Non-asymptotic perspectives on mean field approximations and stochastic control

OR: How to do mean field control without mean field limits

Daniel Lacker

Industrial Engineering and Operations Research, *Columbia University*

March 21, 2024

“Mean field approximations via log-concavity,” joint with:

Sumit Mukherjee
(Columbia)

Lane Chun Yeung
(CMU)

“Approximately optimal distributed stochastic controls beyond the mean field setting,” joint with:

Joe Jackson
(U Chicago)

High-dimensional stochastic control, toy model

Players $i = 1, \ldots, n$ have state processes $\boldsymbol{X} = (X^1, \ldots, X^n)$,

$$dX_t^i = \alpha_i(t, \boldsymbol{X}_t)dt + dW_t^i, \quad \text{valued in } \mathbb{R}^d.$$

$\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n) =$ **Markovian, full-information controls.**

High-dimensional stochastic control, toy model

Players $i = 1, \ldots, n$ have state processes $\boldsymbol{X} = (X^1, \ldots, X^n)$,

$$dX_t^i = \alpha_i(t, \boldsymbol{X}_t)dt + dW_t^i, \quad \text{valued in } \mathbb{R}^d.$$

$\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n) =$ **Markovian, full-information controls.**

Collectively optimize:

$$V := \inf_{\boldsymbol{\alpha}} J(\boldsymbol{\alpha}) = \inf_{\boldsymbol{\alpha}} \mathbb{E}\left[G(\boldsymbol{X}_T) + \frac{1}{2n}\sum_{i=1}^n \int_0^T |\alpha_i(t, \boldsymbol{X}_t)|^2 dt\right]$$

Here $G : (\mathbb{R}^d)^n \to \mathbb{R}$ is arbitrary, say bounded from below.

The usual symmetric case

"Mean field control" case: G takes the form

$$G(\boldsymbol{x}) = \mathcal{G}(m_{\boldsymbol{x}}^n), \qquad m_{\boldsymbol{x}}^n := \frac{1}{n}\sum_{i=1}^{n}\delta_{x_i}, \qquad \mathcal{G} : \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}.$$

The usual symmetric case

"Mean field control" case: G takes the form

$$G(\boldsymbol{x}) = \mathcal{G}(m^n_{\boldsymbol{x}}), \qquad m^n_{\boldsymbol{x}} := \frac{1}{n}\sum_{i=1}^n \delta_{x_i}, \qquad \mathcal{G} : \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}.$$

Mean field limit as $n \to \infty$,

$$V \quad \to \quad \overline{V} := \inf_{\overline{\alpha}} \; \mathcal{G}(\mathrm{Law}(\overline{X}_T)) + \frac{1}{2}\mathbb{E}\int_0^T |\overline{\alpha}(t, \overline{X}_t)|^2 dt,$$
$$d\overline{X}_t = \overline{\alpha}(t, \overline{X}_t)dt + d\overline{W}_t, \quad \text{valued in } \mathbb{R}^d.$$

The usual symmetric case

"Mean field control" case: G takes the form

$$G(\boldsymbol{x}) = \mathcal{G}(m^n_{\boldsymbol{x}}), \qquad m^n_{\boldsymbol{x}} := \frac{1}{n}\sum_{i=1}^n \delta_{x_i}, \qquad \mathcal{G} : \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}.$$

Mean field limit as $n \to \infty$,

$$V \quad \to \quad \overline{V} := \inf_{\overline{\alpha}} \; \mathcal{G}(\mathrm{Law}(\overline{X}_T)) + \frac{1}{2}\mathbb{E}\int_0^T |\overline{\alpha}(t,\overline{X}_t)|^2 dt,$$
$$d\overline{X}_t = \overline{\alpha}(t,\overline{X}_t)dt + d\overline{W}_t, \quad \text{valued in } \mathbb{R}^d.$$

Approximate optimizers for V:

$\alpha_i(t,\boldsymbol{x}) = \overline{\alpha}_*(t,x_i)$, where $\overline{\alpha}_*$ optimal for $\overline{V}$

The usual symmetric case

"Mean field control" case: G takes the form

$$G(\boldsymbol{x}) = \mathcal{G}(m^n_{\boldsymbol{x}}), \qquad m^n_{\boldsymbol{x}} := \frac{1}{n}\sum_{i=1}^n \delta_{x_i}, \qquad \mathcal{G} : \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}.$$

Mean field limit as $n \to \infty$,

$$V \quad \to \quad \overline{V} := \inf_{\overline{\alpha}} \ \mathcal{G}(\mathrm{Law}(\overline{X}_T)) + \frac{1}{2}\mathbb{E}\int_0^T |\overline{\alpha}(t, \overline{X}_t)|^2 dt,$$
$$d\overline{X}_t = \overline{\alpha}(t, \overline{X}_t)dt + d\overline{W}_t, \quad \text{valued in } \mathbb{R}^d.$$

Approximate optimizers for V:
$\alpha_i(t, \boldsymbol{x}) = \overline{\alpha}_*(t, x_i)$, where $\overline{\alpha}_*$ optimal for $\overline{V}$

These approximate optimizers are distributed! (or *decentralized*)

Beyond the symmetric case

For general $G : (\mathbb{R}^d)^n \to \mathbb{R}$, no mean field limit available.

What can be done?

Beyond the symmetric case

For general $G : (\mathbb{R}^d)^n \to \mathbb{R}$, no mean field limit available.

What can be done?

Guiding example: Heterogeneous interactions,

$$G(\boldsymbol{x}) = \frac{1}{n}\sum_{i=1}^{n} G_i(\boldsymbol{x}), \quad G_i(\boldsymbol{x}) := U(x_i) + \frac{1}{2}\sum_{j\neq i} J_{ij} K(x_i - x_j),$$

where $J \in \mathbb{R}^{n\times n}$, and K is an even function.

Beyond the symmetric case

For general $G : (\mathbb{R}^d)^n \to \mathbb{R}$, no mean field limit available.

What can be done?

