Gradient Flow for Regularized Stochastic Control Problems[1]

David Šiška[2]

LNU Stochastic Analysis Seminar, 24th November 2020

Joint work with Kaitong Hu[3], Jean-Francois Jabir[4], Zhenjie Ren[5] and Lukasz Szpruch[6]

[1] https://arxiv.org/abs/2006.05956

[2] University of Edinburgh
[3] CMAP, École Polytechnique
[4] Higher School of Economics, National Research University, Moscow
[5] CEREMADE, Université Paris Dauphine
[6] University of Edinburgh and The Alan Turing Institute

Stochastic Control Problem with Entropic Regularization I

For $\xi \in \mathbb{R}^d$ and $\mu \in \mathcal{V}_q^W$, consider the controlled process

$$X_t(\mu) = \xi + \int_0^t \Phi_r(X_r(\mu), \mu_r)\, dr + \int_0^t \Gamma_r(X_r(\mu), \mu_r)\, dW_r\,, \quad t \in [0, T]\,. \qquad (1)$$

Here

$$\mathcal{V}_q^W := \Big\{ \nu : \Omega^W \to \mathcal{M}_q : \mathbb{E}^W \int_0^T \int |a|^q \, \nu_t(da)\, dt < \infty \\ \text{and } \nu_t \text{ is } \mathcal{F}_t^W\text{-measurable } \forall t \in [0, T] \Big\}$$

and

$$\mathcal{M}_q := \Big\{ \nu \in \mathscr{M}_+([0, T] \times \mathbb{R}^p) : \nu_t \in \mathcal{P}(\mathbb{R}^p)\,, \ \int_0^T \int |a|^q \, \nu_t(da, dt) < \infty\,, \\ \nu(dt, da) = \nu_t(a)\, da\, dt \text{ for a.a. } t \in [0, T] \Big\}\,.$$

Stochastic Control Problem with Entropic Regularization II

If $m \in \mathcal{P}(\mathbb{R}^p)$ is a.c. w.r.t. the Lebesgue measure (so that we can write $m(da) = m(a)\,da$) let

$$\mathrm{Ent}(m) := \int [\log m(a) - \log \gamma(a)]\, m(a)\, da\,,$$

where

$$\gamma(a) = e^{-U(a)} \text{ with } U \text{ s.t. } \int e^{-U(a)}\, da = 1\,.$$

Otherwise let $\mathrm{Ent}(m) := \infty$.

Given F and g we wish to **minimize** the objective functional

$$J^\sigma(\nu,\xi) := \mathbb{E}^W\left[\int_0^T \left[F_t(X_t(\nu),\nu_t) + \frac{\sigma^2}{2}\mathrm{Ent}(\nu_t)\right] dt + g(X_T(\nu)) \Big| X_0(\nu) = \xi\right]. \quad (2)$$

Example: Relaxed Control

$$\Phi_t(x,m) = \int \phi_t(x,a) m(da)\,,$$

$$\Gamma_t(x,m)(\Gamma_t(x,m))^\top = \int \gamma_t(x,a)\gamma_t(x,a)^\top m(da)\,,$$

$$F_t(x,m) = \int f_t(x,a)\, m(da)\,.$$

Why Regularize with Entropy

Several perspectives:

i) Exploration vs. exploitation when solving an episodic control problem with unknown dynamics (learning) Wang, Zariphopoulou and Zhou [7] and Wang and Zhou [8].

ii) Regularity of Markovian controls Reisinger and Zhang [4].

iii) Gradient flow for optimal control Š and Szpruch [6].

Talk outline

i) Introduction

ii) Minimizing Convex Functions of Measures with Gradient Flows (one-hidden layer NNs)
   - Necessary condition for optimality
   - Gradient flow and Free energy balance
   - Convergence to minimum

iii) Regularized Stochastic Control (towards reinforcement learning)
   - Necessary condition for optimality (Pontryagin)
   - Gradient flow and Free energy balance
   - Convergence to optimal control

Minimizing Convex Functions of Measures

Minimizing Convex Functions of Measures

Given $F : \mathcal{P}(\mathbb{R}^p) \to \mathbb{R}$ convex[7], find

$$\inf_{m \in \mathcal{P}(\mathbb{R}^p)} F(m) .$$

Minimum not unique. Consider

$$\inf_{m \in \mathcal{P}(\mathbb{R}^p)} V^\sigma(m) := \inf_{m \in \mathcal{P}(\mathbb{R}^p)} \left(F(m) + \frac{\sigma^2}{2} \mathrm{Ent}(m) \right) .$$

Example: nonlinear regresssion with an idealized 1 hidden layer neural network:

$$V^\sigma(m) := \int_{\mathbb{R} \times \mathbb{R}^D} \left| y - \int_{\mathbb{R}^p} \hat{\varphi}(\theta, z)\, m(d\theta) \right|^2 \nu(dy, dz) + \frac{\sigma^2}{2} \mathrm{Ent}(m) .$$

This has convex + strictly convex part. Observed in the pioneering works of Mei, Misiakiewicz and Montanari [3], Chizat and Bach [1] as well as Rotskoff and Vanden-Eijnden [5].

