Covariance-modulated optimal transport and gradient flows

Part II - Gradient flows

BIRS: Stochastic Mass Transport

André Schlichting

Martin Burger, Matthias Erbar, Franca Hoffmann, Daniel Matthes

Applied Mathematics Münster: Institute for Analysis and Numerics

March 21 2022

Outline

- Motivation
  - Bayesian inverse problems
  - sampling via Covariance-modulated SDE
  - mean-field limit
- Gradient flows induced by Covariance-modulated transport
- Geodesic convexity
- EVI formulations and functional inequality
- Entropy method

Inverse problems for parameter estimation

Parameter estimation: Given data $y \in \mathbb{R}^K$ and noise $\xi \in \mathbb{R}^d$

$$\text{Find parameter } x: \qquad y = G(x) + \xi \qquad \text{for given model } G : \mathbb{R}^d \to \mathbb{R}^K.$$

Posterior density: For Gaussian noise $\xi \sim \mathsf{N}(0, \Gamma)$ and $x \sim \mathsf{N}(0, \Sigma)$

$$\pi(\mathrm{d}x) \propto \exp(-f(x)) \qquad \text{with} \qquad f(x) = \frac{1}{2}|y - G(x)|^2_{\Gamma} + \frac{1}{2}|x|^2_{\Sigma}$$

$G(x) = Ax$ → f Gaussian

Inverse problems for parameter estimation

Parameter estimation: Given data $y \in \mathbb{R}^K$ and noise $\xi \in \mathbb{R}^d$

$$\text{Find parameter } x: \qquad y = G(x) + \xi \qquad \text{for given model } G : \mathbb{R}^d \to \mathbb{R}^K.$$

Posterior density: For Gaussian noise $\xi \sim \mathsf{N}(0, \Gamma)$ and $x \sim \mathsf{N}(0, \Sigma)$

$$\pi(\mathrm{d}x) \propto \exp(-f(x)) \qquad \text{with} \qquad f(x) = \frac{1}{2}|y - G(x)|_{\Gamma}^2 + \frac{1}{2}|x|_{\Sigma}^2$$

Bayesian Inverse problem tasks

(1) Inversion: Find $x^* := \arg\max \pi(x)$ or **(2) Sampling** from π.

Inverse problems for parameter estimation

Parameter estimation: Given data $y \in \mathbb{R}^K$ and noise $\xi \in \mathbb{R}^d$

$$\text{Find parameter } x: \qquad y = G(x) + \xi \qquad \text{for given model } G: \mathbb{R}^d \to \mathbb{R}^K.$$

Posterior density: For Gaussian noise $\xi \sim \mathsf{N}(0, \Gamma)$ and $x \sim \mathsf{N}(0, \Sigma)$

$$\pi(\mathrm{d}x) \propto \exp(-f(x)) \qquad \text{with} \qquad f(x) = \frac{1}{2}|y - G(x)|_{\Gamma}^2 + \frac{1}{2}|x|_{\Sigma}^2$$

Bayesian Inverse problem tasks

(1) Inversion: Find $x^* := \arg\max \pi(x)$ or **(2) Sampling** from π.

(1) Ensemble Kalman Inversion (EKI): Particle algorithm finding x^*

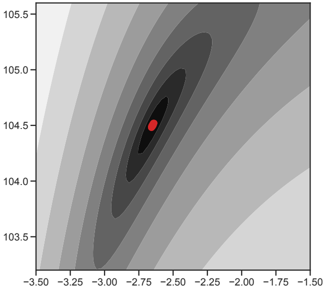
$$\dot{x}^{(j)} = -\frac{1}{J}\sum_{k=1}^{J}\left\langle G(x^{(k)}) - \overline{G}, G(x^{(j)}) - y\right\rangle_{\Gamma} x^{(k)} \qquad \text{with} \qquad \overline{G} := \frac{1}{J}\sum_{k=1}^{J} G(x^{(k)}).$$

[Evensen 1994], [Inglesias, Law, Stuart '13], [Ernst, Sprungk, Starkloff '15], [Schillings, Stuart '17],[Herty, Visconti '19]

Ensemble-Kalman-Sampling (EKS)

Bayesian Inverse problem tasks

(1) Inversion: Find $x^* := \arg\max \pi(x)$ or **(2) Sampling** from π.



