Stochastic neutral fractions

&

the effective population size.

R. FORIEN, E. SCHERTZER, Z. TALYIGAS and J. TOURNIAIRE.

(INRAE) (UNI WIEN) (UMLP)

1) STOCHASTIC NEUTRAL FRACTIONS

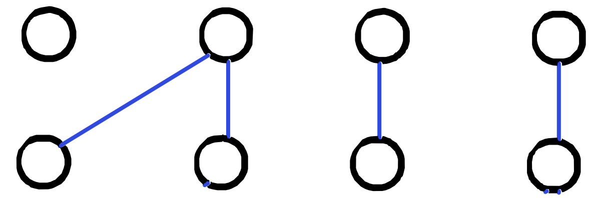
The Wright-Fisher model

- fixed population size $N \in \mathbb{N}$
- discrete time t

<u>ex</u> : $N = 4$.

1) STOCHASTIC NEUTRAL FRACTIONS

The Wright-Fisher model



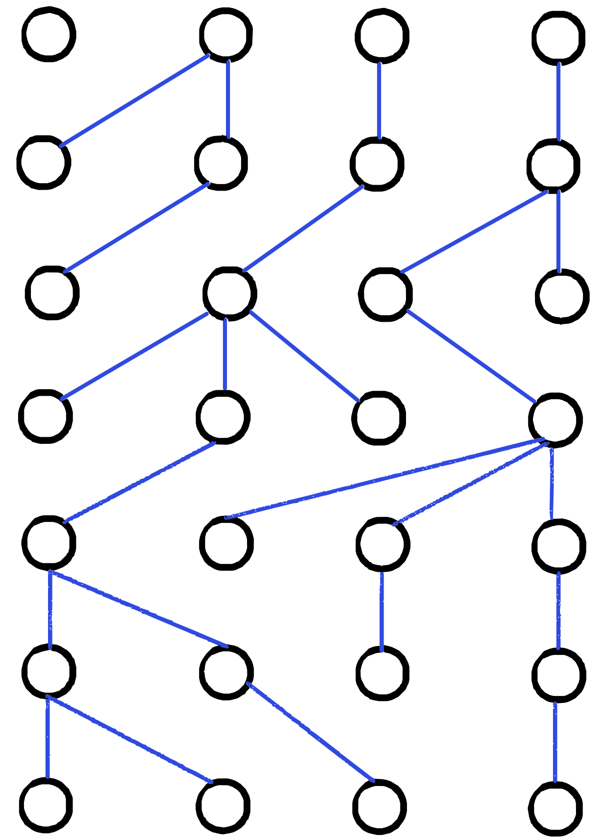
- fixed population size $N \in \mathbb{N}$
- discrete time t
- at each generation $t \to t+1$
  the population is renewed
  each individual uniformly at random, independently of the other individuals

ex: $N = 4$.

1) STOCHASTIC NEUTRAL FRACTIONS

The Wright-Fisher model

- fixed population size $N \in \mathbb{N}$
- discrete time t
- at each generation $t \to t+1$ the population is renewed: each individual uniformly at random, independently of the other individuals



<u>ex</u>: $N = 4$.

1) STOCHASTIC NEUTRAL FRACTIONS

Neutral fractions

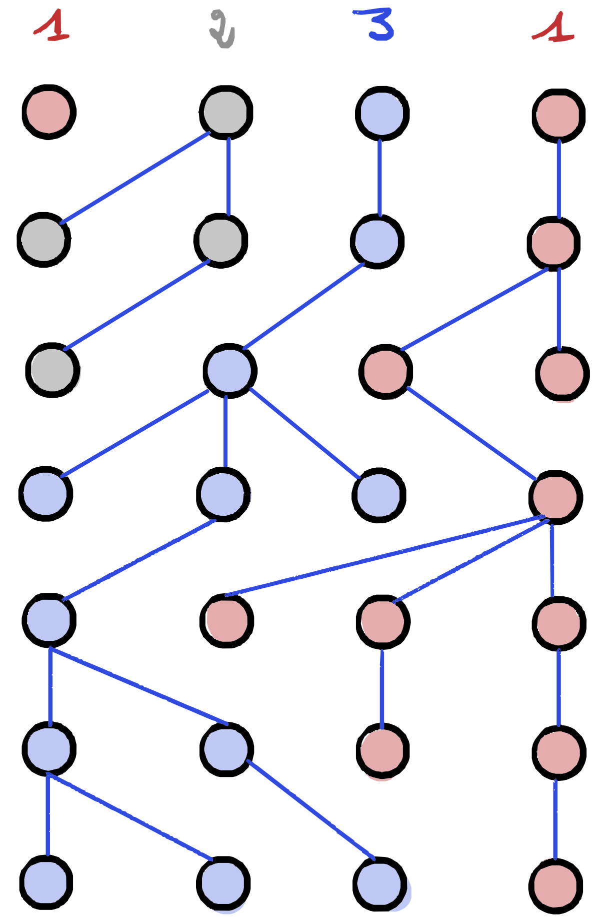
- at $t=0$, each individual is assigned an allele / colour indexed by

