Gradient Flows for Sampling

A Perspective from Invariance

Yifan Chen, Caltech

SOCAMS 2023

The Paper

[Chen, Huang, Huang, Reich, Stuart 2023]

Gradient flows for sampling:
Mean-field models, Gaussian approximations and affine invariance.

By: Yifan Chen, Daniel Zhengyu Huang, Jiaoyang Huang, Sebastian Reich, Andrew M. Stuart. Link: arxiv 2302.11024.

Outline

Outline

Context

The sampling problem

Goal: Draw samples (approximately) from

$$\rho^\star(\theta) \propto \exp(-V(\theta))$$

Context

The sampling problem

Goal: Draw samples (approximately) from

$$\rho^\star(\theta) \propto \exp(-V(\theta))$$

Note: assuming $V(\theta)$ available, in contrast to generative modeling etc.

Context

The sampling problem

Goal: Draw samples (approximately) from

$$\rho^\star(\theta) \propto \exp(-V(\theta))$$

Note: assuming $V(\theta)$ available, in contrast to generative modeling etc.

Applications in

- Bayes inverse problems
- Filtering
- Statistical physics
- ...

Outline

Methodology

Dynamics for sampling

Idea: construct a dynamics of ρ_t that gradually converges to

$$\rho^\star(\theta) \propto \exp(-V(\theta))$$

Note: for simplicity we consider continuous-time

Methodology

Dynamics for sampling

Idea: construct a dynamics of ρ_t that gradually converges to

$$\rho^\star(\theta) \propto \exp(-V(\theta))$$

Note: for simplicity we consider continuous-time

- Finite time dynamics $\rho_1 = \rho^\star$, from a given ρ_0 (e.g., Bayes prior)
  - Sequential Monte Carlo, ...

Methodology

Dynamics for sampling

Idea: construct a dynamics of ρ_t that gradually converges to

$$\rho^\star(\theta) \propto \exp(-V(\theta))$$

Note: for simplicity we consider continuous-time

- Finite time dynamics $\rho_1 = \rho^\star$, from a given ρ_0 (e.g., Bayes prior)
  - Sequential Monte Carlo, ...
- Infinite time dynamics $\rho_\infty = \rho^\star$, from arbitrary ρ_0
  - MCMC, Langevin's dynamics, ...

Methodology

Dynamics for sampling

Idea: construct a dynamics of ρ_t that gradually converges to

$$\rho^\star(\theta) \propto \exp(-V(\theta))$$

Note: for simplicity we consider continuous-time

- Finite time dynamics $\rho_1 = \rho^\star$, from a given ρ_0 (e.g., Bayes prior)
  - Sequential Monte Carlo, ...
- Infinite time dynamics $\rho_\infty = \rho^\star$, from arbitrary ρ_0
  - MCMC, Langevin's dynamics, ...

The focus of this talk: infinite time dynamics

Dynamics through Gradient Flows (GFs)

Gradient flows for sampling

Idea: construct a gradient flow dynamics of ρ_t that converges to

$$\rho^\star(\theta) \propto \exp(-V(\theta))$$

Dynamics through Gradient Flows (GFs)

Gradient flows for sampling

Idea: construct a gradient flow dynamics of ρ_t that converges to

$$\rho^\star(\theta) \propto \exp(-V(\theta))$$

- Langevin's dynamics and Wasserstein GFs
  [Jordan, Kinderlehrer, Otto 1998]
- Stein variantional GD and Stein variational GFs
  [Liu, Wang 2016], [Liu 2017]
- Birth-death dynamics and Wasserstein-Fisher-Rao GFs
  [Lu, Lu, Nolen 2019], [Lu, Slepčev, Wang 2022]
- Interacting Langevin's dynamics and Kalman-Wasserstein GFs
  [Garbuno-Inigo, Hoffmann, Li, Stuart 2020]
- A review paper in Notice of AMS
  [Trillos, Hosseini, Sanz-Alonso 2023]

Gradient Flows

Ingredients in gradient flows

Formally: ($\mathcal{P}$ is the space of probability densities)

