On the stability of genetic algorithms - a variational approach -

Beijing, August 10, 2004

Wilhelm Stannat

Fakultät für Mathematik

Universität Bielefeld

email: stannat@mathematik.uni-bielefeld.de

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

References

- [St1] On the convergence of genetic algorithms - a variational approach, Probab. Theory Relat. Fields 129, 113-132 (2004).
- [St2] On the stability of time-dependent genetic algorithms, Bielefeld (2004), submitted.
- [St3] Stability of the pathwise filter equation on R^d, Bielefeld (2004), submitted.
- [St4] Stability of the pathwise filter equation for a time-dependent signal on R^d, Bielefeld (2004), submitted.

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

Algorithm

$\mu_n :=$ distr. of types in the nth generation

$$\mu_n \overset{\substack{(interactive)\\ selection}}{\longrightarrow} \frac{\sigma(\cdot,\mu_n)\mu_n}{\int \sigma(\cdot,\mu_n)\,d\mu_n} =: \mu'_{n+1} \overset{mutation}{\longrightarrow} \hat{P}\mu'_{n+1} =: \mu_{n+1}$$

Time-continuous analogon

replace P by

Markov process $((X(t))_{t\geq 0}, (P_x))$ with inf. generator A

$\mu_t :=$ distr. of types at time t

$$(1) \quad \frac{d}{dt}\mu_t = \hat{A}\mu_t + \sigma(\cdot, \mu_t) \cdot \mu_t - \int \sigma(\cdot, \mu_t)\, d\mu_t \cdot \mu_t$$

Examples:

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

   Explicit representation of the solution in this particular case:

$$\langle f, \mu_t \rangle := \int f \, d\mu_t = \frac{E_\mu \left[f(X(t)) \exp\left(\int_0^t \sigma(X(s))\, ds \right) \right]}{\underset{\substack{\uparrow \\ \text{ren. Feynman} - \text{Kac semigroup}}}{E_\mu \left[\exp\left(\int_0^t \sigma(X(s))\, ds \right) \right]}}$$

   Remark Simulation of μ_t using finite Moran particle systems more efficient than classical Monte Carlo

   work by: Lyons, Crisan, Miclo, Del Moral, ...

2.

$$\sigma(x,\mu) = \int_{S^n} k(x,y_1,\ldots,y_n)\,\mu(dy_1)\ldots\mu(dy_n)$$

for some $k \in \mathcal{B}_{b,sym}(S^{n+1})_+$

of interest in population genetics

work by Fleming, Donnelly, Kurtz, Dawson, Greven ...

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

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

2. Long-time behaviour

Changes to the population come from two opposing sources:

- selection - encourages conformity

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

  $$\int \sigma \, d\mu_t$$

  hence concentrates more and more on

  $$M(\mu_0) = \{x \in S : \sigma(x) = \mu_0 - ess\, sup\, \sigma\}$$

- mutation - preserves (or even improves) genetic diversity

example in discrete time worked out by Kingman

The variational approach

[St, Probab. Theory Relat. Fields '04]

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

Consider $\mu_0 = h\nu$ with $h \in L^2(\nu)$

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

Replace the mean fitness by the energy

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

Replace (1) by the L^2-normalization

$$(2) \qquad \frac{d}{dt}h_t := Ah_t + \sigma h_t + \lambda(h_t) \cdot h_t$$

Then $\mu_t = \frac{h_t \nu}{\int h_t\,d\nu}$ and

- $\lambda(h_t)$ decreases in time
- critical points of λ coincide with stationary points of (1)

Interacting case

Consider

$$G_\sigma : L^1(\nu) \to R \qquad \text{Frechet differentiable}$$

with differential σ such that

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

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

Theorem 1 Let $h_* \geq 0$ be a critical point of

$$F(h) := -\int Ah\, h\, d\nu - G_\sigma(h^2)\ , h \in \partial B_1(0)\,.$$

Then

$$\boxed{\|h_t - h_*\|_{L^2(\nu)} \leq \Psi\left(\frac{1}{\int hh_*\, d\nu}\right) e^{(L-\frac{\kappa_*}{2})t}\|h - h_*\|_{L^2(\nu)}}$$

Here:

- $\Psi(t) = e^t$

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

- κ_* is the mass gap of the ground state transform

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

  in $L^2(h_*^2\nu)$.

