# On the stability of genetic algorithms - a variational approach -

Oxford, 2 November 2003

Wilhelm Stannat

Fakultät für Mathematik

Universität Bielefeld

email: stannat@mathematik.uni-bielefeld.de

www.mathematik.uni-bielefeld.de/~stannat

# References


- [St 03] On the convergence of genetic algorithms - a variational approach, Bielefeld 2003, submitted.
- [St, PTRF 02] Long-time behaviour and regularity properties of transition semigroups of Fleming-Viot processes, Probab. Theory Relat. Fields, Vol. 122, No. 3, 431-469, 2002.
- [St, AOP 00] On the Validity of the Log-Sobolev inequality for symmetric Fleming–Viot operators. Ann. Probab., Vol. 28, No. 2, 667-684, 2000.

# Contents

# 1. A genetic algorithm with mutation and selection

- $S$ - measurable space *(type space)*
- $\mathcal{M}_1(S) =$ all probability measures on $S$ *(all poss. distr. of types)*
- $A$ - generator of MP $((P_x), (X(t))$ on $S$ *(mutation)*

  $$p_t f(x) := E_x \left[ f \left( X \left( t \right) \right) \right]$$

- $\sigma : S \times \mathcal{M}_1(S) \to R_+$ *(fitness function)*

Fix $n \geq 2$ and consider $n$ independent copies of $(X(t))$

$$X_1^{(n)}(t), \ldots, X_n^{(n)}(t)$$

## Pair-interaction

- each individual has exponential life-time
- once an individual dies, it is replaced by an individual of the type of one of the remaining $n-1$ individuals

(proportional to the fitness $\sigma(x,\mu)$ of type $x$ given the pop. $\mu$)

gives the

*n-particle Moran process with mutation $A$ and selection $\sigma$*

$$Y_1^{(n)}(t),\ldots,Y_n^{(n)}(t)\ ,t\geq 0$$

Generator

$$L^{(n)}f(x) := \underbrace{A^{(n)}f(x)}_{\text{indep. mutation}} + \frac{1}{n}\sum_{i<j}\sigma\left(x_i,\underbrace{\frac{1}{n}\sum_{k=1}^{n}\delta_{x_k}}_{\text{emp. distr.}}\right)\Big(\underbrace{\Phi_{ij}^{(n)}}_{\text{repl. operator}}f(x)-f(x)\Big)$$

where $\Phi_{ij}^{(n)}f(x_1,\ldots,x_n) = f(x_1,\ldots,x_{j-1},x_i,x_{j+1},\ldots,x_n)$.

**Examples:**

1. $\sigma(x,\mu) = \sigma(x)$ (noninteracting case)

2. $\sigma(x,\mu) = \int_{S^{n-1}} k(x_1,\ldots,x_{n-1},x)\,\mu(dx_1)\ldots\mu(dx_{n-1})$

   for some $k \in \mathcal{B}_{b,sym}(S^n)_+$

3. more general:

   $\sigma(x,\mu) = \frac{\partial G_\sigma}{\partial \delta_x}(\mu) := \frac{d}{ds} G_\sigma(\mu + s\delta_x)_{|s=0}$ for some potential

$$G_\sigma : \mathcal{M}_+(S) \to \ R\,.$$

**LLN** (del Moral/Miclo 00): $\sigma(x,\mu) = \sigma(x)$ bounded , $A$ "regular", $(Y_i^{(n)}(0))$ iid $(\mu)$, then

$$\frac{1}{n}\sum_{i=1}^{n} \delta_{Y_i^{(n)}(t)} \stackrel{w}{\to} \psi_t(\mu)$$

where

$$\dot{\psi}_t(\mu) = \hat{A}\psi_t(\mu) + \sigma \ \psi_t(\mu) - \int \sigma \, d\psi_t(\mu) \ \psi_t(\mu) \, , \psi_0(\mu) = \mu$$

Explicit representation of the solution in the **noninteracting** case

$$\langle f, \psi_t(\mu) \rangle = \frac{E_\mu \left[ f(X(t)) \exp \left( \int_0^t \sigma(X(s)) \, ds \right) \right]}{E_\mu \left[ \exp \left( \int_0^t \sigma(X(s)) \, ds \right) \right]}$$

$$\uparrow$$

$$\text{ren. Feynman} - \text{Kac semigroup}$$

# 2. Long-time behaviour

Changes to the population come from two opposing sources:

- selection - encourages conformity by favouring individuals with dominant type

  in fact: if $A = 0$, then $\psi_t(\mu)$ increases the mean fitness

  $$\int \sigma \, d\psi_t(\mu) \, ,$$

  hence concentrates more and more on

  $$M(\mu) = \{x \in S : \sigma(x) = \mu - ess\, sup\, \sigma\}$$

- mutation - preserves (or even improves) genetic diversity by adding fresh variation

**The variational approach**

Assume: $A$ $\nu$-symmetric (for some $\nu \in \mathcal{M}_1(S)$)

Consider $\mu = h\nu$ with $h \in L^2(\nu)$

**Noninteracting case** $\sigma(x,\mu) = \sigma(x)$

A substitute for the mean fitness is the energy

$$\lambda(h) := -\int Ah\,h\,d\nu - \int \sigma h^2\,d\nu; h \in \underbrace{B_1^{L^2(\nu)}(0)_+}_{= \text{unit ball in } L^2(\nu)}$$

Consider the $L^2$-renormalization

(1) $\quad \dot{\eta}_t(h) := A\eta_t(h) + \sigma\eta_t(h) - \lambda(\eta_t(h))\eta_t(h)\,, \eta_0(h) = h$

so that

$$\frac{\eta_t(h)\nu}{\int \eta_t(h)\,d\nu} = \psi_t(h\nu)$$

**Then**

- $\lambda(\eta_t(h))$ decreases in time
- critical points of $\lambda$ coincide with stationary points of $\eta_t$

## Interacting case

Consider a Frechet differentiable potential

$$G_\sigma : L^1(\nu) \to R$$

with differential $\sigma$ such that

$$h \mapsto \sigma(\cdot, h^2)h \qquad\qquad \text{loc. Lipschitz in } L^2(\nu)$$

Then (1) has a unique classical solution for all $h \in B_1^{L^2(\nu)}(0) \cap D(-(-A)^{\frac{1}{2}})$

Define

$$L_1 := \sup_{h,g \in B_1^{L^2(\nu)}(0)_+} \|\sigma(\cdot, h^2) - \sigma(\cdot, g^2)\|_\infty$$

$$L_2 := 2 \sup_{\substack{h,g \in B_1^{L^2(\nu)}(0)_+ \\ h \neq g}} \frac{\int \left(\sigma(\cdot, h^2) - \sigma(\cdot, g^2)\right) h(h-g)\, d\nu}{\|h-g\|^2_{L^2(\nu)}}$$

**Theorem 1** [St 03] Let the function $h_* \geq 0$ be a critical point of

$$F(h) := -\int Ah\, h\, d\nu - G_\sigma(h^2)\ , h \in B_1^{L^2(\nu)}(0)_+\,,$$

and assume that the ground state transform

$$A_* f = \frac{1}{h_*}\left(A + \sigma + \lambda(h_*)\right)(h_* f) = \frac{1}{h_*}\left(A(h_* f) - Ah_*\, f\right)$$

has a mass gap of size $\kappa_*$ (in $L^2(h_*^2\nu)$), then

$$\|\eta_t(h) - h_*\|_{L^2(\nu)}^2 \leq \frac{1}{\int hh_*\, d\nu} e^{(L_1+L_2-\kappa_*)t} \|h - h_*\|_{L^2(\nu)}^2\,.$$

**Remark** In the noninteracting case $L_1 = L_2 = 0$, so that

$$\|\eta_t(h) - h_*\|_{L^2(\nu)}^2 \leq \frac{1}{\int hh_*\, d\nu} e^{-\kappa_* t} \|h - h_*\|_{L^2(\nu)}^2\,.$$

- Theorem 1 implies: If $\kappa_* > L_1 + L_2$ then $F$ has exactly one critical point which coincides with the minimizer of $F$
- How to understand the rate of convergence:

In the noninteracting case

$$\frac{d}{dt}\left(\frac{1}{2}\|\eta_t(h) - h_*\|_{L^2(\nu)}^2\right) \leq \int A_*\left(\frac{\eta_t(h) - h_*}{h_*}\right)\left(\frac{\eta_t(h) - h_*}{h_*}\right) h_*^2 \, d\nu + \underbrace{(\lambda(\eta_t(h)) - \lambda(h_*))}_{\geq 0}\frac{1}{2}\|\eta_t(h) - h_*\|_{L^2(\nu)}^2$$

and

$$\int_0^\infty \lambda(\eta_t(h)) - \lambda(h_*) \, dt \leq -\log\left(\int h h_* \, d\nu\right)$$

## A priori estimates on $\kappa_*$ and $h_*$

- well-known: if $h_*$ is the unique ground state of $A+\sigma$ then $\kappa_* = \lambda_1 - \lambda_0$ where

$$\lambda_0 = \lambda(h_*) = \inf_{h \in B_1^{L^2(\nu)}(0)} \lambda(h)$$

and

$$\lambda_1 = \inf_{\substack{h \in B_1^{L^2(\nu)}(0) \\ \int h h_* \, d\nu = 0}} \lambda(h)$$

- if $A$ has a mass gap of size $\kappa$, then

$$\kappa_* \geq \frac{\kappa}{\|h_*\|_\infty^2 \|h_*^{-1}\|_\infty^2}\,.$$

## Examples

1. If $p_t(x,\cdot) \ll \nu$ and $C^{-1}e^{-\frac{B}{t}} \le \frac{dp_t(x,\cdot)}{d\nu} \le Ce^{\frac{B}{t}}$ then

$$C^{-2}e^{-2\sqrt{2B\mathrm{osc}(\sigma)}} \le h_* \le C^2 e^{2\sqrt{2B\mathrm{osc}(\sigma)}}$$

   where

$$\mathrm{osc}(\sigma) := \sup_{h \in B_1^{L^2(\nu)}(0)} \sup_{x,y \in S} \sigma(x,h^2) - \sigma(y,h^2)$$

2. If

$$Af(x) = \int f(y) - f(x)\, K(x,dy)$$

   with

$$\kappa\ \nu \otimes K \le \nu \otimes \nu \le \kappa^{-1}\ \nu \otimes K$$

   then

$$\frac{\int h_*\, d\nu}{\kappa^{-1}(\kappa^{-1} + \mathrm{osc}(\sigma))} \le h_* \le \begin{cases} \frac{\int h_*\, d\nu}{\min_{x\in S} \nu(x)} & \text{if } |S| < \infty \\ \\ \frac{\kappa^{-2}\int h_*\, d\nu}{1-\kappa^{-1}\mathrm{osc}(\sigma)} & \text{if } \mathrm{osc}(\sigma) < \kappa \end{cases}$$

**Special case of 2** Parent Independent Mutation

(see Kingman 78 for the discrete time analogon)

$$Af(x) = \frac{\theta}{2}\int f(y) - f(x)\,\nu(dy)\,, \theta > 0$$

If $\alpha_*$ is a solution of

$$\int \frac{1}{\alpha_* + 1 - \frac{2}{\theta}\sigma}\,d\nu = 1$$

with $\alpha_* + 1 - \frac{2}{\theta}\sigma \geq 0$, then

$$h_* = \frac{1}{\alpha_* + 1 - \frac{2}{\theta}\sigma} \,/\, \left(\int \frac{1}{(\alpha_* + 1 - \frac{2}{\theta}\sigma)^2}\,d\nu\right)^{\frac{1}{2}}$$

is a stationary point of $\eta_t$ and

$$A_* f = \frac{\theta}{2}\int f(y) - f(x)\frac{h_*(y)}{h_*(x)}\,\nu(dy)$$

has a mass gap of size $\kappa_* \geq \frac{\theta}{2\|h_*\|_\infty^2}$

**Present work**

Extensions of Theorem 1 to the case of

1. unbounded (possibly negative) selection
2. infinite measure $\nu$
3. noninteracting time-dependent case