- An energy functional $\mathcal{E} : \mathcal{P} \to \mathbb{R}$
- A metric $g_\rho : T_\rho\mathcal{P} \times T_\rho\mathcal{P} \to \mathbb{R}$ with $g_\rho(\sigma_1, \sigma_2) = \langle M(\rho)\sigma_1, \sigma_2\rangle_{L^2}$

$$\Longrightarrow \text{ Flow: } \quad \frac{\partial \rho_t}{\partial t} = -\nabla_g \mathcal{E}(\rho_t) = -M(\rho_t)^{-1}\frac{\delta \mathcal{E}}{\delta \rho}|_{\rho=\rho_t}$$

Concepts and notations:

- $T_\rho\mathcal{P}$ (tangent space) is the space of measures integrated to 0
- $\frac{\delta \mathcal{E}}{\delta \rho}$ is the first variation of $\mathcal{E}$ at ρ

Gradient Flows

Ingredients in gradient flows

Formally: ($\mathcal{P}$ is the space of probability densities)

- An energy functional $\mathcal{E} : \mathcal{P} \to \mathbb{R}$
- A metric $g_\rho : T_\rho\mathcal{P} \times T_\rho\mathcal{P} \to \mathbb{R}$ with $g_\rho(\sigma_1, \sigma_2) = \langle M(\rho)\sigma_1, \sigma_2 \rangle_{L^2}$

$$\Longrightarrow \text{ Flow: } \quad \frac{\partial \rho_t}{\partial t} = -\nabla_g \mathcal{E}(\rho_t) = -M(\rho_t)^{-1} \frac{\delta \mathcal{E}}{\delta \rho}\Big|_{\rho=\rho_t}$$

Concepts and notations:

- $T_\rho\mathcal{P}$ (tangent space) is the space of measures integrated to 0
- $\frac{\delta \mathcal{E}}{\delta \rho}$ is the first variation of $\mathcal{E}$ at ρ

Interpretation as a preconditioned dynamics of the density

$$\frac{\partial \rho_t}{\partial t} = - \underbrace{M(\rho_t)^{-1}}_{\text{preconditioner}} \underbrace{\frac{\delta \mathcal{E}}{\delta \rho}\Big|_{\rho=\rho_t}}_{\text{Euclidean gradient}}$$

The Focus of this Talk

Gradient flow equation

$$\frac{\partial \rho_t}{\partial t} = - \underbrace{M(\rho_t)^{-1}}_{\text{preconditioner}} \underbrace{\frac{\delta \mathcal{E}}{\delta \rho}|_{\rho=\rho_t}}_{\text{Euclideangradient}}$$

The question:

Are there any guiding principles for designing $\mathcal{E}$ and $M(\rho)$?

The Focus of this Talk

Gradient flow equation

$$\frac{\partial \rho_t}{\partial t} = - \underbrace{M(\rho_t)^{-1}}_{\text{preconditioner}} \quad \underbrace{\frac{\delta \mathcal{E}}{\delta \rho}|_{\rho=\rho_t}}_{\text{Euclideangradient}}$$

The question:

Are there any guiding principles for designing $\mathcal{E}$ and $M(\rho)$?

We approach this question through the perspective of invariance

- In energy functionals: invariance to normalization consts
- In metrics: invariance to transformations

Outline

On Choosing the Energy Functionals

Recap: Gradient flow equation

$$\frac{\partial \rho_t}{\partial t} = -M(\rho_t)^{-1}\frac{\delta \mathcal{E}}{\delta \rho}|_{\rho=\rho_t}$$

On Choosing the Energy Functionals

Recap: Gradient flow equation

$$\frac{\partial \rho_t}{\partial t} = -M(\rho_t)^{-1}\frac{\delta \mathcal{E}}{\delta \rho}|_{\rho=\rho_t}$$

- Most popular choice of $\mathcal{E}(\rho)$: Kullback–Leibler divergence

$$\mathcal{E}(\rho) = \mathrm{KL}[\rho \| \rho^\star] = \int \rho \log\Big(\frac{\rho}{\rho^\star}\Big)\,\mathrm{d}\theta$$