Guiding example: Heterogeneous interactions,

$$G(\boldsymbol{x}) = \frac{1}{n}\sum_{i=1}^{n} G_i(\boldsymbol{x}), \quad G_i(\boldsymbol{x}) := U(x_i) + \frac{1}{2}\sum_{j \neq i} J_{ij} K(x_i - x_j),$$

where $J \in \mathbb{R}^{n \times n}$, and K is an even function. Alternatively:

$$G(\boldsymbol{x}) = \frac{1}{n}\sum_{i=1}^{n} U(x_i) + \frac{1}{n}\sum_{1 \le i < j \le n} J_{ij} K(x_i - x_j)$$

Ex A: Usual case is $J_{ij} = 1/n$

Ex B: J = scaled adjacency matrix of a graph, $J_{ij} = (1/d_i)1_{i \sim j}$

Beyond the symmetric case

For general $G : (\mathbb{R}^d)^n \to \mathbb{R}$, no mean field limit available.
What can be done?

Guiding example: Heterogeneous interactions,

$$G(\boldsymbol{x}) = \frac{1}{n}\sum_{i=1}^{n} G_i(\boldsymbol{x}), \quad G_i(\boldsymbol{x}) := U(x_i) + \frac{1}{2}\sum_{j\neq i} J_{ij} K(x_i - x_j),$$

where $J \in \mathbb{R}^{n\times n}$, and K is an even function. Alternatively:

$$G(\boldsymbol{x}) = \frac{1}{n}\sum_{i=1}^{n} U(x_i) + \frac{1}{n}\sum_{1\leq i<j\leq n} J_{ij} K(x_i - x_j)$$

Ex A: Usual case is $J_{ij} = 1/n$
Ex B: J = scaled adjacency matrix of a graph, $J_{ij} = (1/d_i)1_{i\sim j}$

Related: recent work on graphon limits of particle systems/games

The distributed optimal control problem

Recall:

$$V = \inf_{\boldsymbol{\alpha}} J(\alpha) = \inf_{\boldsymbol{\alpha}} \mathbb{E}\left[G(\boldsymbol{X}_T) + \frac{1}{2n} \sum_{i=1}^{n} \int_0^T |\alpha_i(t, \boldsymbol{X}_t))|^2 dt \right]$$

Distributed control problem, definition:

$$V_{\mathrm{dstr}} = \inf_{\boldsymbol{\alpha} \ \mathrm{dstr}} J(\alpha)$$

where inf is over controls of the form $\alpha_i(t, \boldsymbol{X}_t) = \tilde{\alpha}_i(t, X_t^i)$.

The distributed optimal control problem

Recall:

$$V = \inf_{\boldsymbol{\alpha}} J(\alpha) = \inf_{\boldsymbol{\alpha}} \mathbb{E}\left[G(\boldsymbol{X}_T) + \frac{1}{2n}\sum_{i=1}^{n}\int_0^T |\alpha_i(t, \boldsymbol{X}_t))|^2 dt \right]$$

Distributed control problem, definition:

$$V_{\mathrm{dstr}} = \inf_{\boldsymbol{\alpha}\ \mathrm{dstr}} J(\alpha)$$

where inf is over controls of the form $\alpha_i(t, \boldsymbol{X}_t) = \tilde{\alpha}_i(t, X_t^i)$.

Questions:

- When are V and V_{dstr} close?
- How do we construct a (near-)optimal distributed control?
- General theory for distributed control problems?

Related litearture

Related perspectives:

- Seguret-Alasseur-Bonnans-De Paola-Oudjane-Trovato '23
- Stochastic teams and information structures. Yüksel, Saldi, Basar...

Related litearture

Related perspectives:

- Seguret-Alasseur-Bonnans-De Paola-Oudjane-Trovato '23
- Stochastic teams and information structures. Yüksel, Saldi, Basar...

Warning: There are different meanings of the term "distributed" in the control literature.

First sentence of a 1973 survey by J.L. Lions defines "distributed systems" as "systems for which the state can be described by a solution of a partial differential equation"...

Usage in this talk is common in mean field game literature, at least.

Comparing the distributed and original problems

Theorem (L.-Mukherjee-Yeung '22)

Let $G : (\mathbb{R}^d)^n \to \mathbb{R}$ *be* C^2 *convex*, $\|\nabla^2 G\|_\infty < \infty$. *Then*

$$0 \le V_{\mathrm{dstr}} - V \le nT^2 \sum_{1 \le i < j \le n} \|\partial_{ij} G\|_\infty^2 =: RHS.$$

Comparing the distributed and original problems

Theorem (L.-Mukherjee-Yeung '22)

Let $G : (\mathbb{R}^d)^n \to \mathbb{R}$ *be* C^2 *convex*, $\|\nabla^2 G\|_\infty < \infty$. *Then*

$$0 \le V_{\mathrm{dstr}} - V \le nT^2 \sum_{1 \le i < j \le n} \|\partial_{ij} G\|_\infty^2 =: RHS.$$

Ex 1: $G(\boldsymbol{x}) = \frac{1}{n} \sum_{i=1}^n U_i(x_i) \rightsquigarrow$ RHS$= 0$

Intuition: RHS measures "how close" the function G is to being additively separable

Side note on deterministic controls

A related result to help with intuition:

Define V_{det} like V_{dstr} but with the further restriciton that controls are deterministic, i.e., solely time-dependent: $\alpha_i(t, \boldsymbol{x}) = \tilde{\alpha}_i(t)$.

Side note on deterministic controls

A related result to help with intuition:

Define V_{det} like V_{dstr} but with the further restriciton that controls are deterministic, i.e., solely time-dependent: $\alpha_i(t, \boldsymbol{x}) = \tilde{\alpha}_i(t)$.

Proposition (L.-Mukherjee-Yeung '22)

Under same assumptions:

$$0 \leq V_{\mathrm{det}} - V \leq \frac{1}{2} n T^2 \sum_{i=1}^{n} \sum_{j=1}^{n} \|\partial_{ij} G\|_\infty^2 =: RHS_2$$

Summation now includes diagonal terms $i = j$!

Side note on deterministic controls

A related result to help with intuition:

Define V_{det} like V_{dstr} but with the further restriciton that controls are deterministic, i.e., solely time-dependent: $\alpha_i(t, \boldsymbol{x}) = \tilde{\alpha}_i(t)$.

Proposition (L.-Mukherjee-Yeung '22)

Under same assumptions:

$$0 \leq V_{\mathrm{det}} - V \leq \frac{1}{2} n T^2 \sum_{i=1}^{n} \sum_{j=1}^{n} \|\partial_{ij} G\|_\infty^2 =: RHS_2$$

Summation now includes diagonal terms $i = j$!