$$\underbrace{[K]}_{\sim} = \{1, 2, \dots, K\}$$

$K \in \mathbb{N}$ number of alleles

- $t \in \mathbb{N}$, each individual inherits the allele of its parent

↳ alleles are neutral



$X_t^{N,1} = 1/4$ $\quad X_t^{N,2} = 0$

$X_t^{N,3} = 3/4$

For $i \in [K]$, $t \in \mathbb{N}_0$

$$X_t^{N,i} = \text{frequency of allele } i \text{ at generation } t.$$

Asymptotic behaviour of $(X_t^{i,N}, i \in [K], t \in \mathbb{N})$ as $N \to +\infty$?

1) STOCHASTIC NEUTRAL FRACTIONS

$X_t^{N,i}$: frequency of type i at time t. / size of the i-th fraction

Assume that the initial frequencies converge to a non degenerate limit $(X_0^{N,i} ; i \in [K]) \Rightarrow x_0 \in [0,1]^K$. Then, as $N \to +\infty$,

$$(X_{\lfloor tN \rfloor}^{N,i} ; i \in [K], t \geq 0) \Longrightarrow (Z_t^i ; i \in [K], t \geq 0)$$

$(K-1)$ dimensional standard Wright-Fisher diffusion (*)

(*) diffusion on the simplex $\{(z_1, \dots, z_K) \in [0,+\infty)^K : \sum z_k = 1\}$ with infinitesimal generator

$$\mathcal{L} f(z) = \frac{1}{2} \sum_{1 \leq i,j \leq K} (\delta_{ij} z_i - z_i z_j) \frac{\partial^2}{\partial z_i \partial z_j} f(z)$$

1) STOCHASTIC NEUTRAL FRACTIONS

WF model : Well-mixed population

equal chances of reproductive success

"fluctuations in the allelic composition only due to genetic drift"

(?) Impact of ecological constraints on genetic compositions

1) STOCHASTIC NEUTRAL FRACTIONS

WF model : Well-mixed population
equal chances of reproductive success] unstructured population

"fluctuations in the allelic composition only due to genetic drift"

(?) Impact of ecological constraints on genetic compositions

structured population

* individuals carry types (spatial locations / fitness)
* each type corresponds to a different ability to reproduce
* non-local reproduction (migrations / mutations)

1) STOCHASTIC NEUTRAL FRACTIONS

WF model : Well-mixed population
equal chances of reproductive success] unstructured population

"fluctuations in the allelic composition only due to genetic drift"

(?) Impact of ecological constraints on genetic compositions

structured population

* individuals carry types (spatial locations / fitness)
* each type corresponds to a different ability to reproduce
* non-local reproduction (migrations / mutations).

Backward-in-time approach

structured coalescents $\xrightarrow[\text{timescales} \quad N \to \infty]{\text{Separation of}}$ Kingman's coalescent on timescale N_e

↓ duality

Wright-Fisher diffusion on timescale N_e

1) STOCHASTIC NEUTRAL FRACTIONS

N_e : effective population size.

"size of the WF diffusion that experiences the same level of genetic drift as the structured population"

Questions ?

* What if there is no separation of timescales ?

  or the population size is not fixed ?

* can we use the backward approach ? Not so easy !

GOAL : Define a general framework to work directly with neutral fractions in structured models.