[7] For any $m, m' \in \mathcal{P}(\mathbb{R}^p)$ we have

$$F\Big((1-\alpha)m + \alpha m'\Big) \le (1-\alpha)F(m) + \alpha F(m') \text{ for all } \alpha \in [0,1] .$$

Convergence when $\sigma \searrow 0$

Proposition 1

Assume that F is continuous in the topology of weak convergence. Then the sequence of functions $V^\sigma = F + \frac{\sigma^2}{2} H$ converges in the sense of Γ-convergence to F as $\sigma \searrow 0$. In particular, given a sequence of minimizers $m^{*,\sigma}$ of V^σ, we have

$$\limsup_{\sigma \to 0} F(m^{*,\sigma}) = \inf_{m \in \mathcal{P}_2(\mathbb{R}^d)} F(m).$$

Characterization of the minimizer

Proposition 2 (First order condition)

Assuming that F is convex, bbd. from below and ∇U dissipative, the function V^σ has a unique minimizer $m^* \in \mathcal{P}_2(\mathbb{R}^d)$ which is absolutely continuous with respect to Lebesgue measure and satisfies

$$\frac{\delta F}{\delta m}(m^*, \cdot) + \frac{\sigma^2}{2}\log(m^*) + \frac{\sigma^2}{2}U \quad \text{is a constant, } m^* - a.s.$$

On the other hand if $m' \in \mathcal{I}_\sigma$ where

$$\mathcal{I}_\sigma := \left\{ m \in \mathcal{P}(\mathbb{R}^d) : \frac{\delta F}{\delta m}(m, \cdot) + \frac{\sigma^2}{2}\log(m) + \frac{\sigma^2}{2}U \quad \text{is a constant} \right\}$$

then $m' = \arg\min_{m \in \mathcal{P}(\mathbb{R}^d)} V^\sigma$.

Corollary 1

The optimal m^* satisfies the functional equation

$$m^*(\theta) = \frac{1}{Z}\exp\left(-\frac{2}{\sigma^2}\left(\frac{\partial F}{\partial m}(m^*, \theta) + U(\theta)\right)\right).$$

where $Z := \int \exp\left(-\frac{2}{\sigma^2}\left(\frac{\partial F}{\partial m}(m^*, \theta) + U(\theta)\right)\right) d\theta$.

Gradient Flow for Convex Optimization on Space of Measures

Due to the form of m^* we "hope" that m^* is the invariant measure of

$$\begin{cases} d\theta_s = -\left(\nabla_\theta \frac{\delta F}{\delta m} F(m_s, \theta_s) + \frac{\sigma^2}{2} \nabla_\theta U(\theta_s)\right) \, ds + \sigma dB_s \,, \;\; s \in [0, \infty)\,, \\ m_s = \mathrm{Law}(\theta_s)\,, \;\; s \in [0, \infty)\,. \end{cases} \tag{3}$$

Fokker–Planck

$$\partial_s m = \nabla_\theta \cdot \left(\left(\nabla_\theta \frac{\delta F}{\delta m}(m, \cdot) + \frac{\sigma^2}{2} \nabla_\theta U\right) m + \frac{\sigma^2}{2} \nabla_\theta m \right) \text{ on } (0, \infty) \times \mathbb{R}^p\,.$$

This can be viewed as a randomized, continuous time version of the classical gradient descent algorithm.

Energy balance

Theorem 2

Let $m_0 \in \mathcal{P}_2(\mathbb{R}^p)$. Under our assumptions on F (growth, smoothness) and ∇U (smoothness, dissipativity), we have for any $s' > s > 0$

$$V^\sigma(m_{s'}) - V^\sigma(m_s)$$
$$= -\int_s^{s'} \int \left| D_m F(m_r, \theta) + \frac{\sigma^2}{2} \frac{\nabla m_r}{m_r}(\theta) + \frac{\sigma^2}{2} \nabla U(\theta) \right|^2 m_r(\theta)\, d\theta\, dr.$$

Proof outline: Follows from a priori estimates and regularity results on the nonlinear Fokker–Planck equation and the chain rule for flows of measures.

Convergence

Theorem 3

Let our assumptions on F (growth, smoothness) and ∇U (smoothness, dissipativity) hold and let $m_0 \in \cup_{p>2}\mathcal{P}_p(\mathbb{R}^d)$. Denote by $(m_s)_{s\geq 0}$ the flow of marginal laws of the solution to (3)*. Then, there exists an invariant measure of* (3) *equal to $m^* := \operatorname{argmin}_m V^\sigma(m)$ and*

$$\mathcal{W}_2(m_s, m^*) \to 0 \text{ as } s \to \infty\,.$$

Proof key ingredients: Tightness of $(m_s)_{s\geq 0}$, Lasalle's invariance principle, Theorem 2, HWI inequality.

All results so far from Hu, Ren, Š and Szpruch [2].