Intuition: RHS_2 measures "how close" G is to being affine

Comparing the distributed and original problems

Theorem (L.-Mukherjee-Yeung '22)

Let $G : (\mathbb{R}^d)^n \to \mathbb{R}$ *be* C^2 *convex*, $\|\nabla^2 G\|_\infty < \infty$. *Then*

$$0 \le V_{\mathrm{dstr}} - V \le nT^2 \sum_{1 \le i < j \le n} \|\partial_{ij} G\|_\infty^2 =: RHS.$$

Ex 2, symmetric case:

Comparing the distributed and original problems

Theorem (L.-Mukherjee-Yeung '22)

Let $G : (\mathbb{R}^d)^n \to \mathbb{R}$ *be* C^2 *convex*, $\|\nabla^2 G\|_\infty < \infty$. *Then*

$$0 \le V_{\mathrm{dstr}} - V \le nT^2 \sum_{1 \le i < j \le n} \|\partial_{ij} G\|_\infty^2 =: \mathit{RHS}.$$

Ex 2, symmetric case: Let $G(\boldsymbol{x}) = F(\frac{1}{n}\sum_{i=1}^n f(x_i))$,

$$\rightsquigarrow \partial_{ij} G(\boldsymbol{x}) = \frac{1}{n^2} F''\Big(\frac{1}{n}\sum_{i=1}^n f(x_i)\Big) f'(x_i) f'(x_j), \quad \text{for } i \neq j$$

$$\rightsquigarrow \mathrm{RHS} \le \frac{T^2}{2n} \|F''\|_\infty^2 \|f\|_\infty^4$$

Comparing the distributed and original problems

Theorem (L.-Mukherjee-Yeung '22)

Let $G : (\mathbb{R}^d)^n \to \mathbb{R}$ *be* C^2 *convex*, $\|\nabla^2 G\|_\infty < \infty$. *Then*

$$0 \leq V_{\mathrm{dstr}} - V \leq nT^2 \sum_{1 \leq i < j \leq n} \|\partial_{ij} G\|_\infty^2 =: RHS.$$

Heterogeneous interactions: U, K convex, K even, $J_{ij} \geq 0$,

$$G(\boldsymbol{x}) = \frac{1}{n} \sum_{i=1}^{n} U(x_i) + \frac{1}{n} \sum_{1 \leq i < j \leq n} J_{ij} K(x_i - x_j)$$

$$\rightsquigarrow \|\partial_{ij} G\|_\infty = \frac{1}{n} J_{ij} \|K''\|_\infty$$

$$\rightsquigarrow \mathrm{RHS} \leq \frac{T^2}{2n} \|K''\|_\infty^2 \mathrm{tr}(J^2)$$

Comparing the distributed and original problems

Theorem (L.-Mukherjee-Yeung '22)

Let $G : (\mathbb{R}^d)^n \to \mathbb{R}$ *be* C^2 *convex,* $\|\nabla^2 G\|_\infty < \infty$*. Then*

$$0 \le V_{\mathrm{dstr}} - V \le nT^2 \sum_{1 \le i < j \le n} \|\partial_{ij} G\|_\infty^2 =: RHS.$$

Heterogeneous interactions: U, K convex, K even, $J_{ij} \ge 0$,

$$G(\boldsymbol{x}) = \frac{1}{n}\sum_{i=1}^{n} U(x_i) + \frac{1}{n}\sum_{1 \le i < j \le n} J_{ij} K(x_i - x_j)$$

$$\rightsquigarrow \mathrm{RHS} \le \frac{T^2}{2n}\|K''\|_\infty^2 \mathrm{tr}(J^2)$$

Key condition: $\mathrm{tr}(J^2) = o(n)$. (cf. Basak-Mukherjee '17)

Comparing the distributed and original problems

Theorem (L.-Mukherjee-Yeung '22)

Let $G : (\mathbb{R}^d)^n \to \mathbb{R}$ *be* C^2 *convex*, $\|\nabla^2 G\|_\infty < \infty$. *Then*

$$0 \le V_{\mathrm{dstr}} - V \le nT^2 \sum_{1 \le i < j \le n} \|\partial_{ij} G\|_\infty^2 =: RHS.$$

Heterogeneous interactions: U, K convex, K even, $J_{ij} \ge 0$,

$$G(\boldsymbol{x}) = \frac{1}{n} \sum_{i=1}^n U(x_i) + \frac{1}{n} \sum_{1 \le i < j \le n} J_{ij} K(x_i - x_j)$$

$$\rightsquigarrow \mathrm{RHS} \le \frac{T^2}{2n} \|K''\|_\infty^2 \mathrm{tr}(J^2)$$

Ex: $J_{ij} = (1/d) 1_{i \sim j}$ in a d-regular graph $\rightsquigarrow \mathrm{tr}(J^2) = n/d$,
so RHS $\to 0$ if $d \to \infty$

Comparing the distributed and original problems

Theorem (L.-Mukherjee-Yeung '22)

Let $G : (\mathbb{R}^d)^n \to \mathbb{R}$ be C^2 convex, $\|\nabla^2 G\|_\infty < \infty$. Then

$$0 \le V_{\mathrm{dstr}} - V \le nT^2 \sum_{1 \le i < j \le n} \|\partial_{ij} G\|_\infty^2 =: RHS.$$

Heterogeneous interactions: U, K convex, K even, $J_{ij} \ge 0$,

$$G(\boldsymbol{x}) = \frac{1}{n}\sum_{i=1}^n U(x_i) + \frac{1}{n}\sum_{1 \le i < j \le n} J_{ij} K(x_i - x_j)$$

$$\rightsquigarrow \mathrm{RHS} \le \frac{T^2}{2n}\|K''\|_\infty^2 \mathrm{tr}(J^2)$$

Interesting point: If J has row sums $= 1$, then $V_{\mathrm{dstr}} = \overline{V} =$ mean field value. $\rightsquigarrow$ Universality of the mean field!