On Choosing the Energy Functionals

Recap: Gradient flow equation

$$\frac{\partial \rho_t}{\partial t} = -M(\rho_t)^{-1}\frac{\delta \mathcal{E}}{\delta \rho}|_{\rho=\rho_t}$$

- Most popular choice of $\mathcal{E}(\rho)$: Kullback–Leibler divergence

$$\mathcal{E}(\rho) = \mathrm{KL}[\rho \| \rho^\star] = \int \rho \log\left(\frac{\rho}{\rho^\star}\right) \mathrm{d}\theta$$

- First variation: (we impose $\int \frac{\delta \mathcal{E}}{\delta \rho} \mathrm{d}\theta = 0$)

$$\frac{\delta \mathcal{E}}{\delta \rho} = \log \rho - \log \rho^\star - \int (\log \rho - \log \rho^\star) \mathrm{d}\theta := \mathcal{F}(\rho, \rho^\star)$$

On Choosing the Energy Functionals

Recap: Gradient flow equation

$$\frac{\partial \rho_t}{\partial t} = -M(\rho_t)^{-1} \frac{\delta \mathcal{E}}{\delta \rho}|_{\rho=\rho_t}$$

- Most popular choice of $\mathcal{E}(\rho)$: Kullback–Leibler divergence

$$\mathcal{E}(\rho) = \mathrm{KL}[\rho \| \rho^\star] = \int \rho \log\Big(\frac{\rho}{\rho^\star}\Big)\,\mathrm{d}\theta$$

- First variation: (we impose $\int \frac{\delta \mathcal{E}}{\delta \rho} \mathrm{d}\theta = 0$)

$$\frac{\delta \mathcal{E}}{\delta \rho} = \log \rho - \log \rho^\star - \int (\log \rho - \log \rho^\star)\mathrm{d}\theta := \mathcal{F}(\rho, \rho^\star)$$

- **Invariance**: $\mathcal{F}(\rho, \rho^\star) = \mathcal{F}(\rho, c\rho^\star)$ for any $c \in \mathbb{R}_+$.

On Choosing the Energy Functionals

Recap: Gradient flow equation

$$\frac{\partial \rho_t}{\partial t} = -M(\rho_t)^{-1}\frac{\delta \mathcal{E}}{\delta \rho}|_{\rho=\rho_t}$$

- Most popular choice of $\mathcal{E}(\rho)$: Kullback–Leibler divergence

$$\mathcal{E}(\rho) = \mathrm{KL}[\rho \| \rho^\star] = \int \rho \log\Big(\frac{\rho}{\rho^\star}\Big)\,\mathrm{d}\theta$$

- First variation: (we impose $\int \frac{\delta \mathcal{E}}{\delta \rho}\mathrm{d}\theta = 0$)

$$\frac{\delta \mathcal{E}}{\delta \rho} = \log \rho - \log \rho^\star - \int (\log \rho - \log \rho^\star)\mathrm{d}\theta := \mathcal{F}(\rho, \rho^\star)$$

- **Invariance**: $\mathcal{F}(\rho, \rho^\star) = \mathcal{F}(\rho, c\rho^\star)$ for any $c \in \mathbb{R}_+$.
- **Implication**: no need to worry about normalization consts of $\rho^\star$

The question

Are there any other choices of $\mathcal{E}$ that have such invariance property?

The question

Are there any other choices of $\mathcal{E}$ that have such invariance property?

The answer is NO

Unique Property of the KL Divergence

Theorem [Chen, Huang, Huang, Reich, Stuart 2023]

Among all f-divergence with continuously differentiable f, KL divergence is the only one, up to scaling, whose first variation $\frac{\delta \mathcal{E}}{\delta \rho}$ is invariant to the normalization consts of ρ^*

- f-divergence: for $f(0) = 1$ and f convex

$$D_f[\rho \| \rho^\star] = \int \rho^\star f\Big(\frac{\rho}{\rho^\star}\Big) \mathrm{d}\theta$$

Examples:

- Kullback–Leibler divergence: $f(x) = x \log x$
- χ^2 divergence: $f(x) = (x-1)^2$
- Hellinger distance: $f(x) = (\sqrt{x} - 1)^2$
- ...

