notions according to the values of

$$c_{11} := \partial^2_{11} f(x^*,x^*) \quad \text{and} \quad c_{22} := \partial^2_{22} f(x^*,x^*)$$

(remind that $\partial^2_{12} f(x^*,x^*) = -(c_{11} + c_{22})/2$).

ESS x^* is an ESS if, for any y close enough to x^*, $f(y,x^*) < 0$, *i.e.* if the function $y \mapsto f(y,x^*)$ admits a local maximum at x^*. Therefore, x^* is an ESS if

$$c_{11} < 0$$

(with an undetermined case when $c_{11} = 0$). Conversely, if $c_{11} > 0$, any mutant trait y sufficiently close to x^* can invade the trait x^*.

CSS x^* is a CSS if $f(y,x) < 0$ for $y < x < x^*$ and $x^* < x < y$ close enough to x^*. Then, $\partial_1 f(x,x) \geq 0$ if $x < x^*$ and $\partial_1 f(x,x) \leq 0$ if $x > x^*$ in a neighborhood of x^*. Therefore, x^* is a CSS if

$$\frac{d}{dx}(x \mapsto \partial_1 f(x,x))(x^*) = c_{11} + \partial^2_{12} f(x^*,x^*) = \frac{c_{11} - c_{22}}{2} < 0$$

(with an undetermined case when $c_{11} = c_{22}$). Conversely, if $c_{11} > c_{22}$, x^* is repulsive for the monomorphic TSS model.

Mutual invasibility Let us first observe that $c_{11} = -2\partial_2\phi(x^*,x^*)$ and $c_{22} = 2\partial_1\phi(x^*,x^*)$. Therefore, if $c_{11} \neq 0$ or $c_{22} \neq 0$, the equation $\phi(x,y) = 0$ defines implicitly a $\mathcal{C}^1$ curve in a neighborhood of (x^*,x^*), inside which, moreover, $\phi(x,y)$ has an opposite sign on each side of this curve. Let us assume that the tangent to this curve at (x^*,x^*), given by the equation $(x - x^*)c_{11} = -(y - y^*)c_{22}$, is not equal to the line $y = 2x^* - x$, *i.e.* $c_{11} + c_{22} \neq 0$. Then, we can assume that the curve $\phi(x,y) = 0$ does not intersect the line $y = 2x^* - x$ in the neighborhood of (x^*,x^*) we consider, except at (x^*,x^*). Therefore, by (14), the fitness has a constant sign on the line $y = 2x^* - x$.

Now, take x and y in this neighborhood such that $f(y,x) > 0$ and $f(x,y) > 0$. Since the segment linking (x,y) and (y,x) cuts the line $y = x$, where the sign of the fitness changes it must necessarily cut the curve $\phi(x,y) = 0$. It is then clear that the fitness has the same sign at (x,y), (y,x) and on the line $y = 2x^* - x$. Therefore, the fitness is positive on this line and there is mutual invasibility at x^* if

$$\frac{d^2}{dx^2}(x \mapsto f(2x^* - x,x))(x^*) = c_{11} - 2\partial^2_{12} f(x^*,x^*) + c_{22} = 2(c_{11} + c_{22}) > 0$$

(with a undetermined case when $c_{11} + c_{22} = 0$). Conversely, if $c_{11} + c_{22} < 0$, there is not mutual invasibility.

What happens after a mutual invasion has occured? Once two trait x_1 and x_2 coexist near x^*, we have to take into account the dimorphic fitness $f(y;x_1,x_2)$. Let us recall that $f(x_1;x_1,x_2) = f(x_2;x_1,x_2) = 0$ et $f(y;x,x) =$

$f(y,x)$ (Property 3). If we assume that this function of three variables is $\mathcal{C}^1$ and that $c_{11} < 0$, then the function $y \mapsto f(y,x^*)$ is strictly concave in the neighborhood of x^* and, as shown in Figure I.4 (a), if x_1 and x_2 are close enough to x^*, the function $y \mapsto f(y;x_1,x_2)$ is also strictly concave in this neighborhood of x^*, containing x_1 and x_2. Similarly, if $c_{11} > 0$, the function $y \mapsto f(y;x_1,x_2)$ is strictly convex in a neighborhood of x^*, x_1 and x_2, as shown in Figure I.4 (b).

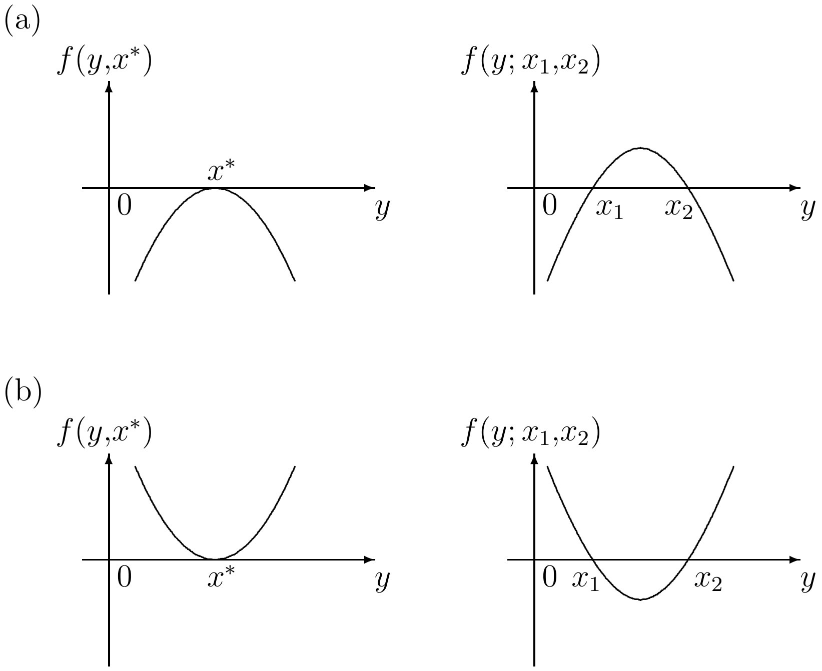


FIG. I.4 – *Possible configurations of the fitness in the neighborhood of an evolutionary singularity with mutual invasibility.*

In the case of a maximum (Figure I.4 (a), $c_{11} < 0$), a mutant trait y belonging to the neighborhood of x^*, x_1 and x_2 obtained above can invade the dimorphic population with traits x_1 and x_2 only when it lies between x_1 and x_2. Then, because of the regularity of the fitness, at least one of the traits x_1 and x_2 must disappear, since otherwise, the function $z \mapsto f(z;y,x_1,x_2)$, that we can assume strictly concave in the neighborhood of x^*, x^1 and x^2 above, would have three zeroes at y, x_1 and x_2, which is impossible. Therefore, even if the population remains dimorphic, the distance between the two traits decreases at each invasion. In this situation, so-called "convergent dimorphism", there is no real diversification in the population.

In the case of a minimum (Figure I.4 (b)), we obtain in a similar way a “divergent dimorphism”, *i.e.* an increase of the distance between the two coexisting trait values. If we exclude the case when x^* is repulsive ($c_{11} > c_{22}$), in which the neighborhood of x^* cannot be reach by the monomorphic TSS model, this case reduces to the condition

$$c_{22} > c_{11} > 0. \tag{21}$$

An evolutionary singularity satisfying this relation is called an *branching point*.

Therefore, the phenomenon of *evolutionary branching* (the appearance and persistence of two coexisting trait values, and the increase of the phenotypic distance between them) can only be observed in the neighborhood of an evolutionary singularity which is a CSS with mutual invasibility.

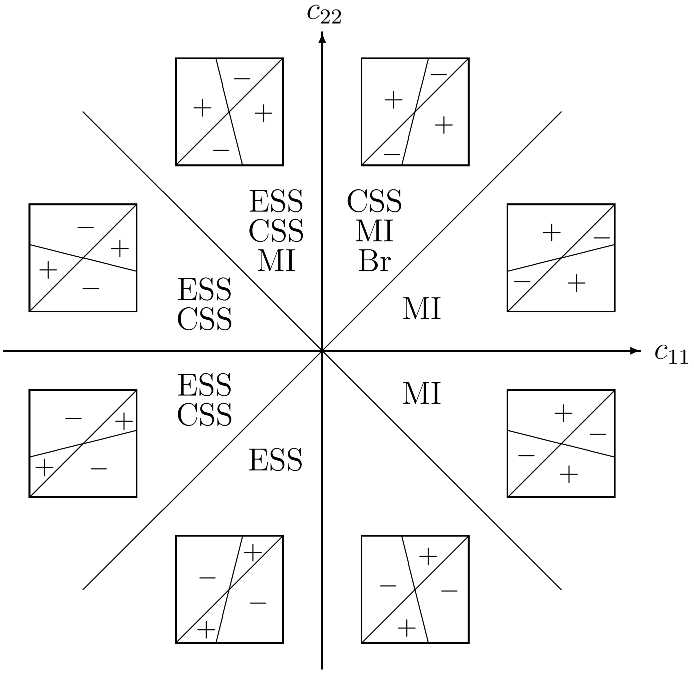


Fig. I.5 – *Classification of the evolutionary singularities according to the values of* c_{11} *and* c_{22}*, and the corresponding PIPs*

The discovery of the evolutionary branching has strong biological implications: on the one hand, it allows to explain the diversification phenomenon in a population, and on the other hand, it might be the first step of a speciation (the split of a species in two distinct species). Until now, the main scenario for speciation proposed by evolutionary biologists required a spatial separation of the species in two (or more) populations (*allopatric speciation*). The phenotypic random divergence or the local selection of these two populations may lead to a reproductive incompatibility when the geographical contact between them is restored. The adaptive dynamics theory proposes a scenario of *sympatric speciation* initiated in a geographically homogeneous population by the evolutionary branching (cf. Dieckmann and Doebeli [16] and Doebeli and Dieckmann [21, 22]).

We sum up the results of this secion in Figure I.5. The nature of the evolutionary singularity is indicated by the abbreviations ESS, CSS, MI (mutual invasibility), and Br (evolutionary branching). We have drawn the shape of the PIP corresponding to each possible case, with the convention that the abscissa is the resident trait x and the ordinate is the mutant trait y. the slope of the tangent to the curve $\phi(x,y)=0$ at (x^*,x^*) is c_{22}/c_{11}.

3.2 The canonical equation of adaptive dynamics

The last biological result of this chapter deals with the behavior of the monomorphic TSS model under the biological assumption (HB3). Dieckmann and Law [18] have introduced heuristically an ODE approximating this model. We will not describe here the nature of this approximation, nor their argument, since we will prove in chapter III a convergence result, in a more precise way, toward a more general ODE, when the amplitude of the mutations converges to 0.

The ODE of [18] writes, in the case of a one-dimensional trait ($\mathcal{X}\subset\mathbb{R}$),

$$\frac{dx}{dt}=\frac{\sigma^2(x)}{2}\mu(x)\bar{n}_x\partial_1 f(x,x), \tag{22}$$

where $\sigma^2(x)$ is the variance of the probability measure $m(x,dh)$, where $\mu(x)$, $\bar{n}_x$, $m(x,dh)$ and $f(y,x)$ are defined page 50, and where the measure $m(x,\cdot)$ is assumed *symmetrical* on $\mathbb{R}$ with respect to 0 for any $x\in\mathcal{X}$.

This ODE is called *canonical equation of adaptives dynamics*, with reference to the canonical equation of evolution, obtained by Lande [52] as a limit of quantitative genetics models, and based on the principle of maximisation of the fitness with respect to its gradient (cf. Wright [72]). The general form of this equation is $\dot{x}=k(x)\nabla W(x)$, where $W(x)$ is the fitness of trait x and where k describes the speed of evolution. The main difference between this equation and (22) lies in the fact that the fitness depends on the mutant trait *and* the resident trait. Moreover, the canonical equation of adaptive dynamics gives an explicit formula for $k(x)$ as a function of individual parameters.

4 Conclusion

Adaptive dynamics has been recently proposed as a framework to model evolution, emphasizing the ecological aspect of the phenomenon. Let us mention the following advantages of this approach:

- the fitness is explicitely defined, and is no more an *ad hoc* notion;
- the adaptive dynamics models are based on a density-dependance in the population dynamics (unlike Wright-Fisher-like models);
- the adaptive dynamics theory takes into account the stochastic and discrete aspects of individual dynamics;
- unlike the game theory approach, the adaptive dynamics models can describe the entire dynamics of evolution (TSS models and canonical equation) in a trait space that can be continuous, and not only the asymptotic evolutionary state in a finite trait space;
- the theory provides a complete classification of evolutionary singularities in dimension 1.

The field of application of this theory is wide. We can mention the study of the evolution of virulence in a host-pathogen population (Pugliese [63]), of the seed size (Geritz *et al.* [38]) or of the social behavior (Le Galliard *et al.* [55]).

Let us also mention two major drawbacks of this approach:

- The assumption of rare mutations leads to a population where the phenotypic variability is *de facto* restricted to a small number of phenotypes, which practically never happens in real populations. The phenotypic variability is a crucial feature of evolution, since it gives to the population the opportunity to react to a sudden environmental change (cf. Kimura [48]).
- The assumption of large population also leads to unrealistic feature for the long term evolution of the population: for various fitness configurations, the population can be trapped in the neighborhood of an evolutionary singularity without any way to go away. The phenomenon of *genetic drift*, outlined by the classical population genetics (cf. [48]), is thus ignored.

Chapter II

A microscopic interpretation for adaptive dynamics trait substitution sequence models

Abstract: We consider an interacting particle Markov process for Darwinian evolution in an asexual population, involving a linear birth rate, a density-dependent logistic death rate, and a probability μ of mutation at each birth event. We introduce a renormalization parameter K scaling the size of the population, which leads, when $K \to +\infty$ and $\mu \equiv 0$, to a deterministic dynamics for the density of individuals holding a given trait. By combining in a non-standard way the limits of large population ($K \to +\infty$) and of small mutations ($\mu \to 0$), we prove that a time scale separation between the birth and death events and the mutation events occurs and that the interacting particle microscopic process converges for finite dimensional distributions to the biological model of evolution known as the "monomorphic trait substitution sequence" model of adaptive dynamics [59, 18], which describes the Darwinian evolution in an asexual population as a Markov jump process in the trait space.

1 Introduction

We will study in this article the link between two biological models of Darwinian evolution in an asexual population. The first one is a system of interacting particles modeling evolution at the *individual* level, referred below as the *microscopic model*, and which has been already studied in Dieckmann [15], Fournier and Méléard [33] and Ferrière *et al* [32]. The second one models the evolution at the *population* level as a jump Markov process in the space of traits characterizing individuals, called "trait substitution sequence", and referred below as the *TSS model.* The TSS model belongs to the recent biological theory of evolution called *adaptive dynamics* (Hofbauer and Sigmund [41], Marrow *et al.* [57] and Metz *et al.* [61]), and has been introduced in 1996 by Metz *et al.* [59] and Dieckmann and Law [18]. This model and its generalizations have revealed a powerful tool for understanding various evolutionary phenomena, such as polymorphism (stable coexistence of different traits) or evolutionary branching (evolution of a monomorphic population to a polymorphic one). The heuristics leading to the TSS model (cf. [59, 18] and chapter I of this thesis) is actually based on the biological assumptions of large population and rare mutations, and on another assumption stating that no two different types of individuals can coexist on a long time scale: the competition eliminates one of them. We propose to prove in this article a convergence result of the microscopic model to this TSS model when the parameters are normalized in a non-standard way, leading to a time scale separation. This limit combines a *large population* asymptotic with a *rare mutations* asymptotic. Such a result will provide a mathematical justification of the TSS model and of the biological heuristics on which it is based.

Let us describe the microscopic model: in a population, Darwinian evolution acts on a set of phenotypes, or *traits*, characterizing each individual's ability to survive and reproduce (e.g. body size, rate of food intake, age at maturity). We will consider a finite number of quantitative traits in an asexual population (clonal reproduction), and we will assume that the trait space $\mathcal{X}$ is a compact subset of $\mathbb{R}^l$ $(l \geq 1)$.

The microscopic model involves the three basic mechanisms of Darwinian evolution: *heredity*, which transmits traits to new offsprings, *mutation*, driving a variation in the trait values in the population, and *selection* between these different trait values, caused by the competition for limited resources or area.

For any $x,y \in \mathcal{X}$, we introduce the following biological parameters

$b(x) \in \mathbb{R}_+^*$ is the rate of birth from an individual holding trait x.

$d(x) \in \mathbb{R}_+^*$ is the rate of "natural" death for an individual holding trait x.

$\alpha(x,y) \in \mathbb{R}_+^*$ is competition kernel representing the pressure felt by an individual holding trait x from an individual holding trait y.

$\mu(x) \in [0,1]$ is the probability that a mutation occurs in a birth from an individual with trait x.

$m(x,dh)$ is the law of $h = y - x$, where the mutant trait y is born from an individual with trait x. It is a probability measure on $\mathbb{R}^l$, and since y

must be in the trait space $\mathcal{X}$, the support of $m(x,\cdot)$ is a subset of

$$\mathcal{X} - x = \{y - x : y \in \mathcal{X}\}.$$

$K \in \mathbb{N}^*$ is a parameter rescaling the competition kernel $\alpha(\cdot,\cdot)$. Biologically, K may be linked to the ressources or area available, and is related to the biological concept of "carrying capacity". As will appear later, this parameter is linked to the size of the population: large K means a large population (provided that the initial condition is proportional to K).

$u_K \in [0,1]$ is a parameter depending on K rescaling the probability of mutation $\mu(\cdot)$. Small u_K means rare mutations.

Let us also introduce the following notations, used throughout this paper :

$$\bar{n}_x = \frac{b(x) - d(x)}{\alpha(x,x)}, \tag{1}$$

$$\beta(x) = \mu(x)b(x)\bar{n}_x \tag{2}$$

$$\text{and} \quad f(y,x) = b(y) - d(y) - \alpha(y,x)\bar{n}_x. \tag{3}$$

We consider, at any time $t \geq 0$, a finite number N_t of individuals, each of them holding a trait value in $\mathcal{X}$. Let us denote by $x_1,\dots,x_{N_t}$ the trait values of these individuals. The state of the population, rescaled by K, at time $t \geq 0$ can be represented by the finite point measure on $\mathcal{X}$

$$\nu_t^K = \frac{1}{K}\sum_{i=1}^{N_t} \delta_{x_i},$$

where δ_x is the Dirac measure at x. Let $\mathcal{M}_F$ denote the set of finite nonnegative measures on $\mathcal{X}$, and define

$$\mathcal{M}^K = \left\{ \frac{1}{K}\sum_{i=1}^{n} \delta_{x_i} : n \geq 0,\, x_1,\dots,x_n \in \mathcal{X} \right\},$$

An individual holding trait x in the population ν_t^K gives birth to another individual with rate $b(x)$ and dies with rate

$$d(x) + \int_{\mathcal{X}} \alpha(x,y)\nu_t^K(dy) = d(x) + \frac{\sum_{i=1}^{N_t} \alpha(x,x_i)}{K}.$$

A newborn holds the same trait value than its progenitor's with probability $1 - u_K\mu(x)$, and with probability $u_K\mu(x)$, the newborn is a mutant which trait value y is chosen according to $y = x + h$, where h is a random variable with law $m(x,dh)$.

In other words, the process $(\nu_t^K, t \geq 0)$ is a $\mathcal{M}^K$-valued Markov process with infinitesimal generator defined for any bounded measurable functions ϕ from $\mathcal{M}^K$ to $\mathbb{R}$ by

$$\begin{aligned} L^K\phi(\nu) = &\int_{\mathcal{X}} \left(\phi\left(\nu + \frac{\delta_x}{K}\right) - \phi(\nu)\right)(1 - u_K\mu(x))b(x)K\nu(dx) \\ &+ \int_{\mathcal{X}}\int_{\mathbb{R}^l} \left(\phi\left(\nu + \frac{\delta_{x+h}}{K}\right) - \phi(\nu)\right) u_K\mu(x)b(x)m(x,dh)K\nu(dx) \\ &+ \int_{\mathcal{X}} \left(\phi\left(\nu - \frac{\delta_x}{K}\right) - \phi(\nu)\right)\left(d(x) + \int_{\mathcal{X}} \alpha(x,y)\nu(dy)\right)K\nu(dx). \end{aligned} \tag{4}$$

The first term (linear) describes the births without mutation, the second term (linear) describes the births with mutation, and the third term (non-linear) describes the deaths by oldness or competition. This logistic density-dependence models the competition in the population, and hence drives the selection process.

Let us denote by (A) the following three assumptions

(A1) b, d and α are measurable functions, and there exists $\bar{b},\bar{d},\bar{\alpha} < +\infty$ such that

$$b(\cdot) \leq \bar{b}, \quad d(\cdot) \leq \bar{d} \quad \text{and} \quad \alpha(\cdot,\cdot) \leq \bar{\alpha}.$$

(A2) $m(x,dh)$ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^l$ with density $m(x,h)$, and there exists a function $m : \mathbb{R}^l \to \mathbb{R}_+$ such that $m(x,h) \leq m(h)$ for any $x \in \mathcal{X}$ and $h \in \mathbb{R}^l$, and $\int m(h)dh < \infty$.

(A3) $\mu(x) > 0$ and $b(x) - d(x) > 0$ for any $x \in \mathcal{X}$, and there exists $\underline{\alpha} > 0$ such that

$$\underline{\alpha} \leq \alpha(\cdot,\cdot).$$

For fixed K, under (A1) and (A2) and assuming that $\mathbf{E}(\langle \nu_0^K, \mathbf{1}\rangle) < \infty$ (where $\langle \nu, f\rangle$ denotes the integral of the measurable function f with respect to the measure ν), the existence and uniqueness in law of a process with infinitesimal generator L^K has been proved by Fournier and Méléard [33].

In this model, the biological assumption of large population corresponds to the limit $K \to +\infty$, and the assumption of rare mutations to $u_K \to 0$. In order to give a precise formulation of the biological assumption on the impossibility of coexistence of two different traits, let us define:

Definition 1.1

(a) *For any $K \geq 1$, $b,d,c \geq 0$ and for any $\mathbb{N}/K$-valued random variable z, we will denote by $\mathbf{P}^K(b,d,c,z)$ the law of the $\mathbb{N}/K$-valued Markov jump* *process with initial state z and with transition rates*

$$\begin{array}{ll} ib & \text{from } i/K \text{ to } (i+1)/K, \\ i(d+ci/K) & \text{from } i/K \text{ to } (i-1)/K. \end{array}$$

(b) *For any $K \geq 1$, $b_k,d_k,c_{kl} \geq 0$ with $k,l \in \{1,2\}$, and for any $\mathbb{N}/K$-valued random variables z_1 and z_2, we will denote by*

$$\mathbf{Q}^K(b_1,b_2,d_1,d_2,c_{11},c_{12},c_{21},c_{22},z_1,z_2)$$

the law of the $(\mathbb{N}/K)^2$-valued Markov jump process with initial state (z_1,z_2) and with transition rates

$$\begin{array}{ll} ib_1 & \text{from } (i/K,j/K) \text{ to } ((i+1)/K,j/K), \\ jb_2 & \text{from } (i/K,j/K) \text{ to } (i/K,(j+1)/K), \\ i(d_1+c_{11}i/K+c_{12}j/K) & \text{from } (i/K,j/K) \text{ to } ((i-1)/K,j/K), \\ j(d_2+c_{21}i/K+c_{22}j/K) & \text{from } (i/K,j/K) \text{ to } (i/K,(j-1)/K). \end{array}$$

Fix x and y in $\mathcal{X}$. The proof of the following two results can be found in chapter 11 of Ethier and Kurtz [28].

Proposition 1.1
(a) *Assume $\mu \equiv 0$ and $\nu_0^K = N_x^K(0)\delta_x$. Then, for any $t \geq 0$, $\nu_t^K = N_x^K(t)\delta_x$, where N_x^K has the law $\mathbf{P}^K(b(x),d(x),\alpha(x,x),N_x^K(0))$. Assume $N_x^K(0) \to n_x(0)$ in probability when $K \to +\infty$. Then, the sequence (N_x^K) converges in probability on $[0,T]$ for the uniform norm to the deterministic function n_x with initial condition $n_x(0)$ solution to*

$$\dot{n}_x = (b(x) - d(x) - \alpha(x,x)n_x)n_x. \tag{5}$$

(b) *Assume $\mu \equiv 0$ and $\nu_0^K = N_x^K(0)\delta_x + N_y^K(0)\delta_y$. Then, for any $t \geq 0$, $\nu_t^K = N_x^K(t)\delta_x + N_y^K(t)\delta_y$, where (N_x^K,N_y^K) has the law*

$$\mathbf{Q}^K(b(x),b(y),d(x),d(y),\alpha(x,x),\alpha(x,y),\alpha(y,x),\alpha(y,y),N_x^K(0),N_y^K(0)).$$

Assume $N_x^K(0) \to n_x(0)$ and $N_y^K(0) \to n_y(0)$ in probability when $K \to +\infty$. Then, (N_x^K,N_y^K) converges in probability when $K \to +\infty$ on $[0,T]$ for the uniform norm to the deterministic function (n_x,n_y) with initial condition $(n_x(0),n_y(0))$ solution to

$$\begin{cases} \dot{n}_x = (b(x) - d(x) - \alpha(x,x)n_x - \alpha(x,y)n_y)n_x \\ \dot{n}_y = (b(y) - d(y) - \alpha(y,x)n_x - \alpha(y,y)n_y)n_y. \end{cases} \tag{6}$$

Note that (5) has two steady states, 0, unstable, and $\bar{n}_x$, defined in (1), stable. System (6) has at least three steady states, $(0,0)$, unstable, $(\bar{n}_x,0)$ and $(0,\bar{n}_y)$.

Let us introduce the following assumption:

(B) Given any $x \in \mathcal{X}$, Lebesgue almost any $y \in \mathcal{X}$ satisfies one of the two following conditions:

$$\text{either} \quad (b(y)-d(y))\alpha(x,x)-(b(x)-d(x))\alpha(y,x)<0, \tag{7}$$

$$\text{or} \quad \begin{cases} (b(y)-d(y))\alpha(x,x)-(b(x)-d(x))\alpha(y,x)>0, \\ (b(x)-d(x))\alpha(y,y)-(b(y)-d(y))\alpha(x,y)<0. \end{cases} \tag{8}$$

Assumption (B) is the mathematical formulation of the impossibility of coexistence of two different traits. Actually, an elementary analysis of system (6) shows that (cf. e.g. Istas [42] p. 25–27 or Theorem 1 of Annex A):

Proposition 1.2 *If x and y satisfy (7), then $(\bar{n}_x,0)$ is a stable steady state of (6). If x and y satisfy (8), then $(\bar{n}_x,0)$ is an unstable steady state of (6), $(0,\bar{n}_y)$ is stable, and any solution to (6) with initial state in $(\mathbb{R}_+^*)^2$ converges to $(0,\bar{n}_y)$ when $t \to +\infty$.*

The TSS model is a Markov jump process in $\mathcal{X}$ with infinitesimal generator given, for any bounded measurable function φ from $\mathcal{X}$ to $\mathbb{R}$, by

$$A\varphi(x)=\int_{\mathbb{R}^l}(\varphi(x+h)-\varphi(x))\beta(x)\frac{[f(x+h,x)]_+}{b(x+h)}m(x,h)dh, \tag{9}$$

where $[a]_+$ denotes the positive part of $a \in \mathbb{R}$, and where $\beta(x)$ and $f(y,x)$ are defined in (2) and (3). The existence and uniqueness in law of a process generated by A holds as soon as $\beta(x)[f(y,x)]_+/b(y)$ is bounded (see e.g. Ethier and Kurtz [28]), which is true under assumption (A) ($[f(y,x)]_+/b(y) \leq 1$).

Our main result writes:

Theorem 1.1 *Assume (A) and (B). Fix a sequence $(u_K)_{K\in\mathbb{N}}$ in $[0,1]^{\mathbb{N}}$ such that*

$$\forall V>0, \quad \exp(-VK) \ll u_K \ll \frac{1}{K\log K} \tag{10}$$

(where $f(K) \ll g(K)$ means that $f(K)/g(K) \to 0$ when $K \to \infty$). Fix also $x \in \mathcal{X}$, $\gamma > 0$ and a sequence of $\mathbb{N}$-valued $\mathbb{L}^1$ random variables $(\gamma_K)_{K\in\mathbb{N}}$ such that γ_K/K converges in law to γ. Consider the process $(\nu_t^K, t \geq 0)$ generated by (4) with initial state $\frac{\gamma_K}{K}\delta_x$. Then, for any $n \geq 1$, $\varepsilon > 0$ and $0 < t_1 < t_2 < \ldots < t_n < \infty$, and for any measurable subsets $\Gamma_1,\ldots,\Gamma_n$ of $\mathcal{X}$,

$$\lim_{K\to+\infty} \mathbf{P}\big(\forall i \in \{1,\ldots,n\},\ \mathrm{Supp}(\nu^K_{t_i/Ku_K}) \text{ is a singleton } \{x_i\},\ x_i \in \Gamma_i$$
$$\text{and } |\langle \nu^K_{t_i/Ku_K},\mathbf{1}\rangle - \bar{n}_{x_i}| < \varepsilon\big) = \mathbf{P}(\forall i \in \{1,\ldots,n\},\ X_{t_i} \in \Gamma_i) \tag{11}$$

where for any $\nu \in \mathcal{M}_F$, $\mathrm{Supp}(\nu)$ *is the support of* ν *and* $(X_t, t \geq 0)$ *is the TSS process generated by (9) with initial state* x.

Corollary 1.1 *With the same notations and assumptions as in Theorem 1.1, when* $K \to +\infty$, *the process* $(\nu^K_{t/Ku_K}, t \geq 0)$ *converges in the sense of the finite dimensional distributions, for the topology on* $\mathcal{M}_F$ *induced by the functions* $\nu \mapsto \langle \nu, f \rangle$ *with* f *bounded and measurable on* $\mathcal{X}$, *to the process* $(Y_t, t \geq 0)$ *defined by*

$$Y_t = \begin{cases} \gamma \delta_x & \text{if } t = 0 \\ \bar{n}_{X_t} \delta_{X_t} & \text{if } t > 0. \end{cases}$$

Proof of Corollary 1.1 Let Γ be a measurable subset of $\mathcal{X}$. Let us prove that

$$\lim_{K \to +\infty} \mathbf{E}(\langle \nu^K_{t/Ku_K}, \mathbf{1}_\Gamma \rangle) = \mathbf{E}(\bar{n}_{X_t} \mathbf{1}_{X_t \in \Gamma}). \tag{12}$$

Fix $\varepsilon > 0$, and observe that $\bar{n}_x \in [0, \bar{b}/\underline{\alpha}]$. Write $[0, \bar{b}/\underline{\alpha}] \subset \cup_{i=1}^p I_i$, where p is the integer part of $\bar{b}/\varepsilon\underline{\alpha}$, and $I_i = [(i-1)\varepsilon, i\varepsilon[$. Define $\Gamma_i = \{x \in \mathcal{X} : \bar{n}_x \in I_i\}$ for $1 \leq i \leq p$, and apply (11) to the sets $\Gamma \cap \Gamma_1, \ldots, \Gamma \cap \Gamma_p$ with $n = 1$, $t_1 = t$ and the constant ε above. Then, for sufficiently large K,

$$\begin{aligned} \limsup_{K \to +\infty} \mathbf{E}(\langle \nu^K_{t/Ku_K}, \mathbf{1}_\Gamma \rangle) &\leq \sum_{i=1}^p \limsup_{K \to +\infty} \mathbf{E}(\langle \nu^K_{t/Ku_K}, \mathbf{1}_{\Gamma \cap \Gamma_i} \rangle) \\ &\leq \sum_{i=1}^p (i+1)\varepsilon \mathbf{P}(X_t \in \Gamma \cap \Gamma_i) \\ &\leq \sum_{i=1}^p \big(\mathbf{E}(\bar{n}_{X_t} \mathbf{1}_{X_t \in \Gamma \cap \Gamma_i}) + 2\varepsilon \mathbf{P}(X_t \in \Gamma_i) \big) \\ &\leq \mathbf{E}(\bar{n}_{X_t} \mathbf{1}_{X_t \in \Gamma}) + 2\varepsilon. \end{aligned}$$

A similar estimate for the lim inf ends the proof of (12), which implies the convergence of one-dimensional laws for the required topology.

The same method gives easily the required limit when we consider a finite number of times $t_1, \ldots, t_n$. □

Remark 1.1

- *The time scale* $1/Ku_K$ *of Theorem 1.1 is the time scale of the mutation events for the process* ν^K. *Assumption (10) is the condition leading to the time scales separation between the mutation events and the birth and death events. The limit (11) means that, when this time scales separation occurs, the population is monomorphic at any time (*i.e. *composed of individuals holding the* same *trait value) with high* *probability, and that the transition periods from a resident trait to a mutant one are infinitesimal on this mutation time scale.*

- *It is not possible to obtain the convergence in law for the Skorohod topology on $\mathbb{D}([0,T],\mathcal{M}_F)$ because of the right discontinuity of the process Y at time 0^+.*

This result is different from usual time scale separation results (averaging principle, cf. Freidlin and Wentzell [34] and Skorohod *et al.* [71]), because no assumption of ergodicity has been made, and because we have to combine two limits simultaneously. Original methods are necessary to prove Theorem 1.1.

Our proof is based on two ingredients: first, when $\mu \equiv 0$ and ν_0^K is monomorphic with trait x, we have seen in Proposition 1.1 (a) the convergence of ν^K to $n(t)\delta_x$, where $n(t)$ is solution to (5). Any solution to this equation with positive initial condition converges for large time to $\bar{n}_x$. The *large deviations* estimates for this convergence will allow us to show that the time during which the stochastic process stays in a neighborhood of its limit (problem of exit from domain, cf. [34]) is of the order of $\exp(KV)$ with $V > 0$.

Now, when u_K is small, the process ν^K with a monomorphic initial condition with trait x is near to the same process with $\mu \equiv 0$, as long as no mutation occurs. Therefore, the left inequality of (10) will allow us to prove that, with high probability, ν^K is near to $\bar{n}_x\delta_x$ when the first mutation occurs.

The second ingredient of our proof is the study of the invasion of a mutant trait y that has just appeared in a monomorphic population with trait x. This invasion can be divided in three steps:

- Firstly, as long as the mutant population size $\langle \nu_t^K, \mathbf{1}_{\{y\}} \rangle$ (initially equal to $1/K$) is smaller than a fixed small $\varepsilon > 0$, the resident dynamics is very close to what it was before the mutation, so $\langle \nu_t^K, \mathbf{1}_{\{x\}} \rangle$ stays close to $\bar{n}_x$. Then, the death rate of a mutant individual is close to the constant $d(y) + \alpha(y,x)\bar{n}_x$. Since its birth rate is constant, equal to $b(y)$, we can approximate the mutant dynamics by a binary branching process. Therefore, the probability that $\langle \nu_t^K, \mathbf{1}_{\{y\}} \rangle$ reaches ε is approximately equal to the probability that this branching process reaches εK, which converges when $K \to +\infty$ to its probability of non-extinction, $[f(y,x)]_+/b(y)$.
- Secondly, once $\langle \nu_t^K, \mathbf{1}_{\{y\}} \rangle$ has reached ε, by Proposition 1.1 (b), for large K, ν^K is close to the solution to (6) with initial state $(\bar{n}_x,\varepsilon)$ with high probability. We will show that Proposition 1.2 implies that this solution reaches the ε-neighborhood of $(0,\bar{n}_y)$ in finite time.
- Finally, once $\langle \nu_t^K, \mathbf{1}_{\{y\}} \rangle$ is close to $\bar{n}_y$ and $\langle \nu_t^K, \mathbf{1}_{\{x\}} \rangle$ is small, $K\langle \nu_t^K, \mathbf{1}_{\{x\}} \rangle$ can be approximated, in a similar way than in the first step, by a binary

branching process, which is subcritical and hence gets extinct a.s. in finite time.

Section 2 will provide the large deviations and branching process results needed to make formal the preceding heuristics. We will also prove several comparison results between $\langle \nu_t^K, \mathbf{1} \rangle$ and the birth and death processes of Definition 1.1. In section 3, the proof of Theorem 1.1 is achieved by computing, for any t, the limit law of ν^K_{t/Ku_K} according to the random number of mutations having occured between 0 and t/Ku_K.

Notations

- $\lceil a \rceil$ denotes the first integer greater or equal to a, and $\lfloor a \rfloor$ denotes the integer part of a.
- For any $K \geq 1$ and $\nu \in \mathcal{M}^K$, we will denote by $\mathbf{P}^K_\nu$ the law of the process ν^K generated by (4) with initial state ν, and by $\mathbf{E}^K_\nu$ the expectation with respect to $\mathbf{P}^K_\nu$.
- The convergence in probability of finite dimensional random variables will be denoted by $\xrightarrow{\mathcal{P}}$.
- We will denote by $\mathcal{L}(Z)$ the law of the stochastic process $(Z_t, t \geq 0)$.
- We will denote by $\preceq$ the following stochastic domination relation: if $\mathbf{Q}_1$ and $\mathbf{Q}_2$ are the laws of $\mathbb{R}$-valued processes, we will write $\mathbf{Q}_1 \preceq \mathbf{Q}_2$ if we can construct on the same filtered probability space $(\Omega, \mathcal{F}_t, \mathbf{P})$ two adapted processes X^1 and X^2 such that $\mathcal{L}(X^i) = \mathbf{Q}_i$ $(i = 1,2)$ and $\forall t \geq 0$, $\forall \omega \in \Omega$, $X^1_t(\omega) \leq X^2_t(\omega)$.
- Finally, if X^1 and X^2 are two random processes and T is a stopping time for X^1, we will write $X^1 \preceq X^2$ for $t \leq T$ if we can construct a process $\hat{X}^2$ on the probability space on which X^1 is constructed, such that $\mathcal{L}(\hat{X}^2) = \mathcal{L}(X^2)$ and $\forall t \leq T$, $\forall \omega \in \Omega$, $X^1_t(\omega) \leq \hat{X}^2_t(\omega)$.

2 Birth and death processes

We will collect in this section various results about the birth and death processes that appeared in Definition 1.1.

2.1 Comparison results

The following theorem gives various stochastic domination results.

Theorem 2.1

(a) *Assume (A). For any* $K \geq 1$ *and any* $\mathbb{L}^1$ *initial condition* ν_0^K *of the process* ν^K,

$$\mathcal{L}(\langle \nu^K, \mathbf{1} \rangle) \preceq \mathbf{P}^K(2\bar{b}, 0, \underline{\alpha}, \langle \nu_0^K, \mathbf{1} \rangle).$$

(b) *With the same assumptions than in (a), let* A_t^K *denote the number of mutations occuring in* ν^K *between times 0 and* t, *and let* $a, a_1, a_2 \geq 0$. *Then, for* $t \leq \inf\{s \geq 0 : \langle \nu_s^K, \mathbf{1} \rangle \geq a\}$,

$$A^K \preceq B^K,$$

where B^K *is a Poisson process with parameter* $K u_K a \bar{b}$.
If moreover $\nu_0^K = \langle \nu_0^K, \mathbf{1} \rangle \delta_x$, *define* $\tau_1 = \inf\{t \geq 0 : A_t^K = 1\}$ *(the first mutation time). Then, for* $t \leq \tau_1 \wedge \inf\{s \geq 0 : \langle \nu_s^K, \mathbf{1} \rangle \not\in [a_1, a_2]\}$,

$$B^K \preceq A^K \preceq C^K,$$

where B^K *and* C^K *are Poisson process with parameter, respectively,* $K u_K a_1 \mu(x) b(x)$ *and* $K u_K a_2 \mu(x) b(x)$.

(c) *Fix* $K \geq 1$ *and take* b, d, c, z *as in Definition 1.1 (a). Then, for any* $\varepsilon_1, \varepsilon_2, \varepsilon_3 \geq 0$ *and any* $\mathbb{N}/K$-*valued random variable* ε_4,

$$\mathbf{P}^K(b, d + \varepsilon_2, c + \varepsilon_3, z) \preceq \mathbf{P}^K(b + \varepsilon_1, d, c, z + \varepsilon_4).$$

(d) *Let* (Z^1, Z^2) *be a stochastic process with law*

$$\mathbf{Q}^K(b_1, b_2, d_1, d_2, c_{11}, c_{12}, c_{21}, c_{22}, z_1, z_2)$$

where the parameters are as in Definition 1.1 (b). Fix $a > 0$ *and define* $T = \inf\{t \geq 0, Z^2 \geq a\}$. *Then, there exists two processes* M^1 *and* M^2 *such that, for* $t \leq T$,

$$\begin{aligned} M^1 \preceq Z^1 &\preceq M^2, \\ \text{where} \quad \mathcal{L}(M^1) &= \mathbf{P}^K(b_1, d_1 + a c_{12}, c_{11}, z_1) \\ \text{and} \quad \mathcal{L}(M^2) &= \mathbf{P}^K(b_1, d_1, c_{11}, z_1). \end{aligned}$$

(e) *Take* (Z^1, Z^2) *as above, fix* $0 \leq a_1 < a_2$ *and* $a > 0$, *and define* $T = \inf\{t \geq 0, Z^1 \not\in [a_1, a_2]$ *or* $Z^2 \geq a\}$. *Then, there exists* M^1 *and* M^2 *such that, for* $t \leq T$,

$$\begin{aligned} M^1 \preceq Z^2 &\preceq M^2, \\ \text{where} \quad \mathcal{L}(M^1) &= \mathbf{P}^K(b_2, d_2 + a_2 c_{21} + a c_{22}, 0, z_2) \\ \text{and} \quad \mathcal{L}(M^2) &= \mathbf{P}^K(b_2, d_2 + a_1 c_{21}, 0, z_2). \end{aligned}$$

Remark 2.1 *Point (a) explains why it is necessary to combine simultaneously the limits* $K \to +\infty$ *and* $u_K \to 0$ *in order to obtain the TSS process in Theorem 1.1. The limit* $K \to +\infty$ *taken alone leads to a deterministic dynamics (cf. Fournier and Méléard [33]), so making the rare mutations limit afterwards cannot lead to a stochastic process. Conversely, taking the limit of rare mutations without making the population larger would lead to an immediate extinction of the population in the mutations time scale, because the stochastic domination of Theorem 2.1 (a) is independent of* u_K *and* $\mu(\cdot)$*, and because a process* Z *with law* $\mathbf{P}^K(2\bar{b},0,\underline{\alpha},\gamma_K/K)$ *gets a.s. extinct in finite time for any* $K \geq 1$*. This last claim can be proved as follows: firstly, 0 is accessible in finite time with a positive probability from any state* $z \in \mathbb{N}/K$*. Secondly, if at some time* t*,* $Z_t > 2\bar{b}/\underline{\alpha}$*, the death rate in* Z_t *is greater than its birth rate* $KZ_t 2\bar{b}$*, and so, as long as this inequality holds,* Z *can be dominated by a critical binary branching process with individual birth and death rates* $2\bar{b}$*. This branching process gets a.s. extinct in finite time, which implies the existence of* $s > t$ *such that* $Z_s \leq 2\bar{b}/\underline{\alpha}$*. We can then conclude to the a.s. extinction of* Z *thanks to the strong Markov property.*

Proof of (a) We will use the construction of the process ν^K given by Fournier and Méléard [33]: let $(\Omega,\mathcal{F},\mathbf{P})$ be a sufficiently large probability space, and consider on this space the following five independent random objects:

(i) a $\mathcal{M}^K$-valued random variable ν_0^K (the initial distribution),

(ii) a Poisson point measure $N_1(ds,di,dv)$ on $[0,\infty[\times\mathbb{N}\times[0,1]$ with intensity measure $q_1(ds,di,dv) = \bar{b}ds\sum_{k\geq 1}\delta_k(di)dv$ (the birth without mutation Poisson point measure),

(iii) a Poisson point measure $N_2(ds,di,dh,dv)$ on $[0,\infty[\times\mathbb{N}\times\mathbb{R}^l\times[0,1]$ with intensity measure $q_2(ds,di,dh,dv) = \bar{b}ds\sum_{k\geq 1}\delta_k(di)m(h)dhdv$ (the birth with mutation Poisson point measure),

(iv) a Poisson point measure $N_3(ds,di,dv)$ on $[0,\infty[\times\mathbb{N}\times[0,1]$ with intensity measure $q_3(ds,di,dv) = \bar{d}ds\sum_{k\geq 1}\delta_k(di)dv$ (the natural death Poisson point measure),

(v) a Poisson point measure $N_4(ds,di,dj,dv)$ on $[0,\infty[\times\mathbb{N}\times\mathbb{N}\times[0,1]$ with intensity measure $q_4(ds,di,dj,dv) = (\bar{\alpha}/K)ds\sum_{k\geq 1}\delta_k(di)\sum_{m\geq 1}\delta_m(dj)dv$ (the competition death Poisson point measure).

We will also need the following function, solving the purely notational problem of associating a number to each individual in the population: for any

$K \geq 1$, let $H = (H^1,\dots,H^k,\dots)$ be the map from $\mathcal{M}^K$ into $(\mathbb{R}^l)^{\mathbb{N}}$ defined by

$$H\left(\frac{1}{K}\sum_{i=1}^{n}\delta_{x_i}\right) = (x_{\sigma(1)},\dots,x_{\sigma(n)},0,\dots,0,\dots),$$

where $x_{\sigma(1)} \preccurlyeq \dots \preccurlyeq x_{\sigma(n)}$ for the lexicographic order $\preccurlyeq$ on $\mathbb{R}^d$. For convenience, we have omitted in our notation the dependence of H and H^i on K.

Then a process ν^K with generator L^K and initial state ν^K_0 can be constructed as follows: for any $t \geq 0$,

$$\begin{aligned}
\nu^K_t = \nu^K_0 &+ \int_0^T\int_{\mathbb{N}}\int_0^1 \mathbf{1}_{\{i\leq K\langle \nu^K_{s-},\mathbf{1}\rangle\}}\frac{\delta_{H^i(\nu^K_{s-})}}{K} \\
&\qquad \mathbf{1}_{\left\{v\leq \frac{[1-u_K\mu(H^i(\nu^K_{s-}))]b(H^i(\nu^K_{s-}))}{\bar{b}}\right\}} N_1(ds,di,dv) \\
&+ \int_0^T\int_{\mathbb{N}}\int_{\mathbb{R}^l}\int_0^1 \mathbf{1}_{\{i\leq K\langle \nu^K_{s-},\mathbf{1}\rangle\}}\frac{\delta_{H^i(\nu^K_{s-})+h}}{K} \\
&\qquad \mathbf{1}_{\left\{v\leq \frac{u_K\mu(H^i(\nu^K_{s-}))b(H^i(\nu^K_{s-}))}{\bar{b}}\frac{m(H^i(\nu^K_{s-}),h)}{m(h)}\right\}} N_2(ds,di,dh,dv) \\
&- \int_0^T\int_{\mathbb{N}}\int_0^1 \mathbf{1}_{\{i\leq K\langle \nu^K_{s-},\mathbf{1}\rangle\}}\frac{\delta_{H^i(\nu^K_{s-})}}{K}\mathbf{1}_{\left\{v\leq \frac{d(H^i(\nu^K_{s-}))}{\bar{d}}\right\}} N_3(ds,di,dv) \\
&- \int_0^T\int_{\mathbb{N}}\int_{\mathbb{N}}\int_0^1 \mathbf{1}_{\{i\leq K\langle \nu^K_{s-},\mathbf{1}\rangle\}}\mathbf{1}_{\{j\leq K\langle \nu^K_{s-},\mathbf{1}\rangle\}}\frac{\delta_{H^i(\nu^K_{s-})}}{K} \\
&\qquad \mathbf{1}_{\left\{v\leq \frac{\alpha(H^i(\nu^K_{s-}),H^j(\nu^K_{s-}))}{\bar{\alpha}}\right\}} N_4(ds,di,dj,dv).
\end{aligned} \tag{13}$$

Although this formula is quite complicated, the principle is very simple: for each type of event, the corresponding Poisson point process jumps faster than ν^K has to. We decide whether a jump of the process ν^K occurs by comparing v to a quantity related to the rates of the various events. The indicator functions involving i and j simply ensures that the i^{th} and j^{th} individuals are alive in the population (since $K\langle\nu^K_t,\mathbf{1}\rangle$ is the number of individuals in the population at time t).

Under (A1), (A2) and the assumption that $\mathbf{E}(\langle\nu^K_0,\mathbf{1}\rangle) < \infty$, Fournier and Méléard [33] prove the existence and uniqueness of the solution to (13), and that this solution is a Markov process with infinitesimal generator (4).

Then, the $\mathbb{N}/K$-valued Markov process Z^K defined by

$$Z_t^K = \langle \nu_0^K, \mathbf{1} \rangle + \frac{1}{K} \int_0^t \int_{\mathbb{N}} \mathbf{1}_{\{i \leq K Z_{s-}^K\}} \bigg(\int_0^1 N_1(ds,di,dv) \\ + \int_{\mathbb{R}^l} \int_0^1 N_2(ds,di,dh,dv) - \int_{\mathbb{N}} \int_0^1 \mathbf{1}_{\{j \leq K Z_{s-}^K,\ v \leq \underline{\alpha}/\bar{\alpha}\}} N_4(ds,di,dj,dv) \bigg) \tag{14}$$

can be easily proved to satisfy $\mathcal{L}(Z^K) = \mathbf{P}^K(2\bar{b},0,\underline{\alpha},\langle \nu_0^K,\mathbf{1}\rangle)$. Moreover, if for some $\omega \in \Omega$, and for some $t \geq 0$, $Z_t^K(\omega) = \langle \nu_t^K(\omega),\mathbf{1}\rangle$, let

$$T^K = \inf\{s \geq t, Z_s^K(\omega) \neq \langle \nu_s^K(\omega),\mathbf{1}\rangle\}.$$

Then, the comparison of (13) and (14) yields that, on the time interval $[t,T^K]$, any birth time (with or without mutation) for ν^K is also a birth time for Z^K, and any death time for Z^K is also a death time for ν^K. Consequently, $Z_{T^K}^K(\omega)$ is necessarily greater than $\langle \nu_{T^K}^K(\omega),\mathbf{1}\rangle$, which implies the required domination result. ☐

Proof of (b) With the same notations than above,

$$A_t^K := \int_0^t \int_{\mathbb{N}} \int_{\mathbb{R}^l} \int_0^1 \mathbf{1}_{\{i \leq K\langle \nu_{s-}^K,\mathbf{1}\rangle\}} \times \\ \times \mathbf{1}_{\left\{v \leq \frac{u_K \mu(H^i(\nu_{s-}^K)) b(H^i(\nu_{s-}^K))}{\bar{b}} \frac{m(H^i(\nu_{s-}^K),h)}{m(h)}\right\}} N_2(ds,di,dh,dv).$$

Therefore, for $t \leq \inf\{s \geq 0 : \langle \nu_s^K,\mathbf{1}\rangle \geq a\}$,

$$A_t^K \leq \int_0^t \int_{\mathbb{N}} \int_{\mathbb{R}^l} \int_0^1 \mathbf{1}_{\{i \leq Ka\}} \mathbf{1}_{\{v \leq u_K\}} N_2(ds,di,dh,dv). \tag{15}$$

Since the intensity measure of N_2 is

$$q_2(ds,di,dh,dv) = \bar{b} ds \sum_{k \geq 1} \delta_k(di) m(h) dh dv, \tag{16}$$

the right-hand side of (15) is a Poisson process with parameter $K u_K a \bar{b}$.

In the case where $\nu_0^K = \langle \nu_0^K,\mathbf{1}\rangle \delta_x$, as long as $t < \tau_1$, $\nu_t^K = \langle \nu_t^K,\mathbf{1}\rangle \delta_x$, therefore, for $t \leq \tau_1 \wedge \inf\{s \geq 0 : \langle \nu_s^K,\mathbf{1}\rangle \notin [a_1,a_2]\}$,

$$\int_0^t \int_{\mathbb{N}} \int_{\mathbb{R}^l} \int_0^1 \mathbf{1}_{\{i \leq K a_1\}} \mathbf{1}_{\left\{v \leq \frac{u_K \mu(x) b(x)}{\bar{b}} \frac{m(x,h)}{m(h)}\right\}} N_2(ds,di,dh,dv) \leq A_t^K \\ \leq \int_0^t \int_{\mathbb{N}} \int_{\mathbb{R}^l} \int_0^1 \mathbf{1}_{\{i \leq K a_2\}} \mathbf{1}_{\left\{v \leq \frac{u_K \mu(x) b(x)}{\bar{b}} \frac{m(x,h)}{m(h)}\right\}} N_2(ds,di,dh,dv).$$

By (16), the left-hand side and the right-hand side of this inequality is are Poisson processes with parameters $K u_K a_1 \mu(x) b(x)$ and $K u_K a_2 \mu(x) b(x)$, respectively. ☐

Proof of (c) Consider, on a probability space $(\Omega,\mathcal{F},\mathbf{P})$ the random variables z and ε_4 as in the statement of Theorem 2.1, and the following independent random objects:

(i) a Poisson point measure $\hat{N}_1(ds,di,dv)$ on $[0,\infty[\times\mathbb{N}\times[0,1]$ with intensity measure $q_1(ds,di,dv)=(b+\varepsilon_1)ds\sum_{k\geq 1}\delta_k(di)dv$,

(ii) a Poisson point measure $\hat{N}_2(ds,di,dv)$ on $[0,\infty[\times\mathbb{N}\times[0,1]$ with intensity measure $q_2(ds,di,dv)=(d+\varepsilon_2)ds\sum_{k\geq 1}\delta_k(di)dv$,

(iii) a Poisson point measure $\hat{N}_3(ds,di,dj,dv)$ on $[0,\infty[\times\mathbb{N}\times\mathbb{N}\times[0,1]$ with intensity $q_3(ds,di,dj,dv)=(c+\varepsilon_3)/K ds\sum_{k\geq 1}\delta_k(di)\sum_{m\geq 1}\delta_m(dj)dv$.

Then, define the processes Z^1 and Z^2 on Ω by, for any $t\geq 0$,

$$Z^1_t = z+\frac{1}{K}\int_0^t\int_{\mathbb{N}}\mathbf{1}_{\{i\leq KZ^1_{s-}\}}\biggl(\int_0^1\mathbf{1}_{\left\{v\leq\frac{b}{b+\varepsilon_1}\right\}}\hat{N}_1(ds,di,dv)\\ -\int_0^1\hat{N}_2(ds,di,dh,dv)-\int_{\mathbb{N}}\int_0^1\mathbf{1}_{\{j\leq KZ^1_{s-}\}}\hat{N}_3(ds,di,dj,dv)\biggr)$$

and

$$Z^2_t = z+\varepsilon_4+\frac{1}{K}\int_0^t\int_{\mathbb{N}}\mathbf{1}_{\{i\leq KZ^2_{s-}\}}\biggl(\int_0^1\hat{N}_1(ds,di,dv)\\ -\int_0^1\mathbf{1}_{\left\{v\leq\frac{d}{d+\varepsilon_2}\right\}}\hat{N}_2(ds,di,dh,dv)-\int_{\mathbb{N}}\int_0^1\mathbf{1}_{\left\{j\leq KZ^2_{s-},\ v\leq\frac{c}{c+\varepsilon_3}\right\}}\hat{N}_3(ds,di,dj,dv)\biggr).$$

Then, $\mathcal{L}(Z^1)=\mathbf{P}^K(b,d+\varepsilon_2,c+\varepsilon_3,z)$ and $\mathcal{L}(Z^2)=\mathbf{P}^K(b+\varepsilon_1,d,c,z+\varepsilon_4)$, and a comparison between the birth and death events of Z^1 and Z^2 in a similar way than in the proof of (a) proves that $Z^1_t(\omega)\leq Z^2_t(\omega)$ for any $t\geq 0$ and $\omega\in\Omega$. □

Proof of (d) Consider, on a sufficiently rich probability space $(\Omega,\mathcal{F},\mathbf{P})$ the random variables z_1 and z_2 as in the statement of Theorem 2.1, and the following independent random objects:

(i) two Poisson point measures $\hat{N}^k_1(ds,di,dv)$ $(k=1,2)$ on $[0,\infty[\times\mathbb{N}\times[0,1]$ with intensity measures $q^k_1(ds,di,dv)=b_k ds\sum_{n\geq 1}\delta_n(di)dv$ $(k=1,2)$,

(ii) two Poisson point measures $\hat{N}^k_2(ds,di,dv)$ $(k=1,2)$ on $[0,\infty[\times\mathbb{N}\times[0,1]$ with intensity measures $q^k_2(ds,di,dv)=d_k ds\sum_{n\geq 1}\delta_n(di)dv$ $(k=1,2)$,

(iii) two Poisson point measures $\hat{N}^k_3(ds,di,dj,dv)$ $(k=1,2)$ on $[0,\infty[\times\mathbb{N}\times\mathbb{N}\times[0,1]$ with intensity measures

$$q^k_3(ds,di,dj,dv)=(c_{k1}/K)ds\sum_{n\geq 1}\delta_n(di)\sum_{m\geq 1}\delta_m(dj)dv\quad(k=1,2).$$

(iv) two Poisson point measures $\hat{N}^k_4(ds,di,dj,dv)$ $(k=1,2)$ on $[0,\infty[\times\mathbb{N}\times\mathbb{N}\times[0,1]$ with intensity measures

$$q^k_4(ds,di,dj,dv)=(c_{k2}/K)ds\sum_{n\geq 1}\delta_n(di)\sum_{m\geq 1}\delta_m(dj)dv \quad (k=1,2).$$

The processes Z^1 and Z^2 can be constructed on Ω as follows: for any $t\geq 0$, and for $k=1,2$,

$$\begin{aligned}Z^k_t=z_k+\frac{1}{K}\int_0^t\int_{\mathbb{N}}\mathbf{1}_{\{i\leq KZ^k_{s-}\}}\biggl(\int_0^1\hat{N}^k_1(ds,di,dv)-\int_0^1\hat{N}^k_2(ds,di,dh,dv)\\-\int_{\mathbb{N}}\int_0^1\Bigl(\mathbf{1}_{\{j\leq KZ^1_{s-}\}}\hat{N}^k_3(ds,di,dj,dv)+\mathbf{1}_{\{j\leq KZ^2_{s-}\}}\hat{N}^k_4(ds,di,dj,dv)\Bigr)\biggr).\end{aligned}$$

Then, we can define on Ω the processes M^1 and M^2 by, for any $t\geq 0$,

$$\begin{aligned}M^1_t=z_1+\frac{1}{K}\int_0^t\int_{\mathbb{N}}\mathbf{1}_{\{i\leq KM^1_{s-}\}}\biggl(\int_0^1\hat{N}^1_1(ds,di,dv)-\int_0^1\hat{N}^1_2(ds,di,dh,dv)\\-\int_{\mathbb{N}}\int_0^1\Bigl(\mathbf{1}_{\{j\leq KM^1_{s-}\}}\hat{N}^1_3(ds,di,dj,dv)+\mathbf{1}_{\{j\leq Ka\}}\hat{N}^1_4(ds,di,dj,dv)\Bigr)\biggr)\end{aligned}$$

and

$$\begin{aligned}M^2_t=z_1+\frac{1}{K}\int_0^t\int_{\mathbb{N}}\mathbf{1}_{\{i\leq KM^2_{s-}\}}\biggl(\int_0^1\hat{N}^1_1(ds,di,dv)-\int_0^1\hat{N}^1_2(ds,di,dh,dv)\\-\int_{\mathbb{N}}\int_0^1\mathbf{1}_{\{j\leq KM^2_{s-}\}}\hat{N}^1_3(ds,di,dj,dv)\biggr),\end{aligned}$$

and a comparison between the birth and death events of M^1, M^2 and Z^1 in a similar way than in the proof of (a) shows the required domination relation. □

Proof of (e) Here, the processes M^1 and M^2 can be defined, for any $t\geq 0$, by

$$\begin{aligned}M^1_t=z_2+\frac{1}{K}\int_0^t\int_{\mathbb{N}}\mathbf{1}_{\{i\leq KM^1_{s-}\}}\biggl(\int_0^1\hat{N}^2_1(ds,di,dv)-\int_0^1\hat{N}^2_2(ds,di,dh,dv)\\-\int_{\mathbb{N}}\int_0^1\Bigl(\mathbf{1}_{\{j\leq Ka_2\}}\hat{N}^2_3(ds,di,dj,dv)+\mathbf{1}_{\{j\leq Ka\}}\hat{N}^2_4(ds,di,dj,dv)\Bigr)\biggr)\end{aligned}$$

and

$$M_t^2 = z_2 + \frac{1}{K}\int_0^t \int_{\mathbb{N}} \mathbf{1}_{\{i \leq K M_{s-}^2\}} \left(\int_0^1 \hat{N}_1^2(ds,di,dv) - \int_0^1 \hat{N}_2^2(ds,di,dh,dv) \right.$$
$$\left. - \int_{\mathbb{N}} \int_0^1 \mathbf{1}_{\{j \leq K a_1\}} \hat{N}_3^2(ds,di,dj,dv) \right).$$

□

2.2 Problem of exit from a domain

Let us give some results on $\mathbf{P}^K(b,d,c,z)$ when $c > 0$. Points (a) and (b) of the following theorem strengthen Proposition 1.1, and point (c) studies the problem of exit from a domain.

Theorem 2.2

(a) *Let* $c,T > 0$ *and* $b,d \geq 0$, *let* C *be a compact subset of* $\mathbb{R}_+^*$, *and write* $\mathbf{P}_z^K = \mathbf{P}^K(b,d,c,z)$ *for* $z \in \mathbb{N}/K$. *Let* ϕ_z *denote the solution to*

$$\dot{\phi} = (b - d - c\phi)\phi \tag{17}$$

with initial condition $\phi_z(0) = z$. *Then*

$$r := \inf_{z \in C} \inf_{0 \leq t \leq T} |\phi_z(t)| > 0 \tag{18}$$

and

$$R := \sup_{z \in C} \sup_{0 \leq t \leq T} |\phi_z(t)| < +\infty. \tag{19}$$

Then, for any $\delta < r$, *(with the convention* $\sup \emptyset = 0$*)*

$$\lim_{K \to +\infty} \sup_{z \in C} \mathbf{P}_z^K \left(\sup_{0 \leq t \leq T} |w_t - \phi_z(t)| \geq \delta \right) = 0, \tag{20}$$

where w_t *is the canonical process on* $\mathbb{D}(\mathbb{R}_+,\mathbb{R})$.

(b) *Let* $T,c_{ij} > 0$ *and* $b_i,d_i \geq 0$ *(*$i,j \in \{1,2\}$*), let* C *be a compact subset of* $(\mathbb{R}_+^*)^2$, *and write* $\mathbf{Q}_{z_1,z_2}^K = \mathbf{Q}^K(b_1,b_2,d_1,d_2,c_{11},c_{12},c_{21},c_{22},z_1,z_2)$ *for* z_1 *and* z_2 *in* $\mathbb{N}/K$. *Let* $\phi_{z_1,z_2} = (\phi_{z_1,z_2}^1,\phi_{z_1,z_2}^2)$ *denote the solution to*

$$\dot{\phi}^1 = (b_1 - d_1 - c_{11}\phi^1 - c_{12}\phi^2)\phi^1$$
$$\dot{\phi}^2 = (b_2 - d_2 - c_{21}\phi^1 - c_{22}\phi^2)\phi^2$$

with initial conditions $\phi_{z_1,z_2}^1(0) = z_1$ *and* $\phi_{z_1,z_2}^2(0) = z_2$. *Then*

$$r := \inf_{z \in C} \inf_{0 \leq t \leq T} \|\phi_{z_1,z_2}(t)\| > 0 \tag{21}$$

and

$$\sup_{z\in C}\sup_{0\leq t\leq T}\|\phi_{z_1,z_2}(t)\|<+\infty.$$

Then, for any $\delta<r$,

$$\lim_{K\to+\infty}\sup_{z\in C}\mathbf{Q}^K_{z_1,z_2}(\sup_{0\leq t\leq T}\|\hat{w}_t-\phi_{z_1,z_2}(t)\|\geq\delta)=0,$$

where $\hat{w}_t=(\hat{w}^1_t,\hat{w}^2_t)$ *is the canonical process on* $\mathbb{D}(\mathbb{R}_+,\mathbb{R}^2)$.