A static reformulation

Relative entropy: $H(\mu \,|\, \nu) = \int \log(d\mu/d\nu)\, d\mu$

Cole-Hopf/Girsanov solution: With $\gamma := N(0, TI)$:

$$V = \inf_{\mu \in \mathcal{P}(\mathbb{R}^n)} \left(\int G\, d\mu + \frac{1}{n} H(\mu \,|\, \gamma) \right) \overset{(*)}{=} -\frac{1}{n} \log \int_{\mathbb{R}^n} e^{-nG}\, d\gamma$$

A static reformulation

Relative entropy: $H(\mu \,|\, \nu) = \int \log(d\mu/d\nu)\, d\mu$

Cole-Hopf/Girsanov solution: With $\gamma := N(0, TI)$:

$$V = \inf_{\mu \in \mathcal{P}(\mathbb{R}^n)} \left(\int G \, d\mu + \frac{1}{n} H(\mu \,|\, \gamma) \right) \overset{(*)}{=} -\frac{1}{n} \log \int_{\mathbb{R}^n} e^{-nG} \, d\gamma$$

$$V_{\mathrm{dstr}} = \inf_{\mu \in \mathcal{P}_{\mathrm{prod}}(\mathbb{R}^n)} \left(\int G \, d\mu + \frac{1}{n} H(\mu \,|\, \gamma) \right)$$

where $\mathcal{P}_{\mathrm{prod}}(\mathbb{R}^n)$ = set of product measures $\mu_1 \otimes \cdots \otimes \mu_n$.

A static reformulation

Relative entropy: $H(\mu \,|\, \nu) = \int \log(d\mu/d\nu)\, d\mu$

Cole-Hopf/Girsanov solution: With $\gamma := N(0, TI)$:

$$V = \inf_{\mu \in \mathcal{P}(\mathbb{R}^n)} \left(\int G\, d\mu + \frac{1}{n} H(\mu \,|\, \gamma) \right) \overset{(*)}{=} -\frac{1}{n} \log \int_{\mathbb{R}^n} e^{-nG}\, d\gamma$$

$$V_{\mathrm{dstr}} = \inf_{\mu \in \mathcal{P}_{\mathrm{prod}}(\mathbb{R}^n)} \left(\int G\, d\mu + \frac{1}{n} H(\mu \,|\, \gamma) \right)$$

where $\mathcal{P}_{\mathrm{prod}}(\mathbb{R}^n)$ = set of product measures $\mu_1 \otimes \cdots \otimes \mu_n$.

Static formulation: Let $P(d\boldsymbol{x}) \propto \exp(-nG(\boldsymbol{x}))\gamma(d\boldsymbol{x})$. Then

$$n(V_{\mathrm{dstr}} - V) = \inf \left\{ H(\mu \,|\, P) : \mu \in \mathcal{P}_{\mathrm{prod}}(\mathbb{R}^n) \right\}.$$

A static reformulation

Relative entropy: $H(\mu \,|\, \nu) = \int \log(d\mu/d\nu)\, d\mu$

Cole-Hopf/Girsanov solution: With $\gamma := N(0, TI)$:

$$V = \inf_{\mu \in \mathcal{P}(\mathbb{R}^n)} \left(\int G\, d\mu + \frac{1}{n} H(\mu \,|\, \gamma) \right) \overset{(*)}{=} -\frac{1}{n} \log \int_{\mathbb{R}^n} e^{-nG}\, d\gamma$$

$$V_{\mathrm{dstr}} = \inf_{\mu \in \mathcal{P}_{\mathrm{prod}}(\mathbb{R}^n)} \left(\int G\, d\mu + \frac{1}{n} H(\mu \,|\, \gamma) \right)$$

where $\mathcal{P}_{\mathrm{prod}}(\mathbb{R}^n)$ = set of product measures $\mu_1 \otimes \cdots \otimes \mu_n$.

Static formulation: Let $P(d\boldsymbol{x}) \propto \exp(-nG(\boldsymbol{x}))\gamma(d\boldsymbol{x})$. Then

$$n(V_{\mathrm{dstr}} - V) = \inf \left\{ H(\mu \,|\, P) : \mu \in \mathcal{P}_{\mathrm{prod}}(\mathbb{R}^n) \right\}.$$

Philosophy: Distributed controls $\Longleftrightarrow$ independent X^i's

$$\mathbb{E}[G(\boldsymbol{X}_T)] \;=\; \int_{\mathbb{R}^n} G(\boldsymbol{x}) \prod_{i=1}^n \mu_i(dx_i), \qquad \mu_i = \mathrm{Law}(X_T^i)$$

A static reformulation

Relative entropy: $H(\mu \,|\, \nu) = \int \log(d\mu/d\nu)\, d\mu$

Static formulation: Let $P(d\boldsymbol{x}) \propto \exp(-nG(\boldsymbol{x}))\gamma(d\boldsymbol{x})$. Then

$$n(V_{\mathrm{dstr}} - V) = \inf\left\{H(\mu \,|\, P) : \mu \in \mathcal{P}_{\mathrm{prod}}(\mathbb{R}^n)\right\}$$

where $\mathcal{P}_{\mathrm{prod}}(\mathbb{R}^n)$ = set of product measures $\mu_1 \otimes \cdots \otimes \mu_n$.

Related literature:

- nonlinear large deviations theory, Chatterjee-Dembo '16,
  also Basak-Mukherjee '17, Eldan '18, Austin '19, Augeri '20...
- mean field variational inference (Wainwright-Jordan '08, Blei et al '17)

A static reformulation

Relative entropy: $H(\mu \,|\, \nu) = \int \log(d\mu/d\nu)\, d\mu$

Static formulation: Let $P(d\boldsymbol{x}) \propto \exp(-nG(\boldsymbol{x}))\gamma(d\boldsymbol{x})$. Then

$$n(V_{\mathrm{dstr}} - V) = \inf\left\{ H(\mu \,|\, P) : \mu \in \mathcal{P}_{\mathrm{prod}}(\mathbb{R}^n) \right\}$$

where $\mathcal{P}_{\mathrm{prod}}(\mathbb{R}^n)$ = set of product measures $\mu_1 \otimes \cdots \otimes \mu_n$.

Related literature:

- nonlinear large deviations theory, Chatterjee-Dembo '16,
  also Basak-Mukherjee '17, Eldan '18, Austin '19, Augeri '20...
- mean field variational inference (Wainwright-Jordan '08, Blei et al '17)

Proof ingredients: first-order condition for μ, Log-Sobolev + Poincaré inequalities for log-concave measures

Toward more general cost functions

Question: Can we generalize beyond the quadratic running cost case, where no static formulation is available?

Toward more general cost functions

Question: Can we generalize beyond the quadratic running cost case, where no static formulation is available?