1) (STOCHASTIC) NEUTRAL FRACTIONS IN THE FKPP EQUATION

The FKPP equation

$$\boxed{\partial_t u = \frac{1}{2}\partial_{xx} u + r(u)u \qquad (PDE_1)}$$

$u(t,x)$ density of individuals at time t at $x \in \mathbb{R}$ = type

$r(u)$: per capita growth rate $\Rightarrow$ density dependent

1) (STOCHASTIC) NEUTRAL FRACTIONS IN THE FKPP EQUATION

The FKPP equation

$$\boxed{\partial_t u = \frac{1}{2}\partial_{xx} u + r(u)u \qquad \text{(PDE}_1\text{)}}$$

$u(t,x)$ density of individuals at time t at $x \in \mathbb{R}$ = Type

$r(u)$: per capita growth rate $\Rightarrow$ density dependent

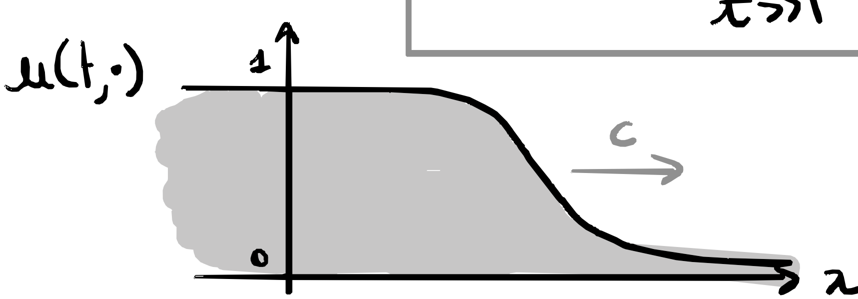
Asymptotic behaviour

There exists a minimal speed c, and an associated profile $\tilde{h}$ s.t

under suitable i.c.,

$$\boxed{u(t,x) \underset{t \gg 1}{\approx} \tilde{h}(x-ct)}$$

travelling wave/front solution to (PDE$_1$)



1) (STOCHASTIC) NEUTRAL FRACTIONS IN THE FKPP EQUATION

The FKPP equation

$$\boxed{\partial_t u = \frac{1}{2}\partial_{xx} u + r(u)\,u \qquad (\mathrm{PDE}_1)}$$

$u(t,x)$ density of individuals at time t at $x \in \mathbb{R}$

How to define neutral fractions ?

Hallatschek and Nelson 2008

Roques, Garnier, Hamel and Klein 2012

initial condition : $u(0,x) = u_0(x) = \sum_{k=1}^{K} v_0^k(x)$

fraction dynamics : for all $k \in [K]$, v^k solution to the PDE

$$\boxed{\begin{cases} \partial_t v^k = \frac{1}{2}\partial_{xx} v^k + r(u)\,v^k \\ v^k(0,x) = v_0^k(x) \end{cases} \qquad (\mathrm{PDE}_2)}$$

1) (STOCHASTIC) NEUTRAL FRACTIONS IN THE FKPP EQUATION

The FKPP equation

$$\boxed{\partial_t u = \frac{1}{2}\partial_{xx} u + r(u)u \qquad (\mathrm{PDE}_1)}$$

$u(t,x)$ density of individuals at time t at $x \in \mathbb{R}$

How to define neutral fractions ?

Hallatschek and Nelson 2008

Roques, Garnier, Hamel and Klein 2012

initial condition : $u(0,x) = u_0(x) = \sum_{k=1}^{K} v_0^k(x)$

fraction dynamics : for all $k \in [K]$, v^k solution to the PDE

$$\boxed{\begin{cases} \partial_t v^k = \frac{1}{2}\partial_{xx} v^k + r(u) v^k \\ v^k(0,x) = v_0^k(x) \end{cases} \qquad (\mathrm{PDE}_2)}$$

$\sum v^k$ solution to (PDE_1)