Ensemble-Kalman-Sampling (EKS)

Bayesian Inverse problem tasks

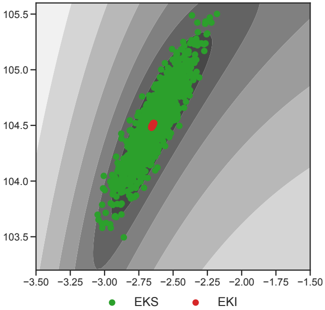
(1) Inversion: Find $x^* := \arg\max \pi(x)$ or **(2) Sampling** from π.

(2) Ensemble Kalman Sampling (EKS): SDE sampling $\pi \propto e^{-f}$ with J particles $\{X^{(j)}\}_{j=1}^J$

$$\dot{X}^{(j)} = -\frac{1}{J}\sum_{k=1}^{J}\left\langle G(X^{(k)}) - \overline{G}, G(X^{(j)}) - y\right\rangle_\Gamma X^{(k)} - \mathrm{C}(\rho^J)\Sigma^{-1}X^{(j)} + \sqrt{2\,\mathrm{C}(\rho^J)}\,\dot{W}^{(j)},$$

with the covariance of the empirical measure $\rho^J = J^{-1}\sum_j \delta_{X^{(j)}}$.

$$\mathrm{C}(\rho^J) = \frac{1}{J}\sum_{k=1}^{J}\left(X^{(k)} - \overline{X}\right)\otimes\left(X^{(k)} - \overline{X}\right) \quad \text{and} \quad \overline{X} = \frac{1}{J}\sum_{k=1}^{J} X^{(k)}.$$



[Reich, Cotter '13], [Garbuno-Iñigo '20], [Li, Hoffmann, Stuart '20]

Ensemble-Kalman-Sampling (EKS)

Bayesian Inverse problem tasks

(1) Inversion: Find $x^* := \arg\max \pi(x)$ or **(2) Sampling** from π.

(2) Ensemble Kalman Sampling (EKS): SDE sampling $\pi \propto e^{-f}$ with J particles $\{X^{(j)}\}_{j=1}^J$

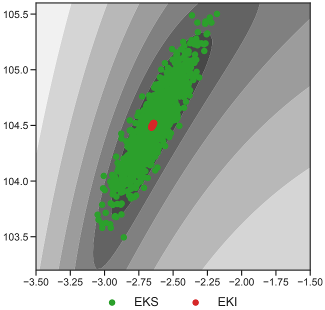
$$\dot{X}^{(j)} = -\frac{1}{J}\sum_{k=1}^{J}\left\langle G(X^{(k)}) - \overline{G}, G(X^{(j)}) - y\right\rangle_\Gamma X^{(k)} - \mathrm{C}(\rho^J)\Sigma^{-1}X^{(j)} + \sqrt{2\,\mathrm{C}(\rho^J)}\,\dot{W}^{(j)},$$

with the covariance of the empirical measure $\rho^J = J^{-1}\sum_j \delta_{X^{(j)}}$.

$$\mathrm{C}(\rho^J) = \frac{1}{J}\sum_{k=1}^{J}\left(X^{(k)} - \overline{X}\right)\otimes\left(X^{(k)} - \overline{X}\right) \quad \text{and} \quad \overline{X} = \frac{1}{J}\sum_{k=1}^{J} X^{(k)}.$$

Mean-field limit of EKS: $J \to \infty$ yields

$$\dot{X} = -\mathrm{C}(\mathrm{law}\,X)\nabla f(X) + \sqrt{2\,\mathrm{C}(\mathrm{law}\,X)}\,\dot{W}.$$



[Reich, Cotter '13], [Garbuno-Iñigo '20], [Li, Hoffmann, Stuart '20], [Ding, Li '21]

Gradient flow for mean-field EKS

Mean-field EKS: $$\dot{X}_t = -\,\mathrm{C}(\rho_t)\nabla f(X_t) + \sqrt{2\,\mathrm{C}(\rho_t)}\,\dot{W}_t$$