Regularized Stochastic Control

Stochastic Control Problem with Entropic Regularization

For $\xi \in \mathbb{R}^d$ and $\mu \in \mathcal{V}_q^W$, consider the controlled process

$$X_t(\mu) = \xi + \int_0^t \Phi_r(X_r(\mu), \mu_r)\, dr + \int_0^t \Gamma_r(X_r(\mu), \mu_r)\, dW_r\,, \quad t \in [0, T]\,,$$

$$\mathcal{V}_q^W := \Big\{ \nu : \Omega^W \to \mathcal{M}_q : \mathbb{E}^W \int_0^T \!\! \int |a|^q\, \nu_t(da)\, dt < \infty \\ \text{and } \nu_t \text{ is } \mathcal{F}_t^W\text{-measurable } \forall t \in [0, T] \Big\}\,,$$

$$\mathcal{M}_q := \Big\{ \nu \in \mathscr{M}_+([0, T] \times \mathbb{R}^p) : \nu_t \in \mathcal{P}(\mathbb{R}^p)\,, \int_0^T \!\! \int |a|^q\, \nu_t(da, dt) < \infty\,, \\ \nu(dt, da) = \nu_t(a)\, da\, dt \text{ for a.a. } t \in [0, T] \Big\}\,.$$

Given F and g we wish to **minimize** the objective functional

$$J^\sigma(\nu, \xi) := \mathbb{E}^W \left[\int_0^T \left[F_t(X_t(\nu), \nu_t) + \frac{\sigma^2}{2} \mathrm{Ent}(\nu_t) \right] dt + g(X_T(\nu)) \Big| X_0(\nu) = \xi \right].$$

Note: $J^\sigma(\nu, \xi)$ is not (necessarily) "convex + strictly convex" function of ν.

Pontryagin optimality

Hamiltonian

$$H_t^\sigma(x,y,z,m) := \Phi_t(x,m)y + \operatorname{tr}(\Gamma_t^\top(x,m)z) + F_t(x,m) + \frac{\sigma^2}{2}\operatorname{Ent}(m)\,. \tag{4}$$

Adjoint process for control μ:

$$\begin{aligned} dY_t(\mu) &= -(\nabla_x H_t^0)(X_t(\mu), Y_t(\mu), Z_t(\mu), \mu_t)\,dt + Z_t(\mu)\,dW_t\,, \quad t \in [0,T]\,,\\ Y_T(\mu) &= (\nabla_x g)(X_T(\mu))\,. \end{aligned} \tag{5}$$

Theorem 4 (Necessary condition for optimality)

Fix $\sigma > 0$. Fix $q > 2$. Let the Assumptions on growth and differentiablity hold. If $\nu \in \mathcal{V}_q^W$ is (locally) optimal for $J^\sigma(\cdot, \xi)$ given by (2), $X(\nu)$ and $Y(\nu)$, $Z(\nu)$ are the associated optimally controlled state and adjoint processes given by (1) and (5) respectively, then for any other $\mu \in \mathcal{V}_q^W$ it holds that

i)

$$\int \left[\frac{\delta H_t^0}{\delta m}(X_t(\nu), Y_t(\nu), Z_t(\nu), \nu_t, a) + \frac{\sigma^2}{2} \log \frac{\nu_t(a)}{\gamma(a)} \right] (\mu_t - \nu_t)(da)$$
$$\geq 0 \ \text{ for a.a. } (\omega, t) \in \Omega^W \times (0, T)\,.$$

ii) *For a.a. $(\omega, t) \in \Omega^W \times (0, T)$ there exists $\varepsilon > 0$ (small and depending on μ_t) such that*

$$H_t^\sigma(X_t(\nu), Y_t(\nu), Z_t(\nu), \nu_t + \varepsilon(\mu_t - \nu_t)) \geq H_t^\sigma(X_t(\nu), Y_t(\nu), Z_t(\nu), \nu_t)\,.$$

In other words, the optimal relaxed control $\nu \in \mathcal{V}_q^W$ locally minimizes the Hamiltonian.

Necessary condition for optimality

Let

$$\mathcal{I}^\sigma := \Big\{ \nu \in \mathcal{V}_q^W : \ \frac{\delta \mathbf{H}_t^\sigma}{\delta m}(a,\nu) \text{ is constant} \quad \text{for a.a. } a \in \mathbb{R}^p, \text{ a.a. } (t,\omega^W) \in (0,T)\times\Omega^W \Big\}. \tag{6}$$

Here

$$\frac{\delta \mathbf{H}_t^0}{\delta m}(\cdot,\nu) := \frac{\delta H^0}{\delta m}(X_t(\nu), Y_t(\nu), Z_t(\nu), \nu_t, \cdot)\,.$$

Corollary 5 (First order condition)

If $\nu \in \mathcal{V}_q^W$ is (locally) optimal for $J^\sigma(\cdot,\xi))$ then $\nu \in \mathcal{I}^\sigma$.

From the first order condition we have that for a.a. $(\omega^W, t) \in \Omega^W \times (0,T)$ we have

$$\mu_t^*(a) = \mathcal{Z}_t^{-1} e^{-\frac{2}{\sigma^2}\frac{\delta \mathbf{H}_t^0}{\delta m}(a,\mu^*)}\gamma(a)\,, \quad \mathcal{Z}_t := \int e^{-\frac{2}{\sigma^2}\frac{\delta \mathbf{H}_t^0}{\delta m}(a,\mu^*)}\gamma(a)da\,. \tag{7}$$

So what is the right gradient flow?