AND

same dispersal + same growth $\Rightarrow$ neutral fractions

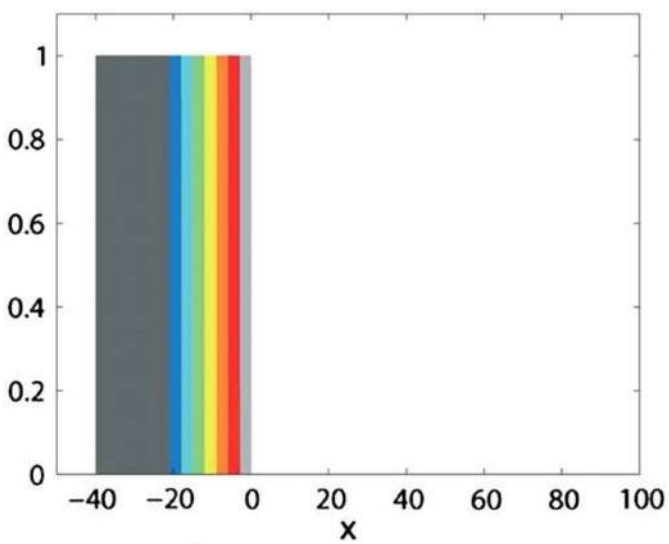
1) (STOCHASTIC) NEUTRAL FRACTIONS IN THE FKPP EQUATION

$$\partial_t u = \frac{1}{2}\partial_{xx} u + r(u)\, u \quad \text{(PDE}_1\text{)}$$ Total population size

$$\partial_t v^k = \frac{1}{2}\partial_{xx} v^k + r(u)\, v^k \quad k \in [K] \quad \text{(PDE}_2\text{)}$$ neutral fractions

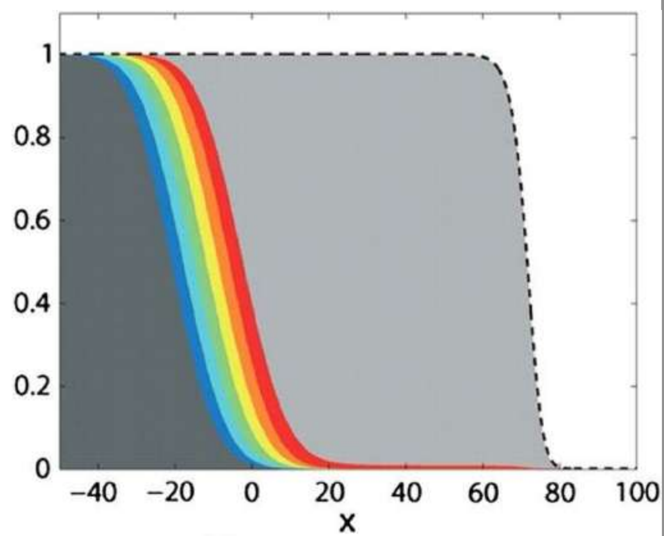
Roques, Garnier, Hamel and Klein 2012

$t = 0$



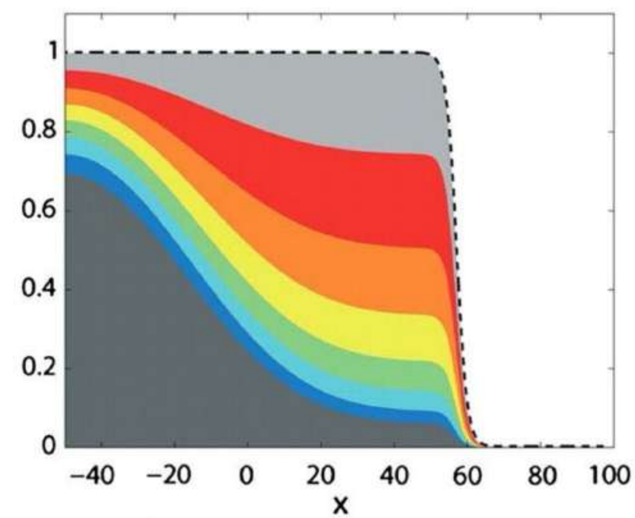
EX 1

$r(u) = (1-u)$



EX 2 : Allee effects

$r(u) = (1-u)(u-u^*)$



1) {STOCHASTIC} NEUTRAL FRACTIONS IN THE FKPP EQUATION

Noisy FKPP equation :

$$\boxed{\partial_t u^N = \frac{1}{2}\partial_{xx} u^N + r(u^N)\,u^N + \sqrt{\frac{u^N}{N}}\,\dot{W}} \quad (SPDE_1)$$