Covariance: $$\mathrm{C}(\rho) = \int (x - \mathrm{M}(\rho)) \otimes (x - \mathrm{M}(\rho))\,\mathrm{d}\rho,$$

Mean: $$\mathrm{M}(\rho) = \int x\,\mathrm{d}\rho.$$

Gradient flow for mean-field EKS

Mean-field EKS: $\dot{X}_t = -\,\mathrm{C}(\rho_t)\nabla f(X_t) + \sqrt{2\,\mathrm{C}(\rho_t)}\,\dot{W}_t$

Covariance: $\mathrm{C}(\rho) = \int (x - \mathrm{M}(\rho)) \otimes (x - \mathrm{M}(\rho))\,\mathrm{d}\rho,$ Mean: $\mathrm{M}(\rho) = \int x\,\mathrm{d}\rho.$

Covariance modulated Gradient Flows

$$\partial_t \rho = \nabla \cdot (\rho\,\mathrm{C}(\rho)\nabla \mathcal{F}'(\rho)), \qquad \text{with energy} \qquad \mathcal{F}(\rho) = \int \log\rho\,\mathrm{d}\rho + \int f\,\mathrm{d}\rho.$$

$$\partial_t \rho = D^2 : (C(\rho_t)\rho_t) + \nabla\cdot(\rho_t \nabla f) = \nabla\cdot(\rho\, C(\rho) \nabla(\log\rho + f))$$

Gradient flow for mean-field EKS

Mean-field EKS: $\dot{X}_t = -\mathrm{C}(\rho_t)\nabla f(X_t) + \sqrt{2\,\mathrm{C}(\rho_t)}\,\dot{W}_t$

Covariance: $\mathrm{C}(\rho) = \int (x - \mathrm{M}(\rho)) \otimes (x - \mathrm{M}(\rho))\,\mathrm{d}\rho,$ Mean: $\mathrm{M}(\rho) = \int x\,\mathrm{d}\rho.$

Covariance modulated Gradient Flows

$$\partial_t \rho = \nabla \cdot (\rho\,\mathrm{C}(\rho)\nabla \mathcal{F}'(\rho)), \qquad \text{with energy} \qquad \mathcal{F}(\rho) = \int \log\rho\,\mathrm{d}\rho + \int f\,\mathrm{d}\rho.$$

Energy-dissipation identity: $\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{F}(\rho_t) = -\int \langle \nabla \mathcal{F}'(\rho_t), \mathrm{C}(\rho_t)\nabla\mathcal{F}'(\rho_t)\rangle\,\mathrm{d}\rho_t$

Gradient flow for mean-field EKS

Mean-field EKS: $\dot{X}_t = -\,\mathrm{C}(\rho_t)\nabla f(X_t) + \sqrt{2\,\mathrm{C}(\rho_t)}\,\dot{W}_t$

Covariance: $\mathrm{C}(\rho) = \int (x - \mathrm{M}(\rho)) \otimes (x - \mathrm{M}(\rho))\,\mathrm{d}\rho,$ Mean: $\mathrm{M}(\rho) = \int x\,\mathrm{d}\rho.$

Covariance modulated Gradient Flows

$$\partial_t \rho = \nabla \cdot (\rho\,\mathrm{C}(\rho)\nabla \mathcal{F}'(\rho)), \qquad \text{with energy} \qquad \mathcal{F}(\rho) = \int \log\rho\,\mathrm{d}\rho + \int f\,\mathrm{d}\rho.$$

Energy-dissipation identity: $\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{F}(\rho_t) = -\int \langle \nabla \mathcal{F}'(\rho_t), \mathrm{C}(\rho_t)\nabla \mathcal{F}'(\rho_t)\rangle\,\mathrm{d}\rho_t$

⇒ **Cov-modulated metric tensor:** Formal metric for gradient flow description

$$\mathcal{W}_{\mathrm{cov}}(\rho^0,\rho^1)^2 = \inf\left\{ \int_0^1 \int \langle \nabla\phi_t, \mathrm{C}(\rho_t)\nabla\phi_t\rangle\,\mathrm{d}\rho_t\,\mathrm{d}t : \partial_t\rho_t + \nabla\cdot(\rho_t\,\underbrace{\mathrm{C}(\rho_t)\nabla\phi_t}_{=V}) = 0, \rho_0 = \rho^0, \rho_1 = \rho^1 \right\}$$