State process: $\boldsymbol{X} = (X^1, \ldots, X^n)$ as before,

$$dX_t^i = \alpha_i(t, \boldsymbol{X}_t)dt + dW_t^i, \quad \text{valued in } \mathbb{R}^d.$$

Cost functional:

$$J(\boldsymbol{\alpha}) := \mathbb{E}\left[G(\boldsymbol{X}_T) + \int_0^T \left(F(\boldsymbol{X}_t) + \frac{1}{n}\sum_{i=1}^n L^i(X_t^i, \alpha_i(t, \boldsymbol{X}_t)) \right) dt \right].$$

Compare: full-information versus distributed values,

$$V := \inf_{\boldsymbol{\alpha}} J(\boldsymbol{\alpha}), \qquad V_{\mathrm{dstr}} := \inf_{\boldsymbol{\alpha}\ \mathrm{dstr}} J(\boldsymbol{\alpha})$$

Toward more general cost functions

Assumptions:

- F, G, and L^i are convex, L^i uniformly in a
- (F, G, L^i) & Hamiltonian $H^i(x,p) = \sup_a(-a \cdot p - L^i(x,a))$ have bounded 2nd order derivatives

Theorem (Jackson-L. '23)

$$V_{\mathrm{dstr}} - V \leq \; Cn \sum_{1 \leq i < j \leq n} \big(\|\partial_{ij} F\|_\infty^2 + \|\partial_{ij} G\|_\infty^2 \big),$$

where C depends only (and explicitly) on T and spectral bounds of Hessians of (F, G, L^i, H^i).

The mean field case, and convexity

Suppose $L^i = L$ does not depend on i, and

$$F(\boldsymbol{x}) = \mathcal{F}(m^n_{\boldsymbol{x}}), \qquad G(\boldsymbol{x}) = \mathcal{G}(m^n_{\boldsymbol{x}}),$$

where $\mathcal{F}, \mathcal{G} : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ are displacement convex with bounded 2nd order Wasserstein (Lions) derivatives. Then

$$|V - \overline{V}| \le |V - V_{\mathrm{dstr}}| + |V_{\mathrm{dstr}} - \overline{V}| = O(1/n).$$

This is optimal! Though not surprising, was essentially folklore.

(Germain-Pham-Warin '22)

The mean field case, and convexity

Suppose $L^i = L$ does not depend on i, and

$$F(\boldsymbol{x}) = \mathcal{F}(m^n_{\boldsymbol{x}}), \qquad G(\boldsymbol{x}) = \mathcal{G}(m^n_{\boldsymbol{x}}),$$

where $\mathcal{F}, \mathcal{G} : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ are displacement convex with bounded 2nd order Wasserstein (Lions) derivatives. Then

$$|V - \overline{V}| \leq |V - V_{\mathrm{dstr}}| + |V_{\mathrm{dstr}} - \overline{V}| = O(1/n).$$

This is optimal! Though not surprising, was essentially folklore.
(Germain-Pham-Warin '22)

Convexity is crucial! Non-convex case is extremely subtle.
Cardaliaguet-Daudin-Jackson-Souganidis '22, Daudin-Delarue-Jackson '23, Cardaliaguet-Jackson-[Mimikos-Stamatopoulos]-Souganidis '23.

General theory of distributed control

Distributed control is not “classical” control!

But...

General theory of distributed control

Distributed control is not "classical" control!

But... it can be viewed as a sort of mean field control problem, with state variable $(\mathcal{L}(X_t^1), \ldots, \mathcal{L}(X_t^n)) \in (\mathcal{P}(\mathbb{R}^d))^n$.

General theory of distributed control

Distributed control is not "classical" control!

But... it can be viewed as a sort of mean field control problem, with state variable $(\mathcal{L}(X_t^1), \dots, \mathcal{L}(X_t^n)) \in (\mathcal{P}(\mathbb{R}^d))^n$.

Philosophy: Distributed controls $\Longleftrightarrow$ independent X^i's

$$\mathbb{E}[G(\boldsymbol{X}_T)] \;=\; \int_{(\mathbb{R}^d)^n} G(\boldsymbol{x}) \prod_{i=1}^n \mu_T^i(dx_i), \quad \mu_t^i = \mathcal{L}(X_t^i)$$

General theory of distributed control

Simpler case: $L^i(x,a) = |a|^2/2$ and $F \equiv 0$

Distributed value function $\mathcal{V}_d(t, \boldsymbol{m})$ on $[0,T] \times (\mathcal{P}(\mathbb{R}^d))^n$ formally satisfies a PDE:

$$-\partial_t \mathcal{V}_d + \frac{1}{2}\sum_{i=1}^n \int_{\mathbb{R}^d} \Big(n|D_{m^i}\mathcal{V}_d|^2 - \mathrm{Tr}(D_y D_{m^i}\mathcal{V}_d) \Big) m^i(dy) = 0,$$

$$\mathcal{V}_d(T, \boldsymbol{m}) = \int G \, d(m^1 \otimes \cdots \otimes m^n),$$

...with a corresponding verification theorem.

General theory of distributed control

Simpler case: $L^i(x,a) = |a|^2/2$ and $F \equiv 0$

Distributed value function $\mathcal{V}_d(t,\boldsymbol{m})$ on $[0,T] \times (\mathcal{P}(\mathbb{R}^d))^n$ formally satisfies a PDE:

$$-\partial_t \mathcal{V}_d + \frac{1}{2}\sum_{i=1}^n \int_{\mathbb{R}^d} \Big(n|D_{m^i}\mathcal{V}_d|^2 - \mathrm{Tr}(D_y D_{m^i}\mathcal{V}_d) \Big) m^i(dy) = 0,$$

$$\mathcal{V}_d(T,\boldsymbol{m}) = \int G \, d(m^1 \otimes \cdots \otimes m^n),$$

...with a corresponding verification theorem.

Also, a stochastic maximum principle $\rightsquigarrow$ FBSDE characterization of optimality:

$$dX_t^i = -nY_t^i dt + dW_t^i, \qquad X_0^i = x^i,$$
$$dY_t^i = Z_t^i dW_t^i, \qquad Y_T^i = \mathbb{E}[G(\boldsymbol{X}_T) \,|\, X_T^i].$$

Lifting the full-info value function

Full-info (ordinary) value function $V : [0, T] \times (\mathbb{R}^d)^n \to \mathbb{R}$.

Lifting the full-info value function

Full-info (ordinary) value function $V : [0, T] \times (\mathbb{R}^d)^n \to \mathbb{R}$.