Necessary condition proof outline I

Let $\mu, \nu \in \mathcal{V}_q^W$ and $\nu_t^\varepsilon := \nu_t + \varepsilon(\mu_t - \nu_t)$. Consider
$\frac{d}{d\varepsilon} J^\sigma \left((\nu_t + \varepsilon(\mu_t - \nu_t)_{t\in[0,T]}, \xi) \right|_{\varepsilon=0}$.

Let X^ε be the solution to (1) with control ν_t^ε and

$$
\begin{aligned}
dV_t = & \left[(\nabla_x \Phi)(X_t, \nu_t) V_t + \int \frac{\delta \Phi}{\delta m}(X_t, \nu_t, a)(\mu_t - \nu_t)(da) \right] dt \\
& + \left[(\nabla_x \Gamma)(X_t, \nu_t) V_t + \int \frac{\delta \Gamma}{\delta m}(X_t, \nu_t, a)(\mu_t - \nu_t)(da) \right] dW_t \, .
\end{aligned}
\tag{8}
$$

Lemma 6

We have

$$
\lim_{\varepsilon \searrow 0} \mathbb{E}^W \left[\sup_{t \le T} \left| \frac{X_t^\varepsilon - X_t}{\varepsilon} - V_t \right|^2 \right] = 0 \, .
$$

Necessary condition proof outline II

Lemma 7

We have that

$$\frac{d}{d\varepsilon} J^0 \left((\nu_t + \varepsilon(\mu_t - \nu_t)_{t\in[0,T]}, \xi \right) \Big|_{\varepsilon=0}$$
$$= \mathbb{E}\bigg[\int_0^T \left[\int \frac{\delta H^0}{\delta m}(X_t, Y_t, Z_t, \nu_t, a)(\mu_t - \nu_t)(da) \right] dt \bigg].$$

Necessary condition proof outline III

Lemma 8

i) *for any* $\varepsilon \in (0,1)$ *we have*

$$\frac{1}{\varepsilon}\int_0^T [Ent(\nu_t^\varepsilon) - Ent(\nu_t)]\ dt \geq \int_0^T \int [\log \nu_t(a) - \log \gamma(a)](\mu_t - \nu_t)(da)\, dt\,,$$

ii)

$$\limsup_{\varepsilon\to 0} \frac{1}{\varepsilon}\int_0^T [Ent(\nu_t^\varepsilon) - Ent(\nu_t)]\ dt \leq \int_0^T \int [\log \nu_t(a) - \log \gamma(a)](\mu_t - \nu_t)(da)\, dt\,.$$

Necessary condition proof outline IV

Proof of Theorem 4. Let $(\mu_t)_{t\in[0,T]}$ be an arbitrary relaxed control Since $(\nu_t)_{t\in[0,T]}$ is optimal we know that

$$J^\sigma\left(\nu_t+\varepsilon(\mu_t-\nu_t)\right)_{t\in[0,T]}\right) \geq J^\sigma(\nu) \quad \text{for any } \varepsilon>0.$$

From this, Lemma 7 and 8 point ii) we get that

$$\begin{aligned}
0 &\leq \limsup_{\varepsilon\to 0}\frac{1}{\varepsilon}\left(J^\sigma(\nu_t+\varepsilon(\mu_t-\nu_t))_{t\in[0,T]}-J^\sigma(\nu)\right)\\
&\leq \mathbb{E}\int_0^T\int\left[\frac{\delta H^0}{\delta m}(X_t,Y_t,Z_t,\nu_t,a)+\frac{\sigma^2}{2}(\log\nu_t(a)-\log\gamma(a))\right](\mu_t-\nu_t)(da)\,dt\,.
\end{aligned}$$

Gradient Flow

Definition 9

We will say that b is a *permissible flow* if $b_{\cdot,t} \in C^{0,1}([0,\infty) \times \mathbb{R}^p; \mathbb{R}^p)$, if for all s, t the function $a \mapsto b_{s,t}(a)$ is of linear growth and if for any $s \geq 0$ and $a \in \mathbb{R}^p$ the random variable $b_{s,t}(a)$ is $\mathcal{F}_t^W$-measurable.

Lemma 10

If b is a permissible flow (c.f. Definition 9) then the linear PDE

$$\partial_s \nu_{s,t} = \nabla_a \cdot \left(b_{s,t}\nu_{s,t} + \frac{\sigma^2}{2}\nabla_a \nu_{s,t} \right), \quad s \in [0,\infty), \quad \nu_{0,t} \in \mathcal{P}_2(\mathbb{R}^p) \qquad (9)$$

has unique solution $\nu_{\cdot,t} \in C^{1,\infty}((0,\infty) \times \mathbb{R}^p; \mathbb{R})$ for each $t \in [0,T]$ and $\omega^W \in \Omega^W$. Moreover for each $s > 0$, $t \in [0,T]$ and $\omega^W \in \Omega^W$ we have $\nu_{s,t}(a) > 0$ and $\nu_{s,t}(a)$ is $\mathcal{F}_t^W$-measurable.