Corollary Let $\nu_* = \frac{h_* \nu}{\int h_* \, d\nu}$. If $\mu_0 = h\,\nu$, then

$$\boxed{\|\mu_t - \nu_*\|_{var} \leq \frac{4}{\int h_* \, d\nu} \Psi\left(\frac{1}{\int h h_* \, d\nu}\right) e^{(L - \frac{\kappa_*}{2})t}}$$

Noninteracting case

$\sigma(x,\mu) = \sigma(x)$ implies $L = 0$ so that

$$\|h_t - h_*\|_{L^2(\nu)} \le \Psi\left(\frac{1}{\int hh_*\,d\nu}\right) e^{-\frac{\kappa_*}{2}t}\|h - h_*\|_{L^2(\nu)}$$

and

$$\|\mu_t - \nu_*\|_{var} \le \frac{4}{\int h_*\,d\nu}\Psi\left(\frac{1}{\int hh_*\,d\nu}\right) e^{-\frac{\kappa_*}{2}t}$$

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

$$\lambda_0 = \lambda(h_*) = \inf_{h \in \partial B_1(0)} \lambda(h)$$

and

$$\lambda_1 = \inf_{\substack{h \in \partial B_1(0) \\ \int hh_*\,d\nu = 0}} \lambda(h)$$

huge literature on properties of h_* and estimates on κ_*

Remark immediate extensions of Theorem 1 to

1. "noncompact" settings
   - unbounded fitness
   - infinite measure ν, etc.
2. noninteracting time-dependent case
3. discrete time analoga
   - to unify some of the previous results obtained in particular examples using particular techniques

3. The noninteracting time-dependent case

Time dependent generalization of (1)

$$\frac{d}{dt}\mu_t = \hat{A}_t \mu_t + \sigma_t \ \mu_t - \int \sigma_t \, d\mu_t \cdot \mu_t \tag{3}$$

where

- (A_t) - family of generators of a time-inhomogeneous Markov process on S
- (σ_t) - family of fitness functions

Let

$$\Gamma(A_t)(f,g) = A_t(fg) - A_t f \, g - f \, A_t g$$

be the carré du champ of A_t

(eg $A_t = \frac{1}{2}\Delta + B \cdot \nabla \Longrightarrow \Gamma(A_t)(f,g) := \nabla f \cdot \nabla g$)

Fix a "suitable" reference measure ν (not necessarily finite)

Theorem 2 Fix $t > 0$, let m_s, $\hat{m}_s$ be solutions of

$$\frac{d}{ds} m_s = -(A_s + \sigma_s) m_s \,, m_t \equiv 1$$

$$\frac{d}{ds} \hat{m}_s = (\hat{A}_s + \sigma_s) \hat{m}_s \,, \hat{m}_0 \equiv 1$$

Let $\mu_{i,s}$, $s \in [0,t]$, be “admissible” solutions of (3) with $\mu_{i,0} = h_{i,0}\, \nu$, $i = 1, 2$, then

$$\boxed{\|\mu_{t,1} - \mu_{t,2}\|_{var} \leq 8 \Psi \left(\frac{1}{\int h_{1,0}\, d\nu_0^*} \right) \cdot \Psi \left(\frac{1}{\int h_{2,0}\, d\nu_0^*} \right) e^{-\int_0^t \kappa_*(s)\, ds}}$$

Here

- $\kappa_*(s)$ is the mass gap of the generator of

$$\int \Gamma(A_s)(f,f)\, d\nu_s^* \qquad \text{in} \qquad L^2(\nu_s^*)$$

- $\nu_s^* = \frac{m_s \hat{m}_s\, \nu}{\int m_s \hat{m}_s\, d\nu}$

Example: Filter theory

(S) $\quad dX(t) = B(X(t))\,dt + \underbrace{C\,dW(t)}_{\textit{systematic error}} \qquad$ on R^d

(O) $\quad dY(t) = GX(t)\,dt + \underbrace{d\tilde{W}(t)}_{\textit{measurement error}},\ Y(0) = 0 \qquad$ on R^p

$(W(t))_{t\geq 0}$, $(\tilde{W}(t))_{t\geq 0}$ independent Brownian motions

Problem Estimate $X(t)$ given $Y(s)$, $s \in [0,t]$, up to time t classical (least square method): compute

$$\eta_t(A) := E\left[1_A\left(X(t)\right)|\mathcal{Y}_t\right] \text{ for meas. } A \subset R^d$$

of the signal $X(t)$ given

$$\mathcal{Y}_t := \sigma\left\{Y(s)|s \in [0,t]\right\}$$

Note η_t depends on distr. μ_0 of the signal X_0

(which is unknown!)

Hence interested in the stability of η_t w.r.t. μ_0

In the following write

$$E_{\mu_0}\left[1_A\left(X(t)\right)|\mathcal{Y}_t\right]$$

to indicate dependence on μ_0

Heuristic $X(t)$ ergodic

$\Longrightarrow$ η_t should become independent of μ_0 for large t

$\Longrightarrow$ stability

corresponding work (for compact state space) by

- Kunita, Stettner
- Ocone, Pardoux (extensions to the the linear case)
- da Prato, Malliavin, Fuhrmann (based on logarithmic derivative of heat-kernel)
- Zeitouni, Atar (introducing new metric: Hilbert distance)

However ergodicity of the signal *not necessary* for stability

The pathwise filter equation

in the following $C = I$ (for simplicity)