(c) *Let* $b,c>0$ *and* $0\leq d<b$. *Observe that* $(b-d)/c$ *is the unique stable steady state of (17). Fix* $0<\eta_1<(b-d)/c$ *and* $\eta_2>0$, *and define on* $\mathbb{D}(\mathbb{R}_+,\mathbb{R})$

$$T^K=\inf\left\{t\geq 0: w_t\notin\left[\frac{b-d}{c}-\eta_1,\frac{b-d}{c}+\eta_2\right]\right\}.$$

Then, there exists $V>0$ *such that, for any compact subset* C *of* $](b-d)/c-\eta_1,(b-d)/c+\eta_2[$,

$$\lim_{K\to+\infty}\sup_{z\in C}\mathbf{P}^K_z(T^K<e^{KV})=0. \tag{22}$$

Proof of (a) and (b) Observe that any solution to (17) with positive initial condition is bounded, since $\dot{\phi}<0$ as soon as $\phi>(b-d)/c$. This implies easily (19).

Moreover, since a solution to (17) can be written as

$$\phi(t)=\phi(0)\exp\left(\int_0^t(b-d-c\phi(s))ds\right),$$

it cannot reach 0 in finite time if $\phi(0)>0$. Therefore, (18) follows from the continuity of the flow, which is a classical consequence of the fact that $z\mapsto(b-d-cz)z$ is locally Lipschitz and of Gronwall's Lemma (cf. e.g. Queffélec and Zuily [64] p. 356).

Finally, (20) is a consequence of large deviations estimates for the sequence of laws $(\mathbf{P}^K_z)_{K\geq1}$. As can be seen in Theorem 10.2.6 in chapter 10 of Dupuis and Ellis [24], a large deviations principle on $[0,T]$ with a good rate function I_T holds for $\mathbb{Z}/K$-valued Markov jump processes with transition rates

$$\begin{array}{ll} Kp(i/K) & \text{from } i/K \text{ to } (i+1)/K,\\ Kq(i/K) & \text{from } i/K \text{ to } (i-1)/K,\end{array}$$

where p and q are functions defined on $\mathbb{R}$ and with positive values, bounded, Lipschitz and uniformly bounded away from 0. The rate function I_T writes

$$I_T(\phi) = \begin{cases} \displaystyle\int_0^T L(\phi(t),\dot{\phi}(t))dt & \text{if } \phi \text{ is absol. cont. on } [0,T] \\ +\infty & \text{otherwise,} \end{cases} \tag{23}$$

where $L(y,z) = 0$ if $z = p(y) - q(y)$ and $L(y,z) > 0$ otherwise. Therefore, $I_T(\phi) = 0$ if and only if ϕ is absolutely continuous and

$$\dot{\phi} = p(\phi) - q(\phi). \tag{24}$$

Moreover, this large deviation is *uniform* with respect to the initial condition. This means that, if $\mathbf{R}_z^K$ denotes the law of this process with initial condition z, for any compact set $C \subset \mathbb{R}$, for any closed set F and any open set G of $\mathbb{D}([0,T],\mathbb{R})$,

$$\liminf_{K\to+\infty} \frac{1}{K} \log \inf_{z\in C} \mathbf{R}_z^K(G) \geq -\sup_{z\in C} \inf_{\psi\in G,\ \psi(0)=z} I_T(\psi) \tag{25}$$

$$\text{and} \quad \limsup_{K\to+\infty} \frac{1}{K} \log \sup_{z\in C} \mathbf{R}_z^K(F) \leq -\inf_{\psi\in F,\ \psi(0)\in C} I_T(\psi). \tag{26}$$

Our birth and death process does not satisfy these asumptions. However, if we define

$$p(z) = b\chi(z) \quad \text{and} \quad q(z) = d\chi(z) + c\chi(z)^2,$$

where $\quad \chi(z) = z$ if $z \in [r-\delta,R+\delta]$; $r-\delta$ if $z < r-\delta$; $R+\delta$ if $z > R+\delta$,

then $\mathbf{R}_z^K = \mathbf{P}_z^K$ on the time interval $[0,\tau]$, where $\tau = \inf\{t \geq 0, w_t \notin [r-\delta,R+\delta]\}$, and p and q satisfy the assumptions above. Therefore, by (26),

$$\limsup_{K\to+\infty} \frac{1}{K} \log \sup_{z\in C} \mathbf{P}_z^K \left(\sup_{0\leq t\leq T} |w_t - \phi_z(t)| \geq \delta \right) \leq -\inf_{\psi\in F^\delta} I_T(\psi), \quad \text{where}$$

$$F^\delta := \left\{ \psi \in \mathbb{D}([0,T],\mathbb{R}) : \psi(0) \in C \text{ and } \exists t \in [0,T],\ |\psi(t) - \phi_{\psi(0)}(t)| \geq \delta \right\}$$

By the continuity of the flow of (24), the set F^δ is closed. Since I_T is a good rate function, the infimum of I_T over this set is attained at some function belonging to F^δ, which cannot be a solution to (24), and thus non-zero. This ends the proof of (20).

The proof of (b) can be made in a very similar way, using the large deviations estimates for two-dimensional jump processes of Theorem 10.2.6 in chapter 10 of Dupuis and Ellis [24]. □

Proof of (c) Define the function χ on $\mathbb{R}$ by

$$\chi(z) = \begin{cases} z & \text{if } z \in [(b-d)/c - \eta_1,(b-d)/c + \eta_2] \\ (b-d)/C - \eta_1 & \text{if } z < (b-d)/c - \eta_1 \\ (b-d)/C + \eta_2 & \text{if } z > (b-d)/c - \eta_2. \end{cases}$$

As in the proof of (a), we can construct from the functions $p(z) = b\chi(z)$ and $q(z) = d\chi(z) + c\chi(z)^2$ a family of laws $(\mathbf{R}^K_z)$ such that $\mathbf{R}^K_z = \mathbf{P}^K_z$ on the time interval $[0,T^K]$, and such that (25) and (26) hold for the good rate function I_T defined in (23).

Observe that any solution to (24) are monotonic and converge to $(b-d)/c$ when $t \to +\infty$. Therefore, the following estimates for the time of exit from an attracting domain are classical (cf. Freidlin and Wentzell [34], chapter 5, section 4): there exists $\bar{V} \geq 0$ such that, for any $\delta > 0$,

$$\lim_{K\to+\infty} \inf_{z\in C} \mathbf{R}^K_z \left(e^{K(\bar{V}-\delta)} < T^K < e^{K(\bar{V}+\delta)} \right) = 1,$$

which implies (22) if we can prove that $\bar{V} > 0$.

The constant $\bar{V}$ is obtained as follows (see [34] pp. 108–109): for any $y,z \in \mathbb{R}$, define

$$V(y,z) := \inf_{t>0,\ \varphi(0)=y,\ \varphi(t)=z} I_t(\varphi).$$

Then

$$\bar{V} := V\left(\frac{b-d}{c},\frac{b-d}{c} - \eta_1\right) \wedge V\left(\frac{b-d}{c},\frac{b-d}{c} + \eta_2\right).$$

Now, Theorem 5.4.3. of [34] states that, for any $y,z \in \mathbb{R}$, the infimum defining $V(y,z)$ is attained at some function ϕ linking y to z, in the sense that, either there exists an absolutely continuous function ϕ defined on $[0,T]$ for some $T > 0$ such that $\phi(0) = y$, $\phi(T) = z$ and $V(y,z) = I_T(\phi) = \int_0^T L(\phi(t),\dot{\phi}(t))dt$, or there exists an absolutely continuous function ϕ defined on $]-\infty,T]$ for some $T > -\infty$ such that $\lim_{t\to-\infty} \phi(t) = y$, $\phi(T) = z$ and $V(y,z) = \int_{-\infty}^T L(\phi(t),\dot{\phi}(t))dt$.

Since any solution to (24) is decreasing on $[(b-d)/c, +\infty[$, a function ϕ defined on $[0,T]$ or $]-\infty,T]$ linking $(b-d)/c$ to $(b-d)/c + \eta_2$ cannot be a solution to (24), and thus $V((b-d)/c,(b-d)/c + \eta_2) > 0$. Similarly, $V((b-d)/c,(b-d)/c - \eta_1) > 0$, and so $\bar{V} > 0$, which concludes the proof of Theorem 2.2. □

2.3 Some results on branching processes

Observe that, when $c = 0$, $\mathbf{P}^K(b,d,0,z)$ is the law of a binary branching process divided by K. Let us give some results on these processes.

Theorem 2.3 *Let* $b,d > 0$. *As in Theorem 2.2, define, for any* $K \geq 1$ *and any* $z \in \mathbb{N}/K$, $\mathbf{P}^K_z = \mathbf{P}^K(b,d,0,z)$. *Define also, for any* $z \in \mathbb{R}$, *on* $\mathbb{D}(\mathbb{R}_+,\mathbb{R})$, *the stopping time*

$$T_z = \inf\{t \geq 0 : w_t = z\}.$$

Finally, let $(t_K)_{K\geq 1}$ *be a sequence of positive numbers such that* $\log K \ll t_K$.
(a) *If* $b < d$ *(sub-critical case), for any* $\varepsilon > 0$,

$$\lim_{K\to+\infty} \mathbf{P}^K_{1/K}(T_0 \leq t_K \wedge T_{\lceil \varepsilon K\rceil/K}) = 1, \tag{27}$$

$$\text{and} \quad \lim_{K\to+\infty} \mathbf{P}^K_{\lfloor \varepsilon K\rfloor/K}(T_0 \leq t_K) = 1. \tag{28}$$

Moreover, for any $K \geq 1$, $k \geq 1$ *and* $n \geq 1$,

$$\mathbf{P}^K_{n/K}(T_{kn/K} \leq T_0) \leq \frac{1}{k}. \tag{29}$$

(b) *If* $b > d$ *(super-critical case), for any* $\varepsilon > 0$,

$$\lim_{K\to+\infty} \mathbf{P}^K_{1/K}(T_0 \leq t_K \wedge T_{\lceil \varepsilon K\rceil/K}) = \frac{d}{b} \tag{30}$$

$$\text{and} \quad \lim_{K\to+\infty} \mathbf{P}^K_{1/K}(T_{\lceil \varepsilon K\rceil/K} \leq t_K) = 1 - \frac{d}{b}. \tag{31}$$

Proof Let us denote by $\mathbf{Q}_n$ the law of the binary branching process with initial state $n \in \mathbb{N}$, with individual birth rate b and individual death rate d. Then (27), (28), (29), (30) and (31) rewrite respectively

$$\lim_{K\to+\infty} \mathbf{Q}_1(T_0 \leq t_K \wedge T_{\lceil \varepsilon K\rceil}) = 1, \tag{32}$$

$$\lim_{K\to+\infty} \mathbf{Q}_{\lfloor \varepsilon K\rfloor}(T_0 \leq t_K) = 1, \tag{33}$$

$$\mathbf{Q}_n(T_{kn} \leq T_0) \leq \frac{1}{k}, \tag{34}$$

$$\lim_{K\to+\infty} \mathbf{Q}_1(T_0 \leq t_K \wedge T_{\lceil \varepsilon K\rceil}) = \frac{d}{b} \tag{35}$$

$$\text{and} \quad \lim_{K\to+\infty} \mathbf{Q}_1(T_{\lceil \varepsilon K\rceil} \leq t_K) = 1 - \frac{d}{b}. \tag{36}$$

The limit (33) follows easily from the distribution of the extinction time for binary branching processes when $b \neq d$ (cf. Athreya and Ney [3] p. 109): for any $t \geq 0$ and $n \in \mathbb{N}$,

$$\mathbf{Q}_n(T_0 \leq t) = \left(\frac{d\left(1 - e^{-(b-d)t}\right)}{b - de^{-(b-d)t}}\right)^n. \tag{37}$$

It is known that there is no accumulation of jumps for branching processes. Therefore, under $\mathbf{Q}_1$, when $K \to +\infty$, $T_{\lceil \varepsilon K \rceil} \to +\infty$ a.s. , and thus $\mathbf{Q}_1(T_0 \leq T_{\lceil \varepsilon K \rceil}, T_0 < \infty) \to \mathbf{Q}_1(T_0 < \infty)$. Therefore, (32) and (35) follow easily from (37).

The inequality (34) follows from the fact that, if $(Z_t,t \geq 0)$ is a process with law $\mathbf{Q}_n$, $(Z_t \exp(-(b-d)t),t \geq 0)$ is a martingale (cf. [3] p. 111). Then, Doob's stopping theorem applied to the stopping time $T_0 \wedge T_{kn}$ yields,

$$\mathbf{E}_n(kne^{(d-b)T_{kn}}\mathbf{1}_{\{T_{kn}<T_0\}}) = n,$$

where $\mathbf{E}_n$ is the expectation with respect to $\mathbf{Q}_n$. Therefore, when $b < d$, $kn\mathbf{Q}_n(T_{kn} < T_0) \leq n$, and the proof of (34) is completed.

The limit (36) follows from the fact that, if $(Z_t,t \geq 0)$ is a process with law $\mathbf{Q}_1$, the martingale $(Z_t \exp(-(b-d)t),t \geq 0)$ converges a.s. when $t \to +\infty$ to a random variable W, where $W = 0$ on the event $\{T_0 < \infty\}$ and $W > 0$ on the event $\{T_0 = \infty\}$ (cf. [3] p. 112). Hence, on the event $\{T_0 = \infty\}$, when $b > d$,

$$\log Z_t/t \to b - d > 0,$$

and so, for almost any $\omega \in \{T_0 = \infty\}$, there exists $S(\omega) < \infty$ such that for any $t \geq S(\omega)$,

$$Z_t \geq \exp((b-d)t/2).$$

Therefore, since $\log K \ll t_K$, for any $\varepsilon > 0$, $\mathbf{Q}_1(T_0 = \infty, T_{\lceil \varepsilon K \rceil} \geq t_K) \to 0$ when $K \to +\infty$. Then, (36) follows from the fact that $\mathbf{Q}_1(T_0 = \infty) = 1-d/b$, which is an immediate consequence of (37). □

3 Proof of Theorem 1.1

Let us assume that ν^K is constructed by (13) on a sufficiently rich probability space $(\Omega,\mathcal{F},\mathbf{P})$.

Let us introduce the following sequences of stopping times: for all $n \geq 1$, let τ_n be the first mutation time after time τ_{n-1}, with $\tau_0 = 0$ (*i.e.* τ_n is the n^{th} mutation time), and for any $n \geq 0$, let θ_n be the first time after τ_n when the population gets monomorphic. Observe that $\theta_0 = 0$ if the initial population is monomorphic. For any $n \geq 1$, define the random variable U_n as the new trait value appearing at the mutation time τ_n, and, when $\theta_n < \infty$, define V_n by $\mathrm{Supp}(\nu^K_{\theta_n}) = \{V_n\}$. When $\theta_n = +\infty$, define $V_n = +\infty$.

Our proof of Theorem 1.1 is based on the following two lemmas. The first lemma proves that there is no accumulation of mutations on the time scale of Theorem 1.1, and studies the asymptotic behavior of τ_1 starting from a monomorphic population, when $K \to +\infty$.

Lemma 3.1
(a) *Assume that the initial condition of* ν^K *satisfies* $\sup_K \mathbf{E}(\langle \nu_0^K, \mathbf{1}\rangle) < +\infty$. *Then, for any* $\eta > 0$, *there exists* $\varepsilon > 0$ *such that, for any* $t > 0$,

$$\limsup_{K\to+\infty} \mathbf{P}^K_{\nu_0^K}\left(\exists n \geq 0 : \frac{t}{Ku_K} \leq \tau_n \leq \frac{t+\varepsilon}{Ku_K}\right) < \eta. \tag{38}$$

Let $x \in \mathcal{X}$ *and let* $(z_K)_{K\geq 1}$ *be a sequence of integers such that* $z_K/K \to z > 0$.
(b) *For any* $\varepsilon > 0$,

$$\lim_{K\to+\infty} \mathbf{P}^K_{\frac{z_K}{K}\delta_x}\left(\tau_1 > \log K, \sup_{t\in[\log K,\tau_1]} |\langle \nu_t^K, \mathbf{1}\rangle - \bar{n}_x| > \varepsilon\right) = 0. \tag{39}$$

Observe that, by (a) with $t = 0$, *since* $\log K \ll 1/Ku_K$,

$$\lim_{K\to+\infty} \mathbf{P}^K_{\frac{z_K}{K}\delta_x}(\tau_1 < \log K) = 0.$$

In particular, under $\mathbf{P}^K_{\frac{z_K}{K}\delta_x}$, $\nu^K_{\log K} \xrightarrow{\mathcal{P}} \bar{n}_x\delta_x$ *and* $\nu^K_{\tau_1-} \xrightarrow{\mathcal{P}} \bar{n}_x\delta_x$.
If, moreover, $z = \bar{n}_x$, *then, for any* $\varepsilon > 0$,

$$\lim_{K\to+\infty} \mathbf{P}^K_{\frac{z_K}{K}\delta_x}\left(\sup_{t\in[0,\tau_1]} |\langle \nu_t^K, \mathbf{1}\rangle - \bar{n}_x| > \varepsilon\right) = 0. \tag{40}$$

(c) *For any* $t > 0$,

$$\lim_{K\to+\infty} \mathbf{P}^K_{\frac{z_K}{K}\delta_x}\left(\tau_1 > \frac{t}{Ku_K}\right) = \exp(-\beta(x)t),$$

where $\beta(\cdot)$ *has been defined in (2).*

The second lemma studies the asymptotic behavior of θ_0 and V_0 starting from a dimorphic population, when $K \to +\infty$.

Lemma 3.2 *Fix* $x,y \in \mathcal{X}$ *satisfying (7) or (8), and let* $(z_K)_{K\geq 1}$ *be a sequence of integers such that* $z_K/K \to \bar{n}_x$. *Then,*

$$\lim_{K\to+\infty} \mathbf{P}^K_{\frac{z_K}{K}\delta_x+\frac{1}{K}\delta_y}(V_0 = y) = \frac{[f(y,x)]_+}{b(y)}, \tag{41}$$

$$\lim_{K\to+\infty} \mathbf{P}^K_{\frac{z_K}{K}\delta_x+\frac{1}{K}\delta_y}(V_0 = x) = 1 - \frac{[f(y,x)]_+}{b(y)}, \tag{42}$$

$$\forall \eta > 0, \quad \lim_{K\to+\infty} \mathbf{P}^K_{\frac{z_K}{K}\delta_x+\frac{1}{K}\delta_y}\left(\theta_0 > \frac{\eta}{Ku_K} \wedge \tau_1\right) = 0 \tag{43}$$

$$\text{and} \quad \forall \varepsilon > 0, \quad \lim_{K\to+\infty} \mathbf{P}^K_{\frac{z_K}{K}\delta_x+\frac{1}{K}\delta_y}\left(|\langle \nu^K_{\theta_0}, \mathbf{1}\rangle - \bar{n}_{V_0}| < \varepsilon\right) = 1, \tag{44}$$

where $f(y,x)$ *has been defined in (3).*

Observe that (43) implies in particular that

$$\lim_{K\to+\infty}\mathbf{P}^K_{\frac{z_K}{K}\delta_x+\frac{1}{K}\delta_y}(\theta_0<\tau_1)=1.$$

The proofs of these lemmas are postponed at the end of this section.

Proof of Theorem 1.1 Observe that the generator A, defined in (9), of the TSS process $(X_t,t\geq 0)$ of Theorem 1.1 can be written as

$$A\varphi(x)=\int_{\mathbb{R}^l}(\varphi(x+h)-\varphi(x))\beta(x)\kappa(x,dh), \tag{45}$$

where the probability measure $\kappa(x,dh)$ is defined by

$$\kappa(x,dh)=\left(1-\int_{\mathbb{R}^l}\frac{[f(x+v,x)]_+}{b(x+v)}m(x,v)dv\right)\delta_0(dh) \\ +\frac{[f(x+h,x)]_+}{b(x+h)}m(x,h)dh. \tag{46}$$

This means that the TSS model X with initial state x can be constructed as follows: let $(Z(k),k=0,1,2,\dots)$ be a Markov chain in $\mathcal{X}$ with initial state x and with transition kernel $\kappa(x,dh)$, and let $(N(t),t\geq 0)$ be and independent standard Poisson process. Let also $(T_n)_{n\geq 1}$ denote the sequence of jump times of the Poisson process N. Then, the process $(X_t,t\geq 0)$ defined by

$$X_t:=Z\left(N\left(\int_0^t\beta(X_s)ds\right)\right)$$

is a Markov process with infinitesimal generator (45) (cf. [28] chapter 6).

Let $\mathbf{P}_x$ denote its law, and define $(S_n)_{n\geq 1}$ by $T_n=\int_0^{S_n}\beta(X_s)ds$. By (A1) and (A3), $\beta(\cdot)>0$, and so S_n is a.s. finite for any $n\geq 1$. Observe that any jump of the process X occurs at some time S_n, but that all S_n may not be effective jump times for X, because of the Dirac mass at 0 appearing in (46).

Fix $t>0$, $x\in\mathcal{X}$ and a measurable subset Γ of $\mathcal{X}$. Under $\mathbf{P}_x$, S_1 and X_{S_1} are independent, S_1 is an exponential random variable with parameter $\beta(x)$, and $X_{S_1}-x$ has law $\kappa(x,\cdot)$. Therefore, for any $n\geq 1$, the strong Markov property applied to X at time S_1 yields

$$\mathbf{P}_x(S_n\leq t<S_{n+1},\,X_t\in\Gamma) \\ =\int_0^t\beta(x)e^{-\beta(x)s}\int_{\mathbb{R}^l}\mathbf{P}_{x+h}(S_{n-1}\leq t-s<S_n,\,X_{t-s}\in\Gamma)\kappa(x,dh)ds. \tag{47}$$

Moreover,

$$\mathbf{P}_x(0 \leq t < S_1,\ X_t \in \Gamma) = \mathbf{1}_{\{x\in\Gamma\}} e^{-\beta(x)t}. \tag{48}$$

The idea of our proof of Theorem 1.1 is to show that the same relations hold when we replace S_n by τ_n and X_t by the support of ν^K_{t/Ku_K} (when it is a singleton) and when $K \to +\infty$.

More precisely, fix $x \in \mathcal{X}$, $t > 0$ and a measurable subset Γ of $\mathcal{X}$, and observe that

$$\begin{aligned}\big\{\mathrm{Supp}(\nu^K_{t/Ku_K}) \text{ is a singleton } \{y\},\ y \in \Gamma \text{ and } |\langle \nu^K_{t/Ku_K},\mathbf{1}\rangle - \bar{n}_y| < \varepsilon\big\}\\ = \bigcup_{n\geq 0} A^K_n(t,\Gamma,\varepsilon),\end{aligned} \tag{49}$$

where

$$A^K_n(t,\Gamma,\varepsilon) := \left\{\theta_n \leq \frac{t}{Ku_K} < \tau_{n+1},\ V_n \in \Gamma,\ |\langle \nu^K_{t/Ku_K},\mathbf{1}\rangle - \bar{n}_{V_n}| < \varepsilon\right\}.$$

Let us define, for any $z \in \mathbb{N}$ and $n \geq 0$,

$$\begin{aligned}p^K_n(t,x,\Gamma,\varepsilon,z) := \mathbf{P}^K_{\frac{z}{K}\delta_x}\Bigg(\theta_n \leq \frac{t}{Ku_K} < \tau_{n+1},\ V_n \in \Gamma\\ \text{and} \sup_{s\in[\theta_n,\tau_{n+1}]} |\langle \nu^K_s,\mathbf{1}\rangle - \bar{n}_{V_n}| < \varepsilon\Bigg),\end{aligned}$$

and define also,

$$\begin{aligned}q^K_0(t,x,\Gamma,\varepsilon,z) &:= \mathbf{P}^K_{\frac{z}{K}\delta_x}\left(\frac{t}{Ku_K} < \tau_1,\ V_0 \in \Gamma,\ \sup_{s\in[\log K,\tau_1]} |\langle \nu^K_s,\mathbf{1}\rangle - \bar{n}_{V_0}| < \varepsilon\right)\\ &= \mathbf{1}_{\{x\in\Gamma\}} \mathbf{P}^K_{\frac{z}{K}\delta_x}\left(\frac{t}{Ku_K} < \tau_1,\ \sup_{s\in[\log K,\tau_1]} |\langle \nu^K_s,\mathbf{1}\rangle - \bar{n}_x| < \varepsilon\right).\end{aligned}$$

Let us also extend these definitions to $\varepsilon = \infty$ by suppressing the condition involving the supremum of $|\langle \nu^K,\mathbf{1}\rangle - \bar{n}_{V_n}|$.

Then

Lemma 3.3

(a) *For any $x \in \mathcal{X}$, $n \geq 1$, $t > 0$, $\varepsilon \in]0,\infty]$ and for any sequence of integers (z_K) such that $z_K/K \to z > 0$, $p_n(t,x,\Gamma) := \lim_{K\to+\infty} p^K_n(t,x,\Gamma,\varepsilon,z_K)$ exists, and is independent of (z_K), $z > 0$ and ε.*

Similarly, $p_0(t,x,\Gamma) := \lim_{K\to+\infty} q^K_0(t,x,\Gamma,\varepsilon,z_K)$ exists, and is independent of (z_K), $z > 0$ and ε, and, if $z = \bar{n}_x$, $\lim_{K\to+\infty} p^K_0(t,x,\Gamma,\varepsilon,z_K)$ exists and is also equal to $p_0(t,x,\Gamma)$.

Finally, if we assume that (z_K) *is a sequence of* $\mathbb{N}$*-valued random variables such that* z_K/K *converge in probability to a deterministic* $z > 0$*, then the limits above hold in probability (with the same restriction that* z *has to be equal to* $\bar{n}_x$ *for* p_0^K*).*

(b) *The functions* $p_n(t,x,\Gamma)$ *are continuous with respect to* t *and measurable with respect to* x*, and satisfy*

$$\begin{aligned} p_0(t,x,\Gamma) &= \mathbf{1}_{\{x\in\Gamma\}} e^{-\beta(x)t} \quad \textit{and} \quad \forall n \geq 0, \\ p_{n+1}(t,x,\Gamma) &= \int_0^t \beta(x) e^{-\beta(x)s} \int_{\mathbb{R}^l} p_n(t-s,x+h,\Gamma)\kappa(x,dh)ds. \end{aligned}$$

Let us postpone the proof of this lemma after the proof of Theorem 1.1.

Observe that, because of (47) and (48), Lemma 3.3 (b) implies that $\mathbf{P}_x(S_n \leq t < S_{n+1},\ X_t \in \Gamma) = p_n(t,x,\Gamma)$.

Now, let $\tilde{\mathbf{P}}^K_\nu$ denote the law of the process ν^K with *random* initial state ν. Since ν^K is Markov, $\tilde{\mathbf{P}}^K_{\gamma_K/K\delta_x} = \mathbf{E}[\mathbf{P}^K_{\gamma_K(\omega)/K\delta_x}]$. By (49),

$$\begin{aligned} &\tilde{\mathbf{P}}^K_{\frac{\gamma_K}{K}\delta_x}\big(\mathrm{Supp}(\nu^K_{t/Ku_K}) \text{ is a singleton } \{y\},\ y\in\Gamma \\ &\qquad\qquad \text{and } |\langle \nu^K_{t/Ku_K},\mathbf{1}\rangle - \bar{n}_y| < \varepsilon\big) = \sum_{n\geq 0} \tilde{\mathbf{P}}^K_{\frac{\gamma_K}{K}\delta_x}(A_n^K(t,\Gamma,\varepsilon)), \end{aligned}$$

where (γ_K) is the sequence of $\mathbb{N}$-valued random variables of Theorem 1.1.

For any $K \geq 1$ and $n \geq 1$,

$$\begin{aligned} p_n^K(t,x,\Gamma,\varepsilon,\gamma_K) &\leq \mathbf{P}^K_{\frac{\gamma_K}{K}\delta_x}(A_n^K(t,\Gamma,\varepsilon)) \leq p_n^K(t,x,\Gamma,\infty,\gamma_K), \\ \text{and} \quad q_0^K(t,x,\Gamma,\varepsilon,\gamma_K) &\leq \mathbf{P}^K_{\frac{\gamma_K}{K}\delta_x}(A_n^K(t,\Gamma,\varepsilon)) \leq p_n^K(t,x,\Gamma,\infty,\gamma_K), \end{aligned}$$

so, by Lemma 3.3 (a), for any $n \geq 0$, $\mathbf{P}^K_{\gamma_K/K\delta_x}(A_n^K(t,\Gamma,\varepsilon)) \xrightarrow{\mathcal{P}} p_n(t,x,\Gamma)$, and therefore, $\lim_{K\to+\infty} \tilde{\mathbf{P}}^K_{\gamma_K/K\delta_x}(A_n^K(t,\Gamma,\varepsilon)) = p_n(t,x,\Gamma)$.

Now, by (49), for any $K \geq 1$,

$$\sum_{n=0}^{+\infty} \tilde{\mathbf{P}}^K_{\frac{\gamma_K}{K}\delta_x}(A_n^K(t,\Gamma,\varepsilon)) \leq 1,$$

so, by (49) and the dominated convergence theorem,

$$\begin{aligned} &\lim_{K\to+\infty} \tilde{\mathbf{P}}^K_{\frac{\gamma_K}{K}\delta_x}\big(\mathrm{Supp}(\nu^K_{t/Ku_K}) \text{ is a singleton } \{y\},\ y\in\Gamma \\ &\qquad\qquad \text{and } |\langle \nu^K_{t/Ku_K},\mathbf{1}\rangle - \bar{n}_y| < \varepsilon\big) = \sum_{n\geq 0} p_n(t,x,\Gamma) = \mathbf{P}_x(X_t\in\Gamma), \end{aligned}$$

which is (11) in the case of a single time t.

In order to complete the proof of Theorem 1.1, we have to generalize this limit to any sequence of times $0 < t_1 < \ldots < t_n$.

We will specify the method only in the case of two times $0 < t_1 < t_2$. It can be easily generalized to a sequence of n times. We introduce for any integers $0 \leq n_1 \leq n_2$ the probabilities

$$\begin{aligned}
&p^K_{n_1,n_2}(t_1,t_2,x,\Gamma_1,\Gamma_2,\varepsilon,z)\\
&:= \mathbf{P}^K_{\frac{z}{K}\delta_x}\Bigg(\theta_{n_1} \leq \frac{t_1}{Ku_K} < \tau_{n_1+1},\ V_{n_1} \in \Gamma_1,\ \sup_{s\in[\theta_{n_1},\tau_{n_1+1}]} |\langle \nu^K_s,\mathbf{1}\rangle - \bar{n}_{V_{n_1}}| < \varepsilon,\\
&\qquad \theta_{n_2} \leq \frac{t_2}{Ku_K} < \tau_{n_2+1},\ V_{n_2} \in \Gamma_2 \text{ and } \sup_{s\in[\theta_{n_2},\tau_{n_2+1}]} |\langle \nu^K_s,\mathbf{1}\rangle - \bar{n}_{V_{n_2}}| < \varepsilon\Bigg),
\end{aligned}$$

and

$$\begin{aligned}
&q^K_{0,n_2}(t_1,t_2,x,\Gamma_1,\Gamma_2,\varepsilon,z)\\
&\qquad := \mathbf{1}_{\{\mathbf{x}\in\mathbf{\Gamma_1}\}}\mathbf{P}^K_{\frac{z}{K}\delta_x}\Bigg(\frac{t_1}{Ku_K} < \tau_1,\ \sup_{s\in[\log K,\tau_1]} |\langle \nu^K_s,\mathbf{1}\rangle - \bar{n}_x| < \varepsilon,\\
&\qquad \theta_{n_2} \leq \frac{t_2}{Ku_K} < \tau_{n_2+1},\ V_{n_2} \in \Gamma_2 \text{ and } \sup_{s\in[\theta_{n_2},\tau_{n_2+1}]} |\langle \nu^K_s,\mathbf{1}\rangle - \bar{n}_{V_{n_2}}| < \varepsilon\Bigg).
\end{aligned}$$

Then, we can use a calculation very similar to the proof of Lemma 3.3 to prove that, as $K \to +\infty$, $p^K_{n_1,n_2}(t_1,t_2,x,\Gamma_1,\Gamma_2,\varepsilon,z_K)$ converges to a limit $p_{n_1,n_2}(t_1,t_2,x,\Gamma_1,\Gamma_2)$ independent of $\varepsilon \in]0,\infty]$, z_K and the limit $z > 0$ of z_K/K (with the restriction that z has to be equal to $\bar{n}_x$ if $n_1 = 0$), and that $\lim_K q^K_{0,n_2}(t_1,t_2,x,\Gamma_1,\Gamma_2,\varepsilon,z) = p_{0,n_2}(t_1,t_2,x,\Gamma_1,\Gamma_2)$, where

$$\begin{cases}
p_{0,n_2}(t_1,t_2,x,\Gamma_1,\Gamma_2) = \mathbf{1}_{\{x\in\Gamma_1\}}e^{-\beta(x)t_1}p_{n_2}(t_2-t_1,x,\Gamma_2);\\
p_{n_1+1,n_2+1}(t_1,t_2,x,\Gamma_1,\Gamma_2)\\
\qquad = \displaystyle\int_0^{t_1} \beta(x)e^{-\beta(x)s}\int_{\mathbb{R}^l} p_{n_1,n_2}(t_1-s,t_2-s,x+h,\Gamma_1,\Gamma_2)\kappa(x,dh)ds.
\end{cases}$$

As above, we obtain equation (11) for $n = 2$ by observing that the same relation holds for the TSS process X.

This completes the proof of Theorem 1.1. □

Proof of Lemma 3.3 First, let us prove that the convergence property of $p^K_n(t,x,\Gamma,\varepsilon,z_K)$ when $z_K \in \mathbb{N}$ in Lemma 3.3 (a) implies the convergence in probability of these quantities when z_K are random variables. Actually, if (z_K) is a sequence of random variables such that $z_K/K \xrightarrow{\mathcal{P}} z$, by Skorohod's Theorem, we can construct on an auxiliary probability space $\hat{\Omega}$ a sequence of random variables $(\hat{z}_K)$ such that $\mathcal{L}(\hat{z}_K) = \mathcal{L}(z_K)$ and $\hat{z}_K(\hat{\omega})/K \to z$ for any $\hat{\omega} \in \hat{\Omega}$. Then, $\lim p_n^K(t,x,\Gamma,\varepsilon,\hat{z}_K(\hat{\omega})) = p_n(t,x,\Gamma)$ for any $\hat{\omega} \in \hat{\Omega}$, which implies that $p_n^K(t,x,\Gamma,\varepsilon,z_K) \xrightarrow{\mathcal{P}} p_n(t,x,\Gamma)$. The same method applies to $q_0^K(t,x,\Gamma,\varepsilon,z_k)$.

We will prove Lemma 3.3 (a) and (b) by induction over $n \geq 0$.

First, when $t > 0$, it follows from the fact that $t/Ku_K > \log K$ for sufficiently large K, and from Lemma 3.1 (b) and (c), that

$$\lim_{K\to+\infty} q_0^K(t,x,\Gamma,\varepsilon,z_K) = \mathbf{1}_{\{x\in\Gamma\}} e^{-\beta(x)t},$$

and that, if $z = \bar{n}_x$,

$$\lim_{K\to+\infty} p_0^K(t,x,\Gamma,\varepsilon,z_K) = \mathbf{1}_{\{x\in\Gamma\}} e^{-\beta(x)t}.$$

Then, fix $n \geq 0$ and assume that Lemma 3.3 (a) holds for n. We intend to prove the convergence of $p_{n+1}^K(t,x,\Gamma,\varepsilon,z_K)$ to $p_{n+1}(t,x,\Gamma)$ such that

$$p_{n+1}(t,x,\Gamma) = \int_0^t \beta(x)e^{-\beta(x)s} \int_{\mathbb{R}^l} p_n(t-s,x+h,\Gamma)\kappa(x,dh)ds,$$

by applying the strong Markov property at time τ_1, in a similar way than when we obtained (47). However, the induction assumption only gives the convergence of $p_n^K(t,x,\Gamma,\varepsilon,z_K)$ to $p_n(t,x,\Gamma)$ for *non-random* t. Therefore, we will divide the time interval $[0,t]$ in a finite number of small intervals and use the Markov property at time τ_1 when τ_1 is in each of these intervals. Moreover, we will also use the Markov property at time θ_1 and, in order to be able to apply Lemma 3.2 (which holds for a *non-random* mutant trait y) after this time, we will use the fact that U_1 is independent of τ_1 and $\nu^K_{\tau_1-}$ and that $U_1 - x$ is a random variable with law $m(x,h)dh$.

Following this program, we can bound $p_{n+1}^K(t,x,\Gamma,\varepsilon,z_K)$ from above as follows: fix $\eta > 0$; using Lemma 3.1 (a) in the first inequality, for sufficiently large $k \geq 0$ and $K \geq 1$,

$$\begin{aligned} p_{n+1}^K(t,x,\Gamma,\varepsilon,z_K) &\leq \mathbf{P}^K_{\frac{z_K}{K}\delta_x}\left(\theta_{n+1} \leq \frac{t}{Ku_K},\ \tau_{n+2} > \frac{t+2/2^k}{Ku_K},\ V_{n+1} \in \Gamma\right) + \eta \\ &\leq \sum_{i=0}^{\lceil t2^k\rceil -1} \mathbf{P}^K_{\frac{z_K}{K}\delta_x}\left(\frac{i}{2^kKu_K} \leq \tau_1 \leq \frac{i+1}{2^kKu_K},\ \theta_{n+1} \leq \frac{t}{Ku_K},\right. \\ &\qquad \left.\tau_{n+2} > \frac{t+2/2^k}{Ku_K} \text{ and } V_{n+1} \in \Gamma\right) + \eta \end{aligned}$$

$$\leq \sum_{i=0}^{\lceil t2^k\rceil-1} \mathbf{E}^K_{\frac{z_K}{K}\delta_x}\left[\mathbf{1}_{\left\{\frac{i}{2^kKu_K}\leq\tau_1\leq\frac{i+1}{2^kKu_K}\right\}}\mathbf{P}^K_{\nu^K_{\tau_1-}+\frac{1}{K}\delta_{U_1}}\left(\theta_n\leq\frac{t-i/2^k}{Ku_K},\right.\right.$$

$$\left.\left.\tau_{n+1}>\frac{t-(i-1)/2^k}{Ku_K}\text{ and }V_n\in\Gamma\right)\right]+\eta$$

$$\leq \sum_{i=0}^{\lceil t2^k\rceil-1} \mathbf{E}^K_{\frac{z_K}{K}\delta_x}\left[\mathbf{1}_{\left\{\frac{i}{2^kKu_K}\leq\tau_1\leq\frac{i+1}{2^kKu_K}\right\}}\int_{\mathbb{R}^l}\mathbf{E}^K_{\nu^K_{\tau_1-}+\frac{1}{K}\delta_{x+h}}\left(\mathbf{1}_{\left\{\theta_0\geq\frac{1}{2^kKu_K}\wedge\tau_1\right\}}\right.\right.$$

$$\left.\left.+\mathbf{1}_{\left\{\theta_0<\frac{1}{2^kKu_K}\wedge\tau_1\right\}}\mathbf{P}^K_{\nu^K_{\theta_0}}\left(\theta_n\leq\frac{t-i/2^k}{Ku_K}<\tau_{n+1},\,V_n\in\Gamma\right)\right)m(x,h)dh\right]+\eta.$$

$$\leq \sum_{i=0}^{\lceil t2^k\rceil-1} \mathbf{E}^K_{\frac{z_K}{K}\delta_x}\left[\mathbf{1}_{\left\{\frac{i}{2^kKu_K}\leq\tau_1\leq\frac{i+1}{2^kKu_K}\right\}}\int_{\mathbb{R}^l}\mathbf{E}^K_{\nu^K_{\tau_1-}+\frac{1}{K}\delta_{x+h}}\left(\mathbf{1}_{\left\{\theta_0\geq\frac{1}{2^kKu_K}\wedge\tau_1\right\}}\right.\right.$$

$$\left.\left.+\mathbf{1}_{\left\{\theta_0<\frac{1}{2^kKu_K}\wedge\tau_1\right\}}p^K_n(t-i/2^k,V_0,\Gamma,\infty,K\langle\nu^K_{\theta_0},\mathbf{1}\rangle)\right)m(x,h)dh\right]+\eta. \tag{50}$$

Now, since $\nu^K_{\tau_1-}=\langle\nu^K_{\tau_1-},\mathbf{1}\rangle\delta_x$, under $\mathbf{P}^K_{\nu^K_{\tau_1-}+\frac{1}{K}\delta_{x+h}}$, on the event $\{\theta_0<\tau_1\}$,

$$\begin{aligned}p^K_n(t-i/2^k,V_0,\Gamma,\infty,K\langle\nu^K_{\theta_0},\mathbf{1}\rangle)&=\mathbf{1}_{\{V_0=x\}}p^K_n(t-i/2^k,x,\Gamma,\infty,K\langle\nu^K_{\theta_0},\mathbf{1}\rangle)\\&\quad+\mathbf{1}_{\{V_0=x+h\}}p^K_n(t-i/2^k,x+h,\Gamma,\infty,K\langle\nu^K_{\theta_0},\mathbf{1}\rangle).\end{aligned} \tag{51}$$

By Lemma 3.1 (b), $\nu^K_{\tau_1-}\xrightarrow{\mathcal{P}}\bar{n}_x\delta_x$ under $\mathbf{P}_{\frac{z_K}{K}\delta_x}$, so we can use Skorohod's Theorem to construct random variables $\hat{N}_K$ on an auxiliary probability space $\hat{\Omega}$ with the same law that $\langle\nu^K_{\tau_1-},\mathbf{1}\rangle$ and converging to $\bar{n}_x$ for any $\hat{\omega}\in\hat{\Omega}$.

Fix $\hat{\omega}\in\hat{\Omega}$. Under $\mathbf{P}^K_{\hat{N}_K(\hat{\omega})\delta_x+\frac{1}{K}\delta_{x+h}}$, define

$$Z^K_1=\langle\nu^K_{\theta_0},\mathbf{1}\rangle\mathbf{1}_{\{V_0=x,\ \theta_0<\tau_1\}}+\frac{\lceil K\bar{n}_x\rceil}{K}\mathbf{1}_{\{V_0\neq x\}\cup\{\theta_0\geq\tau_1\}}.$$

It follows from Lemma 3.2 (43) and (44), and from assumption (B) that, for Lebesgue almost every h, under $\mathbf{P}^K_{\hat{N}_K(\hat{\omega})\delta_x+1/K\delta_{x+h}}$, $Z^K_1\xrightarrow{\mathcal{P}}\bar{n}_x$, and so, by the induction assumption, when $K\to+\infty$,

$$p^K_n(t-i/2^k,x,\Gamma,\infty,KZ^K_1)\xrightarrow{\mathcal{P}}p_n(t-i/2^k,x,\Gamma).$$

Now, given two sequences $(X_K)_{K\geq1}$ and $(Y_K)_{K\geq0}$ of uniformly bounded random variables such that X_K and Y_K are defined on the same probability space for any $K\geq1$, and such that, when $K\to+\infty$, X_K converges in probability to a constant C and $\lim_K\mathbf{E}(Y_K)$ exists, it is easy to prove that

$$\lim_{K\to+\infty}\mathbf{E}(X_KY_K)=C\lim_{K\to+\infty}\mathbf{E}(Y_K). \tag{52}$$

Applying this property to $X_K = p_n^K(t - i/2^k, x, \Gamma, \infty, KZ_1^K)$ and $Y_K = \mathbf{1}_{\{V_0=x,\ \theta_0<\tau_1\}}$, we obtain, by Lemma 3.2 (42) and (43) and assumption (B), for Lebesgue almost any h, and for any $\hat{\omega} \in \hat{\Omega}$,

$$\begin{aligned}\lim_{K\to+\infty} \mathbf{E}^K_{\hat{N}_K(\hat{\omega})\delta_x + \frac{1}{K}\delta_{x+h}} &\left(\mathbf{1}_{\{V_0=x,\ \theta_0<\tau_1\}} p_n^K(t - i/2^k, x, \Gamma, \infty, K\langle \nu^K_{\theta_0}, \mathbf{1}\rangle)\right) \\ &= \left(1 - \frac{[f(x+h,x)]_+}{b(x+h)}\right) p_n(t - i/2^k, x, \Gamma).\end{aligned}$$

Since this holds for any $\hat{\omega} \in \hat{\Omega}$, we finally obtain that, for Lebesgue almost any h, under $\mathbf{P}^K_{\frac{z_K}{K}\delta_x}$,

$$\begin{aligned}\mathbf{E}^K_{\nu^K_{\tau_1-} + \frac{1}{K}\delta_{x+h}} &\left(\mathbf{1}_{\{V_0=x,\ \theta_0<\tau_1\}} p_n^K(t - i/2^k, x, \Gamma, \infty, K\langle \nu^K_{\theta_0}, \mathbf{1}\rangle)\right) \\ &\xrightarrow{\mathcal{P}} \left(1 - \frac{[f(x+h,x)]_+}{b(x+h)}\right) p_n(t - i/2^k, x, \Gamma). \qquad (53)\end{aligned}$$

Similarly, we can use Lemma 3.2 (41) and the random variable

$$Z_2^K = \langle \nu^K_{\theta_0}, \mathbf{1}\rangle \mathbf{1}_{\{V_0=x+h,\ \theta_0<\tau_1\}} + \bar{n}_{x+h}\mathbf{1}_{\{V_0\neq x+h\}\cup\{\theta_0\geq\tau_1\}}$$

to prove that, for Lebesgue almost any h, under $\mathbf{P}^K_{\frac{z_K}{K}\delta_x}$,

$$\begin{aligned}\mathbf{E}^K_{\nu^K_{\tau_1-} + \frac{1}{K}\delta_{x+h}} &\left(\mathbf{1}_{\{V_0=x+h,\ \theta_0<\tau_1\}} p_n^K(t - i/2^k, x+h, \Gamma, \infty, K\langle \nu^K_{\theta_0}, \mathbf{1}\rangle)\right) \\ &\xrightarrow{\mathcal{P}} \frac{[f(x+h,x)]_+}{b(x+h)} p_n(t - i/2^k, x+h, \Gamma). \qquad (54)\end{aligned}$$

Moreover, by Lemma 3.2 (43), for Lebesgue almost any h, under $\mathbf{P}^K_{(z_K/K)\delta_x}$,

$$\mathbf{P}^K_{\nu^K_{\tau_1-} + \frac{1}{K}\delta_{x+h}}\left(\theta_0 \geq \frac{1}{2^k K u_K} \wedge \tau_1\right) \xrightarrow{\mathcal{P}} 0. \qquad (55)$$

Collecting these results together, it follows from Lemma 3.1 (c), (51) and from the fact that all the quantities in (53), (54) and (55) are bounded by 1, that, for Lebesgue almost any h,

$$\begin{aligned}\lim_{K\to+\infty} \mathbf{E}^K_{\frac{z_K}{K}\delta_x} &\Bigg[\mathbf{1}_{\left\{\frac{i}{2^k K u_K} \leq \tau_1 \leq \frac{i+1}{2^k K u_K}\right\}} \mathbf{E}^K_{\nu^K_{\tau_1-} + \frac{1}{K}\delta_{x+h}}\Bigg(\mathbf{1}_{\left\{\theta_0 \geq \frac{1}{2^k K u_K} \wedge \tau_1\right\}} \\ &\quad + \mathbf{1}_{\left\{\theta_0 < \frac{1}{2^k K u_K} \wedge \tau_1\right\}} p_n^K(t - i/2^k, V_0, \Gamma, \infty, K\langle \nu^K_{\theta_0}, \mathbf{1}\rangle)\Bigg)\Bigg] \\ &= \left(e^{-\beta(x)\frac{i}{2^k}} - e^{-\beta(x)\frac{i+1}{2^k}}\right)\Bigg[\frac{[f(x+h,x)]_+}{b(x+h)} p_n(t - i/2^k, x+h, \Gamma) \\ &\quad + \left(1 - \frac{[f(x+h,x)]_+}{b(x+h)}\right) p_n(t - i/2^k, x, \Gamma)\Bigg].\end{aligned}$$

Finally, taking the integral of both sides with respect to $m(x,h)dh$, the dominated convergence theorem and (50) yield

$$\begin{aligned}&\limsup_{K\to+\infty}\ p^K_{n+1}(x,t,\Gamma,\varepsilon,z_K)\\&\quad\le\sum_{i=0}^{\lceil t2^k\rceil-1}\left(e^{-\beta(x)\frac{i}{2^k}}-e^{-\beta(x)\frac{i+1}{2^k}}\right)\int_{\mathbb{R}^l}p_n(t-i/2^k,x+h,\Gamma)\kappa(x,dh)+\eta.\end{aligned}$$

Taking the limit $k\to+\infty$ first and then $\eta\to 0$, it follows from the fact that

$$e^{-\beta(x)i/2^k}-e^{-\beta(x)(i+1)/2^k}=e^{-\beta(x)i/2^k}(\beta(x)/2^k+O(1/2^{2k}))$$

and from the convergence of Riemann sums that

$$\limsup_{K\to+\infty}\ p^K_{n+1}(x,t,\Gamma,\varepsilon,z_K)\le\int_0^t\beta(x)e^{-\beta(x)s}\int_{\mathbb{R}^l}p_n(t-s,x+h,\Gamma)\kappa(x,dh)ds.$$

Using the same method than for (50), we can give a lower bound for p^K_n as follows: for any $\eta>0$, for sufficiently large $k\ge 0$ and $K\ge 1$,

$$\begin{aligned}p^K_{n+1}(t,x,\Gamma,\varepsilon,z_K)&\ge\mathbf{P}^K_{\frac{z_K}{K}\delta_x}\Bigg(\theta_{n+1}\le\frac{t}{Ku_K},\ \tau_{n+2}>\frac{t-2/2^k}{Ku_K},\ V_{n+1}\in\Gamma\\&\qquad\text{and}\sup_{s\in[\theta_{n+1},\tau_{n+2}]}|\langle\nu^K_s,\mathbf{1}\rangle-\bar{n}_{V_{n+1}}|<\varepsilon\Bigg)-\eta\\&\ge\sum_{i=0}^{\lfloor t2^k\rfloor-3}\mathbf{E}^K_{\frac{z_K}{K}\delta_x}\Bigg[\mathbf{1}_{\left\{\frac{i}{2^kKu_K}\le\tau_1\le\frac{i+1}{2^kKu_K}\right\}}\mathbf{P}^K_{\nu^K_{\tau_1-}+\frac{1}{K}\delta_{U_1}}\Bigg(\theta_n\le\frac{t-(i+1)/2^k}{Ku_K},\\&\qquad\tau_{n+1}>\frac{t-(i+2)/2^k}{Ku_K},\ V_n\in\Gamma\text{ and}\sup_{s\in[\theta_n,\tau_{n+1}]}|\langle\nu^K_s,\mathbf{1}\rangle-\bar{n}_{V_n}|<\varepsilon\Bigg)\Bigg]-\eta\\&\ge\sum_{i=0}^{\lfloor t2^k\rfloor-3}\mathbf{E}^K_{\frac{z_K}{K}\delta_x}\Bigg[\mathbf{1}_{\left\{\frac{i}{2^kKu_K}\le\tau_1\le\frac{i+1}{2^kKu_K}\right\}}\int_{\mathbb{R}^l}\mathbf{E}^K_{\nu^K_{\tau_1-}+\frac{1}{K}\delta_{x+h}}\Bigg(\mathbf{1}_{\left\{\theta_0<\frac{1}{2^kKu_K}\wedge\tau_1\right\}}\\&\qquad p^K_n(t-(i+2)/2^k,V_0,\Gamma,\varepsilon,K\langle\nu^K_{\theta_0},\mathbf{1}\rangle)\Bigg)m(x,h)dh\Bigg]-\eta.\end{aligned}$$

Then, using the same method as above, letting $K\to+\infty$, then $k\to+\infty$ and finally $\eta\to 0$,

$$\liminf_{K\to+\infty}p^K_{n+1}(x,t,\Gamma,\varepsilon,z_K)\ge\int_0^t\beta(x)e^{-\beta(x)s}\int_{\mathbb{R}^l}p_n(t-s,x+h,\Gamma)\kappa(x,dh)ds,$$

which completes the proof of Lemma 3.3 by induction. □

Proof of Lemma 3.1 (a) Fix $\eta > 0$. By Theorem 2.1 (a) and (c), for any $K \geq 1$,

$$\langle \nu^K,\mathbf{1}\rangle \preceq Z^K,$$
$$\text{where}\quad \mathcal{L}(Z^K) = \mathbf{P}^K(2\bar{b},0,\underline{\alpha},\langle \nu_0^K,\mathbf{1}\rangle + 1).$$

Since $\sup_K \mathbf{E}(\langle \nu_0^K,\mathbf{1}\rangle) < +\infty$, we can choose $M < +\infty$ such that

$$\sup_{K\geq 1} \mathbf{P}(\langle \nu_0^K,\mathbf{1}\rangle + 1 > M) < \eta/3.$$

Then, applying Theorem 2.2 (c) to $\mathbf{P}^K(2\bar{b},0,\underline{\alpha},\langle \nu_0^K,\mathbf{1}\rangle+1)$ with $C = [1,M]$, $\eta_2 = M$ and η_1 such that $0 < 2\bar{b}/\underline{\alpha} - \eta_1 < 1/2$, there exists $V > 0$ such that

$$\limsup_{K\to+\infty} \mathbf{P}(T^K < e^{KV}) < \eta/3, \tag{56}$$
$$\text{where}\quad T^K = \inf\{t \geq 0, Z_t^K \notin [1/2, M + 2\bar{b}/\underline{\alpha}]\}.$$

Fix $t,\varepsilon > 0$. Since, for $s \leq T^K$, $\langle \nu_s^K,\mathbf{1}\rangle \leq M + 2\bar{b}/\underline{\alpha}$, if we apply Theorem 2.1 (b) to the process $(\nu^K_{s+(t/Ku_K)}, s \geq 0)$, we obtain, for $s \leq T^K - t/Ku_K$,

$$A^K_{t/Ku_K+s} - A^K_{t/Ku_K} \preceq B^K_s,$$

where A_s^K is the number of mutations occuring between 0 and s, and where B^K is a Poisson process with parameter $Ku_K\bar{b}(M + 2\bar{b}/\underline{\alpha})$. Therefore, combining (56) with the fact that $1/Ku_K \ll e^{KV}$, we obtain that, for sufficiently large K

$$\begin{aligned}\mathbf{P}(A^K_{(t+\varepsilon)/Ku_K} - A^K_{t/Ku_K} \geq 1) &\leq \mathbf{P}(B^K_{\varepsilon/Ku_K} \geq 1) + 2\eta/3\\ &= 1 - \exp(-\bar{b}(M + 2\bar{b}/\underline{\alpha})\varepsilon) + 2\eta/3,\end{aligned}$$

which can be made smaller than η if ε is sufficiently small. This ends the proof of (38). □

Proof of Lemma 3.1 (b) Fix $\varepsilon > 0$. It follows from the formula (13) for ν^K that, for $t < \tau_1$, under $\mathbf{P}^K_{\frac{z_K}{K}\delta_x}$,

$$\nu_t^K = Z_t^K \delta_x,$$
$$\text{where}\quad \mathcal{L}(Z^K) = \mathbf{P}^K((1 - u_K\mu(x))b(x),d(x),\alpha(x,x),z_K/K).$$

Therefore, by Theorem 2.1 (c), for sufficiently large K such that $u_K < \varepsilon$ and for $t \leq \tau_1$,

$$Z^{K,1} \preceq \langle \nu^K,\mathbf{1}\rangle \preceq Z^{K,2}, \tag{57}$$
$$\text{where}\quad \mathcal{L}(Z^{K,1}) = \mathbf{P}^K((1 - \varepsilon)b(x),d(x),\alpha(x,x),z_K/K)$$
$$\text{and}\quad \mathcal{L}(Z^{K,2}) = \mathbf{P}^K(b(x),d(x),\alpha(x,x),z_K/K).$$

Now, let ϕ^1_y, resp. ϕ^2_y, be the solution to

$$\begin{aligned}&\dot{\phi} = ((1-\varepsilon)b(x) - d(x) - \alpha(x,x)\phi)\phi,\\ \text{resp.}\quad &\dot{\phi} = (b(x) - d(x) - \alpha(x,x)\phi)\phi,\end{aligned}$$

with initial state y, and observe that, for any $y > 0$, when $t \to +\infty$, $\phi^1_y(t) \to e^1 := \bar{n}_x - \varepsilon b(x)/\alpha(x,x)$ and $\phi^2_y(t) \to e^2 := \bar{n}_x$.

Define, for any $y > 0$, $t^{i,y}_\varepsilon$ the first time such that $\forall s \geq t^{i,y}_\varepsilon$, $\phi^i_y(s) \in [e^i - \varepsilon, e^i + \varepsilon]$ $(i = 1,2)$. Because of the continuity of the flow of these ODEs (see [64] p. 356),

$$t^i_\varepsilon := \sup_{y \in [z/2,2z]} t^{i,y}_\varepsilon < +\infty.$$

Let us apply Theorem 2.2 (a) to $Z^{K,1}$ and $Z^{K,2}$ on $[0,t_\varepsilon]$, where $t_\varepsilon = t^1_\varepsilon \vee t^2_\varepsilon$: since $z_K/K \to z$, for sufficiently small $\delta > 0$, and for $i = 1,2$,

$$\lim_{K\to+\infty} \mathbf{P}\left(\sup_{0\leq t\leq t_\varepsilon} |Z^{K,i}_t - \phi^i_{z_K/K}(t)| > \delta\right) = 0.$$

If we choose $\delta < \varepsilon$, we obtain, for $i = 1,2$,

$$\lim_{K\to+\infty} \mathbf{P}(|Z^{K,i}_{t_\varepsilon} - e^i| < 2\varepsilon) = 1,$$

and so, because of the expression of e^i, for $i = 1,2$,

$$\lim_{K\to+\infty} \mathbf{P}(|Z^{K,i}_{t_\varepsilon} - \bar{n}_x| < M\varepsilon) = 1, \tag{58}$$

where $M = 2 + b(x)/\alpha(x,x)$.

Now, assuming ε sufficiently small for $(M+1)\varepsilon < \bar{n}_x$, define the stopping times

$$T^{K,i}_\varepsilon = \inf\{t \geq t_\varepsilon : |Z^{K,i}_t - \bar{n}_x| > (M+1)\varepsilon\}$$

for $i = 1,2$, and $T^K_\varepsilon = T^{K,1}_\varepsilon \wedge T^{K,2}_\varepsilon$.

For any $z \in \mathbb{N}/K$, define also

$$\mathbf{P}^{K,1}_z := \mathbf{P}^K((1-\varepsilon)b(x),d(x),\alpha(x,x),z).$$

Then, applying Theorem 2.2 (c) to $\mathbf{P}^{K,1}_z$ with $C = [\bar{n}_x - M\varepsilon, \bar{n}_x + M\varepsilon]$, $\eta_1 = 3\varepsilon$ and $\eta_2 = (2M-1)\varepsilon$, there exists $V_1 > 0$ such that

$$\lim_{K\to+\infty} \inf_{z\in C} \mathbf{P}^{K,1}_z(\hat{Z}_\varepsilon > e^{KV_1}) = 1, \tag{59}$$

$$\text{where}\quad \hat{T}_\varepsilon = \inf\{t \geq 0 : |w_t - \bar{n}_x| > (M+1)\varepsilon\}.$$

Since, by the Markov property, for any $K \geq 1$,

$$\mathbf{P}(T_\varepsilon^{K,1} > e^{KV_1} + t_\varepsilon) = \mathbf{E}\left(\mathbf{P}^{K,1}_{Z^{K,1}_{t_\varepsilon}}(\hat{T}_\varepsilon > e^{KV_1})\right),$$

it follows from (58) that

$$\lim_{K\to+\infty} \mathbf{P}(T_\varepsilon^{K,1} > e^{KV_1} + t_\varepsilon) = 1.$$

Similarly, applying Theorem 2.2 (c) to $\mathbf{P}^K(b(x),d(x),\alpha(x,x),y)$ with $C = [\bar{n}_x - M\varepsilon,\bar{n}_x + M\varepsilon]$ and $\eta_1 = \eta_2 = (M+1)\varepsilon$, there exists $V_2 > 0$ such that

$$\lim_{K\to+\infty} \mathbf{P}(T_\varepsilon^{K,2} > e^{KV_2} + t_\varepsilon) = 1.$$

Therefore,

$$\lim_{K\to+\infty} \mathbf{P}(T_\varepsilon^{K} > e^{KV}) = 1 \tag{60}$$

where $V := V_1 \wedge V_2$.

Now, because of (57),

$$\forall t \in [t_\varepsilon,T_\varepsilon^K \wedge \tau_1], \quad |\langle \nu_s^K,\mathbf{1}\rangle - \bar{n}_x| < (M+1)\varepsilon. \tag{61}$$

Therefore, since $\log K > t_\varepsilon$ for sufficiently large K, in order to complete the proof of (39), it suffices to show that

$$\lim_{K\to+\infty} \mathbf{P}(\tau_1 < T_\varepsilon^K) = 1. \tag{62}$$

If we denote by A_t^K the number of mutations occuring between t_ε and $t + t_\varepsilon$, by Theorem 2.1 (b), for t such that $t_\varepsilon + t \leq T_\varepsilon^K \wedge \tau_1$,

$$B^K \preceq A^K,$$

where B^K is a Poisson process with parameter $Ku_K(\bar{n}_x - (M+1)\varepsilon)\mu(x)b(x)$.

Therefore, if we denote by S^K the first time when $B_t^K = 1$, on the event $\{t_\varepsilon + S^K < T_\varepsilon^K\}$,

$$\tau_1 \leq t_\varepsilon + S^K.$$

Since $\exp(-KV) \ll Ku_K$, $\lim_K \mathbf{P}(t_\varepsilon + S^K < e^{KV}) = 1$, and hence, by (60),

$$\lim_{K\to+\infty} \mathbf{P}(t_\varepsilon + S^K < T_\varepsilon^K) = 1,$$

which implies (62).

In the case where $z_K/K \to \bar{n}_x$, using (59) as above, we obtain easily

$$\lim_{K\to+\infty} \mathbf{P}(S_\varepsilon^K > e^{KV}) = 1,$$

$$\text{where}\quad S_\varepsilon^K = \inf\{t \geq 0 : |Z_t^{K,i} - \bar{n}_x| > (M+1)\varepsilon,\ i = 1,2\}.$$

Then, the proof of (40) can be completed using a method similar to the one we used above. □

Proof of Lemma 3.1 (c) Fix $t>0$ and $\varepsilon>0$. Take K large enough for $\log K<t/Ku_K$. The Markov property at time $\log K$ for ν^K yields

$$\begin{aligned}\mathbf{P}^K_{\frac{z_K}{K}\delta_x}\left(\tau_1>\frac{t}{Ku_K},\ \sup_{t\in[\log K,\tau_1]}|\langle\nu^K_t,\mathbf{1}\rangle-\bar n_x|<\varepsilon\right)&\\ =\mathbf{E}^K_{\frac{z_K}{K}\delta_x}\Bigg[\mathbf{1}_{\{\tau_1>\log K\}}\mathbf{P}^K_{\nu^K_{\log K}}\Bigg(\tau_1>\frac{t}{Ku_K}-\log K,&\\ \sup_{t\in[0,\tau_1]}|\langle\nu^K_t,\mathbf{1}\rangle-\bar n_x|<\varepsilon\Bigg)\Bigg].&\qquad(63)\end{aligned}$$

For any initial condition $\nu^K_0=\langle\nu^K_0,\mathbf{1}\rangle\delta_x$ of ν^K, by Theorem 2.1 (b), the number A^K_t of mutations of ν^K between 0 and t satisfies, for any $t\leq\tau_1$ such that $\sup_{s\in[0,t]}|\langle\nu^K_s,\mathbf{1}\rangle-\bar n_x|<\varepsilon$,

$$B^K\preceq A^K\preceq C^K,$$

where B^K_t and C^K_t are Poisson processes with respective parameters $Ku_K(\bar n_x-\varepsilon)\mu(x)b(x)$ and $Ku_K(\bar n_x+\varepsilon)\mu(x)b(x)$.

