

Convergence of Policy Improvement for Entropy-Regularized Stochastic Control Problems

Yu-Jui Huang
University of Colorado, Boulder

Joint work with
Zhenhua Wang (*University of Michigan*)
Zhou Zhou (*University of Sydney*)



Probability Seminar @ National Central University
June 9, 2023

CLASSICAL STOCHASTIC CONTROL

► **Controlled dynamics:**

$$dX_s^\alpha = b(X_s^\alpha, \alpha_s)ds + \sigma(X_s^\alpha)dW_s, \quad X_0 = x \in \mathbb{R}^d,$$

► $\alpha \in \mathcal{A}$: U -valued adapted process, with $U \subset \mathbb{R}^\ell$.

► **Problem formulation:**

$$V^*(x) := \sup_{\alpha \in \mathcal{A}} V^\alpha(x),$$

where

$$V^\alpha(x) := \mathbb{E} \left[\int_0^\infty e^{-\rho s} r(X_s^\alpha, \alpha_s) ds \right].$$

ENTROPY REGULARIZATION

► **Controlled dynamics:**

$$dX_s^\pi = \left(\int_U b(X_s^\pi, u) \pi_s(u) du \right) ds + \sigma(X_s^\pi) dW_s, \quad X_0 = x \in \mathbb{R}^d, \quad (1)$$

- $\pi \in \mathcal{A}$: $\mathcal{P}(U)$ -valued adapted process (i.e., *relaxed control*).
- $\mathcal{P}(U)$: the set of density functions on U .

► **Problem formulation:**

$$V^*(x) := \sup_{\pi \in \mathcal{A}} V^\pi(x), \quad (2)$$

where

$$V^\pi(x) := \mathbb{E} \left[\int_0^\infty e^{-\rho s} \left(\int_U r(X_s^\pi, u) \pi_s(u) du - \lambda \int_U \pi_s(u) \ln \pi_s(u) du \right) ds \right]$$

- $-\int_U \pi_s(u) \ln \pi_s(u) du$: entropy of density $\pi_s \in \mathcal{P}(U)$.
- $\lambda > 0$: weight for *exploration* (versus *exploitation*).

LITERATURE

- ▶ **Principle of maximum entropy:**
 - ▶ Jaynes (1957 (a), 1957 (b)); Shannon (1948).
- ▶ **Applications to reinforcement learning:**
 - ▶ The “*soft-max*” criterion in
Ziebart et al. (2008), Ziebart et al. (2010), Fox et al. (2016),
Haarnoja et al. (2017), and Haarnoja et al. (2018), ...
 - ▶ discrete-time algorithms, limited theoretic justification
- ▶ **Continuous-time HJB analysis:**
 - ▶ Initiated by Wang, Zariphopoulou, & Zhou (2020).
 - ▶ Actively researched by
Wang & Zhou (2020), Firoozi & Jaimungal (2022), Guo et al.
(2022), Reisinger & Zhang (2021), Tang et al. (2022), ...

THE EXPLORATORY HJB

- ▶ The (exploratory) HJB equation for V^* in (2) is

$$\rho v = \sup_{\varpi \in \mathcal{P}(U)} \int_U \left(b(x, u) \cdot D_x v(x) + r(x, u) - \lambda \ln \varpi(u) \right) \varpi(u) du + \frac{1}{2} \text{tr}((\sigma \sigma') D_x^2 v)(x). \quad (3)$$

- ▶ Candidate optimizer $\pi^* \in \mathcal{A}$ is in the **Gibbs form**

$$\pi^*(x, u) := \Gamma(x, D_x v(x), u), \quad (4)$$

where $\Gamma : \mathbb{R}^d \times \mathbb{R}^d \times U \rightarrow \mathbb{R}$ is defined by

$$\Gamma(x, y, u) = \frac{\exp(\frac{1}{\lambda} [b(x, u) \cdot y + r(x, u)])}{\int_U \exp(\frac{1}{\lambda} [b(x, \tilde{u}) \cdot y + r(x, \tilde{u})]) d\tilde{u}}. \quad (5)$$

THE PIA

Policy Improvement Algorithm (PIA)

1. Take an arbitrary $v^0 \in \mathcal{C}_{\text{unif}}^{1,1}(\mathbb{R}^d)$.
2. For each $n \in \mathbb{N}$, introduce

$$\pi^n(x, u) := \Gamma(x, D_x v^{n-1}(x), u) \in \mathcal{A} \quad \text{and} \quad v^n := V^{\pi^n}. \quad (6)$$

Our Goal:

Convergence of PIA to optimality.