Assume that the Kallianpur-Striebel formula holds. Then

$$E\left[1_A(X(t))\,|\mathcal{Y}_t\right] = \frac{\int 1_A(x)\, e^{G^T Y(t)\cdot x}\, \mu_t^Y(dx)}{\int e^{G^T Y(t)\cdot x}\, \mu_t^Y(dx)} \quad a.s.$$

where:

(4) $$\frac{d}{dt}\mu_t^Y = \hat{A}_t^Y \mu_t^Y + \sigma_t^Y\ \mu_t^Y - \int_{R^d} \sigma_t^Y\, d\mu_t^Y \cdot \mu_t^Y$$

for

- $A_t^Y f(x) := \frac{1}{2}\Delta f(x) + \left(B(x) - G^T Y(t)\right) \cdot \nabla f(x)$
- $\sigma_t^Y(x) := -G^T Y(t) \cdot B(x) + \frac{1}{2}\|G^T Y(t)\|^2 - \frac{1}{2}\|Gx\|^2$

The variational approach

Assume that

A.1 $B(x) = \frac{\nabla\varphi}{\varphi}(x)$ for some bounded smooth log-concave φ.

$\Longrightarrow A$ has symm. measure $\nu(dx) := \varphi^2 \, dx$

A.2 $W(x) := \|Gx\|^2 + \frac{\Delta\varphi}{\varphi}(x)$

is uniformly strictly convex

$$\exists\kappa_* > 0 \text{ such that } W'' \geq \kappa_*^2 \cdot I$$

Remark $\frac{1}{2}\Delta - \frac{1}{2}W$ has unique ground state $h_0 > 0$,

$$-(\log h_0)'' \geq \kappa_* \cdot I > 0 \qquad \text{Brascamp, Lieb 76}$$

and

$$h_*(x) := h_0(x)\varphi(x)^{-1}$$

is unique ground state of $A + \sigma$ (mass gap $\geq \kappa_*$ below λ_0)

Long-time behaviour of (4)

Heuristic time-homogeneous case $Y \equiv 0$

$$A_t^y f(x) = Af(x) := \frac{1}{2}\Delta f(x) + B(x) \cdot \nabla f(x)$$

and

$$\sigma_t^y(x) = \sigma(x) := -\frac{1}{2}\|Gx\|^2$$

and coefficients of (4) become time-independent

$$\frac{d}{dt}\mu_t = \hat{A}\mu_t + \sigma\mu_t - \int \sigma \, d\mu_t \cdot \mu_t$$

Theorem 1 implies

$$\lim_{t\to\infty} \mu_t(dx) = \nu_*(dx) := \frac{h_*\varphi^2 \, dx}{\int h_*\varphi^2 \, dx}$$

with exp. rate $\frac{\kappa_*}{2}$ (w.r.t. $\|\cdot\|_{L^2}$ and $\|\cdot\|_{var}$)

Theorem 3 Let

- $Y \in C([0,\infty[; R^p)$, $Y(0) = 0$
- $\mu_i \ll \nu_* = \frac{h_* \, \nu}{\int h_* \, d\nu}$ with densities bounded from above and from below

Then

$$\limsup_{t\to\infty} e^{\frac{\kappa_*}{2}t} \|\mu_{1,t}^Y - \mu_{2,t}^Y\|_{var} < \infty$$

Moreover

$$\limsup_{t\to\infty} e^{\frac{\kappa_*}{2}t} \left\| \frac{e^{G^T Y(t)\cdot x} \mu_{1,t}^Y(dx)}{\int e^{G^T Y(t)\cdot x} \mu_{1,t}^Y(dx)} - \frac{e^{G^T Y(t)\cdot x} \mu_{2,t}^Y(dx)}{\int e^{G^T Y(t)\cdot x} \mu_{2,t}^Y(dx)} \right\|_{var} < \infty$$

Remark

- rate *independent* of Y
- precise quantitative bounds
- (4) can be stable even if (S) is nonergodic - first proof in the nonlinear time-continuous case

Corollary Assume that (4) is unique

for all $Y \in C([0,\infty[; R^p)$, $Y(0) = 0$

then

$$\limsup_{t\to\infty} e^{\frac{\kappa_*}{2}t} \left| E_{\mu_1}\left[1_A\left(X(t)\right)|\mathcal{Y}_t\right] - E_{\mu_2}\left[1_A\left(X(t)\right)|\mathcal{Y}_t\right]\right|$$

$$< \infty \quad a.s.$$

Remark

- simple sufficient criterion for stability:

$$\boxed{\begin{array}{c} W(x) = \|Gx\|^2 + \dfrac{\Delta\varphi}{\varphi}(x)\,, \qquad B = \dfrac{\nabla\varphi}{\varphi} \\ W'' \geq \kappa_*^2 \cdot I \text{ implies exp. stability with rate } \dfrac{\kappa_*}{2} \end{array}}$$

- more precise and explicit results in the linear case (Kalman-Bucy filter) - also in the nongradient case

- similar results for time-dependent signals

- work in progress for the time-discrete case