Therefore, on the event $\{\sup_{s\in[0,\tau_1]}|\langle\nu^K_s,\mathbf{1}\rangle-\bar n_x|<\varepsilon\}$, $S^K\leq\tau_1\leq T^K$, where T^K is the first time when $B^K_t=1$, and S^K the first time when $C^K_t=1$.

Now, by Lemma 3.1 (b), under $\mathbf{P}^K_{(z_K/K)\delta_x}$, $\nu^K_{\log K}\xrightarrow{\mathcal{P}}\bar n_x\delta_x$, so, by Skorohod's Theorem, we can construct $\hat N^K$ with the same law than $\langle\nu^K_{\log K},\mathbf{1}\rangle$ on an auxiliary probability space $\hat\Omega$ such that $\hat N^K(\hat\omega)\to\bar n_x$ for any $\hat\omega\in\hat\Omega$. Fix $\hat\omega\in\hat\Omega$. Then, by Lemma 3.1 (b),

$$\lim_{K\to+\infty}\mathbf{P}^K_{\hat N(\hat\omega)\delta_x}\left(\sup_{t\in[0,\tau_1]}|\langle\nu^K_t,\mathbf{1}\rangle-\bar n_x|<\varepsilon\right)=1,$$

and so,

$$\begin{aligned}\limsup_{K\to+\infty}\mathbf{P}^K_{\hat N(\hat\omega)\delta_x}\left(\tau_1>\frac{t}{Ku_K}-\log K,\ \sup_{t\in[0,\tau_1]}|\langle\nu^K_t,\mathbf{1}\rangle-\bar n_x|<\varepsilon\right)&\\ \leq\limsup_{K\to+\infty}\mathbf{P}^K_{\hat N(\hat\omega)\delta_x}\left(T^K>\frac{t}{Ku_K}-\log K\right)=\exp(-t(\bar n_x-\varepsilon)\mu(x)b(x)).&\end{aligned}$$

Therefore, under $\mathbf{P}^K_{(z_K/K)\delta_x}$,

$$\begin{aligned}\limsup_{K\to+\infty}\mathbf{P}^K_{\nu^K_{\log K}}\left(\tau_1>\frac{t}{Ku_K}-\log K,\ \sup_{t\in[0,\tau_1]}|\langle\nu^K_t,\mathbf{1}\rangle-\bar n_x|<\varepsilon\right)&\\ \leq\exp(-t(\bar n_x-\varepsilon)\mu(x)b(x))&\end{aligned}$$

in probability (where $\limsup X_n \leq a$ in probability means that, for any $\eta > 0$, $\mathbf{P}(X_n > a + \eta) \to 0$).

Similarly, under $\mathbf{P}^K_{(z_K/K)\delta_x}$,

$$\liminf_{K\to+\infty} \mathbf{P}^K_{\nu^K_{\log K}}\left(\tau_1 > \frac{t}{Ku_K} - \log K,\ \sup_{t\in[0,\tau_1]} |\langle \nu^K_t,\mathbf{1}\rangle - \bar{n}_x| < \varepsilon\right) \geq \exp(-t(\bar{n}_x + \varepsilon)\mu(x)b(x))$$

in probability.

Now, by Lemma 3.1 (a) and (b),

$$\lim_{K\to+\infty} \mathbf{P}^K_{\frac{z_K}{K}\delta_x}(\tau_1 > \log K) = 1$$
$$\text{and} \quad \lim_{K\to+\infty} \mathbf{P}^K_{\frac{z_K}{K}\delta_x}\left(\sup_{t\in[\log K,\tau_1]} |\langle \nu^K_t,\mathbf{1}\rangle - \bar{n}_x| < \varepsilon\right) = 1.$$

So, using results similar to (52), it follows from (63) that

$$\limsup_{K\to+\infty} \mathbf{P}^K_{\frac{z_K}{K}\delta_x}\left(\tau_1 > \frac{t}{Ku_K}\right) \leq \exp(-t(\bar{n}_x - \varepsilon)\mu(x)b(x))$$
$$\text{and} \quad \liminf_{K\to+\infty} \mathbf{P}^K_{\frac{z_K}{K}\delta_x}\left(\tau_1 > \frac{t}{Ku_K}\right) \geq \exp(-t(\bar{n}_x + \varepsilon)\mu(x)b(x)).$$

Since this holds for any $\varepsilon > 0$, we have completed the proof of Lemma 3.1 (c). □

Proof of Lemma 3.2 The proof of this lemma follows the three steps of the invasion of a mutant described at the end of the introduction of this chapter.

Fix $\eta > 0$, $\varepsilon_0 > 0$ and $0 < \varepsilon < \varepsilon_0$. By Lemma 3.1 (a), there exists a constant $\rho > 0$ that we can assume smaller than η, such that, for sufficiently large K,

$$\mathbf{P}^K_{\frac{z_K}{K}\delta_x + \frac{1}{K}\delta_y}\left(\tau_1 < \frac{\rho}{Ku_K}\right) < \varepsilon. \tag{64}$$

Observe that, under $\mathbf{P}^K_{(z_K/K)\delta_x+(1/K)\delta_y}$, for $t \leq \tau_1$,

$$\mathcal{L}((\langle \nu^K,\mathbf{1}_{\{x\}}\rangle,\langle \nu^K,\mathbf{1}_{\{y\}}\rangle)) = \mathbf{Q}^K((1 - u_K\mu(x))b(x),(1 - u_K\mu(y))b(y), d(x),d(y),\alpha(x,x),\alpha(x,y),\alpha(y,x),\alpha(y,y),z_K/K,1/K).$$

Fix K large enough for $u_K < \varepsilon$ and define

$$S^K_\varepsilon := \inf\{s \geq 0 : \langle \nu^K_s,\mathbf{1}_{\{y\}}\rangle \geq \varepsilon\}$$

By Theorem 2.1 (c) and (d), for $t<\tau_1\wedge S^K_\varepsilon$,

$$\begin{gathered}
Z^{K,1}\preceq\langle\nu^K,\mathbf{1}_{\{x\}}\rangle\preceq Z^{K,2}, \qquad (65)\\
\text{where}\quad \mathcal{L}(Z^{K,1})=\mathbf{P}^K((1-\varepsilon)b(x),d(x)+\varepsilon\alpha(x,y),\alpha(x,x),z_K/K)\\
\text{and}\quad \mathcal{L}(Z^{K,2})=\mathbf{P}^K(b(x),d(x),\alpha(x,x),z_K/K).
\end{gathered}$$

Using exactly the same method than led us to (60), we can deduce from Theorem 2.2 (c) that there exists $V>0$ such that

$$\begin{gathered}
\lim_{K\to+\infty}\mathbf{P}(R^K_\varepsilon>e^{KV})=1, \qquad (66)\\
\text{where}\quad R^K_\varepsilon=\inf\{t\geq 0:|Z^{K,i}_t-\bar n_x|>M\varepsilon,\ i=1,2\},
\end{gathered}$$

with $M=3+(b(x)+\alpha(x,y))/\alpha(x,x)$.

Now, observe that, by (65),

$$\forall t\leq\tau_1\wedge S^K_\varepsilon\wedge R^K_\varepsilon,\quad \langle\nu^K_t,\mathbf{1}_{\{x\}}\rangle\in[\bar n_x-M\varepsilon,\bar n_x+M\varepsilon].$$

Therefore, by Theorem 2.1 (b) and (d), for $t\leq\tau_1\wedge S^K_\varepsilon\wedge R^K_\varepsilon$

$$\begin{gathered}
Z^{K,3}\preceq\langle\nu^K,\mathbf{1}_{\{y\}}\rangle\preceq Z^{K,4},\quad\text{where} \qquad (67)\\
\mathcal{L}(Z^{K,3})=\mathbf{P}^K((1-\varepsilon)b(y),d(y)+(\bar n_x+M\varepsilon)\alpha(y,x)+\varepsilon\alpha(y,y),0,1/K)\\
\text{and}\quad \mathcal{L}(Z^{K,4})=\mathbf{P}^K(b(y),d(y)+(\bar n_x-M\varepsilon)\alpha(y,x),0,1/K).
\end{gathered}$$

Define, for any $K\geq 1$, $n\in\mathbb{N}$ and $i\in\{3,4\}$, the stopping time

$$T^{K,i}_{n/K}=\inf\{t\geq 0:Z^{K,i}_t=n/K\}.$$

Observe that, if $T^{K,3}_{\lceil\varepsilon K\rceil/K}<\tau_1\wedge R^K_\varepsilon$,

$$T^{K,4}_{\lceil\varepsilon K\rceil/K}\leq S^K_\varepsilon\leq T^{K,3}_{\lceil\varepsilon K\rceil/K} \qquad (68)$$

and that, if $T^{K,4}_0<T^{K,4}_{\lceil\varepsilon K\rceil/K}\wedge\tau_1\wedge R^K_\varepsilon$,

$$\theta_0\leq T^{K,4}_0.$$

If $Z^{K,4}$ is sub-critical, apply Theorem 2.3 (27), and if $Z^{K,4}$ is super-critical, apply Theorem 2.3 (30) (the critical case can be excluded by slightly changing the value of ε). Since $\log K\ll 1/Ku_K$, we obtain

$$\begin{aligned}
&\lim_{K\to+\infty}\mathbf{P}\left(T^{K,4}_0\leq\frac{\rho}{Ku_K}\wedge T^{K,4}_{\lceil\varepsilon K\rceil/K}\right)\\
&\quad=\frac{d(y)+(\bar n_x-M\varepsilon)\alpha(y,x)}{b(y)}\wedge 1\geq 1-\frac{[f(y,x)]_+}{b(y)}-\frac{\alpha(y,x)}{b(y)}M\varepsilon. \qquad (69)
\end{aligned}$$

Combining (64), (66), (67) and (69), and using the facts that $\rho < \eta$, $\varepsilon < \varepsilon_0$ and $\exp(KV) > \rho/Ku_K$ for sufficiently large K, we obtain, taking K larger if necessary,

$$\begin{aligned}\mathbf{P}\biggl(\theta_0 < \tau_1 \wedge \frac{\eta}{Ku_K},\ V_0 = x \text{ and } |\langle \nu^K_{\theta_0}, \mathbf{1}\rangle - \bar{n}_x| < M\varepsilon_0\biggr) \\ \geq \mathbf{P}\left(\theta_0 < \tau_1 \wedge S^K_\varepsilon \wedge R^K_\varepsilon \wedge \frac{\rho}{Ku_K} \text{ and } V_0 = x\right) \\ \geq \mathbf{P}\left(T^{K,4}_0 < \tau_1 \wedge T^{K,4}_{\lceil \varepsilon K\rceil/K} \wedge R^K_\varepsilon \wedge \frac{\rho}{Ku_K}\right) \\ \geq 1 - \frac{[f(y,x)]_+}{b(y)} - \left(\frac{\alpha(y,x)}{b(y)}M + 3\right)\varepsilon. \qquad (70)\end{aligned}$$

This ends the proof of Lemma 3.2 in the case where $f(y,x) \leq 0$.

Let us assume that $f(y,x) > 0$, *i.e.* that $b(y) - d(y) - \bar{n}_x\alpha(y,x) > 0$. If we choose $\varepsilon > 0$ sufficiently small, then $Z^{K,3}$ is super-critical. By Theorem 2.3 (31),

$$\begin{aligned}\lim_{K\to+\infty} \mathbf{P}\left(T^{K,3}_{\lceil \varepsilon K\rceil/K} < \frac{\rho}{3Ku_K}\right) \\ = \frac{(1-\varepsilon)b(y) - d(y) - (\bar{n}_x + M\varepsilon)\alpha(y,x) - \varepsilon\alpha(y,y)}{(1-\varepsilon)b(y)} \\ \geq \frac{f(y,x)}{(1-\varepsilon)b(y)} - \varepsilon\frac{b(y) + M\alpha(y,x) + \alpha(y,y)}{(1-\varepsilon)b(y)}.\end{aligned}$$

Therefore, by (66) and (64), assuming (without loss of generality) that $\varepsilon < 1/2$, for sufficiently large K,

$$\mathbf{P}\left(T^{K,3}_{\lceil \varepsilon K\rceil/K} < \tau_1 \wedge R^K_\varepsilon \wedge \frac{\rho}{3Ku_K}\right) \geq \frac{f(y,x)}{(1-\varepsilon)b(y)} - M'\varepsilon,$$

where $M' := 2(b(y) + M\alpha(y,x) + \alpha(y,y))/b(y) + 3$. Then, it follows from (68) that

$$\mathbf{P}\left(S^K_\varepsilon < \tau_1 \wedge R^K_\varepsilon \wedge \frac{\rho}{3Ku_K}\right) \geq \frac{f(y,x)}{(1-\varepsilon)b(y)} - M'\varepsilon. \qquad (71)$$

Observe that, on the event $\{S^K_\varepsilon < \tau_1 \wedge R^K_\varepsilon \wedge (\rho/3Ku_K)\}$,

$$\langle \nu^K_{S^K_\varepsilon}, \mathbf{1}_{\{y\}}\rangle = \lceil \varepsilon K\rceil/K \quad \text{and} \quad |\langle \nu^K_{S^K_\varepsilon}, \mathbf{1}_{\{x\}}\rangle - \bar{n}_x| < M\varepsilon. \qquad (72)$$

Now, since we have assumed $f(y,x) > 0$, x and y satisfy (8) and, by Proposition 1.2, any solution to (6) with initial state in the compact set

$[\bar{n}_x - M\varepsilon,\bar{n}_x + M\varepsilon] \times [\varepsilon/2,2\varepsilon]$ converges to $(0,\bar{n}_y)$ when $t \to +\infty$. As in the proof of Lemma 3.1 (b), because of the continuity of the flow of system (6), we can find $t_\varepsilon < +\infty$ large enough such that any of these solutions do not leave the set $[0,\varepsilon^2/2] \times [\bar{n}_y - \varepsilon/2,\bar{n}_y + \varepsilon/2]$ after time t_ε.

Apply Theorem 2.2 (b) on $[0,t_\varepsilon]$, with $C = [\bar{n}_x - M\varepsilon,\bar{n}_x + M\varepsilon] \times [\varepsilon/2,2\varepsilon]$ and with a constant $\delta < \varepsilon^2/2 \wedge r$, where r is defined in (21) (with $T = t_\varepsilon$). Then, with the notations of Theorem 2.2 (b), because of (71) and (72), the Markov property at time S^K_ε yields

$$\liminf_{K\to+\infty} \mathbf{P}\Bigg(S^K_\varepsilon < \tau_1 \wedge R^K_\varepsilon \wedge \frac{\rho}{3Ku_K}, \sup_{S^K_\varepsilon \leq s \leq S^K_\varepsilon + t_\varepsilon} \big\| \big(\langle \nu^K_s,\mathbf{1}_{\{x\}}\rangle,\langle \nu^K_s,\mathbf{1}_{\{y\}}\rangle\big) - \phi_{\langle \nu^K_{S^K_\varepsilon},\mathbf{1}_{\{x\}}\rangle,\langle \nu^K_{S^K_\varepsilon},\mathbf{1}_{\{y\}}\rangle}(s) \big\| \leq \delta \Bigg) \geq \frac{f(y,x)}{(1-\varepsilon)b(y)} - M'\varepsilon. \tag{73}$$

Now, observe that, since $\delta < r$, on the event

$$\Bigg\{ S^K_\varepsilon < \tau_1 \wedge R^K_\varepsilon, \sup_{S^K_\varepsilon \leq s \leq S^K_\varepsilon + t_\varepsilon} \big\| \big(\langle \nu^K_s,\mathbf{1}_{\{x\}}\rangle,\langle \nu^K_s,\mathbf{1}_{\{y\}}\rangle\big) - \phi_{\langle \nu^K_{S^K_\varepsilon},\mathbf{1}_{\{x\}}\rangle,\langle \nu^K_{S^K_\varepsilon},\mathbf{1}_{\{y\}}\rangle}(s) \big\| \leq \delta \Bigg\},$$

for any $t \in [S^K_\varepsilon,S^K_\varepsilon + t_\varepsilon]$, $\langle \nu^K_t,\mathbf{1}_{\{x\}}\rangle \geq r - \delta > 0$ and $\langle \nu^K_t,\mathbf{1}_{\{y\}}\rangle \geq r - \delta > 0$, and thus

$$\theta_0 > S^K_\varepsilon + t_\varepsilon.$$

Therefore, since $\delta < \varepsilon^2/2 < \varepsilon/2$, by (64) and (73), for sufficiently large K,

$$\mathbf{P}\Bigg(S^K_\varepsilon < R^K_\varepsilon \wedge \frac{\rho}{3Ku_K}, \tau_1 > \frac{\rho}{3Ku_K} + t_\varepsilon, \theta_0 > S^K_\varepsilon + t_\varepsilon, \langle \nu^K_{S^K_\varepsilon + t_\varepsilon},\mathbf{1}_{\{x\}}\rangle < \varepsilon^2 \text{ and } \langle \nu^K_{S^K_\varepsilon + t_\varepsilon},\mathbf{1}_{\{y\}}\rangle \in [\bar{n}_y - \varepsilon,\bar{n}_y + \varepsilon] \Bigg) \geq \frac{f(y,x)}{(1-\varepsilon)b(y)} - (M'+2)\varepsilon. \tag{74}$$

Now, we will compare $\langle \nu^K,\mathbf{1}_{\{x\}}\rangle$ with a branching process after time $S^K_\varepsilon + t_\varepsilon$ in order to prove that trait x gets extinct with a very high probability. We will use a method very similar to the one we used in the beginning of the proof of Lemma 3.2. First, on the event inside the probability in (74),

$\langle \nu^K_{S^K_\varepsilon+t_\varepsilon}, \mathbf{1}_{\{x\}}\rangle < \varepsilon^2$. In order to prove that the population with trait x stays small after $S^K_\varepsilon + t_\varepsilon$, since $\varepsilon^2 < \varepsilon$ (we have assumed that $\varepsilon < 1/2$), let us define the stopping time

$$\hat{S}^K_\varepsilon = \inf\{t \geq S^K_\varepsilon + t_\varepsilon : \langle \nu^K_t, \mathbf{1}_{\{x\}}\rangle > \varepsilon\}.$$

Using Theorem 2.1 (c) and (d) again, we see that, on the event

$$F^{K,\varepsilon} := \big\{\langle \nu^K_{S^K_\varepsilon+t_\varepsilon}, \mathbf{1}_{\{x\}}\rangle < \varepsilon^2,\ \langle \nu^K_{S^K_\varepsilon+t_\varepsilon}, \mathbf{1}_{\{y\}}\rangle \in [\bar{n}_y - \varepsilon, \bar{n}_y + \varepsilon]\big\},$$

for any $t \geq 0$ such that $S^K_\varepsilon + t_\varepsilon + t \leq \hat{S}^K_\varepsilon \wedge \tau_1$,

$$\begin{gathered} Z^{K,5}_t \preceq \langle \nu^K_{S^K_\varepsilon+t_\varepsilon+t}, \mathbf{1}_{\{y\}}\rangle \preceq Z^{K,6}_t, \\ \text{where} \quad \mathcal{L}(Z^{K,5}) = \mathbf{P}^K((1-\varepsilon)b(y), d(y) + \varepsilon\alpha(y,x), \alpha(y,y), \lfloor (\bar{n}_y - \varepsilon)K\rfloor/K) \\ \text{and} \quad \mathcal{L}(Z^{K,6}) = \mathbf{P}^K(b(y), d(y), \alpha(y,y), \lceil (\bar{n}_y + \varepsilon)K\rceil/K). \end{gathered}$$

We can apply Theorem 2.2 (c) to $Z^{K,5}$ and $Z^{K,6}$ as above to obtain a constant $V' > 0$ such that

$$\lim_{K\to+\infty} \mathbf{P}(\hat{R}^K_\varepsilon > e^{KV'}) = 1, \tag{75}$$
$$\text{where} \quad \hat{R}^K_\varepsilon = \inf\{t \geq 0 : |Z^{K,i}_t - \bar{n}_y| > M''\varepsilon,\ i = 5,6\},$$

with $M'' = 3 + (b(y) + \alpha(y,x))/\alpha(y,y)$.

Observe that, on the event $F^{K,\varepsilon}$, for any $t \leq \hat{R}^K_\varepsilon$ such that $S^K_\varepsilon + t_\varepsilon + t \leq \hat{S}^K_\varepsilon \wedge \tau_1$,

$$|\langle \nu^K_{S^K_\varepsilon+t_\varepsilon+t}, \mathbf{1}_{\{y\}}\rangle - \bar{n}_y| \leq M''\varepsilon,$$

and so, by Theorem 2.1 (c) and (e), on $F^{K,\varepsilon}$ and for t as above,

$$\begin{gathered} \langle \nu^K_{S^K_\varepsilon+t_\varepsilon+t}, \mathbf{1}_{\{x\}}\rangle \preceq Z^{K,7}_t \\ \text{where} \quad \mathcal{L}(Z^{K,7}) = \mathbf{P}^K(b(x), d(x) + (\bar{n}_y - M''\varepsilon)\alpha(x,y), 0, \lceil \varepsilon^2 K\rceil/K). \end{gathered}$$

Now, since x and y satisfy (8), $b(x) - d(x) - \bar{n}_y\alpha(x,y) < 0$, and thus $Z^{K,7}$ is sub-critical for sufficiently small ε. Fix such an $\varepsilon > 0$ and define for any $n \geq 0$

$$\hat{T}^K_{n/K} = \inf\{t \geq 0 : Z^{K,7}_t = n/K\}.$$

If $\hat{T}^K_{\lceil \varepsilon K\rceil/K} \leq \hat{R}^K_\varepsilon$ and $S^K_\varepsilon + t_\varepsilon + \hat{T}^K_{\lceil \varepsilon K\rceil/K} \leq \tau_1$, then

$$\hat{S}^K_\varepsilon \geq S^K_\varepsilon + t_\varepsilon + \hat{T}^K_{\lceil \varepsilon K\rceil/K}$$

and if $\hat{T}^K_0 \leq \hat{R}^K_\varepsilon$ and $S^K_\varepsilon + t_\varepsilon + \hat{T}^K_0 \leq \hat{S}^K_\varepsilon \wedge \tau_1$, then

$$\theta_0 \leq \hat{T}^K_0.$$

Moreover, by Theorem 2.3 (28) and (29), for sufficiently large K,

$$\mathbf{P}\left(\hat{T}_0^K \leq \frac{\rho}{3Ku_K}\right) \geq 1-\varepsilon$$
$$\text{and} \quad \mathbf{P}(\hat{T}^K_{\lceil K\varepsilon\rceil/K} \leq \hat{T}_0^K) \leq 2\varepsilon.$$

Combining the last two inequalities with (64), (74) and (75), and reminding that $\rho<\eta$ and $\varepsilon<\varepsilon_0$, we finally obtain, for sufficiently large K,

$$\begin{aligned}
&\mathbf{P}\left(\theta_0<\tau_1\wedge\frac{\eta}{Ku_K},\, V_0=y \text{ and } |\langle\nu^K_{\theta_0},\mathbf{1}\rangle-\bar{n}_y|<M''\varepsilon_0\right)\\
&\geq \mathbf{P}\left(S^K_\varepsilon<R^K_\varepsilon\wedge\frac{\rho}{3Ku_K},\theta_0>S^K_\varepsilon+t_\varepsilon,\tau_1>\frac{2\rho}{3Ku_K}+t_\varepsilon,\langle\nu^K_{S^K_\varepsilon+t_\varepsilon},\mathbf{1}_{\{x\}}\rangle<\varepsilon^2,\right.\\
&\qquad\left.\langle\nu^K_{S^K_\varepsilon+t_\varepsilon},\mathbf{1}_{\{y\}}\rangle\in[\bar{n}_y-\varepsilon,\bar{n}_y+\varepsilon],\,\hat{T}^K_0<\frac{\rho}{3Ku_K}\wedge\hat{T}^K_{\lceil K\varepsilon\rceil/K} \text{ and } \hat{R}^K_\varepsilon>\frac{\rho}{Ku_K}\right)\\
&\geq\frac{f(y,x)}{(1-\varepsilon)b(y)}-(M'+7)\varepsilon.
\end{aligned}$$

Adding this inequality with (70), we obtain

$$\mathbf{P}\left(\theta_0<\tau_1\wedge\frac{\eta}{Ku_K}\right)\geq 1-\frac{\varepsilon}{1-\varepsilon}\frac{f(y,x)}{b(y)}-\left(M\frac{\alpha(y,x)}{b(y)}+M'+10\right)\varepsilon\geq 1-M'''\varepsilon,$$

where $M'''=2f(y,x)/b(y)+M\alpha(y,x)/b(y)+M'+10$ (remind that $\varepsilon<1/2$), which implies (43), and

$$\mathbf{P}\big(|\langle\nu^K_{\theta_0},\mathbf{1}\rangle-\bar{n}_{V_0}|<(M\vee M'')\varepsilon_0\big)\geq 1-M'''\varepsilon,$$

which implies (44).

Therefore,

$$\mathbf{P}(V_0=x)\geq 1-\frac{f(y,x)}{b(y)}-2M'''\varepsilon \quad\text{and}\quad \mathbf{P}(V_0=y)\geq\frac{f(y,x)}{(1-\varepsilon)b(y)}-2M'''\varepsilon.$$

Since $\mathbf{P}(V_0=x)\leq 1-\mathbf{P}(V_0=y)$, we finally obtain (41) and (42). □

Chapter III

Convergence of adaptive dynamics jump processes to the canonical equation and degenerate diffusion approximation

Abstract: We consider a jump process describing the evolution of a n-morphic population (where n types co-exist), generalizing the biological model of trait substitution sequence [59, 18]. Scaling the size of the jumps by a small parameter $\varepsilon > 0$, a non-standard rescaling of time (by $1/\varepsilon^2$) induces a convergence to an EDO generalizing the biological canonical equation of adaptive dynamics [18]. A first order development of the infinitesimal generator of this jump process when $\varepsilon \to 0$ leads to a biologically motivated diffusion process. Its coefficients have non-standard regularity (discontinuous drift and degenerate diffusion at the same points). The weak existence of such a diffusion is a difficult problem, which we can only solve in particular cases (cf. chapter IV).

1 Introduction

Adaptive dynamics (Hofbauer and Sigmund [41], Marrow *et al.* [57] and Metz *et al.* [61]) is a recent branch of evolutionary ecology that proposes new models of Darwinian evolution, involving three mechanisms: reproduction, that transmits the trait through generations, mutations, that generate variability in the trait values, and selection between traits, that results from ecological interactions between individuals and their environment.

We consider a population subject to *clonal reproduction* (*e.g.* reproduction in asexual) where each individual is characterized by a quantitative *adaptive trait*: a real vector representing individual features affecting reproduction and survival, and subject to mutation, such as individual size, age at maturity, size of prefered preys, pathogen virulence, etc. Each clone has the same trait as its progenitor, unless affected by a *mutation*. When a mutant appears, the background population is called *resident*, and the *invasion* of the mutant trait occurs if the mutant population significantly increases in frequency with respect to the resident one. On a long time scale, when mutations are rare, the evolution of the population can be described by a succession of invasions of mutants in the population, occuring at a very short time scale compared to the time scale of mutations. This heuristic, which has been mathematically justified in a simple case in chapter II, leads to biological models known as “trait substitution sequences” (Metz *et al.* [59]), in which the *population state* is described by the set of trait values represented in the population, and the evolution proceeds by jumps of the population state.

We assume that the trait space $\mathcal{X}$ is a *convex open* subset of $\mathbb{R}^d$ for some $d \geq 1$. A population in which all individuals bear the same trait value is called *monomorphic*. A non-monomorphic population is called *polymorphic*. Among these are n-*morphic* populations, in which the individual trait assumes exactly n different values. We will study in this article a trait substitution sequence model generalizing the monomorphic one (cf. Metz *et al.* [59] and Dieckmann and Law [18]) to a n-morphic population (cf. chapter I). This $\mathcal{X}^n$-valued Markov jump process is described in section 2, where a precise construction is provided.

An important advance of adaptive dynamics theory is the so-called “canonical equation” (Dieckmann and Law [18]), proposed as an approximation of the trait substitution sequence model in the monomorphic case under the assumption of small mutations (*i.e.* the trait of a mutant is only slightly different from the trait of its progenitor). In section 3, we introduce a small parameter $\varepsilon > 0$ scaling the size of the mutation jumps, and we prove the convergence of the jump process of section 2 to a deterministic EDO when $\varepsilon \to 0$ and when the time is rescaled in an unusual way (by $1/\varepsilon^2$), because of the particularities of the biological model. This provides a mathematical justification and a generalization of the canonical equations of [18] and Champagnat *et al.* [10]. This convergence result will be obtained by, first, proving the tightness of the laws of the jump processes, and then, identifying the limit law thanks to the martingale problem method.

In section 4, we obtain a new model of adaptive dynamics, a biologically motivated diffusion process, by expending to the first order the infinitesimal generator of the jump process of section 2 when $\varepsilon \to 0$. We obtain a diffusion operator with *degenerate* diffusion parameter at isolated points of the space and *non-continuous* drift parameter at the same points. The classical theory of diffusion processes does not apply, and in particular, the weak existence is a very difficult problem. We are able to prove it only in the monomorphic case ($n = 1$ but $d \geq 1$), and when the mutation law is symmetrical, in chapter IV.

Notations 1.1 *$\mathcal{B}_b(E)$ (resp. $\mathcal{C}^k(E)$,$\mathcal{C}^k_b(E)$,$\mathcal{C}^k_c(E)$) denotes the set of functions from E to $\mathbb{R}$ that are Borel and bounded (resp. of class $\mathcal{C}^k$, resp. bounded, of class $\mathcal{C}^k$ with bounded i^{th} order derivatives, for $1 \leq i \leq k$, resp. of class $\mathcal{C}^k$ with compact support).*

2 n-morphic trait substitution sequence

Our n-morphic trait substitution sequence model is based on the same heuristic that classical trait substitution sequences models (see [59] and chapter I). We will not describe it here.

Consider a n-morphic population with trait values $x_1,\dots,x_n$. We will denote by $\mathbf{x} = (x_1,\dots,x_n) \in \mathcal{X}^n$ the state of the population. Let use introduce the following biological parameters:

$u_i(\mathbf{x}) \geq 0$ is the equilibrium size (density of individuals) of the sub-population of trait $x_i \in \mathcal{X}$ in the population $\mathbf{x} \in \mathcal{X}^n$ ($1 \leq i \leq n$).

$b(y;\mathbf{x}) \geq 0$ is the rate of birth for a single individual with trait $y \in \mathcal{X}$ in a population $\mathbf{x} \in \mathcal{X}^n$ at its equilibrium $(u_1(\mathbf{x}),\dots,u_n(\mathbf{x}))$.

$d(y;\mathbf{x}) > 0$ is the rate of death of a single individual with trait $y \in \mathcal{X}$ in a resident population $\mathbf{x} \in \mathcal{X}^n$ at its equilibrium $(u_1(\mathbf{x}),\dots,u_n(\mathbf{x}))$.

$\mu(y) \in [0,1]$ is the probability that a mutation occurs in a birth from an individual with trait $y \in \mathcal{X}$

$p(y,dh)$ is the law of $h = y' - y$, where $y' \in \mathcal{X}$ is a mutant trait born from an individual with trait $y \in \mathcal{X}$. It is a probability measure on $\mathbb{R}^d$, and since y' must belong to $\mathcal{X}$, the support of $p(x,\cdot)$ is a subset of $\overline{\mathcal{X}} - y = \{y' - y; y' \in \overline{\mathcal{X}}\}$.

We have assumed $\mathcal{X}$ open, like in many biological models. This corresponds to the biological assumption that mutations are possible in any direction of the trait space from any viable traits. Extreme traits are non-viable, so that, when one trait in the population reaches the boundary of $\mathcal{X}$, this corresponds to an extinction of this trait. Therefore, we will assume that $b(y,\mathbf{x}) = 0$ for any $y \in \partial\mathcal{X}$, $u_i(\mathbf{x}) = 0$ if $x_i \in \partial\mathcal{X}$, and $p(y,dh) = \delta_0(dh)$ when $y \in \underline{\mathcal{X}}$. Moreover, in order to be able to define the process on the Polish space $\overline{\mathcal{X}}^n$, we will assumed that b and d are actually defined on $\overline{\mathcal{X}} \times \overline{\mathcal{X}}^n$, u_i on $\overline{\mathcal{X}}^n$, and μ on $\overline{\mathcal{X}}$.

We will not give the precise biological definition of these parameters. However, the following relation will be important below: when the population $\mathbf{x}$ is at its equilibrium $(u_1(\mathbf{x}), \dots, u_n(\mathbf{x}))$, the birth and death rate of each individual are equal, *i.e.*

$$\forall \mathbf{x} = (x_1, \dots, x_n) \in \overline{\mathcal{X}}^n, \quad \forall i \in \{1, \dots, n\}, \quad b(x_i; \mathbf{x}) = d(x_i; \mathbf{x}). \tag{1}$$

Let us define, for $1 \leq i \leq n$,

$$g_i(y; \mathbf{x}) = \mu(x_i) b(x_i; \mathbf{x}) u_i(\mathbf{x}) \frac{b(y; \mathbf{x}) - d(y; \mathbf{x})}{b(y; \mathbf{x})}$$

when $b(y; \mathbf{x}) > 0$, and $g_i(y; \mathbf{x}) = 0$ otherwise. Biologically, $g_i(y; \mathbf{x})$ represents a measure of the selective advantage of a mutant y born from x_i in the population $\mathbf{x} = (x_1, \dots, x_n)$ (see [59]). It follows from (1) that

$$\forall \mathbf{x} = (x_1, \dots, x_n) \in \overline{\mathcal{X}}^n, \quad \forall i \in \{1, \dots, n\}, \quad g_i(x_i; \mathbf{x}) = 0. \tag{2}$$

The biological heuristic (see chapter I) leads to a description of the evolution by a $\overline{\mathcal{X}}^n$-valued jump process with infinitesimal generator given, for any bounded and measurable function $\varphi : \overline{\mathcal{X}}^n \to \mathbb{R}$, by

$$L\varphi(\mathbf{x}) = \sum_{i=1}^{n} \int_{\mathbb{R}^d} (\varphi(\mathbf{x} + (h)_i) - \varphi(\mathbf{x}))[g_i(x_i + h; \mathbf{x})]_+ p(x_i, dh). \tag{3}$$

with the notation $(h)_i = (0, \dots, 0, h, 0, \dots, 0)$ where h is at the i^{th} coordinate.

If, for $1 \leq i \leq n$, $[g_i]_+$ is bounded, this generator defines a unique semigroup, so that the process is unique in law (see Ethier and Kurtz [28]). We will now prove its existence by giving an explicit pathwise representation under the following assumptions:

(Ha) $p(x, \cdot)$ has finite and bounded first order moment, and is absolutely continuous with respect to some measure ν on $\mathbb{R}^d$ for any $x \in \overline{\mathcal{X}}$, *i.e.*

$$\forall x \in \mathcal{X}, \quad p(x, dh) = p(x, h)\nu(dh).$$

Moreover, there is some function $p : \mathbb{R}^d \to \mathbb{R}$ such that $p(h)\nu(dh)$ is a σ-finite measure on $\mathbb{R}^d$ and such that

$$\forall x \in \mathcal{X}, \quad p(x, h) \leq p(h).$$

(Hb) For all $i \in \{1, \dots, n\}$, $[g_i]_+$ is bounded by some constant κ on $\overline{\mathcal{X}} \times \overline{\mathcal{X}}^n$.

Note that, under (Ha), when $\mathcal{X} \neq \mathbb{R}^d$, $\nu(\{0\}) > 0$ since $p(x,dh) = \delta_0(dh)$ when $x \in \partial\mathcal{X}$.

In these conditions, following on Fournier and Méléard [33], it is possible to give an explicit construction of a process $(X(t),t \geq 0)$ generated by (3):

Definition 2.1 *Let* $(\Omega,\mathcal{F}_t,\mathbf{P})$ *be a sufficiently rich filtered probability space. Let us consider on this probability space* n *independent Poisson point measures* $N_i(dh,d\theta,ds)$ *on* $\mathbb{R}^d \times [0,1] \times \mathbb{R}_+$ *with intensity*

$$q_i(dh,d\theta,ds) = p(h)\nu(dh)\kappa d\theta ds \quad \text{for } 0 \leq i \leq n.$$

Let us also consider a random variable X_0 *on* $\overline{\mathcal{X}}^n$*, independent of* N_i *for* $1 \leq i \leq n$*.*

Then, assuming (Ha) and (Hb), we can define for any $\omega \in \Omega$ *and* $t \geq 0$

$$X(t,\omega) = X_0(\omega) + \sum_{i=1}^{n} \int_0^t \int_0^1 \int_{\mathbb{R}^d} (h)_i \mathbf{1}_{\left\{\theta \leq \frac{[g_i(x_i(s-,\omega)+h;X(s-,\omega))]_+}{\kappa} \frac{p(x_i(s-,\omega),h)}{p(h)}\right\}} N_i(dh,d\theta,ds)(\omega),$$

where $X(t,\omega) = (x_1(t,\omega),\dots,x_n(t,\omega))$*. Note that, in the expression above,* g_i *may be computed at points out of its domain, when* h *is such that* $x_i(s-,\omega)+h \notin \overline{\mathcal{X}}$*. But then,* $p(x_i(s-,\omega),h) = 0$*, so let us admit the convention that, in this case,* $[g_i(x_i(s-,\omega)+h;X(s-,\omega))]_+ p(x_i(s-,\omega),h) = 0$*.*

Note that the Poisson point measure N_i is well defined since $p(h)\nu(dh)$ is a σ-finite measure, and that the process $(X(t),t \geq 0)$ is well defined under (Ha) and (Hb) since

$$\begin{aligned}\int_0^t \int_0^1 \int_{\mathbb{R}^d} \|h\| \mathbf{1}_{\left\{\theta \leq \frac{[g_i(x_i(s-)+h;X(s-))]_+}{\kappa} \frac{p(x_i(s-),h)}{p(h)}\right\}} q_i(dh,d\theta,ds) \\ \leq \int_0^t \int_{\mathbb{R}^d} \|h\| [g_i(x_i(s-)+h;X(s-))]_+ p(x_i(s-),h)\nu(dh)ds < +\infty.\end{aligned}$$

We can show:

Proposition 2.1 *Assume (Ha) and (Hb). Then,* L*, defined in (3), is the infinitesimal generator of the Markov process* $(X(t),t \geq 0)$ *of Definition 2.1.*

Proof For any $\varphi \in \mathcal{B}(\overline{\mathcal{X}}^n)$ and for $X_0 = \mathbf{x} \in \overline{\mathcal{X}}^n$, Itô's formula for jump processes writes

$$\varphi(X(t)) = \varphi(\mathbf{x}) + \sum_{i=1}^{n} \int_0^t \int_0^1 \int_{\mathbb{R}^d} [\varphi(X(s-)+(h)_i) - \varphi(X(s-))] \times \mathbf{1}_{\left\{\theta \leq \frac{[g_i(x_i(s-)+h;X(s-))]_+}{\kappa} \frac{p(x_i(s-),h)}{p(h)}\right\}} N_i(dh,d\theta,ds).$$

Taking the expectation, we get

$$\begin{aligned}\mathbf{E}[\varphi(X(t))-\varphi(\mathbf{x})] &= \sum_{i=1}^{n}\mathbf{E}\left[\int_0^t\int_0^1\int_{\mathbb{R}^d}[\varphi(X(s-)+(h)_i)-\varphi(X(s-))]\right.\\ &\qquad\left.\mathbf{1}_{\left\{\theta\leq\frac{[g_i(x_i(s-)+h;X(s-))]_+}{\kappa}\frac{p(x_i(s-),h)}{p(h)}\right\}}q_i(dh,d\theta,ds)\right]\\ &= \sum_{i=1}^{n}\mathbf{E}\left[\int_0^t\int_{\mathbb{R}^d}[\varphi(X(s-)+(h)_i)-\varphi(X(s-))]\right.\\ &\qquad\left.[g_i(x_i(s-)+h;X(s-))]_+p(x_i(s-),dh)ds\right]\end{aligned}$$

and it is straightforward to show that $(\mathbf{E}[\varphi(X(t))]-\varphi(\mathbf{x}))/t\to L\varphi(\mathbf{x})$ when $t\to 0$. $\square$

3 Polymorphic canonical equation of adaptive dynamics

We now intend to apply to the process X the biological limit of small mutations.

3.1 Rescaled process

Let us assume that $p(y,\cdot)$ has finite second order moments, and let us denote by $K(y)$ its covariance matrix:

$$K(y)=[k_{ij}(y)]_{1\leq i,j\leq d},$$
$$\text{where } k_{ij}(y)=\int_{\mathbb{R}^d}h_ih_jp(y,dh)-\int_{\mathbb{R}^d}h_ip(y,dh)\int_{\mathbb{R}^d}h_jp(y,dh)$$
$$\text{for } i \text{ and } j \text{ in } \{1,\dots,d\},\text{ with } h=(h_1,\dots,h_d).$$

We introduce a parameter $\varepsilon>0$ and we replace $p(y,dh)$ by $p(y,dh/\varepsilon)$ (*i.e.* the image of the measure $p(y,\cdot)$ by the function $h\mapsto\varepsilon h$). Then,

$$\int_{\mathbb{R}^d}h_ih_jp(y,dh/\varepsilon)=\varepsilon^2k_{ij}(y)$$

goes to zero as $\varepsilon\to 0$.

So, assuming (Ha) and (Hb), let us consider, on the probability space $(\Omega,\mathcal{F},\mathbf{P})$ of Definition 2.1, a family of random variables $\{X_0^\varepsilon\}_{0<\varepsilon\leq 1}$, and a family of Markov jump processes $\{(X^\varepsilon_t, t \geq 0)\}_{0<\varepsilon\leq 1}$, with paths in the Skorohod's space $\mathbb{D}(\mathbb{R}_+, \overline{\mathcal{X}}^n)$ of right continuous and left limited functions from $\mathbb{R}_+$ to $\overline{\mathcal{X}}^n$, defined by

$$X^\varepsilon_t = X^\varepsilon_0 + \sum_{i=1}^{n} \int_0^t \int_0^1 \int_{\mathbb{R}^d} (\varepsilon h)_i \mathbf{1}_{\left\{\theta \leq \frac{[g_i(x^\varepsilon_i(s-)+\varepsilon h; X^\varepsilon_{s-})]_+}{\kappa} \frac{p(x^\varepsilon_i(s-),h)}{p(h)}\right\}} N_i\left(dh, d\theta, \frac{ds}{\varepsilon^2}\right),$$

where we used the notation $X^\varepsilon_t = (x^\varepsilon_1(t), \ldots, x^\varepsilon_n(t))$. Since the jumps of X^ε get smaller as $\varepsilon \to 0$, we need to accelerate time, by replacing $N_i(dh, d\theta, ds)$ by $N_i(dh, d\theta, ds/\varepsilon^2)$. This unusual rescaling of time comes from the fundamental biological property (2).

Since $\overline{\mathcal{X}}$ is a convex set, as long as $\varepsilon \leq 1$, the jumps governed by the measure $p(x, dh/\varepsilon)$ cannot make the process go out of the state space $\overline{\mathcal{X}}^n$.

Remark 3.1 *This is the only place where the assumption that $\mathcal{X}$ is convex is used. In fact, it would be sufficient to assume that $\exists \varepsilon_0 > 0$ such that $\forall \varepsilon < \varepsilon_0$, $\forall x \in \overline{\mathcal{X}}$, $\varepsilon Supp(p(x,\cdot)) \subset \overline{\mathcal{X}}$, where $\varepsilon Supp(p(x,\cdot)) = \{\varepsilon y; y \in Supp(p(x,\cdot))\}$, and where $Supp(p(x,\cdot))$ is the support of the measure $p(x,\cdot)$. However, almost all the biological models involve convex sets of traits.*

One can prove, exactly as in Proposition 2.1, that:

Proposition 3.1 *Assuming (Ha) and (Hb), the infinitesimal generator of X^ε writes for $\varphi \in \mathcal{B}_b(\overline{\mathcal{X}}^n)$*

$$L^\varepsilon \varphi(\mathbf{x}) = \frac{1}{\varepsilon^2} \sum_{i=1}^{n} \int_{\mathbb{R}^d} (\varphi(\mathbf{x} + (\varepsilon h)_i) - \varphi(\mathbf{x}))[g_i(x_i + \varepsilon h; \mathbf{x})]_+ p(x_i, dh). \qquad (4)$$

3.2 Convergence result

We first need to define metrics on sets of probability measures:

Definition 3.1 *For any $k \geq 1$, let ρ_k be the Kantorovich metric (see Rachev [65]) defined on the set of probability measures on some closed subset S of $\mathbb{R}^d$ with finite k^{th} order moments, given by*

$$\rho_k(P_1, P_2) = \inf \int_{S^2} c_k(x, y) R(dx, dy),$$

where the infimum is taken over the set of measures $R(dx, dy)$ on S^2 with marginales $P_1(dx)$ and $P_2(dy)$, and where

$$c_k(x, y) = \|x - y\| \sup\{\|x\|^{k-1}, \|y\|^{k-1}\}.$$

We will use the following property of the metric ρ_k:

Proposition 3.2 *Consider a family* $\{q(x,\cdot),x \in D\}$ *of probability measures on a closed* $S \subset \mathbb{R}^d$ *having finite* k^{th} *order moments, where* $D \subset \mathbb{R}^m$*. If* $x \mapsto q(x,\cdot)$ *is continuous (resp. Lipschitz) on* D *for the metric* ρ_k*, then, for any continuous function* $\psi : S \to \mathbb{R}_+$ *such that* $|\psi(h) - \psi(h')| \leq Kc_k(h,h')$ *for some constant* K*, the function* $x \to \int_S \psi(h)q(x,dh)$ *is continuous (resp. Lipschitz) on* D*.*

Proof Remind the following consequence of the dual form of the Kantorovich-Rubinstein metrics (Rachev [65]): for any probability measures P_1 and P_2 on S having finite k^{th} order moments,

$$\sup\left\{\int_S \psi d(P_1 - P_2) : \psi \in \mathcal{C}_b(S) \text{ such that} \right.$$
$$\left. \forall x,y \in S,\ |\psi(x) - \psi(y)| \leq c_k(x,y)\right\} \leq \rho_k(P_1,P_2).$$

Take q and ψ as in the statement. For all $N > 0$ and $a \in \mathbb{R}$, set $\xi_N(a) = a\mathbf{1}_{-N\leq a\leq N} + N\mathbf{1}_{a>N} - N\mathbf{1}_{a<-N}$ and $\psi_N = \xi_N \circ \psi$. ψ_N is bounded and continuous, and satisfies $|\psi_N(h) - \psi_N(h')| \leq Kc_k(h,h')$ for all h and h' in S, so, by the preceding inequality, for all x and x' in D and $N > 0$,

$$\left|\int_S \psi_N(h)(q(x,dh) - q(x',dh))\right| \leq K\rho_k(q(x,\cdot),q(x',\cdot)).$$

Now, by the dominated convergence theorem (since $q(x,\cdot)$ and $q(x',\cdot)$ have finite k^{th} order moments), letting $N \to +\infty$, we get

$$\left|\int_S \psi(h)(q(x,dh) - q(x',dh))\right| \leq K\rho_k(q(x,\cdot),q(x',\cdot)),$$

and this gives the result both when $x \mapsto q(x,\cdot)$ is continuous or Lipschitz. $\square$

We will need the following additional assumptions. We will, according to cases, assume (Hd) or (He).

(Hc) For all $i \in \{1,\dots,n\}$, $g_i(y;\mathbf{x})$ is continuous on $\overline{\mathcal{X}}^{n+1}$ and $\mathcal{C}^1$ with respect to the first vector y, and $\nabla_1 g_i$, the gradient vector of g_i with respect to the first variable, is bounded and Lipschitz on $\overline{\mathcal{X}}^{n+1}$.

(Hd) The probability measure $p(y,\cdot)$ has finite and bounded third order moments on $\overline{\mathcal{X}}$, and $y \mapsto p(y,dh)$ is Lipschitz from $\overline{\mathcal{X}}$ to the set of probability measures on $\mathbb{R}^d$ for the metric ρ_2.

(He) The probability measure $p(y,\cdot)$ has finite and bounded third order moments, and its covariance matrix $K(y)$ has Lipschitz entries on $\mathcal{X}$.

Proposition 3.2 shows that (Hd) implies (He).

In the case where $\mathcal{X} = \mathbb{R}^d$ and ν is the Lebesgue measure on $\mathbb{R}^d$, assumption (Hd) is implied by the following Lipschitz-like simpler condition

$$\forall x,x',h \in \mathbb{R}^d,\ |p(x,h) - p(x',h)| \leq \|x - x'\| m(h),$$
$$\text{with } \int_{\mathbb{R}^d} (1 \vee \|h\|^3) m(h)dh < +\infty.$$

Now, let us state the following convergence result:

Theorem 3.1 *Assume (Ha), (Hb), (Hc) and (Hd). Suppose also that the family of initial states* $\{X_0^\varepsilon\}_{0<\varepsilon\leq 1}$ *is bounded in* $\mathbb{L}^1(\overline{\mathcal{X}}^n)$ *and converges in law to a random variable* X_0 *as* $\varepsilon \to 0$.

Then X^ε *converges when* $\varepsilon \to 0$*, for the Skorohod topology of* $\mathbb{D}(\mathbb{R}_+,\overline{\mathcal{X}}^n)$*, to the process* $(\mathbf{x}(t) = (x_1(t), \dots, x_n(t)), t \geq 0)$ *with initial state* X_0 *and which deterministic sample paths are the unique solution to the system of differential equations*

$$\frac{dx_i}{dt} = \int_{\mathbb{R}^d} h[h \cdot \nabla_1 g_i(x_i;\mathbf{x})]_+ p(x_i,dh) \qquad \text{for } 1 \leq i \leq n. \tag{5}$$

In the case where $p(y,\cdot)$ *is a symmetrical measure on* $\mathbb{R}^d$ *for all* $y \in \mathcal{X}$*, this convergence holds under the less restrictive assumptions (Ha), (Hb), (Hc) and (He), and equation (5) gets the simpler form*

$$\frac{dx_i}{dt} = \frac{1}{2} K(x_i) \nabla_1 g_i(x_i;\mathbf{x}) \qquad \text{for } 1 \leq i \leq n. \tag{6}$$

In the case of monomorphic evolution ($n = 1$), we recover the classical form of the canonical equation heuristically introduced in [18] (cf. also chapter I).

3.3 Proof of Theorem 3.1

The proof of Theorem 3.1 is divided in four steps: first, we prove that the solutions to (5) exist and are unique, second, that the laws of X^ε form a tight family, third, that a uniform convergence result holds for the generators, and finally, using the martingale problem associated with X^ε, that any accumulation point of the laws of X^ε when $\varepsilon \to 0$ is solution to (5). This will yield the required convergence (see e.g. Jacod and Shiryaev [43]).

Let us first introduce the infinitesimal generator L^0 of $\mathbf{x}(t)$: for $\varphi \in \mathcal{C}^1(\overline{\mathcal{X}}^n)$ and for a non-random $X_0 \in \overline{\mathcal{X}}^n$,

$$L^0\varphi(\mathbf{x}) = A(\mathbf{x}) \cdot \nabla\varphi(\mathbf{x}), \tag{7}$$

where $A(\mathbf{x}) = (a_1(\mathbf{x}), \ldots, a_n(\mathbf{x})) \in (\mathbb{R}^d)^n$ is defined by

$$a_i(\mathbf{x}) = \int_{\mathbb{R}^d} h[h \cdot \nabla_1 g_i(x_i;\mathbf{x})]_+ p(x_i,dh).$$

Existence and uniqueness for (5) in $\overline{\mathcal{X}}^n$ By the Cauchy-Lipschitz Theorem, it suffices to prove that $\mathbf{x} \mapsto A(\mathbf{x})$ is Lipschitz on $\overline{\mathcal{X}}^n$. If this is established, since $a_i(\mathbf{x}) = 0$ as soon as $x_i \in \partial\mathcal{X}$ ($p(x_i,dh) = \delta_0(dh)$), the unique solution to (5) cannot leave $\overline{\mathcal{X}}^n$.

$$\begin{aligned}|a_i(\mathbf{x}) - a_i(\mathbf{x}')| \leq \int_{\mathbb{R}^d} \|h\| \times \big|[h \cdot \nabla_1 g_i(x_i;\mathbf{x})]_+ - [h \cdot \nabla_1 g_i(x_i';\mathbf{x}')]_+\big| p(x_i,dh) \\ + \left|\int_{\mathbb{R}^d} h[h \cdot \nabla_1 g_i(x_i';\mathbf{x}')]_+ (p(x_i,dh) - p(x_i',dh))\right|.\end{aligned} \tag{8}$$

Using the facts that $|[a]_+ - [b]_+| \leq |a-b|$, that $\nabla_1 g_i(x_i;\mathbf{x})$ is bounded by some constant K and K-Lipschitz, and that $p(x,\cdot)$ has second order moments bounded by M_2, we get

$$\int_{\mathbb{R}^d} \|h\| \times \big|h \cdot (\nabla_1 g_i(x_i;\mathbf{x}) - \nabla_1 g_i(x_i';\mathbf{x}'))\big| p(x_i,dh) \leq K(2\|\mathbf{x} - \mathbf{x}'\|)M_2,$$

where M_2 is a bound for the second order moments of $p(y,\cdot)$.

If we denote by ξ the vector $\nabla_1 g_i(x_i;\mathbf{x})$, then

$$\begin{aligned}\|h[h\cdot\xi]_+ - h'[h'\cdot\xi]_+\| &\leq \|(h-h')[h\cdot\xi]_+\| + \|h'([h\cdot\xi]_+ - [h'\cdot\xi]_+)\| \\ &\leq \|h-h'\| \times |h\cdot\xi| + \|h'\| \times |[h\cdot\xi]_+ - [h'\cdot\xi]_+| \\ &\leq \|\xi\| c_2(h,h') + \|h'\| \times |h\cdot\xi - h'\cdot\xi| \leq 2\|\xi\| c_2(h,h').\end{aligned}$$

Therefore, by Proposition 3.2 for $k = 2$, the second term of the right-hand side of (8) is bounded by some constant times $\|x_i - x_i'\|$. Hence $(\mathbf{x}(t), t \geq 0)$ is well defined.

In the case of symmetrical mutations, a simple computation shows that

$$\begin{aligned}\int_{\mathbb{R}^d} h[h \cdot \nabla_1 g_i(x_i;\mathbf{x})]_+ p(x_i,dh) &= \frac{1}{2}\int_{\mathbb{R}^d} h(h \cdot \nabla_1 g_i(x_i;\mathbf{x})) p(x_i,dh) \\ &= \frac{1}{2} K(x_i) \nabla_1 g_i(x_i;\mathbf{x}),\end{aligned}$$

so that equation (5) writes (6). To prove that this function is Lipschitz, one only needs assumptions (Hc) and (He) instead of (Hc) and (Hd).

Tightness of $\{\mathbf{P}^{\varepsilon}_{X_0^{\varepsilon}}\}_{\varepsilon>0}$ Let us denote by $\mathbf{P}^{\varepsilon}_{X_0^{\varepsilon}}$ the law on $\mathbb{D}(\mathbb{R}_+,\overline{\mathcal{X}}^n)$ of the jump process X^{ε} with initial random state X_0^{ε}. We will use the Aldous criterion [1] to prove the tightness of $\{\mathbf{P}^{\varepsilon}_{X_0^{\varepsilon}}\}_{\varepsilon>0}$.

Let $\{\tau_{\varepsilon}\}$ be a family of stopping times and $\{\delta_{\varepsilon}\}$ be a family of positive numbers converging to 0 when $\varepsilon \to 0$. Using the fact that $|g_i(x_i+\varepsilon h;\mathbf{x})| \leq \varepsilon K\|h\|$ (because of (2) and of the fact that $\nabla_1 g_i$ is bounded by K),

$$\begin{aligned}
&\mathbf{E}[\|X^{\varepsilon}_{\tau_\epsilon+\delta_\varepsilon}-X^{\varepsilon}_{\tau_\varepsilon}\|] \\
&\leq \sum_{i=1}^{n}\mathbf{E}\left[\int_{\tau_\varepsilon}^{\tau_\varepsilon+\delta_\varepsilon}\int_0^1\int_{\mathbb{R}^d}\|\varepsilon h\|\mathbf{1}_{\left\{\theta\leq\frac{[g_i(x_i^{\varepsilon}(s-)+\varepsilon h;X^{\varepsilon}_{s-})]_+}{\kappa}\frac{p(x_i^{\varepsilon}(s-),h)}{p(h)}\right\}}q_i\left(dh,d\theta,\frac{ds}{\varepsilon^2}\right)\right] \\
&= \sum_{i=1}^{n}\mathbf{E}\left[\int_{\tau_\varepsilon}^{\tau_\varepsilon+\delta_\varepsilon}\int_{\mathbb{R}^d}\|\varepsilon h\|[g_i(x_i^{\varepsilon}(s-)+\varepsilon h;X^{\varepsilon}_{s-})]_+p(x_i^{\varepsilon}(s-),h)\nu(dh)\frac{ds}{\varepsilon^2}\right] \\
&\leq nKM_2\delta_\varepsilon. \qquad (9)
\end{aligned}$$

Then, for any $\alpha>0$, it follows from the Markov inequality that

$$\mathbf{P}(\|X^{\varepsilon}_{\tau_\epsilon+\delta_\varepsilon}-X^{\varepsilon}_{\tau_\varepsilon}\|>\alpha)\leq\frac{nKM_2}{\alpha}\delta_\varepsilon\to 0$$

when $\varepsilon\to 0$. This gives the first part of the Aldous criterion. For the second part, we have to prove the tightness of $\{X_t^{\varepsilon}\}_{\varepsilon>0}$ for any t in $\mathbb{R}_+$. This follows trivially from the fact that $\mathbf{E}(\|X_t^{\varepsilon}\|)\leq\mathbf{E}(\|X_0^{\varepsilon}\|)+nKM_2t$ (this is inequality (9) with $\tau_\varepsilon=0$ and $\delta_\varepsilon=t$) and from the fact that $(X_0^{\varepsilon})_{0<\varepsilon\leq 1}$ is bounded in $\mathbb{L}^1(\overline{\mathcal{X}}^n)$.

Convergence of the generators Let us now prove that

$$\forall\varphi\in\mathcal{C}_b^2(\overline{\mathcal{X}}^n),\ L^{\varepsilon}\varphi\to L^0\varphi \text{ uniformly on } \overline{\mathcal{X}}^n, \qquad (10)$$

where L^{ε} is defined in (4) and L^0 is defined in (7).

Fix $\mathbf{x}\in\overline{\mathcal{X}}^n$ and $\varphi\in\mathcal{C}_b^2(\overline{\mathcal{X}}^n)$. We can write

$$L^0\varphi(\mathbf{x})=\sum_{i=1}^{n}\int_{\mathbb{R}^d}(h\cdot\nabla_i\varphi(\mathbf{x}))[h\cdot\nabla_1 g_i(x_i;\mathbf{x})]_+p(x_i,dh),$$

where $\nabla_i\varphi(\mathbf{x})$ is the gradient vector of $\varphi(\mathbf{x})$ considered as a fonction of the i^{th} coordinate x_i of $\mathbf{x}$. Then,

$$\begin{aligned}
|L^{\varepsilon}\varphi(\mathbf{x})-L^0\varphi(\mathbf{x})| &\leq \sum_{i=1}^{n}\int_{\mathbb{R}^d}\left|\frac{\varphi(\mathbf{x}+(\varepsilon h)_i)-\varphi(\mathbf{x})}{\varepsilon}\right|\times \\
&\qquad\left|\left[\frac{g_i(x_i+\varepsilon h;\mathbf{x})}{\varepsilon}\right]_+-[h\cdot\nabla_1 g_i(x_i;\mathbf{x})]_+\right|p(x_i,dh) \\
&+\sum_{i=1}^{n}\int_{\mathbb{R}^d}[h\cdot\nabla_1 g_i(x_i;\mathbf{x})]_+\times\left|\frac{\varphi(\mathbf{x}+(\varepsilon h)_i)-\varphi(\mathbf{x})}{\varepsilon}-h\cdot\nabla_i\varphi(\mathbf{x})\right|p(x_i,dh).
\end{aligned} \qquad (11)$$

Let us call B_i and C_i the quantities inside the integral in the i^{th} term of the first and the second sum, respectively. Now, φ is $\mathcal{C}^1$, $g_i(x_i;\mathbf{x})=0$ for all i in $\{1,\dots,n\}$ and, by (Hc), for $1\leq i\leq n$, $g_i(x;\mathbf{x})$ is $\mathcal{C}^1$ with respect to the first variable x. So, we can find θ_1, θ_2 and θ_3 in [0,1] depending on $\mathbf{x}$, h and i such that

$$\begin{aligned} B_i &= |h\cdot\nabla_i\varphi(\mathbf{x}+(\theta_1\varepsilon h)_i)|\times|[h\cdot\nabla_1 g_i(x_i+\theta_2\varepsilon h;\mathbf{x})]_+ - [h\cdot\nabla_1 g_i(x_i;\mathbf{x})]_+| \\ \text{and}\quad C_i &= [h\cdot\nabla_1 g_i(x_i;\mathbf{x})]_+ \times |h\cdot\nabla_i\varphi(\mathbf{x}+(\theta_3\varepsilon h)) - h\cdot\nabla_i\varphi(\mathbf{x})|. \end{aligned}$$

Now, since φ is $\mathcal{C}^2$ with bounded first and second order derivatives, and, because of (Hc), we can choose a number K such that for all $i\in\{1,\dots,n\}$, $\nabla_i\varphi$ and $\nabla_1 g_i$ are both K-Lipschitz and bounded by K on $\overline{\mathcal{X}}^n$ and $\overline{\mathcal{X}}^{n+1}$ respectively. Then

$$\begin{aligned} B_i &\leq K\|h\|\times|h\cdot\nabla_1 g_i(x_i+\theta_2\varepsilon h;\mathbf{x}) - h\cdot\nabla_1 g_i(x_i;\mathbf{x})| \leq \varepsilon K^2\|h\|^3 \\ \text{and}\quad C_i &\leq K\|h\|\times\|h\|K\|\theta_3\varepsilon h\| \leq \varepsilon K^2\|h\|^3. \end{aligned}$$

It remains to put these bounds in equation (11) to obtain:

$$|L^\varepsilon\varphi(\mathbf{x}) - L^0\varphi(\mathbf{x})| \leq 2\varepsilon K^2\sum_{i=1}^n\int_{\mathbb{R}^d}\|h\|^3 p(x_i,h)\nu(dh),$$

and the integrals of the right hand side are bounded, by (Hd) (or (He) in the symmetrical case), which ends the proof of (10).