- 1) $v^{n+1} \geq v^n \forall n \in \mathbb{N}$. (*Policy improvement works*)
- 2) $v^n \uparrow V^*$ as $n \rightarrow \infty$. (*Policy improvement achieves optimum*)
- 3) $\pi^*(x, u) := \Gamma(x, D_x V^*(x), u) \in \mathcal{A}$ is optimal.

FUNCTION SPACES

- Given $E \subseteq \mathbb{R}^d$, $k \in \mathbb{N}_0 := \{0, 1, 2, \dots\}$, and $0 < \alpha \leq 1$,

$$\|f\|_{C^k(E)} := \sum_{j=0, \dots, k} \sum_{|a|_{l_1}=j} \|D_x^a f\|_{L^\infty(E)},$$

$$\|f\|_{C^{k,\alpha}(E)} := \|f\|_{C^k(E)} + \sum_{j=0, \dots, k} \sum_{|a|_{l_1}=j} \sup_{x,y \in E} \frac{|D_x^a f(x) - D_x^a f(y)|}{|x - y|^\alpha}.$$

- For $E = \mathbb{R}^d$, we additionally consider

$$\|f\|_{C_{\text{unif}}^{k,\alpha}(\mathbb{R}^d)} := \sup_{x \in \mathbb{R}^d} \|f\|_{C^{k,\alpha}(B_1(x))}.$$

Assume:

- ▶ $U \subset \mathbb{R}^\ell, 0 < \text{Leb}(U) < \infty$.
- ▶ There exists $k \in \mathbb{N}_0 := \{0, 1, 2, \dots\}$ such that

$$\Lambda_k := \sup_{u \in U} \left\{ \|r(\cdot, u)\|_{C^k(\mathbb{R}^d)} + \|b(\cdot, u)\|_{C^k(\mathbb{R}^d)} + \|\sigma\|_{C^k(\mathbb{R}^d)} \right\} \quad (7)$$

is a finite number.

Focus on:

- ▶ Gibbs-form relaxed control π , i.e.,

$$\boxed{\pi(x, u) := \Gamma(x, p(x), u)} \quad \text{for some } p : \mathbb{R}^d \rightarrow \mathbb{R}^d. \quad (8)$$

- ▶ Is π a well-defined relaxed control (or, *admissible*)?
- ▶ What are the properties of V^π ?

Definition

A relaxed control $\pi = (\pi_t)_{t \geq 0}$ is admissible if

- (i) \exists unique strong solution X^π to (1),
- (ii) $|V^\pi(x)| < \infty$ for all $x \in \mathbb{R}^d$.

We denote by \mathcal{A} the set of all admissible relaxed controls.

Proposition

Let $\Lambda_1 < \infty$ in (7). For any $p : \mathbb{R}^d \rightarrow \mathbb{R}^d$ in $C_{\text{unif}}^{0,1}(\mathbb{R}^d)$,

$$\pi(x, u) := \Gamma(x, p(x), u) \in \underline{\underline{\mathcal{A}}}.$$

Moreover, V^π is bounded and continuous on \mathbb{R}^d , with

$$|V^\pi(x)| \leq 3\Lambda_0(1 + \|p\|_{C^0(\mathbb{R}^d)}) + \lambda |\ln(\text{Leb}(U))| \quad \forall x \in \mathbb{R}^d.$$

PDE CHARACTERIZATION

- ▶ Consider

$$\mathcal{H}(x, y, u) := \lambda \ln \Gamma(x, y, u), \quad \forall (x, y, u) \in \mathbb{R}^d \times \mathbb{R}^d \times U.$$

- ▶ For any $f : \mathbb{R}^d \times \mathbb{R}^d \times U \rightarrow \mathbb{R}$, define

$$\hat{f}(x, y) := \int_U f(x, y, u) \Gamma(x, y, u) du \quad \forall (x, y) \in \mathbb{R}^d \times \mathbb{R}^d. \quad (9)$$

- ▶ For $\boxed{\pi(x, u) := \Gamma(x, p(x), u) \in \mathcal{A}}$,

- ▶ V^π can be expressed as

$$V^\pi(x) = \mathbb{E} \left[\int_0^\infty e^{-\rho s} \left(\hat{r}(X_s^\pi, p(X_s^\pi)) - \hat{\mathcal{H}}(X_s^\pi, p(X_s^\pi)) \right) ds \right].$$