N : large demographic parameter local population density

$\dot{W}$: white noise

1) {STOCHASTIC} NEUTRAL FRACTIONS IN THE FKPP EQUATION

Noisy FKPP equation :

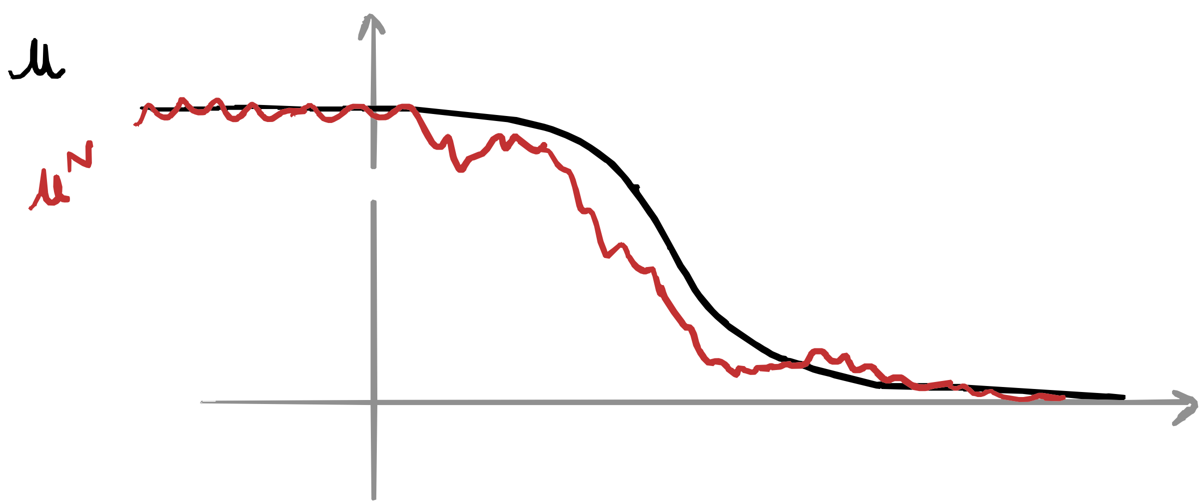
$$\boxed{\partial_t u^N = \frac{1}{2}\partial_{xx} u^N + r(u^N)\, u^N + \sqrt{\frac{u^N}{N}}\,\dot{W} \quad (\mathrm{SPDE}_1)}$$

N : large demographic parameter   local population density

$\dot{W}$: white noise

1st step : fluctuations of the total population u^N around the deterministic limit $(N \to +\infty)$?

↳ HARD PROBLEM !



2) THE MODEL

$E = (x_1, \dots, x_d)$ finite type space

We consider the d-dimensional stochastic differential equation

$$\boxed{dU_t^N = b(U_t^N)\,dt + \frac{1}{\sqrt{N}}\, a(U_t^N)\, dW_t} \quad (SDE_1)$$

$U_t \in \mathbb{R}^E$, $\forall x \in E$ $(U_t)_x$ "number" of individuals with type x

$b : \mathbb{R}^E \longrightarrow \mathbb{R}^E$ growth of the system

$a : \mathbb{R}^E \longrightarrow M_d(\mathbb{R})$ noise correlation matrix

W_t d-dimensional Wiener process.

a and b are smooth enough to have existence + uniqueness of weak sol.

How to define neutral fractions in this population model?

2) THE MODEL

Neutral fractions. Can we define $\forall K \in \mathbb{N}$ $V_t^1, \dots, V_t^K : \mathbb{R}^+ \times \mathbb{R}^E \longrightarrow \mathbb{R}^E$ s.t.