Martingale problem for P Finally, let us show that any accumulation point $\mathbf{P}$ of the family of laws $\{\mathbf{P}^\varepsilon_{X_0^\varepsilon}\}$ on $\mathbb{D}(\mathbb{R}_+,\overline{\mathcal{X}}^n)$ is the law of the process $\mathbf{x}$ solution to (5) with initial state X_0. Fix such a $\mathbf{P}$. Let us endow the space $\mathbb{D}(\mathbb{R}_+,\overline{\mathcal{X}}^n)$ with the canonical filtration $\mathcal{G}_t$, and for any $\varphi\in\mathcal{C}^2(\overline{\mathcal{X}}^n)$, let us define on this space the processes

$$\begin{aligned} M_t^\varphi(w) &= \varphi(w_t)-\varphi(w_0)-\int_0^t L^0\varphi(w_s)ds \\ \text{and}\quad M_t^{\varepsilon,\varphi}(w) &= \varphi(w_t)-\varphi(w_0)-\int_0^t L^\varepsilon\varphi(w_s)ds. \end{aligned}$$

We have to show that $M^\varphi=0$ $\mathbf{P}$-a.s. Fix $\varphi\in\mathcal{C}_c^2(\overline{\mathcal{X}}^n)$. It is standard, using Itô formula for jump processes, to show that, under $\mathbf{P}^\varepsilon_{X_0^\varepsilon}$, $M^{\varepsilon,\varphi}$ is a square-integrable $\mathcal{G}_t$-martingale and that

$$\begin{aligned} M_t^{\varepsilon,\varphi}(X^\varepsilon) = \sum_{i=1}^n\int_0^t\int_0^1\int_{\mathbb{R}^d} &(\varphi(X_s^\varepsilon+(\varepsilon h)_i)-\varphi(X_s^\varepsilon)) \\ &\mathbf{1}_{\left\{\theta\leq\frac{[g_i(x_i^\varepsilon(s)+\varepsilon h;X_s^\varepsilon)]_+}{\kappa}\frac{p(x_i^\varepsilon(s),h)}{p(h)}\right\}}\tilde{N}_i\left(dh,d\theta,\frac{ds}{\varepsilon^2}\right) \end{aligned}$$

where $\tilde{N}_i = N_i - q_i$ is the compensated Poisson measure associated with N_i. Thus, using a calculation similar to (9),

$$\begin{aligned}\langle M^{\varepsilon,\varphi}\rangle_t &= \frac{1}{\varepsilon^2}\int_0^t \int_{\mathbb{R}^d} (\varphi(X^\varepsilon_s + (\varepsilon h)_i) - \varphi(X^\varepsilon_s))^2 \\ &\qquad\qquad [g_i(x^\varepsilon_i(s) + \varepsilon h; X^\varepsilon_s)]_+ p(x^\varepsilon_i(s),h)\nu(dh)ds \\ &\leq KK'M_3 t\varepsilon, \end{aligned} \tag{12}$$

where K' is a bound for $\nabla\varphi$, and M_3 is a bound for the third-order moment of $p(y,\cdot)$.

Using (12) and the fact that $M^\varphi_t(w) = M^{\varepsilon,\varphi}_t(w) + \int_0^t (L^\varepsilon\varphi(w_s) - L^0\varphi(w_s))ds$, it follows that

$$\mathbf{E}(|M^\varphi_t(X^\varepsilon)|^2) \leq 2t^2\|L^\varepsilon\varphi - L^0\varphi\|^2_\infty + 2K^2K'^2M_3^2t^2\varepsilon^2$$

which converges to 0 when $\varepsilon \to 0$ thanks to (10). Since $L^0\varphi$ is continuous on $\overline{\mathcal{X}}^n$ for the weak topology and since $\mathbf{P}$ is the weak limit of an extracted sequence of $(\mathbf{P}^\varepsilon_{X^\varepsilon_0})$, it follows that, under $\mathbf{P}$, $M^\varphi(w) = 0$ a.s. Note that this implies in particular that $\varphi(w)$ is $\mathbf{P}$-a.s. continuous for any $\varphi \in \mathcal{C}^2_c(\overline{\mathcal{X}}^n)$, and therefore that $\mathbf{P}$ gives a null mass on the set of discontinuous functions.

It remains to extend this result to any $\varphi \in \mathcal{C}^2(\overline{\mathcal{X}}^n)$. Fix such a φ and let (φ_k) be a sequence of compact supported functions such that $\varphi_k(x) = \varphi(x)$ if $\|x\| \leq k$. Let τ_k be the stopping time $\inf\{t \geq 0, \|w_t\| \geq k\}$. Then, for any $t \leq \tau_k$, $M^\varphi_t(w) = M^{\varphi_k}_t(w) = 0$ $\mathbf{P}$-a.s. Since the canonical process w is $\mathbf{P}$-a.s. continuous and since $\tau_k(w) \to +\infty$ when $k \to +\infty$ for any continuous w, we finally obtain that for any $t \geq 0$, $M^\varphi_t = 0$ $\mathbf{P}$-a.s., which completes the proof of Theorem 3.1. ☐

4 Diffusion model of adaptive dynamics

We will obtain here the infinitesimal generator of a diffusion process by expanding to the first order the generator L^ε of X^ε in (3). The diffusion processes given by this method are called by Ethier and Kurtz [28] (chapter 11) "diffusion approximation". We will see that its coefficients have bad regularity properties (degenerate diffusion and discontinuous drift), preventing us to prove the weak existence in the general case.

The biological motivation of this work is to provide a biological model of evolution in which the population can evolve in any direction of the trait space, unlike the jump process of section 2, where the invasion of a mutant is only possible for particular mutants, because of the positive part of g_i appearing in (3). For biological reasons, a diffusion is more realistic in many cases (see Kimura [46]), and in particular in finite populations, where any mutant, even deleterious, has a chance to invade the population.

4.1 Diffusion generator

Proposition 4.1 *Assume that for all* $i \in \{1,\dots,n\}$, $g_i(y;\mathbf{x})$ *is* $\mathcal{C}^2$ *with respect to the first variable* x, *and that its Hessian matrix* $H_1 g_i$ *with respect to the first variable is bounded and has Lipschitz entries on* $\overline{\mathcal{X}}^{n+1}$. *Assume also that* $p(y,\cdot)$ *has finite and bounded fourth order moments, and is absolutely continuous with respect to the Lebesgue measure on* $\mathbb{R}^d$ *for any* $y \in \mathcal{X}$.

Fix $\varphi \in \mathcal{C}_b^3(\overline{\mathcal{X}}^n)$. *Then, for any* $\mathbf{x} \in \overline{\mathcal{X}}^n$ *such that* $\nabla_1 g_i(x_i;\mathbf{x}) \neq 0$,

$$\begin{aligned}L^\varepsilon\varphi(\mathbf{x}) = \sum_{i=1}^n \Bigg(&\int_{\mathbb{R}^d} (h\cdot\nabla_i\varphi(\mathbf{x}))[h\cdot\nabla_1 g_i(x_i;\mathbf{x})]_+ p(x_i,dh) \\ &+ \frac{\varepsilon}{2}\int_{\mathbb{R}^d} h^* H_i\varphi(\mathbf{x}) h [h\cdot\nabla_1 g_i(x_i;\mathbf{x})]_+ p(x_i,dh) \\ &+ \frac{\varepsilon}{2}\int_{\mathbb{R}^d} (h\cdot\nabla_i\varphi(\mathbf{x}))\mathbf{1}_{\{h\cdot\nabla_1 g_i(x_i;\mathbf{x})>0\}} h^* H_1 g_i(x_i;\mathbf{x}) h p(x_i,dh)\Bigg) + o(\varepsilon),\end{aligned} \tag{13}$$

and for any $\mathbf{x} \in \mathcal{X}^n$ *such that* $\nabla_1 g_i(x_i;\mathbf{x}) = 0$,

$$L^\varepsilon\varphi(\mathbf{x}) = \frac{\varepsilon}{2}\sum_{i=1}^n \int_{\mathbb{R}^d} (h\cdot\nabla_i\varphi(\mathbf{x}))\,[h^* H_1 g_i(x_i;\mathbf{x}) h]_+\, p(x_i,dh) + o(\varepsilon). \tag{14}$$

Proof Fix $\varphi \in \mathcal{C}_b^3(\overline{\mathcal{X}}^n)$ and $\mathbf{x} \in \overline{\mathcal{X}}^n$. Expanding to the second order φ and g_i in equation (4), the Taylor-Lagrange formula gives θ_1 and θ_2 in [0,1] depending on $\mathbf{x}$, h and ε, such that

$$\begin{aligned}L^\varepsilon\varphi(\mathbf{x}) = \sum_{i=1}^n \int_{\mathbb{R}^d} &\Big(h\cdot\nabla_i\varphi(\mathbf{x}) + \frac{\varepsilon}{2} h^* H_i\varphi(\mathbf{x} + (\theta_1\varepsilon h)_i) h\Big) \\ &\Big[h\cdot\nabla_1 g_i(x_i;\mathbf{x}) + \frac{\varepsilon}{2} h^* H_1 g_i(x_i + \theta_2\varepsilon h;\mathbf{x}) h\Big]_+ p(x_i,dh),\end{aligned}$$

where $H_i\varphi$ is the Hessian matrix of $\varphi(\mathbf{x})$ considered as a function of the i^{th} coordinate of $\mathbf{x} = (x_1,\dots,x_n)$.

Using the fact that $H_1 g_i$ and $H_i\varphi$ are bounded and Lipschitz, and that $p(y,\cdot)$ has bounded fourth order moments, one can do the same computation as in the part “convergence of generators” of the proof of Theorem 3.1 in section 3.3, in order to obtain:

$$L^\varepsilon\varphi(\mathbf{x})=\sum_{i=1}^n\int_{\mathbb{R}^d}\Big(h\cdot\nabla_i\varphi(\mathbf{x})+\frac{\varepsilon}{2}h^*H_i\varphi(\mathbf{x})h\Big)\\ \Big[h\cdot\nabla_1 g_i(x_i;\mathbf{x})+\frac{\varepsilon}{2}h^*H_1g_i(x_i;\mathbf{x})h\Big]_+p(x_i,dh)+o(\varepsilon).$$

Now, (14) is immediate when $\nabla_1 g_i(x_i;\mathbf{x})=0$.

When $\nabla_1 g_i(x_i;\mathbf{x})\neq 0$, the proof of (13) is more technical. Define

$$C_\varepsilon=\{h\in\mathbb{R}^d;h\cdot\nabla_1 g_i(x_i;\mathbf{x})+(\varepsilon/2)h^*H_1g_i(x_i;\mathbf{x})h>0\}\\ \text{and } D=\{h\in\mathbb{R}^d;h\cdot\nabla_1 g_i(x_i;\mathbf{x})>0\}.$$

Let us rewrite

$$\begin{aligned}L^\varepsilon\varphi(\mathbf{x})&=\sum_{i=1}^n\int_{C_\varepsilon}\Big(h\cdot\nabla_i\varphi(\mathbf{x})+\frac{\varepsilon}{2}h^*H_i\varphi(\mathbf{x})h\Big)\\ &\qquad\Big(h\cdot\nabla_1 g_i(x_i;\mathbf{x})+\frac{\varepsilon}{2}h^*H_1g_i(x_i;\mathbf{x})h\Big)\,p(x_i,dh)+o(\varepsilon)\\ &=\sum_{i=1}^n\Big(\int_{C_\varepsilon}(h\cdot\nabla_i\varphi(\mathbf{x}))(h\cdot\nabla_1 g_i(x_i;\mathbf{x}))p(x_i,dh)\\ &+\frac{\varepsilon}{2}\int_{C_\varepsilon}(h^*H_i\varphi(\mathbf{x})h)(h\cdot\nabla_1 g_i(x_i;\mathbf{x}))p(x_i,dh)\\ &+\frac{\varepsilon}{2}\int_{C_\varepsilon}(h\cdot\nabla_i\varphi(\mathbf{x}))(h^*H_1g_i(x_i;\mathbf{x})h)p(x_i,dh)\Big)+o(\varepsilon).\end{aligned}$$

We have to show that each integral over C_ε in the right-hand side differs from the same integral taken over D by a $o(\varepsilon)$ quantity. Firstly,

$$\begin{aligned}&\Big|\int_{C_\varepsilon}(h\cdot\nabla_i\varphi(\mathbf{x}))(h\cdot\nabla_1 g_i(x_i;\mathbf{x}))p(x_i,dh)\\ &\qquad-\int_D(h\cdot\nabla_i\varphi(\mathbf{x}))(h\cdot\nabla_1 g_i(x_i;\mathbf{x}))p(x_i,dh)\Big|\\ &\leq\int_{C_\varepsilon\cap D^c}|h\cdot\nabla_i\varphi(\mathbf{x})|\times|h\cdot\nabla_1 g_i(x_i;\mathbf{x})|p(x_i,dh)\\ &\qquad+\int_{C_\varepsilon^c\cap D}|h\cdot\nabla_i\varphi(\mathbf{x})|\times|h\cdot\nabla_1 g_i(x_i;\mathbf{x})|p(x_i,dh).\end{aligned}\tag{15}$$

On $C_\varepsilon\cap D^c$, $h\cdot\nabla_1 g_i(x_i;\mathbf{x})\leq 0$ and $h\cdot\nabla_1 g_i(x_i;\mathbf{x})+\varepsilon/2(h^*H_1g_i(x_i;\mathbf{x})h)>0$, so $|h\cdot\nabla_1 g_i(x_i;\mathbf{x})|\leq\varepsilon/2(h^*H_1g_i(x_i;\mathbf{x})h)$. Similarly, the same is true on $C_\varepsilon^c\cap D$.

Since we assumed that there is some constant K bounding $H_1 g_i$ and $\nabla_i \varphi$, the right-hand side of (15) is smaller than

$$\frac{\varepsilon}{2}K^2 \int_{C_\varepsilon \cap D^c} \|h\|^3 p(x_i,dh) + \frac{\varepsilon}{2}K^2 \int_{C_\varepsilon^c \cap D} \|h\|^3 p(x_i,dh).$$

When $x_i \in \partial\mathcal{X}$ this quantity is 0 ($p(x_i,dh) = \delta_0(dh)$). When $x_i \notin \mathcal{X}$, $p(x_i,dh)$ is absolutely continuous with respect to the Lebesgue measure. Now the set $C_\varepsilon \cap D^c$ converges to the set $\{h \cdot \nabla_1 g_i(x_i;\mathbf{x}) = 0 \text{ and } h^* H_1 g_i(x_i;\mathbf{x})h > 0\}$ as $\varepsilon \to 0$, which has Lebesgue measure 0, and the set $C_\varepsilon^c \cap D$ converges to $\emptyset$ as $\varepsilon \to 0$. Since $p(x_i,\cdot)$ has finite third order moments, by the dominated convergence theorem, the quantity in the right-hand side of (15) is $o(\varepsilon)$ as $\varepsilon \to 0$.

The same method proves that

$$\begin{aligned}\frac{\varepsilon}{2}\int_{C_\varepsilon} &(h^* H_i\varphi(\mathbf{x})h)(h \cdot \nabla_1 g_i(x_i;\mathbf{x}))p(x_i,dh) \\ &= \frac{\varepsilon}{2}\int_D (h^* H_i\varphi(\mathbf{x})h)(h \cdot \nabla_1 g_i(x_i;\mathbf{x}))p(x_i,dh) + o(\varepsilon^2)\end{aligned}$$

and that

$$\begin{aligned}\frac{\varepsilon}{2}\int_{C_\varepsilon} &(h \cdot \nabla_i\varphi(\mathbf{x}))(h^* H_1 g_i(x_i;\mathbf{x})h)p(x_i,dh) \\ &= \frac{\varepsilon}{2}\int_D (h \cdot \nabla_i\varphi(\mathbf{x}))(h^* H_1 g_i(x_i;\mathbf{x})h)p(x_i,dh) + o(\varepsilon),\end{aligned}$$

as required. □

Neglecting the terms of order greater than one in (13) and (14), we obtain the following diffusion operator: for any $\varphi \in \mathcal{C}^2(\overline{\mathcal{X}}^n)$,

$$\tilde{L}^\varepsilon \varphi(\mathbf{x}) = b^\varepsilon(\mathbf{x}) \cdot \nabla\varphi(\mathbf{x}) + \frac{\varepsilon}{2}\sum_{i=1}^n \sum_{1\le k,l\le d} a^i_{kl}(\mathbf{x}) \frac{\partial_i^2 \varphi}{\partial_i x_k \partial_i x_l}(\mathbf{x}). \tag{16}$$

where

$$b^\varepsilon(\mathbf{x}) = b(\mathbf{x}) + \varepsilon\tilde{b}(\mathbf{x}),$$

where $b = (b^1,\dots,b^n)$, and $\tilde{b} = (\tilde{b}^1,\dots,\tilde{b}^n)$, with, for any $\mathbf{x} \in \overline{\mathcal{X}}^n$,

$$b^i_k(\mathbf{x}) = \int_{\mathbb{R}^d} h_k[\nabla_1 g_i(x_i;\mathbf{x}) \cdot h]_+ p(x_i,dh),$$

$$\tilde{b}^i_k(\mathbf{x}) = \begin{cases} \dfrac{1}{2}\displaystyle\int_{\{h\cdot\nabla_1 g_i(x_i;\mathbf{x})>0\}} h_k(h^* H_1 g_i(x_i;\mathbf{x})h)p(x_i,dh) \\ \qquad\qquad \text{when } \nabla_1 g_i(x_i;\mathbf{x}) \neq 0, \\ \dfrac{1}{2}\displaystyle\int_{\mathbb{R}^d} h_k[h^* H_1 g_i(x_i;\mathbf{x})h]_+ p(x_i,dh) \\ \qquad\qquad \text{when } \nabla_1 g_i(x_i;\mathbf{x}) = 0, \end{cases}$$

$$\text{and } a^i_{kl}(\mathbf{x}) = \int_{\mathbb{R}^d} h_k h_l [h\cdot\nabla_1 g_i(x_i;\mathbf{x})]_+ p(x_i,dh). \tag{17}$$

Define also $a^i = [a^i_{kl}]_{1\leq k,l\leq d}$ for all $i \in \{1,\dots,n\}$, and a the block diagonal matrix with blocks a^i for $1 \leq i \leq n$.

4.2 Regularity of the coefficients a, b and $\tilde{b}$

We will need the assumptions

(H1) $g_i(x;\mathbf{x})$ is $\mathcal{C}^2$ with respect to the first variable x on $\overline{\mathcal{X}}^{n+1}$, and $\nabla_1 g_i$ and $H_1 g_i$ are continuous and bounded on $\overline{\mathcal{X}}^{n+1}$.

(H2) $x \mapsto p(x,\cdot)$ is continuous on $\overline{\mathcal{X}}$ for the distance ρ_3 of Definition 3.1. Moreover, for any $x \in \mathcal{X}$, $p(x,dh) = p(x,h)dh$, and there exists some measurable function $p(h)$ such that for all $x \in \mathcal{X}$ and $h \in \mathbb{R}^d$, $p(x,h) \leq p(h)$, and such that $\int_{\mathbb{R}^d} \|h\|^3 p(h)dh < \infty$.

(H3) Let us define for $1 \leq i \leq n$

$$\Gamma_i = \{\mathbf{x} \in \mathcal{X}^n ; \nabla_1 g_i(x_i;\mathbf{x}) = 0\} \quad \text{and} \quad \Gamma = \bigcup_{i=1}^n \Gamma_i.$$

Proposition 4.2 *Assume (H1) and (H2). Then, a, b and $\tilde{b}$, defined in (17), are bounded, and a and b are continuous on $\overline{\mathcal{X}}^n$. Moreover, the matrix $a(\mathbf{x})$ is symmetrical non-negative for $\mathbf{x} \in \mathcal{X}^n$, non-degenerate for $\mathbf{x} \in \mathcal{X}^n \setminus \Gamma$, and $\tilde{b}^i$ is continuous on $\overline{\mathcal{X}}^n \setminus \Gamma_i$.*

Proof The boundedness properties are obvious. Moreover, for $\mathbf{x} \in \mathcal{X}^n$, a is obviously symmetrical, and, given any vector $v = (v_1,\dots,v_d)$ in $\mathbb{R}^d$ and any $i \in \{1,\dots,n\}$, an easy calculation shows that

$$v^* a^i(\mathbf{x})v = \int_{\mathbb{R}^d} (h\cdot v)^2 [h\cdot\nabla_1 g_i(x_i;\mathbf{x})]_+ p(x_i,h)dh \geq 0.$$

This quantity is non-zero when $\nabla_1 g_i(x_i;\mathbf{x}) \neq 0$.

For any $\mathbf{x}$ and $\mathbf{x}'$ in $\overline{\mathcal{X}}^n$, and for $1 \leq k \leq d$ and $1 \leq i \leq n$,

$$|b_k^i(\mathbf{x}) - b_k^i(\mathbf{x}')| \leq \left| \int_{\mathbb{R}^d} h_k \left([\nabla_1 g_i(x_i;\mathbf{x}) \cdot h]_+ - [\nabla_1 g_i(x_i';\mathbf{x}') \cdot h]_+ \right) p(x_i,dh) \right| \\ + \left| \int_{\mathbb{R}^d} h_k [\nabla_1 g_i(x_i;\mathbf{x}) \cdot h]_+ (p(x_i,dh) - p(x_i',dh)) \right|.$$

The first term of the right-hand side converges to 0 when $\mathbf{x}' \to \mathbf{x}$ because $\nabla_1 g_i$ is continuous on $\overline{\mathcal{X}}^n$ and $p(x,\cdot)$ has bounded second order moments, and the second term also converges to 0 by Proposition 3.2 and (H2). A similar computation shows that the same holds for a.

It remains to show the continuity of $\tilde{b}^i$ on $\overline{\mathcal{X}}^n \setminus \Gamma_i$. Fix $\mathbf{x}$ and $\mathbf{x}'$ in $\mathcal{X}^n \setminus \Gamma_i$ and define $S = \{h; h \cdot \nabla_1 g_i(x_i;\mathbf{x}) > 0\}$ and $S' = \{h; h \cdot \nabla_1 g_i(x_i';\mathbf{x}') > 0\}$. Then

$$\begin{aligned}
|\tilde{b}_k^i(\mathbf{x}) - \tilde{b}_k^i(\mathbf{x}')| \leq \frac{1}{2} \Bigg| & \int_{S \cap S'} h_k [h^* (H_1 g_i(x_i;\mathbf{x}) - H_1 g_i(x_i';\mathbf{x}'))h] p(x_i',h) dh \\
& + \int_S h_k (h^* H_1 g_i(x_i;\mathbf{x}) h)(p(x_i,h) - p(x_i',h)) dh \\
& - \int_{S \cap S'^c} h_k (h^* H_1 g_i(x_i;\mathbf{x}) h) p(x_i',h) dh \\
& - \int_{S^c \cap S'} h_k (h^* H_1 g_i(x_i';\mathbf{x}') h) p(x_i',h) dh \Bigg| \\
\leq \frac{1}{2} \Bigg(& \int_{\mathbb{R}^d} \|h\|^3 \|H_1 g_i(x_i;\mathbf{x}) - H_1 g_i(x_i';\mathbf{x}')\| p(x_i',h) dh \\
& + \left| \int_S h_k (h^* H_1 g_i(x_i;\mathbf{x}) h)(p(x_i,h) - p(x_i',h)) dh \right| \\
& + \int_{S \cap S'^c} \|h\|^3 \|H_1 g_i(x_i;\mathbf{x})\| p(x_i',h) dh \\
& + \int_{S^c \cap S'} \|h\|^3 \|H_1 g_i(x_i';\mathbf{x}')\| p(x_i',h) dh \Bigg).
\end{aligned}$$

The first term goes to 0 when $\mathbf{x}' \to \mathbf{x}$ since $H_1 g_i$ is continuous and $p(x_i,\cdot)$ has finite third order moments. The second term goes to 0 by Proposition 3.2. Using (H2) and the fact that $H_1 g_i$ is bounded, the third term is bounded by

$$\int_{S \cap S'^c} K \|h\|^3 p(h) dh$$

and the fourth term by

$$\int_{S^c \cap S'} K \|h\|^3 p(h) dh$$

for some constant K. Now, when $\mathbf{x}' \to \mathbf{x}$, the sets $S \cap S'^c$ and $S^c \cap S'$ converge respectively to $\emptyset$ and $\{h; h \cdot \nabla_1 g_i(x_i; \mathbf{x}) = 0\}$, so the dominated convergence theorem gives the required result, since $\{h : h \cdot \nabla_1 g_i(x_i; \mathbf{x}) = 0\}$ is neglectable for the Lebesgue measure when $\mathbf{x} \notin \Gamma$.

Finally, when $\mathbf{x} \to \mathbf{x}' \in \partial(\mathcal{X}^n)$, it is straightforward to prove that $\tilde{b}(\mathbf{x}) \to 0$. □

Note that in general, $\tilde{b}$ is not continuous at the points of Γ. For example, if $\mathcal{X} = \mathbb{R}$, $n = 1$ and $p(x,\cdot)$ is symmetrical for all $x \in \mathcal{X}$, a fair computation shows that

$$\tilde{b}(x) = \frac{1}{2}\operatorname{sign}[\partial_1 g(x;x)]\partial_1^2 g(x;x) \int_{\mathbb{R}} |h|^3 p(x,dh),$$

with the convention $\operatorname{sign}(0) = 0$.

4.3 Concluding remarks and comments

Since, by Proposition 4.2, this diffusion operator combines a degenerate diffusion parameters with a discontinuous drift *at the same points*, the weak existence of a diffusion solution to

$$dY_t^\varepsilon = b^\varepsilon(Y_t^\varepsilon)dt + \sqrt{\varepsilon}\sigma(Y_t^\varepsilon)dW_t, \tag{18}$$

where $\sigma(\mathbf{x})$ is a $nd \times k$ matrix such that $\sigma(\mathbf{x})\sigma^*(\mathbf{x}) = a(\mathbf{x})$, and where W_t is a standard k-dimensional Brownian motion, is a very difficult problem.

We have considered several methods in order to solve this problem, none of which had allowed to prove the weak existence in the general case. The first one is to extend in a continuous way the parameters a, b and $\tilde{b}$ to $\mathbb{R}^{nd}$, and then to define $a_p(\mathbf{x}) = a(\mathbf{x}) + \mathrm{Id}/p$ for $p \geq 1$. Since a_p is continuous and non degenerate, the weak existence of the solutions to

$$dY_t^{\varepsilon,p} = b^\varepsilon(Y_t^{\varepsilon,p})dt + \sqrt{\varepsilon}\sigma_p(Y_t^\varepsilon)dW_t,$$

where $\sigma_p\sigma_p^* = a_p$ follows from the Girsanov's formula (see problem 5.3.13 p. 305 of Karatzas and Shreve [45]) and from the theorem of Stroock and Varadhan (cf. [45] p. 323).

It is not difficult to deduce from the Burkholder-Davis-Gundy inequality the tightness of the law of $Y^{\varepsilon,p}$ with fixed initial state $x \in \mathbb{R}^{nd}$ when $p \to +\infty$, but, given an accumulation point $\mathbf{P}$ of this sequence of laws, it is very difficult to determine the behavior of a process with law $\mathbf{P}$ at Γ, so that we were not able to tell if it is solution to the (18).

Another method consists to construct a process solution to (18) starting from $\mathbf{x} \notin \Gamma$ by an approximation technique until the time when it reaches Γ (this idea is made formal in a simpler case in chapter IV). If this time is a.s. infinite, then it is easy to prove that there is weak existence, but whether this time is a.s. infinite or not is a difficult problem, that we partly solve in a simpler case in chapter IV. The main difficulty is that $a^i(\mathbf{x}) = 0$ for $\mathbf{x} \in \Gamma_i$, but $a^j(\mathbf{x})$ is non-degenerate when $\mathbf{x} \notin \Gamma_j$, so that the diffusion matrix is partly degenerate at points of Γ.

However, in the case where mutations are symmetrical (*i.e.*, for any $x \in \overline{\mathcal{X}}$, the measure $p(x,dh)$ is invariant under the application $h \mapsto -h$ on $\mathbb{R}^d$), it follows from (17) that $b^i(\mathbf{x}) = \tilde{b}^i(\mathbf{x}) = a^i(\mathbf{x}) = 0$ for any $\mathbf{x} \in \Gamma_i$.

Therefore, in the case where $n = 1$ (but $d \geq 1$), and when the mutations are symmetrical, we could define a solution to (18) until the time when it reaches Γ, and assume it constant after this time. We use this idea in chapter IV to prove the existence of the diffusion (18) in this special case. We are also able to study the uniqueness in law and the strong Markov property for (18). These properties hold under the assumption that a solution X^ε to (18) starting from $\mathcal{X} \setminus \Gamma$ does not reach Γ in finite time. We also found, for any dimension d, conditions under which X^ε never reaches Γ, and other conditions under which X^ε can reach Γ in finite or infinite time.

Let us finally mention an important biological application of this diffusion process, which is our principal motivation for this study. Evolution, when it reaches a steady state, can proceed by different strategies (see Rand and Wilson [66]), among which is the phenomenon of *punctualism* (rapid evolution from one evolutionary steady state to another). Our diffusion model of section 4 could help study the process of punctualism, by means of a large deviations principle, through the study of the problem of exit from a domain (Freidlin and Wentzell [34]). This study forms the main part of chapter IV, under the assumption of symmetrical mutations and in the monomorphic case.

Chapter IV

Existence, uniqueness, strong Markov property and large deviations for degenerate diffusion models of evolution

Abstract: In the course of Darwinian evolution of a population, punctualism is an important phenomenon whereby long periods of genetic stasis alternate with short periods of rapid evolutionary change. This paper provides a mathematical interpretation of punctualism as a sequence of change of basin of attraction for a diffusion model of the theory of adaptive dynamics obtained in chapter III. The main difficulty lies in the fact that this diffusion process has degenerate and non-Lipschitz diffusion part at isolated points of the space, called "evolutionary singularities", and non-continuous drift part at the same points. Nevertheless, we are able to prove the weak existence of these diffusions, and to give conditions under which the uniqueness in law and strong Markov property hold. These conditions rely on the question whether this diffusion can reach the evolutionary singularities, which we partly solve. Finally, we prove a large deviation principle under general assumptions, allowing to obtain asymptotic estimates for the time needed to exit an attracting domain, and to identify the points where this exit is more likely to occur.

1 Introduction

The Darwinian evolution of an asexual population is controlled by demographic (birth and death) rates, which are typically influenced by quantitative characters: morphological traits like body size, physiological traits like the rate of food intake, life-history traits like the age at maturity. Such traits are heritable yet not perfectly transmitted from parents to offsprings, due to mutations of genes involved in their expression. The resulting variation of traits is then exposed to selection caused by ecological interactions between individuals competing for limited resources.

Adaptive dynamics theory has been developed since the early 1990's, by Hofbauer and Sigmund [41], Marrow *et al.* [57] and Metz *et al.* [61], as a mathematical framework to model the Darwinian evolution of a population, as a stochastic process in the trait space. This new modelling approach has shed light on polymorphism (stable coexistence of different traits), on evolutionary branching (evolution of a monomorphic population to a polymorphic one), leading eventually to speciation (Metz *et al.* [59], Doebeli and Dieckmann [21] and Dieckmann and Doebeli [16]), and on the process of "evolutionary suicide" whereby the extinction of a population is caused by Darwinian evolution itself (Dieckmann and Ferrière [17]).

One evolutionary pattern that remains poorly understood among biologists is that of "punctualism": the phenomenon of Darwinian evolution whereby long periods of trait stasis alternate with periods of global, rapid changes in the trait values of the population, which can be due to a large mutation or to successive invasions of slightly disadvantaged mutants in the population (see Rand and Wilson [66]). In this paper, we show that punctualism can be interpreted as quick changes of basin of attraction for the evolutionary process, between which this process stays for long time scales near the evolutionary equilibrium inside the current basin of attraction. We use a diffusion model of adaptive dynamics obtained in chapter III for which we prove a large deviation principle that allows to estimate the time needed to exit a given basin of attraction, and to determine where the exit happens.

Let us assume that the quantitative trait characterizing individuals belongs to a convex open subset $\mathcal{X}$ of $\mathbb{R}^d$. The classical adaptive dynamics models are based on two biological assumptions: (1) Mutations are very rare (evolutionary and ecological time scales are separated, see [59]), which allows one to assume that the population remains *monomorphic* (*i.e.* composed of individuals holding the same trait value) at any time, so that the state of the population is described by a single trait value $x \in \mathcal{X}$; (2) The population is large, so that changes in the population size are nearly deterministic. This leads to a jump process, so-called *trait substitution sequence* (Metz *et al.* [59], Dieckmann and Law [18]) mathematically studied in chapter III, in which evolution is possible only in particular directions in the trait space. This pattern is unrealistic on long time scales, where, in large yet finite populations, slightly disadvantaged mutants could by chance become fixed in the resident population. To account for this feature, we have introduced in chapter III a new model of adaptive dynamics in the form of a diffusion process, that can be interpreted as an approximation of the trait substitution sequence. The infinitesimal generator of this diffusion process is a first-order approximation of the generator of the trait substitution sequence process, in the limit of small mutation jumps.

In this diffusion process, the trait value of the population is solution on $\overline{\mathcal{X}}$ to the following stochastic differential equation, which coefficients are obtained explicitly in terms of biological parameters (see section 2):

$$dX_t^\varepsilon = (b(X_t^\varepsilon) + \varepsilon\tilde{b}(X_t^\varepsilon))dt + \sqrt{\varepsilon}\sigma(X_t^\varepsilon)dW_t, \tag{1}$$

where $b(x)$ and $\tilde{b}(x)$ are in $\mathbb{R}^d$, where $\sigma(x)$ is a $d \times d$ symmetric positive real matrix, and where $\varepsilon > 0$ is a small parameter scaling the size of the mutation jumps. The term $\varepsilon\tilde{b}$ is a correction term for b and $\sqrt{\varepsilon}\sigma$ is the coefficient driving the fluctuations around the unperturbed dynamics.

The main difficulty of this model is that the standard regularity assumptions for stochastic processes are not satisfied by these coefficients: the function b is (globally) Lipschitz, but $\tilde{b}$ is discontinuous at isolated points of the trait space, called *evolutionary singularities*, and σ is not globally Lipschitz, but is only 1/2-Hölder near the set Γ of evolutionary singularities. Moreover, $b(x) = \tilde{b}(x) = \sigma(x) = 0$ for $x \in \Gamma$.

Despite these difficulties, we are able to partly study the existence, uniqueness and strong markov property for these diffusion processes, to prove a large deviations principle as $\varepsilon \to 0$, and to partly solve the problem of diffusion exit from an attracting domain (see Freidlin and Wentzell [34]), which is the key question for punctualism: what are the time and the point of exit from such a domain for the diffusion process X^ε?

In this paper, we first give a precise description of the model and study the regularity of the parameters $a = \sigma\sigma^*$, b and $\tilde{b}$ (section 2). In section 3.1, we establish the weak existence for (1) by constructing explicitely a solution to the martingale problem associated with (1), under which the solution X^ε to (1) is constant after hitting Γ. We will study in the rest of the paper this particular diffusion process. Because of the bad regularity properties of $\tilde{b}$ and σ, the uniqueness in law and strong Markov property for (1) are difficult problems: we are only able to solve them under the assumption that X^ε a.s. never reaches Γ (section 3.2), and we give in sections 3.3 and 3.4 technical conditions under which this assumption holds, involving precise estimates on the coefficients a, b and $\tilde{b}$. In section 4, we prove the main result of this paper, without assuming that X^ε does not reach Γ: a large deviation principle for the paths of X^ε as $\varepsilon \to 0$. Our proof is inspired from Doss and

Priouret [23] (who follow Azencott [4]), and requires a delicate and technical study of the paths of the diffusion X^ε near Γ. This method leads to an original form of the rate function, which is not lower semicontinuous. This study is made possible by the facts that the diffusion coefficient σ degenerates at the same points where $\tilde{b}$ is discontinuous, and that the diffusion process is stopped after reaching these points. Finally (section 5), we apply this result to the problem of diffusion exit from an attracting domain. Since the process stays constant after hitting Γ, we have to introduce an original form of the quasi-potential of Freidlin and Wentzell. We obtain a lower bound for the time of exit and we prove that the exit occurs with high probability near points of the boundary minimizing this quasi-potential.

2 Description of the model

The coefficients b, $\tilde{b}$ and $\sigma\sigma^* = a$ are expressed in terms of two biological parameters: a "fitness" function, and the mutation law. In this section, we will first describe these parameters, and then study the regularity of the coefficients of the SDE (1).

2.1 The fitness function

The first biological parameter, $g(x',x)$, is a function from $\overline{\mathcal{X}} \times \overline{\mathcal{X}}$ to $\mathbb{R}$ which measures the *fitness*, *i.e.* the selective advantage (or disadvantage), of a single individual with trait x' in a monomorphic population made of individuals holding trait x (see chapter III).

For biological reasons, we assume that g satisfies

$$\forall x \in \mathcal{X}, \quad g(x,x) = 0, \tag{2}$$

(a classical property for fitness functions, cf. chapter I) and that $g = 0$ on the boundary of $\overline{\mathcal{X}} \times \overline{\mathcal{X}}$.

When g is $\mathcal{C}^1$ with respect to the first variable, we will denote by $\nabla_1 g$ the gradient vector of $g(x',x)$ with respect to the first variable x', and, when g is $\mathcal{C}^2$, by $H_{i,j}g$ the Hessian matrix of $g(x',x)$ with respect to the i^{th} and j^{th} variables ($1 \leq i,j \leq 2$).

Let us define the following sets, needed to control the regularity of the coefficients:

$$\Gamma = \partial\mathcal{X} \cup \{x \in \mathcal{X} : \nabla_1 g(x,x) = 0\}, \tag{3}$$

$$\text{and} \quad \forall \alpha > 0,\ \Gamma_\alpha = \{x \in \mathcal{X} : d(x,\Gamma) \geq \alpha\}, \tag{4}$$

Γ defines the set of *evolutionary singularities*.

We assume moreover that

(H1) $g(x',x)$ is $\mathcal{C}^2$ on $\overline{\mathcal{X}} \times \overline{\mathcal{X}}$ with respect to the first variable x', and $\nabla_1 g$ and $H_{1,1}g$ are bounded and Lipschitz on $\mathcal{X}^2$.

(H2) $\forall \alpha > 0,\ \inf_{x\in\Gamma_\alpha} \|\nabla_1 g(x,x)\| > 0$.

(H2) is obviously satisfied when (H1) is true and $\overline{\mathcal{X}}$ is compact. In the case where $\mathcal{X}$ is not bounded, it only states that $\nabla_1 g(x,x)$ does not converge too fast to 0 when $\|x\| \to +\infty$.

2.2 The mutation law

The second biological parameter, $p(x,dh)$, is the law of $h = x' - x$, where x' is a mutant trait born out of an individual with trait $x \in \overline{\mathcal{X}}$. Since x' must be in the trait space $\mathcal{X}$, the support of $p(x,\cdot)$ is a subset of

$$\overline{\mathcal{X}} - x = \{y - x; y \in \overline{\mathcal{X}}\}. \tag{5}$$

We assume that, for $x \in \mathcal{X}$, $p(x,dh) = p(x,h)dh$, where $p(x,h)$ is defined on $\mathcal{X} \times \mathbb{R}^d$ and $p(x,h)dh$ is symmetrical with respect to 0. When x is in the boundary $\partial\mathcal{X}$ of $\mathcal{X}$, we assume $p(x,dh) = \delta_0(dh)$, where δ_0 is the Dirac measure at 0.

The assumption that $p(x,h)dh$ is a symmetrical measure is required for the construction of solutions to (1). This is almost always assumed in adaptive dynamics models (see e.g. Dieckmann and Law [18], Doebeli and Dieckmann [21] or Kisdi [49]). See chapter III for a discussion of the difficulties arising when $p(x,h)dh$ is asymmetric. The assumption that $p(x,dh) = \delta_0(dh)$ for $x \in \partial\mathcal{X}$ is natural in view of the assumption of symmetry for the measure $p(x,dh)$ and of the fact that the support of this measure is a subset of $\overline{\mathcal{X}} - x$.

We assume moreover that

(H3) (1) In the case where $\mathcal{X} = \mathbb{R}^d$, we assume that $p(x,h)dh$ has finite and bounded third-order moment, and the following Lipschitz-like condition: there exists a measurable function $m : \mathbb{R}^d \to \mathbb{R}_+$ with

$$\int (\|h\|^2 \vee \|h\|^3) m(h)dh < +\infty,$$

such that, for any $x,x' \in \mathcal{X}$ and $h \in \mathbb{R}^d$,

$$|p(x,h) - p(x',h)| \leq \|x - x'\| m(h). \tag{6}$$

(2) When $\mathcal{X} \neq \mathbb{R}^d$, since $p(x,dh)$ is degenerate at the boundary of $\mathcal{X}$, we have to change this assumption in the following technical way: there

exists $\alpha_0 > 0$ and a measurable function $m : \mathbb{R}^d \to \mathbb{R}_+$ with

$$\int (\|h\|^2 \vee \|h\|^3) m(h) dh < +\infty, \tag{7}$$

such that, for all x and x' in $\mathcal{X}$ and $h \in \mathbb{R}^d$, using the notation $\alpha := d(x,\partial\mathcal{X})$ and $\alpha' := d(x',\partial\mathcal{X})$,

$$\text{if } \alpha \geq \alpha_0 \text{ and } \alpha' \geq \alpha_0, \quad |p(x,h) - p(x',h)| \leq \|x - x'\| m(h), \tag{8}$$

$$\text{if } \alpha \leq \alpha_0 \text{ and } \alpha' \leq \alpha_0, \; |\alpha^d p(x,\alpha h) - \alpha'^d p(x',\alpha' h)| \leq \|x - x'\| m(h), \tag{9}$$

$$\text{and} \quad \int_{\mathbb{R}^d} (\|h\|^2 \vee \|h\|^3) p(x,h) dh < K(\alpha \wedge 1) = K(d(x,\partial\mathcal{X}) \wedge 1). \tag{10}$$

for some constant K.

(H4) $\forall \alpha > 0, \inf_{\{d(x,\partial\mathcal{X}) \geq \alpha, \; u,v \in \mathbb{R}^d : \|u\| = \|v\| = 1\}} \int_{\mathbb{R}^d} |h \cdot u|^2 |h \cdot v| p(x,h) dh > 0.$

(H3) ensures a sufficient regularity for $p(x,h)$. When $\mathcal{X} = \mathbb{R}^d$, (6) is usually satisfied for classical jump measures: for example, when $p(x,h)dh$ is Gaussian for all $x \in \mathcal{X}$ with covariance matrix $K(x)$ uniformly non-degenerate, bounded and Lipschitz on $\mathbb{R}^d$. When $\mathcal{X} \neq \mathbb{R}^d$, assumption (9) means that for x close to x' and close to $\partial\mathcal{X}$, the image of $p(x,h)$ by the application $h \mapsto h/\alpha$ is very close to the image of $p(x',h)$ by the application $h \mapsto h/\alpha'$. Condition (10) is true (for example) when the support of $p(x,\cdot)$ is a subset of $B(0,d(x,\partial\mathcal{X}))$.

(H4) is a technical condition needed to control the non-degeneracy of the matrix a. For example, it is true if there exists $\beta > 0$ such that for any $h \in \mathbb{R}^d$ and $x \in \mathcal{X}$, $p(x,h) \geq \beta \mathbf{1}_{\{\|x-h\| \leq d(x,\partial\mathcal{X})/2\}}$.

Assumption (H3) implies the following lemma:

Lemma 2.1 *Assume (H3). Let* $S = \mathbb{R}^d$ *or* $S = \{h : h \cdot u > 0\}$ *for some* $u \in \mathbb{R}^d \setminus \{0\}$, *and let* f *be a function from* $\mathbb{R}^d$ *to* $\mathbb{R}$ *such that* $f(0) = 0$ *and*

$$\forall x,y \in \mathbb{R}^d, \; |f(x) - f(y)| \leq K\|x - y\| \max\{\|x\|,\|y\|,\|x\|^2,\|y\|^2\} \tag{11}$$

for some constant K. *Then, the function* $\phi(x) = \int_S f(h) p(x,dh)$ *is globally Lipschitz on* $\overline{\mathcal{X}}$.

Proof of Lemma 2.1 Observe that (11) implies that

$$|f(h)| \leq K(\|h\|^2 \vee \|h\|^3). \tag{12}$$

The fact that ϕ is globally Lipschitz on $\{x \in \mathcal{X} : d(x,\partial\mathcal{X}) \geq \alpha_0\}$ follows easily from (11), (12) and from assumptions (7) and (8).

Fix $x \in \partial\mathcal{X}$ and x' such that $d(x',\partial\mathcal{X}) \leq \alpha_0$. A uniform Lipschitz bound for $|\phi(x) - \phi(x')|$ easily follows from (12) and assumption (10).

Now fix x and x' in $\mathcal{X}$ such that $\alpha := d(x,\partial\mathcal{X}) \leq \alpha_0$ and $\alpha' := d(x',\partial\mathcal{X}) \leq \alpha_0$, and assume that $\alpha \leq \alpha'$. Then, using the changes of variable $\alpha u = h$ for $p(x,h)dh$ and $\alpha' u = h$ for $p(x',h)dh$, using that S is a positive cone, and using (9), we can write

$$\begin{aligned}
|\phi(x) - \phi(x')| &\leq \int_S |f(\alpha u)||\alpha^d p(x,\alpha u) - \alpha'^d p(x',\alpha' u)|du \\
&\quad + \int_S |f(\alpha u) - f(\alpha' u)|\alpha'^d p(x',\alpha' u)du \\
&\leq K\|x - x'\| \int_{\mathbb{R}^d} ((\alpha_0\|u\|) \vee (\alpha_0^2\|u\|^2))m(u)du \\
&\quad + K|\alpha - \alpha'| \int_{\mathbb{R}^d} \|u\| \max\{\alpha'\|u\|,\alpha'^2\|u\|^2\} p(x',\alpha' u)du \\
&\leq K\|x - x'\| + K|\alpha - \alpha'|\frac{1}{\alpha'} \int_{\mathbb{R}^d} (\|h\|^2 \vee \|h\|^3)p(x',h)dh,
\end{aligned}$$

where the constant K may change from line to line. Using the inequalities $|\alpha - \alpha'| \leq \|x - x'\|$ and (10), we obtain a uniform Lipschitz bound when $\alpha \leq \alpha_0$ and $\alpha' \leq \alpha_0$, which completes the proof of Lemma 2.1. ☐

Remark 2.1 *This lemma is the only consequence of (H3) that will be used in the following. One could replace assumption (H3) with any condition ensuring Lemma 2.1. In chapter III, we have given a less explicit condition on* $p(x,h)dh$ *implying Lemma 2.1 based on Kantorovich metrics (see Rachev [65]).*

2.3 The diffusion approximation

The diffusion approximation of chapter III leads to the following formulas for b, $\tilde{b}$ and $a = \sigma\sigma^*$: let $b(x) = (b_1(x),\dots,b_d(x))$, $\tilde{b}(x) = (\tilde{b}_1(x),\dots,\tilde{b}_d(x))$, and $a(x) = (a_{kl}(x))_{1 \leq k,l \leq d}$. Then, for $x \in \overline{\mathcal{X}}$,

$$b_k(x) = \int_{\mathbb{R}^d} h_k[\nabla_1 g(x,x) \cdot h]_+ p(x,dh),$$

$$\tilde{b}_k(x) = \begin{cases} \dfrac{1}{2} \displaystyle\int_{\{h \cdot \nabla_1 g(x,x) > 0\}} h_k(h^* H_{1,1} g(x,x) h) p(x,dh) & \\ & \text{if} \quad \nabla_1 g(x,x) \neq 0, \\ 0 & \text{otherwise,} \end{cases}$$

$$\text{and}\quad a_{kl}(x)=\int_{\mathbb{R}^d} h_k h_l[h\cdot\nabla_1 g(x,x)]_+ p(x,dh). \tag{13}$$

We will take for the coefficient σ appearing in (1), satisfying $a=\sigma\sigma^*$, the unique real symmetrical positive $d\times d$ square root of a.

Observe that, when $x\in\partial\mathcal{X}$, since $p(x,dh)=\delta_0(dh)$, $a(x)=b(x)=\tilde{b}(x)=0$, and this also holds when $\nabla_1 g(x,x)=0$. Thus, all points in Γ (defined in (3)) are possible rest points of solutions to (1).

Let us also define

$$b^\varepsilon=b+\varepsilon\tilde{b}.$$

The following proposition studies the regularity of a, b, $\tilde{b}$ and σ.

Proposition 2.1

(i) *Assume (H1) and (H3). Then a and b are Lipschitz and bounded on $\mathcal{X}$, and $\tilde{b}$ is bounded on $\mathcal{X}$ and continuous on $\mathcal{X}\setminus\Gamma$.*

(ii) *The matrix a is symmetrical and non-negative on $\mathcal{X}$, $a(x)=0$ if $x\in\Gamma$, and $a(x)$ is definite positive if $x\in\mathcal{X}\setminus\Gamma$. Moreover, under assumptions (H1) and (H3), the symmetrical square root σ of a is bounded and Hölder with exponent $1/2$ on $\overline{\mathcal{X}}$.*

(iii) *Assume (H2) and (H4). Then, $\forall\alpha>0$, $\exists c>0$ such that $\Gamma_\alpha\subset\{x\in\mathcal{X}:\forall s\in\mathbb{R}^d,\ s^*a(x)s\geq c\|s\|^2\}$ (where Γ_α is defined in (4)). Under the additional assumptions (H1) and (H3), σ is Lipschitz on Γ_α for all $\alpha>0$.*

Proof of (i) a, b and $\tilde{b}$ are trivially bounded.

Let x and x' be in $\overline{\mathcal{X}}$. For $1\leq k\leq d$,

$$\begin{aligned}|b_k(x)-b_k(x')|\leq&\left|\int_{\mathbb{R}^d} h_k([\nabla_1 g(x,x)\cdot h]_+-[\nabla_1 g(x',x')\cdot h]_+)p(x,dh)\right|\\&+\left|\int_{\mathbb{R}^d} h_k[\nabla_1 g(x,x)\cdot h]_+(p(x,dh)-p(x',dh))\right|.\end{aligned}$$

Using the fact that $|[a]_+-[b]_+|\leq|a-b|$ and that $\nabla_1 g$ is K-Lipschitz, the first term of the right-hand side is less than $K(2\|x-x'\|)M_2$, where M_2 is a bound for the second-order moments of $p(x,dh)$ for $x\in\mathcal{X}$. The second term can be bounded by $K'\|\nabla_1 g(x,x)\|\|x-x'\|$ thanks to Lemma 2.1 (the function $h\mapsto h_k(\nabla_1 g(x,x)\cdot h)$ can easily be shown to satisfy (11) by bounding its differential, and one should then choose $S=\{h:\nabla_1 g(x,x)\cdot h>0\}$). Since, by (H1), $\nabla_1 g$ is bounded on $\mathcal{X}^2$, it follows that b is Lipschitz on $\overline{\mathcal{X}}$. Similarly, a is Lipschitz on $\overline{\mathcal{X}}$.

Fix x and x' in $\mathcal{X} \setminus \Gamma$. Define $S = \{h \in \mathbb{R}^d : h \cdot \nabla_1 g(x,x) > 0\}$ and $S' = \{h : h \cdot \nabla_1 g(x',x') > 0\}$. Then,

$$\begin{aligned}
&|\tilde{b}_k(x) - \tilde{b}_k(x')| \\
&\qquad \leq \frac{1}{2} \left| \int_{S \cap S'} h_k[h^*(H_{1,1}g(x,x) - H_{1,1}g(x',x'))h]p(x',h)dh \right. \\
&\qquad\quad + \int_S h_k(h^* H_{1,1}g(x,x)h)(p(x,h) - p(x',h))dh \\
&\qquad\quad - \int_{S \cap S'^c} h_k(h^* H_{1,1}g(x,x)h)p(x',h)dh \\
&\qquad\quad \left. - \int_{S^c \cap S'} h_k(h^* H_{1,1}g(x',x')h)p(x',h)dh \right| .
\end{aligned} \tag{14}$$

In view of (H1), (H3) and Lemma 2.1, the first two terms of the right-hand side both converge to 0 when $x' \to x$. Observe that, as $x' \to x$, the sets $S^c \cap S'$ and $S \cap S'^c$ converge to $\{h : h \cdot \nabla_1 g(x,x) = 0\}$, which has Lebesgue measure 0 since $\nabla_1 g(x,x) \neq 0$ ($x \notin \Gamma$), and to $\emptyset$, respectively. So, by the dominated convergence Theorem, the last two terms of the right-hand side of (14) converge to 0 as $x' \to x$, and $\tilde{b}$ is continuous on $\mathcal{X} \setminus \Gamma$.

Proof of (ii) a is obviously symmetrical, and $\forall s = (s_1, \ldots, s_d) \in \mathbb{R}^d$, using the symmetry of $p(x,h)dh$, an easy calculation (the second line is obtained by the change of variable $h' = -h$) shows that

$$\begin{aligned}
s^* a(x) s &= \int_{\mathbb{R}^d} (h \cdot s)^2 [h \cdot \nabla_1 g(x,x)]_+ p(x,h)dh \\
&= \frac{1}{2} \int_{\mathbb{R}^d} (h \cdot s)^2 |h \cdot \nabla_1 g(x,x)| p(x,h)dh.
\end{aligned}$$

This is non-negative for all $s \in \mathbb{R}^d$, and is non zero if $s \neq 0$ and $x \notin \Gamma$. The fact that σ is 1/2-Hölder follows immediately from the following lemma, which proof is postponed after the proof of Proposition 4.2 :

Lemma 2.2 *Let $\mathcal{S}_+$ be the set of symmetrical non-negative $d \times d$ real matrices, define, for any $c > 0$*

$$\mathcal{S}_c = \{a \in \mathcal{S}_+ : \forall s \in \mathbb{R}^d,\, s^* a s \geq c\|s\|^2\}, \tag{15}$$

and let ζ be the function from $\mathcal{S}_+$ to $\mathcal{S}_+$ that maps the matrix a on its unique non-negative symmetrical square root. Then ζ is Hölder with exponent 1/2 on $\mathcal{S}_+$, and Lipschitz on $\mathcal{S}_c$ for any $c > 0$.

Proof of (iii) Fix $\alpha > 0$, $x \in \Gamma_\alpha$, and $s = (s_1, \dots, s_d) \in \mathbb{R}^d$. Denote by u and v the unit vectors of $\mathbb{R}^d$ such that $s = \|s\| u$ and $\nabla_1 g(x,x) = \|\nabla_1 g(x,x)\| v$. Then

$$\begin{aligned} s^* a(x) s &= \frac{1}{2} \|s\|^2 \|\nabla_1 g(x,x)\| \int_{\mathbb{R}^d} |h \cdot u|^2 |h \cdot v| p(x,dh) \\ &\geq C_\alpha \|s\|^2 \|\nabla_1 g(x,x)\| \end{aligned} \tag{16}$$

where C_α is given by (H4), and it follows from (H2) that there exists some $c > 0$ such that $a(x) \in \mathcal{S}_c$ for all $x \in \Gamma_\alpha$. Lemma 2.2 ends the proof of (iii).□

Proof of Lemma 2.2 Let $\tilde{\zeta} : \mathcal{S}_+ \to \mathcal{S}_+$ be defined by $\tilde{\zeta}(a) = a^2$. Then $\zeta \circ \tilde{\zeta} = \tilde{\zeta} \circ \zeta = \mathrm{Id}_{\mathcal{S}_+}$. The differential of $\tilde{\zeta}$ at a writes $d_a \tilde{\zeta}(h) = ah + ha$ and, when it is invertible, its inverse is the differential $d_{a^2}\zeta$ of ζ at a^2. Let $(ah+ha)_{ij}$ denote the i,j coefficient of the matrix $ah+ha$. Using the symmetry of a and h, an easy calculation gives that

$$\sum_{i,j} h_{ij} (ah + ha)_{ij} = \sum_{i=1}^{d} h_i^* a h_i + \sum_{j=1}^{d} h_j^* a h_j \tag{17}$$

where h_i is the i^{th} column (and, by symmetry, row) of the symmetrical matrix h. Let $\|\cdot\|$ be the spectral norm on $\mathcal{S}_+$. If $a \in \mathcal{S}_c$, the quantity (17) is greater than $2c \sum_i \|h_i\|^2 \geq Kc \|h\|^2$ for some constant K, and it is obviously smaller than $K' \|h\| \|ah + ha\|$ for another constant K'.

Hence, $\|ah + ha\| \geq \frac{Kc}{K'} \|h\|$ for any $h \in \mathcal{S}_+$ and $a \in \mathcal{S}_c$. In particular, $\|h\| = \|d_a \tilde{\zeta} \circ d_{a^2} \zeta(h)\| \geq \frac{Kc}{K'} \|d_{a^2} \zeta(h)\|$, which shows that ζ has a bounded differential on $\mathcal{S}_{c^2}$ for all $c > 0$. Therefore, ζ is Lipschitz on $\mathcal{S}_c$ for any $c > 0$.

The following proof that ζ is Hölder on $\mathcal{S}_+$ is taken from Serre [70]. Let us consider two matrices a and b in $\mathcal{S}_+$. Then, there exists an orthonormal basis in which a is diagonal, and another one where b is diagonal. So, there are two orthogonal matrices U and V, and two diagonal matrices D_1 and D_2 with respective non-negative diagonal elements $\lambda_1, \dots, \lambda_d$ and $\mu_1, \dots, \mu_d$, such that $a = U D_1 U^*$ and $b = U V D_2 V^* U^*$.

Define $A = D_1$ and $B = V D_2 V^*$, and denote by $\|\cdot\|_F$ the Schur-Frobenius norm on $d \times d$ matrices, given by $\|M\|_F = \sqrt{\sum_{i,j} m_{ij}^2}$ where $M = (m_{ij})_{1 \leq i,j \leq d}$. Then,

$$\begin{aligned} \|B - A\|_F^2 &= \sum_{i,j} \left(\sum_k v_{ik} \mu_k v_{jk} - \lambda_i \delta_{ij} \right)^2 \\ &= \sum_{i,j} \sum_{k,l} v_{ik} v_{il} v_{jk} v_{jl} \mu_k \mu_l - 2 \sum_i \sum_k v_{ik}^2 \lambda_i \mu_k + \sum_i \lambda_i^2 \\ &= \sum_k \mu_k^2 - 2 \sum_i \sum_k v_{ik}^2 \lambda_i \mu_k + \sum_i \lambda_i^2, \end{aligned}$$

where δ_{ij} is the Kronecker delta symbol, and where we used the fact that

$V = (v_{ij})_{1\leq i,j\leq d}$ is an orthogonal matrix to obtain the last line. This can be rewritten as

$$\|B-A\|_F^2 = \sum_{i,j} v_{ij}^2(\lambda_i-\mu_j)^2.$$

Now, observe that $\zeta(a) = U\zeta(A)U^* = U\zeta(D_1)U^*$, that $\zeta(b) = U\zeta(B)U^* = UV\zeta(D_2)V^*U^*$, and that $\zeta(D_1)$ and $\zeta(D_2)$ are diagonal matrices with respective elements $\sqrt{\lambda_1},\dots,\sqrt{\lambda_d}$ and $\sqrt{\mu_1},\dots,\sqrt{\mu_d}$. Then, it follows from the fact that $(\sqrt{\lambda}-\sqrt{\mu})^2 \leq |\lambda-\mu|$ for $\lambda\geq 0$ and $\mu\geq 0$, and from the Cauchy-Schwartz inequality, that

$$\begin{aligned}\|\zeta(A)-\zeta(B)\|_F^4 &= \left(\sum_{i,j} v_{ij}^2(\sqrt{\lambda_i}-\sqrt{\mu_j})^2\right)^2 \leq \left(\sum_{i,j} v_{ij}^2|\lambda_i-\mu_j|\right)^2\\ &\leq \sum_{i,j} v_{ij}^2 \sum_{i,j} v_{ij}^2(\lambda_i-\mu_j)^2 = d\|A-B\|_F^2.\end{aligned}$$

Observing that, for any orthogonal matrix U, $\|UMU^*\|_F = \|M\|_F$, it finally follows that $\|\zeta(a)-\zeta(b)\|_F \leq d^{1/4}\sqrt{\|a-b\|_F}$ for any a and b in $\mathcal{S}_+$. ☐

3 Weak existence, uniqueness in law and Markov property

We will in this section construct a particular solution to the martingale problem associated with (1), identify the difficulty for uniqueness and for the strong Markov property, and give some conditions solving this difficulty, both in dimension $d = 1$, and for greater dimensions.

Let us fix $\varepsilon > 0$ until the end of this section.

3.1 Construction of a particular solution to (1)

Let us define, on the canonical space $\mathcal{C}(\mathbb{R}_+,\mathbb{R}^d)$, the stopping time

$$\tau = \inf\{t\geq 0 : w_t\in\Gamma\}, \tag{18}$$

where w_t is the canonical process on $\mathcal{C}(\mathbb{R}_+,\mathbb{R}^d)$.

Theorem 3.1 *Under assumptions (H1)–(H4), for any $x\in\overline{\mathcal{X}}$, there exists a probability measure on $\mathcal{C}(\mathbb{R}_+,\mathbb{R}^d)$ solution to the martingale problem associated with the SDE (1) with initial state x, under which the canonical process is constant after τ.*

Proof By Proposition 2.1, the function a (resp. $\tilde{b}$) extended by 0 to $\mathbb{R}^d\setminus\overline{\mathcal{X}}$ is globally Lipschitz, bounded and uniformly non-degenerate (resp. bounded and continuous) on Γ_α for any $\alpha>0$. Therefore, there exists a globally Lipschitz, bounded and uniformly non-degenerate (resp. bounded and continuous) function $a^\alpha:\mathbb{R}^d\rightarrow\mathcal{S}_+$ (resp. $\tilde{b}^\alpha:\mathbb{R}^d\rightarrow\mathbb{R}^d$) such that $a^\alpha(x)=a(x)$ (resp. $\tilde{b}^\alpha(x)=\tilde{b}(x)$) for any $x\in\Gamma_\alpha$. By Lemma 2.2, the square root $\sigma^\alpha\in\mathcal{S}_+$ of a^α is also globally Lipschitz, bounded and uniformly non-degenerate.

If we also extend by 0 the function b out of $\overline{\mathcal{X}}$, by Girsanov's theorem (cf. e.g. Theorem IX.1.10 of Revuz and Yor [68]), the weak existence and uniqueness in law on $\mathbb{R}^d$ for the SDE

$$dX^{\varepsilon,\alpha}_t=(b(X^{\varepsilon,\alpha}_t)+\varepsilon\tilde{b}^\alpha(X^{\varepsilon,\alpha}_t))dt+\sqrt{\varepsilon}\sigma^\alpha(X^{\varepsilon,\alpha}_t)dW_t \tag{19}$$

is equivalent to the weak existence and uniqueness in law on $\mathbb{R}^d$ for

$$dX_t=\sqrt{\varepsilon}\sigma^\alpha(X_t)dW_t,$$

which hold since σ^α is globally Lipschitz.

Fix $x\in\overline{\mathcal{X}}$. Let us denote by $\mathbf{P}^\alpha$ the unique solution to the martingale problem associated with (19) with initial state x. Let $\{\mathcal{G}_t\}$ denote the canonical filtration in $\mathcal{C}(\mathbb{R}_+,\mathbb{R}^d)$ ($\mathcal{G}_t=\sigma(w_s,0\leq s\leq t)$), and remind that it is left-continuous (cf. Karatzas and Shreve [45] problem 2.7.1 p. 89). Since $a_\alpha=a_{\alpha'}$ on Γ_α for any $\alpha'<\alpha$, by the uniqueness of (19), then

$$\forall\alpha'<\alpha,\quad \mathbf{P}^\alpha=\mathbf{P}^{\alpha'}\quad\text{on }\mathcal{G}_{\tau_\alpha},$$
$$\text{where}\quad \tau_\alpha=\inf\{t\geq 0: w_t\notin\Gamma_\alpha\}.$$

Since $\mathcal{G}_{\tau_\alpha}\subset\mathcal{G}_{\tau_{\alpha'}}$ when $\alpha'<\alpha$, we can define, for any $\alpha>0$,

$$\mathbf{P}:=\mathbf{P}^\alpha\quad\text{on}\quad\mathcal{G}_{\tau_\alpha}. \tag{20}$$

By Caratheodory's theorem, this relation defines a probability measure $\mathbf{P}$ on $\sigma(\mathcal{G}_{\tau_\alpha},\alpha>0)=\mathcal{G}_{\tau-}=\mathcal{G}_\tau$, where τ has been defined in (18).