- ▶ The elliptic operator for X^π in (1) is

$$\mathcal{L}^\pi w := -\rho w + \hat{b}(\cdot, p(\cdot)) \cdot D_x w + \frac{1}{2} \text{tr}(\sigma \sigma' D_x^2 w), \quad w \in C^2(\mathbb{R}^d).$$

PDE CHARACTERIZATION

Proposition

Fix $\alpha \in (0, 1)$. Let $\Lambda_{k+1} < \infty$ in (7) with $k \in \mathbb{N}_0$ and suppose

$$\text{tr}(\sigma(x)\sigma(x)'\xi'\xi) \geq \eta_0|\xi|^2 \quad \forall \xi, x \in \mathbb{R}^d, \quad (10)$$

for some $\eta_0 > 0$. For any $p : \mathbb{R}^d \rightarrow \mathbb{R}^d$ in $\mathcal{C}_{\text{unif}}^{0,1}(\mathbb{R}^d) \cap \mathcal{C}_{\text{unif}}^{k,\alpha}(\mathbb{R}^d)$,

- (i) $\pi(x, u) := \Gamma(x, p(x), u) \in \mathcal{A}$,
- (ii) $V^\pi \in \mathcal{C}_{\text{unif}}^{k+2,\alpha}(\mathbb{R}^d)$ and satisfies

$$\mathcal{L}^\pi V^\pi(x) + \hat{r}(x, p(x)) - \hat{\mathcal{H}}(x, p(x)) = 0 \quad \forall x \in \mathbb{R}^d.$$

► Based on Gilbarg & Trudinger (1998).

REGULARITY IN PIA

Corollary 1

Let $\Lambda_k < \infty$ in (7) with $k \in \mathbb{N}$ and (10) hold. In the PIA (6),

- (i) $\pi^n \in \mathcal{A}$ for all $n \in \mathbb{N}$;
- (ii) $\{v^n\}_{n \geq 0}$ satisfy, for any $0 < \alpha < 1$,

$$\begin{aligned} v^n &\in C_{\text{unif}}^{n+1, \alpha}(\mathbb{R}^d), & \text{for } 1 \leq n \leq k; \\ v^n &\in C_{\text{unif}}^{k+1, \alpha}(\mathbb{R}^d), & \text{for } n \geq k+1. \end{aligned} \tag{11}$$

Moreover, for each $n \in \mathbb{N}$,

$$\mathcal{L}^{\pi^n} v^n(x) + \hat{r}(x, D_x v^{n-1}(x)) - \hat{\mathcal{H}}(x, D_x v^{n-1}(x)) = 0 \quad \forall x \in \mathbb{R}^d. \tag{12}$$

POLICY IMPROVEMENT

Proposition

Let $\Lambda_1 < \infty$ in (7) and assume (10). For $p : \mathbb{R}^d \rightarrow \mathbb{R}^d$ in $\mathcal{C}_{\text{unif}}^{0,1}(\mathbb{R}^d)$,

$$\pi(x, u) := \Gamma(x, p(x), u) \quad \text{and} \quad \tilde{\pi}(x, u) := \Gamma(x, D_x V^\pi(x), u).$$

satisfy

- (i) $\pi \in \mathcal{A}$ and $V^\pi \in \mathcal{C}_{\text{unif}}^{2,\alpha}(\mathbb{R}^d) \forall 0 < \alpha < 1$;
- (ii) $\tilde{\pi} \in \mathcal{A}$ and $V^{\tilde{\pi}} \in \mathcal{C}_{\text{unif}}^{2,\alpha}(\mathbb{R}^d) \forall 0 < \alpha < 1$;
- (iii) $V^{\tilde{\pi}}(x) \geq V^\pi(x)$ for all $x \in \mathbb{R}^d$.

Consequence: In the PIA (6),

$$v^{n+1} \geq v^n \quad \forall n \in \mathbb{N}.$$

CONVERGENCE OF PIA

Lemma 1

For any $\pi \in \mathcal{A}$, $V^\pi(x) \leq \frac{1}{\rho} (\|r\|_{C^0(\mathbb{R}^d)} + \lambda |\ln(\text{Leb}(U))|) \forall x \in \mathbb{R}^d$.