1) $\sum_{k=1}^{K} V_t^k$ is solution to (SDE$_1$) "fractions"

2) the fractions are exchangeable : $\forall \sigma \in S_K$

$$\boxed{\left(V_t^{\sigma(1)}, \dots, V_t^{\sigma(K)}\right) \overset{\mathcal{L}}{=} \left(V_t^1, \dots, V_t^K\right)}$$

"neutral"

3) the fractions are consistent

$$\boxed{\left(V_t^1, \dots, V_t^{K-1} + V_t^K\right) \overset{\mathcal{L}}{=} \left(V_t^1, \dots, V_t^{K-1}\right)}$$

1) + 2) + 3) : the system is ∞-decomposable

2) THE MODEL

Sufficient conditions for (SDE$_1$) to be ∞-decomposable

(1) DRIFT TERM

$\exists F : \mathbb{R}^E \to \mathcal{L}(\mathbb{R}^E, \mathbb{R}^E) : \forall U \in \mathbb{R}^E,$ $\boxed{b(U) = F(U)\,U}$

remark : density-dependent growth rate $\approx$ FKPP $r(u)\,u$

(2) DEMOGRAPHIC FLUCTUATIONS

$\exists C : \mathbb{R}^E \to \mathcal{L}(\mathbb{R}^E, \mathcal{L}(\mathbb{R}^E, \mathbb{R}^E))$ and $\sigma : \mathbb{R}^E \times \mathbb{R}^E \to M_{d,K}(\mathbb{R})$.

such that $\forall U_1, U_2 \in \mathbb{R}^E,$

$$\boxed{aa^*(U_1) = C(U_1).U_1 \quad \text{and} \quad \sigma\sigma^*(U_1, U_2) = C(U_1)\,U_2}$$

Neutral fractions $(V_t^1, \dots, V_t^K)$ solutions to

$$\boxed{dV_t^{N,k} = F\left(\sum V_t^{N,k}\right) V_t^{N,k} + \frac{1}{\sqrt{N}}\, \sigma\left(\sum V_t^{N,k}, V_t^{N,k}\right) dW_t^k \;;\; k \in [K] \quad (\text{SDE}_2)}$$

(W_t^k) independent d-dimensional Wiener processes

$$\boxed{dV_t^{N,k} = F\left(\sum V_t^{N,k}\right) V_t^{N,k} + \frac{1}{\sqrt{N}}\, \sigma\left(\sum V_t^{N,k}, V_t^{N,k}\right) dW_t^k \;;\; k \in [K]} \quad (SDE_2)$$

Katzenberger 90':

Assume that the deterministic system

$$\boxed{dV_t^k = F\left(\sum_{k=1}^{K} V_t^k\right) V_t^k \quad k \in [K],\ t>0} \quad (ODE_2)$$

has an attractive stable manifold Γ.

Then, as $N \to +\infty$, on a suitable time scale, $(V_t^{N,k};\ k \in [K])$ converges to a diffusion on Γ

$$\boxed{dV_t^{N,k} = F\left(\sum V_t^{N,k}\right) V_t^{N,k} + \frac{1}{\sqrt{N}}\, \sigma\left(\sum V_t^{N,k}, V_t^{N,k}\right) dW_t^k \;;\; k \in [K]} \quad (SDE_2)$$

Katzenberger 90':

Assume that the deterministic system

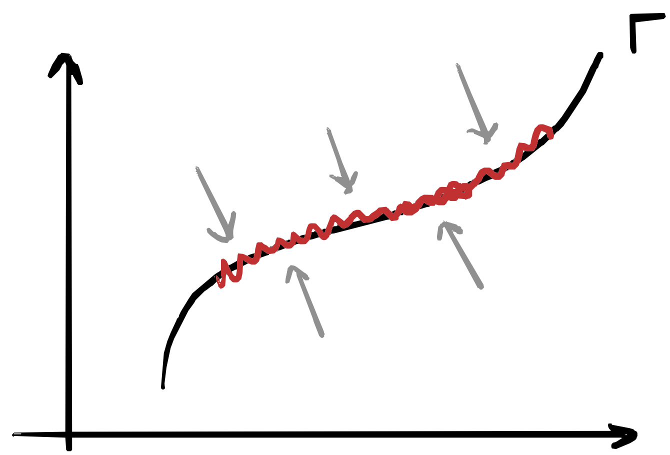
$$\boxed{dV_t^k = F\left(\sum_{k=1}^{K} V_t^k\right) V_t^k \qquad k \in [K],\ t > 0} \quad (ODE_2)$$

has an attractive stable manifold Γ.