Define

$$\tilde{\mathcal{C}}:=\{w\in\mathcal{C}(\mathbb{R}_+,\mathbb{R}^d):\forall t\geq\tau,\ w_t=w_\tau\}.$$

Since $\tilde{\mathcal{C}}^c=\{w\in\mathcal{C}(\mathbb{R}_+,\mathbb{R}^d):\exists t>\tau,\ w_t\neq w_\tau\}\in\mathcal{G}_\infty\setminus\mathcal{G}_\tau$, we can put

$$\mathbf{P}(\tilde{\mathcal{C}}^c):=0. \tag{21}$$

Now, observe that, for any $A\in\mathcal{G}_\infty$,

$$\tilde{\mathcal{C}}\cap A=\tilde{\mathcal{C}}\cap\{w_{\cdot\wedge\tau}\in A\}\quad\text{and}\quad\{w_{\cdot\wedge\tau}\in A\}\in\mathcal{G}_\tau.$$

Therefore, by (21), for any $A \in \mathcal{G}_\infty$, we can put

$$\mathbf{P}(A) = \mathbf{P}\{w_{\cdot\wedge\tau} \in A\},$$

which is properly defined by (20).

Therefore, we have defined a probability measure $\mathbf{P}$ on $(\mathcal{C}(\mathbb{R}_+,\mathbb{R}^d),\mathcal{G}_\infty)$. We will prove below that this probability measure solves the martingale problem associated with (1).

Define, for any $\mathcal{C}^2$ function with compact support $\varphi : \mathbb{R}^d \to \mathbb{R}$,

$$M_t^\varphi := \varphi(w_t) - \varphi(w_0) - \int_0^t L\varphi(w_s)dt,$$

$$\text{where} \quad L\varphi(x) = (b(x) + \varepsilon\tilde{b}(x)) \cdot \nabla\varphi(x) + \frac{\varepsilon}{2}\sum_{1\leq i,j\leq d} a_{ij}(x)\frac{\partial^2\varphi}{\partial x_i \partial x_j}(x).$$

Let us prove that M^φ is a $\mathbf{P}$-martingale. Since $\mathbf{P}^\alpha$ is a solution to the martingale problem associated with (19), since a, b and $\tilde{b}$ are bounded, and since $a^\alpha = a$ and $\tilde{b}^\alpha = \tilde{b}$ on Γ_α, $M^\varphi_{\cdot\wedge\tau_\alpha}$ is a $\mathbf{P}^\alpha$-martingale.

Let us denote by $\mathbf{E}$ (resp. $\mathbf{E}^\alpha$) the expectation with respect to $\mathbf{P}$ (resp. $\mathbf{P}^\alpha$). Observe that, since $\mathbf{P}(\tilde{\mathcal{C}}) = 1$, $\mathbf{1}_{\{t\geq\tau\}}w_t = \mathbf{1}_{\{t\geq\tau\}}w_\tau$ $\mathbf{P}$-a.s. for any $t \geq 0$. Therefore, for any $t > s$ and $\alpha > 0$,

$$\begin{aligned}
&\mathbf{E}(M_t^\varphi \mid \mathcal{G}_s) \\
&\quad= \mathbf{E}\left(\mathbf{1}_{\{s<\tau_\alpha\}}M_t^\varphi \mid \mathcal{G}_s\right) + \mathbf{E}\left(\mathbf{1}_{\{\tau_\alpha\leq s<\tau\}}M_t^\varphi \mid \mathcal{G}_s\right) + \mathbf{E}\left(\mathbf{1}_{\{s\geq\tau\}}M_t^\varphi \mid \mathcal{G}_s\right) \\
&\quad= \mathbf{E}^\alpha\left(\mathbf{1}_{\{s<\tau_\alpha\}}M_t^\varphi \mid \mathcal{G}_s\right) + \mathbf{E}\left(\mathbf{1}_{\{\tau_\alpha\leq s<\tau\}}M_t^\varphi \mid \mathcal{G}_s\right) + \mathbf{E}\left(\mathbf{1}_{\{s\geq\tau\}}M_s^\varphi \mid \mathcal{G}_s\right) \\
&\quad= \mathbf{1}_{\{s<\tau_\alpha\}}M_s^\varphi + \mathbf{E}\left(\mathbf{1}_{\{\tau_\alpha\leq s<\tau\}}M_t^\varphi \mid \mathcal{G}_s\right) + \mathbf{1}_{\{s\geq\tau\}}M_s^\varphi.
\end{aligned}$$

Since $\tau_\alpha \to \tau$ a.s. when $\alpha \to 0$, $\mathbf{1}_{\{s<\tau_\alpha\}}M_s^\varphi \to \mathbf{1}_{\{s<\tau\}}M_s^\varphi$. Moreover, since φ and $L\varphi$ are bounded, M_t^φ is bounded, and so $\mathbf{E}(\mathbf{1}_{\{\tau_\alpha\leq s<\tau\}}M_t^\varphi \mid \mathcal{G}_s) \to 0$ when $\alpha \to 0$. Therefore, $\mathbf{E}(M_t^\varphi \mid \mathcal{G}_s) = M_s^\varphi$, and thus M^φ is a martingale.

Now, fix a $\mathcal{C}^2$ function φ, define the sequence of stopping times

$$T^n = \inf\{t \geq 0 : \|w_t\| \geq n\}$$

and take, for any $n \geq 1$, a $\mathcal{C}^2$ function φ_n with compact support equal to φ on $B(0,n) = \{x \in \mathbb{R}^d : \|x\| \leq n\}$. Then, $M^\varphi_{t\wedge T^n} = M^{\varphi_n}_{t\wedge T^n}$ is a $\mathbf{P}$-martingale. Since $T^n \to +\infty$ a.s. when $n \to +\infty$, we finally obtain that M^φ is a local martingale for any $\mathcal{C}^2$ function φ.

Therefore, $\mathbf{P}$ solves the martingale problem associated with (1) with initial state $x \in \overline{\mathcal{X}}$, which ends the proof of Theorem 3.1. □

3.2 Uniqueness and strong Markov property

No standard technique applies directly to prove the uniqueness in law or the strong Markov property for solutions to (1). This comes from the fact that σ degenerates at points of Γ, *and* that $\tilde{b}$ is not continuous at these points. Uniqueness is known to hold when only one of these difficulties arises, but the combination of both of them leads to great difficulties.

The strong Markov property for solutions to SDEs is known to be linked to the uniqueness of solutions to the corresponding martingale problem. Here, we are only able to solve these questions under particular assumptions.

Let us first recall the following classical result (cf. e.g. Karatzas and Shreve [45] Proposition 5.4.6 p. 315): let $\mathbf{P}_x$ be the solution to the martingale problem associated with (1) obtained in the previous section. One can classically construct from $\mathbf{P}_x$ a Brownian motion W on $(\Omega,\mathcal{F}_t,\mathbf{Q}_x) = (\mathcal{C}(\mathbb{R}_+,\mathbb{R}^d)\times\mathcal{C}(\mathbb{R}_+,\mathbb{R}^d),\mathcal{G}_t\otimes\mathcal{G}_t,\mathbf{P}_x\otimes\mathbf{W}^d)$, where $\mathbf{W}^d$ is the d-dimensional Wiener measure, such that the canonical process $X_t(w,w')=w(t)$ is solution to (1) with respect to W under $\mathbf{Q}_x$.

Theorem 3.2 *Assume (H1)–(H4) and consider the canonical process solution to (1) on $(\Omega,\mathcal{F}_t,W,\mathbf{Q}_x)$ with initial state x constructed above. Let us denote it by X^ε. Its law is still denoted by $\mathbf{P}_x$.*

(a) *Suppose that the stopping time τ defined in (18) satisfies for some $x\in\mathcal{X}\setminus\Gamma$*

$$\mathbf{P}_x(\tau=\infty)=1. \tag{22}$$

Then, uniqueness in law holds for solutions to (1) with initial state x.

(b) *For any $\mathcal{F}_t$-stopping time $S<\tau$, for any $t>0$, for any continuous and bounded function $\varphi:\overline{\mathcal{X}}\to\mathbb{R}$ and for any $x\in\mathcal{X}\setminus\Gamma$,*

$$\mathbf{E}_x(\varphi(X^\varepsilon_{S+t})|\mathcal{F}_S)=\mathbf{E}_x(\varphi(X^\varepsilon_{S+t})|X^\varepsilon_S), \tag{23}$$

where $\mathbf{E}_x$ denotes the expectation with respect to $\mathbf{Q}_x$.

(c) *Assume (22) for all $x\in\mathcal{X}\setminus\Gamma$. Then X^ε satisfies the strong Markov property.*

Proof As we have seen in the proof of Theorem 3.1, the martingale problem associated with (19) is well-posed, and thus, the strong Markov property holds for (19) (cf. [45] Theorem 5.4.20 p. 322). Let us denote by $\mathbf{P}^\alpha_x$ the (unique) law of the solution to (19) with initial state x.

Let us construct for any $\alpha>0$ a process $X^{\varepsilon,\alpha}$ solution to (19) on the canonical space Ω with initial state x, as we did for X^ε: $X^{\varepsilon,\alpha}$ is the canonical process $X_t(w,w') = w(t)$, and it is solution to (19) under the probability measure $\mathbf{Q}^\alpha_x = \mathbf{P}^\alpha_x \times \mathbf{W}^d$ and with respect to a $\mathbf{Q}^\alpha_x$-Brownian motion W^α.

Now, define $\tau_\alpha = \inf\{t \geq 0 : X^\varepsilon_t \notin \Gamma_\alpha\}$. Then, for $t \leq \tau_\alpha$, $\tilde{b}(X^\varepsilon_t) = \tilde{b}_\alpha(X^\varepsilon_t)$ and $\sigma(X^\varepsilon_t) = \sigma_\alpha(X^\varepsilon_t)$, so X^ε is solution to (19) for $t \leq \tau_\alpha$, for which uniqueness in law holds. Therefore, uniqueness in law holds for solution to (1) for $t \leq \tau_\alpha$ for any $\alpha > 0$. In particular, when (22) holds, $\tau_\alpha \to +\infty$ when $\alpha \to 0$, and thus point (a) is proved.

Note also that this implies that $(\mathbf{Q}_x)_{|\mathcal{F}_{\tau_\alpha}} = (\mathbf{Q}^\alpha_x)_{|\mathcal{F}_{\tau_\alpha}}$.

Let us come to the proof of (b). Fix S, t, φ and x as in the statement of (b). Since X^ε is $\mathbf{Q}_x$-a.s. constant after time τ, (23) is equivalent to the existence of a Lebesgue-measurable function $f : \mathbb{R}^d \to \mathbb{R}$ such that

$$\mathbf{E}_x(\varphi(X^\varepsilon_{(S+t)\wedge\tau})|\mathcal{F}_S) = f(X^\varepsilon_S). \tag{24}$$

The strong Markov property for $X^{\varepsilon,\alpha}$ implies that, for any $\alpha > 0$, there is a Lebesgue-measurable function f_α from $\mathbb{R}^d$ to $[0,1]$ such that

$$\mathbf{E}^\alpha_x(\mathbf{1}_{\tau_\alpha > S}\varphi(X^{\varepsilon,\alpha}_{(S+t)\wedge\tau_\alpha})|\mathcal{F}_S) = f_\alpha(X^{\varepsilon,\alpha}_S),$$

where $\mathbf{E}^\alpha_x$ denotes the expectation with respect to $\mathbf{Q}^\alpha_x$. Therefore, since $X^{\alpha,\varepsilon} = w = X^\varepsilon$ and $(\mathbf{Q}_x)_{|\mathcal{F}_{\tau_\alpha}} = (\mathbf{Q}^\alpha_x)_{|\mathcal{F}_{\tau_\alpha}}$,

$$\mathbf{E}_x(\mathbf{1}_{\tau_\alpha > S}\varphi(X^\varepsilon_{(S+t)\wedge\tau_\alpha})|\mathcal{F}_S) = f_\alpha(X^\varepsilon_S).$$

Now, it follows from the fact that $\tau_\alpha \to \tau > S$ a.s. as $\alpha \to 0$ and from the dominated convergence Theorem that, when $\alpha \to 0$,

$$\mathbf{E}_x(\mathbf{1}_{\tau_\alpha > S}\varphi(X^\varepsilon_{(S+t)\wedge\tau_\alpha})|\mathcal{F}_S) \to \mathbf{E}_x(\varphi(X^\varepsilon_{(S+t)\wedge\tau})|\mathcal{F}_S) \text{ a.s.}$$

Thus $f_\alpha(X^\varepsilon_S)$ converges almost surely to a $\sigma(X^\varepsilon_S)$-measurable function (as a limit of $\sigma(X^\varepsilon_S)$-measurable function), a.s. equal to $\mathbf{E}_x(\varphi(X^\varepsilon_{(S+t)\wedge\tau})|\mathcal{F}_S)$. This completes the proof of (24).

Finally, under the assumption (22), (c) is a trivial consequence of (b). □

3.3 Study of (22): the dimension 1 case

As we saw above, the uniqueness and the strong Markov property of X^ε rely on the fact that $\mathbf{P}_x(\tau = \infty) = 1$, where τ has been defined in (18). It is possible, in dimension $d = 1$, to give conditions under which this is true. In this case, an elementary calculation gives the following formulas for a, b and $\tilde{b}$:

$$b(x) = \frac{M_2(x)}{2}\partial_1 g(x,x), \quad \tilde{b}(x) = \frac{M_3(x)}{4}\text{sign}[\partial_1 g(x,x)]\partial^2_{1,1}g(x,x),$$
$$a(x) = \frac{M_3(x)}{2}|\partial_1 g(x,x)|, \quad \text{where} \quad M_k(x) = \int_{\mathbb{R}} |h|^k p(x,h)dh$$
$$\text{and} \quad \text{sign}(x) = -1 \text{ if } x < 0;\ 0 \text{ if } x = 0;\ 1 \text{ if } x > 0.$$

Theorem 3.3 *Assume (H1), (H3), that $d = 1$, and that g is $\mathcal{C}^3$ with bounded third-order derivatives. Let X^ε be a solution to (1) starting at $x \not\in \Gamma$. Define $c = \sup\{y \in \Gamma, y < x\}$, $c' = \inf\{y \in \Gamma, y > x\}$, and assume that c and c' do not belong to $\partial\mathcal{X}$, that $-\infty < c < c' < \infty$, and that $\partial^2_{1,1}g(c,c) + \partial^2_{1,2}g(c,c) \neq 0$ and $\partial^2_{1,1}g(c',c') + \partial^2_{1,2}g(c',c') \neq 0$. Then we can define*

$$\begin{aligned} \alpha &:= \frac{\partial^2_{1,1}g(c,c)}{\partial^2_{1,1}g(c,c) + \partial^2_{1,2}g(c,c)} = \frac{2\partial^2_{1,1}g(c,c)}{\partial^2_{1,1}g(c,c) - \partial^2_{2,2}g(c,c)} \\ \beta &:= \frac{\partial^2_{1,1}g(c',c')}{\partial^2_{1,1}g(c',c') + \partial^2_{1,2}g(c',c')} = \frac{2\partial^2_{1,1}g(c',c')}{\partial^2_{1,1}g(c',c') - \partial^2_{2,2}g(c',c')}. \end{aligned} \tag{25}$$

These equalities follow from the fact that $\partial^2_{1,1}g + 2\partial^2_{1,2}g + \partial^2_{2,2}g = 0$, obtained by differentiating equation (2). Then, we distinguish four cases:

(a) *If $\alpha \geq 1$ and $\beta \leq -1$, then $\mathbf{P}(\tau = \infty) = 1$ and the process X^ε is recurrent in (c,c').*

(b) *If $\alpha \geq 1$ and $\beta > -1$, then $\mathbf{P}(\tau < \infty) = 1$ and $\mathbf{P}(\lim_{t\to\tau} X^\varepsilon_t = c') = 1$.*

(c) *If $\alpha < 1$ and $\beta \leq -1$, then $\mathbf{P}(\tau < \infty) = 1$ and $\mathbf{P}(\lim_{t\to\tau} X^\varepsilon_t = c) = 1$.*

(d) *If $\alpha < 1$ and $\beta > -1$, then $\mathbf{P}(\tau < \infty) = 1$, $\mathbf{P}(\lim_{t\to\tau} X^\varepsilon_t = c) > 0$ and $\mathbf{P}(\lim_{t\to\tau} X^\varepsilon_t = c') = 1 - \mathbf{P}(\lim_{t\to\tau} X^\varepsilon_t = c) > 0$.*

Remark 3.1 *The conditions $\partial^2_{1,1}g(c,c) + \partial^2_{1,2}g(c,c) \neq 0$ and $\partial^2_{1,1}g(c',c') + \partial^2_{1,2}g(c',c') \neq 0$ are only technical. A higher order calculation is possible if one of them does not hold.*

Remark 3.2 *When $c = -\infty$ or $c' = \infty$, the calculation below depends on technical properties of g and M_k, and no simple general result can be stated.*

Biological remark 3.3 *The biological theory of adaptive dynamics gives a classification of the evolutionary singularities in dimension $d = 1$, depending on the values of $\partial^2_{1,1}g$ and $\partial^2_{2,2}g$ at these points. Here, the condition $\alpha \geq 1$ corresponds, when $\partial^2_{1,1}g(c,c) - \partial^2_{2,2}g(c,c) > 0$, to the case $\partial^2_{1,1}g(c,c) + \partial^2_{2,2}g(c,c) \geq 0$, which corresponds in the biological terminology (see* e.g. *Diekmann [19]) to a* converging stable strategy with mutual invasibility*, which include the* evolutionary branching *condition; and when $\partial^2_{1,1}g(c,c) - \partial^2_{2,2}g(c,c) < 0$, to the case $\partial^2_{1,1}g(c,c) + \partial^2_{2,2}g(c,c) \leq 0$, which corresponds biologically to a* repelling strategy without mutual invasibility*.*

Proof of Theorem 3.3 We will use the classical methods of removal of drift of Engelbert and Schmidt and the explosion criterion of Feller (see [45] pp. 345–351). They can be applied to X^ε, considered as a process with value in (c,c') killed when it hits c or c', under the following assumptions, obviously fulfilled by our process:

$$\begin{gathered}\forall x \in (c,c'),\ \sigma(x) > 0,\\ \text{and } \forall x \in (c,c'),\ \exists \delta > 0 \text{ such that } \int_{x-\delta}^{x+\delta} \frac{1+|b^\varepsilon(y)|}{\varepsilon\sigma^2(y)}dy < \infty.\end{gathered}$$

These methods involve the two following functions, defined for a fixed $\gamma \in (c,c')$:

$$\begin{aligned} p(x) &= \int_\gamma^x \exp\left[-2\int_\gamma^y \frac{b^\varepsilon(z)dz}{\varepsilon\sigma^2(z)}\right]dy,\ \forall x \in (c,c'),\\ \text{and}\quad v(x) &= \int_\gamma^x p'(y)\int_\gamma^y \frac{2dz}{\varepsilon p'(z)\sigma^2(z)}dy,\ \forall x \in (c,c'). \end{aligned} \tag{26}$$

Then, as can be seen in [45] pages 345 (Proposition 5.5.22), 348 (Theorem 5.5.29) and 351 (Proposition 5.5.32), the statements about the limit of the process X^ε_t when $t \to \tau$ and about the recurrence of X^ε depend on whether $p(x)$ is finite or not when $x \to c$ and $x \to c'$ (point (a) corresponds to $p(c+) = -\infty$ and $p(c'-) = +\infty$, point (b) to $p(c+) = -\infty$ and $p(c'-) < +\infty$, etc.), and the statements about τ depends on whether $v(x) < \infty$ or not when $x \to c$ and c' ($\mathbf{P}(\tau = \infty) = 1$ if and only if $v(c+) = v(c'-) = \infty$, and $\mathbf{P}(\tau < \infty) = 1$ if and only if $v(c+) < \infty$ and $v(c'-) < \infty$, or $v(c+) < \infty$ and $p(c'-) = +\infty$, or $v(c'-) < \infty$ and $p(c+) = -\infty$).

Let us compute these limits.

$$\frac{b^\varepsilon(x)}{\varepsilon\sigma^2(x)} = \frac{b^\varepsilon(x)}{\varepsilon a(x)} = \frac{M_2(x)}{\varepsilon M_3(x)}\operatorname{sign}[\partial_1 g(x,x)] + \frac{1}{2}\frac{\partial^2_{1,1}g(x,x)}{\partial_1 g(x,x)}, \tag{27}$$

so, for $x < y < \gamma$, the quantity inside the exponential appearing in the definition of p writes

$$\int_y^\gamma \frac{2M_2(z)}{\varepsilon M_3(z)}\operatorname{sign}[\partial_1 g(z,z)]dz + \int_y^\gamma \frac{\partial^2_{1,1}g(z,z)}{\partial_1 g(z,z)}dz.$$

Since $c \notin \partial\mathcal{X}$, the first term is bounded for $c < y < \gamma$ (by assumption (H2), M_3 is positive and continuous on $[c,\gamma]$, so it is bounded away of 0 on this interval), so we only have to study the second term.

When $y \to c$, an easy calculation gives

$$\frac{\partial^2_{1,1}g(z,z)}{\partial_1 g(z,z)} = \frac{\alpha}{z-c} + C + o(1),$$

where α is defined in (25), and where C is a constant depending on the derivatives of g at $(c{,}c)$ up to order 3. Consequently, when $y \to c$,

$$\begin{aligned}\exp\left[-2\int_\gamma^y \frac{b^\varepsilon(z)dz}{\varepsilon\sigma^2(z)}\right] &= \exp\left[C' + o(1) + \int_y^\gamma \left(\frac{\alpha}{y-c} + C + o(1)\right) dz\right] \\ &\sim e^C (y-c)^{-\alpha}, \qquad (28)\end{aligned}$$

where the constant C may change from line to line.

Therefore, if $\alpha < 1$, $p(c+) > -\infty$, and if $\alpha \geq 1$, $p(c+) = -\infty$. The same computation gives the required result when $x \to c'$.

Now let us compute the limit of v at c and c'. Since $p(c'-) = \infty \Rightarrow v(c'-) = \infty$ and $p(c+) = -\infty \Rightarrow v(c+) = \infty$ (see [45] page 348), we only have to deal with the cases $\alpha < 1$ and $\beta > -1$.

Equation (28) implies that $p'(y) \sim e^C(y-c)^{-\alpha}$, so, there exists a positive constant C such that

$$\frac{2}{\varepsilon p'(z)a(z)} \sim C(z-c)^{\alpha-1},$$

since

$$a(z) = \frac{M_3(z)}{2}|\partial_1 g(z{,}z)| \sim \frac{M_3(c)}{2}|\partial^2_{1,1} g(c{,}c) + \partial^2_{1,2} g(c{,}c)|(z-c). \qquad (29)$$

If $\alpha < 0$, when $y \to c$, $p'(y)\int_y^\gamma \frac{2dz}{\varepsilon p'(z)a(z)} \sim -Cp'(y)(y-c)^\alpha$ is bounded on $(c{,}\gamma)$, and so $v(c+) < \infty$. If $\alpha = 0$, $p'(y)\int_y^\gamma \frac{2dz}{\varepsilon p'(z)a(z)} \sim C\log(y-c)$, which has a finite integral on $(c{,}\gamma)$, so $v(c+) < \infty$. Finally, if $0 < \alpha < 1$, $\int_y^\gamma \frac{2dz}{\varepsilon p'(z)a(z)}$ is bounded, so $v(c+) < \infty$ is equivalent to the convergence of the integral $\int_c^\gamma p'(y)dy$, which holds since $p'(y) \sim \frac{C}{(y-a)^\alpha}$ and $\alpha < 1$. ☐

In the case where c or c' belong to $\partial\mathcal{X}$, what changes in the calculation above? Assume for example that $c \in \partial\mathcal{X}$. The problem is that $M_2(x)$ and $M_3(x)$ are not bounded away from zero in the neighborhood of c. Indeed, the support of $p(x{,}h)dh$ is a subset of $\mathcal{X} - x$ which is symmetrical with respect to 0, so it is a subset of $(-(x-c){,}x-c)$, which converges to $\{0\}$ when $x \to c$. So the quantity $2\frac{b(x)}{a(x)} = \frac{M_2(x)}{\varepsilon M_3(x)}\mathrm{sign}[\partial_1 g(x{,}x)]$ appearing in the equation (27) may not be bounded in the neighborhood of c.

If $x \in \mathcal{X}$, the support of the symmetrical measure $p(x{,}h)dh$ is a subset of $(-(x-c){,}x-c)$, so $\int |h|^3 p(x{,}h)dh \leq (x-c)\int |h|^2 p(x{,}h)dh$, *i.e.*

$$\frac{M_2(x)}{M_3(x)} \geq \frac{1}{x-c}.$$

Since $\Gamma \cap (c,c') = \emptyset$, the sign of $\partial_1 g(x,x)$ is constant on (c,c'), equal to the sign of $\partial^2_{1,1} g(c,c) + \partial^2_{1,2} g(c,c)$ (by expanding $\partial_1 g(x,x)$ in the neighborhood of c), that is non-null by assumption. Let us call s this sign.

If $s = +1$, then, for $c < y < \gamma$, there is a constant $C > 0$ such that

$$\exp\left[\frac{2}{\varepsilon}\int_y^\gamma \frac{M_2(z)}{M_3(z)}\text{sign}[\partial_1 g(z,z)]dz\right] \geq \frac{C}{(y-c)^{2/\varepsilon}}.$$

Combining this fact with the estimations obtained in the proof of Theorem 3.3 for the term $\frac{\partial^2_{1,1} g(z,z)}{\partial_1 g(z,z)}$, we see that

$$p'(y) = \exp\left[-2\int_\gamma^y \frac{b^\varepsilon(z)dz}{\varepsilon\sigma^2(z)}\right] \geq \frac{C}{(y-c)^{\alpha+2/\varepsilon}}, \tag{30}$$

and we finally obtain that $p(c+) > -\infty$ if $\alpha + \frac{2}{\varepsilon} < 1$.

In the case where $s = -1$, we obtain

$$p'(y)\exp\left[-2\int_\gamma^y \frac{b^\varepsilon(z)dz}{\varepsilon\sigma^2(z)}\right] \leq \frac{C}{(y-c)^{\alpha-2/\varepsilon}}, \tag{31}$$

and so $p(c+) = -\infty$ if $\alpha - \frac{2}{\varepsilon} \geq 1$.

Observe that if, for example,

$$\exists \rho > 0,\ \exists 0 < \theta < 1,\ \forall x \in \mathcal{X},\ p(x,h) \geq \rho \mathbf{1}_{B(0,\theta(x-c))}, \tag{32}$$

we easily obtain that $\frac{M_2(x)}{M_3(x)} \geq \frac{C}{x-c}$, and, consequently, the converse inequalities in (30) and (31) hold. So, under this assumption, if $s = +1$, $p(c+) > -\infty \Leftrightarrow \alpha + \frac{2}{\varepsilon} < 1$, and if $s = -1$, $p(c+) > -\infty \Leftrightarrow \alpha - \frac{2}{\varepsilon} < 1$.

Concerning the limit of the function v at c, observe that assumption (H2) implies that M_3 is a Lipschitz function, so there is some constant K such that $M_3(x) \leq K(x-c)$. In equation (29), this gives that $a(z) \leq C(z-c)^2$ for some constant C.

Then, in the case $s = +1$, if we assume (32), we obtain

$$\frac{2}{\varepsilon p'(z)a(z)} \geq C(z-c)^{\alpha-2+2/\varepsilon}.$$

Thus, if $\alpha + 2/\varepsilon < 1$,

$$p'(y)\int_y^\gamma \frac{2dz}{\varepsilon p'(z)a(z)} \geq \frac{C}{(y-c)^{\alpha+2/\varepsilon}} - \frac{C'}{y-c},$$

so the first term of the right-hand side has a finite integral on (c,γ), and the second term has a divergent integral, so $v(c+) = \infty$.

In the case where $s = -1$, the same calculation can be made replacing $\alpha + 2/\varepsilon$ by $\alpha - 2/\varepsilon$, and gives that $v(c+) = \infty$ when $\alpha - 2/\varepsilon < 1$.

Let us collect all these results in the following theorem:

Theorem 3.4 *With the same notations and assumptions as in Theorem 3.3, except that* $c \in \partial\mathcal{X}$ *and* $c' \notin \partial\mathcal{X}$*, and with the notation* $s = sign[\partial^2_{1,1}g(c,c) + \partial^2_{1,2}g(c,c)]$,

(a) *If* $s = -1$, $\alpha - 2/\varepsilon \geq 1$ *and* $\beta \leq -1$*, then the process* X^ε *is recurrent in* (c,c')*. The same holds if* $s = 1$, $\alpha + 2/\varepsilon \geq 1$ *and* $\beta \leq -1$*, under the additional assumption (32). In both cases,* $\mathbf{P}(\tau = \infty) = 1$.

(b) *If* $s = -1$, $\alpha - 2/\varepsilon \geq 1$ *and* $\beta > -1$*, then* $\mathbf{P}(\lim_{t\to\tau} X^\varepsilon_t = c') = 1$*. The same holds if* $s = 1$, $\alpha + 2/\varepsilon \geq 1$ *and* $\beta > -1$*, under the additional assumption (32). In both cases,* $\mathbf{P}(\tau < \infty) = 1$.

(c) *If* $s = 1$, $\alpha + 2/\varepsilon < 1$ *and* $\beta \leq -1$*, then* $\mathbf{P}(\lim_{t\to\tau} X^\varepsilon_t = c) = 1$*. The same holds if* $s = -1$, $\alpha - 2/\varepsilon \geq 1$ *and* $\beta \leq -1$*, under the additional assumption (32). In both cases, under assumption (32),* $\mathbf{P}(\tau = \infty) = 1$.

(d) *If* $s = 1$, $\alpha + 2/\varepsilon < 1$ *and* $\beta > -1$*, then* $\mathbf{P}(\lim_{t\to\tau} X^\varepsilon_t = c) > 0$ *and* $\mathbf{P}(\lim_{t\to\tau} X^\varepsilon_t = c') = 1 - \mathbf{P}(\lim_{t\to\tau} X^\varepsilon_t = c) > 0$*. The same holds if* $s = -1$, $\alpha - 2/\varepsilon \geq 1$ *and* $\beta > -1$*, under the additional assumption (32). In both cases,* $0 < \mathbf{P}(\tau < \infty) < 1$.

Remark 3.4 *A similar result can be stated in the case where* $c \notin \partial\mathcal{X}$ *and* $c' \in \partial\mathcal{X}$*, and in the case where* c *and* c' *are both in* $\partial\mathcal{X}$.

3.4 Study of (22): the dimension $d \geq 2$ case

Let us turn now to the case $d \geq 2$. We will consider in this section the process X^ε constructed on the filtered probability space $(\Omega,\mathcal{F}_t,\mathbf{Q}_x)$ of section 3.2 for which part (b) of Theorem 3.2 holds. If we restrict ourself to the case $\mathcal{X} = \mathbb{R}^d$, it is possible to give conditions under which $\mathbf{P}_x(\tau = \infty) = 1$, based on a comparison of $d(X^\varepsilon,\Gamma)$ with Bessel processes:

Theorem 3.5 *Assume (H1)–(H4), that* $\mathcal{X} = \mathbb{R}^d$*, that* g *is* $\mathcal{C}^2$ *on* $\mathbb{R}^d \times \mathbb{R}^d$ *and that the points of* Γ *are isolated. For any* $y \in \Gamma$*, let* $\mathcal{U}_y$ *be a neighborhood of* y*, and take two constants* $a^y > 0$ *and* $a_y > 0$ *such that* a *is* a^y*-Lipschitz on* $\mathcal{U}_y$*, and such that* $\forall x \in \mathcal{U}_y$, $\forall s \in \mathbb{R}^d$, $s^* a(x) s \geq a_y \|s\|^2 \|x - y\|$*. Define also*

$$\tilde{b}_y = \inf_{x \in \mathcal{U}_y \setminus \{y\}} \frac{x - y}{\|x - y\|} \cdot \tilde{b}(x)$$

$$\textit{and} \quad \tilde{b}^y = \sup_{x \in \mathcal{U}_y \setminus \{y\}} \frac{x - y}{\|x - y\|} \cdot \tilde{b}(x).$$

Then

(a) *If for any* $y \in \Gamma$, $\frac{\tilde{b}_y + da_y/2}{a^y} \geq 1$, *then,* $\forall x \notin \Gamma$, $\mathbf{P}_x(\tau = \infty) = 1$ *and* $\mathbf{P}_x(\lim_{t\to+\infty} X_t^\varepsilon \in \Gamma) = 0$.

(b) *If there exists* $y \in \Gamma$ *such that* $\frac{\tilde{b}^y + da^y/2}{a_y} < 1$, *then,* $\forall x \notin \Gamma$, $\mathbf{P}_x(\lim_{t\to\tau} X_t^\varepsilon = y) > 0$.

Remark 3.5 *In the case where* $\mathcal{X} \neq \mathbb{R}^d$, *the method of this proof applies only if* $x \mapsto d(x,\partial\mathcal{X})$ *is* $\mathcal{C}^2$ *on* $\{x \in \mathcal{X} : d(x,\partial\mathcal{X}) < \alpha\}$ *for some* $\alpha > 0$ *(this holds in particular when* $\partial\mathcal{X}$ *is compact and* $\mathcal{C}^2$*). This gives conditions for* $\mathbf{P}(\tau = \infty) = 1$ *involving very technical constants governing the behaviour of* a, b *and* $\tilde{b}$ *near* $\partial\mathcal{X}$.

Before proving Theorem 3.5, let us give some bounds for the constants involved in this Theorem:

Proposition 3.1 *Assume (H1), (H3), (H4), that* $\mathcal{X} = \mathbb{R}^d$, *that* g *is* $\mathcal{C}^2$ *on* $\mathbb{R}^d \times \mathbb{R}^d$ *and that the points of* Γ *are isolated. Fix* $y \in \Gamma$, *and fix* $\alpha > 0$ *such that* $B(y,\alpha) \cap \Gamma = \{y\}$. *Define*

$$C = \inf_{x\in\mathbb{R}^d,\ u,v\in\mathbb{R}^d : \|u\|=\|v\|=1} \int |h\cdot u|^2 |h\cdot v| p(x,h)dh.$$

$C > 0$ *by (H4). Let* M_3 *be a bound for the third-order moment of* $p(x,h)dh$ *on* $\mathcal{X}$, *given by (H3). Let* $D = H_{1,1}g(y,y) + H_{1,2}g(y,y)$, *and denote by* λ^y *(resp.* λ_y*) the greatest (resp. the smallest) eigenvalue of* D^*D. *Suppose that the kernel of* D *is* $\{0\}$ *(*$\lambda_y > 0$*). Then, for any* $\delta > 0$ *there exists a neighborhood* $\mathcal{U}_y$ *of* y *such that, in the statement of Theorem 3.5, we can take*

$$a^y = M_3\sqrt{\lambda^y} + \delta, \quad a_y = C\sqrt{\lambda_y} - \delta,$$

$$\tilde{b}^y < \frac{M_3}{2}\|H_{1,1}g(y,y)\| + \delta \quad \text{and} \quad \tilde{b}_y > -\frac{M_3}{2}\|H_{1,1}g(y,y)\| - \delta.$$

Remark 3.6 *There are cases where* $\tilde{b}_y \geq 0$. *This holds for example when* $D = cId$ *for some constant* $c > 0$ *and* $H_{1,1}g(y,y)$ *is a positive symmetrical matrix, or* $D = -cId$ *and* $H_{1,1}g(y,y)$ *is a negative matrix. This will appear clearly in the following proof.*

Proof of Proposition 3.1 It follows from the definition (13) of $\tilde{b}$ that for $x \neq y$,

$$\frac{x-y}{\|x-y\|}\cdot\tilde{b}(x) = \int_{\{\nabla_1 g(x,x)\cdot h>0\}} \left(\frac{x-y}{\|x-y\|}\cdot h\right)(h^* H_{1,1}g(x,x)h)p(x,h)dh, \quad (33)$$

and, because of (H1), the quantity inside the integral can be bounded by $\|h\|^3[\|H_{1,1}g(y,y)\| + O(\|x-y\|)]p(x,h)$. So

$$\frac{x-y}{\|x-y\|} \cdot \tilde{b}(x) \leq [\|H_{1,1}g(y,y)\| + O(\|x-y\|)] \int_{\{\nabla_1 g(x,x)\cdot h>0\}} \|h\|^3 p(x,h)dh$$
$$= \frac{M_3}{2}[\|H_{1,1}g(y,y)\| + O(\|x-y\|)].$$

Hence, for any $\delta > 0$, $\tilde{b}^y$ can be made smaller than $\frac{M_3}{2}\|H_{1,1}g(y,y)\| + \delta$ if we choose $\mathcal{U}_y$ sufficiently small. Similarly, $\tilde{b}_y > -\frac{M_3}{2}\|H_{1,1}g(y,y)\| - \delta$ if $\mathcal{U}_y$ is sufficiently small.

To prove Remark 3.6, is suffices to notice that, if $H_{1,1}g(y,y)$ is symmetrical positive, the quantity inside the integral (33) is positive for all h such that $(x-y)\cdot h > 0$, and that $D = c\mathrm{Id}$ implies that $\nabla_1 g(x,x) \sim c(x-y)$ when $x \to y$. So the set $\{\nabla_1 g(x,x)\cdot h > 0\} \setminus \{(x-y)\cdot h > 0\}$ converges to $\emptyset$, and we can conclude thanks to the dominated convergence theorem.

It follows from equation (16) in the proof of Proposition 2.1, that $\forall s \in \mathbb{R}^d$ and $\forall x \in \mathbb{R}^d$

$$C\|s\|^2\|\nabla_1 g(x,x)\| \leq s^* a(x) s \leq M_3\|s\|^2\|\nabla_1 g(x,x)\|.$$

Considering an orthonormal basis of $\mathbb{R}^d$ in which D^*D is diagonal, one can easily see that $\lambda_y\|v\|^2 \leq \|Dv\|^2 = v\cdot D^*Dv \leq \lambda^y\|v\|^2$ for any $v \in \mathbb{R}^d$. It remains to observe that $\nabla_1 g(x,x) \sim D(x-y)$ when $x \to y$ to obtain the required bounds for a^y and a_y. □

Proof of Theorem 3.5 Fix $y \in \Gamma$. Let us assume, for convenience, that $y = 0$. By assumption, to this point of Γ is associated a neighborhood $\mathcal{U}_0$ of 0 and four finite constants $a_0 > 0$, $a^0 > 0$, $\tilde{b}_0$ and $\tilde{b}^0$. Let ρ be small enough for $B(\rho) := \{x \in \mathbb{R}^d : \|x\| \leq \rho\} \subset \mathcal{U}_0$ and $\Gamma \cap B(2\rho) = \{0\}$, and define $\tau_\rho := \inf\{t \geq 0 : \|X^\varepsilon_t\| = \rho\}$ and $\tau_0 = \inf\{t \geq 0 : X^\varepsilon_t = 0\}$.

Theorem 3.5 follows from the following lemma, which proof is postponed after the proof of Theorem 3.5.

Lemma 3.1

(a) *If* $\frac{\tilde{b}_0 + da_0/2}{a^0} \geq 1$, *then, for all* $x \in B(\rho)\setminus\{0\}$, $\mathbf{P}_x(\tau_\rho \leq \tau_0) = 1$.

(b) *If* $\frac{\tilde{b}^0 + da^0/2}{a_0} < 1$, *then, there exists a constant* $c > 0$ *such that, for all* $x \in B(\rho/2)\setminus\{0\}$, $\mathbf{P}_x(\{\tau_0 < \tau_\rho\} \cup \{\tau_0 = \tau_\rho = \infty$ *and* $\lim_{t\to+\infty} X^\varepsilon_t = 0\}) \geq c$.

Together with the incomplete strong Markov property of Theorem 3.2 (b), part (a) of this lemma easily implies Theorem 3.5 (a), and part (b) implies

Theorem 3.5 (b) if we can prove that for any $x \in \mathcal{X} \setminus \Gamma$, $\mathbf{P}_x(\tau_{\rho/2} < \infty) > 0$. This can be proved as follows.

Fix $x \in \mathcal{X} \setminus \Gamma$. If $x \in B(\rho/2)$, there is nothing to prove, so let us assume that $x \notin B(\rho/2)$, and let $\alpha < d(x,\Gamma) \wedge (\rho/4)$. Let ϕ be a $\mathcal{C}^1$ function from an interval $[0,T]$ to Γ_α such that $\phi(0) = x$ and $\|\phi(T)\| = \alpha$. Remember the definition of the process $X^{\varepsilon,\alpha/2}$ with initial state x in the proof of Theorem 3.1. This process has bounded drift part and uniformly non-degenerate and Lipschitz diffusion part, and it is well-known that, for such a process, $\mathbf{P}^\alpha_x(\|X^{\varepsilon,\alpha/2} - \phi\|_{0,T} \leq \alpha/2) > 0$ (this can be seen as a consequence of the Girsanov's formula). Since $X^{\varepsilon,\alpha/2}$ and X^ε has the same law for $t \leq \inf\{t \geq 0 : d(X^\varepsilon_t,\partial\mathcal{X}) \leq \alpha/2\}$, this implies that $\mathbf{P}_x(\exists t \in [0,T] : X^\varepsilon_t \in B(3\alpha/2)) > 0$, which gives the required result. This ends the proof of Theorem 3.5. $\square$

Before coming to the proof of Lemma 3.1, we need to introduce a few notations and processes: a standard computation using the Itô's formula gives that, for any $t < \tau$,

$$\|X^\varepsilon_t\| = \|x\| + \int_0^t \frac{1}{\|X^\varepsilon_s\|}\Bigg[X^\varepsilon_s \cdot (b(X^\varepsilon_s) + \varepsilon\tilde{b}(X^\varepsilon_s)) \\ + \frac{\varepsilon}{2}\mathrm{Tr}(a(X^\varepsilon_s)) - \frac{\varepsilon}{2}\frac{(X^\varepsilon_s)^*}{\|X^\varepsilon_s\|}a(X^\varepsilon_s)\frac{X^\varepsilon_s}{\|X^\varepsilon_s\|}\Bigg]ds + M_t,$$

where Tr is the trace operator on $d \times d$ matrices, and where, for $t < \tau$,

$$M_t := \sqrt{\varepsilon}\int_0^t \frac{(X^\varepsilon_s)^*}{\|X^\varepsilon_s\|}\sigma(X^\varepsilon_s)dW_s.$$

Let us extend M_t to $t \geq \tau$ by setting $M_t = M_{t\wedge\tau}$ for all $t \geq 0$. Since σ is bounded, M_t is a $\mathbb{L}^2$-martingale in $\mathbb{R}$ with quadratic variation

$$\langle M\rangle_t = \varepsilon\int_0^{t\wedge\tau} \frac{(X^\varepsilon_s)^*}{\|X^\varepsilon_s\|}a(X^\varepsilon_s)\frac{X^\varepsilon_s}{\|X^\varepsilon_s\|}ds. \tag{34}$$

It follows from the Dubins-Schwartz Theorem that for any $t \geq 0$, $M_t = B_{\langle M\rangle_t}$, where B is a one-dimensional Brownian motion. Since a is Lipschitz and bounded, $\langle M\rangle$ is $\mathcal{C}^1$ on $[0,\tau)$ with bounded derivative $\varepsilon\frac{(X^\varepsilon_s)^*}{\|X^\varepsilon_s\|}a(X^\varepsilon_s)\frac{X^\varepsilon_s}{\|X^\varepsilon_s\|}$, and constant after τ.

Define the time change $T_t = \inf\{s \geq 0 : \langle M\rangle_s > t\}$ for all $t \geq 0$. If $t < \langle M\rangle_\tau := \lim_{t\to 0}\langle M\rangle_t \in [0,+\infty]$, then $T_t < \infty$ and $\langle M\rangle_{T_t} = t$. For $t < \langle M\rangle_\infty$, define $Y_t = X^\varepsilon_{T_t}$. An easy change of variable shows that for $t < \langle M\rangle_\infty = \langle M\rangle_\tau$, $Y_t \notin \Gamma$, and

$$\|Y_t\| = \|x\| + \int_0^t c(Y_s)ds + B_t,$$

where

$$c(z) = \|z\| \frac{z \cdot (b(z) + \varepsilon \tilde{b}(z)) + \varepsilon \mathrm{Tr}(a(z))/2}{\varepsilon z^* a(z) z} - \frac{1}{2\|z\|}.$$

Using the constants defined in the statement of Theorem 3.5, the fact that b is K-Lipschitz on $\mathbb{R}^d$, and the fact that $\mathrm{Tr}(a) = \sum_{i=1}^d e_i^* a e_i$, where e_i is the i^{th} vector of the canonical basis of $\mathbb{R}^d$, one easily obtains that, for $z \in \mathcal{U}_0$,

$$\forall z \in \mathbb{R}^d,\ c_1(\|z\|) < c(z) < c_2(\|z\|),$$

where, for $u > 0$,

$$\begin{aligned} c_1(u) &= \left(\frac{da_0/2 + \tilde{b}_0}{a^0} - \frac{1}{2} \right) \frac{1}{u} - \frac{2K}{\varepsilon a_0} \\ \text{and} \quad c_2(u) &= \left(\frac{da^0/2 + \tilde{b}^0}{a_0} - \frac{1}{2} \right) \frac{1}{u} + \frac{2K}{\varepsilon a_0}. \end{aligned}$$

Define also the processes Z^1 and Z^2 strong solutions in $(0,\infty)$ to the SDEs

$$Z_t^i = \|x\| + \int_0^t c_i(Z_s^i) ds + B_t$$

for $i = 1,2$, and stopped when they reach 0. As strong solutions, these processes can be constructed on the same probability space than X^ε (and Y). Finally, define for $1 \leq i \leq 2$ the stopping times

$$\begin{aligned} \theta_0^i &= \inf\{t \geq 0 : Z^i = 0\} \\ \text{and} \quad \theta_\rho^i &= \inf\{t \geq 0 : Z^i = \rho\}. \end{aligned}$$

The proof of Lemma 3.1 relies on the following three lemmas. The first one is a comparison result between Z^1, Z^2 and Y.

Lemma 3.2 *Almost surely,* $\forall t < \theta_\rho^1 \wedge \langle M \rangle_\infty$, $Z_t^1 \leq \|Y_t\|$, *and* $\forall t < \theta_\rho^2 \wedge \langle M \rangle_\infty$, $\|Y_t\| \leq Z_t^2$.

The processes Z^1 and Z^2 are Bessel processes with additional drifts. The second lemma examines whether these processes hit 0 in finite time or not.

Lemma 3.3

(a) Z^1 *is recurrent in* $(0, +\infty)$ *if and only if* $\frac{\tilde{b}_0 + da_0/2}{a^0} \geq 1$.

(b) *Let* $\tilde{\mathbf{P}}_u$ *be the law of* Z^2 *with initial state* $u > 0$. *If* $\frac{\tilde{b}^0 + da^0/2}{a_0} < 1$, *then, for any* $u < \rho$, $\tilde{\mathbf{P}}_u(\theta_0^2 < \theta_\rho^2) > 0$.

The last lemma states that, when $\langle M\rangle_\infty < \infty$, X^ε reaches Γ in finite or infinite time.

Lemma 3.4 $\{\langle M\rangle_\infty < \infty\} \subset \{\tau < \infty\} \cup \{\tau = \infty$ *and* $\lim_{t\to+\infty} X^\varepsilon_t \in \Gamma\}$ *a.s.*

Proof of Lemma 3.1 Assume first that $\frac{\tilde{b}_0 + da_0/2}{a^0} \geq 1$, and fix $x \in B(\rho) \setminus \{0\}$. Then, by Lemma 3.3 (a), $\theta^1_0 = \infty$ and $\theta^1_\rho < \infty$ a.s. Moreover, by Lemma 3.2, $\forall t < T_{\theta^1_\rho}$, $\|X^\varepsilon_t\| = \|Y_{\langle M\rangle_t}\| \geq Z^1_{\langle M\rangle_t}$.

Therefore, $\langle M\rangle_\infty = \infty$ a.s. implies that $\tau_\rho < \tau_0$, and, by Lemma 3.4, $\langle M\rangle_\infty < \infty$ a.s. implies that $\lim_{t\to\tau} X^\varepsilon_t \in \Gamma \setminus \{0\}$, and thus that $\tau_\rho < \tau_0$. This completes the proof of Lemma 3.1 (a).

Now, assume that $\frac{\tilde{b}^0 + da^0/2}{a_0} < 1$ and fix $x \in B(\rho/2)$. By Lemma 3.2, $\forall t < T_{\theta^2_\rho}$, $\|X^\varepsilon_t\| = \|Y_{\langle M\rangle_t}\| \leq Z^2_{\langle M\rangle_t}$.

Therefore, on the event $\{\theta^2_0 < \theta^2_\rho\}$, $\langle M\rangle_\infty = \infty$ a.s. implies that $\tau_0 < \tau_\rho$, and, by Lemma 3.4, $\langle M\rangle_\infty < \infty$ a.s. implies that $\lim_{t\to\tau} X^\varepsilon_t = 0$ (where τ may be finite or infinite), and thus that $\tau_0 < \tau_\rho$ or that $\tau_0 = \tau_\rho = \infty$ and $\lim_{t\to+\infty} X^\varepsilon_t = 0$. Hence,

$$\mathbf{P}_x(\{\tau_0 < \tau_\rho\} \cup \{\tau_0 = \tau_\rho = \infty \text{ and } \lim_{t\to+\infty} X^\varepsilon_t = 0\}) \geq \tilde{\mathbf{P}}_{\|x\|}(\theta^2_0 < \theta^2_\rho),$$

and, by the Markov property for Z^2, for any $x \in B(\rho/2)$, $\tilde{\mathbf{P}}_{\|x\|}(\theta^2_0 < \theta^2_\rho) \geq \tilde{\mathbf{P}}_{\rho/2}(\theta^2_0 < \theta^2_\rho)$, which is positive thanks to Lemma 3.3 (b). This completes the proof of Lemma 3.1 (b). $\square$

Proof of Lemma 3.2 Firstly, remind that Y_t is defined only for $t < \langle M\rangle_\infty$. Observe that for $t < \theta^1_0 \wedge \langle M\rangle_\infty$,

$$\|Y_t\| - Z^1_t = \int_0^t (c(Y_s) - c_1(Z^1_s))ds.$$

If there exists $t_0 < \theta^1_\rho \wedge \theta^1_0 \wedge \langle M\rangle_\infty$ such that $\|Y_{t_0}\| = Z^1_{t_0}$, then $(\|Y\| - Z^1)'(t_0) = c(Y_{t_0}) - c_1(Z^1_{t_0}) = c(Y_{t_0}) - c_1(\|Y_{t_0}\|) > 0$, and therefore, $\|Y_t\| > Z^1_t$ for $t > t_0$ in a neighborhood of t_0. Consequently, $Z^1_t \leq \|Y_t\|$ for any $t < \theta^1_\rho \wedge \theta^1_0 \wedge \langle M\rangle_\infty$. Since $Z^1_t = 0$ for $t \geq \theta^1_0$, this inequality actually holds for $t < \theta^1_\rho \wedge \langle M\rangle_\infty$. The proof of the other inequality is similar. $\square$

Proof of Lemma 3.3 The proof relies on the same functions p and v than in the proof of Theorem 3.3. They are defined by equation (26), where b^ε has to be replaced in our case by c_i, and εa by 1. For the process Z^1, if we fix $\gamma > 0$, then, for any $x > 0$,

$$p(y) = \int_\gamma^y \exp\left[-2\int_\gamma^u c_1(z)dz\right] du$$

$$= -\int_y^\gamma \exp\left[2k\int_u^\gamma \frac{dz}{z} - k'(\gamma - u)\right] du = -C\int_y^\gamma u^{-2k}e^{k'u}du,$$

where we have used the constants $k = \frac{\tilde{b}_0 + da_0/2}{a^0} - \frac{1}{2}$ and $k' = \frac{4K}{\varepsilon a_0}$. Consequently, $p(0+) = -\infty$ if and only if $2k \geq 1$, and $p(+\infty) = +\infty$, which yields (a). A similar computation for Z^2 gives that $p(0+) > -\infty$ if and only if $\frac{\tilde{b}^0 + da^0/2}{a_0} < 1$, which yields Lemma 3.3 (b). ☐

Proof of Lemma 3.4 Assume that $\mathbf{P}(\{\langle M\rangle_\infty < \infty\} \cap \{\lim_{t\to+\infty} X_t^\varepsilon \in \Gamma\}^c) > 0$. Then, there exists $\alpha > 0$ such that

$$\delta := \mathbf{P}(\langle M\rangle_\infty < \infty,\ \limsup_{t\to+\infty} d(X_t^\varepsilon,\Gamma) \geq \alpha) > 0.$$

Define for any $t > 0$ the stopping time $\tau_{\alpha,t} = \inf\{s \geq t : d(X_s^\varepsilon,\Gamma) \geq \alpha\}$. Then, for any $t > 0$,

$$\mathbf{P}(\langle M\rangle_\infty < \infty,\ \tau_{\alpha,t} < \infty) \geq \delta. \tag{35}$$

We will obtain a contradiction from this statement thanks to the following lemma:

Lemma 3.5 *Given an a.s. finite stopping time S, and $\varepsilon < 1$, for any $h \in$ (0,1)*

$$\mathbf{E}\left[\sup_{0<u<h} \|X_{S+u}^\varepsilon - X_S^\varepsilon\|^2\right] \leq 10C^2h,$$

where C is a bound for b, $\tilde{b}$ and σ on $\mathcal{X}$.

Proof of Lemma 3.5 This is a straightforward consequence of the inequality

$$\|X_{S+u}^\varepsilon - X_S^\varepsilon\|^2 \leq 2\left(\int_S^{S+u} \|b(X_s^\varepsilon) + \varepsilon\tilde{b}(X_s^\varepsilon)\|ds\right)^2 + 2\sqrt{\varepsilon}\left\|\int_S^{S+u} \sigma(X_s^\varepsilon)dW_s\right\|^2$$

and of Doob's inequality. □

Set $h = \delta\alpha^2/80C^2$. Then, it follows from Lemma 3.5 and Tchebichev's inequality that for any a.s. finite stopping time S

$$\mathbf{P}\left(\sup_{0<u<h} \|X^\varepsilon_{S+u} - X^\varepsilon_S\| > \frac{\alpha}{2}\right) \leq \frac{\delta}{2}.$$

Applying this to $\tau_{\alpha,t} \wedge T$ and letting $T \to +\infty$ yields

$$\mathbf{P}\left(\tau_{\alpha,t} < \infty,\ \sup_{0<u<h} \|X^\varepsilon_{\tau_{\alpha,t}+u} - X^\varepsilon_{\tau_{\alpha,t}}\| > \frac{\alpha}{2}\right) \leq \frac{\delta}{2}.$$

Together with inequality (35), this yields the first line of the following inequality, and the last line makes use of a constant $C > 0$ such that $s^* a(x) s \geq C\|s\|^2$ for any $s \in \mathbb{R}^d$ and $x \in \Gamma_{\alpha/2}$ (given by Proposition 2.1 (iii)) and of the formula (34) for $\langle M \rangle$.

$$\begin{aligned}\frac{\delta}{2} &\leq \mathbf{P}\left(\langle M\rangle_\infty < \infty,\ \sup_{0<u<h} \|X^\varepsilon_{\tau_{\alpha,t}+u} - X^\varepsilon_{\tau_{\alpha,t}}\| \leq \frac{\alpha}{2}\right)\\ &\leq \mathbf{P}\left(\langle M\rangle_\infty < \infty,\ \inf_{0<u<h} \|X^\varepsilon_{\tau_{\alpha,t}+u}\| \geq \frac{\alpha}{2}\right)\\ &\leq \mathbf{P}\left(\langle M\rangle_\infty < \infty,\ \langle M\rangle_{\tau_{\alpha,t}+h} - \langle M\rangle_{\tau_{\alpha,t}} \geq \varepsilon C h\right).\end{aligned}$$

So

$$\mathbf{P}\left(\langle M\rangle_\infty < \infty,\ \langle M\rangle_\infty - \langle M\rangle_t \geq \varepsilon C h\right) \geq \frac{\delta}{2}$$

holds for any $t > 0$, which is impossible. □

4 Large deviations for X^ε as $\varepsilon \to 0$

This result will be obtained using a transfer technique to carry the LDP from the family $\{\sqrt{\varepsilon}W\}_{\varepsilon>0}$, where W is a standard d-dimensional Brownian motion (Schilder's Theorem, see for example Dembo and Zeitouni [13]) to the family $\{X^\varepsilon\}_{\varepsilon>0}$, where X^ε is the solution to the SDE (1) defined in section 3.1. The method of the proof, adapted from Doss and Priouret [23] (inspired from Azencott[4]), consists in obtaining a function S that maps in some sense the paths of $\sqrt{\varepsilon}W$ on the corresponding paths of X^ε when ε is small.

The following notations will be used throughout this section.

Notations

- $\mathcal{C}(I,E)$ (resp. $\mathcal{C}_x(I,E)$, $\mathcal{C}_x^{ac}(I,E)$, $\mathcal{C}_x^1(I,E)$) is the set of continuous functions from an interval I of $\mathbb{R}_+$ to a subset E of $\mathbb{R}^d$ (resp. with value $x \in E$ at 0, resp. absolutely continuous with value x at 0, resp. $\mathcal{C}^1$ with value x at 0), endowed with the uniform norm.
- For $\varphi \in \mathcal{C}([0,T],\mathbb{R}^d)$ and $0 \leq a < b \leq T$, define

$$\|\varphi\|_{a,b} = \sup_{a \leq t \leq b} \|\varphi(t)\|, \tag{36}$$

  and

$$B_b(\varphi,\delta) = \{\tilde{\varphi} \in \mathcal{C}([0,T],\mathbb{R}^d) : \|\tilde{\varphi} - \varphi\|_{0,b} \leq \delta\}. \tag{37}$$

  When $a = 0$ and $b = T$, $\|\cdot\|_{0,T}$ is the usual norm of uniform convergence in $\mathcal{C}([0,T],\mathbb{R}^d)$, and when $b = T$, $B_T(\varphi,\delta)$ is the usual closed ball centered at φ with radius δ in $\mathcal{C}([0,T],\mathbb{R}^d)$ for the uniform norm.

4.1 Statement of the result

Let us recall the definition of the (good) rate function of the large deviation principle for the Brownian motion (Shilder's Theorem):

$$J_T(\varphi) = \begin{cases} \dfrac{1}{2}\displaystyle\int_0^T \|\dot{\varphi}(t)\|^2 dt & \text{if} \quad \varphi \in \mathcal{C}_0^{ac}([0,T],\mathbb{R}^d) \\ +\infty & \text{otherwise,} \end{cases}$$

where $\dot{\varphi}$ denotes the derivative of φ.

Let us now give the expression of the rate function involved in our result. Fix $T > 0$ and $x \in \overline{\mathcal{X}}$, and define

$$\begin{aligned} \forall \psi \in \mathcal{C}\left([0,T],\overline{\mathcal{X}}\right), \quad & t_\psi = \inf\{t \in [0,T] : \psi(t) \in \Gamma\} \wedge T \\ \text{and} \quad \tilde{\mathcal{C}}_x^{ac}([0,T],\overline{\mathcal{X}}) &= \{\psi \in \mathcal{C}_x^{ac}([0,T],\overline{\mathcal{X}}) \text{ constant on } [t_\psi,T]\}. \end{aligned} \tag{38}$$

Then, we can define for $\psi \in \mathcal{C}([0,T],\overline{\mathcal{X}})$

$$I_{T,x}(\psi) = \begin{cases} \dfrac{1}{2}\displaystyle\int_0^{t_\psi} [\dot{\psi}(t) - b(\psi(t))]^* a^{-1}(\psi(t))[\dot{\psi}(t) - b(\psi(t))]dt & \\ & \text{if} \quad \psi \in \tilde{\mathcal{C}}_x^{ac}([0,T],\overline{\mathcal{X}}) \\ +\infty & \text{otherwise.} \end{cases} \tag{39}$$

The inverse matrix $a^{-1}(x)$ of $a(x)$ is, by Proposition 2.1 (ii), defined for $x \notin \Gamma$, so the quantity inside the integral is well-defined. Moreover, since $a(x)$ is a non-negative symmetrical matrix, this quantity is positive, so the integral is well-defined and $I_{T,x}(\psi)$ belongs to $\mathbb{R}_+ \cup \{+\infty\}$. When $t_\psi = T$, $I_{T,x}(\psi)$ takes the classical form of rate functions for diffusion processes. Note also that $\tilde{b}$ does not appear in these expressions. This comes from the fact that b^ε uniformly converges to b when $\varepsilon \to 0$.

This original form of rate function will appear naturally in the proof and follows from the bad regularity properties of the coefficients. Moreover, we will see that this function is *not* lower semicontinuous (see Proposition 4.1).

Theorem 4.1 *Assume (H1)–(H4). Suppose also that the points* $x \in \mathcal{X}$ *such that* $\nabla_1 g(x,x) = 0$ *are isolated points of* $\mathbb{R}^d$*. For any* $y \in \overline{\mathcal{X}}$ *and* $\varepsilon > 0$*, let* $X^{\varepsilon,y}$ *be the solution to (1) with initial state* y *constructed in section 3.2, and let* $\mathbf{P}^\varepsilon_y$ *denote its law. Fix* $T > 0$ *and* $x \in \overline{\mathcal{X}}$*. Then, for any open subset* O *of* $\mathcal{C}\left([0,T],\overline{\mathcal{X}}\right)$*, and for any closed subset* C *of* $\mathcal{C}\left([0,T],\overline{\mathcal{X}}\right)$ *such that* $\mathcal{C}^1_x([0,T],\mathcal{X} \setminus \Gamma)$ *is dense in* $C \cap \mathcal{C}_x([0,T],\overline{\mathcal{X}})$*,*

$$\liminf_{\varepsilon\to 0, y\to x} \varepsilon \ln \mathbf{P}^\varepsilon_y(O) = \liminf_{\varepsilon\to 0, y\to x} \varepsilon \ln \mathbf{P}(X^{\varepsilon,y} \in O) \geq - \inf_{\psi\in O} I_{T,x}(\psi) \tag{40}$$

$$\limsup_{\varepsilon\to 0, y\to x} \varepsilon \ln \mathbf{P}^\varepsilon_y(C) = \limsup_{\varepsilon\to 0, y\to x} \varepsilon \ln \mathbf{P}(X^{\varepsilon,y} \in C) \leq - \inf_{\psi\in C} I_{T,x}(\psi). \tag{41}$$

Remark 4.1 *If we take* $y = x$ *in (40) and (41), we recover the classical form of large deviations bounds. The more general form of Theorem 4.1 is necessary to handle the problem of exit from a domain of section 5.*

Remark 4.2 *Since we cannot prove that the upper bound holds for any closed set* C*, Theorem 4.1 is actually an incomplete large deviation principle. This comes from the degeneracy of* a *at points of* Γ *and the fact that* σ *is not Lipschitz near* Γ*. Note also that the fact that the process* X^ε *stays constant when it reaches* Γ *plays an important role in the proof. However, the condition on* C *that we obtain covers* all *the closed sets of interest in usual applications of large deviations (in particular the closed sets involved in the problem of exit from a domain, see section 5).*

Before proving this theorem, let us prove two corollaries. The first one will be useful in section 5:

Corollary 4.1 *Assume the conditions of Theorem 4.1. Then, for any compact set* $K \subset \overline{\mathcal{X}}$*, for any open set* $O \subset \mathcal{C}([0,T],\overline{\mathcal{X}})$*, and for any closed set* $C \subset \mathcal{C}([0,T],\overline{\mathcal{X}})$ *such that* $\mathcal{C}^1_y([0,T],\mathcal{X} \setminus \Gamma)$ *is dense in* $C \cap \mathcal{C}_y([0,T],\overline{\mathcal{X}})$ *for any* $y \in K$*,*

$$\liminf_{\varepsilon\to 0} \varepsilon \ln \inf_{y\in K} \mathbf{P}(X^{\varepsilon,y} \in O) \geq - \sup_{y\in K} \inf_{\psi\in O} I_{T,y}(\psi),$$

$$\text{and} \quad \limsup_{\varepsilon\to 0} \varepsilon \ln \sup_{y\in K} \mathbf{P}(X^{\varepsilon,y} \in C) \leq - \inf_{y\in K,\ \psi\in C} I_{T,y}(\psi).$$

This result can be easily deduced from Theorem 4.1 (for details, see the proof of Corollary 5.6.15 in Dembo and Zeitouni [13]).