Idea: Define $\nu \in \mathcal{P}(U)$ by $\nu(u) \equiv 1/\text{Leb}(U)$. For any $f \in \mathcal{P}(U)$,

$$0 \leq D_{KL}(f\|\nu) := \int_U f \ln \left(\frac{f}{\nu} \right) du = \int_U f \ln f du + \ln(\text{Leb}(U)). \quad (13)$$

Corollary 1

Let $\Lambda_1 < \infty$ in (7) and (10) hold. Then, in the PIA (6),

$$v^*(x) := \lim_{n \rightarrow \infty} v^n(x) \quad \forall x \in \mathbb{R}^d \quad (14)$$

is well-defined.

Question: *How to show $v^* = V^*$?*

- ▶ **Idea:** If we have

$$\sup_{n \in \mathbb{N}} \|v^n\|_{\mathcal{C}^2(\mathbb{R}^d)} < \infty, \quad (15)$$

- ▶ v^* will inherit regularity from v^n , i.e., $v^* \in \mathcal{C}^2(\mathbb{R}^d)$.
- ▶ As $n \rightarrow \infty$ in (12), v^* satisfies HJB (3).
- ▶ By verification argument, $v^* = V^*$.
- ▶ In classical stochastic control,
 - ▶ (15) holds by standard estimates; see e.g., Krylov (1980).
 - ▶ “ $v^* = V^*$ ” established (Puterman (1981), Jacka & Mijatović (2017), Kerimkulov, Siska, & Szpruch (2020)).
- ▶ **Challenge:** With entropy regularization,
no standard estimates to use...

- ▶ **Already known:** $\sup_{n \in \mathbb{N}} \|v^n\|_{C^0(\mathbb{R}^d)} < \infty$.
- ▶ **Remains to show:**

$$\sup_{n \in \mathbb{N}} \|D_x v^n\|_{C^1(\mathbb{R}^d)} = \sup_{n \in \mathbb{N}} \left(\|D_x v^n\|_{C^0(\mathbb{R}^d)} + \|D_x^2 v^n\|_{C^0(\mathbb{R}^d)} \right) < \infty.$$

- ▶ **Observe:** *Schauder estimates don't help!*

$$\begin{aligned} \|D_x v^n\|_{C^1(\mathbb{R}^d)} &\leq \|D_x v^n\|_{C_{\text{unif}}^{1,\alpha}(\mathbb{R}^d)} \\ &\leq K_n \left(1 + \|\hat{\mathcal{H}}(\cdot, D_x v^{n-1}(\cdot))\|_{C_{\text{unif}}^{0,\alpha}(\mathbb{R}^d)} \right) \\ &\leq K_n \left(1 + \|\hat{\mathcal{H}}(\cdot, D_x v^{n-1}(\cdot))\|_{C^1(\mathbb{R}^d)} \right). \end{aligned}$$

- ▶ $K_n = g(\|D_x v^{n-1}\|_{C_{\text{unif}}^{0,\alpha}(\mathbb{R}^d)})$, with g unknown.
- ▶ $\|\hat{\mathcal{H}}(\cdot, D_x v^{n-1}(\cdot))\|_{C^1(\mathbb{R}^d)}$ grows linearly in $\|D_x v^{n-1}\|_{C^1(\mathbb{R}^d)}$.

THE GRAND PLAN

Set $s := \lfloor d/2 \rfloor + 1$. For $n > s$,

1. Bound $|D_x v^n(x)| + |D_x^2 v^n(x)|$ by $\|v^n\|_{W^{s+2,2}(B_1(x))}$.
2. Bound $\|v^n\|_{W^{s+2,2}(B_1(x))}$ by a power function of the sum of

$$\|v^{n-i}\|_{L^q(B_\eta(x))}, \quad i = 0, 1, \dots, s-1, \quad \text{and} \quad \|\hat{\mathcal{H}}(\cdot, D_x v^{n-s-1}(\cdot))\|_{L^q(B_\eta(x))}.$$

3. Show that for some ψ of logarithmic growth,

$$|\hat{\mathcal{H}}(\cdot, D_x v^{n-s-1}(\cdot))| \leq \psi(|D_x v^{n-s-1}(\cdot)|).$$

4. Combine Steps 1-3 to get

$$|D_x v^n(x)| + |D_x^2 v^n(x)| \leq C(1 + |D_x^2 v^{n-s-1}(x)|).$$

Turn this local estimate into a global one, i.e.,

$$\|D_x v^n\|_{C^1(\mathbb{R}^d)} \leq C(1 + \|D_x v^{n-s-1}\|_{C^1(\mathbb{R}^d)}).$$

STEP 1

- Evans (1998): for any $w \in W^{s,2}(\mathbb{R}^d)$ with $s > d/2$,

$$\|w\|_{L^\infty(\mathbb{R}^d)} \leq C\|w\|_{W^{s,2}(\mathbb{R}^d)}, \quad (16)$$

where $C > 0$ depends on only s and d .