Energy balance

Theorem 11

Fix $\sigma \geq 0$ and assume enough differentiability / integrability. Let b be a permissible flow (c. f. Definition 9) such that $a \mapsto |\nabla_a b_{s,t}(a)|$ is bounded uniformly in s, t and $\omega^W \in \Omega^W$. Let $\nu_{s,t}$ be the solution to (9). Assume that $X_{s,\cdot}, Y_{s,\cdot}, Z_{s,\cdot}$ are the forward and backward processes arising from control $\nu_{s,\cdot} \in \mathcal{V}_2^W$ and data $\xi \in \mathbb{R}^d$ given by (1) and (5). Then

$$\begin{aligned}\frac{d}{ds} J^\sigma(\nu_{s,\cdot}) = \\ -\mathbb{E}^W \int_0^T \int \left[\left(\nabla_a \frac{\delta \mathbf{H}_t^0}{\delta m} \right)(a, \nu_{s,\cdot}) + \frac{\sigma^2}{2} \nabla_a U(a) + \frac{\sigma^2}{2} \nabla_a \log(\nu_{s,t}(a)) \right] \\ \cdot \left(b_{s,t} + \frac{\sigma^2}{2} \nabla_a \log \nu_{s,t} \right) \nu_{s,t}(da)\, dt\,.\end{aligned} \tag{10}$$

We can take

$$b_{s,t} = \left(\nabla_a \frac{\delta \mathbf{H}_t^0}{\delta m} \right)(a, \nu_{s,\cdot}) + \frac{\sigma^2}{2} \nabla_a U(a)$$

so that $\frac{d}{ds} J^\sigma(\nu_{s,\cdot}) \leq 0$ for all $s \geq 0$.

Energy balance proof outline I

Lemma 12 (Properties of Gradient Flow, Hu, Ren, Š, Szpruch [2])

Let b be a permissible flow such that $a \mapsto |\nabla_a b_{s,t}(a)|$ is bounded uniformly in $s > 0$, $t \in [0, T]$, $\omega^W \in \Omega^W$. Then

i) *For all $s > 0$, $t \in [0, T]$, $\omega^W \in \Omega^W$ and $a \in \mathbb{R}^p$ we have $\nu_{s,t}(a) > 0$ and $Ent(\nu_{s,t}) < \infty$.*

ii) *For all $s > 0$, $t \in [0, T]$ and $\omega^W \in \Omega^W$ we have $\int |\nabla_a \log \nu_{s,t}(a)|^2 \nu_{s,t}(a)(da) < \infty$.*

iii) *For all $s > 0$, $t \in [0, T]$ and $\omega^W \in \Omega^W$ we have*

$$\int |\nabla_a \nu_{s,t}(a)|\, da + \int |a \cdot \nabla_a \nu_{s,t}(a)|\, da + \int |\Delta_a \nu_{s,t}(a)|\, da < \infty\,.$$

Energy balance proof outline II

Let

$$d\theta_{s,t} = -b_{s,t}(\theta_{s,t})\, ds + \sigma\, dB_s\,.$$

With the above estimates we can use Itô formula on $\log(\theta_{s,t})$ and take expectation:

Lemma 13

Fix $\sigma \geq 0$. Let b be a permissible flow (c. f. Definition 9) such that $a \mapsto |\nabla_a b_{s,t}(a)|$ is bounded uniformly in s, t and $\omega^W \in \Omega^W$. Let $\nu_{s,t}$ be the solution to (9). Then

$$d\,Ent(\nu_{s,t}) = -\int \left(\nabla_a \log \nu_{s,t} + \nabla_a U\right) \cdot \left(b_{s,t} + \frac{\sigma^2}{2}\nabla_a \log \nu_{s,t}\right)\, \nu_{s,t}(da)\, ds\,.$$

SDE / BSDE System Representation for Gradient Flow

Let $(\theta^0_t)_{t\in[0,T]}$ be an $(\mathcal{F}^W_t)$-adapted, $\mathbb{R}^p$-valued stochastic process on $(\Omega, \mathcal{F}, \mathbb{P})$ such that $(\mathcal{L}(\theta^0_t|\mathcal{F}^W_t))_{t\in[0,T]} \in \mathcal{V}^W_2$ and consider with $\theta_{t,0} = \theta^0_t$ and $s \geq 0$:

$$d\theta_{s,t} = -\Big((\nabla_a \frac{\delta H^0_t}{\delta m})(X_{s,t}, Y_{s,t}, Z_{s,t}, \nu_{s,t}, \theta_{s,t}) + \frac{\sigma^2}{2}(\nabla_a U)(\theta_{s,t})\Big)\, ds + \sigma dB_s\,, \tag{11}$$

coupled with

$$\begin{cases} \nu_{s,t} &= \mathcal{L}(\theta_{s,t}|\mathcal{F}^W_t)\,, \\ X_{s,t} &= \xi + \int_0^t \Phi_r(X_{s,r}, \nu_{s,r})\, dr + \int_0^t \Gamma_r(X_{s,r}, \nu_{s,r}(da))\, dW_r\,, \quad t \in [0,T]\,, \\ dY_{s,t} &= -(\nabla_x H^0_t)(X_{s,t}, Y_{s,t}, Z_{s,t}, \nu_{s,t})\, dt + Z_{s,t}\, dW_t\,, \\ Y_{s,T} &= (\nabla_x g)(X_T)\,. \end{cases} \tag{12}$$

Theorem 14

Let Assumptions regularity / integrability assumption hold. Moreover, assume that for any $\mu^0 \in \mathcal{V}_q^W$ *the MFLD* (11)-(12) *has unique solution* $P_s\mu^0$ *and that it admits unique invariant measure* $\mu^* \in \mathcal{V}_q^W$ *such that for any* $\mu^0 \in \mathcal{V}_q^W$, $\lim_{s\to\infty} \rho_q(P_s\mu^0, \mu^*) = 0$. *Then*

i) *We have* $J^\sigma(\mu^*) < \infty$ *and* $\mathcal{I}^\sigma = \{\mu^*\}$. *In other words,* μ^* *is the only control which satisfies the first order condition in* (6).

ii) *The unique minimizer of* J^σ *is* μ^*.