The second corollary of Theorem 4.1 states that X^ε converges as $\varepsilon \to 0$ to a deterministic function, solution to a differential equation called in the biological litterature *canonical equation of adaptive dynamics* (see [18] and [59]). A similar result for jump processes has been obtained in chapter III.

Corollary 4.2 *Assume the conditions of Theorem 4.1. Then, for any* $T > 0$ *and* $x \in \overline{\mathcal{X}}$, $X^{\varepsilon,x}$ *converges in probability as* $\varepsilon \to 0$ *to the solution* ϕ *of* $\dot{\phi} = b(\phi)$ *with initial state* x *on* $[0,T]$.

Proof of Corollary 4.2 This follows immediately from (41) applied to the set $C_\eta = \{\psi : \|\psi - \phi\|_{0,T} \geq \eta\}$ for $\eta > 0$ (where $\|\cdot\|_{a,b}$ has been defined in (36)), if we can prove that $\inf_{\psi\in C_\eta} I_{T,x}(\psi) > 0$ for each $\eta > 0$. This can be proved as below:

For any $\psi \in C_\eta$ such that $I_{0,T}(\psi) < \infty$, there exists $t \in [0,T]$ such that $\|\psi_t - \phi_t\| \geq \eta$. Note that ϕ (resp. ψ) is constant after the time t_ϕ (resp. t_ψ) where it reaches Γ, and that $b(x) = 0$ for $x \in \Gamma$. Note also that, by Proposition 2.1 (i), there exists $K > 0$ such that b is K-Lipschitz, and a is bounded by K on $\overline{\mathcal{X}}$. The Cauchy-Schwartz inequality yields

$$\begin{aligned}
\|\psi_t - \phi_t\|^2 &\leq 2\left\|\int_0^{t\wedge t_\psi}(\dot{\psi}_s - b(\psi_s))ds\right\|^2 + 2\left\|\int_0^t (b(\psi_s) - b(\phi_s))ds\right\|^2 \\
&\leq 2T\int_0^{t_\psi}\|\dot{\psi}_s - b(\psi_s)\|^2 ds + 2T\int_0^t \|b(\phi_s) - b(\psi_s)\|^2 ds
\end{aligned}$$

$$\begin{aligned}
&\leq \frac{2T}{K}\int_0^{t_\psi}(\dot{\psi}_s - b(\psi_s))^* a^{-1}(\psi_s)(\dot{\psi}_s - b(\psi_s))ds \\
&\qquad + 2TK^2\int_0^t \|\phi_s - \psi_s\|^2 ds \\
&\leq \frac{2T}{K} I_{T,x}(\psi) + 2TK^2\int_0^t \|\phi_s - \psi_s\|^2 ds,
\end{aligned}$$

and, by Gronwall's Lemma, $I_{T,x}(\psi) \geq K\eta^2 e^{-2T^2K^2}/2T > 0$. □

Finally, let us make observations about the lower semicontinuity of $I_{T,x}$.

Define, when $d \geq 2$,

$$\tilde{I}_{T,x}(\psi) = \begin{cases} \displaystyle\frac{1}{2}\int_0^T \mathbf{1}_{\{\psi(t)\notin\Gamma\}}[\dot{\psi}(t) - b(\psi(t))]^* a^{-1}(\psi(t))[\dot{\psi}(t) - b(\psi(t))]dt \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{if} \quad \psi \in \mathcal{C}_x^{ac}([0,T],\overline{\mathcal{X}}) \\ +\infty \qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad \text{otherwise,} \end{cases} \tag{42}$$

and when $d = 1$, define $\tilde{I}_{T,x}$ by the same formula, except that the condition $\psi \in \mathcal{C}_x^{ac}([0,T],\overline{\mathcal{X}})$ is replaced by $\psi \in \mathcal{C}_x^{ac}([0,T],\overline{C_x})$, where C_x is the connected component of $\mathcal{X} \setminus \Gamma$ containing x.

Define also for $u \geq 0$

$$\Phi(u) = \{\varphi \in \mathcal{C}_0([0,T],\mathbb{R}^d) : J_T(\varphi) \leq u\} \tag{43}$$

$$\Psi(u) = \{\psi \in \mathcal{C}_x([0,T],\overline{\mathcal{X}}) : I_{T,x}(\psi) \leq u\} \tag{44}$$

$$\tilde{\Psi}(u) = \{\psi \in \mathcal{C}_x([0,T],\overline{\mathcal{X}}) : \tilde{I}_{T,x}(\psi) \leq u\}. \tag{45}$$

By Schilder's Theorem, $\Phi(u)$ is compact for any $u \geq 0$.

Proposition 4.1 *Assume the conditions of Theorem 4.1. Then, for any* $u \geq 0$, $\tilde{\Psi}(u)$ *is closed (*i.e. $\tilde{I}_{T,x}$ *is lower semicontinuous), and* $\overline{\Psi(u)} \subset \tilde{\Psi}(u)$. *Assume additionally that there exists an isolated point* y *of* Γ *such that* g *is* $\mathcal{C}^2$ *at* y, *and that* $H_{1,1}g(y,y) + H_{1,2}g(y,y)$ *has a null kernel. Then, for any* $x \notin \Gamma$, *there exists* $u > 0$ *such that* $\Psi(u) \subsetneq \overline{\Psi(u)}$, *and so* $I_{T,x}$ *is* not *lower semicontinuous.*

Remark 4.3 *It is always possible to obtain a large deviation principle with a* rate *function* $\bar{I}_{T,x}$ *from our large deviation principle with the non-lower semincontinuous "rate" function* $I_{T,x}$ *by putting* $\bar{I}_{T,x}(\psi) = \liminf_{\hat{\psi}\to\psi} I_{T,x}(\hat{\psi})$. *Indeed,* $\bar{I}_{T,x}$ *is lower semicontinuous,* $\bar{I}_{T,x} \leq I_{T,x}$, *and for any open subset* O *of* $\mathcal{C}_x([0,T],\overline{\mathcal{X}})$, $\inf_{\psi\in O} \bar{I}_{T,x}(\psi) = \inf_{\psi\in O} I_{T,x}(\psi)$. *So (40) and (41) hold with* $\bar{I}_{T,x}$ *instead of* $I_{T,x}$.

Proposition 4.1 shows that $\tilde{I}_{T,x}$ *is a good candidate for* $\bar{I}_{T,x}$, *since* $\tilde{I}_{T,x} \leq I_{T,x}$ *and* $\tilde{I}_{T,x}$ *is lower semicontinuous. Unfortunately, we are not able to prove that, for any open set* O *in* $\mathcal{C}_x([0,T],\overline{\mathcal{X}})$, $\inf_{\psi\in O} \tilde{I}_{T,x}(\psi) = \inf_{\psi\in O} I_{T,x}(\psi)$, *or, equivalently, that* $\overline{\Psi(u)} = \tilde{\Psi}(u)$. *However,* $\tilde{I}_{T,x} \leq \bar{I}_{T,x} \leq I_{T,x}$, *so* $\bar{I}_{T,x}$ *can differ from* $I_{T,x}$ *only on* $\mathcal{C}_x^{ac} \setminus \tilde{\mathcal{C}}_x^{ac}$.

Proof of Proposition 4.1 Proposition 4.1 relies on the following lemma, which is Proposition 3.1 of Doss and Priouret [23]:

Lemma 4.1 *Let $\hat{\sigma}$ be a bounded and Lipschitz function from $\mathbb{R}^d$ to $\mathcal{S}_c$ for some $c > 0$, and let $\hat{b}$ be a bounded and Lipschitz function from $\mathbb{R}^d$ to $\mathbb{R}^d$. Define $\hat{I}_T$ on $\mathcal{C}^{ac}([0,T],\mathbb{R}^d)$ as follows:*

$$\hat{I}_T(\psi) = \frac{1}{2}\int_0^T [\dot{\psi}(t) - \hat{b}(\psi(t))]^* \hat{a}^{-1}(\psi(t))[\dot{\psi}(t) - \hat{b}(\psi(t))]dt,$$

where $\hat{a} = \hat{\sigma}\hat{\sigma}^*$. Then $\hat{I}_T$ is lower semicontinuous on $\mathcal{C}^{ac}([0,T],\mathbb{R}^d)$ for the norm of uniform convergence. Moreover, for any compact set K and positive u, $\{\psi \in \mathcal{C}([0,T],\mathbb{R}^d) : \hat{I}_T(\psi) \leq u, \psi(0) \in K\}$ is compact.

Let (ψ_n) be a sequence of functions of $\tilde{\Psi}(u)$ uniformly converging to a function $\psi \in \mathcal{C}_x([0,T],\overline{\mathcal{X}})$. Note that, in the case when $d = 1$, since $\tilde{I}_{T,x}(\psi_n) \leq u$ implies that $\psi_n \in \mathcal{C}_x^{ac}([0,T],\overline{C_x})$, we actually have $\psi \in \mathcal{C}_x([0,T],\overline{C_x})$.

For any $\delta > 0$, define $K_\delta = \{t \in [0,T] : d(t,\psi^{-1}(\Gamma)) \geq \delta\}$. K_δ is a compact set, made of a finite union of intervals (since between each interval, there is at least a distance of 2δ). By compactness, there exists $\alpha > 0$ such that, for all $t \in K_\delta$, $d(\psi(t),\Gamma) \geq \alpha$. Consequently, for n sufficiently large, $d(\psi_n(t),\Gamma) \geq \alpha/2$ for all $t \in K_\delta$.

Define $\hat{a} = a + \chi\mathrm{Id}$ with χ Lipschitz, $\chi \equiv 0$ on $\Gamma_{\alpha/2}$, and $\chi \equiv 1$ on $\Gamma_{\alpha/4}$. Then, by Proposition 2.1 (i) and (iii), $\hat{a}$ is Lipschitz and uniformly nondegenerate and b is Lipschitz, and, therefore, $\hat{\sigma} = \zeta(\hat{a})$ and $\hat{b} = b$ satisfy the assumptions of Lemma 4.1.

Let $[s,t]$ be an interval included in K_δ. Since $\hat{\sigma} = \sigma$ on $\Gamma_{\alpha/2}$, if we replace $[0,T]$ by $[s,t]$ in the statement of Lemma 4.1, we obtain that ψ is a.c. on $[s,t]$, and that

$$\begin{aligned}&\frac{1}{2}\int_s^t [\dot{\psi}(v) - b(\psi(v))]^* a^{-1}(\psi(v))[\dot{\psi}(v) - b(\psi(v))]dv \\ &\qquad \leq \liminf \frac{1}{2}\int_s^t [\dot{\psi}_n(v) - b(\psi_n(v))]^* a^{-1}(\psi_n(v))[\dot{\psi}_n(v) - b(\psi_n(v))]dv.\end{aligned}$$

Consequently, since K_δ is a finite union of such intervals,

$$\begin{aligned}&\frac{1}{2}\int_{K_\delta} [\dot{\psi}(v) - b(\psi(v))]^* a^{-1}(\psi(v))[\dot{\psi}(v) - b(\psi(v))]dv \\ &\qquad \leq \liminf \frac{1}{2}\int_{K_\delta} [\dot{\psi}_n(v) - b(\psi_n(v))]^* a^{-1}(\psi_n(v))[\dot{\psi}_n(v) - b(\psi_n(v))]dv \leq u.\end{aligned}$$

Finally, since K_δ converges to $\psi^{-1}(\Gamma)$ when $\delta \to 0$, it follows from the monotone convergence Theorem that

$$\tilde{I}_{T,x}(\psi) = \frac{1}{2}\int_{\psi(t)\notin\Gamma} [\dot{\psi}(v) - b(\psi(v))]^* a^{-1}(\psi(v))[\dot{\psi}(v) - b(\psi(v))]dv \leq u,$$

and, therefore, $\tilde{\Psi}(u)$ is closed.

Since $\tilde{I}_{T,x} \leq I_{T,x}$, this implies immediately that $\overline{\Psi(u)} \subset \tilde{\Psi}(u)$.

Now, take y as in the statement of Proposition 4.1. By translation, we can suppose that $y = 0$. Then, Proposition 3.1 implies that there exists a neighborhood $\mathcal{N}_0$ of 0 and a constant $a_0 > 0$ such that for all $s \in \mathbb{R}^d$ and $x \in \mathcal{N}_0$, $s^* a(x) s \geq a_0 \|x\| \|s\|^2$, *i.e.* each eigenvalue of $a(x)$ is greater than $a_0\|x\|$. Therefore, for all $s \in \mathbb{R}^d$ and $x \in \mathcal{N}_0$,

$$s^* a^{-1}(x) s \leq \frac{\|s\|^2}{a_0 \|x\|}. \tag{46}$$

Firstly, take $x_0 \in \mathcal{X} \setminus \Gamma$ such that the segment $(0,x_0]$ is included in $\mathcal{X} \setminus \Gamma$ and in $\mathcal{N}_0$, and define for $0 \leq t \leq T$

$$\psi(t) = \left(1 - \frac{2t}{T}\right)^2 x_0,$$

and for all $n \geq 1$

$$\psi_n(t) = \begin{cases} \psi(t) & \text{if} \quad t \in \left[0,\frac{T}{2} - \frac{1}{n}\right] \cup \left[\frac{T}{2} + \frac{1}{n},T\right] \\ \psi\left(\frac{T}{2} - \frac{1}{n}\right) & \text{otherwise.} \end{cases}$$

Since $\psi(T/2 - 1/n) = \psi(T/2 + 1/n)$, ψ_n is continuous and piecewise differentiable. Note that the values of ψ and ψ_n belong to the segment $[0,x_0]$, that $\psi(t) \notin \Gamma$ except if $t = T/2$, and that $\psi_n(t) \notin \Gamma$ for any $t \in [0,T]$. Therefore, $I_{T,x_0}(\psi) = \infty$, and $I_{T,x_0}(\psi_n) < \infty$.

It follows from (46) and from the fact that b is K-Lipschitz that

$$\begin{aligned} \tilde{I}_{T,x_0}(\psi) &\leq \frac{1}{2a_0} \int_0^T \frac{\|(1 - 2t/T)2x_0/T + b(\psi(t))\|^2}{\|\psi(t)\|} dt \\ &\leq \frac{1}{2a_0} \int_0^T \frac{2(1 - 2t/T)^2 4\|x_0\|^2/T^2 + 2K^2\|\psi(t)\|^2}{\|\psi(t)\|} dt \\ &\leq \frac{1}{2a_0} \int_0^T \left(\frac{8}{T^2}\|x_0\| + 2K^2\|\psi(t)\|\right) dt < \infty. \end{aligned} \tag{47}$$

Therefore, for all $n \geq 1$,

$$\begin{aligned} I_{T,x_0}(\psi_n) &\leq \tilde{I}_{T,x_0}(\psi) + \frac{1}{2a_0} \int_{T/2-1/n}^{T/2+1/n} \frac{\|b(\psi_n(t))\|^2}{\|\psi_n(t)\|} dt \\ &\leq \tilde{I}_{T,x_0}(\psi) + \frac{1}{2a_0} \int_{T/2-1/n}^{T/2+1/n} K^2 \|\psi_n(t)\| dt, \end{aligned}$$

which is bounded by some constant u. Hence $\limsup I_{T,x_0}(\psi_n) \leq u < \infty$, and so $\Psi(u) \subsetneq \overline{\Psi}(u)$.

Now, let us extend this result to an arbitrary $x \notin \Gamma$. Since the points of Γ are isolated in $\mathcal{X}$, there exists $\alpha > 0$ and $\phi \in \mathcal{C}^1([0,T],\Gamma_\alpha)$ such that $\phi(0) = x$ and $\phi(T) = x_0$. Since, by Proposition 2.1 (iii), a is uniformly non-degenerate on Γ_α, $I_{T,x}(\phi) < \infty$. Then, it suffices to concatenate ϕ and ψ to obtain a function $\tilde{\psi}$ defined on $[0,2T]$ such that $\tilde{I}_{2T,x}(\tilde{\psi}) < \infty$ and $I_{2T,x}(\tilde{\psi}) = \infty$, and the same procedure as above ends the proof of Proposition 4.1. □

4.2 Proof of Theorem 4.1

Let us first give some notations. For any $\varepsilon > 0$ and $y \in \overline{\mathcal{X}}$, let $X^{\varepsilon,y}$ be a weak solution to (1) with initial state y, defined on some filtered probability space $(\Omega^{\varepsilon,y},\mathcal{F}_t^{\varepsilon,y},W^{\varepsilon,y},\mathbf{P}^{\varepsilon,y})$, where $W^{\varepsilon,y}$ is a standard d-dimensional $\mathbf{P}^{\varepsilon,y}$-Brownian motion.

Let us define the function S that transfers the LDP for the Brownian motion to the processes X^ε: for any $\varphi \in \mathcal{C}_0^{ac}([0,T],\mathbb{R}^d)$, let $S(\varphi)$ be the solution on $[0,T]$ to

$$S(\varphi)_t = x + \int_0^t b(S(\varphi)_s)ds + \int_0^t \sigma(S(\varphi)_s)\dot{\varphi}_s ds, \tag{48}$$

obtained in the following way: by Proposition 2.1 (i) and (iii), b and σ are bounded and locally Lipschitz on $\mathcal{X}\setminus\Gamma$, so, by the Cauchy-Lipschitz Theorem, there is local existence and uniqueness in $\mathcal{X} \setminus \Gamma$ of an absolutely continuous function solution to $\dot{y} = b(y) + \sigma(y)\dot{\varphi}$. This defines properly $S(\varphi)$ until the time $t_{S(\varphi)}$ where it reaches Γ. In the case where $t_{S(\varphi)} < T$, set $S(\varphi)_t = S(\varphi)_{t_{S(\varphi)}}$ for $t_{S(\varphi)} \leq t \leq T$. Since for any $x \in \Gamma$, $b(x) = 0$ and $\sigma(x) = 0$, this function $S(\varphi)$ is actually a solution to (48) on $[0,T]$. Hence, we have defined properly the function S from $\mathcal{C}_0^{ac}([0,T],\mathbb{R}^d)$ to $\mathcal{C}_x^{ac}\left([0,T],\overline{\mathcal{X}}\right)$.

The proof of Theorem 4.1 is based on the following three lemmas. Their proof is postponed after the proof of the theorem.

The first lemma precises in which sense the function S maps the paths of $\sqrt{\varepsilon}W$ to the paths of $X^{\varepsilon,y}$ when ε is small and y is close to x.

Lemma 4.2

(i) *Fix $\varphi \in \mathcal{C}_0^{ac}([0,T],\mathbb{R}^d)$ such that $\psi := S(\varphi)$ takes no value in Γ and such that $J_T(\varphi) < +\infty$. Then, $\forall\eta > 0$, $\forall R > 0$, $\exists\delta > 0$ such that*

$$\limsup_{\varepsilon\to 0, y\to x} \varepsilon \ln \mathbf{P}_y^\varepsilon \left(\|X^{\varepsilon,y} - S(\varphi)\|_{0,T} \geq \eta, \|\sqrt{\varepsilon}W^{\varepsilon,y} - \varphi\|_{0,T} \leq \delta\right) \leq -R. \tag{49}$$

(ii) *With the same φ as in (i), $\forall\delta > 0$, $\forall R > 0$, $\exists\eta > 0$ such that*

$$\limsup_{\varepsilon\to 0, y\to x} \varepsilon \ln \mathbf{P}_y^\varepsilon \left(\|X^{\varepsilon,y} - S(\varphi)\|_{0,T} \leq \eta, \|\sqrt{\varepsilon}W^{\varepsilon,y} - \varphi\|_{0,T} \geq \delta \right) \leq -R. \quad (50)$$

(iii) *Fix* $\tilde{\varphi} \in \mathcal{C}_0^{ac}([0,T],\mathbb{R}^d)$ *such that* $\psi_t := S(\tilde{\varphi})_t \in \Gamma$ *for some* $t \in [0,T]$. *Define* $\varphi_t = \tilde{\varphi}_t$ *for* $t < t_\psi$ *and* $\varphi_t = \tilde{\varphi}_{t_\psi}$ *for* $t_\psi \leq t \leq T$. *Then* $S(\varphi) = S(\tilde{\varphi}) = \psi$. *Suppose that* $J_T(\varphi) < +\infty$. *Then,* $\forall \eta > 0$, $\forall R > 0$, $\exists \delta > 0$ *such that (49) holds.*

Let us briefly comment this lemma. In Doss and Priouret [23], $b^\varepsilon = b + \varepsilon\tilde{b}$ and σ are both supposed Lipschitz on $\mathcal{X}^n$, which is, by Proposition 2.1, only true for b in our case. Moreover, the process X^ε stays constant after the time where it reaches Γ. Because of these difficulties, the method of [23] has to be adapted in order to obtain (i). A more careful study is necessary to obtain (iii), and this only gives the lower bound (40). In order to establish (41), we have firstly to adapt the method of the proof of (i) to prove (ii), and then to obtain the upper bound thanks to a different estimate.

The second lemma specifies the usual relation between S, $I_{T,x}$ and J_T for transfers of large deviation principles.

Lemma 4.3

(i) *For all* $\psi \in \mathcal{C}_x\left([0,T],\overline{\mathcal{X}}\right)$, $I_{T,x}(\psi) = \inf\{J_T(\varphi), S(\varphi) = \psi\}$, *and when* $I_{T,x}(\psi) < +\infty$, *there is unique* $\varphi \in \mathcal{C}_0^{ac}([0,T],\mathbb{R}^d)$ *that realizes this infimum, and this function is constant after* t_ψ.

(ii) $\mathcal{C}^1([0,T],\mathcal{X} \setminus \Gamma)$ *is dense in* $S(\{J_T < \infty\})$.

The last lemma gives a uniform exponential tightness estimate.

Lemma 4.4 *Define for any* $k > 0$ *and* $y \in \overline{\mathcal{X}}$ *the compact set*

$$K_k^y = \left\{ \psi \in \mathcal{C}_y([0,T],\overline{\mathcal{X}}) : \forall l \geq k,\ \omega\left(\psi, \frac{1}{l^3}\right) \leq \frac{1}{l} \right\}, \quad (51)$$

where $\omega(\psi,\delta) = \sup_{|t-s|\leq\delta} \|\psi(t) - \psi(s)\|$. *Then, there exists* k_0 *and* ε_0, *such that for all* $y \in \overline{\mathcal{X}}$, $k \geq k_0$ *and* $\varepsilon \leq \varepsilon_0$,

$$\varepsilon \ln \mathbf{P}(X^{\varepsilon,y} \notin K_k^y) \leq -\frac{k}{64d\Sigma^2}, \quad (52)$$

where Σ *is a bound for* σ *on* $\overline{\mathcal{X}}$.

All the preliminary steps required for the proof of Theorem 4.1 have now been completed.

Proof of Theorem 4.1 (40) The lower bound (40) for any open set O is classically equivalent to the fact that $\forall \psi \in \mathcal{C}_x([0,T],\overline{\mathcal{X}})$ and $\forall \eta > 0$,

$$\liminf_{\varepsilon \to 0, y \to x} \varepsilon \ln \mathbf{P}^{\varepsilon,y}(\|X^{\varepsilon,y} - \psi\|_{0,T} \leq \eta) \geq -I_{T,x}(\psi). \tag{53}$$

Fix ψ and η as above, and assume that $I_{T,x}(\psi) < +\infty$ (otherwise, there is nothing to prove). By Lemma 4.3 (i), there is a unique $\varphi \in \mathcal{C}_0^{ac}([0,T],\mathbb{R}^d)$ such that $S(\varphi) = \psi$ and $u := J_T(\varphi) = I_{T,x}(\psi)$. Choose $R > u$. If the image of ψ has empty intersection with Γ, apply Lemma 4.2 (i). Otherwise, apply Lemma 4.2 (iii). In both cases, there exists $\delta > 0$ such that

$$\limsup_{\varepsilon \to 0, y \to x} \varepsilon \ln \mathbf{P}^{\varepsilon,y}\left(\|X^{\varepsilon,y} - \psi\|_{0,T} \geq \eta, \|\sqrt{\varepsilon} W^{\varepsilon,y} - \varphi\|_{0,T} \leq \delta\right) \leq -R.$$

Write

$$\begin{aligned} \mathbf{P}^{\varepsilon,y}(\|\sqrt{\varepsilon}W^{\varepsilon,y} - \varphi\|_{0,T} \leq \delta) &\leq \mathbf{P}^{\varepsilon,y}(\|X^{\varepsilon,y} - \psi\|_{0,T} < \eta) \\ &+ \mathbf{P}^{\varepsilon,y}(\|X^{\varepsilon,y} - \psi\|_{0,T} \geq \eta, \|\sqrt{\varepsilon}W^{\varepsilon,y} - \varphi\|_{0,T} \leq \delta) \end{aligned}$$

and observe that $\mathbf{P}^{\varepsilon,y}(\|\sqrt{\varepsilon}W^{\varepsilon,y} - \varphi\|_{0,T} \leq \delta)$ is independent of ε and y. Take the lim inf of ε times the log of both sides of this inequality: using Schilder's Theorem on the second line, we obtain

$$\begin{aligned} -u = -J_T(\varphi) &\leq -\inf\{J_T(\tilde{\varphi}), \tilde{\varphi} \in B_T(\varphi,\delta)\} \\ &\leq \liminf_{\varepsilon \to 0, y \to x} \varepsilon \ln \mathbf{P}^{\varepsilon,y}(\|\sqrt{\varepsilon}W^{\varepsilon,y} - \varphi\|_{0,T} < \delta) \\ &\leq \sup\Bigg\{\liminf_{\varepsilon \to 0, y \to x} \varepsilon \ln \mathbf{P}^{\varepsilon,y}(\|X^{\varepsilon,y} - \psi\|_{0,T} < \eta), \\ &\quad \liminf_{\varepsilon \to 0, y \to x} \varepsilon \ln \mathbf{P}^{\varepsilon,y}(\|X^{\varepsilon,y} - \psi\|_{0,T} \geq \eta, \|\sqrt{\varepsilon}W^{\varepsilon,y} - \varphi\|_{0,T} \leq \delta)\Bigg\} \\ &\leq \sup\left\{\liminf_{\varepsilon \to 0, y \to x} \varepsilon \ln \mathbf{P}^{\varepsilon,y}(\|X^{\varepsilon,y} - \psi\|_{0,T} < \eta), -R\right\}, \end{aligned}$$

and since $R > u$, (53) is established.

Proof of Theorem 4.1 (41) Let us first study the case where $x \in \Gamma$. In this case, $\forall \varepsilon > 0$ and $\forall t \geq 0$, $X_t^\varepsilon = x$, and $I_{T,x}(\psi) = +\infty$ as soon as $\psi \not\equiv x$. Inequality (41) is trivial when $x \in C$, so assume that $x \notin C$. Then, there exists $\eta > 0$ such that $B_T(x,\eta) \cap C = \emptyset$, where $B_T(\varphi,\eta)$ has been defined in (37). In particular, if $\|y - x\| < \eta$, there is no function in C with initial value y, so $\mathbf{P}(X^{\varepsilon,y} \in C) = 0$. This yields (41) when $x \in \Gamma$.

Now, fix $x \notin \Gamma$. Let us first establish (41) for particular compact sets.

Let K be a non-empty compact set of $\mathcal{C}\left([0,T],\overline{\mathcal{X}}\right)$ such that $S(\{J_T < +\infty\})$ is dense in K_x, where $K_x := K \cap \mathcal{C}_x([0,T],\overline{\mathcal{X}})$. By Lemma 4.3 (i), $S(\{J_T < +\infty\}) = \{I_{T,x} < +\infty\}$, and so $u := \inf\{I_{T,x}(\psi),\psi \in K\} < +\infty$.

Fix $\rho > 0$. For any $\psi \in K \cap S(\{J_T < +\infty\})$, by Lemma 4.3 (i), there exists a unique $\varphi \in \mathcal{C}_0^{ac}([0,T],\mathbb{R}^d)$ such that $S(\varphi) = \psi$ and $I_{T,x}(\psi) = J_T(\varphi) < \infty$. We intend to use Lemma 4.2 (ii), which holds only if ψ takes no value in Γ. So we have to introduce $\alpha_\psi > 0$ such that

$$\frac{1}{2}\int_{t_\psi-\alpha_\psi}^{t_\psi} \|\dot{\varphi}_s\|^2 ds < \frac{\rho}{2},$$

so that $J_T(\varphi) \le J_{t_\psi-\alpha_\psi}(\varphi) + \rho/2$ (by Lemma 4.3 (i), $\dot{\varphi}_t = 0$ for $t > t_\psi$), and, since $J_{t_\psi-\alpha_\psi}$ is lower semicontinuous, there exists $\delta_\psi > 0$ such that

$$\forall \tilde{\varphi} \in B_{t_\psi-\alpha_\psi}(\varphi,\delta_\psi), \quad J_{t_\psi-\alpha_\psi}(\tilde{\varphi}) \ge J_{t_\psi-\alpha_\psi}(\varphi) - \frac{\rho}{2} \ge J_T(\varphi) - \rho, \tag{54}$$

where $B_t(\varphi,\delta)$ has been defined in (37).

Since $\psi_t \notin \Gamma$ for any $t \in [0,t_\psi - \alpha_\psi]$, we can apply Lemma 4.2 (ii) with $T = t_\psi - \alpha_\psi$, $\delta = \delta_\psi$ and $R > u$: there exists $\eta_\psi > 0$ such that

$$\limsup_{\varepsilon\to 0, y\to x} \varepsilon \ln \mathbf{P}^{\varepsilon,y}\left(\|X^{\varepsilon,y} - \psi\|_{0,t_\psi-\alpha_\psi} \le \eta_\psi, \|\sqrt{\varepsilon}W^{\varepsilon,y} - \varphi\|_{0,t_\psi-\alpha_\psi} \ge \delta_\psi\right) \le -R.$$

Since we have assumed that $K_x = \overline{K_x \cap S(\{J_T < +\infty\})}$, we can write

$$K_x \subset \bigcup_{\psi \in K_x \cap S(\{J_T<+\infty\})} B_T(\psi,\eta_\psi),$$

so, from the compactness of K_x follows the existence of a finite number of functions $\psi_1,\ldots,\psi_n$ in $K_x \cap S(\{J_T < +\infty\})$ such that

$$K_x \subset \bigcup_{i=1}^n B_T(\psi_i,\eta_i),$$

where we wrote η_i instead of η_{ψ_i}. It easily follows from the compactness of K that there exists a neighborhood $\mathcal{N}_x$ of x such that

$$K_{\mathcal{N}_x} \subset \bigcup_{i=1}^n B_T(\psi_i,\eta_i),$$

where $K_{\mathcal{N}_x} = \{\psi \in K : \psi(0) \in \mathcal{N}_x\}$. Define

$$U = \bigcup_{i=1}^n B_{t_i-\alpha_i}(\varphi_i,\delta_i),$$

where $t_i = t_{\psi_i}$, $\alpha_i = \alpha_{\psi_i}$ and $\delta_i = \delta_{\psi_i}$, and where φ_i is the function satisfying $S(\varphi_i) = \psi_i$ and $I_{Tx}(\psi_i) = J_T(\varphi_i)$.

Then, it remains to write for $y \in \mathcal{N}_x$

$$\begin{aligned}
\mathbf{P}^{\varepsilon,y}(X^{\varepsilon,y} \in K) &\le \mathbf{P}^{\varepsilon,y}(\sqrt{\varepsilon}W^{\varepsilon,y} \in U) + \mathbf{P}^{\varepsilon,y}(\sqrt{\varepsilon}W^{\varepsilon,y} \notin U,\, X^{\varepsilon,y} \in K_{\mathcal{N}_x})\\
&\le \sum_{i=1}^{n} \mathbf{P}^{\varepsilon,y}(\sqrt{\varepsilon}W^{\varepsilon,y} \in B_{t_i-\alpha_i}(\varphi_i,\delta_i))\\
&\quad + \sum_{i=1}^{n} \mathbf{P}^{\varepsilon,y}(\|X^{\varepsilon,y} - \psi_i\|_{0,T} \le \eta_i, \sqrt{\varepsilon}W^{\varepsilon,y} \notin U)\\
&\le \sum_{i=1}^{n} \mathbf{P}^{\varepsilon,y}(\|\sqrt{\varepsilon}W^{\varepsilon,y} - \varphi_i\|_{0,t_i-\alpha_i} < \delta_i)\\
&\quad + \sum_{i=1}^{n} \mathbf{P}^{\varepsilon,y}(\|X^{\varepsilon,y} - \psi_i\|_{0,t_i-\alpha_i} \le \eta_i, \|\sqrt{\varepsilon}W^{\varepsilon,y} - \varphi_i\|_{0,t_i-\alpha_i} \ge \delta_i),
\end{aligned}$$

and to observe that, by Schilder's Theorem and (54),

$$\begin{aligned}
&\limsup_{\varepsilon\to 0, y\to x} \varepsilon \ln \mathbf{P}^{\varepsilon,y}(\|\sqrt{\varepsilon}W^{\varepsilon,y} - \varphi_i\|_{0,t_i-\alpha_i} \le \delta_i)\\
&\qquad \le - \inf_{\varphi \in B_{t_i-\alpha_i}(\varphi_i,\delta_i)} J_{t_i-\alpha_i}(\varphi) \le -J_T(\varphi_i) + \rho,
\end{aligned}$$

to obtain

$$\begin{aligned}
&\limsup_{\varepsilon\to 0, y\to x} \varepsilon \ln \mathbf{P}^{\varepsilon,y}(X^{\varepsilon,y} \in K)\\
&\qquad \le \sup \Bigg\{ \sup_{1\le i\le n} \limsup_{\varepsilon\to 0, y\to x} \varepsilon \ln \mathbf{P}^{\varepsilon,y}(\|\sqrt{\varepsilon}W^{\varepsilon,y} - \varphi_i\|_{0,t_i-\alpha_i} < \delta_i),\\
&\sup_{1\le i\le n} \limsup_{\varepsilon\to 0, y\to x} \varepsilon \ln \mathbf{P}^{\varepsilon,y}(\|X^{\varepsilon,y} - \psi_i\|_{0,t_i-\alpha_i} \le \eta_i, \|\sqrt{\varepsilon}W^{\varepsilon,y} - \varphi_i\|_{0,t_i-\alpha_i} \ge \delta_i)\Bigg\}\\
&\qquad \le \sup \left\{ - \inf_{1\le i\le n} (J_T(\varphi_i) - \rho), -R \right\}\\
&\qquad \le \sup \{ -\inf\{I_{T,x}(\psi), \psi \in K\} + \rho, -R\} \le -u + \rho.
\end{aligned}$$

Since this holds for any $\rho > 0$, the proof of (41) for the set K is complete.

Now, let C be a closed subset of $\mathcal{C}([0,T],\overline{\mathcal{X}})$ such that $\mathcal{C}^1_x([0,T],\mathcal{X}\setminus\Gamma)$ is dense in $C \cap \mathcal{C}_x([0,T],\overline{\mathcal{X}})$. Define the compact set

$$\begin{aligned}
K_k &= \{\psi \in \mathcal{C}([0,T],\overline{\mathcal{X}}) : \|\psi(0) - x\| \le 1,\ \forall l \ge k,\ \omega(\psi, 1/l^3) \le 1/l\}\\
&= \cup_{\|y-x\|\le 1}\, K^y_k.
\end{aligned}$$

By Lemma 4.4,

$$\limsup_{\varepsilon\to 0, y\to x} \varepsilon \ln \mathbf{P}(X^{\varepsilon,y} \notin K_k) \leq -k/64d\Sigma^2. \tag{55}$$

The exponential tightness method requires to decompose C as $(C \cap (K_k)^c) \cup (C \cap K_k)$, to use inequality (55) to bound the probability of the first set, and to use the large deviations upper bound for compact sets to bound the probability of the second set. Unfortunately, $S\{(J_T < \infty\})$ may not be dense in $C \cap K_k \cap \mathcal{C}_x([0,T],\overline{\mathcal{X}})$. However, (55) still holds if we replace K_k by any bigger compact set. Therefore, we will introduce a compact set $\tilde{K}_k \supset K_k$ so that $S\{(J_T < \infty\})$ should be dense in $C \cap \tilde{K}_k \cap \mathcal{C}_x([0,T],\overline{\mathcal{X}})$.

Let us construct $\tilde{K}_k$ as follows. $C \cap K_k \cap \mathcal{C}_x([0,T],\overline{\mathcal{X}})$ is compact, so it is separable. Let ψ_n be a sequence of functions dense in this set. For all $n \geq 0$, $\psi_n \in C$, so, by assumption, there exists a sequence $(\psi_{n,p})_{p\geq 0}$ in $C \cap \mathcal{C}^1_x([0,T],\mathcal{X} \setminus \Gamma)$ converging to ψ_n. We can moreover assume that $\|\psi_{n,p} - \psi_n\|_{0,T} \leq 2^{-p}$ for all $p \geq 0$. Let us define

$$\tilde{K}_k = K_k \cup \left(\bigcup_{n\geq 0} \{\psi_{n,p} : p \geq n\} \right),$$

and let us prove that $\tilde{K}_k$ is compact. Let (ϕ_m) be a sequence of $\tilde{K}_k$, and let us extract a converging subsequence. The only problem is when $\{m : \phi_m \in K_k\}$ is finite, and when for all $n \geq 0$, $\{m : \phi_m \in \{\psi_{n,p} : p \geq n\}\}$ is finite. In this case, there exists two (strictly) increasing sequences of integers (α_m) and (β_m) such that for all $m \geq 0$, $\phi_{\alpha_m} \in \{\psi_{\beta_m,p} : p \geq \beta_m\}$. For all $m \geq 0$, ψ_{β_m} belongs to the compact set $C \cap K_k \cap \mathcal{C}_x([0,T],\overline{\mathcal{X}})$, so, extracting a subsequence from (α_m) and (β_m), we can suppose that $\psi_{\beta_m} \to \psi \in C \cap K_k \cap \mathcal{C}_x([0,T],\overline{\mathcal{X}})$. Then

$$\|\phi_{\alpha_m} - \psi\|_{0,T} \leq \|\phi_{\alpha_m} - \psi_{\beta_m}\|_{0,T} + \|\psi_{\beta_m} - \psi\|_{0,T} \leq 2^{-\beta_m} + \|\psi_{\beta_m} - \psi\|,$$

which converge to 0 when $m \to \infty$. Hence $\phi_{\alpha_m} \to \psi$, and $\tilde{K}_k$ is compact.

Moreover, since $\tilde{K}_k \supset K_k$, it follows from (55) that

$$\limsup_{\varepsilon\to 0, y\to x} \varepsilon \ln \mathbf{P}(X^{\varepsilon,y} \notin \tilde{K}_k) \leq -k/64d\Sigma^2, \tag{56}$$

and $\tilde{K}_k$ has been constructed in such a way that $\mathcal{C}^1_x([0,T],\mathcal{X} \setminus \Gamma)$ is dense in $C \cap \tilde{K}_k \cap \mathcal{C}_x([0,T],\overline{\mathcal{X}})$. So, by Lemma 4.3 (ii), $S(\{J_T < \infty\})$ is dense in $C \cap \tilde{K}_k \cap \mathcal{C}_x([0,T],\overline{\mathcal{X}})$, and we can apply the first part of the proof to the compact set $C \cap \tilde{K}_k$:

$$\limsup_{\varepsilon\to 0, y\to x} \varepsilon \ln \mathbf{P}(X^{\varepsilon,y} \in C \cap \tilde{K}_k) \leq -\inf\{I_{T,x}(\psi), \psi \in C \cap \tilde{K}_k\}.$$

Together with (56), this yields for sufficiently large k

$$\begin{aligned}\limsup_{\varepsilon\to 0, y\to x} \varepsilon \ln \mathbf{P}(X^{\varepsilon,y} \in C) &\leq \sup \left\{ \limsup_{\varepsilon\to 0, y\to x} \varepsilon \ln \mathbf{P}(X^{\varepsilon,y} \in C \cap \tilde{K}_k), \right. \\ &\qquad \left. \limsup_{\varepsilon\to 0, y\to x} \varepsilon \ln \mathbf{P}(X^{\varepsilon,y} \notin \tilde{K}_k) \right\} \\ &\leq - \inf_{\psi\in C} I_{T,x}(\psi),\end{aligned}$$

and the proof of (41) is complete □

Proof of Lemma 4.2 Let φ be as in any point of Lemma 4.2. We will first use the Girsanov's Theorem to restrict ourselves to the case $\varphi = 0$.

Define on $(\Omega^{\varepsilon,y}, \mathcal{F}_T^{\varepsilon,y})$ the probability measure $\tilde{\mathbf{P}}^{\varepsilon,y}$ by

$$\frac{d\tilde{\mathbf{P}}^{\varepsilon,y}}{d\mathbf{P}^{\varepsilon,y}} = \exp\left(\frac{1}{\sqrt{\varepsilon}} \int_0^t \dot{\varphi}_s dW_s^{\varepsilon,y} - \frac{1}{2\varepsilon} \int_0^t \|\dot{\varphi}_s\|^2 ds \right).$$

Since in all cases $J_T(\varphi) = 1/2 \int_0^T \|\dot{\varphi}_t\|^2 dt < +\infty$, by Novikov's criterion, Girsanov's Theorem is applicable and yields that $\tilde{W}_t^{\varepsilon,y} := W_t^{\varepsilon,y} - \varphi_t/\sqrt{\varepsilon}$ is a $\tilde{\mathbf{P}}^{\varepsilon,y}$-Brownian motion for $t \leq T$. If we denote by $\tilde{X}^{\varepsilon,y}$ the process $X^{\varepsilon,y}$ on $(\Omega^{\varepsilon,y}, \mathcal{F}_t^{\varepsilon,y}, \tilde{W}^{\varepsilon,y}, \tilde{\mathbf{P}}^{\varepsilon,y})$, then, $\tilde{\mathbf{P}}^{\varepsilon,y}$-a.s., for $t \leq T$

$$\tilde{X}_t^{\varepsilon,y} = y + \int_0^t (b^\varepsilon(\tilde{X}_s^{\varepsilon,y}) + \sigma(\tilde{X}_s^{\varepsilon,y})\dot{\varphi}_s) ds + \sqrt{\varepsilon} \int_0^t \sigma(\tilde{X}_s^{\varepsilon,y}) d\tilde{W}_s^{\varepsilon,y}. \tag{57}$$

Let $F^{\varepsilon,y}$ denote the event $\{\|X^{\varepsilon,y} - S(\varphi)\|_{0,T} \geq \eta,\ \|\sqrt{\varepsilon}W^{\varepsilon,y} - \varphi\|_{0,T} \leq \delta\} = \{\|\tilde{X}^{\varepsilon,y} - S(\varphi)\|_{0,T} \geq \eta,\ \|\sqrt{\varepsilon}\tilde{W}^{\varepsilon,y}\|_{0,T} \leq \delta\}$. It follows from Cauchy-Schwartz's Theorem that

$$\mathbf{P}^{\varepsilon,y}(F^{\varepsilon,y}) = \int \mathbf{1}_{F^{\varepsilon,y}} \frac{d\mathbf{P}^{\varepsilon,y}}{d\tilde{\mathbf{P}}^{\varepsilon,y}} d\tilde{\mathbf{P}}^{\varepsilon,y} \leq \left(\tilde{\mathbf{P}}^{\varepsilon,y}(F^{\varepsilon,y}) \right)^{\frac{1}{2}} \left(\int \left(\frac{d\mathbf{P}^{\varepsilon,y}}{d\tilde{\mathbf{P}}^{\varepsilon,y}} \right)^2 d\tilde{\mathbf{P}}^{\varepsilon,y} \right)^{\frac{1}{2}}. \tag{58}$$

Since $W_t^{\varepsilon,y} = \tilde{W}_t^{\varepsilon,y} + \varphi_t/\sqrt{\varepsilon}$, we can write

$$\begin{aligned}\left(\frac{d\mathbf{P}^{\varepsilon,y}}{d\tilde{\mathbf{P}}^{\varepsilon,y}} \right)^2 &= \exp\left(-\frac{2}{\sqrt{\varepsilon}} \int_0^T \dot{\varphi}_s d\tilde{W}_s^{\varepsilon,y} - \frac{1}{\varepsilon} \int_0^T \|\dot{\varphi}_s\|^2 ds \right) \\ &= \exp\left(\int_0^T \left(-\frac{2\dot{\varphi}_s}{\sqrt{\varepsilon}} \right) d\tilde{W}_s^{\varepsilon,y} - \frac{1}{2} \int_0^T \left\| \frac{2\dot{\varphi}_s}{\sqrt{\varepsilon}} \right\|^2 ds \right) \\ &\qquad \times \exp\left(\frac{1}{\varepsilon} \int_0^T \|\dot{\varphi}_s\|^2 ds \right).\end{aligned}$$

The first term in the product of the right-hand side is a $\tilde{\mathbf{P}}^{\varepsilon,y}$-martingale (by Novikov's criterion), and the second term is equal to $\exp(2J_T(\varphi)/\varepsilon)$. Therefore, (58) yields

$$\varepsilon \ln \mathbf{P}^{\varepsilon,y}(F^{\varepsilon,y}) \le \frac{\varepsilon}{2} \ln \tilde{\mathbf{P}}^{\varepsilon,y}(F^{\varepsilon,y}) + J_T(\varphi).$$

Therefore, the proof of Lemma 4.2 is completed if we can prove the following lemma. □

Lemma 4.5 *The three points of Lemma 4.2 hold when (49) and (50) are replaced respectively by*

$$\limsup_{\varepsilon\to 0, y\to x} \varepsilon \ln \tilde{\mathbf{P}}^{\varepsilon,y}\left(\|\tilde{X}^{\varepsilon,y} - S(\varphi)\|_{0,T} \ge \eta, \|\sqrt{\varepsilon}\tilde{W}^{\varepsilon,y}\|_{0,T} \le \delta\right) \le -R \tag{59}$$

and

$$\limsup_{\varepsilon\to 0, y\to x} \varepsilon \ln \tilde{\mathbf{P}}^{\varepsilon,y}\left(\|\tilde{X}^{\varepsilon,y} - S(\varphi)\|_{0,T} \le \eta, \|\sqrt{\varepsilon}\tilde{W}^{\varepsilon,y}\|_{0,T} \ge \delta\right) \le -R. \tag{60}$$

In order to keep notations simple, we will write throughout the proof of this lemma W instead of $\tilde{W}^{\varepsilon,y}$ and $\mathbf{P}$ instead of $\tilde{\mathbf{P}}^{\varepsilon,y}$, for events involving the process $\tilde{X}^{\varepsilon,y}$ solution to (57).

Lemma 4.5 relies on the following lemma, which proof is postponed after the proof of Lemma 4.5.

Lemma 4.6 *Let $\tilde{X}^{\varepsilon,y}$ be defined by (57) with φ satisfying $J_T(\varphi) < \infty$. Let Y_t be a $\mathcal{F}_t^{\varepsilon,y}$-martingale such that $\sup_{t\le T} \|\langle Y\rangle_t\| \le A$, let τ be a $\mathcal{F}_t^{\varepsilon,y}$-stopping time, and let ξ be uniformly continuous bounded function defined on $\overline{\mathcal{X}}$. Then, $\forall \eta > 0$, $\forall R > 0$, $\exists \delta > 0$, $\exists \varepsilon_0 > 0$ both depending on Y only through A, and both independent of τ, such that $\forall y \in \overline{\mathcal{X}}$, $\forall \varepsilon < \varepsilon_0$,*

$$\varepsilon \ln \mathbf{P}\left(\left\|\sqrt{\varepsilon}\int_0^{\cdot\wedge\tau} \xi(\tilde{X}_s^{\varepsilon,y})dY_s\right\|_{0,T} \ge \eta,\ \|\sqrt{\varepsilon}Y\|_{0,T} \le \delta\right) \le -R. \tag{61}$$

Proof of Lemma 4.5 (i) The function $\psi = S(\varphi)$ does not take any value in Γ on $[0,T]$, so there exists $\alpha > 0$ such that $\forall t \in [0,T]$, $\psi_t \in \Gamma_\alpha$. Suppose without loss of generality that $\eta < \alpha/2$, and define for $y \in \overline{\mathcal{X}}$

$$\tau^{\varepsilon,y} = \inf\{t : d(\tilde{X}_t^{\varepsilon,y},\Gamma) \le \alpha/2\} \wedge T.$$

When $\tau^{\varepsilon,y} < T$, $\|\tilde{X}^{\varepsilon,y}_{\tau^{\varepsilon,y}} - S(\varphi)_{\tau^{\varepsilon,y}}\| \ge d(S(\varphi)_{\tau^{\varepsilon,y}},\Gamma) - d(\tilde{X}^{\varepsilon,y}_{\tau^{\varepsilon,y}},\Gamma) \ge \alpha/2 > \eta$, so, in any case,

$$\|\tilde{X}^{\varepsilon,y} - S(\varphi)\|_{[0,T]} \ge \eta \Rightarrow \|\tilde{X}^{\varepsilon,y} - S(\varphi)\|_{[0,\tau^{\varepsilon,y}]} \ge \eta.$$

Consequently, (59) will be proved if we find $\delta > 0$ such that

$$\limsup_{\varepsilon\to 0, y\to x} \varepsilon \ln \mathbf{P}(\|\tilde{X}^{\varepsilon,y} - S(\varphi)\|_{0,\tau^{\varepsilon,y}} \geq \eta,\ \|\sqrt{\varepsilon}W\|_{0,T} \leq \delta) \leq -R.$$

Now, take C such that σ is C-Lipschitz on $\Gamma_{\alpha/2}$, b is C-Lipschitz and $\tilde{b}$ is bounded by C on $\overline{\mathcal{X}}$ (see Proposition 2.1). It follows from (57) that, for $t \leq \tau^{\varepsilon,y}$,

$$\begin{aligned}
&\|\tilde{X}^{\varepsilon,y}_t - S(\varphi)_t\| \\
&\quad \leq \sqrt{\varepsilon}\left\|\int_0^t \sigma(\tilde{X}^{\varepsilon,y}_s)dW_s\right\| + \varepsilon\int_0^t \|\tilde{b}(\tilde{X}^{\varepsilon,y}_s)\|ds + \|x-y\| \\
&\qquad + \int_0^t \|b(\tilde{X}^{\varepsilon,y}_s) - b(S(\varphi)_s)\|ds + \int_0^t \|\sigma(\tilde{X}^{\varepsilon,y}_s) - \sigma(S(\varphi)_s)\| \times \|\dot{\varphi}_s\|ds \\
&\quad \leq \sqrt{\varepsilon}\left\|\int_0^t \sigma(\tilde{X}^{\varepsilon,y}_s)dW_s\right\| + \varepsilon CT + \|x-y\| \\
&\qquad + C\int_0^t (1+\|\dot{\varphi}_s\|)\|\tilde{X}^{\varepsilon,y}_s - S(\varphi)_s\|ds.
\end{aligned}$$

Remind that $u := \int_0^T \|\dot{\varphi}_s\|^2 ds < +\infty$. Gronwall's Lemma and the Cauchy-Schwartz inequality yield for $t \leq \tau^{\varepsilon,y}$

$$\begin{aligned}
&\|\tilde{X}^{\varepsilon,y}_t - S(\varphi)_t\| \\
&\quad \leq \left[\sqrt{\varepsilon}\left\|\int_0^t \sigma(\tilde{X}^{\varepsilon,y}_s)dW_s\right\| + \varepsilon CT + \|x-y\|\right] \exp\left[C\left(T+\sqrt{uT}\right)\right].
\end{aligned}$$

Therefore, it suffices to find $\delta > 0$ such that

$$\limsup_{\varepsilon\to 0, y\to x} \varepsilon \ln \mathbf{P}\left(\sqrt{\varepsilon}\left\|\int_0^t \sigma(\tilde{X}^{\varepsilon,y}_s)dW_s\right\|_{0,\tau^{\varepsilon,y}} \geq \eta\beta,\ \sqrt{\varepsilon}\|W\|_{0,T} \leq \delta\right) \leq -R,$$

where $\beta = \exp[-C(T+\sqrt{uT})]/2$. This is an direct consequence of Lemma 4.6 with $Y = W$, $A = 1$, $\xi = \sigma$ and $\tau = \tau^{\varepsilon,y}$ (by Proposition 2.1 (ii), σ is uniformly Hölder, so ξ is uniformly continuous). □

Proof of Lemma 4.5 (ii) As above, take $\alpha > 0$ such that $\forall t \in [0,T]$, $S(\varphi)_t \in \Gamma_\alpha$. Fix $\eta \leq \alpha/2$. Then, on the event $\{\|\tilde{X}^{\varepsilon,y} - S(\varphi)\|_{0,T} \leq \eta\}$, for any $t \in [0,T]$, $\tilde{X}^{\varepsilon,y}_t \in \Gamma_{\alpha/2}$. Take C such that b is C-Lipschitz, $\tilde{b}$ is bounded by C, and σ is C-Lipschitz on $\Gamma_{\alpha/2}$. It follows from equation (57) that, on the event $\{\|\tilde{X}^{\varepsilon,y} - S(\varphi)\|_{[0,T]} \leq \eta\}$, for any $t \in [0,T]$,

$$\begin{aligned}
&\sqrt{\varepsilon}\left\|\int_0^t \sigma(\tilde{X}^{\varepsilon,y}_s)dW_s\right\| \\
&\quad = \left\|\tilde{X}^{\varepsilon,y}_t - S(\varphi)_t + y - x - \int_0^t [\sigma(\tilde{X}^{\varepsilon,y}_s) - \sigma(S(\varphi)_s)]\dot{\varphi}_s ds \right.\\
&\qquad \left. - \int_0^t [b(\tilde{X}^{\varepsilon,y}_s) - b(S(\varphi)_s)]ds - \varepsilon \int_0^t \tilde{b}(\tilde{X}^{\varepsilon,y}_s)ds\right\|
\end{aligned}$$

$$\begin{aligned}
&\leq \|\tilde{X}^{\varepsilon,y}_t - S(\varphi)_t\| + \|x-y\| + C\int_0^T (1+\|\dot{\varphi}_s\|)\|\tilde{X}^\varepsilon_s - S(\varphi)_s\|ds + \varepsilon CT \\
&\leq \eta + \|x-y\| + C\eta(T+\sqrt{uT}) + \varepsilon CT \leq \eta(2+2CT+C\sqrt{uT})
\end{aligned}$$

for $\varepsilon < \eta$ and $\|x-y\| \leq \eta$. Therefore, using the notation $\beta = 2+2CT+C\sqrt{uT}$,

$$\begin{gathered}
\{\|\tilde{X}^{\varepsilon,y} - S(\varphi)\|_{0,T} \leq \eta,\ \|\sqrt{\varepsilon}W\|_{0,T} \geq \delta\} \subset \qquad (62)\\
\left\{\forall t \in [0,T],\ \tilde{X}^{\varepsilon,y}_t \in \Gamma_{\frac{\alpha}{2}},\ \sqrt{\varepsilon}\left\|\int_0^t \sigma(\tilde{X}^{\varepsilon,y}_s)dW_s\right\|_{0,T} \leq \eta\beta,\ \sqrt{\varepsilon}\|W\|_{0,T} \leq \delta\right\}.
\end{gathered}$$

Define

$$\begin{aligned}
\tau^{\varepsilon,y} &= \inf\{t : d(\tilde{X}^{\varepsilon,y}_t,\Gamma) \leq \alpha/2\} \wedge T,\\
Y^{\varepsilon,y}_t &= \int_0^t \sigma(\tilde{X}^{\varepsilon,y}_s)dW_s,\\
\xi &= \chi\sigma^{-1},
\end{aligned}$$

where χ is a Lipschitz function from $\overline{\mathcal{X}}$ to [0,1] such that $\chi(x) = 0$ if $d(x,\Gamma) \leq \alpha/4$ and $\chi(x) = 1$ if $d(x,\Gamma) \geq \alpha/2$. So $\xi(x) = \sigma^{-1}(x)$ if $x \in \Gamma_{\alpha/2}$. With these notations, a small computation shows that (62) rewrites

$$\begin{gathered}
\{\|\tilde{X}^{\varepsilon,y} - S(\varphi)\|_{0,T} \leq \eta,\ \|\sqrt{\varepsilon}W\|_{0,T} \geq \delta\} \subset\\
\left\{\sqrt{\varepsilon}\|Y^{\varepsilon,y}\|_{0,T} \leq \eta\beta,\ \sqrt{\varepsilon}\left\|\int_0^{t\wedge\tau^{\varepsilon,y}} \xi(\tilde{X}^{\varepsilon,y}_s)dY^{\varepsilon,y}_s\right\|_{0,T} \geq \delta\right\}
\end{gathered}$$

and (60) is now a direct consequence of Lemma 4.6: ξ is Lipschitz and bounded on $\overline{\mathcal{X}}$ (by Proposition 2.1 (iii), $\Gamma_{\alpha/4} \subset \{x \in \mathcal{X} : a(x) \in \mathcal{S}_c\}$ for some $c > 0$, so $\sigma(x) \in \mathcal{S}_{\sqrt{c}}$ for $x \in \Gamma_{\alpha/4}$, and it remains to observe that the inverse matrix application is Lipschitz and bounded on $\mathcal{S}_{\sqrt{c}}$), and for any $t \leq \tau$, $\langle Y^{\varepsilon,y}\rangle_t = \int_0^t a(\tilde{X}^{\varepsilon,y}_s)ds$ which is bounded, by Proposition 2.1 (i), by a constant A independent of y and ε. $\square$

Proof of Lemma 4.5 (iii) In Lemma 4.5 (iii), φ is defined from $\tilde{\varphi}$ by $\varphi_t = \tilde{\varphi}_t$ for $t \leq t_\psi$, and $\varphi_t = \tilde{\varphi}_{t_\psi}$ otherwise (*i.e.* $\dot{\varphi}_t = 0$ for $t > t_\psi$), where $\psi = S(\tilde{\varphi})$. Then $\psi = S(\varphi) = S(\tilde{\varphi})$ since $S(\varphi)$ does not depend on φ_t for $t \geq t_\psi$. By the Cauchy-Schwartz inequality, $\int_0^{t_\psi} \|\dot{\varphi}_s\| ds \leq (2T J_T(\varphi))^{1/2} < +\infty$, so there exists $\rho > 0$ small enough such that

$$\int_{t_\psi - \rho}^{t_\psi} \|\dot{\varphi}_s\| ds \leq \frac{\eta e^{-CT}}{8C}, \tag{63}$$

where C is a constant bounding b, $\tilde{b}$ and σ, and such that b is C-Lipschitz.

Distinguishing when $\|\tilde{X}^{\varepsilon,y} - \psi\|_{0,t_\psi-\rho} \geq \eta e^{-CT}/4$ or not, we can write

$$\{\|\tilde{X}^{\varepsilon,y} - \psi\|_{0,T} \geq \eta,\ \|\sqrt{\varepsilon} W\|_{0,T} \leq \delta\} \subset D^{\varepsilon,y} \cup E^{\varepsilon,y},$$

where

$$D^{\varepsilon,y} = \left\{ \|\tilde{X}^{\varepsilon,y} - \psi\|_{0,t_\psi-\rho} \leq \frac{\eta e^{-CT}}{4},\ \|\tilde{X}^{\varepsilon,y} - \psi\|_{t_\psi-\rho,T} \geq \eta, \|\sqrt{\varepsilon} W\|_{0,T} \leq \delta \right\}$$

$$\text{and} \quad E^{\varepsilon,y} = \left\{ \|\tilde{X}^{\varepsilon,y} - \psi\|_{0,t_\psi-\rho} \geq \frac{\eta e^{-CT}}{4},\ \|\sqrt{\varepsilon} W\|_{0,t_\psi-\rho} \leq \delta \right\}.$$

Part (i) of Lemma 4.5 shows that $\mathbf{P}(E^{\varepsilon,y})$ has the required exponential decay if δ is small enough. Let us estimate the probability of $D^{\varepsilon,y}$.

It follows from (57) and from the fact that $\dot{\varphi}_t = 0$ for $t > t_\psi$ that, for any $t \geq t_\psi - \rho$

$$\begin{aligned} \|\tilde{X}_t^{\varepsilon,y} - \psi_t\| &\leq \|\tilde{X}_{t_\psi-\rho}^{\varepsilon,y} - \psi_{t_\psi-\rho}\| + \sqrt{\varepsilon} \left\| \int_{t_\psi-\rho}^t \sigma(\tilde{X}_s^{\varepsilon,y}) dW_s \right\| \\ &+ C \int_{t_\psi-\rho}^t \|\tilde{X}_s^{\varepsilon,y} - \psi_s\| ds + \varepsilon CT + \int_{t_\psi-\rho}^{t_\psi \wedge t} \|\sigma(\tilde{X}_s^{\varepsilon,y}) - \sigma(\psi_s)\| \times \|\dot{\varphi}_s\| ds. \end{aligned}$$

On the event $D^{\varepsilon,y}$, the first term of the right-hand side is smaller than $\eta e^{-CT}/4$, and, since σ is bounded by C, the last term can be bounded by $2C \int_{t_\psi-\rho}^{t_\psi} \|\dot{\varphi}\| ds$, which is smaller than $\eta e^{-CT}/4$ by (63). Moreover, we can suppose ε small enough to have $\varepsilon CT \leq \eta e^{-CT}/4$. So, on the event $D^{\varepsilon,y}$, by Gronwall's Lemma, for $t \geq t_\psi - \rho$,

$$\|\tilde{X}_t^{\varepsilon,y} - \psi_t\| \leq \left[\frac{3}{4} \eta e^{-CT} + \sqrt{\varepsilon} \left\| \int_{t_\psi-\rho}^t \sigma(\tilde{X}_s^{\varepsilon,y}) dW_s \right\| \right] e^{CT}.$$

Since $\|\tilde{X}^{\varepsilon,y} - \psi\|_{t_\psi-\rho,T} \geq \eta$ on $D^{\varepsilon,y}$, we finally can write

$$D^{\varepsilon,y} \subset$$
$$\left\{ \left\| \sqrt{\varepsilon} \int_{t_\psi-\rho}^{\cdot} \sigma(\tilde{X}^{\varepsilon,y}_s) dW_s \right\|_{t_\psi-\rho,T} \geq \frac{\eta e^{-CT}}{4},\ \|\sqrt{\varepsilon}(W_\cdot - W_{t_\psi-\rho})\|_{t_\psi-\rho,T} \leq 2\delta \right\}$$

and (59) is now a consequence of Lemma 4.6. □

Now let us come to the proof of Lemma 4.6. It is adapted from the proof of Lemma 1.3 in Doss and Priouret [23], and makes use of the following classical exponential inequalities for stochastic integrals, which proof is omitted, and where $\mathcal{M}_{d,d}$ denotes the set of real $d \times d$ matrices, and $\|\cdot\|$ is the norm on $\mathcal{M}_{d,d}$ defined by $\|M\| = \sup_{\|\zeta\|=1} \|M\zeta\|$:

Lemma 4.7 *Let Y_t be a $\mathcal{F}_t$-martingale with values in $\mathbb{R}^d$ on a filtered probability space $(\Omega,\mathcal{F},\mathcal{F}_t,\mathbf{P})$, and suppose that its quadratic covariation process $\langle Y\rangle_t$ satisfies $\sup_{t\leq T} \|\langle Y\rangle_t\| \leq A$. Let τ be a $\mathcal{F}_t$ stopping time, and let $Z : \mathbb{R}_+ \times \Omega \to \mathcal{M}_{d,d}$ be a progressively measurable process such that $\sup_{t\leq\tau} \|Z^*_t\| \leq B$ (where Z^*_t is the transpose matrix of Z_t). Then for any $R > 0$,*

$$\mathbf{P}\left(\sup_{t\leq T} \left\| \int_0^{t\wedge\tau} Z_s dY_s \right\| \geq R \right) \leq 2d \exp\left(-\frac{R^2}{2dTAB^2}\right).$$

The proof of this lemma follows immediately the proof of Lemma 4.6.