- With $s := \lfloor d/2 \rfloor + 1$, a *localized version* of (16) gives

$$|D_x v^n(x)| + |D_x^2 v^n(x)| \leq C\|v^n\|_{W^{s+2,2}(B_1(x))}, \quad (17)$$

where $C > 0$ depends on only s and d .

STEP 2: ORDER REDUCTION

Fix $E \subseteq \mathbb{R}^d$. Consider the elliptic operator

$$L^{(\gamma, \beta, \delta)} w := -\gamma w + \beta \cdot D_x w + \frac{1}{2} \operatorname{tr}(\delta \delta' D_x^2 w), \quad \forall w \in C^2(E). \quad (18)$$

Theorem (Chen & Wu (1998))

Fix $\beta \in C^0(E)$ and $\delta \in C^0(\bar{E})$ that is uniformly elliptic. Let $w \in W^{2,q}(E)$, $q \geq 1$, be a solution to $L^{(\gamma, \beta, \delta)} w = f$ in E .

Then, for any $E' \subset\subset E$,

$$\|w\|_{W^{2,q}(E')} \leq C(\|w\|_{L^q(E)} + \|f\|_{L^q(E)}), \quad (19)$$

where $C > 0$ depends on d , q , $\operatorname{dist}(E', \partial E)$, γ , β and δ .

► **Order reduction:** order 2 \implies order 0.

Lemma 2

There exist three nondecreasing functions

$$A(k, q) : \mathbb{N}_0 \times [1, \infty) \rightarrow \mathbb{R}_+ \text{ with } A(k, q) \geq q,$$

$$Q(k, q) : \mathbb{N}_0 \times [1, \infty) \rightarrow \mathbb{N}_0,$$

$$R(k, \eta) : \mathbb{N}_0 \times \mathbb{R}_+ \rightarrow \mathbb{R}_+ \text{ with } R(k, \eta) \geq 2\eta,$$

such that whenever $\beta \in \mathcal{C}^k(\mathbb{R}^d)$ and $\delta \in \mathcal{C}^{k \vee 1}(\mathbb{R}^d)$ for some $k \in \mathbb{N}_0$ and δ is uniformly elliptic, the following holds:

Given $f \in \mathcal{C}^k(\mathbb{R}^d)$, a solution $w \in \mathcal{C}^{k+2}(\mathbb{R}^d)$ to $L^{(\gamma, \beta, \delta)}w = f$ in \mathbb{R}^d fulfills

$$\begin{aligned} \|w\|_{W^{k+2, q}(B_\eta(x))} &\leq H_k \left(1 + \|\beta\|_{W^{k, A(k, q)}(B_{R(k, \eta)}(x))} \right)^{Q(k, q)} \\ &\quad \cdot \left(\|w\|_{L^{A(k, q)}(B_{R(k, \eta)}(x))} + \|f\|_{W^{k, A(k, q)}(B_{R(k, \eta)}(x))} \right), \quad (20) \end{aligned}$$

for all $x \in \mathbb{R}^d$ and $(q, \eta) \in [1, \infty) \times \mathbb{R}_+$, where $H_k > 0$ depends on $k, q, \eta, d, \gamma, \beta$, and δ .

- ▶ Lemma 2 generalizes standard estimate (19) in two important ways:
 - ▶ **Order reduction:** order $k + 2 \implies$ order k .
 - ▶ **Explicit dependence on drift** β , through the term

$$\left(1 + \|\beta\|_{W^{k,A(k,q)}(B_{R(k,\eta)}(x))}\right)^{Q(k,q)}.$$

Apply Lemma 2 recursively to PIA (6):

- ▶ $(\gamma, \beta, \delta) \implies (\rho, \hat{b}(\cdot, D_x v^{n-1}(\cdot)), \sigma)$.
- ▶ As v^n is a solution to (12),

$$w \implies v^n, \quad f(\cdot) \implies -\hat{r}(\cdot, D_x v^{n-1}(\cdot)) + \hat{\mathcal{H}}(\cdot, D_x v^{n-1}(\cdot))$$