Lift: $\mathcal{V} : [0, T] \times (\mathcal{P}(\mathbb{R}^d))^n \to \mathbb{R}$,

$$\mathcal{V}(t, \boldsymbol{m}) = \int_{(\mathbb{R}^d)^n} V(t, \boldsymbol{x}) \prod_{i=1}^{n} m^i(dx^i)$$

Lifting the full-info value function

Full-info (ordinary) value function $V : [0, T] \times (\mathbb{R}^d)^n \to \mathbb{R}$.

Lift: $\mathcal{V} : [0, T] \times (\mathcal{P}(\mathbb{R}^d))^n \to \mathbb{R}$,

$$\mathcal{V}(t, \boldsymbol{m}) = \int_{(\mathbb{R}^d)^n} V(t, \boldsymbol{x}) \prod_{i=1}^{n} m^i(dx^i)$$

...turns out to obey a PDE:

$$-\partial_t \mathcal{V} + \frac{1}{2} \sum_{i=1}^{n} \int_{\mathbb{R}^d} \Big(n|D_{m^i}\mathcal{V}|^2 - \mathrm{Tr}(D_y D_{m^i}\mathcal{V}) \Big) m^i(dy) = -E(t, \boldsymbol{m}),$$

$$\mathcal{V}(T, \boldsymbol{m}) = \int G \, d(m^1 \otimes \cdots \otimes m^n).$$

Same PDE as $\mathcal{V}_d$ except E term! (to be defined)

Comparing the value functions

Comparison principle $\Longrightarrow$

$$0 \le \mathcal{V}_d(t, \boldsymbol{m}) - \mathcal{V}(t, \boldsymbol{m}) \le \int_t^T E(s, \widehat{\boldsymbol{m}}_s)\, ds,$$

Comparing the value functions

Comparison principle $\Longrightarrow$

$$0 \leq \mathcal{V}_d(t, \boldsymbol{m}) - \mathcal{V}(t, \boldsymbol{m}) \leq \int_t^T E(s, \widehat{\boldsymbol{m}}_s)\, ds,$$

where $(\widehat{\boldsymbol{m}}_s)_{s \in [t,T]}$ is TBD

Comparing the value functions

Comparison principle $\Longrightarrow$

$$0 \le \mathcal{V}_d(t, \boldsymbol{m}) - \mathcal{V}(t, \boldsymbol{m}) \le \int_t^T E(s, \widehat{\boldsymbol{m}}_s)\, ds,$$

where $(\widehat{\boldsymbol{m}}_s)_{s \in [t,T]}$ is TBD

Def: For $\boldsymbol{m} = (m^1, \dots, m^n) \in (\mathcal{P}(\mathbb{R}^d))^n$ and $\boldsymbol{\xi} = (\xi^1, \dots, \xi^n)$ with $\xi^i \sim m^i$ independent,

$$E(t, \boldsymbol{m}) := \frac{n}{2} \sum_{i=1}^{n} \mathbb{E}\, \mathrm{Var}(D_i V(t, \boldsymbol{\xi}) \,|\, \xi_i).$$

Comparing the value functions

Comparison principle $\Longrightarrow$

$$0 \le \mathcal{V}_d(t, \boldsymbol{m}) - \mathcal{V}(t, \boldsymbol{m}) \le \int_t^T E(s, \widehat{\boldsymbol{m}}_s)\, ds,$$

where $(\widehat{\boldsymbol{m}}_s)_{s\in[t,T]}$ is TBD

Def: For $\boldsymbol{m} = (m^1, \ldots, m^n) \in (\mathcal{P}(\mathbb{R}^d))^n$ and $\boldsymbol{\xi} = (\xi^1, \ldots, \xi^n)$ with $\xi^i \sim m^i$ independent,

$$E(t, \boldsymbol{m}) := \frac{n}{2} \sum_{i=1}^n \mathbb{E}\, \mathrm{Var}(D_i V(t, \boldsymbol{\xi}) \,|\, \xi_i).$$

Not obvious how to bound it! Uniform (in $\boldsymbol{m}$) bounds don't work: $\|D_i V\|_\infty \lesssim \|D_i G\|_\infty = O(1/n)$ at best (e.g., mean field case), gives only $E = O(1)$.

Comparing the value functions

The right idea: Just bound E along $(\widehat{\boldsymbol{m}}_s)_{s\in[t,T]}$.

- **Key calculation**, where convexity of V is crucial:

$$\frac{d}{ds}E(s,\widehat{\boldsymbol{m}}_s) \geq 0.$$

Comparing the value functions

The right idea: Just bound E along $(\widehat{\boldsymbol{m}}_s)_{s\in[t,T]}$.

- **Key calculation**, where convexity of V is crucial:

$$\frac{d}{ds}E(s,\widehat{\boldsymbol{m}}_s) \ge 0.$$

- Bound by time-T value:

$$E(s,\widehat{\boldsymbol{m}}_s) \le E(T,\widehat{\boldsymbol{m}}_T) = \frac{n}{2}\sum_{i=1}^n \mathbb{E}\,\mathrm{Var}(D_i G(\widehat{\boldsymbol{X}}_T)\,|\,\widehat{X}^i_T).$$

Comparing the value functions

The right idea: Just bound E along $(\widehat{\boldsymbol{m}}_s)_{s\in[t,T]}$.

- **Key calculation**, where convexity of V is crucial:

$$\frac{d}{ds}E(s,\widehat{\boldsymbol{m}}_s) \ge 0.$$

- Bound by time-T value:

$$E(s,\widehat{\boldsymbol{m}}_s) \le E(T,\widehat{\boldsymbol{m}}_T) = \frac{n}{2}\sum_{i=1}^n \mathbb{E}\,\mathrm{Var}(D_iG(\widehat{\boldsymbol{X}}_T)\,|\,\widehat{X}^i_T).$$

- Poincaré inequality:

$$\mathrm{Var}(D_iG(\widehat{\boldsymbol{X}}_T)\,|\,\widehat{X}^i_T) \le (T-t)\sum_{j\neq i}\mathbb{E}\big[|D_{ij}G(\widehat{\boldsymbol{X}}_T)|^2\,|\,\widehat{X}^i_T\big], \quad \forall i.$$