Outline

Example: The Fisher-Rao Metric

Recap: gradient flow of KL divergence

First variation: $$\frac{\delta \mathcal{E}}{\delta \rho} = \log \rho - \log \rho^\star - \int (\log \rho - \log \rho^\star) \mathrm{d}\theta$$

Flow: $$\frac{\partial \rho_t}{\partial t} = -M(\rho_t)^{-1} \left(\log \rho - \log \rho^\star - \int (\log \rho - \log \rho^\star) \mathrm{d}\theta \right)$$

Example: The Fisher-Rao Metric

Recap: gradient flow of KL divergence

First variation: $$\frac{\delta \mathcal{E}}{\delta \rho} = \log \rho - \log \rho^\star - \int (\log \rho - \log \rho^\star) \mathrm{d}\theta$$

Flow: $$\frac{\partial \rho_t}{\partial t} = -M(\rho_t)^{-1} \left(\log \rho - \log \rho^\star - \int (\log \rho - \log \rho^\star) \mathrm{d}\theta \right)$$

A renowned metric: Fisher-Rao metric [Rao 1945], [Amari 1985]

Metric: $$M(\rho)^{-1}\psi = \rho(\psi - \mathbb{E}_\rho[\psi]) \in T_\rho \mathcal{P}$$

Flow: $$\frac{\partial \rho_t}{\partial t} = \rho_t \big(\log \rho^\star - \log \rho_t\big) - \rho_t \mathbb{E}_{\rho_t}[\log \rho^\star - \log \rho_t]$$

Invariance to Diffeomorphism

Fisher-Rao gradient flow

$$\frac{\partial \rho_t}{\partial t} = \rho_t\big(\log \rho^\star - \log \rho_t\big) - \rho_t \mathbb{E}_{\rho_t}[\log \rho^\star - \log \rho_t]$$

Invariance to Diffeomorphism

Fisher-Rao gradient flow

$$\frac{\partial \rho_t}{\partial t} = \rho_t\big(\log \rho^\star - \log \rho_t\big) - \rho_t \mathbb{E}_{\rho_t}[\log \rho^\star - \log \rho_t]$$

Apply transformations: given any diffeomorphism $\varphi : \mathbb{R}^{d_\theta} \to \mathbb{R}^{d_\theta}$

- $\tilde{\rho}_t = \varphi \# \rho_t$ is the transformed distribution at time t
- $\tilde{\rho}^\star = \varphi \# \rho^\star$ is the transformed target distribution

Push-forward

$$\tilde{\rho}_t(\theta) = \rho_t(\varphi^{-1}(\theta))|\det \nabla \varphi^{-1}|$$
$$\tilde{\rho}^\star(\theta) = \rho^\star(\varphi^{-1}(\theta))|\det \nabla \varphi^{-1}|$$

Invariance to Diffeomorphism

Fisher-Rao gradient flow

$$\frac{\partial \rho_t}{\partial t} = \rho_t\big(\log \rho^\star - \log \rho_t\big) - \rho_t \mathbb{E}_{\rho_t}[\log \rho^\star - \log \rho_t]$$

Apply transformations: given any diffeomorphism $\varphi : \mathbb{R}^{d_\theta} \to \mathbb{R}^{d_\theta}$

- $\tilde{\rho}_t = \varphi \# \rho_t$ is the transformed distribution at time t
- $\tilde{\rho}^\star = \varphi \# \rho^\star$ is the transformed target distribution

Push-forward

$$\tilde{\rho}_t(\theta) = \rho_t(\varphi^{-1}(\theta))|\det \nabla \varphi^{-1}|$$
$$\tilde{\rho}^\star(\theta) = \rho^\star(\varphi^{-1}(\theta))|\det \nabla \varphi^{-1}|$$

Then, the form of the flow equation remains invariant

$$\frac{\partial \tilde{\rho}_t}{\partial t} = \tilde{\rho}_t\big(\log \tilde{\rho}^\star - \log \tilde{\rho}_t\big) - \tilde{\rho}_t \mathbb{E}_{\tilde{\rho}_t}[\log \tilde{\rho}^\star - \log \tilde{\rho}_t]$$