- ▶ (20) becomes

$$\begin{aligned} & \|v^n\|_{W^{s+2,q}(B_\eta(x))} \\ & \leq H_s \left(1 + \|\hat{b}(\cdot, D_x v^{n-1}(\cdot))\|_{W^{s,A(s,q)}(B_{R(s,\eta)}(x))} \right)^{Q(s,q)} \\ & \quad \cdot \left(\|v^n\|_{L^{A(s,q)}(B_{R(s,\eta)}(x))} \right. \\ & \quad \left. + \|\hat{r}(\cdot, D_x v^{n-1}(\cdot)) - \hat{\mathcal{H}}(\cdot, D_x v^{n-1}(\cdot))\|_{W^{s,A(s,q)}(B_{R(s,\eta)}(x))} \right). \end{aligned} \tag{21}$$

Lemma 3

Let $\Lambda_k < \infty$ in (7) and $p : \mathbb{R}^d \rightarrow \mathbb{R}^d$ be in $C^k(\mathbb{R}^d)$ for some $k \in \mathbb{N}_0$. Consider the function

$$f(x, y, u) := r(x, u) - \mathcal{H}(x, y, u) \quad (\text{or, } f(x, y, u) := b(x, u)).$$

Then, for any $x_0 \in \mathbb{R}^d$, $q \geq 1$, and $\eta > 0$,

$$\|\hat{f}(\cdot, p(\cdot))\|_{W^{k,q}(B_\eta(x_0))} \leq C_k \left(1 + \|p\|_{W^{k,(k+1)2^{k+1}q}(B_\eta(x_0))} \right)^{(k+1)2^{k+1}},$$

where $C_k > 0$ depends on only k , Λ_k , λ , d , q , and η .

- Apply Lemma 3 to (21) gives

$$\begin{aligned} & \|v^n\|_{W^{s+2,q}(B_\eta(x))} \\ & \leq \tilde{H}_s \left(1 + \|v^n\|_{L^{A(s,q)}(B_{R(s,\eta)}(x))} \right. \\ & \quad \left. + \|v^{n-1}\|_{W^{s+1,q_s A(s,q)}(B_{R(s,\eta)}(x))} \right)^{a_s(Q(s,q)+1)}. \end{aligned}$$

- Repeating this procedure s times yields

$$\begin{aligned} \|v^n\|_{W^{s+2,q}(B_\eta(x))} & \leq H^* \left(1 + \sum_{i=0}^{s-1} \|v^{n-i}\|_{L^{q_{i+1}}(B_{\eta_{i+1}}(x))} \right. \\ & \quad \left. + \|v^{n-s}\|_{W^{2,q_s}(B_{\eta_s}(x))} \right)^{\theta_1 \theta_2 \dots \theta_s}. \quad (22) \end{aligned}$$

- Apply Lemma 2 to $\|\vartheta^{n-s}\|_{W^{2,q_s}(B_{\eta_s}(x))}$:

$$\begin{aligned} & \|\vartheta^{n-s}\|_{W^{2,q_s}(B_{\eta_s}(x))} \\ & \leq H_0 \left(1 + \|\hat{b}(\cdot, D_x \vartheta^{n-s-1}(\cdot))\|_{L^{A(0,q_s)}(B_{R(0,\eta_s)}(x))} \right)^{Q(0,q_s)} \\ & \quad \cdot \left(\|\vartheta^{n-s}\|_{L^{A(0,q_s)}(B_{R(0,\eta_s)}(x))} \right. \\ & \quad \left. + \|\hat{r}(\cdot, D_x \vartheta^{n-s-1}(\cdot)) - \hat{\mathcal{H}}(\cdot, D_x \vartheta^{n-s-1}(\cdot))\|_{L^{A(0,q_s)}(B_{R(0,\eta_s)}(x))} \right) \\ & \leq C(1 + \|\hat{\mathcal{H}}(\cdot, D_x \vartheta^{n-s-1}(\cdot))\|_{C^0(\mathbb{R}^d)}), \end{aligned}$$

where the last line holds if $\Lambda_{s+1} = \Lambda_{\lfloor d/2 \rfloor + 2} < \infty$.