Proof of Lemma 4.6 We use a discretization technique: for any $p \in \mathbb{N}$, let us define $\tilde{X}^{\varepsilon,y,p}_t = \tilde{X}^{\varepsilon,y}_{k2^{-p}}$, where $k \in \mathbb{N}$ is such that $k \leq t2^{-p} < k+1$. Given $\gamma > 0$, $p \geq 1$ and $\delta > 0$, to be determined next, we can write

$$\left\{ \left\| \sqrt{\varepsilon} \int_0^{\cdot\wedge\tau} \xi(\tilde{X}^{\varepsilon,y}_s) dY_s \right\|_{0,T} \geq \eta,\ \|\sqrt{\varepsilon} Y\|_{0,T} \leq \delta \right\} \subset A^\varepsilon \cup B^\varepsilon \cup C^\varepsilon,$$

where

$$A^\varepsilon = \{\|\tilde{X}^{\varepsilon,y} - \tilde{X}^{\varepsilon,y,p}\|_{0,\tau} \geq \gamma\},$$
$$B^\varepsilon = \left\{ \|\tilde{X}^{\varepsilon,y} - \tilde{X}^{\varepsilon,y,p}\|_{0,\tau} \leq \gamma,\ \left\| \sqrt{\varepsilon} \int_0^{\cdot\wedge\tau} [\xi(\tilde{X}^{\varepsilon,y}_s) - \xi(\tilde{X}^{\varepsilon,y,p}_s)] dY_s \right\|_{0,T} \geq \frac{\eta}{2} \right\}$$
$$\text{and} \quad C^\varepsilon = \left\{ \left\| \sqrt{\varepsilon} \int_0^{\cdot\wedge\tau} \xi(\tilde{X}^{\varepsilon,y,p}_s) dY_s \right\|_{0,T} \geq \frac{\eta}{2},\ \|\sqrt{\varepsilon} Y\|_{0,T} \leq \delta \right\}.$$

We will choose first γ such that $\mathbf{P}(B^\varepsilon)$ is sufficiently small, then choose $p \geq 1$ to control $\mathbf{P}(A^\varepsilon)$, and finally choose $\delta > 0$ such that $C^\varepsilon = \emptyset$.

Firstly, let us apply Lemma 4.7 with $Z_t = \sqrt{\varepsilon}[\xi(\tilde{X}_t^{\varepsilon,y}) - \xi(\tilde{X}_t^{\varepsilon,y,p})]$. If we define $B_\gamma := \sup_{\|x-y\| \leq \gamma} \|\xi(x) - \xi(y)\|$, then, on B^ε, for $t \leq \tau$, $\|Z_t^*\| \leq \sqrt{\varepsilon} B_\gamma$. Therefore, Lemma 4.7 gives

$$\mathbf{P}(B^\varepsilon) \leq 2d \exp\left(-\frac{\eta^2/4}{2dTA\varepsilon B_\gamma^2}\right).$$

Now, since ξ is uniformly continuous, $B_\gamma \to 0$ when $\gamma \to 0$. Therefore, for $\varepsilon < 1$, $\varepsilon \ln \mathbf{P}(B^\varepsilon)$ can be made smaller than $-2R$ choosing γ small enough.

Secondly, $\gamma > 0$ being fixed as above, equation (57) yields

$$\begin{aligned}
&\mathbf{P}(\|\tilde{X}^{\varepsilon,y} - \tilde{X}^{\varepsilon,y,p}\|_{0,\tau} \geq \gamma) \\
&\quad \leq \sum_{k=0}^{T2^p-1} \mathbf{P}\left(\sup_{k2^{-p} \leq t \leq (k+1)2^{-p}} \left\| \int_{k2^{-p}\wedge\tau}^{t\wedge\tau} \sqrt{\varepsilon}\sigma(\tilde{X}_s^{\varepsilon,y})dW_s \right\| \geq \frac{\gamma}{2}\right) \\
&\qquad + \sum_{k=0}^{T2^p-1} \mathbf{P}\left(\sup_{k2^{-p} \leq t \leq (k+1)2^{-p}} \left\| \int_{k2^{-p}\wedge\tau}^{t\wedge\tau} \left[b^\varepsilon(\tilde{X}_s^{\varepsilon,y}) + \sigma(\tilde{X}_s^{\varepsilon,y})\dot{\varphi}_s\right] ds \right\| \geq \frac{\gamma}{2}\right) \\
&\quad \leq \sum_{k=0}^{T2^p-1} \mathbf{P}\left(\sup_{k2^{-p} \leq t \leq (k+1)2^{-p}} \left\| \int_{k2^{-p}\wedge\tau}^{t\wedge\tau} \sqrt{\varepsilon}\sigma(\tilde{X}_s^{\varepsilon,y})dW_s \right\| \geq \frac{\gamma}{2}\right) \\
&\qquad + \sum_{k=0}^{T2^p-1} \mathbf{P}\left(C2^{-p} + C2^{-p/2}\sqrt{u} \geq \frac{\gamma}{2}\right),
\end{aligned}$$

where C is a bound for b^ε and σ, where $u = 2J_T(\varphi) = \int_0^T \|\dot{\varphi}_s\|^2 ds < +\infty$, and where we have used the Cauchy-Schwartz inequality to obtain $\sqrt{u}$ in the last line of this inequality. For p big enough, the second sum of the right-hand side equals 0. For the first sum, Lemma 4.7 with $\tau = T = 2^{-p}$, $Y = W$, $A = 1$, $R = \gamma/2$ and $B = \sqrt{\varepsilon}C$ gives that, for $0 \leq k < T2^p$,

$$\mathbf{P}\left(\sup_{k2^{-p} \leq t \leq (k+1)2^{-p}} \left\| \int_{k2^{-p}\wedge\tau}^{t\wedge\tau} \sqrt{\varepsilon}\sigma(\tilde{X}_s^{\varepsilon,y})dW_s \right\| \geq \frac{\gamma}{2}\right) \leq 2d \exp\left(-\frac{\gamma^2/4}{2d2^{-p}C^2\varepsilon}\right).$$

Therefore, taking p large enough, $\varepsilon \ln \mathbf{P}(A^\varepsilon) \leq -2R$ for all $\varepsilon < 1$.

Finally, with $p \geq 1$ and $\gamma > 0$ as above, for $t \leq T$,

$$\sqrt{\varepsilon} \int_0^{t\wedge\tau} \xi(\tilde{X}_s^{\varepsilon,y,p})dY_s = \sum_{i=0}^{T2^p-1} \sqrt{\varepsilon}\xi(\tilde{X}_{i2^{-p}\wedge\tau}^{\varepsilon,y})[Y_{(i+1)2^{-p}\wedge t\wedge\tau} - Y_{i2^{-p}\wedge t\wedge\tau}].$$

On C^ε, $\|\sqrt{\varepsilon}Y\|_{[0,T]} \leq \delta$, so, for $t \leq T$,

$$\left\| \sqrt{\varepsilon} \int_0^{t\wedge\tau} \xi(\tilde{X}_s^{\varepsilon,y,p}) dY_s \right\| \leq \sum_{i=0}^{T2^p-1} 2\delta C,$$

where C is a bound for ξ. Hence $C^\varepsilon = \emptyset$ as soon as $\delta < \eta 2^{-(p+2)}/CT$.

We obtain that $\varepsilon \ln \mathbf{P}(A^\varepsilon \cup B^\varepsilon \cup C^\varepsilon) \leq \varepsilon \ln 2 - 2R$, which yields (61) as soon as $\varepsilon < R/\ln 2 \wedge 1$.

This argument is true for any $y \in \overline{\mathcal{X}}$ and for any stopping time τ. It remains to observe that A is the only information about Y that we used to estimate $\mathbf{P}(B^\varepsilon)$, that Y does not appear in A^ε, and that no assumption about Y is necessary to obtain $C^\varepsilon = \emptyset$. Hence, the constant A is the only information about Y required to obtain δ and ε_0. $\square$

Proof of Lemma 4.7 For any $v \in \mathbb{R}^d$, let $M(v)$ be the exponential local martingale defined for $t > 0$ by

$$M_t(v) = \exp\left[v^* \int_0^{t\wedge\tau} Z_s dY_s - \frac{1}{2} \int_0^{t\wedge\tau} (v^* Z_s) d\langle Y \rangle_s (v^* Z_s)^* \right].$$

Since $\int_0^{t\wedge\tau} (v^* Z_s) d\langle Y \rangle_s (v^* Z_s)^* \leq TAB^2 \|v\|^2$, by Novikov's criterion, $M_t(v)$ is actually a martingale.

Then Doob's inequality gives that if $\|v\| = 1$ and $\lambda > 0$,

$$\begin{aligned} &\mathbf{P}\left(\sup_{0\leq t\leq T} v^* \int_0^{t\wedge\tau} Z_s dY_s \geq \frac{R}{\sqrt{d}} \right) \\ &\quad \leq \mathbf{P}\left(\sup_{0\leq t\leq T} M_t(\lambda v) \geq \exp\left(\frac{\lambda R}{\sqrt{d}} - \frac{\lambda^2 AB^2 T}{2} \right) \right) \\ &\quad \leq \exp\left(-\frac{\lambda R}{\sqrt{d}} + \frac{\lambda^2 AB^2 T}{2} \right). \end{aligned}$$

The infimum of the right-hand side is obtained when $\lambda = \frac{R}{\sqrt{d}AB^2T}$, which gives

$$\mathbf{P}\left(\sup_{0\leq t\leq T} v^* \int_0^{t\wedge\tau} Z_s dY_s \geq \frac{R}{\sqrt{d}} \right) \leq \exp\left(-\frac{R^2}{2dAB^2T} \right).$$

Finally, let us consider an orthonormal basis $\{v_1,\dots,v_d\}$ of $\mathbb{R}^d$. Then

$$\begin{aligned}\mathbf{P}\left(\sup_{t\le T}\left\|\int_0^{t\wedge\tau} Z_s dY_s\right\|\ge R\right) &= \mathbf{P}\left(\sup_{t\le T}\sum_{i=1}^d\left(v_i\cdot\int_0^{t\wedge\tau} Z_s dY_s\right)^2\ge R^2\right)\\ &\le \sum_{i=1}^d \mathbf{P}\left(\sup_{t\le T}\left|v_i^*\int_0^{t\wedge\tau} Z_s dY_s\right|\ge\frac{R}{\sqrt{d}}\right)\\ &\le 2d\exp\left(-\frac{R^2}{2dTAB^2}\right)\end{aligned}$$

which is exactly the required result. □

The proof of Lemma 4.2 is now completed. Let us come to the proof of Lemmas 4.3 and 4.4.

Proof of Lemma 4.3 Take $\psi\in\tilde{\mathcal{C}}_x^{ac}([0,T],\overline{\mathcal{X}})$, where $\tilde{\mathcal{C}}^{ac}([0,T],\overline{\mathcal{X}})$ has been defined in (38), and assume that there exists $\varphi\in\mathcal{C}_0^{ac}([0,T],\mathbb{R}^d)$ such that $S(\varphi)=\psi$. Then for any $t\in[0,T]$

$$\dot\psi_t = b(\psi_t)+\sigma(\psi_t)\dot\varphi_t.$$

For $t<t_\psi$, $\psi_t\notin\Gamma$, so $\sigma(\psi_t)$ is invertible, and

$$\dot\varphi_t=\sigma^{-1}(\psi_t)[\dot\psi_t-b(\psi_t)]. \tag{64}$$

This defines uniquely the function φ on $[0,t_\psi)$, and φ_{t_ψ} exists if and only if $\sigma^{-1}(\psi_t)[\dot\psi_t-b(\psi_t)]$ is $\mathbb{L}^1$ on $[0,t_\psi]$. Assume this is true. Then, since for $t\ge t_\psi$, $\psi_t\in\Gamma$, and so $b(\psi_t)=\sigma(\psi_t)=0$, $S(\varphi)_t=\psi_t=\psi_{t_\psi}$ for $t>t_\psi$ whatever is the function φ_t on $(t_\psi,T]$. Consequently,

$$\begin{aligned}J_T(\varphi)&=\frac12\int_0^T\|\dot\varphi_t\|^2dt\\ &\ge\frac12\int_0^{t_\psi}\|\dot\varphi_t\|^2dt=\frac12\int_0^{t_\psi}\|\sigma^{-1}(\psi_t)[\dot\psi_t-b(\psi_t)]\|^2dt=I_{T,x}(\psi),\end{aligned}$$

and, when $I_{T,x}(\psi)<+\infty$, a solution φ to $S(\varphi)=\psi$ satisfies $J_T(\varphi)=I_{T,x}(\psi)$ if and only if φ is constant for $t\ge t_\psi$.

Conversely, when $\sigma^{-1}(\psi_t)[\dot\psi_t-b(\psi_t)]$ is not $\mathbb{L}^1$ on $[0,t_\psi]$ or ψ does not belong to $\tilde{\mathcal{C}}_x^{ac}([0,T],\overline{\mathcal{X}})$, there is no solution to $S(\varphi)=\psi$. Moreover, in this case $I_{T,X}(\psi)=+\infty$, so the proof of (i) is completed.

Since by Proposition 2.1 (iii) σ is uniformly non-degenerate on Γ_α for any $\alpha>0$, it follows trivially from (64) that $\mathcal{C}^1([0,T],\mathcal{X}\setminus\Gamma)\subset S(\{J_T<\infty\})$. (ii) is now clear. □

Proof of Lemma 4.4 It follows from (1) that, for any $y \in \overline{\mathcal{X}}$, $s > 0$ and $t \in [0,T]$,

$$\|X^{\varepsilon,y}_{t+s} - X^{\varepsilon,y}_t\| \leq Cs + \sqrt{\varepsilon}\left\|\int_t^{t+s} \sigma(X^{\varepsilon,y}_u)dW_u\right\|$$

where C is a bound for b^ε (for $\varepsilon < 1$). So, for a given $h > 0$, we can apply Lemma 4.7 with $R \geq Ch$ to obtain

$$\mathbf{P}\left(\sup_{0\leq s\leq h} \|X^{\varepsilon,y}_{t+s} - X^{\varepsilon,y}_t\| \geq R\right) \leq 2d\exp\left(-\frac{(R-Ch)^2}{2dh\varepsilon\Sigma^2}\right),$$

where Σ is a bound for σ. Writing this for $t = ih$ for $0 \leq i < T/h$, we easily deduce that, for $R \geq Ch$,

$$\mathbf{P}\left(\omega(X^\varepsilon,h) \geq 2R\right) \leq 2d\left(\frac{T}{h}+1\right)\exp\left(-\frac{(R-Ch)^2}{2d\varepsilon\Sigma^2 h}\right), \tag{65}$$

where $\omega(\psi,h)$ has been defined in the statement of Lemma 4.4.

For any $l \geq 1$, set $R_l = 1/2l$ and $h_l = 1/l^3$. Then, for sufficiently large l, $R_l \geq Ch_l$ and

$$\frac{(R_l - Ch_l)^2}{2d\varepsilon\Sigma^2 h_l} = \frac{(\sqrt{l} - 2C/l^{3/2})^2}{8d\varepsilon\Sigma^2} \geq \frac{(\sqrt{l}/2)^2}{8d\varepsilon\Sigma^2} = \frac{l}{32d\varepsilon\Sigma^2}. \tag{66}$$

Now, observe that the set K^y_k, defined in (51) in the statement of Lemma 4.4 satisfies

$$K^y_k = \{\psi \in \mathcal{C}_y([0,T],\mathcal{X}) : \forall l \geq k, \omega(\psi,h_l) \leq 2R_l\}.$$

This is a compact set by Ascoli's Theorem, and a simple computation, using (65) and (66) shows that $\varepsilon \ln \mathbf{P}(X^{\varepsilon,y} \notin K^y_k) \leq -k/64d\Sigma^2$ for sufficiently large k. □

5 Application to the problem of exit from a domain

We propose to study in this section the phenomenon of punctualism. We consider a bounded open subset G of $\mathcal{X}$ containing a unique, stable equilibrium of the canonical equation of adaptive dynamics $\dot{\phi} = b(\phi)$ (Dieckmann and Law [18], Champagnat *et al.* [10]). Such an equilibrium is a point of Γ. As we saw in Corollary 4.2, when ε is small, $X^{\varepsilon,x}$ is close to the solution of this ODE with initial state x with high probability. Yet, the diffusion phenomenon may almost surely drive $X^{\varepsilon,x}$ out of G in long time. Our main result gives estimates for the time of exit of X^ε from G, and determines where the exit is more likely to occurs in ∂G.

We will assume $d \geq 2$. Otherwise, the problem has few interest: the process $X^{\varepsilon,y}$ can exit from an open domain $G = (c,c')$ of $\mathcal{X}$ containing a unique point x of Γ, with $y > x$, only from the right side c' of G, and the probability of reaching x before c' can be computed explicitly using Proposition 5.5.22 of Karatzas and Shreve [45].

In this section, we will consider the process X^ε constructed on the filtered space $(\Omega,\mathcal{F},\mathbf{Q}_x)$ of section 3.2 for which part (b) of Theorem 3.2 holds, and we will use the notations $\mathbf{P}_x$ and $\mathbf{E}_x$ used therein.

We will follow closely the treatment of section 5.7 of Dembo and Zeitouni [13]. Fix an open bounded domain $G \subset \mathcal{X}$, and suppose that the boundary of G is smooth enough for

$$\tau^\varepsilon = \inf\{t > 0 : X^\varepsilon_t \in \partial G\}$$

to be a well-defined stopping time.

Define

$$V(y,z,t) = \inf_{\{\psi \in \mathcal{C}([0,t],\overline{\mathcal{X}}) : \psi(0)=y, \psi(t)=z\}} I_{t,y}(\psi),$$

which is, heuristically, the cost of forcing $X^{\varepsilon,y}$ to be at z at time t. Define also

$$V(y,z) = \inf_{t>0} V(y,z,t).$$

Observe that the *quasi-potential* of Freidlin and Wentzell [34] $V(0,z)$ has no interest in our setting, since for all $t > 0$, $V(0,z,t) = \infty$ if $z \neq 0$ ($X^{\varepsilon,0}$ is constant, equal to 0). Instead, let us define

$$\bar{V}(0,z) := \lim_{\rho \to 0} \inf_{y \in S(\rho)} V(y,z),$$

where $S(\rho) = \{y \in \mathbb{R}^d : \|y\| = \rho\}$. Define also the sphere $B(\rho) = \{y \in \mathbb{R}^d : \|y\| \leq \rho\}$. Note that the limit above exists since it is an increasing limit: if $\rho < \rho'$ and if $z \notin B(\rho')$, one can obtain from a path from $S(\rho)$ to z a new path from $S(\rho')$ to z by "cutting" the beginning of the former path.

Six assumptions are required for our result:

Hypotheses

(Ha) *G is a bounded open subset of $\mathcal{X}$ such that $\overline{G} \subset \partial\mathcal{X}$. The unique stable equilibrium point in G of the d-dimensional ordinary differential equation*

$$\dot{\phi} = b(\phi) \tag{67}$$

is at $0 \in G$, and $\phi(0) \in G \Rightarrow \forall t > 0, \phi(t) \in G$ and $\lim_{t\to\infty} \phi(t) = 0$.

(Hb) $\bar{V} := \inf_{z\in\partial G} \bar{V}(0,z) < \infty$.

(Hc) *For any* $\varepsilon > 0$ *and* $y \in G \setminus \{0\}$, $\mathbf{P}_y\left(\lim_{t\to\infty} X_t^\varepsilon = 0\right) = 0$.

(Hd) *The points* $x \in \mathcal{X}$ *such that* $\partial_1 g(x,x) = 0$ *are isolated points of* $\mathbb{R}^d$.

(He) *For any* $y \in \overline{G} \cap \Gamma$, g *is* $\mathcal{C}^2$ *at* (y,y) *and* $H_{1,1}g(y,y) + H_{1,2}g(y,y)$ *has a null kernel.*

(Hf) *All the trajectories of the deterministic system (67) with initial value in* ∂G *converges to* 0 *as* $t \to \infty$.

(Ha) states that the domain G is an *attracting* domain. Assumption (Hb) is natural, since otherwise all points on ∂G are equally unlikely on the large deviations scale. We have given in Theorem 3.5 (sections 3.4) conditions under which (Hc) holds, and, by Theorem 3.2 (b), assumption (Hc) implies that X^ε is strong Markov as long as it stays inside G. Assumption (Hd) is required in the large deviation Theorem 4.1. We have already encountered an assumption similar to (He) in Proposition 3.1 (section 3.4) and Proposition 4.1 (section 4.1). It allows to bound below the eigenvalues of $a(x)$ for x near $y \in \overline{G}\cap\Gamma$. Finally, assumption (Hf) is an auxiliary assumption that prevents consideration of situations in which ∂G is the charateristic boundary of the domain of attraction of 0.

Observe that when (Hf) is true, $\partial G \cap \Gamma = \emptyset$ (since any equilibrium of the canonical equation (67) belong to Γ). Hence, in this case, (He) reduces to the assumption that D_0 has a null kernel.

The treatment to follow is guided by the heuristics that, as $\varepsilon \to 0$, X^ε wanders around 0 for an exponentially long time, during which its chance of hitting a closed set $N \subset \partial G$ is determined by $\inf_{z\in N} \bar{V}(0,z)$. Any excursion off the stable point 0 has an overwhelmingly high probability of being pulled back near 0, and it is not the time spent near any part of ∂G that matters but the *a priori* chance for a direct, fast exit due to a rare segment in the Brownian motion's path. Guided by this heuristics, we can prove

Theorem 5.1

(a) *Assume (H1)–(H4) and (Ha)–(He). Then, for all* $x \in G \setminus \{0\}$ *and* $\delta > 0$,

$$\lim_{\varepsilon\to 0} \mathbf{P}_x(\tau^\varepsilon > e^{(\bar{V}-\delta)/\varepsilon}) = 1. \tag{68}$$

(b) *Assume (H1)–(H4) and (Ha)–(Hf). If* N *is a closed subset of* ∂G *and if* $\inf_{z\in N} \bar{V}(0,z) > \bar{V}$, *then for any* $x \in G \setminus \{0\}$,

$$\lim_{\varepsilon\to 0} \mathbf{P}_x(X^\varepsilon_{\tau^\varepsilon} \in N) = 0. \tag{69}$$

In particular, if there exists $z^* \in \partial G$ *such that* $\bar{V}(0,z^*) < \bar{V}(0,z)$ *for all* $z \in \partial G \setminus \{z^*\}$, *then*

$$\forall \delta > 0, \forall x \in G \setminus \{0\}, \quad \lim_{\varepsilon \to 0} \mathbf{P}_x(\|X^\varepsilon_{\tau^\varepsilon} - z^*\| < \delta) = 1. \tag{70}$$

Note that assumption (Hf) is not necessary to prove (a).

Remark 5.1 *Part (a) of this kind of results usually includes an upper bound for* τ^ε. *We are not able to achieve this because of the singularity at* 0 *of the process* X^ε*: we are only able to obtain an uniform exponential lower bound on* $\mathbf{P}_x(\tau^\varepsilon \leq T)$ *for* $x \in G \setminus B(\rho)$, *and not for any* $x \in G \setminus \{0\}$. *This uniform lower estimate is crucial in the classical proofs of the fact that* $\mathbf{P}_x(\tau^\varepsilon < e^{(\bar{V}+\delta)/\varepsilon}) \to 1$.

The proof of a similar result in Dembo and Zeitouni [13] (Theorem 5.7.11 and Corollary 5.7.16) is based on the strong Markov property for X^ε, which holds in our case for any stopping time smaller than τ^ε thanks to assumption (Hc) and Theorem 3.2 (b), and on several lemmas, which have to be adapted to our degenerate case. Some of them will be very close to the lemmas of [13], and some of them will require a different treatment. In particular, part (a) of Theorem 5.1 will be obtained in a very similar way than in [13], whereas part (b) has to be obtained without using any upper bound on τ^ε.

The first lemma implies the continuity of $V(x, \cdot ,t)$ on $\overline{G}$.

Lemma 5.1 *Assume (H1)–(H4), (Hd) and (He). For any* $\delta > 0$, *there exists* $\rho > 0$ *small enough such that*

$$\sup_{x \in B(\rho)\setminus\{0\},\ y \in B(\rho)} \inf_{t \in [0,1]} V(x,y,t) < \delta \tag{71}$$

and

$$\sup_{\{x \in \mathcal{X}\setminus\Gamma,\ y \in \mathcal{X} : \inf_{z \in \partial G}(\|y-z\|+\|x-z\|) \leq \rho\}} \inf_{t \in [0,1]} V(x,y,t) < \delta. \tag{72}$$

The second lemma states that the diffusion wanders in G for an arbitrary long time without hitting a small neighborhood of 0 with an exponentially negligible probability. Assumption (Hf) is necessary to prove this lemma.

Lemma 5.2 *Assume (H1)–(H4) and (Ha)–(Hf). Let*

$$\sigma_\rho := \inf\{t \geq 0 : X^\varepsilon \in B(\rho) \cup \partial G\},$$

for ρ *small enough to have* $B(\rho) \subset G$ *(mind that* σ_ρ *depends on* ε*; we do not mention it to keep notations simple). Then*

$$\lim_{t \to \infty} \limsup_{\varepsilon \to 0} \varepsilon \ln \sup_{x \in G} \mathbf{P}_x(\sigma_\rho > t) = -\infty.$$

The third lemma gives a uniform lower bound on the probability of an exit from G starting from a small sphere around 0 before hitting an ever smaller sphere.

Lemma 5.3 *Assume (H1)–(H4) and (Ha)–(He). Then*

$$\lim_{\rho\to 0}\liminf_{\varepsilon\to 0}\varepsilon\ln\inf_{y\in S(2\rho)}\mathbf{P}_y(X^\varepsilon_{\sigma_\rho}\in\partial G)\geq -\bar{V}=-\inf_{z\in\partial G}\bar{V}(0,z).$$

The following upper bound relates our quasi-potential $\bar{V}(0,\cdot)$ with the probability that an excursion starting from a small sphere around 0 hits a given subset of ∂G before hitting an even smaller sphere.

Lemma 5.4 *Assume (H1)–(H4) and (Ha)–(Hf). For any closed set* $N\subset\partial G$*,*

$$\lim_{\rho\to 0}\limsup_{\varepsilon\to 0}\varepsilon\ln\sup_{y\in S(2\rho)}\mathbf{P}_y(X^\varepsilon_{\sigma_\rho}\in N)\leq -\inf_{z\in N}\bar{V}(0,z)$$

The following lemma is used to extend the upper bound to hold for every initial point $x\in G$.

Lemma 5.5 *Assume (H1)–(H4) and (Ha). Then, for every* $\rho>0$ *satisfying* $B(\rho)\subset G$ *and all* $x\in G$*,*

$$\lim_{\varepsilon\to 0}\mathbf{P}_x(X^\varepsilon_{\sigma_\rho}\in B(\rho))=1.$$

Finally, we need a uniform estimate stating that over short time intervals, the process X^ε with initial state x has an exponentially negligible probability of getting too far from x.

Lemma 5.6 *Assume (H1)–(H4) and (Ha)–(He). For every* $\rho>0$ *and every* $c>0$*, there exists a constant* $T(c,\rho)<\infty$ *such that*

$$\limsup_{\varepsilon\to 0}\varepsilon\ln\sup_{x\in G}\mathbf{P}_x\left(\|X^\varepsilon-x\|_{0,T(c,\rho)}\geq\rho\right)<-c.$$

Lemma 5.3 is the only one having a different statement than in [13]. We will prove Lemmas 5.2, 5.4, 5.5 and 5.6 by extending the methods of Lemmas 5.7.19, 5.7.21, 5.7.22 and 5.7.23 of [13] to our case. Lemma 5.1 has to be proved in a different way than in [13], because of the degeneracy of a at points of Γ, and the new Lemma 5.3 will allow to prove Theorem 5.1 (b), but is not sufficient to prove the usual upper bound for τ^ε.

We will first prove Theorem 5.1, and postpone the proof of all the preceding lemmas to the end of the section.

Proof of Theorem 5.1 (a) We will first prove (a) under the additional assumption (Hf). Let $\rho > 0$ be small enough for $B(2\rho) \subset G$ (ρ is to be specified later). Let $\theta_0 = 0$ and for $m = 0{,}1,\ldots$ define the stopping times

$$\begin{aligned}\tau_m &= \inf\{t \geq \theta_m : X_t^\varepsilon \in B(\rho) \cup \partial G\},\\ \theta_{m+1} &= \inf\{t > \tau_m : X_t^\varepsilon \in S(2\rho)\},\end{aligned}$$

with the convention that $\theta_{m+1} = \infty$ if $X_{\tau_m}^\varepsilon \in \partial G$. Each interval $[\tau_m,\tau_{m+1}]$ represents one significant excursion off $B(\rho)$. Note that, necessarily, $\tau^\varepsilon = \tau_m$ for some integer m.

Moreover, assumption (Hc) implies that $\theta_{m+1} < \infty$ as soon as $X_{\tau_m}^\varepsilon \in B(\rho)$. This can be proved as follows. Since for any $\alpha > 0$, X^ε is a diffusion with bounded drift part and uniformly non-degenerate diffusion part in $B(2\rho) \cap \Gamma_{\alpha/2}$, X^ε has an uniformly positive probability to reach $S(2\rho)$ before $S(\alpha/2)$ starting from any point of $S(\alpha)$. Hence, by the strong Markov property, for all $x \in S(\rho)$, $\mathbf{P}_x(\theta_1 < \infty | \limsup_{t\to+\infty} \|X_t^\varepsilon\| \geq \alpha) = 1$. Since assumption (Hc) implies that for all $x \in S(\rho)$, $\mathbf{P}_x(\limsup_{t\to+\infty} \|X_t^\varepsilon\| \geq \alpha) \to 1$ when $\alpha \to 0$, we actually have, for all $x \in S(\rho)$, $\mathbf{P}_x(\theta_1 < \infty) = 1$. By the strong Markov property, this implies that, almost surely, $X_{\tau_m}^\varepsilon \in B(\rho) \Rightarrow \theta_{m+1} < \infty$.

For $\bar{V} = 0$, the lower bound on τ^ε is an easy consequence of Lemmas 5.5 and 5.6. Hence, assume hereafter that $\bar{V} > 0$, and fix $\delta > 0$ arbitrarily small. Note that ∂G is a closed set and choose $\rho > 0$ small enough as needed by Lemma 5.4 for

$$\limsup_{\varepsilon\to 0} \varepsilon \ln \sup_{y\in S(2\rho)} \mathbf{P}_y(X_{\sigma_\rho}^\varepsilon \in \partial G) < -\bar{V} + \frac{\delta}{2}$$

to hold. Now, let $c = \bar{V}$ and let $T_0 = T(c,\rho)$ be as determined by Lemma 5.6. Then, there exists $\varepsilon_0 > 0$ such that for all $\varepsilon \leq \varepsilon_0$ and all $m \geq 1$,

$$\sup_{x\in G\setminus\{0\}} \mathbf{P}_x(\tau^\varepsilon = \tau_m) \leq \sup_{y\in S(2\rho)} \mathbf{P}_y(X_{\sigma_\rho}^\varepsilon \in \partial G) \leq e^{-(\bar{V}-\delta/2)/\varepsilon}$$

and

$$\sup_{x\in G\setminus\{0\}} \mathbf{P}_x(\theta_m - \tau_{m-1} \leq T_0) \leq \sup_{x\in G} \mathbf{P}_x(\|X^\varepsilon - x\|_{0,T_0} \geq \rho) \leq e^{-(\bar{V}-\delta/2)/\varepsilon}.$$

The event $\{\tau^\varepsilon \leq kT_0\}$ implies that either one of the mutually exclusive events $\{\tau^\varepsilon = \tau_m\}$ for $0 \leq n \leq k$ occurs, or else that at least one of the first k excursions $[\tau_m,\tau_{m+1}]$ off $B(\rho)$ is of length at most T_0. Thus, utilizing the preceding worst-case estimates, for all $x \in G \setminus \{0\}$ and any integer k,

$$\begin{aligned}\mathbf{P}_x(\tau^\varepsilon \leq kT_0) &\leq \sum_{m=0}^{k} \mathbf{P}_x(\tau^\varepsilon = \tau_m) + \mathbf{P}_x(\min_{1\leq m\leq k}\{\theta_m - \tau_{m-1}\} \leq T_0)\\ &\leq \mathbf{P}_x(\tau^\varepsilon = \tau_0) + 2ke^{-(\bar{V}-\delta/2)/\varepsilon}.\end{aligned}$$

Recall the identity $\{\tau^\varepsilon = \tau_0\} \equiv \{X^\varepsilon_{\sigma_\rho} \notin B(\rho)\}$ and apply the preceding inequality with $k = [T_0^{-1} e^{(\bar{V}-\delta)/\varepsilon}] + 1$ to obtain, for small enough ε,

$$\mathbf{P}_x(\tau^\varepsilon \le e^{(\bar{V}-\delta)/\varepsilon}) \le \mathbf{P}_x(\tau^\varepsilon \le kT_0) \le \mathbf{P}_x(X^\varepsilon_{\sigma_\rho} \notin B(\rho)) + 4T_0^{-1} e^{-\delta/2\varepsilon}.$$

By Lemma 5.5, the left side of this inequality converges to 0 as $\varepsilon \to 0$; hence, the proof of (68) is complete.

It remains to study the case where assumption (Hf) is removed. In this case, let $G^{-\rho} := \{x \in G : d(x,\partial G) > \rho\}$. Observe that $G^{-\rho}$ are open sets for which assumption (Hd) holds. Therefore, (a) is true for these sets. Moreover, the stopping times τ^ε related to $G^{-\rho}$ are increasing when ρ decreases to 0. The announced lower bound on τ^ε results easily from this fact and from the continuity of the quasi-potential at any point of ∂G, implied by (72). □

Proof of Theorem 5.1 (b) Fix a closed set $N \subset G$ such that $\bar{V}_N := \inf_{z \in N} \bar{V}(0,z) > \bar{V}$ (if $\bar{V}_N = \infty$,, then simply use throughout the proof an arbitrary large constant as $\bar{V}_N$). Fix $\eta > 0$ such that $\eta < (\bar{V}_N - \bar{V})/3$, and set $\rho > 0$ and $\varepsilon_0 > 0$ as needed in Lemmas 5.3 and 5.4 for

$$\inf_{y \in S(2\rho)} \mathbf{P}_y(X^\varepsilon_{\sigma_\rho} \in \partial G) \ge e^{-(\bar{V}+\eta)/\varepsilon}, \quad \forall \varepsilon \le \varepsilon_0 \tag{73}$$

and

$$\sup_{y \in S(2\rho)} \mathbf{P}_y(X^\varepsilon_{\sigma_\rho} \in N) \le e^{-(\bar{V}_N-\eta)/\varepsilon}, \quad \forall \varepsilon \le \varepsilon_0$$

to hold. Fix $y \in B(\rho)$ and remind the definition of the stopping times τ_m and θ_m in the proof of Theorem 5.1 (a). Observe that

$$\mathbf{P}_y(X^\varepsilon_{\tau^\varepsilon} \in N) \le \mathbf{P}_y(\tau^\varepsilon > \tau_l) + \sum_{m=1}^{l} \mathbf{P}_y(\tau^\varepsilon = \tau_m \text{ and } X^\varepsilon_{\tau^\varepsilon} \in N). \tag{74}$$

Firstly, let us bound the second term: for $m \ge 1$, $y \in B(\rho)$ and $\varepsilon \le \varepsilon_0$, it follows from the strong Markov property that

$$\begin{aligned}\mathbf{P}_y(\tau^\varepsilon = \tau_m \text{ and } X^\varepsilon_{\tau^\varepsilon} \in N) &= \mathbf{P}_y(\tau^\varepsilon > \tau_{m-1}) \mathbf{P}_y(X^\varepsilon_{\tau_m} \in N | \tau^\varepsilon > \tau_{m-1}) \\ &= \mathbf{P}_y(\tau^\varepsilon > \tau_{m-1}) \mathbf{E}_y[\mathbf{P}_{X^\varepsilon_{\theta_m}}(X^\varepsilon_{\sigma_\rho} \in N) | \tau^\varepsilon > \tau_{m-1}] \\ &\le \sup_{x \in S(2\rho)} \mathbf{P}_x(X^\varepsilon_{\sigma_\rho} \in N) \le e^{-(\bar{V}_N-\eta)/\varepsilon}.\end{aligned}$$

Secondly, let us bound the first term of the right-hand side of (74): for $l \ge 1$ and $y \in B(\rho)$,

$$\mathbf{P}_y(\tau^\varepsilon > \tau_l) = \mathbf{E}_y[\mathbf{P}_{X^\varepsilon_{\theta_1}}(\tau^\varepsilon > \tau_{l-1})] \le \sup_{x \in S(2\rho)} \mathbf{P}_x(\tau^\varepsilon > \tau_{l-1}). \tag{75}$$

Now, for $x \in S(2\rho)$ and $k \geq 1$,

$$\begin{aligned}\mathbf{P}_x(\tau^\varepsilon > \tau_k) &= [1 - \mathbf{P}_x(\tau^\varepsilon = \tau_k | \tau^\varepsilon > \tau_{k-1})]\mathbf{P}_x(\tau^\varepsilon > \tau_{k-1}) \\ &= [1 - \mathbf{E}_x[\mathbf{P}_{X^\varepsilon_{\theta_k}}(X^\varepsilon_{\sigma_\rho} \in \partial G) | \tau^\varepsilon > \tau_{k-1}]]\mathbf{P}_x(\tau^\varepsilon > \tau_{k-1}) \\ &\leq (1-q)\mathbf{P}_x(\tau^\varepsilon > \tau_{k-1}),\end{aligned}$$

where $q := \inf_{y \in S(2\rho)} \mathbf{P}_y(X^\varepsilon_{\sigma_\rho} \in \partial G) \geq e^{-(\bar{V}+\eta)/\varepsilon}$ by 73. Iterating over $k = 1,2,\ldots$ gives for $k \geq 0$

$$\sup_{y \in S(2\rho)} \mathbf{P}_x(\tau^\varepsilon > \tau_k) \leq (1-q)^k.$$

In (75), this yields for all $l \geq 1$ and $y \in B(\rho)$

$$\mathbf{P}_y(\tau^\varepsilon > \tau_l) \leq (1-q)^{l-1}.$$

Putting together these estimates in (74) gives finally for all $y \in B(\rho)$ and $\varepsilon \leq \varepsilon_0$

$$\mathbf{P}_y(X^\varepsilon_{\tau^\varepsilon} \in N) \leq \left(1 - e^{-\frac{\bar{V}+\eta}{\varepsilon}}\right)^{l-1} + l e^{-\frac{\bar{V}_N - \eta}{\varepsilon}}.$$

Choosing $l = [2e^{(\bar{V}+2\eta)/\varepsilon}]$, for ε small enough, $l - 1 > e^{(\bar{V}+2\eta)/\varepsilon}$, and hence

$$\mathbf{P}_y(X^\varepsilon_{\tau^\varepsilon} \in N) \leq \left[\left(1 - \frac{1}{u_\varepsilon}\right)^{u_\varepsilon}\right]^{e^{\eta/\varepsilon}} + 2e^{\frac{\bar{V} - \bar{V}_N + 3\eta}{\varepsilon}},$$

where $u_\varepsilon := e^{(\bar{V}+\eta)/\varepsilon} \to +\infty$. So $(1 - 1/u_\varepsilon)^{u_\varepsilon} \to 1/e$, and, finally, $\mathbf{P}_y(X^\varepsilon_{\tau^\varepsilon} \in N) \to 0$ for $y \in B(\rho)$ (recall that $0 < \eta < (\bar{V}_N - \bar{V})/3$). The proof of (69) is now completed by combining Lemma 5.5 and the inequality

$$\mathbf{P}_x(X^\varepsilon_{\tau^\varepsilon} \in N) \leq \mathbf{P}_x(X^\varepsilon_{\sigma_\rho} \notin B(\rho)) + \sup_{y \in B(\rho)} \mathbf{P}_y(X^\varepsilon_{\tau^\varepsilon} \in N).$$

Applying (69) to $N = \{z \in \partial G : \|z - z^*\| \geq \delta\}$ and observing that Lemma 5.1 (72) implies the continuity of $z \mapsto \bar{V}(0,z)$ on ∂G we easily obtain (70). $\square$

Proof of Lemma 5.1 (71) Fix $\delta > 0$, let ρ be small enough for $B(\rho) \subset G$ to hold, and fix x in $B(\rho) \setminus \{0\}$ and y in $B(\rho)$. In order to simplify the notations, we will use the complex notation for the coordinates of the points of the plane of $\mathbb{R}^d$ containing 0, x and y, and we will assume that $x = r \in \mathbb{R}$ and $y = r'e^{i\theta}$, with $0 < r \leq \rho$ and $0 \leq r' \leq \rho$. Define $\psi \in \mathcal{C}([0,1],B(\rho))$ by

$$\psi(t) = \begin{cases} (1-(3t)^2)r + (3t)^2\rho & \text{if } 0 \leq t \leq 1/3 \\ \rho e^{i\theta(3t-1)} & \text{if } 1/3 \leq t \leq 2/3 \\ (1-(3-3t)^2)r'e^{i\theta} + (3-3t)^2\rho e^{i\theta} & \text{if } 2/3 \leq t \leq 1. \end{cases}$$

Then $\psi(0) = x$ and $\psi(1) = y$, and $\psi(t) \in B(\rho) \setminus \{0\}$ for any $t \in [0,1)$. Moreover, for $0 \leq t \leq 1/3$, $\psi(t) = r + 9t^2(\rho - r)$, so that $\|\psi(t)\| \geq 9t^2(\rho - r)$, and, similarly, for $2/3 \leq t \leq 1$, $\|\psi(t)\| \geq 9(1-t)^2(\rho - r')$. Thanks to assumption (He), a calculation similar to equation (47) in the proof of Proposition 4.1 gives, with the same K, $\mathcal{N}_0$ and a_0 as therein, if $B(\rho) \subset \mathcal{N}_0$,

$$\begin{aligned}
I_{1,x}(\psi) \leq \frac{1}{2a_0}\Bigg(&\int_0^{1/3} \frac{2(18t(\rho-r))^2 + 2K^2\|\psi(t)\|^2}{\|\psi(t)\|}dt \\
&+ \int_{1/3}^{2/3} \frac{2(3\theta\rho)^2 + 2K^2\|\psi(t)\|^2}{\|\psi(t)\|} \\
&+ \int_{2/3}^{1} \frac{2(18(1-t)(\rho-r'))^2 + 2K^2\|\psi(t)\|^2}{\|\psi(t)\|}dt\Bigg) \\
\leq \frac{1}{2a_0}\Bigg(&\int_0^{1/3}(648(\rho-r) + 2K^2\|\psi(t)\|)dt + \int_{1/3}^{2/3}(18\theta^2 + 2K^2)\rho dt \\
&+ \int_{2/3}^{1}(648(\rho-r') + 2K^2\|\psi(t)\|)dt\Bigg) \\
\leq \frac{(216 + 2K^2/3)\rho + (6\theta^2 + 2K^2/3)\rho + (216 + 2K^2/3)\rho}{2a_0}&.
\end{aligned}$$

Consequently, for sufficiently small $\rho > 0$ not depending on x and y in $B(\rho) \setminus \{0\}$, $I_{1,x}(\psi) \leq \delta/2$, which yields (71). □

Proof of Lemma 5.1 (72) Fix $\delta > 0$. Thanks to assumption (He) and since $\partial G \cap \partial\mathcal{X} = \emptyset$, using the same method as in the proof of (71), one can find for any $z \in \partial G \cap \Gamma$ a positive ρ_z such that

$$\sup_{x \in B(z,\rho_z)\setminus\{0\},\ y \in B(z,\rho_z)} \ \inf_{t \in [0,1]} V(x,y,t) < \delta/2, \tag{76}$$

where $B(z,\rho)$ is the ball of radius ρ cenetered at z.

Let $\bar\rho_0$ be the infimum of the ρ_z for $z \in \partial G \cap \Gamma$. Since G is bounded and $\partial G \cap \partial\mathcal{X}$, because of assumption (Hd), this set is finite, and so $\bar\rho_0 > 0$. By reducing $\bar\rho_0$ if necessary, we can assume that $B(\bar\rho_0) \subset G$ and that $d(\Gamma \setminus \overline{G}, \overline{G}) > \bar\rho_0$ (this distance is positive because of assumption (Hd) and because G is bounded).

Fix x and y in $\mathcal{X} \setminus \bigcup_{z \in \partial G \cap \Gamma} B(z,\bar\rho_0)$ and assume that there exists $z \in \partial G$ with $\|x - z\| + \|y - z\| \leq \bar\rho_0/3$. Then $d(x,\Gamma) > 2\bar\rho_0/3$, $d(y,\Gamma) > 2\bar\rho_0/3$, and, since $\|x - y\| \leq \bar\rho_0/3$, the segment $[x,y]$ is included in $\Gamma_{\bar\rho_0/3}$.

For any $t_0 > 0$, define $\psi_{t_0} \in \mathcal{C}([0,t_0],\Gamma_{\bar\rho_0/3})$ by

$$\psi_{t_0}(t) = \left(1 - \frac{t}{t_0}\right) x + \frac{t}{t_0} y$$

for $0 \le t \le t_0$. Then $\psi_{t_0}(0) = x$ and $\psi_{t_0}(t_0) = y$.

Since a is uniformly non-degenerate on $\Gamma_{\bar\rho_0/3}$, there exists a constant C bounding a^{-1} on this set. Then

$$\begin{aligned} I_{t_0,x}(\psi_{t_0}) &\le \frac{C}{2} \int_0^{t_0} (2\|\dot\psi(t)\|^2 + 2\|b(\psi(t))\|^2) dt \\ &\le C \int_0^{t_0} \left(\frac{\|x-y\|^2}{t_0^2} + B^2 \right) dt \\ &\le C \left(\frac{\|x-y\|^2}{t_0} + B^2 t_0 \right), \end{aligned}$$

where B is a bound for b on $\overline{\mathcal{X}}$. The infimum of the right-hand side is obtained for $t_0 = \|x-y\|/B$, and gives

$$I_{\rho/B,x}(\psi_{\rho/B}) \le 2BC\|x-y\|.$$

Therefore, there exists $\bar\rho_1 > 0$ such that $\inf_{t\in[0,1]} V(x,y,t) < \delta/2$ for any x and y in $\mathcal{X} \setminus \cup_{z\in\partial G\cap\Gamma} B(z,\bar\rho_0)$ such that $\|x-y\| \le \bar\rho_1$ and such that there exists $z \in \partial G$ with $\|x-z\| + \|y-z\| \le \bar\rho_0/3$. In view of (76), $\rho = \bar\rho_0 \wedge (\bar\rho_0/3)$ is an appropriate constant in (72). □

Proof of Lemma 5.2 If $x \in B(\rho)$, then $\sigma_\rho = 0$ and the lemma trivially holds. Otherwise, consider the closed sets

$$C_t = \mathcal{C}([0,t],\overline{G \setminus B(\rho)}),$$

and observe that, for $x \in G$, the event $\{\sigma_\rho > t\}$ is contained in $\{X^\varepsilon \in C_t\}$. Corollary 4.1 yields, for all $t > 0$,

$$\limsup_{\varepsilon\to 0} \varepsilon \ln \sup_{x\in\overline{G\setminus B(\rho)}} \mathbf{P}_x(X^\varepsilon \in C_t) \le - \inf_{\psi\in C_t} I_t(\psi),$$

where, throughout this proof, $I_t(\psi)$ stands for $I_{t,\psi(0)}(\psi)$. Hence, in order to complete the proof of the lemma, it suffices to show that

$$\lim_{t\to\infty} \inf_{\psi\in C_t} I_t(\psi) = \infty. \tag{77}$$

Let ψ^x denote the trajectory of (67) starting at $x \in \overline{G \setminus B(\rho)}$. By assumption (Ha) and (Hf), ψ^x hits $S(\rho/3)$ in a finite time, denoted T_x. Moreover, by the uniform Lipschitz continuity of b and Gronwall's lemma, there exists an open neighborhood $\mathcal{W}_x$ of x such that, for all $y \in \mathcal{W}_x$, the path ψ^y hits $S(2\rho/3)$ before T_x. Extracting a finite cover of the compact set $\overline{G \setminus B(\rho)}$ by such sets, it follows that there exists $T < \infty$ such that for all $y \in \overline{G \setminus B(\rho)}$, ψ^y hits $S(2\rho/3)$ before time T.

Assume now that (77) does not hold true. Then , for some $M < \infty$ and every integer n, there exists $\psi^n \in C_{nT}$ such that $I_{nT}(\psi^n) \leq M$. Consequently, for some $\psi^{n,k} \in C_T$,

$$M \geq I_{nT}(\psi^n) = \sum_{k=1}^{n} I_T(\psi^{n,k}) \geq n \min_{1 \leq k \leq n} I_T(\psi^{n,k}).$$

Hence, there exists a sequence $\phi^n \in C_T$ with $\lim I_T(\phi^n) = 0$. It follows from the fact that a is uniformly non-degenerate and uniformly Lipschitz on $\overline{G \setminus B(\rho)}$ and from Lemma 4.1, that the set $\{\phi \in C_T : I_T(\phi) \leq 1\}$ is compact, so the sequence ϕ^n has a limit point ϕ^* in C_T. It follows also that I_T is lower semicontinuous on C_T, and therefore, $I_T(\phi^*) = 0$. Consequently, ϕ^* is a trajectory of (67) staying inside $\overline{G \setminus B(\rho)}$ on $[0,T]$, which yields a contradiction with the definition of T. $\square$

Proof of Lemma 5.3 Fix $\eta > 0$ and let $\rho > 0$ be small enough for $B(2\rho) \subset G$ and for Lemma 5.1 to hold with $\delta = \eta/3$ and 2ρ instead of ρ. Note that the definition of $\bar{V}(0,z)$ yields the inequality $\inf_{y \in S(2\rho)} V(y,z) \leq \bar{V}(0,z)$ as soon as $z \notin B(2\rho)$. Then, by (72) and assumption (Hb), there exists $x \in S(2\rho)$, $z \notin \overline{G}$, $T_1 < \infty$ and $\psi \in \mathcal{C}([0,T_1],\mathcal{X})$ such that $\psi(0) = x$, $\psi(T_1) = z$ and $I_{T_1,x}(\psi) \leq \bar{V} + \eta/3$. Moreover, by properly "cutting" the beginning of the path ψ, we can suppose that for all $t > 0$, $\psi(t) \notin B(2\rho)$. Since $z \in \mathcal{X} \setminus \overline{G}$, the constant $\Delta := d(z,\partial G \cup \partial \mathcal{X})$ is (strictly) positive.

Thanks to (71), for any $y \in S(2\rho)$, there exists a continuous path ψ^y of length $t_y \leq 1$ such that $\psi^y(0) = y$, $\psi^y(t_y) = x$, and $I_{t_y,y}(\psi^y) \leq \eta/3$. Moreover, the construction of such a function in the proof of Lemma 5.1 allows us to assume that $\|\psi^y(t)\| = 2\rho$ for all $t \in [0,t_y]$. Let ϕ^y denote the path obtained by concatenating ψ^y and ψ (in that order) and extending the resulting function to be of length $T_0 = T_1 + 1$ by following the trajectory of (67) after reaching z. Since the latter path does not contribute to the rate function, it follows that $I_{T_0,y}(\phi^y) \leq \bar{V} + 2\eta/3$.

Consider the set

$$O := \bigcup_{y \in S(2\rho)} \left\{ \psi \in \mathcal{C}([0,T_0],\overline{\mathcal{X}}) : \|\psi - \phi^y\|_{0,T_0} \leq \frac{\Delta \wedge \rho}{2} \right\}.$$

Observe that O is an open subset of $\mathcal{C}([0,T_0],\overline{\mathcal{X}})$ that contains the functions $\{\phi^y\}_{y\in S(2\rho)}$. Therefore, by Corollary 4.1,

$$\begin{aligned}\liminf_{\varepsilon\to 0}\varepsilon\ln\inf_{y\in S(2\rho)}\mathbf{P}_y(X^\varepsilon\in O)&\geq-\sup_{y\in S(2\rho)}\inf_{\psi\in O}I_{T_0,y}(\psi)\\&\geq-\sup_{y\in S(2\rho)}I_{T_0,y}(\phi^y)>-(\bar V+\eta).\end{aligned}$$

If $\psi\in O$, then ψ reaches the open ball of radius $\Delta/2$ centered at z before hitting $B(\rho)$, so ψ hits ∂G before hitting $B(\rho)$. Hence, for $X^\varepsilon_0=y\in S(2\rho)$, the event $\{X^\varepsilon\in O\}$ is contained in $\{X^\varepsilon_{\sigma_\rho}\in\partial G\}$, and the proof is complete. □

Proof of Lemma 5.4 Let us first notice that, since $V(x,z)\leq V(x,y)+V(y,z)$ for all x, y and z in $\mathcal{X}$, we have

$$\forall y,z\in\mathcal{X},\quad \bar V(0,z)\leq\bar V(0,y)+V(y,z). \tag{78}$$

Fix a closed set $N\subset\partial G$, fix $\delta>0$ and define $\bar V^\delta_N:=(\inf_{z\in N}\bar V(0,z)-\delta)\wedge 1/\delta$. Then, it follows from (78) and Lemma 5.1 (71) that, for $\rho>0$ small enough,

$$\inf_{y\in S(2\rho),z\in N}V(y,z)\geq\inf_{z\in N}\bar V(0,z)-\sup_{y\in S(2\rho)}\bar V(0,y)\geq\bar V^\delta_N.$$

Moreover, by Lemma 5.2, there exists $T<\infty$ large enough for getting

$$\limsup_{\varepsilon\to 0}\varepsilon\ln\sup_{y\in S(2\rho)}\mathbf{P}_y(\sigma_\rho>T)<-\bar V^\delta_N.$$

Consider the following closed subset of $\mathcal{C}([0,T],\overline{\mathcal{X}})$:

$$C:=\{\psi\in\mathcal{C}([0,T],\overline{\mathcal{X}}):\exists t\in[0,T]\text{ such that }\psi(t)\in N\}.$$

Note that C obviously satisfies the assumptions of Corollary 4.1, and that

$$\inf_{y\in S(2\rho),\psi\in C}I_{y,T}(\psi)\geq\inf_{y\in S(2\rho),z\in N}V(y,z)\geq\bar V^\delta_N.$$

Thus

$$\limsup_{\varepsilon\to 0}\varepsilon\ln\sup_{y\in S(2\rho)}\mathbf{P}_y(X^\varepsilon\in C)\leq-\inf_{y\in S(2\rho),\psi\in C}I_{T,y}(\psi)\leq-\bar V^\delta_N.$$

Since $\mathbf{P}_y(X^\varepsilon_{\sigma_\rho}\in N)\leq\mathbf{P}_y(\sigma_\rho>T)+\mathbf{P}_y(X^\varepsilon\in C)$, it follows that

$$\limsup_{\varepsilon\to 0}\varepsilon\ln\sup_{y\in S(2\rho)}\mathbf{P}_y(X^\varepsilon_{\sigma_\rho}\in N)\leq-\bar V^\delta_N.$$

Taking $\delta\to 0$ completes the proof of the lemma. □

Proof of Lemma 5.5 Let ρ be small enough for $B(\rho) \subset G$. For $x \in B(\rho)$, there is nothing to prove. Thus, fix $x \in G \setminus B(\rho)$, let ϕ denote the trajectory of (67) with initial state x, and let $T := \inf\{t : \phi(t) \in S(\rho/2)\}$. Because of assumption (Ha), $T < \infty$ and there exists a positive distance between $\{\phi(t)\}_{t \leq T}$ and ∂G. Let $\Delta := \rho \wedge d(\{\phi(t)\}_{t\leq T}, \partial G)$, then

$$\|X^{\varepsilon,x} - \phi\|_{0,T} \leq \Delta/2 \Rightarrow X^{\varepsilon,x}_{\sigma_\rho} \in B(\rho).$$

By the uniform Lipschitz continuity of b, for $t \leq T$,

$$\|X^{\varepsilon,x}_t - \phi(t)\| \leq K \int_0^t \|X^{\varepsilon,x}_s - \phi(s)\| ds + \varepsilon BT + \sqrt{\varepsilon} \left\| \int_0^t \sigma(X^{\varepsilon,x}_s) dW_s \right\|,$$

where B is a bound for $\tilde{b}$. Hence, by Gronwall's lemma,

$$\|X^{\varepsilon,x} - \phi\|_{0,T} \leq \sqrt{\varepsilon} e^{KT} \left(\sqrt{\varepsilon} BT + \left\| \int_0^{\cdot} \sigma(X^{\varepsilon,x}_s) dW_s \right\|_{0,T} \right),$$

and, by Lemma 4.7, for sufficiently small $\varepsilon > 0$,

$$\begin{aligned} \mathbf{P}_x(X^\varepsilon_{\sigma_\rho} \in \partial G) &\leq \mathbf{P}_x(\|X^\varepsilon - \phi\|_{0,T} > \Delta/2) \\ &\leq \mathbf{P}_x \left(\left\| \int_0^{\cdot} \sigma(X^\varepsilon_s) dW_s \right\|_{0,T} > \frac{\Delta}{4\sqrt{\varepsilon}} e^{-KT} \right) \\ &\leq 2d \exp\left(-\frac{\Delta^2 e^{-2KT}}{32 d \varepsilon T \Sigma^2} \right) \to_{\varepsilon \to 0} 0, \end{aligned}$$

where Σ is a uniform bound for σ. □

Proof of Lemma 5.6 Let B be a uniform bound for b and $\tilde{b}$. For any $t \leq \rho/4B$, and for any $x \in G$, (1) yields

$$\|X^{\varepsilon,x}_t - x\| \leq \frac{\rho}{4} + \varepsilon \frac{\rho}{4} + \sqrt{\varepsilon} \left\| \int_0^t \sigma(X^{\varepsilon,x}_s) dW_s \right\|.$$

Therefore, for any $\varepsilon \leq 1$ and any $t \leq \rho/4B$, by Lemma 4.7,

$$\begin{aligned} \mathbf{P}_x(\|X^\varepsilon - x\|_{0,t} \geq \rho) &\leq \mathbf{P}_x \left(\sqrt{\varepsilon} \left\| \int_0^{\cdot} \sigma(X^\varepsilon_s) dW_s \right\|_{0,t} \geq \frac{\rho}{2} \right) \\ &\leq 2d \exp\left(-\frac{\rho^2}{8 \varepsilon d t \Sigma^2} \right), \end{aligned}$$

where Σ is a uniform bound for σ. Therefore,

$$T(c,\rho) = \frac{\rho}{4B} \wedge \frac{\rho^2}{8 d T \Sigma^2 c}$$

is an appropriate constant for Lemma 5.6. □

Annexe A

Analyse du système de compétition logistique dimorphique

Rappelons d'abord la définition de quelques notions concernant les équilibres des systèmes d'équations différentielles ordinaires (EDO) homogènes en temps. Un tel système s'écrit

$$\dot{X} = F(X) = (F^1(X), \dots ,F^d(X)), \tag{1}$$

où $F : \mathbb{R}^d \to \mathbb{R}^d$ sera supposée localement lipshitzienne. Nous allons considérer des solutions à valeurs dans un domaine D qui pourra être soit $\mathbb{R}^d$, soit $\mathbb{R}^d_+$. Ce dernier cas intervient par exemple lorsque (1) modélise la dynamique des effectifs d'une ou plusieurs populations. Pour de tels systèmes, $F^i(X) = 0$ dès que la $i^{\text{ème}}$coordonnée de X est nulle, de façon que, du fait de l'unicité des solutions, aucune solution de (1) ne puisse sortir de $\mathbb{R}^d_+$.

Un **équilibre** est une solution de $F = 0$. L'équilibre $X^* \in D$ est **asymptotiquement stable** s'il existe $\varepsilon > 0$ tel que toute solution partant de $D \cap B(X^*,\varepsilon)$ (où $B(U,\rho)$ est la boule de centre U et de rayon ρ) converge vers X^* quand $t \to +\infty$. Nous appelerons **instable** un équilibre non asymptotiquement stable. L'équilibre X^* est **répulsif** s'il existe $\varepsilon > 0$ tel que toute solution partant de $D \cap B(X^*,\varepsilon)$ sorte de cet ensemble en temps fini. Un équilibre répulsif est instable. Enfin, l'équilibre X^* est **globalement asymptotiquement stable** (**GAS**) s'il est asymptotiquement stable, et si le **domaine d'attraction** de X^* (défini comme l'ensemble des $X_0 \in D$ tels que la solution $X(t)$ de (1) satisfaisant $X(0) = X_0$ converge vers X^* quand $t \to +\infty$) est dense dans D. Notons que, lorsque F est localement lipschitzienne, du fait de la continuité du flot (cf. Queffélec and Zuily [64]

p. 356), le domaine d'attraction d'un équilibre asymptotiquement stable est un ouvert de D.

L'objet de cette annexe est de prouver un théorème examinant le comportement asymptotique du système logistique

$$\begin{cases} \dot{z}_1 = z_1(r_1 - \alpha_{11}z_1 - \alpha_{12}z_2) \\ \dot{z}_2 = z_2(r_2 - \alpha_{21}z_1 - \alpha_{22}z_2), \end{cases} \tag{2}$$

où $r_i > 0$ et $\alpha_{ij} > 0$ ($1 \leq i,j \leq 2$). Ce système possède trois équilibres triviaux $(0,0)$, $(\bar{z}_1,0)$ et $(0,\bar{z}_2)$, où

$$\bar{z}_i = \frac{r_i}{\alpha_{ii}}, \quad i = 1,2.$$

Théorème 1

(a) *Toute solution de (2) de condition initiale dans $\mathbb{R}_+^2$ converge vers une limite finie dans $\mathbb{R}_+^2$ quand $t \to +\infty$.*

(b) *Si la condition initiale d'une solution de (2) appartient à $(\mathbb{R}_+^*)^2$, alors, cette solution reste dans cet ensemble pour tout temps $t > 0$.*

(c) *L'équilibre $(\bar{z}_1,0)$ est asymptotiquement stable si et seulement si*

$$\begin{aligned} & r_2\alpha_{11} - r_1\alpha_{21} < 0, \\ \textit{ou} \quad & r_2\alpha_{11} - r_1\alpha_{21} = 0 \quad \textit{et} \quad r_1\alpha_{22} - r_2\alpha_{12} > 0. \end{aligned}$$

Il est GAS si et seulement s'il est asymptotiquement stable et si $(0,\bar{z}_2)$ ne l'est pas, c'est-à-dire si

$$\begin{aligned} & r_2\alpha_{11} - r_1\alpha_{21} < 0 \quad \textit{et} \quad r_1\alpha_{22} - r_2\alpha_{12} \geq 0, \\ \textit{ou} \quad & r_2\alpha_{11} - r_1\alpha_{21} = 0 \quad \textit{et} \quad r_1\alpha_{22} - r_2\alpha_{12} > 0, \end{aligned} \tag{3}$$

et son domaine d'attraction est $\mathbb{R}_+ \times \mathbb{R}_+^*$.