Then, as $N \to +\infty$, on a suitable time scale, $\left(V_t^{N,k} ; k \in [K]\right)$ converges to a diffusion on Γ

⊕ (SDE_2) does not necessarily satisfies the ∞-decomposability condition

⊖ the limiting diffusion is not explicit.



Assumptions

$$du_t^N = b(u_t^N)\,dt + \frac{1}{\sqrt{N}}\,a(u_t^N)\,dW_t \qquad \text{(SDE}_1\text{)}$$

(∞-decomposability) $\quad b(u) = F(u)u$ and $aa^*(u) = C(u)u$.

(regularity) $\quad F$ and C are smooth

(equilibrium) $\quad \frac{du_t}{dt} = b(u_t)$ (ODE$_1$) has an exponentially locally attractive equilibrium $\tilde{h} > 0$ $\quad [b(\tilde{h}) = F(\tilde{h})\tilde{h} = 0]$

(PF) $\quad F(\tilde{h})$ is Metzler and $\exists t_0 > 0 : \forall x, y \in E \ (e^{t_0 F(\tilde{h})})_{xy} > 0$

↳ $\tilde{h}$ is the right principal eigenvector associated to $F(\tilde{h})$

↳ there exists an associated positive left eigenvector s.t.

$$\langle h . \tilde{h} \rangle = 1$$

<u>Remark</u>: $\langle \tilde{h} . 1 \rangle \neq 1$ carrying capacity of the system.

3) RESULT

$$\boxed{dV_t^{N,k} = F\left(\sum V_t^{N,k}\right) V_t^{N,k} + \frac{1}{\sqrt{N}}\, \sigma\left(\sum V_t^{N,k}, V_t^{N,k}\right) dW_t^k \;, \quad k \in [K]} \quad (SDE_2)$$

Stable manifold : $\Gamma^K = \left\{ \theta \otimes \tilde{h} \,,\ \sum_{k=1}^{K} \theta^k = 1 \,,\ \theta^k \geq 0 \right\}$

Scaling parameter $\quad \Sigma^2 = \langle h, aa^*(\tilde{h})\, h \rangle$

> **THEOREM** : Under natural assumptions on the initial condition,
>
> $$\left(V^{N,k}_{tN/\Sigma^2} \;;\ k \in [K],\ t \in [0,T] \right) \underset{N\infty}{\Longrightarrow} \left(\theta_t \otimes \tilde{h},\ t \in [0,T] \right)$$
>
> where (θ_t) is a standard $(K-1)$-dimensional Wright–Fisher diffusion

$N_e = \dfrac{N}{\Sigma^2}$ $\qquad$ effective population size.

3) RESULT

$$\boxed{du_t^N = b(u_t^N)\,dt + \frac{1}{\sqrt{N}}\,a(u_t^N)\,dW_t} \quad (SDE_1)$$

$$\boxed{dv_t^{N,k} = F\left(\sum v_t^{N,k}\right) v_t^{N,k} + \frac{1}{\sqrt{N}}\,\sigma\left(\sum v_t^{N,k}, v_t^{N,k}\right) dW_t^k \;;\; k \in [K]} \quad (SDE_2)$$

<u>EX</u> $\approx$ Noisy FKPP equation

$$a(u) = \begin{pmatrix} \sqrt{u_1} & & 0 \\ & \ddots & \\ 0 & & \sqrt{u_d} \end{pmatrix}$$

then

$$C(u)\,u = \begin{pmatrix} u_1 & & 0 \\ & \ddots & \\ 0 & & u_d \end{pmatrix} \quad \text{and} \quad \sigma(u,w) = \begin{pmatrix} \sqrt{w_1} & & 0 \\ & \ddots & \\ 0 & & \sqrt{w_d} \end{pmatrix}$$

$$\Sigma^2 = \sum_{x \in E} h_x^2\,\tilde{h}_x \qquad (N_e = N/\Sigma^2)$$

reproductive variance of the structured system.

$\tilde{h}_x$: mean number of children of an individual with type x.