Comparing the value functions

Combined:

$$\begin{aligned}\mathcal{V}_d(t,\boldsymbol{m}) - \mathcal{V}(t,\boldsymbol{m}) &\le \int_t^T E(s,\widehat{\boldsymbol{m}}_s)\,ds \\ &\le (T-t)E(T,\widehat{\boldsymbol{m}}_T) \\ &\le n(T-t)^2 \sum_{1\le i<j\le n} \mathbb{E}|D_{ij}G(\widehat{\boldsymbol{X}}_T)|^2\end{aligned}$$

Note: Omitted constant factor $\propto$ Poincaré constant of initial $\boldsymbol{m}$

Additional results

- Optional controls α^i (full-info) and $\overline{\alpha}^i$ (distributed) are close:

$$\frac{1}{n}\sum_{i=1}^{n}\int_0^T \mathbb{E}|\alpha_t^i - \overline{\alpha}_t^i|^2\,dt \leq \text{ [same bound]}$$

Additional results

- Optimal controls α^i (full-info) and $\overline{\alpha}^i$ (distributed) are close:

$$\frac{1}{n}\sum_{i=1}^{n}\int_0^T \mathbb{E}|\alpha_t^i - \overline{\alpha}_t^i|^2\,dt \leq \text{ [same bound]}$$

Proof idea: Look at associated FBSDEs (X, Y, Z) and $(\overline{X}, \overline{Y}, \overline{Z})$, compute $d(X_t - \overline{X}_t)\cdot(Y_t - \overline{Y}_t)$, and use convexity.

Additional results

- Optimal controls α^i (full-info) and $\overline{\alpha}^i$ (distributed) are close:

$$\frac{1}{n}\sum_{i=1}^{n}\int_0^T \mathbb{E}|\alpha_t^i - \overline{\alpha}_t^i|^2\,dt \le \text{ [same bound]}$$

  Proof idea: Look at associated FBSDEs (X, Y, Z) and $(\overline{X}, \overline{Y}, \overline{Z})$, compute $d(X_t - \overline{X}_t)\cdot(Y_t - \overline{Y}_t)$, and use convexity.

- Most low-dimensional marginals are close:

$$\frac{1}{\binom{n}{k}}\sum_{S\subset[n],|S|=k}\mathcal{W}_2^2(\mathrm{Law}(X^S),\mathrm{Law}(\widehat{X}^S)) \le k\cdot\text{[same bound]}$$

Additional results

- Optimal controls α^i (full-info) and $\overline{\alpha}^i$ (distributed) are close:

$$\frac{1}{n}\sum_{i=1}^{n}\int_0^T \mathbb{E}|\alpha_t^i - \overline{\alpha}_t^i|^2\,dt \le \text{ [same bound]}$$

Proof idea: Look at associated FBSDEs (X, Y, Z) and $(\overline{X}, \overline{Y}, \overline{Z})$, compute $d(X_t - \overline{X}_t) \cdot (Y_t - \overline{Y}_t)$, and use convexity.

- Most low-dimensional marginals are close:

$$\frac{1}{\binom{n}{k}} \sum_{S \subset [n], |S|=k} \mathcal{W}_2^2(\mathrm{Law}(X^S), \mathrm{Law}(\widehat{X}^S)) \le k \cdot \text{[same bound]}$$

Implies (quantitative) concentration of empirical measure,

$$\frac{1}{n}\sum_{i=1}^{n}\delta_{X^i} \approx \frac{1}{n}\sum_{i=1}^{n}\mathrm{Law}(\widehat{X}^i)$$

The independent projection

Recall: $V \leq V_{\mathrm{dstr}}$ trivially, because every distributed control is also a full-info control.

The independent projection

Recall: $V \le V_{\mathrm{dstr}}$ trivially, because every distributed control is also a full-info control.

Key problem: Given a full-info control $\alpha(t, \boldsymbol{x})$, how to construct a "comparable" distributed control?

The independent projection

Recall: $V \leq V_{\mathrm{dstr}}$ trivially, because every distributed control is also a full-info control.

Key problem: Given a full-info control $\alpha(t, \boldsymbol{x})$, how to construct a "comparable" distributed control?

The independent projection will approximate a given state process $\boldsymbol{X} = (X^1, \ldots, X^n)$ by another one $\boldsymbol{Y} = (Y^1, \ldots, Y^n)$ in which components are independent, i.e., control is distributed.

The independent projection

Recall: $V \leq V_{\mathrm{dstr}}$ trivially, because every distributed control is also a full-info control.

Key problem: Given a full-info control $\alpha(t, \boldsymbol{x})$, how to construct a "comparable" distributed control?

The independent projection will approximate a given state process $\boldsymbol{X} = (X^1, \ldots, X^n)$ by another one $\boldsymbol{Y} = (Y^1, \ldots, Y^n)$ in which components are independent, i.e., control is distributed.

In comparison principle: $\widehat{\boldsymbol{m}}_s = (\mathrm{Law}(Y_s^1), \ldots, \mathrm{Law}(Y_s^n))$.

The independent projection

Given process $\boldsymbol{X}_t = (X_t^1, \ldots, X_t^n)$:

$$dX_t^i = \quad \alpha_i(t, \boldsymbol{X}_t) \qquad dt + dW_t^i$$

Independent projection $\boldsymbol{Y}_t = (Y_t^1, \ldots, Y_t^n)$:

$$dY_t^i = \mathbb{E}[\alpha_i(t, \boldsymbol{Y}_t) \mid Y_t^i]\, dt + dW_t^i$$

where $\mathbb{E}[\,\cdot \mid Y_t^i]$ is really integration w.r.t. law of $(Y_t^k)_{k \neq i}$.
(Assume iid initialization, for simplicity.)

The independent projection

Given process $\boldsymbol{X}_t = (X_t^1, \dots, X_t^n)$:

$$dX_t^i = \quad \alpha_i(t, \boldsymbol{X}_t) \qquad dt + dW_t^i$$

Independent projection $\boldsymbol{Y}_t = (Y_t^1, \dots, Y_t^n)$:

$$dY_t^i = \mathbb{E}[\alpha_i(t, \boldsymbol{Y}_t) \mid Y_t^i]\, dt + dW_t^i$$

where $\mathbb{E}[\,\cdot \mid Y_t^i]$ is really integration w.r.t. law of $(Y_t^k)_{k \neq i}$.
(Assume iid initialization, for simplicity.)

Among all ways of approximating $\boldsymbol{X}$ by a process with independent components, this choice $\boldsymbol{Y}$ is natural in a few senses.