Gradient flow for mean-field EKS

WWU MÜNSTER

Mean-field EKS: $\dot{X}_t = -\,\mathrm{C}(\rho_t)\nabla f(X_t) + \sqrt{2\,\mathrm{C}(\rho_t)}\,\dot{W}_t$

Covariance: $\mathrm{C}(\rho) = \int (x - \mathrm{M}(\rho)) \otimes (x - \mathrm{M}(\rho))\,\mathrm{d}\rho,$ Mean: $\mathrm{M}(\rho) = \int x\,\mathrm{d}\rho.$

Covariance modulated Gradient Flows

$\partial_t \rho = \nabla \cdot (\rho\,\mathrm{C}(\rho)\nabla \mathcal{F}'(\rho)),$ with energy $\mathcal{F}(\rho) = \int \log \rho\,\mathrm{d}\rho + \int f\,\mathrm{d}\rho.$

Energy-dissipation identity: $\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{F}(\rho_t) = -\int \langle \nabla \mathcal{F}'(\rho_t), \mathrm{C}(\rho_t)\nabla \mathcal{F}'(\rho_t)\rangle\,\mathrm{d}\rho_t$

$\Rightarrow$ **Cov-modulated metric tensor:** Formal metric for gradient flow description

$$\mathcal{W}_{\mathrm{cov}}(\rho^0, \rho^1)^2 = \inf\left\{ \int_0^1 \int \langle \nabla \phi_t, \mathrm{C}(\rho_t)\nabla \phi_t\rangle\,\mathrm{d}\rho_t\,\mathrm{d}t : \partial_t \rho_t + \nabla \cdot (\rho_t\,\mathrm{C}(\rho_t)\nabla \phi_t) = 0, \rho_0 = \rho^0, \rho_1 = \rho^1 \right\}$$

Question: Can the covariance-modulation improve convergence rates?

Convexity of energy (formal)

Splitting of Cov-OT:

$$\mathcal{W}_{\text{cov}}(\rho^0,\rho^1)^2 = \mathcal{W}_{0,\mathbb{1}}\left(\overline{\rho}^0,\overline{\rho}^1\right)^2 + D(\rho^0,\rho^1)^2.$$

shape moments

Cov-constrained OT:

$$\mathcal{W}_{0,\mathbb{1}}(\rho^0,\rho^1) = \inf_{(\rho,\psi)}\left\{\int_0^1\int|\nabla\psi|^2\,\mathrm{d}\rho_t\,\mathrm{d}t : \substack{\mathrm{M}(\rho_t)=0,\\ \mathrm{C}(\rho_t)=\mathbb{1}}\right\}.$$

Constrained geodesics:

$$\partial_t\rho_t + \nabla\cdot(\rho_t\nabla\psi_t) = 0$$

$$\partial_t\psi_t + \tfrac{1}{2}|\psi_t|^2 + x\cdot\Theta x = 0$$

$$\Theta = \int \nabla\psi\otimes\nabla\psi\,\mathrm{d}\rho$$

$$|\dot\rho_0|^2_{\mathcal{W}_{0,\mathbb{1}}} = \int|\nabla\psi|^2\,\mathrm{d}\rho = \operatorname{tr}\Theta$$

Convexity of energy (formal)

Theorem (Improved constrained convexity)

Let $U : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ be an internal energy satisfying the *McCann*-conditions and set

$$\mathcal{U}(\rho) = \int U(\rho)\,\mathrm{d}x$$

then for any constrained geodesic follows

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\mathcal{U}(\rho_t)\Big|_{t=0} \geq |\dot{\rho}_0|_{\mathcal{W}_{0,\mathbb{1}}} \int P(\rho) \geq 0,$$

with the *pressure* $P(r) = rU'(r) - U(r)$.

$\Rightarrow$ Entropy $U(r) = r \log r$ $(P(r) = r)$ is **1-convex**.