Proof outline for Theorem 14 part i):

Since μ^* is invariant $\partial_s \mu^*_{s,t} = 0$ and so for $t \in [0, T]$

$$0 = \nabla_a \cdot \left(\Big((\nabla_a \frac{\delta \mathbf{H}^0_t}{\delta m})(\cdot, \mu^*) + \frac{\sigma^2}{2} (\nabla_a U) \Big) \mu^*_t + \frac{\sigma^2}{2} \nabla_a \mu^*_t \right) . \qquad (13)$$

This implies that $\mu^* \in \mathcal{I}^\sigma$.

Consider now some $\nu \in \mathcal{I}^\sigma$. Then from (6) we get that

$$\nu_t(a) = \mathcal{Z}_t^{-1} e^{-\frac{2}{\sigma^2} \frac{\delta \mathbf{H}^0_t}{\delta m}(a, \nu_t(a))} \gamma(a) \, , \quad \mathcal{Z}_t := \int e^{-\frac{2}{\sigma^2} \frac{\delta \mathbf{H}^0_t}{\delta m}(a, \nu_t(a))} \gamma(a) da \, .$$

From this we see that almost all $t \in [0, T]$ and $\omega^W \in \Omega^W$ we have that ν_t solves (13).

Proof outline for Theorem 14 part ii):

Let $\mu^0 \in \mathcal{V}_2^W$ s.t. $J^\sigma(\mu^0) < J^\sigma(\mu^*)$. By assumption $\lim_{s\to\infty} P_s\mu^0 = \mu^*$.

From this and Theorem 11 and from lower semi-continuity of J^σ we get

$$
\begin{aligned}
&J^\sigma(\mu^*) - J^\sigma(\mu^0) \leq \liminf_{s\to\infty} J^\sigma(P_s\mu^0) - J^\sigma(\mu^0) \\
&= -\liminf_{s\to\infty} \int_0^s \mathbb{E}^W \int_0^T \left[\int \left| \left(\nabla_a \frac{\delta \mathbf{H}^\sigma}{\delta m} \right) (a, (P_s\mu^0)_t) \right|^2 (P_s\mu^0)_t(da) \right] dt\, ds \\
&\leq 0
\end{aligned}
$$

which is a contradiction so μ^* is (locally) optimal.

Any other (locally) optimal control $\nu^* \in \mathcal{V}_2^W$ we have for any $\nu \in \mathcal{V}_2^W$, due to Theorem 4 that

$$
0 \leq \mathbb{E}^W \left[\int_0^T \int \frac{\delta \mathbf{H}_t^\sigma}{\delta m}(a, \nu^*)(\nu_t - \nu_t^*)(da)\, dt \right].
$$

Due to Corollary 5 this implies that $\nu^* \in \mathcal{I}^\sigma$. But part i) says $\mathcal{I}^\sigma = \{\mu^*\}$.

Structural Assumptions for Convergence to Inv. Meas.

Assumption 3

Let $\nabla_a U$ be Lipschitz continuous in a, let there be $\kappa > 0$ such that:

$$\big(\nabla_a U(a') - \nabla_a U(a)\big) \cdot \big(a' - a\big) \geq \kappa |a' - a|^2,\ a, a' \in \mathbb{R}^p\,.$$

Assumption 4

Assume that there exists $\eta_1, \eta_2 \in \mathbb{R}$, $\bar{\eta} \in L^{q/2}(\Omega^W \times (0,T);\mathbb{R})$ and $\mathcal{E} : \mathcal{V}_q^W \times \mathcal{V}_q^W \to [0,\infty)$ s. t. for any $a \in \mathbb{R}^p$, any $\mu \in \mathcal{V}_2^W$

$$\Big(\nabla_a \frac{\delta \mathbf{H}_t^0}{\delta m}\Big)(a,\mu)a \geq \eta_1 |a|^2 - \eta_2 \mathcal{E}_t(\mu,\delta_0)^2 - \bar{\eta}_t\,,\ \ t \in [0,T]$$

and for all $\mu, \mu' \in \mathcal{V}_q^W$ we have $\mathbb{E}^W\big[\int_0^T \mathcal{E}_t(\mu,\mu')^q\,dt\big] \leq \rho_q(\mu,\mu')^q$.