- Plugging the above into (22) leads to

$$\|\vartheta^n\|_{W^{s+2,q}(B_{\eta}(x))} \leq C(1 + \|\hat{\mathcal{H}}(\cdot, D_x \vartheta^{n-s-1}(\cdot))\|_{C^0(\mathbb{R}^d)})^{\theta_1 \theta_2 \dots \theta_s}, \quad (23)$$

where r.h.s. is independent of $x \in \mathbb{R}^d$.

Combining Steps 1 & 2 (i.e., (17) and (23)) gives

$$\|D_x v^n\|_{C^1(\mathbb{R}^d)} \leq C^* (1 + \|\hat{\mathcal{H}}(\cdot, D_x v^{n-s-1}(\cdot))\|_{C^0(\mathbb{R}^d)})^{\theta_1 \theta_2 \dots \theta_s}. \quad (24)$$

Question: *How to bound the entropy term $\hat{\mathcal{H}}$?*

- ▶ **Lower bound:** By (13),

$$\hat{\mathcal{H}}(x, y) = \lambda \int_U \ln(\Gamma(x, y, u)) \Gamma(x, y, u) du \geq -\ln(\text{Leb}(U)).$$

- ▶ **Upper bound:** *A known mathematical challenge...*

STEP 3: BOUNDING THE ENTROPY

Assumption 1

$u \mapsto r(x, u), b(x, u)$ are **Lipschitz**, uniformly in x , i.e.,

$$\Theta := \sup_{u_1, u_2 \in U, x \in \mathbb{R}^d} \frac{|r(x, u_1) - r(x, u_2)| + |b(x, u_1) - b(x, u_2)|}{|u_1 - u_2|} < \infty. \quad (25)$$

Assumption 2

There exist $\zeta > 0$ and $\alpha \in (0, \pi/2]$ such that:

$$\forall u \in U, \exists \text{ cone}(u, \alpha) \text{ with } (\text{cone}(u, \alpha) \cap B_\zeta(u)) \subseteq U,$$

where $\text{cone}(u, \alpha)$ is a cone with vertex u and angle α .

- ▶ This is a uniform cone condition on U :
 - ▶ A cone with a fixed size fits into U at any $u \in U$.

Lemma 4

Under Assumptions 1 and 2, $\Gamma(x, y, u)$ in (5) satisfies

$$\Gamma(x, y, u) \leq C^\ell (1 + |y|)^\ell \quad \forall (x, y, u) \in \mathbb{R}^d \times \mathbb{R}^d \times U,$$

where $C > 0$ depends on $\ell, \lambda, \Theta, \zeta$, and α .

► **Proof ideas:**

- As $r(x, u), b(x, u)$ are Lipschitz in u (i.e., Assumption 1),

$$\begin{aligned} \Gamma(x, y, u_0) &= \frac{\exp\left(\frac{1}{\lambda}[r(x, u_0) + b(x, u_0) \cdot y]\right)}{\int_U \exp\left(\frac{1}{\lambda}[r(x, u) + b(x, u) \cdot y]\right) du} \\ &\leq \frac{1}{\int_U \exp\left(-\frac{\Theta}{\lambda}(1 + |y|)|u - u_0|\right) du}. \end{aligned}$$

► **Proof ideas (conti.):**

- By **Assumption 2**,

$$\begin{aligned} \int_U e^{-\frac{\Theta}{\lambda}(1+|y|)|u-u_0|} du &\geq \int_{\text{cone}(u_0, \alpha) \cap B_\zeta(u_0)} e^{-\frac{\Theta}{\lambda}(1+|y|)|u-u_0|} du \\ &= \int_{\text{cone}(0, \alpha) \cap B_\zeta(0)} e^{-\frac{\Theta}{\lambda}(1+|y|)|u|} du \\ &= K_1 \int_0^\zeta r^{\ell-1} e^{-\frac{\Theta}{\lambda}(1+|y|)r} dr \quad (\text{by polar coordinates}) \\ &= K_1 \left(\frac{\lambda}{\Theta(1+|y|)} \right)^\ell \int_0^{\frac{\Theta}{\lambda}(1+|y|)\zeta} z^{\ell-1} e^{-z} dz \\ &\geq K_1 K_2 \left(\frac{\lambda}{\Theta(1+|y|)} \right)^\ell, \quad \text{with } K_2 := \int_0^{\frac{\Theta}{\lambda}\zeta} z^{\ell-1} e^{-z} dz, \end{aligned}$$

where $K_1 > 0$ depends on only ℓ and α .