Invariance seems Useful

Consequence of diffeomorphism invariance

Convergence rate of the flow are the same for
general and Gaussian $\rho^\star$

- For any density $\rho^\star$, there always exists a φ such that

$$\varphi \# \rho^\star = \text{Gaussian}$$

Invariance and Convergence

Convergence of Fisher-Rao gradient flows

[Chen, Huang, Huang, Reich, Stuart 2023], [Lu, Slepčev, Wang 2022]

Let ρ_t solve the Fisher-Rao gradient flow. Assume that there exist constants $K, B > 0$ such that the initial density ρ_0 satisfies

$$e^{-K(1+|\theta|^2)} \leq \frac{\rho_0(\theta)}{\rho^\star(\theta)} \leq e^{K(1+|\theta|^2)}$$

and the second moments of $\rho_0, \rho^\star$ are both bounded by B. Then, for any $t \geq \log\big((1+B)K\big)$,

$$\mathrm{KL}[\rho_t \| \rho^\star] \leq (2 + B + eB)Ke^{-t}.$$

- Unconditional uniform exponential convergence

Invariance and Convergence

Convergence of Fisher-Rao gradient flows

[Chen, Huang, Huang, Reich, Stuart 2023], [Lu, Slepčev, Wang 2022]

Let ρ_t solve the Fisher-Rao gradient flow. Assume that there exist constants $K, B > 0$ such that the initial density ρ_0 satisfies

$$e^{-K(1+|\theta|^2)} \leq \frac{\rho_0(\theta)}{\rho^\star(\theta)} \leq e^{K(1+|\theta|^2)}$$

and the second moments of $\rho_0, \rho^\star$ are both bounded by B. Then, for any $t \geq \log\big((1+B)K\big)$,

$$\mathrm{KL}[\rho_t \| \rho^\star] \leq (2 + B + eB)Ke^{-t}.$$

- Unconditional uniform exponential convergence
- Simulating the flow takes additional efforts
    - Birth-death dynamics [Lu, Lu, Nolen 2019], [Lu, Slepčev, Wang 2022]
    - Gaussian projection [Chen, Huang, Huang, Reich, Stuart 2023]
      Kalman methodology [Huang, Huang, Reich, Stuart 2022]

The question

Any other choices of metric that have such invariance property?

The question

Any other choices of metric that have such invariance property?

The answer is again, NO

Geometric Viewpoint and Uniqueness of Fisher-Rao Metric

Invariance via a geometric viewpoint [Chen, Huang, Huang, Reich, Stuart 2023]

The following two conditions are equivalent:

1. The gradient flow under Riemannian metric g is diffeomorphism-invariant for any $\mathcal{E}$;
2. The Riemannian metric g is diffeomorphism-invariant, namely $\varphi^{\#} g = g$ for any diffeomorphism g.

Geometric Viewpoint and Uniqueness of Fisher-Rao Metric

Invariance via a geometric viewpoint [Chen, Huang, Huang, Reich, Stuart 2023]

The following two conditions are equivalent:

1. The gradient flow under Riemannian metric g is diffeomorphism-invariant for any $\mathcal{E}$;
2. The Riemannian metric g is diffeomorphism-invariant, namely $\varphi^{\#} g = g$ for any diffeomorphism g.

Unique property of Fisher-Rao metric

[Cencov 2000], [Ay, Jost, Lê, Schwachhöfer 2015], [Bauer, Bruveris, Michor 2016]

The Fisher-Rao metric is the only Riemannian metric on smooth positive densities (up to scaling) that is invariant under any diffeomorphism of the parameter space.