**Motivation for 3** long-time behaviour of the pathwise filter equation

## Example: quasi-stationary case

Let

$$\sigma : R_+ \times \mathcal{M}_1(S) \to R_+$$

be such that

$$M := \int_0^\infty \operatorname{osc}\left(\sigma(t,\cdot) - \sigma_\infty\right)\, dt < \infty$$

for some $\sigma_\infty : S \to R_+$

In this case

$$\|\eta_t(h) - \eta_t(g)\|_{L^2(\nu)}^2 \leq \frac{e^{12M}}{\int hh_*\, d\nu \int gh_*\, d\nu} e^{-\int_0^t \kappa_*(s)\, ds} \|h - g\|_{L^2(\nu)}^2$$

# 3. The Fleming-Viot Process

From now on

- $S$ compact metric space
- $A$ generator of a Feller process $((P_x), (X(t)))$ on $S$

  (recall $p_t f(x) := E_x[f(X(t))]$)

Back to the approximating Moran particle process

Consider n-particle Moran process with mutation $A$ and neutral selection

$$Y_1^{(n)}(t), \ldots, Y_n^{(n)}(t)$$

with generator

$$L^{(n)} f(x) = A^{(n)} f(x) + \sum_{i<j} \left( \Phi_{ij}^{(n)} f(x) - f(x) \right)$$

(Note different scaling in $n$)

**The FV-Limit** (Fleming/Viot 79, Donnelly/Kurtz 99)

If $(Y_i^{(n)}(0))$ iid $(\mu)$ (w.r.t. $P_\mu$), then

$$\frac{1}{n}\sum_{i=1}^{n}\delta_{Y_i^{(n)}(t)} \xrightarrow{w} Z(t)$$

$Z = (Z(t))$ is called the *FV-process with mutation* $A$

Generator:

$$LF(\mu) := \int \mu(dx) A\left(\frac{\partial F}{\partial \delta_\cdot}(\mu)\right)(x) + \frac{1}{2}\int\int \mu(dx)\left(\delta_x(dy) - \mu(dy)\right)\frac{\partial^2 F}{\partial \delta_x \partial \delta_y}(\mu)$$

For Comparison: Generator of $\eta_t$

$$\tilde{L}F(\mu) := \int \mu(dx) A\left(\frac{\partial F}{\partial \delta_\cdot}(\mu)\right)(x) + \int\int \mu(dx)\mu(dy)\sigma(x,\mu)\left(\frac{\partial F}{\partial \delta_x}(\mu) - \frac{\partial F}{\partial \delta_y}(\mu)\right)$$

# 4. Long-time behaviour

$Z$ is a Feller process on $\mathcal{M}_1(S)$

that is

$$T_t F(\mu) := E_\mu \left[ F\left( Z(t) \right) \right] = \int F\left( Z(t) \right) \, dP_\mu$$

induces Markovian $C_0$-semigroup on $C(\mathcal{M}_1(S))$.

In particular: invariant measure $\Pi$ exists

(and unique, if invariant measure $\nu$ of $(X(t))$ is unique)

## Example: Parent Independent Mutation

Fix $\nu \in \mathcal{M}_1(S)$

$$Af(x) = \frac{\theta}{2} \qquad f(y) - f(x)\,\nu(dy)\,,\theta > 0$$

Then $\Pi =$ Dirichlet Distribution ass. with $\theta\nu$, that is,

for every measurable partition $(A_k)_{1 \le k \le d+1}$ of S

$$\begin{aligned} \Pi\left[\mu(A_1) \in dx_1,\ \ldots, \mu(A_{d+1}) \in dx_{d+1}\right] & \\ &= \frac{\Gamma(\theta)}{\prod_{k=1}^{d+1} \Gamma(\theta\nu(A_k))} \prod_{k=1}^{d+1} x_k^{\theta\nu(A_k)-1}\,dx \\ &= \text{ Dirichlet Distribution on } \Delta_d \text{ w.r.t. } \theta(\nu(A_1), \ldots, \nu(A_{d+1})) \end{aligned}$$

Special feature:

$\Pi$ is symm. measure for $L$ (even $L$ (ess.) self-adjoint)

Known results

Convergence to equilibrium with exponential rate

$$\|T_t F - \langle F, \Pi \rangle\|_{L^2(\Pi)} \leq e^{-\frac{\theta}{2}t} \|F\|_{L^2(\Pi)}$$

([Shimakura, 76] for $|S| < \infty$ and [St, AOP 00] for general $S$)