Corollary

Under Assumptions 1 and 2,

$$\sup_{x \in \mathbb{R}^d} |\hat{\mathcal{H}}(x, y)| \leq \kappa + \lambda \ell \ln(1 + |y|), \quad \forall y \in \mathbb{R}^d, \quad (26)$$

where $\kappa > 0$ depends on only $\ell, \lambda, \text{Leb}(U), \Theta, \zeta$, and α .

► **Proof ideas:**

By (13) and Lemma 4, for any $y \in \mathbb{R}^d$,

$$\begin{aligned} -\lambda \ln(\text{Leb}(U)) \leq \hat{\mathcal{H}}(x, y) &= \lambda \int_U \ln(\Gamma(x, y, u)) \Gamma(x, y, u) du \\ &\leq \lambda(\ell \ln C + \ell \ln(1 + |y|)), \quad \forall x \in \mathbb{R}^d. \end{aligned}$$

By (24) and (26), for any $n \geq s + 2 = \lfloor d/2 \rfloor + 3$,

$$\begin{aligned} \|D_x v^n\|_{C^1(\mathbb{R}^d)} &\leq C^* \left(1 + \|\hat{\mathcal{H}}(\cdot, D_x v^{n-s-1}(\cdot))\|_{C^0(\mathbb{R}^d)} \right)^{\theta_1 \theta_2 \dots \theta_s} \\ &\leq \tilde{C}^* \left(1 + \ln \left(1 + \|D_x v^{n-s-1}\|_{C^0(\mathbb{R}^d)} \right) \right)^{\theta_1 \theta_2 \dots \theta_s} \\ &\leq \phi \left(\|D_x v^{n-s-1}\|_{C^0(\mathbb{R}^d)} \right) \\ &\leq \phi \left(\|D_x v^{n-s-1}\|_{C^1(\mathbb{R}^d)} \right). \end{aligned}$$

- ▶ $\phi(\cdot)$ grows sublinearly (i.e., $\phi(z)/z \rightarrow 0$ as $z \rightarrow \infty$).
 - ▶ Any power of $\ln(\cdot)$ remains sublinear!

Consider $z_0 := \sup\{z \geq 0 : z \leq \phi(z)\} < \infty$ and

$$z^* := \max \left\{ \max_{1 \leq n \leq s+1} \|D_x v^n\|_{C^1(\mathbb{R}^d)}, z_0 \right\} < \infty.$$

Recursively, we have

$$\sup_{s+2 \leq n \leq 2(s+1)} \|D_x v^n\|_{C^1(\mathbb{R}^d)} \leq \phi \left(\sup_{1 \leq n \leq s+1} \|D_x v^n\|_{C^1(\mathbb{R}^d)} \right) \leq \phi(z^*) \leq z^*.$$

$$\sup_{2s+3 \leq n \leq 3(s+1)} \|D_x v^n\|_{C^1(\mathbb{R}^d)} \leq \phi \left(\sup_{s+2 \leq n \leq 2(s+1)} \|D_x v^n\|_{C^1(\mathbb{R}^d)} \right) \leq \phi(z^*) \leq z^*.$$

⋮

Conclude: $\sup_{n \in \mathbb{N}} \|D_x v^n\|_{C^1(\mathbb{R}^d)} \leq z^* < \infty$

Theorem

Let $\Lambda_{\lfloor d/2 \rfloor + 2} < \infty$ in (7), σ be uniformly elliptic, and Assumptions 1 and 2 hold. Then, in the PIA (6),

$$v^* := \lim_{n \rightarrow \infty} v^n$$

satisfies

- (i) v^* is the unique solution in $C^2(\mathbb{R}^d)$ to HJB equation (3);
- (ii) $v^* = V^*$ on \mathbb{R}^d and $\pi^*(x, u) := \Gamma(x, D_x v^*(x), u) \in \mathcal{A}$ is an optimal relaxed control;
- (iii) $v^* \in C_{\text{unif}}^{\lfloor d/2 \rfloor + 3, \alpha}(\mathbb{R}^d)$ for all $0 < \alpha < 1$.

- By Arzela-Ascoli theorem, verification arguments, and (11).

THANK YOU!!

Q & A

Preprint available @ arXiv: 2209.07059

*“Convergence of Policy Improvement for Entropy-Regularized
Stochastic Control Problems”*