(d) *Le système (2) possède un unique équilibre non trivial dans $(\mathbb{R}_+^*)^2$ si et seulement si*

$$(r_2\alpha_{11} - r_1\alpha_{21})(r_1\alpha_{22} - r_2\alpha_{12}) > 0. \tag{4}$$

Il est instable si

$$r_2\alpha_{11} - r_1\alpha_{21} < 0 \quad \textit{et} \quad r_1\alpha_{22} - r_2\alpha_{12} < 0. \tag{5}$$

Il est globalement asymptotiquement stable si et seulement si

$$r_2\alpha_{11} - r_1\alpha_{21} > 0 \quad \textit{et} \quad r_1\alpha_{22} - r_2\alpha_{12} > 0, \tag{6}$$

et son domaine d'attraction est $(\mathbb{R}_+^*)^2$.

Preuve de (a) Découpons le quart de plan $\mathbb{R}_+^2$ selon le signe de $\dot{z}_1$ et $\dot{z}_2$: $\dot{z}_1$ est positif sous la droite $r_1 - \alpha_{11}z_1 - \alpha_{12}z_2$, et $\dot{z}_2$ est positif sous la droite $r_2 - \alpha_{21}z_1 - \alpha_{22}z_2$. Chacune de ces deux droites coupent les axes des coordonnées dans $\mathbb{R}_+ \times \{0\}$ et $\{0\} \times \mathbb{R}_+$. De plus, le système (5) possède un équilibre dans $(\mathbb{R}_+^*)^2$ si et seulement si ces deux droites s'intersectent dans $(\mathbb{R}_+^*)^2$.

Nous obtenons quatre configurations de signes possibles représentées figures A.1 (a) à (d), où les petites flèches représentent la direction de la tangente aux solutions sur les deux droite. Il est facile de se convaincre (cf. e.g. Istas [42] pages 25–27) que la figure A.1 (a) correspond au cas $r_2\alpha_{11} - r_1\alpha_{21} < 0$ et $r_1\alpha_{22} - r_2\alpha_{12} > 0$, la figure (b) au cas $r_2\alpha_{11} - r_1\alpha_{21} > 0$ et $r_1\alpha_{22} - r_2\alpha_{12} < 0$, la figure (c) au cas $r_2\alpha_{11} - r_1\alpha_{21} < 0$ et $r_1\alpha_{22} - r_2\alpha_{12} < 0$ et la figure (d) au cas $r_2\alpha_{11} - r_1\alpha_{21} > 0$ et $r_1\alpha_{22} - r_2\alpha_{12} > 0$. Les cas de nullité correspondent aux cas où les deux droites s'intersectent sur l'un des axes de coordonnées, et les deux

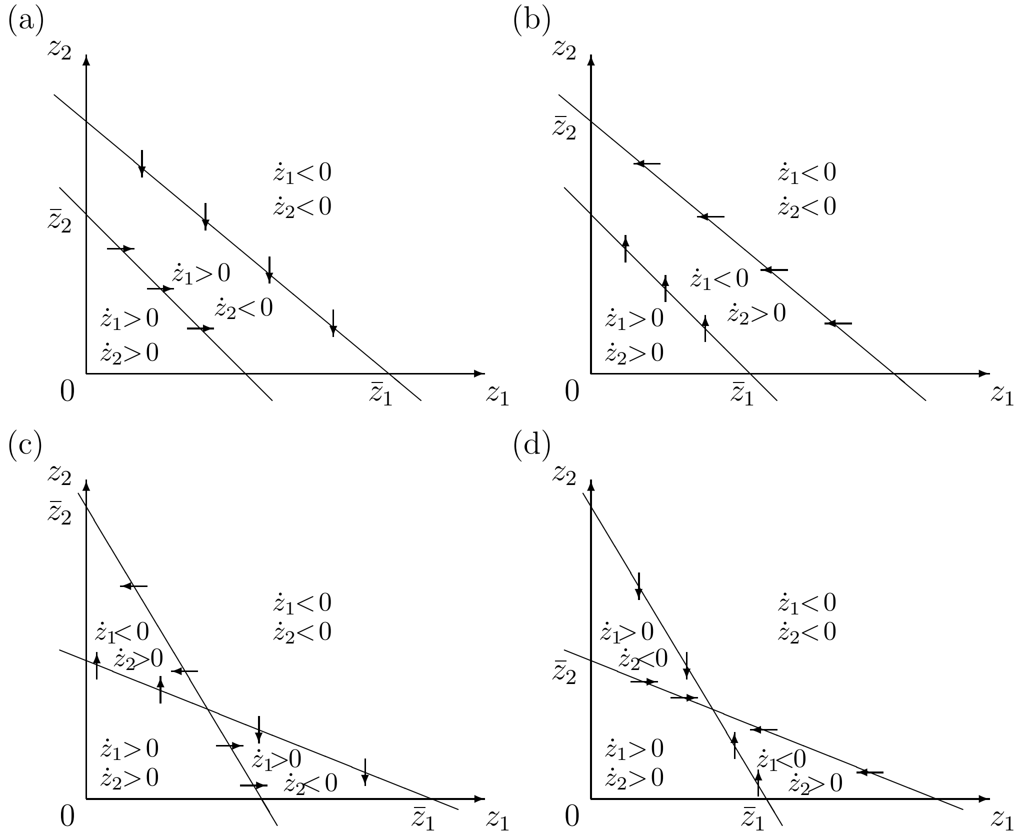


FIG. A.1 – *Configurations possibles des signes de $\dot{z}_1$ et $\dot{z}_2$ pour le système (5).*

droites sont confondues si et seulement si $r_2\alpha_{11} - r_1\alpha_{21} = r_1\alpha_{22} - r_2\alpha_{12} = 0$.

Considérons maintenant une solution quelconque de (5) dans l'un des quatre cas (a) à (d). Si cette solution part du domaine de $\mathbb{R}^2_+$ où $\dot{z}_1 \leq 0$ et $\dot{z}_2 \geq 0$, en examinant le signe des dérivées de z_1 et z_2 au bord de ce domaine, on se convainc qu'elle ne peut en sortir. Ses deux coordonnées sont donc monotones et convergent quand $t \to +\infty$ vers une limite finie appartenant à ce domaine. De même si la condition initiale appartient au domaine où $\dot{z}_1 \geq 0$ et $\dot{z}_2 \leq 0$. Dans le cas où sa condition initiale vérifie $\dot{z}_1(0) > 0$ et $\dot{z}_2(0) > 0$, la solution de (5) peut, soit rester dans ce domaine pour tout $t \geq 0$, et donc converger vers une limite finie quand $t \to +\infty$, soit quitter ce domaine au bout d'un temps fini, et atteindre l'un des deux domaines considérés plus haut. En faisant le même raisonnement dans le cas où la condition initiale vérifie $\dot{z}_1(0) < 0$ et $\dot{z}_2(0) < 0$, on obtient finalement que toute solution de (5) converge vers une limite finie, qui est nécessairement un équilibre du système. □

Preuve de (b) En particulier, toute solution $(z_1(t),z_2(t))$ de (2) est bornée. Puisqu'elle peut s'écrire

$$\begin{aligned} z_1(t) &= z_1(0)\exp\left(\int_0^t (r_1 - \alpha_{11}z_1(s) - \alpha_{12}z_2(s))ds\right) \\ z_2(t) &= z_2(0)\exp\left(\int_0^t (r_2 - \alpha_{21}z_1(s) - \alpha_{22}z_2(s))ds\right), \end{aligned}$$

z_1 et z_2 ne peuvent atteindre 0 en temps fini si $z_1(0) > 0$ et $z_2(0) > 0$. □

Preuve de (c) La matrice de stabilité de $(\bar{z}_1,0)$ s'écrit

$$\begin{pmatrix} -\alpha_{11} & -\alpha_{12}r_1/\alpha_{11} \\ 0 & r_2 - \alpha_{21}r_1/\alpha_{11} \end{pmatrix}.$$

L'examen du signe des valeurs propres de cette matrice montre que $(\bar{z}_1,0)$ est instable si $r_2\alpha_{11} - r_1\alpha_{21} > 0$, et asymptotiquement stable si $r_2\alpha_{11} - r_1\alpha_{21} < 0$.

Le cas $r_2\alpha_{11} - r_1\alpha_{21} = 0$ nécessite un développement au second ordre : réécrivons le système (6) sous la forme

$$(\dot{z}_1,\dot{z}_2) = F(z_1,z_2) = (F_1(z_1,z_2),F_2(z_1,z_2)).$$

On a alors

$$\begin{aligned} F(\bar{z}_1 + h_1,h_2) = \Big(-\frac{r_1}{\alpha_{11}}(\alpha_{11}h_1 + \alpha_{12}h_2) + o(|h_1| + |h_2|), \\ - h_2(\alpha_{21}h_1 + \alpha_{22}h_2)\Big). \quad (7) \end{aligned}$$

Remarquons de plus que $\alpha_{11}\alpha_{22} - \alpha_{12}\alpha_{21} = r_1\alpha_{22} - r_2\alpha_{12}$.

Dans le cas où $\alpha_{11}\alpha_{22} > \alpha_{12}\alpha_{21}$, prenons ε_1 et ε_2 tels que $\frac{\alpha_{12}}{\alpha_{11}}\varepsilon_2 < \varepsilon_1 < \frac{\alpha_{22}}{\alpha_{21}}\varepsilon_2$. Si ε_1 et ε_2 sont assez petits, par (7), une solution de (6) issue de $[\bar{z}_1 - \varepsilon_1, \bar{z}_1 + \varepsilon_1] \times [0, \varepsilon_2]$ reste dans ce domaine pour tout $t \geq 0$ (examiner le signe des dérivées d'une solution à la frontière de ce domaine). D'après (a), cette solution converge nécessairement vers un équilibre, qui ne peut être que $(\bar{z}_1, 0)$. Il s'agit donc, par définition, d'un équilibre asymptotiquement stable.

Supposons maintenant $\alpha_{11}\alpha_{22} = \alpha_{12}\alpha_{21}$, et définissons $\beta = r_2/r_1 = \alpha_{22}/\alpha_{12} = \alpha_{21}/\alpha_{11}$. Dans ce cas, tous les points de la droite $\alpha_{12}z_1 + \alpha_{11}z_2 = r_1$ sont des équilibres non isolés de (6), par conséquent instables.

Enfin, lorsque $\alpha_{11}\alpha_{22} < \alpha_{12}\alpha_{21}$, d'après (7), pour $\varepsilon > 0$ suffisamment petit, tout point (z_1, z_2) à l'intérieur du cône tronqué

$$D_\varepsilon := \left\{ (\bar{z}_1 - v_1, v_2) \in (\mathbb{R}_+^*)^2 : 0 < v_2 < \frac{\alpha_{21}}{\alpha_{22}} v_1 \right\} \cap B((\bar{z}_1, 0), \varepsilon),$$

(où $B(u, \rho) \subset \mathbb{R}^2$ est la boule centrée en u de rayon ρ), vérifie $F_2(z_1, z_2) > 0$.

De plus, $F_1(\bar{z}_1 - v_1, \frac{\alpha_{21}}{\alpha_{22}} v_1) < 0$ pour $v_1 > 0$ assez petit (voir figure A.2). Par conséquent, il existe $\varepsilon_0 > 0$ tel que toute solution de (2) issue de D_{ε_0} quitte cet ensemble par la partie en arc de cercle de sa frontière. Par suite, $(\bar{z}_1, 0)$ est instable.

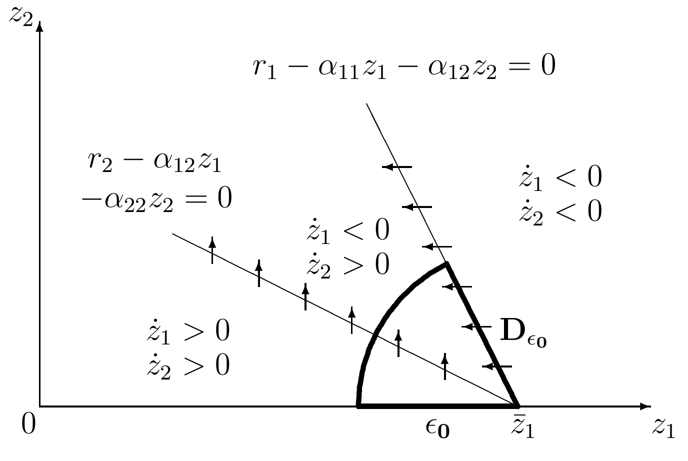


FIG. A.2 – *L'ensemble D_ε et le signe de $\dot{z}_1$ et $\dot{z}_2$.*

Nous avons établi le critère de stabilité asymptotique de $(\bar{z}_1, 0)$. En ce qui concerne la stabilité asymptotique globale, observons que $(0,0)$ est toujours répulsif ($r_1 > 0$ et $r_2 > 0$). D'autre part, si le système (2) admet un équilibre GAS, aucun autre équilibre ne peut être asymptotiquement stable, et, en particulier, $(0, \bar{z}_2)$ est instable. Le critère (3) constitue donc une condition nécessaire pour que $(\bar{z}_1,0)$ soit GAS. Or, ce critère interdit l'existence d'équilibres de (2) dans $(\mathbb{R}_+^*)^2$. En effet, puisque $r_2\alpha_{11}-r_1\alpha_{21}$ et $r_1\alpha_{22}-r_2\alpha_{12}$ ne sont pas tous les deux nuls, il ne peut y avoir au plus qu'un seul équilibre non trivial (les deux droites des figures A.1 ne sont pas confondues). Si $\alpha_{11}\alpha_{22}=\alpha_{12}\alpha_{21}$, les deux droites sont parallèles, et il n'y a pas d'équilibre non trivial. Sinon, l'intersection des deux droites est donnée par

$$\left(\frac{r_1\alpha_{22}-r_2\alpha_{12}}{\alpha_{11}\alpha_{22}-\alpha_{12}\alpha_{21}},\frac{r_2\alpha_{11}-r_1\alpha_{21}}{\alpha_{11}\alpha_{22}-\alpha_{12}\alpha_{21}}\right). \qquad (8)$$

Or, ce point ne peut avoir ses deux coordonnées négatives, puisque aucune des droites n'intersecte l'ensemble $\mathbb{R}_-^2$. Par suite, ce quatrième équulibre appartient à $(\mathbb{R}_+^*)^2$ si et seulement si

$$(r_1\alpha_{22}-r_2\alpha_{12})(r_2\alpha_{11}-r_1\alpha_{21})>0$$

(les deux coordonnées sont de même signe). Or, ceci est incompatible avec la condition (3).

Par conséquent, sous (3), $(\bar{z}_1,0)$ et $(0,\bar{z}_2)$ sont les seules limites possibles des solutions de (2) (excepté si la condition initiale est nulle). Il suffit donc de montrer qu'aucune solution issue de $(\mathbb{R}_+^*)^2$ ne peut converger vers $(0,\bar{z}_2)$ pour conclure la preuve de (c).

Dans le cas où $r_1\alpha_{22}-r_2\alpha_{12}>0$, $\dot{z}_1>0$ en tout point suffisamment proche de $(0,\bar{z}_2)$ avec $z_1>0$. Le résultat désiré découle alors aisément de (b).

Le cas où $r_2\alpha_{11}-r_1\alpha_{21}<0$ et $r_1\alpha_{22}-r_2\alpha_{12}=0$ est plus délicat. Considérons une solution $(z_1(t),z_2(t))$ de (2) avec $z_1(0)>0$ and $z_2(0)>0$, et supposons que $z_1(t)\to\bar{z}_1$ and $z_2(t)\to 0$ when $t\to+\infty$. Soit t_0 suffisamment grand pour que $(z_1(t),z_2(t))\in B((\bar{z}_1,0),\varepsilon_0)$ pour tout $t\geq t_0$. Comme nous l'avons vu plus haut, toute solution de (2) issue de D_{ε_0} quitte $B((\bar{z}_1,0),\varepsilon_0)$. Par conséquent, la solution que nous étudions ne peut atteindre D_{ε_0} après t_0. Si nous posons $h_1(t)=z_1(t)-\bar{z}_1$ et $h_2(t)=z_2(t)$, et si nous changeons l'origine des temps de telle sorte que $t_0=0$, ceci s'écrit :

$$\forall t\geq 0,\quad h_2(t)\geq -\frac{\alpha_{21}}{\alpha_{22}}h_1(t). \qquad (9)$$

Nous allons obtenir une contradiction de cette inégalité, ce qui achèvera la preuve de (c).

Définissons

$$q=\begin{cases}\frac{h_2(0)}{h_1(0)} & \text{if } h_1(0)>0\\ 1 & \text{if } h_1(0)\leq 0.\end{cases}$$

D'après (b), $z_2(0) > 0$, et, par suite, $q > 0$ et $h_2(0) \geq qh_1(0)$. Soit $v_1,v_2 \in (0,\varepsilon_0)$ tel que $v_2 = qv_1$. Quitte à résuire ε_0 si nécessaire, on a

$$\begin{aligned} -F_1(\bar{z}_1 + v_1,v_2) &\geq \frac{r_1}{2\alpha_{11}}(\alpha_{11} + q\alpha_{12})v_1 \\ -F_2(\bar{z}_1 + v_1,v_2) &= qv_1^2(\alpha_{21} + q\alpha_{22}), \end{aligned}$$

et $\varepsilon_0 < r_1/2\alpha_{21}$. Puisque $\alpha_{11}\alpha_{22} < \alpha_{12}\alpha_{21}$, on peut alors écrire

$$v_1 < \varepsilon_0 < \frac{r_1}{2\alpha_{21}} \frac{1 + \frac{\alpha_{12}}{\alpha_{11}}q}{1 + \frac{\alpha_{22}}{\alpha_{21}}q} = \frac{\frac{r_1}{2\alpha_{11}}(\alpha_{11} + q\alpha_{12})}{\alpha_{21} + q\alpha_{22}},$$

ce qui implique l'inégalité $F_2(\bar{z}_1+v_1,v_2) \geq qF_1(\bar{z}_1+v_1,v_2)$, vraie pour $v_2 = qv_1$ et $v_1,v_2 \in (0,\varepsilon_0)$. En particulier, aucune solution de (2) issue de $B((\bar{z}_1,0),\varepsilon_0)$ ne peut couper la droite $z_2 = q(z_1 - \bar{z}_1)$ vers le bas à l'intérieur de cette boule. Nous obtenons ainsi, puisque $h_2(0) \geq qh_1(0)$,

$$\forall t \geq 0, \quad h_2(t) \geq qh_1(t). \tag{10}$$

Maintenant, fixons $\varepsilon < \varepsilon_0$, et changeons de nouveau l'origine des temps de telle sorte que $|h_1(t)| < \varepsilon$ et $h_2(t) < \varepsilon$ pour tout $t \geq 0$. Les propriétés (9) et (10) sont toujours vérifiées. Or, d'après (7), pour tout $t \geq 0$,

$$-h_2(t)(\alpha_{21} + \alpha_{22})\varepsilon \leq \dot{h}_2(t) = F_2(\bar{z}_1 + h_1(t),h_2(t)) \leq 0,$$

et, par suite,

$$h_2(0)e^{-\varepsilon(\alpha_{21}+\alpha_{22})t} \leq h_2(t) \leq h_2(0). \tag{11}$$

Observons que les inégalités $\alpha_{11}\alpha_{22} < \alpha_{12}\alpha_{21}$ et (9) impliquent que $\alpha_{11}h_1(t)+\alpha_{12}h_2(t) > 0$ pour tout $t \geq 0$. Par conséquent, en combinant (7) avec l'inégalité de gauche de (11), nous obtenons, pour ε assez petit,

$$-\dot{h}_1(t) = -F_1(\bar{z}_1 + h_1(t),h_2(t)) \geq \frac{r_1}{2}\left(h_1(t) + \frac{\alpha_{12}}{\alpha_{11}}h_2(0)e^{-\varepsilon(\alpha_{21}+\alpha_{22})t}\right).$$

Cette inégalité différentielle s'intègre comme suit :

$$\begin{aligned} - h_1(t) \geq{}& -h_1(0)e^{-r_1t/2} \\ &+ \frac{r_1/2}{r_1/2 - \varepsilon(\alpha_{21} + \alpha_{22})}\frac{\alpha_{12}}{\alpha_{11}}h_2(0)\left(e^{-\varepsilon(\alpha_{21}+\alpha_{22})t} - e^{-r_1t/2}\right). \end{aligned} \tag{12}$$

En combinant (9), (10), (11) and (12) et en utilisant la notation $C = 2(\alpha_{21}+\alpha_{22})/r_1$, il suit

$$\begin{aligned} h_2(0) \geq h_2(t) \geq -\frac{\alpha_{21}}{\alpha_{22}} h_1(t) \geq -\frac{\alpha_{21}}{\alpha_{22}} \frac{h_2(0)}{q} e^{-r_1 t/2} \\ + \frac{\alpha_{12}\alpha_{21}}{\alpha_{11}\alpha_{22}} \frac{h_2(0)}{1-C\epsilon} \left(e^{-\varepsilon(\alpha_{21}+\alpha_{22})t} - e^{-r_1 t/2} \right). \end{aligned}$$

Or, nous avons supposé $\alpha_{11}\alpha_{22} < \alpha_{12}\alpha_{21}$, donc il existe $\theta \in (0,1)$ tel que $\alpha_{12}\alpha_{21}/\alpha_{11}\alpha_{22} = 1/(1-\theta)$. De plus, $1 < 1/(1-C\varepsilon) < 2$ pour ε assez petit. Donc, après simplification,

$$1 \geq -\frac{\alpha_{21}}{q\alpha_{22}} e^{-r_1 t/2} + \frac{1}{1-\theta} \left(e^{-\varepsilon(\alpha_{21}+\alpha_{22})t} - 2e^{-r_1 t/2} \right). \tag{13}$$

Fixons d'abord $t > 0$ (indépendamment de ε) tel que

$$\frac{\alpha_{21}}{q\alpha_{22}} e^{-r_1 t/2} < \frac{\theta}{4(1-\theta)} \quad \text{and} \quad e^{-r_1 t/2} < \frac{\theta}{4}.$$

Choisissons ensuite $\varepsilon < \varepsilon_0$ de telle sorte que

$$e^{-\varepsilon(\alpha_{21}+\alpha_{22})t} > 1 - \frac{\theta}{4}.$$

Pour de tels t et ε, d'après (13),

$$1 > -\frac{\theta}{4(1-\theta)} + \frac{1}{1-\theta} \left(1 - \frac{\theta}{4} - \frac{\theta}{2} \right) = 1,$$

ce qui donne la contradiction voulue. □

Preuve de (d) Nous avons déjà vu plus haut que le système (2) possède un unique équilibre (z_1^*,z_2^*) dans $(\mathbb{R}_+^*)^2$ si et seulement si (4) est vérifiée. L'expression de (z_1^*,z_2^*) est donnée par (8), et la matrice de stabilité de cet équilibre est

$$\begin{pmatrix} -\alpha_{11} z_1^* & -\alpha_{12} z_1^* \\ -\alpha_{21} z_2^* & -\alpha_{22} z_2^* \end{pmatrix}. \tag{14}$$

Son déterminant, $(\alpha_{11}\alpha_{22} - \alpha_{12}\alpha_{21}) z_1^* z_2^*$, est strictement négatif si $\alpha_{11}\alpha_{22} - \alpha_{12}\alpha_{21} < 0$. D'après (8), ce cas est équivalent à (5). En outre, les valeurs propres de (14) sont de signe opposées, d'où l'instabilité de (z_1^*,z_2^*).

Si (z_1^*,z_2^*) est GAS, $(\bar{z}_1,0)$ et $(0,\bar{z}_2)$ ne peuvent pas être asymptotiquement stables, donc (6) est une condition nécessaire de stabilité asymptotique globale. Réciproquement, si (6) est vraie, comme nous l'avons vu dans la preuve de (c), une solution de (2) issue de $(\mathbb{R}_+^*)^2$ ne peut converger vers $(0,0)$, $(\bar{z}_1,0)$ ou $(0,\bar{z}_2)$. Or, d'après (a), elle doit nécessairement converger vers un équilibre. Le seule possibilité est (z_1^*,z_2^*), ce qui achève la preuve du théorème 1. □

Annexe B

Code C de simulation du modèle individu-centré

Voici le source du programme simulant le modèle de Kisdi [49]. Ce programme crée directement une image `.bmp` semblable aux figures de la section 1.3 de l'introduction. La plus grande partie du programme s'occupe de l'affichage de la simulation sur le fichier sortie (procédures `format_bmp`, `aff_fond`, `aff_grad`, `aff_couleur`, `init_grad`, `tri_branche`, `nbre_indiv` et `aff_popu`). Seules la procédure `evenement` et le corps du programme effectuent la simulation proprement dite.

Pour simuler le modèle de Dieckmann et Doebeli [16], il suffit de changer, dans le prologue du programme, les définitions de `RHO(s)` et de `ALPHA(s1,s2)` en

```
#define RHO(s) (exp(-(s*s*SIGMAK)))
```

et

```
#define ALPHA(s1,s2) (exp(-(((s1)-(s2))*((s1)-(s2))*SIGMAC)))
```

et de donner la valeur `1.0` aux constantes `RHO_BAR` et `ALPHA_BAR`, la valeur `-2.0` à `TRAIT_MIN` et la valeur `2.0` à `TRAIT_MAX` (on peut aussi changer le nom du fichier `NOM[200]` en `char NOM[200]="Dieckmann_Doebeli_";`). Les variables `SIGMAK` et `SIGMAC` valent alors, respectivement, $1/(2\sigma_k^2)$ et $1/(2\sigma_c^2)$ avec les notations de l'exemple 2 de la section 1.3 de l'introduction, et peuvent être définies dans le préambule du programme ou entrées avec les autres paramètres dans la ligne de commande (en modifiant le préambule et le corps du programme en conséquence).

```
/************************************************************************/
/*                                                                      */
/*      Programme de simulation du modèle individu centré de Kisdi      */
/*                                                                      */
/*     Prend quatre nombres en argument : la carrying capacity K, la    */
/*     probabilité de mutation à la naissance MU, l'écart-type des      */
/*     sauts de mutation SIGMA et la durée de la simulation T_MAX.      */
/*                                                                      */
/************************************************************************/


#include <stdio.h>
#include <math.h>
#include <string.h>                 /* Manipulation des chaînes de char */
#include <stdlib.h>         /* Permet de générer des nombres aléatoires */

#define RHO(s) (4-(s))        /* Taux de naissance en fonction du trait */
#define RHO_BAR 4.0                      /* Majorant uniforme de RHO(s) */
#define ALPHA(s1,s2) (2*(1-1/(1+1.2*exp(-4*((s1)-(s2))))))
                                                  /* Noyau de mortalité */
#define ALPHA_BAR 2.0              /* Majorant uniforme de ALPHA(s1,s2) */

#define TRAIT_INIT 1.2                                 /* Trait initial */
#define TRAIT_MAX 4                         /* Valeur maximale du trait */
#define TRAIT_MIN 0                         /* Valeur minimale du trait */
#define POPU_MAX 100000             /* Taille maximale de la population */
#define PAS_AFF_TRAITS 0.01               /* Pas d'affichage des traits */
                  /* (vaut 0.01 si l'image a une largeur de 400 pixels) */
#define N_AFF_MAX 3             /* N_AFF_MAX*K=effectif maximal affiché */
#define FOND -1                     /* Couleur du fond : 0=noir -1=blanc */
#define GRAD 0                                 /* Couleur de graduation */
#define SEUIL_BRANCHE 8   /* Nombre de pixels vides minimal pour que le */
                                   /* programme distingue deux branches */
#define TAILLE_BRANCHE 0.05     /* Densité minimale pour qu'une branche */
                                                /* soit prise en compte */

int taille=1713600;               /* Taille du fichier sortie (en char) */
FILE *fp;                                      /* Fichier sortie (.bmp) */
char NOM[200]="Kisdi_";                        /* Nom du fichier sortie */
char d=11, e=-11;          /* Définissent la clarté ou le contraste des
                                         couleurs (de manière empirique) */
int K;                                             /* Carrying capacity */
float MU;                                    /* Probabilité de mutation */
float SIGMA;                                    /* Ecart-type des sauts */
float POPU[POPU_MAX];                     /* Liste des traits présents */
int N;                           /* Effectif total courant : < POPU_MAX */
int N_BRANCHE[7];            /* Effectifs dans chaque branche (au plus,
                                                6 branches distinguées) */
int COUL_BRANCHE[6];         /* Couleurs correspondant à chaque branche */
```

```
float PENTE_COULEUR;                /* Pente réglant le code de couleur */
int INDICE_BRANCHE;                      /* Nombre de branches courant */
long double T_MAX;                       /* Temps de fin de simulation */
long double T=0.0;                                    /* Temps courant */
long double T_AFF;                     /* Temps entre chaque affichage */
int CPTR_AFF=0;                          /* Nombre de lignes affichées */
int lignes_grad;                    /* Nombre de lignes affichées entre
                                          chaque graduation horizontale */
float temp;                                  /* Variable intermédiaire */

/************************************************************************/
/*  Procédure codant le prologue d’un fichier .bmp 820*620 (empirique)  */
/************************************************************************/

void format_bmp(void)
{
  char c;
  int i;

  c=’B’; fputc(c, fp);
  c=’M’; fputc(c, fp);
  c=’ö’; fputc(c, fp);
  c=’%’; fputc(c, fp);
  c=26; fputc(c, fp);
  for(c=0, i=0; i<5; i++) fputc(c, fp);
  c=’6’; fputc(c, fp);
  for(c=0, i=0; i<3; i++) fputc(c, fp);
  c=’(’; fputc(c, fp);
  for(c=0, i=0; i<3; i++) fputc(c, fp);
  c=’H’; fputc(c, fp);
  c=3; fputc(c, fp);
  for(c=0, i=0; i<2; i++) fputc(c, fp);
  c=-88; fputc(c, fp);
  c=2; fputc(c, fp);
  for(c=0, i=0; i<2; i++) fputc(c, fp);
  c=1; fputc(c, fp);
  c=0; fputc(c, fp);
  c=24; fputc(c, fp);
  for(c=0, i=0; i<5; i++) fputc(c, fp);
  c=-64; fputc(c, fp);
  c=37; fputc(c, fp);
  c=26; fputc(c, fp);
  c=0; fputc(c, fp);
  c=’Ä’; fputc(c, fp);
  c=14; fputc(c, fp);
  for(c=0, i=0; i<2; i++) fputc(c, fp);
  c=’Ä’; fputc(c, fp);
  c=14; fputc(c, fp);
  for(c=0, i=0; i<10; i++) fputc(c, fp);
}
```

```
/**********************************************************************/
/*    Procédures affichant un pixel (trois char) de FOND ou de GRAD.    */
/**********************************************************************/

void aff_fond(void)
{
  fputc(FOND, fp); fputc(FOND, fp); fputc(FOND, fp);
  taille -=3;
}

void aff_grad(void)
{
  fputc(GRAD, fp); fputc(GRAD, fp); fputc(GRAD, fp);
  taille -=3;
}

/**********************************************************************/
/* Procédure affichant un pixel d'une couleur (empirique) correspondant */
/* à la densité de population dens, dans le fichier sortie.           */
/**********************************************************************/

void aff_couleur(float dens)
{
  int ent;                               /* Indice du code de couleur */
  char b,c;                                        /* Chars à afficher */
  float tmp;                                /* Variable intermédiaire */

  /* dens devrait varier entre 0 et 3, avec principalement des
     valeurs entre 0 et 0.1. On va donc utiliser une échelle de couleur
     en arctan avec une pente PENTE_COULEUR. */
  /* Les couleurs vont de bleu clair à rouge pour les faibles
     effectifs, et de orange à vert pour les grands effectifs. */
  if (dens == 0)
    {
     aff_fond();
     return;
    }
  dens *= PENTE_COULEUR;
  if (dens < 3) b = e;
  else
    {
     if (dens > 23) b = -111;
     else b = e-(char)(5*(dens-3));
    }
  dens = atan(dens)*5.5/1.57;                      /* Couleur max=5.5 */
  ent = floor(dens);
  dens -= ent;
  dens *= (256+((int)e-d));     /* Nombre de char dans chaque catégorie */
  switch(ent)
    {
    case 0 :          /* Premier groupe de couleurs : BLEU CLAIR - BLEU */
```

```
      c = e-(char)floor(dens);
      fputc(b, fp); fputc(c, fp); fputc(d, fp);
      break;
    case 1 :       /* Deuxième groupe de couleurs : BLEU MARINE - MAUVE */
      c = d+(char)floor(dens);
      fputc(b, fp); fputc(d, fp); fputc(c, fp);
      break;
    case 2 :            /* Troisième groupe de couleurs : MAUVE - ROUGE */
      c = e-(char)floor(dens);
      fputc(c, fp); fputc(d, fp); fputc(b, fp);
      break;
    case 3 :            /* Quatrième groupe de couleurs : ROUGE - JAUNE */
      c = d+(char)floor(dens);
      fputc(d, fp); fputc(c, fp); fputc(b, fp);
      break;
    case 4 :             /* Cinquième groupe de couleurs : JAUNE - VERT */
      c = e-(char)floor(dens);
      fputc(d, fp); fputc(b, fp); fputc(c, fp);
      break;
    case 5 :           /* Sixième groupe de couleurs : VERT - TURQUOISE */
      c = d+(char)floor(dens);
      fputc(c, fp); fputc(b, fp); fputc(d, fp);
      break;
    }
  taille -= 3;        /* Nombre d'octets restant dans le fichier sortie */
}

/*********************************************************************/
/*        Procédure de graduation horizontale du fichier sortie.       */
/*********************************************************************/

void init_grad(void)
{
  int i, j;

  /* 50 lignes vierges, plus le début de la suivante */
  for (i=0; i<42019; i++) aff_fond();

  /* Graduation horizontale : ligne 1 */
  aff_grad();
  for (i=0; i<4; i++)
    {
      for (j=0; j<99; j++) aff_fond();
      aff_grad();
    }
  for (i=0; i<54; i++) aff_fond();
  aff_grad();
  for (i=0; i<3; i++)
    {
      for (j=0; j<99; j++) aff_fond();
      aff_grad();
```

```
    }
  for (i=0; i<84; i++) aff_fond();
  /* Ligne 2 */
  aff_grad();
  for (i=0; i<8; i++)
    {
      for (j=0; j<49; j++) aff_fond();
      aff_grad();
    }
  for (i=0; i<54; i++) aff_fond();
  aff_grad();
  for (i=0; i<6; i++)
    {
      for (j=0; j<49; j++) aff_fond();
      aff_grad();
    }
  for (i=0; i<84; i++) aff_fond();
  /* Ligne 3 */
  aff_grad();
  for (i=0; i<40; i++)
    {
      for (j=0; j<9; j++) aff_fond();
      aff_grad();
    }
  for (i=0; i<54; i++) aff_fond();
  aff_grad();
  for (i=0; i<30; i++)
    {
      for (j=0; j<9; j++) aff_fond();
      aff_grad();
    }
  for (i=0; i<84; i++) aff_fond();
  /* Ligne 4 */
  for (i=0; i<401; i++) aff_grad();
  for (i=0; i<54; i++) aff_fond();
  for (i=0; i<301; i++) aff_couleur((300.1-i)/100.0);
  for (i=0; i<84; i++) aff_fond();
}

/***********************************************************************/
/* Procédure de tri rapide (quicksort) de N_BRANCHE[] (effecifs de      */
/* chaque branche de la population) entre les indices u et v.          */
/* Le tableau COUL_BRANCHE est trié en même temps afin de garder trace  */
/* de l'ordre des branches dans l'espace des traits.                   */
/***********************************************************************/

void tri_branche(int u, int v)
{
  int interm;

  /* Tests de fin de récursion */
```

```
  if (u >= v) return;
  if (u == v-1)
    {
      if (N_BRANCHE[u]>N_BRANCHE[v])
        {
          temp=N_BRANCHE[u];
          N_BRANCHE[u]=N_BRANCHE[v];
          N_BRANCHE[v]=temp;
          interm=COUL_BRANCHE[u];
          COUL_BRANCHE[u]=COUL_BRANCHE[v];
          COUL_BRANCHE[v]=interm;
        }
      return;
    }

  int i=u, j=v-1;
  float pivot=N_BRANCHE[v];

  /* Tri */
  while (1)
    {
      while ((N_BRANCHE[i]<=pivot) && (i<v)) i++;
      if (j<=i) break;
      while ((N_BRANCHE[j]>=pivot) && (j>i)) j--;
      if (j<=i) break;
      temp=N_BRANCHE[i];
      N_BRANCHE[i]=N_BRANCHE[j];
      N_BRANCHE[j]=temp;
      interm=COUL_BRANCHE[i];
      COUL_BRANCHE[i]=COUL_BRANCHE[j];
      COUL_BRANCHE[j]=interm;
      i++;
      j--;
    }
  N_BRANCHE[v]=N_BRANCHE[i];
  N_BRANCHE[i]=pivot;
  interm=COUL_BRANCHE[v];
  COUL_BRANCHE[v]=COUL_BRANCHE[i];
  COUL_BRANCHE[i]=interm;
  tri_branche(u,i-1);
  tri_branche(i+1,v);
  return;
}
```

```
/*********************************************************************/
/* Procédure calculant le nombre d'individus ayant un trait entre s et  */
/* s-0.01 dans la population POPU[] de taille N.                        */
/*********************************************************************/

int nbre_indiv(float s)
{
  int nbre=0;
  int i;

  for (i=0; i<=N; i++)
    {
      if ((POPU[i]<s) && (POPU[i]>=s-PAS_AFF_TRAITS)) nbre++;
    }
  return(nbre);
}

/*********************************************************************/
/*  Procédure d'affichage d'une nouvelle ligne. Cette procédure tient   */
/*  compte de l'affichage de la graduation. Elle remplit le tableau     */
/*  N_BRANCHE[] et affiche les effectifs dans chaque branche.           */
/*********************************************************************/

void aff_popu(void)
{
  int TRAIT;                              /* 100*(trait courant affiché) */
  int n;                                /* Nombre d'individus à afficher */
  int i, j;                                               /* Compteurs */
  float trait_moyen=0;                                  /* Trait moyen */

  /* Affichage des traits de la population */
  INDICE_BRANCHE=0;
  for (i=0; i<6; i++) N_BRANCHE[i]=0;
  for (TRAIT=TRAIT_MAX/PAS_AFF_TRAITS+1, i=SEUIL_BRANCHE;
       PAS_AFF_TRAITS*TRAIT>TRAIT_MIN; TRAIT--)
    {
      n=nbre_indiv(TRAIT*PAS_AFF_TRAITS);      /* Nombre d'individus de */
      if (n == 0)             /* trait entre TRAIT/100 et (TRAIT-1)/100 */
        {
          i--;
          if ((i == 0) && (INDICE_BRANCHE<6) &&       /* On conserve la */
              (N_BRANCHE[INDICE_BRANCHE]>(float)K*TAILLE_BRANCHE))
            INDICE_BRANCHE++;    /* branche si sa taille est suffisante */
        }
      else
        {
          i=SEUIL_BRANCHE;
          N_BRANCHE[INDICE_BRANCHE] += n;
        }
      aff_couleur((float)n/K);         /* Affiche la couleur associée à */
    }                                                /* la densité n/K */
```

```
  if (N_BRANCHE[INDICE_BRANCHE]<(float)K*TAILLE_BRANCHE)
    INDICE_BRANCHE--;                                 /* Voir plus haut */

  /* Affichage de la graduation */
  aff_grad();
  if (CPTR_AFF % lignes_grad == 0)
    {
      aff_grad();
      for (j=0; j<=N; j++) trait_moyen += POPU[j];
      printf("Temps= %8.1Lf   Taille de la population= %6d   \
Trait moyen= %2.3f\n", T, N+1, trait_moyen/(N+1));
      for (i=1; i<lignes_grad; i++) printf("#");
    }
  else
    {
      aff_fond();
      printf("\b \b");
    }
  if (CPTR_AFF % (5*lignes_grad) == 0) aff_grad();
  else aff_fond();
  if (CPTR_AFF % (10*lignes_grad) == 0) aff_grad();
  else aff_fond();
  for (i=0; i<50; i++) aff_fond();

  /* Affichage des effectifs totaux et de chaque branche */
  n=(int)((300*N)/(K*N_AFF_MAX));           /* Nombre de pixels restant */
  if (n>300) n=300;               /* Taille max du graphe des effectifs */
  else
    {
      for (i=300; i>n; i--) aff_fond();
      aff_grad();                        /* On affiche l'effectif total */
      n--;
    }
  if (INDICE_BRANCHE >= 6) INDICE_BRANCHE=5;
  if (INDICE_BRANCHE <= 0) INDICE_BRANCHE=-1;
  for (i=0; i<6; i++) COUL_BRANCHE[i]=i;
  tri_branche(0, INDICE_BRANCHE);     /* Tri des effectifs des branches */
  for (j=INDICE_BRANCHE; j>=0; j--)
    {
      if (n == -1) break;
      for (i=(int)((300*N_BRANCHE[j])/(K*N_AFF_MAX)); n>i; n--)
        aff_fond();
      aff_couleur(tan((COUL_BRANCHE[j]+0.5)*1.57/5.5)/PENTE_COULEUR);
      n--;
    }
  for ( ; n>-1; n--) aff_fond();                        /* Fin de ligne */

  /* Affichage de la graduation */
  aff_grad();
  if (CPTR_AFF % lignes_grad == 0) aff_grad();
  else aff_fond();
```

```
  if (CPTR_AFF % (5*lignes_grad) == 0) aff_grad();
  else aff_fond();
  if (CPTR_AFF % (10*lignes_grad) == 0) aff_grad();
  else aff_fond();
  for (i=0; i<80; i++) aff_fond();
  /* Ici, je suis arrivé au début de la nouvelle ligne du graphique */
}

/************************************************************************/
/*   Procédure simulant un événement de la population (naissance sans   */
/*   mutation, naissance avec mutation ou mort) et incrémentant le      */
/*   temps T. Cette procédure stoppe le programme si la population      */
/*   dépasse la taille maximale POPU_MAX.                               */
/************************************************************************/

void evenement(void)
{
  float s, saut;                       /* Trait et saut de trait mutant */
  int i1;                                             /* Indice de mort */
  float taux_tot;                                               /* Taux */

  taux_tot=N*(RHO_BAR+ALPHA_BAR*N/K);
  T += (long double)-log(1-drand48())/(long double)taux_tot;
  if (drand48()<N*RHO_BAR/taux_tot)           /* Evénement de naissance */
    {
      s=POPU[(int)floor(N*drand48())];
      if (drand48()<RHO(s)/RHO_BAR)
        {
          N++;
          if (N>=POPU_MAX)       /* En cas de dépassement, on interromp */
            {
              printf("\nDepassement d'effectif de population !\nChanger\
le parametre POPU_MAX dans le source.\n");
              for ( ; taille>0; taille--) fputc(FOND, fp);
              fclose(fp);
              exit(1);
            }
          if (drand48()<MU)                  /* Naissance avec mutation */
            {
              do
                {
                  saut = sqrt(-2.0*log(drand48()))*cos(6.2852*drand48());
                  if (drand48()<0.5) saut=s+SIGMA*saut;
                  else saut=s-SIGMA*saut;
                }
              while (saut<=TRAIT_MIN || saut>=TRAIT_MAX);
              POPU[N]=saut;
            }
          else POPU[N]=s;                    /* Naissance sans mutation */
        }
    }
```

```
  else                                             /* Evénement de mort */
    {
      i1=(int)floor(N*drand48());
      s=POPU[(int)floor(N*drand48())];
      if (drand48()<ALPHA(POPU[i1],s)/ALPHA_BAR)
        {
          POPU[i1]=POPU[N];
          N--;
        }
    }
}

/************************************************************************/
/*                          Corps du programme                          */
/************************************************************************/

int main(int argc, char *argv[])
{
  srand48(time(0));   /* Initialise le générateur de nombres aléatoires */

  /* On prend les arguments dans la ligne de commande */
  if (argc!=5)
    {
      printf("Nombre de parametres incorrect.\nArguments requis :\n\
\tCarrying capacity (entiere)\n\tProbabilite de mutation lors d'une \
naissance\n\tEcart-type de la distribution du nouveau trait lors \
d'une mutation\n\tDuree de la simulation\n");
      exit(1);
    }
  K = atoi(argv[1]);
  MU = atof(argv[2]);
  SIGMA = atof(argv[3]);
  T_MAX = (long double)atof(argv[4]);
  if (K<=0 || MU<0 || MU>1 || SIGMA<=0 || T_MAX<=0)
    {
      printf("Parametres incorrects\n");
      exit(1);
    }
  if ((4*K)>POPU_MAX)
    {
      printf("Carrying capacity trop grande.\nChanger le parametre \
POPU_MAX dans le source\n");
      exit(1);
    }

  /* On prépare le fichier graphique */
  strcat(NOM, argv[1]);
  strcat(NOM, "_");
  strcat(NOM, argv[2]);
  strcat(NOM, "_");
  strcat(NOM, argv[3]);
```

```
  strcat(NOM, "_");
  strcat(NOM, argv[4]);
  strcat(NOM, ".bmp");
  if ((fp = fopen(NOM, "w")) == NULL)          /* Ouverture en écriture */
    {
      fprintf(stderr,"Impossible d'ouvrir le fichier Kisdi.bmp\n");
      exit(1);
    }
  format_bmp();
  PENTE_COULEUR = 5.0*log10(K);            /* Pente du code de couleurs */
  init_grad();                                /* Graduations verticales */

  temp = pow(10,floor(log10(T_MAX))-1);           /* Temps entre chaque
                                                           graduation */
  lignes_grad = floor(600*temp/T_MAX);          /* Nombre de lignes par
                                                           graduation */
  T_AFF = (long double)temp/lignes_grad;    /* Temps entre chaque ligne
                                                             affichée */

  /* Population initiale = K individus de trait TRAIT_INIT */
  for (N=0; N<K; N++) POPU[N]=TRAIT_INIT;
  N--;                                      /* N=K-1, i.e. K individus */

  /* Simulation */
  while (taille>63000)        /* 25 lignes vierges à la fin du fichier */
    {
      evenement();                            /* On simule un événement */
      if (N == -1)                                  /* Cas d'extinction */
        {
          for ( ; taille>0; taille--) fputc(FOND, fp);
          printf("\nExtinction !\n");
          fclose(fp);
          return(0);
        }
      if (T-CPTR_AFF*T_AFF>0)    /* On affiche une nouvelle ligne si le */
        {                                  /* pas de temps est franchi */
          aff_popu();           /* Affiche l'ensemble de la population */
          CPTR_AFF++;      /* Incrément du compteur de lignes affichées */
        }
    }

  /* Fermeture du fichier */
  for ( ; (CPTR_AFF % lignes_grad) != 0; CPTR_AFF++) printf("\b \b");
  for ( ; taille>0; taille--) fputc(FOND, fp);
  fclose(fp);
  return(0);
}
```

Bibliographie

[1] D. Aldous. Stopping times and tightness. *Ann. Probab.*, **6**:335–340, 1978.

[2] C. B. Anfinsen. Principles that govern the folding of protein chains. *Science*, **181**:223–230, 1973.

[3] K. B. Athreya and P. E. Ney. *Branching processes.* Springer-Verlag, Berlin, 1972.

[4] R. Azencott. Grandes déviations et applications. In *Summer school of St Flour VII, 78*, Lecture Note in Mathematics. Springer-Verlag, 1980.

[5] B. Bolker and S. W. Pacala. Using moment equations to understand stochastically driven spatial pattern formation in ecological systems. *Theor. Pop. Biol.*, **52**:179–197, 1997.

[6] B. M. Bolker and S. W. Pacala. Spatial moments equations for plant competition: Understanding spatial strategies and the advantage of short dispersal. *Am. Nat.*, **153**:575–602, 1999.

[7] R. Bürger. *The Mathematical Theory of Selection, Recombination and Mutation.* John Wiley & Sons, 2000.

[8] R. Bürger and I. M. Bomze. Stationary distributions under mutation-selection balance: structure and properties. *Adv. Appl. Probab.*, **28**:227–251, 1996.

[9] N. Champagnat. Large deviations for diffusion models of adaptive dynamics. Prépublication Université Paris 10 – Nanterre No. 03/23, soumis à Ann. Appl. Proba., 2003.

[10] N. Champagnat, R. Ferrière, and G. Ben Arous. The canonical equation of adaptive dynamics: A mathematical view. *Selection*, **2**:73–83, 2001.

[11] A. S. Cherny and H.-J. Engelbert. Isolated singular points of stochastic differential equations. In R. Buckdahn, H.-J. Engelbert, and M. Yor, editors, *Stochastic Processes and Related Topics, Proceedings of the 12th Winter School on Stochastic Processes, Seigmundsburg (Germany), February 27–March 4, 2000.* Stochastics Monographs Vol. 12, Taylor & Francis, London New-York, 2002.

[12] O. De Feo and R. Ferrière. Bifurcation analysis of population invasion: On-off intermittency and basin riddling. *International Journal of Bifurcation and Chaos*, **10**:443–452, 2000.

[13] A. Dembo and O. Zeitouni. *Large Deviation Techniques and Applications.* Springer-Verlag, New York, second edition, 1998.

[14] F. Dercole, J. O. Irisson, and S. Rinaldi. Bifurcation analysis of a prey-predator coevolution model. *SIAM J. Appl. Math.*, **64**:1378–1391, 2003.

[15] U. Dieckmann. Coevolutionary dynamics of stochastic replicator systems. Masters thesis, internal report jül-3018, University of Jülich, 1995.

[16] U. Dieckmann and M. Doebeli. On the origin of species by sympatric speciation. *Nature*, **400**:354–357, 1999.

[17] U. Dieckmann and R. Ferrière. Adaptive dynamics and evolving biodiversity. In R. Ferrière, U. Dieckmann, and D. Couvet, editors, *Evolutionary Conservation Biology*, pages 188–224, Cambridge, 2004. Cambridge University Press.

[18] U. Dieckmann and R. Law. The dynamical theory of coevolution: A derivation from stochastic ecological processes. *J. Math. Biol.*, **34**:579–612, 1996.

[19] O. Diekmann. A beginner's guide to adaptive dynamics. *Banach Center Publications*, **63**:47–86, 2003.

[20] O. Diekmann, P.-E. Jabin, S. Mischler, and B. Perthame. The dynamics of adaptation: an illuminating example and a Hamilton-Jacobi approach. Prépublication du DMA No. DMA-04-14, Ecole Normale Supérieure, Paris, 2004.

[21] M. Doebeli and U. Dieckmann. Evolutionary branching and sympatric speciation caused by different types of ecological interactions. *Am. Nat.*, **156**:S77–S101, 2000.

[22] M. Doebeli and U. Dieckmann. Speciation along environmental gradients. *Nature*, **421**:259–264, 2003.

[23] H. Doss and P. Priouret. Petites perturbations de systèmes dynamiques avec reflexion. In *Seminar of Probability XVII, Lecture Notes in Mathematics **986***, pages 353–370. Springer-Verlag, 1983.

[24] P. Dupuis and R. S. Ellis. *A Weak Convergence Approach to the Theory of Large Deviations.* Wiley, New-York, 1997.

[25] R. Durrett. Interacting particle systems. In *Summer School of Saint Flour, XXII*, Lecture Notes in Mathematics, 1608. Springer-Verlag, Berlin, 1993.

[26] R. Durrett. *Probability Models for DNA Sequence Evolution.* Springer-Verlag, 2002.

[27] R. Durrett and S. Levin. The importance of being discrete (and spatial). *Theor. Pop. Biol.*, **46**:363–394, 1994.

[28] S. N. Ethier and T. G. Kurtz. *Markov Processes, characterization and convergence*. John Wiley & Sons, New York, 1986.

[29] S. N. Ethier and T. G. Kurtz. Fleming viot processes in population genetics. *SIAM J. Control Optim.*, **31**(2):345–386, 1993.

[30] R. Ferrière. Adaptive response to environmental threats: evolutionary suicide, insurance and rescue. In *Options, spring 2000*, pages 12–16. IIASA, Laxenburg, Austria, 2000.

[31] R. Ferrière and M. Gatto. Lyapunov exponents and the mathematics of invasion in oscillatory or chaotic populations. *Theor. Pop. Biol.*, **48**:126–171, 1995.

[32] R. Ferrière, S. Méléard, N. Fournier, and N. Champagnat. The mathematics of darwinian evolution: from stochastic individual processes to adaptive dynamics. Prépublication Université Paris 10 – Nanterre No. 04/1. Soumis à J. Math. Biol., 2004.

[33] N. Fournier and S. Méléard. A microscopic probabilistic description of a locally regulated population and macroscopic approximations. *Ann. Appl. Probab.*, sous presse, 2004.

[34] M. I. Freidlin and A. D. Wentzell. *Random Perturbations of Dynamical Systems*. Springer-Verlag, Berlin, 1984.

[35] S. A. H. Geritz. Resident-invader dynamics and the coexistence of similar strategies. *J. Math. Biol.*, sous presse, 2004.

[36] S. A. H. Geritz, M. Gyllenberg, F. J. A. Jacobs, and K. Parvinen. Invasion dynamics and attractor inheritance. *J. Math. Biol.*, **44**:548–560, 2002.

[37] S. A. H. Geritz, J. A. J. Metz, E. Kisdi, and G. Meszena. The dynamics of adaptation and evolutionary branching. *Phys. Rev. Lett.*, **78**:2024–2027, 1997.

[38] S. A. H. Geritz, E. van der Meijden, and J. A. J. Metz. Evolutionary dynamics of seed size and seedling competitive ability. *Theor. Pop. Biol.*, **55**:324–343, 1999.

[39] A. Gragnani, F. Dercole, R. Ferrière, and S. Rinaldi. Coevolution of slow-fast populations: The case of predator-prey interactions. Prépublication, Politecnico di Milano, soumis à Am. Nat., 2003.

[40] M. Gyllenberg, K. Parvinen, and U. Dieckmann. Evolutionary suicide and evolution of dispersal in structured metapopulations. *J. Math. Biol.*, **45**:79–105, 2002.

[41] J. Hofbauer and R. Sigmund. Adaptive dynamics and evolutionary stability. *Appl. Math. Lett.*, **3**:75–79, 1990.

[42] J. Istas. *Introduction aux Modélisations Mathématiques pour les Sciences du Vivant*. Springer-Verlag, Berlin, 2000.

[43] J. Jacod and A. N. Shiryaev. *Limit Theorems for Stochastic Processes.* Springer-Verlag, Berlin, 1987.

[44] A. Joffe and M. Métivier. Weak convergence of sequences of semimartingales with applications to multitype branching processes. *Adv. Appl. Probab.*, **18**:20–65, 1986.

[45] I. Karatzas and S. Shreve. *Brownian Motion and Stochastic Calculus.* Springer-Verlag, New York, 1988.

[46] M. Kimura. Diffusion models in population genetics. J. Appl. Probab., **1**:177–232, 1964.

[47] M. Kimura. A stochastic model concerning the maintenance of genetic variability in quantitative characters. *Proc. Natl. Acad. Sci. USA*, **54**:731–736, 1965.

[48] M. Kimura. *The Neutral Theory of Molecular Evolution.* Cambridge University Press, 1983.

[49] E. Kisdi. Evolutionary branching under asymmetric competition. *J. Theor. Biol.*, **197**:149–162, 1999.

[50] E. Kisdi and S. A. H. Geritz. Evolutionary disarmament in intraspecific competition. *Proc. R. Soc. Lond. B*, **268**:2589–2594, 2001.

[51] A. Lambert. The branching process with logistic growth. À paraître dans *Ann. Appl. Probab.*, 2003.

[52] R. Lande. Quantitative genetic analysis of multivariate evolution, applied to brain: body size allometry. *Evolution*, **33**:402–416, 1979.

[53] R. Law and U. Dieckmann. Moment approximations of individual-based models. In U. Dieckmann, R. Law, and J. A. J. Metz, editors, *The geometry of ecological interactions: simplifying spatial complexity*, pages 252–270. Cambridge University Press, 2000.

[54] R. Law, D. J. Murrell, and U. Dieckmann. Population growth in space and time: spatial logistic equations. *Ecology*, **84**(1):252–262, 2003.

[55] J.-F. Le Galliard, R. Ferrière, and U. Dieckmann. The adaptive dynamics of altruism in spatially heterogeneous populations. *Evolution*, **57**, 2003.

[56] O. Leimar. Multidimensional continuous stability and the canonical adaptive dynamics. In J. A. J. Metz and U. Dieckmann, editors, *Elements of Adaptive Dynamics.* Cambridge University Press, 2001.

[57] P. Marrow, R. Law, and C. Cannings. The coevolution of predator-prey interactions: ESSs and red queen dynamics. *Proc. R. Soc. Lond. B*, **250**:133–141, 1992.

[58] J. Maynard Smith. *Evolution and the Theory of Games.* Cambridge University Press, 1982.

[59] J. A. J. Metz, S. A. H. Geritz, G. Meszéna, F. A. J. Jacobs, and J. S. van Heerwaasden. Adaptive dynamics, a geometrical study of the consequences of nearly faithful reproduction. In S. J. van Strien and S. M. Verduyn Lunel, editors, *Stochastic and Spatial Structures of Dynamical Systems*, pages 183–231, Amsterdam, 1996.

[60] J. A. J. Metz and M. Gyllenberg. How should we define fitness in structured metapopulation models? Including an application to the calculation of evolutionarily stable dispersal strategies. *Proc. R. Soc. Lond. B*, **268**:499–508, 2001.

[61] J. A. J. Metz, R. M. Nisbet, and S. A. H. Geritz. How should we define 'fitness' for general ecological scenarios? *TREE*, **7**:198–202, 1992.

[62] S. D. Mylius and O. Diekmann. The resident strickes back: Invader-induced switching of resident attractor. *J. Theor. Biol.*, **211**:297–311, 2001.

[63] A. Pugliese. On the evolutionary coexistence of parasite strains. *Mathematical Biosciences*, **177–178**:355–375, 2002.

[64] H. Queffélec and G. Zuily. *Éléments d'analyse.* Masson, Paris, 1995.

[65] S. T. Rachev. *Probability Metrics and the Stability of Stochastic Models.* John Wiley & Sons, 1991.

[66] D. A. Rand and H. B. Wilson. Evolutionary catastrophes, punctuated equilibria and gradualism in ecosystem evolution. *Proc. R. Soc. Lond. B*, **253**:137–141, 1993.

[67] E. Renshaw. A survey on stepping-stone models in population dynamics. *Adv. Appl. Probab.*, **18**:515–528, 1973.

[68] D. Revuz and M. Yor. *Continuous Martingales and Brownian Motion.* Springer-Verlag, Berlin, 1991.

[69] E. Seneta and D. Vere-Jones. On quasi-stationary distributions in discrete-time markov chains with a denumerable infinity of states. *J. Appl. Probab.*, **3**:403–434, 1966.

[70] D. Serre. *Matrices: Theory and Applications.* Springer-Verlag, Berlin, 2002.

[71] A. V. Skorohod, F. C. Hoppensteadt, and H. Salehi. *Random Perturbation Methods.* Springer-Verlag, 2002.

[72] S. Wright. Evolution in Mendelian populations. *Genetics*, **16**:97–159, 1931.

Glossaire des notions biologiques

Adaptatif : se dit d'un trait phénotypique soumis a l'action de la sélection naturelle.

Adaptation : valeur ou état d'un phénotype sur lequel s'exerce la sélection naturelle, dans une population dont l'environnement est stationnaire.

Coexistence : maintien à long terme de plusieurs entités (phénotypes, génotypes, espèces) partageant une même ressource spatiale, énergetique ou informationnelle (NB : ce partage n'implique pas necessairement compétition).

Compétition interspécifique : compétition entre individus d'espèces différentes, par opposition à la compétition intra-spécifique.

Compétition intraspécifique : interaction entre individus d'une même espèce, mettant en jeu une ressource limitante (énergie, temps, espace, information) et dont les effets directs sur le succès reproducteur d'un individu sont négatifs.

Densité-dépendance : processus par lequel l'effectif de la population ou d'une partie de celle-ci affecte certaines composantes du cycle de vie de l'individu (la compétition intraspécifique peut en être un mécanisme ; la densité-dépendance est alors négative).

Dimorphique : se dit d'une population dont les individus ne présentent que deux phénotypes distincts, ou deux valeurs différentes du (ou des) trait(s) considéré(s).

Dynamique d'une population : trajectoire temporelle de la population dans un espace d'états ; des variables d'états typiques sont les effectifs de classes structurant la population, ou l'effectif total.

Écologie : science des interactions biotiques, plus particulièrement entre un "être vivant" (être unicellulaire, être multicellulaire, organisme multicellulaire différencié) et son "environnement" (conspécifiques, autres espèces, milieu abiotique).

Environnement : ensemble des conditions extérieures a un être vivant qui sont susceptibles d'influencer son cycle de vie.

Évolution : modification dans le temps et/ou l'espace de la composition génétique d'une population. On distingue évolution neutre et évolution adaptative. Cette dernière se caractérise par l'opération simultanée de deux mécanismes : variabilité héritable du phénotype, différentiel de succès reproducteur des phénotypes en présence.

Fitness : mesure du succès démographique d'un phénotype, défini par la capacité de croissance à long terme d'une population engendrée par un seul individu mutant.

Fixation : processus par lequel un génotype envahit une population de génotypes différents et se substitue complètement à eux (cf. chapitre I section 2).

Gène : portion du génome codant pour une protéine.

Génome : suite des paires de base de l'ADN (A,C,G,T) présentes dans chaque cellule de l'individu.

Génotype : vecteur génétique impliqué dans l'expression des traits phénotypiques considérés. Les composantes de ce vecteur sont les loci, et les états possibles de ces loci sont les allèles.

Invasion : expansion d'une population récemment apparue, à partir d'un effectif très faible.

Monomorphique : se dit d'une population dont tous les individus portent le même phénotype, ou la même valeur du (ou des) trait(s) considéré(s).

Mutation : erreur de copie du génome d'un individu lors d'une division cellulaire.

Mutant : individu dont le génome a subi une mutation par rapport à celui de ses parents.

n-morphique : se dit d'une population dont les individus présentent n phénotypes distincts, ou n valeurs différentes du (ou des) trait(s) considéré(s).

Phénotype : ensemble ou sous-ensemble des caractéristiques d'un individu, hors génome. Le phénotype résulte de l'expression du génotype en interaction avec l'environnement extérieur au génome (du milieu cellulaire à l'habitat de l'individu).

Polymorphique : se dit d'une population non monomorphique.

Protéine : brique élémentaire du métabolisme d'un organisme vivant.

Sélection naturelle : phénomène de tri des phénotypes qui résulte des interactions des individus avec leur environnemement. Les traits phénotypiques soumis a sélection sont ceux qui influencent les paramètres démographiques (donc la *fitness*) des individus.

Trait ou trait phénotypique : composante du phénotype. Les traits phénotypiques soumis a sélection sont dits adaptatifs. Par exemple, la taille corporelle à un âge donné est un trait morphologique ; les vitesses d'absorption de nutriments sont des traits physiologiques ; les paramètres démographiques eux-mêmes (probabilité de survie, vitesse de maturation reproductive, fécondité) définissent les traits d'histoire de vie.

Trait mutant : par abus de langage, désigne un trait phénotypique exprimé par un génotype ayant subi une mutation.

Trait(s) résident(s) : trait(s) présent(s) dans la population au moment d'une mutation, à l'exclusion du trait mutant qui vient d'apparaître.

Valeur sélective : voir *fitness*.

Variabilité phénotypique : présence simultanée de plusieurs phénotypes dans une même population, résultant de la variabilité génétique produite par mutation ou recombinaison, ou de l'expression différentielle d'un même génotype dans des microenvironnements différents (plasticité phénotypique).