5) What do we expect for stochastic FKPP equations ?

$$\boxed{\partial_t u^N = \frac{1}{2}\partial_{xx} u^N + r(u^N)u^N + \sqrt{\frac{u^N}{N}}\,\dot{W}(t)} \quad \text{(SPDE}_1\text{)}$$

(equilibrium) there exists a stable profile $\tilde{h}$ such that

$$\| u^\infty(t, \cdot + ct) - \tilde{h}(\cdot) \|_\infty \leq C e^{-\lambda t}$$ [√ Allee case]

(finite variance) $$\Sigma^2 = \int_{\mathbb{R}} h(x)^2 \tilde{h}(x)\, dx < +\infty$$ [√ Allee case]

Then $\exists (\xi_t^N)_{t\geq 0}$ s.t

$$\boxed{\begin{array}{l} \sup_{t\in[0,TN]} \| u^N(t,\cdot) - \tilde{h}(\cdot - \xi_t^N) \| \xrightarrow[N\infty]{} 0 \\ (\xi_t^N) \text{ converges to a Brownian motion} \end{array}}$$

Etheridge, Forien, Penington 2025+

and

$$\boxed{\sup_{t\in[0,TN]} \sup_{k\in[K]} \| v^{k,N}(t,\cdot) - \underbrace{\theta^{k,N}(t)}_{\text{WF ?}} \tilde{h}(\cdot - \xi_t^N) \| \xrightarrow[N\infty]{} 0}$$

4) SKETCH OF PROOF

notation $\quad V_t^N = (V_t^{N,1} \dots V_t^{N,K})$ composition matrix $\in M_{d,K}(\mathbb{R}^+)$

Katzenberger projection π

let $\Phi : [0,+\infty) \times M_{d,K}(\mathbb{R}^+) \longrightarrow M_{d,K}(\mathbb{R}^+)$

$$(t, v_0) \longmapsto \Phi(t, v_0) = v_t$$

where v_t is solution of the ODE

"flow of the deterministic dynamics"

$$\boxed{\begin{array}{l} \dot{v}_t = F\left(\sum_{j=1}^{K} v_t^j\right) v_t \\ \text{initial condition } v_0 . \ (ODE_2) \end{array}}$$

Define $\pi(v_0) = \lim_{t \to +\infty} \Phi(t, v_0)$

Fact 1 : $\pi(v_0) \in \Gamma^K \Rightarrow \pi(V_t^N)$ natural candidate for the limiting diffusion.

$\lightning$ It usually requires a lot of info on the flow to work with π.

4) SKETCH OF PROOF

notation $V_t^N = (V_t^{N,1} \dots V_t^{N,k})$ composition matrix $\in M_{d,k}(\mathbb{R}^+)$

Katzenberger projection $\pi(\cdot) = \lim_{t \to +\infty} \Phi(t, \cdot)$

Fact 1 : $\pi(v_0) \in \Gamma^K$ $\Rightarrow$ $\pi(V_t^N)$ natural candidate for the limiting diffusion.

Fact 2 $\exists H : \mathbb{R}^d \longrightarrow \mathbb{R}^d$ such that

$\forall v_0 \in M_{d,k}(\mathbb{R}^+)$ $\boxed{\pi(v_0)^j = \langle H(u_0), v_0^j \rangle \tilde{h}}$ $u_0 = \sum v_0^j$

Proof. decomposability $\Rightarrow$ each column is solution of a linear ODE

$\rightsquigarrow$ Reduction of the problem's dimension

Next steps of the proof : 1) show that $V_t^N \approx \pi(V_t^N)$ STABILITY.

2) show that $(\langle H(U_t^N), V_t^{N,j} \rangle)_j \Rightarrow WF$

ITÔ CALCULUS.

4) SKETCH OF PROOF

Stability: show that $\pi(V_t^N) \approx V_t^N$

$\mathbb{R}^d \ni w$ right e.v

Define the Perron-Frobenius projection $P : w \longmapsto \langle w, h \rangle \tilde{h}$

left e.v.

0) Stability of the total pop. size $U_t^N \approx \tilde{h}$

1) Stability of Perron-Frobenius dynamic projections

$$\left(P(V_t^{N,i})\right)_i \approx V_t^N$$

Proof. Standard Perron-Frobenius theory + martingale arguments.

2) Comparison with Katzemberger projections

Recall that $\pi(V_t^N)^i = \langle H(U_t^N), V_t^{N,i} \rangle \tilde{h}$.

$$\| H(v_0) - h \| \leq C \| v_0 - \tilde{h} \|$$