Weaker Condition: Affine Invariance

Idea: restrict the diffeomorphism to invertible affine mappings

Weaker Condition: Affine Invariance

Idea: restrict the diffeomorphism to invertible affine mappings

- Affine invariance is useful in optimization and sampling
  - Newton's methods; preconditioning in numerical analysis, ...
  - affine-invariant ensemble sampler [Goodman, Weare 2010]

Weaker Condition: Affine Invariance

Idea: restrict the diffeomorphism to invertible affine mappings

- Affine invariance is useful in optimization and sampling
  - Newton's methods; preconditioning in numerical analysis, ...
  - affine-invariant ensemble sampler [Goodman, Weare 2010]
- Interacting Langevin dynamics [Garbuno-Inigo, Hoffmann, Li, Stuart 2020]

$$\mathrm{d}\theta_t = C(\rho_t)\nabla_\theta \log \rho^\star \mathrm{d}t + \sqrt{2C(\rho_t)}\mathrm{d}W_t$$

Flow equation:

$$\frac{\partial \rho_t}{\partial t} = -\nabla_\theta \cdot (\rho_t C(\rho_t) \nabla_\theta \log \rho^\star) + \nabla \cdot (C(\rho_t)\nabla \rho_t)$$

Gradient flow structure:

Kalman-Wasserstein metric: $M(\rho)^{-1} = -\nabla \cdot (\rho C(\rho) \nabla \cdot)$

Weaker Condition: Affine Invariance

Idea: restrict the diffeomorphism to invertible affine mappings

- Affine invariance is useful in optimization and sampling
  - Newton's methods; preconditioning in numerical analysis, ...
  - affine-invariant ensemble sampler [Goodman, Weare 2010]
- Interacting Langevin dynamics [Garbuno-Inigo, Hoffmann, Li, Stuart 2020]

$$\mathrm{d}\theta_t = C(\rho_t)\nabla_\theta \log \rho^\star \mathrm{d}t + \sqrt{2C(\rho_t)}\mathrm{d}W_t$$

Flow equation:

$$\frac{\partial \rho_t}{\partial t} = -\nabla_\theta \cdot (\rho_t C(\rho_t)\nabla_\theta \log \rho^\star) + \nabla \cdot (C(\rho_t)\nabla \rho_t)$$

Gradient flow structure:

Kalman-Wasserstein metric: $M(\rho)^{-1} = -\nabla \cdot (\rho C(\rho)\nabla \cdot)$

[Chen, Huang, Huang, Reich, Stuart 2023]

Preconditioning recipes to *produce* affine invariant gradient flows

Numerical Examples

- **2D Potential**: $\theta = (\theta^{(1)}, \theta^{(2)})$

$$V(\theta) = \frac{\lambda(\theta^{(2)} - (\theta^{(1)})^2)^2}{20} + \frac{(1-\theta^{(1)})^2}{20} \quad \text{with} \quad \lambda = 0.01,\, 0.1,\, 1$$

  This example is known as the Rosenbrock function

- **Goal**: sample $\rho^\star \sim \exp(-V(\theta))$
- **Method**: Wasserstein GF and its affine invariant modification
- **Configuration**: we initialize the gradient flows from

$$\theta_0 \sim \mathcal{N}\Big(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}4 & 0\\0 & 4\end{bmatrix}\Big)$$

  with 1000 particles. We integrate the dynamics to $t = 15$

A Illustration by Numerical Examples

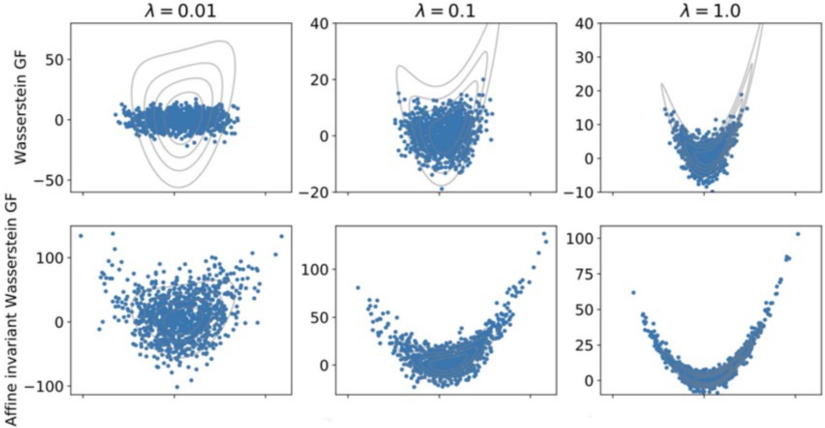

Figure: 1000 particles obtained by different gradient flows at $t = 15$. Grey lines represent the contour of the true posterior.