Assumption 5

There exists $\eta_1, \eta_2 \in \mathbb{R}$ and $\mathcal{E} : \mathcal{V}_q^W \times \mathcal{V}_q^W \to [0,\infty)$ s. t. for all $t \in [0,T]$, for all a, a' and for all $\mu, \mu' \in \mathcal{V}_q^W$ we have $\mathbb{E}^W\big[\int_0^T \mathcal{E}_t(\mu,\mu')^q\,dt\big] \leq \rho_q(\mu,\mu')^q$ and

$$2\bigg((\nabla_a \frac{\delta \mathbf{H}_t^0}{\delta m})(a',\mu') - (\nabla_a \frac{\delta \mathbf{H}_t^0}{\delta m})(a,\mu)\bigg)(a'-a) \geq \eta_1 |a'-a|^2 - \eta_2 \mathcal{E}_t(\mu',\mu)^2\,.$$

Lemma 15 (Existence and uniqueness)

Let Assumptions 3, 4 and 5 hold. If $\frac{q}{2}\left(\sigma^2\kappa + \eta_1\right) > 0$ *then there is a unique solution to* (11)-(12) *for any* $s \geq 0$*. Moreover if* $\lambda := \frac{q}{2}\left(\frac{\sigma^2\kappa}{4} + \eta_1 - \eta_2\right) > 0$ *then there is* $c = c_{T,q,\sigma,\kappa,\eta_1,\bar{\eta}}$ *such that for any* $s \geq 0$ *we have*

$$\int_0^T \mathbb{E}[|\theta_{s,t}|^q]\, dt \leq e^{-\lambda s} \int_0^T \mathbb{E}[|\theta_t^0|^q]\, dt + c \int_0^s e^{-\lambda(s-v)}\, dv\,. \tag{14}$$

For $\mu, \mu' : \Omega^W \to \mathcal{V}_2^W$ let

$$\rho_q(\mu,\mu') = \left(\mathbb{E}^W\left[|\mathcal{W}_q^T(\mu,\mu')|^q\right]\right)^{1/q}$$

Theorem 16 (Exponential convergence to invariant measure)

Let Assumptions 3 and 5 hold. Moreover, assume that $\lambda = \frac{q}{2}\left(\sigma^2\kappa + \eta_1 - \eta_2\right) > 0$*. Then there is* $\mu^* \in \mathcal{V}_q^W$ *such that for any* $s \geq 0$ *we have* $P_s\mu^* = \mu^*$ *and* μ^* *is unique. For any* $\mu^0 \in \mathcal{V}_q^W$ *we have that*

$$\rho_q(P_s\mu^0,\mu^*) \leq e^{-\frac{1}{q}\lambda s}\rho_q(\mu^0,\mu^*)\,. \tag{15}$$

Proof outline for Lemma 15: Show that $\left(\mathcal{V}_q^W, \rho_q\right)$ is a complete metric space. Use Banach's Fixed point theorem on the linearised solution map Ψ given by $\mu \mapsto \{\mathcal{L}(\theta_{s,\cdot}(\mu) \,|\, W(\omega^W)) : \omega^W \in \Omega^W, s \in I\}$ with

$$d\theta_{s,t}(\mu) = -\left((\nabla_a \frac{\delta \mathbf{H}_t^0}{\delta m})(\theta_{s,t}(\mu), \mu_{s,t}) + \frac{\sigma^2}{2}(\nabla_a U)(\theta_{s,t}(\mu)) \right) ds + \sigma\, dB_s\,. \quad (16)$$

To get contraction apply Itô's formula:

$$\begin{aligned} d\Big(e^{\lambda s}|\theta_{s,t}(\mu) - \theta_{s,t}(\mu')|^q\Big) = \; & e^{\lambda s}\Big[\lambda|\theta_{s,t}(\mu) - \theta_{s,t}(\mu')|^q \\ & - \frac{q}{2}(\theta_{s,t}(\mu) - \theta_{s,t}(\mu'))\Big(\sigma^2\Big[(\nabla_a U)(\theta_{s,t}(\mu)) - (\nabla_a U)(\theta_{s,t}(\mu'))\Big] \\ & + 2\Big[(\nabla_a \frac{\delta \mathbf{H}_t^0}{\delta m})(\theta_{s,t}(\mu), \mu_{s,\cdot}) - (\nabla_a \frac{\delta \mathbf{H}_t^0}{\delta m})(\theta_{s,t}(\mu'), \mu'_{s,\cdot})\Big]\Big)|\theta_{s,t}(\mu) - \theta_{s,t}(\mu')|^{q-2}\Big] ds\,. \end{aligned}$$

Assumption 5 is needed. Get

$$e^{\lambda S}\rho_q(\Psi(\mu)_s, \Psi(\mu')_s)^q \le c_{q,\kappa,\sigma,\eta_1,\eta_2} \int_0^S e^{\lambda s}\rho_q(\mu_{s,\cdot}, \mu'_{s,\cdot})^q ds\,. \quad (17)$$

Lemma 17

Let Assumptions 3 and 5 hold. If $\lambda = \frac{q}{2}\left(\sigma^2\kappa + \eta_1 - \eta_2\right) \geq 0$ *and if* $\mu^0, \bar{\mu}^0 \in \mathcal{V}_q^W$, *then for all* $s \geq 0$ *we have*

$$\rho_q(P_s\mu^0, P_s\bar{\mu}^0) \leq e^{-\frac{1}{q}\lambda s}\rho_q(\mu^0, \bar{\mu}^0)\,. \tag{18}$$

Proof outline: similar calculation with Itô formula and using Assumption 5 as above.