The independent projection

$$
\begin{aligned}
dX_t^i &= \quad \alpha_i(t, \boldsymbol{X}_t) \qquad\quad dt + dW_t^i \\
dY_t^i &= \mathbb{E}[\alpha_i(t, \boldsymbol{Y}_t) \mid Y_t^i]\, dt + dW_t^i
\end{aligned}
$$

Example: Mean field interacting particle systems.

$$
\alpha_i(t, \boldsymbol{x}) = b_0(x_i) + \frac{1}{n-1} \sum_{k \neq i} b(x_i, x_k)
$$

The independent projection

$$dX_t^i = \quad \alpha_i(t, \boldsymbol{X}_t) \qquad dt + dW_t^i$$
$$dY_t^i = \mathbb{E}[\alpha_i(t, \boldsymbol{Y}_t) \mid Y_t^i] \, dt + dW_t^i$$

Example: Mean field interacting particle systems.

$$\alpha_i(t, \boldsymbol{x}) = b_0(x_i) + \frac{1}{n-1} \sum_{k \neq i} b(x_i, x_k)$$

$\Longrightarrow Y^1, \ldots, Y^n$ iid copies of McKean-Vlasov SDE,

$$dY_t^i = \left(b_0(Y_t^i) + \int_{\mathbb{R}^d} b(Y_t^i, \cdot) \, d\mu_t \right) dt + dW_t^i, \quad \mu_t = \mathrm{Law}(Y_t^i),$$

the well-known large-n limit in law of X^1

The independent projection

$$\begin{aligned} dX_t^i &= \quad \alpha_i(t, \boldsymbol{X}_t) \qquad dt + dW_t^i \\ dY_t^i &= \mathbb{E}[\alpha_i(t, \boldsymbol{Y}_t) \mid Y_t^i] \, dt + dW_t^i \end{aligned}$$

Example: Non-exchangeable interacting particle systems, interaction matrix $J = (J_{ij})$.

$$\alpha_i(t, \boldsymbol{x}) = b_0(x_i) + \sum_{k \neq i} J_{ik} b(x_i, x_k)$$

The independent projection

$$dX_t^i = \quad \alpha_i(t, \boldsymbol{X}_t) \qquad dt + dW_t^i$$
$$dY_t^i = \mathbb{E}[\alpha_i(t, \boldsymbol{Y}_t) \mid Y_t^i]\, dt + dW_t^i$$

Example: Non-exchangeable interacting particle systems, interaction matrix $J = (J_{ij})$.

$$\alpha_i(t, \boldsymbol{x}) = b_0(x_i) + \sum_{k \neq i} J_{ik}\, b(x_i, x_k)$$

J row sums $= 1 \implies Y^1, \ldots, Y^n$ iid copies of McKean-Vlasov SDE,

$$dY_t^i = \left(b_0(Y_t^i) + \int_{\mathbb{R}^d} b(Y_t^i, \cdot)\, d\mu_t \right) dt + dW_t^i, \quad \mu_t = \mathrm{Law}(Y_t^i)$$

cf. Jabin-Poyato-Soler '21

The independent projection

$$\begin{aligned} dX_t^i &= \quad \alpha_i(t, \boldsymbol{X}_t) \qquad dt + dW_t^i \\ dY_t^i &= \mathbb{E}[\alpha_i(t, \boldsymbol{Y}_t) \mid Y_t^i] \, dt + dW_t^i \end{aligned}$$

Optimality principle #1: $\boldsymbol{Y}$ minimizes the rate of entropy production

$$\frac{d}{dt}\bigg|_{t=0^+} H\big[\mathrm{Law}((\boldsymbol{Y}_s)_{s\le t}) \,\big|\, \mathrm{Law}((\boldsymbol{X}_s)_{s\le t})\big]$$

over all processes with independent components.

The independent projection

Special case: $\alpha_i(t, \boldsymbol{x}) = \partial_i f(\boldsymbol{x})$, smooth $f : \mathbb{R}^n \to \mathbb{R}$.

$$\begin{aligned} dX_t^i &= \quad \partial_i f(\boldsymbol{X}_t) \qquad dt + dW_t^i \\ dY_t^i &= \mathbb{E}[\partial_i f(\boldsymbol{Y}_t) \mid Y_t^i]\, dt + dW_t^i \end{aligned}$$

Optimality principle #2: Gradient flow.

The independent projection

Special case: $\alpha_i(t, \boldsymbol{x}) = \partial_i f(\boldsymbol{x})$, smooth $f : \mathbb{R}^n \to \mathbb{R}$.

$$
\begin{aligned}
dX_t^i &= \quad \partial_i f(\boldsymbol{X}_t) \qquad\quad dt + dW_t^i \\
dY_t^i &= \mathbb{E}[\partial_i f(\boldsymbol{Y}_t) \mid Y_t^i]\, dt + dW_t^i
\end{aligned}
$$

Optimality principle #2: Gradient flow.

- $(\mathrm{Law}(\boldsymbol{X}_t))_{t \geq 0}$ is curve of steepest descent for relative entropy functional $H(\cdot \mid e^{f(\boldsymbol{x})} d\boldsymbol{x})$ in Wasserstein space $\mathcal{P}_2((\mathbb{R}^d)^n)$.

  Jordan-Kinderlehrer-Otto '98

The independent projection

Special case: $\alpha_i(t, \boldsymbol{x}) = \partial_i f(\boldsymbol{x})$, smooth $f : \mathbb{R}^n \to \mathbb{R}$.

$$\begin{aligned} dX_t^i &= \quad \partial_i f(\boldsymbol{X}_t) \qquad\quad dt + dW_t^i \\ dY_t^i &= \mathbb{E}[\partial_i f(\boldsymbol{Y}_t) \mid Y_t^i]\, dt + dW_t^i \end{aligned}$$

Optimality principle #2: Gradient flow.

- $(\mathrm{Law}(\boldsymbol{X}_t))_{t \geq 0}$ is curve of steepest descent for relative entropy functional $H(\cdot \mid e^{f(\boldsymbol{x})} d\boldsymbol{x})$ in Wasserstein space $\mathcal{P}_2((\mathbb{R}^d)^n)$.

  Jordan-Kinderlehrer-Otto '98
- $(\mathrm{Law}(\boldsymbol{Y}_t))_{t \geq 0}$ is curve of steepest descent for same entropy functional but in the submanifold of product measures.