Outline

Take-away messages

Gradient flows for sampling [Chen, Huang, Huang, Reich, Stuart 2023]

- **Energy functional**: KL divergence
  - invariant to normalization consts
  - unique property up to scaling, among all f divergences
- **Metric**: Fisher-Rao metric
  - invariant to any diffeomorphism of the parameters
  - unique up to scaling among all metrics on probability space
  - unconditional uniform exponential convergence
- **Affine invariance** in the metric
  - unconditional uniform exponential convergence for Gaussian target
  - examples: affine invariant Wasserstein metric and others
- Ongoing work: efficient approximations of Fisher-Rao gradient flows
  - Gaussian projection and variational inference [Chen, Huang, Huang, Reich, Stuart 2023]
  - Kalman methodology [Huang, Huang, Reich, Stuart 2022]

Thanks

https://yifanc96.github.io

Back Up Slides

General Affine Invariance

Affine Invariance in the density level

Consider $\mathcal{E}(\rho) = \mathrm{KL}[\rho \| \rho^\star]$ (general $\mathcal{E}$ in the paper)

- a gradient flow
$$\frac{\partial \rho_t}{\partial t} = -\nabla_g \mathcal{E}(\rho_t)$$
- any affine transformation $\tilde{\theta} = \varphi(\theta) = A\theta + b$

Let

- $\tilde{\rho}_t = \varphi \# \rho_t$ is distribution of $\tilde{\theta}$ at time t
- $\tilde{\mathcal{E}} = \varphi \# \mathcal{E}$ such that $\tilde{\mathcal{E}}(\tilde{\rho}) = \mathcal{E}(\varphi^{-1} \# \tilde{\rho}) = \mathrm{KL}[\tilde{\rho} \| \varphi \# \rho^\star]$

The gradient flow is affine invariant if we have

$$\frac{\partial \tilde{\rho}_t}{\partial t} = -\nabla_g \tilde{\mathcal{E}}(\tilde{\rho}_t)$$

The above holds for affine invariant metrics: $\varphi^\# g = g$

Construct New Affine Invariant Metrics

Stein's metric

$$M(\rho)^{-1}\psi = -\nabla_\theta \cdot \Big(\rho(\theta)\int \kappa(\theta,\theta',\rho)\rho(\theta')\nabla_{\theta'}\psi(\theta')\mathrm{d}\theta'\Big)$$

Flow equation:

$$\frac{\partial \rho_t}{\partial t} = \nabla_\theta \cdot \Big(\rho_t(\theta)\int \kappa(\theta,\theta',\rho_t)\rho_t(\theta')\nabla_{\theta'}\big(\log\rho_t(\theta') - \log\rho^\star(\theta')\big)\mathrm{d}\theta'\Big)$$

Mean field model:

$$\frac{\mathrm{d}\theta_t}{\mathrm{d}t} = \int \kappa(\theta_t,\theta',\rho_t)\rho_t(\theta')\nabla_{\theta'}\log\rho^\star(\theta') + \rho_t(\theta')\nabla_{\theta'}\kappa(\theta_t,\theta',\rho_t)\mathrm{d}\theta'$$

- Affine invariant Stein's metric:
  $M(\rho)^{-1}\psi = -\nabla_\theta \cdot \Big(\rho(\theta)\int \kappa(\theta,\theta',\rho)\rho(\theta')P(\theta,\theta',\rho)\nabla_{\theta'}\psi(\theta')\mathrm{d}\theta'\Big)$
- Sufficient and necessary condition for affine invariance:

$$\kappa(\tilde\theta,\tilde\theta',\tilde\rho)P(\tilde\theta,\tilde\theta',\tilde\rho) = \kappa(\theta,\theta',\rho)AP(\theta,\theta',\rho)A^T$$

  for any $\tilde\theta = \varphi(\theta) = A\theta + b$ and $\tilde\theta' = \varphi(\theta')$
- Example: $P = C(\rho),\ \kappa(\theta,\theta',\rho) \propto \exp\{-\frac{1}{2}(\theta-\theta')^T C(\rho)^{-1}(\theta-\theta')\}$