Proof outline for Theorem 16 (unique invariant measure exists and we have exponential convergence):

Choose $s_0 > 0$ such that $e^{-\frac{1}{q}\lambda s_0} < 1$. Then $P_{s_0} : \mathcal{V}_q^W \to \mathcal{V}_q^W$ is a contraction due to Lemma 17. By Banach's fixed point theorem there is a (unique) $\tilde{\mu} \in \mathcal{V}_q^W$ such that $P_{s_0}\tilde{\mu} = \tilde{\mu}$.

Let $\mu^* := \int_0^{s_0} P_s\tilde{\mu}\, ds$. Take an arbitrary $r \geq 0$ and show that

$$P_r\mu^* = \mu^* \,.$$

Consider $\nu^* \neq \mu^*$ such that $P_r\nu^* = \nu^*$ for any $r \geq 0$. Then from Lemma 17 we have, for any $r > s_0$, that

$$\rho_q(\mu^*, \nu^*) = \rho_q(P_r\mu^*, P_r\nu^*) \leq e^{-\frac{1}{q}\lambda r}\rho_q(\mu^*, \nu^*)$$

which is a contradiction as $e^{-\frac{1}{q}\lambda r} < 1$.

When are Structural Conditions Met

Example:

$$X_t(\mu) = \xi + \int_0^t \Phi_r(X_r(\mu), \mu_r)\, dr + \Gamma\, W_t\,, \quad t \in [0, T]\,.$$

The BSDE is (no dependence on Z in driver)

$$\begin{aligned} dY_t(\mu) &= -(\nabla_x H_t^0)(X_t(\mu), Y_t(\mu), \mu_t)\, dt + Z_t(\mu)\, dW_t\,, \quad t \in [0, T]\,, \\ Y_T(\mu) &= (\nabla_x g)(X_T(\mu))\,. \end{aligned}$$

Objective

$$J^\sigma(\nu, \xi) := \mathbb{E}^W \left[\int_0^T \left[\tilde{F}_t(X_t(\nu), \nu_t) + \bar{F}_t(\nu_t) + \frac{\sigma^2}{2} \mathrm{Ent}(\nu_t) \right] dt + g(X_T(\nu)) \Big| X_0(\nu) = \xi \right]$$

with $\bar{F}$ strictly convex.

Lemma 18

Assume sufficient regularity and bounds on coefficients. Let $T \geq s > t \geq 0$. Then there exists constant $c_{q,T} > 0$ such that

$$\mathbb{E}\left[|X_s(\mu) - X_s(\nu)|^q \mid \mathcal{F}_t\right] \leq c_{q,T}\left(|X_t(\mu) - X_t(\nu)|^q + \int_t^s \mathbb{E}\left[(\mathcal{W}_1(\mu_r, \nu_r))^q \mid \mathcal{F}_t^W\right] dr\right)$$

Lemma 19 (BSDE Estimates)

Assume sufficient regularity and bounds on coefficients. Then

$$\sup_{\mu \in \mathcal{V}_q^W} \sup_{t \in [0,T]} \|Y_t(\mu)\|_\infty < \infty .$$

Furthermore, there exists a constant $c > 0$ such that

$$\begin{aligned}|Y_t(\mu) - Y_t(\nu)| + \leq c\,\mathbb{E}\Bigg[&|X_T(\mu) - X_T(\nu)| \\ &+ \int_t^T \left[\mathcal{W}_1(\nu_r, \mu_r) + |X_r(\mu) - X_r(\nu)|\right] dr \Bigg| \mathcal{F}_t^W\Bigg] .\end{aligned} \tag{19}$$

References

[1] Chizat, L., and Bach, F. On the global convergence of gradient descent for over-parameterized models using optimal transport. In *Advances in neural information processing systems* (2018), pp. 3040–3050.

[2] Hu, K., Ren, Z., Šiška, D., and Szpruch, L. Mean-field Langevin dynamics and energy landscape of neural networks. *arXiv:1905.07769* (2019).

[3] Mei, S., Montanari, A., and Nguyen, P.-M. A mean field view of the landscape of two-layer neural networks. *Proceedings of the National Academy of Sciences 115*, 33 (2018), E7665–E7671.

[4] Reisinger, C., and Zhang, Y. Regularity and stability of feedback relaxed controls. *arXiv preprint arXiv:2001.03148* (2020).

[5] Rotskoff, G. M., and Vanden-Eijnden, E. Neural networks as interacting particle systems: Asymptotic convexity of the loss landscape and universal scaling of the approximation error. *arXiv:1805.00915* (2018).

[6] Šiška, D., and Szpruch, Ł. Gradient flows for regularized stochastic control problems. *arXiv preprint arXiv:2006.05956* (2020).

[7] Wang, H., Zariphopoulou, T., and Zhou, X. Y. Exploration versus exploitation in reinforcement learning: a stochastic control approach. *Available at SSRN 3316387* (2019).

[8] Wang, H., and Zhou, X. Y. Continuous-time mean-variance portfolio selection: A reinforcement learning framework. *Mathematical Finance 30* (2020), 1273–1308.

Thank you!