In particular $\mathcal{E}(F, G) := -\int LF\, G\, d\Pi$ satisfies a Poincaré-inequality with constant $\frac{2}{\theta}$

$$(F - \langle F, \Pi \rangle)^2 \, d\Pi \leq \frac{2}{\theta} \mathcal{E}(F, F)$$

## Logarithmic Sobolev inequality

**Theorem 2** [St, AOP 00] $\mathcal{E}$ satisfies a log. Sobolev inequality iff $|S| < \infty$. In this case

$$\int F^2 \log\left(\frac{F^2}{\|F\|_{L^2(\Pi)}}\right) d\Pi \leq \frac{320}{\theta\nu_*}\mathcal{E}(F,F)$$

where $\nu_* := \inf_{x\in S} \nu(x)$.

Moreover: No other inequalities of hypercontractive type if $|S| = \infty$

**Remark** [Miclo, 01] $\exists\ c_1, c_2$ (independent of $\nu$) s.th. the optimal constant $c_\nu$ satisfies

$$\frac{c_1}{\theta}\log\left(\frac{\theta\wedge 1}{(\theta\nu_*)\wedge e^{-1}}\right) \leq c_\nu \leq \frac{c_2}{\theta}\log\left(\frac{\theta\wedge 1}{(\theta\nu_*)\wedge e^{-1}}\right)$$

**Remark** Both results typical

- for FV-processes with general bounded mutation ([St, PTRF 02])
- for corresponding superprocesses (cf. [St, JFA 03 & AOP 03])

**Theorem 3** [St, PTRF 02] Assume that

$$\|p_t f - \langle f, \nu \rangle\|_\infty \le M e^{-\lambda t} \|f\|_\infty \, .$$

Then

$$\|T_t F - \langle F, \Pi \rangle\|_{L^\infty(\Pi)} \le C e^{-\lambda t} \|F\|_{L^\infty(\Pi)} \, , t \ge 0 \; .$$

## **Further global properties of** $(T_t)$

Consider:

(U) $\exists M, \alpha > 0$ s.th. $\|p_t f\|_\infty \leq M t^{-\alpha} \|f\|_{L^1(\nu)}$, $t \in (0, 1]$

(L) $\exists \delta, m, \beta > 0$ s.th. $p_t(x, y) \geq m e^{-\frac{\beta}{t}}$, $t \in (0, \delta)$, $x, y \in S$

(Lip) $\exists c \geq 0, \alpha \in (0, 1)$ s.th. $\|p_t f\|_{\mathrm{Lip}(d_S)} \leq c t^{-\alpha} \|f\|_\infty$, $t > 0$.

**Example** $S = T^d$ (d-dim Torus)

$$Af(x) = \sum_{ij} a_{ij}(x) \partial_{ij} f(x) + \sum_i b_i(x) \partial_i f(x)$$

strictly elliptic, $a_{ij}$, $b_i \in C^3(T^d)$

Harnack inequality $\Longrightarrow$ $(p_t)$ satisfies (U), (L) & (Lip)

**Theorem 4** [St, PTRF 02]

**(a)** (U) & (L) imply

$$\|T_t F\|_{L^\infty(\Pi)} \le e^{c(1+t^{-3})} \|e^{|F|}\|_{L^1(\Pi)}$$
$$\int (T_t F)^2 \log\left(\frac{(T_t F)^2}{\|T_t F\|^2_{L^2(\Pi)}}\right) d\Pi \le e^{c(1+t^{-3})} \|F\|^2_{L^2(\Pi)}.$$

**(b)** Let $d_w$ be the Wasserstein metric

$$d_w(\mu_1, \mu_2) := \sup_{\substack{f: S \to R \\ \|f\|_{\mathsf{Lip}(d_S)} \le 1}} \left| \int f \, d\mu_1 - \int f \, d\mu_2 \right|.$$

Then: (U), (L) & (Lip) imply

$$\|T_t F\|_{\mathsf{Lip}(d_w)} \le (1 + t^{1-\alpha}) e^{c(1+t^{-3})} \|e^{|F|}\|_{L^1(\Pi)}$$

In particular: $(T_t)$ is strong Feller.

**(c)** (U), (L) & (Lip) imply: $(T_t)$ is compact in $L^2(\Pi)$.