Splitting of Cov-OT:

$$\mathcal{W}_{\mathrm{cov}}(\rho^0, \rho^1)^2 = \mathcal{W}_{0,\mathbb{1}}\left(\bar{\rho}^0, \bar{\rho}^1\right)^2 + D(\rho^0, \rho^1)^2.$$

Cov-constrained OT:

$$\mathcal{W}_{0,\mathbb{1}}(\rho^0, \rho^1) = \inf_{(\rho,\psi)} \left\{ \int_0^1 \int |\nabla\psi|^2\,\mathrm{d}\rho_t\,\mathrm{d}t : \substack{\mathrm{M}(\rho_t)=0,\\ \mathrm{C}(\rho_t)=\mathbb{1}} \right\}.$$

Constrained geodesics:

$$\partial_t \rho_t + \nabla \cdot (\rho_t \nabla \psi_t) = 0$$

$$\partial_t \psi_t + \tfrac{1}{2}|\psi_t|^2 + x \cdot \Theta x = 0$$

$$\Theta = \int \nabla\psi \otimes \nabla\psi\,\mathrm{d}\rho$$

$$|\dot{\rho}_0|^2_{\mathcal{W}_{0,\mathbb{1}}} = \int |\nabla\psi|^2\,\mathrm{d}\rho = \operatorname{tr}\Theta$$

Convexity of energy (formal)

Theorem (Improved constrained convexity)

Let $U : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ be an internal energy satisfying the *McCann*-conditions and set

$$\mathcal{U}(\rho) = \int U(\rho)\,\mathrm{d}x$$

then for any constrained geodesic follows

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\mathcal{U}(\rho_t)\Big|_{t=0} \geq |\dot\rho_0|_{\mathcal{W}_{0,\mathbb{1}}} \int P(\rho) \geq 0,$$

with the *pressure* $P(r) = rU'(r) - U(r)$.

$\Rightarrow$ Entropy $U(r) = r\log r$ $(P(r) = r)$ is **1-convex**.

Quadratic potential energies $\mathcal{V}(\rho) = \frac{1}{2}\int |x|_B^2\,\mathrm{d}\rho$ are constant along constrained geodesics.

Splitting of Cov-OT:

$$\mathcal{W}_{\mathrm{cov}}(\rho^0,\rho^1)^2 = \mathcal{W}_{0,\mathbb{1}}\left(\overline{\rho}^0,\overline{\rho}^1\right)^2 + D(\rho^0,\rho^1)^2.$$

Cov-constrained OT:

$$\mathcal{W}_{0,\mathbb{1}}(\rho^0,\rho^1) = \inf_{(\rho,\psi)}\left\{\int_0^1\int |\nabla\psi|^2\,\mathrm{d}\rho_t\,\mathrm{d}t : \begin{smallmatrix}\mathrm{M}(\rho_t)=0,\\ \mathrm{C}(\rho_t)=\mathbb{1}\end{smallmatrix}\right\}.$$

Constrained geodesics:

$$\begin{aligned}
\partial_t\rho_t + \nabla\cdot(\rho_t\nabla\psi_t) &= 0\\
\partial_t\psi_t + \tfrac{1}{2}|\psi_t|^2 + x\cdot\Theta x &= 0\\
\Theta &= \int \nabla\psi\otimes\nabla\psi\,\mathrm{d}\rho\\
|\dot\rho_0|^2_{\mathcal{W}_{0,\mathbb{1}}} &= \int |\nabla\psi|^2\,\mathrm{d}\rho = \operatorname{tr}\Theta
\end{aligned}$$

Normalization and splitting

Setting: $f(x) = \frac{1}{2}|x - x_0|_B^2$

$$\partial_t \rho_t = \nabla \cdot \big(\mathrm{C}(\rho) \nabla \big(\rho + \rho B^{-1}(x - x_0)\big)\big).$$

Normalization and splitting

Setting: $f(x) = \frac{1}{2}|x - x_0|_B^2$

$$\partial_t \rho_t = \nabla \cdot \big(\mathrm{C}(\rho)\nabla\big(\rho + \rho B^{-1}(x - x_0)\big)\big).$$

Square-root: Given $C_t = \mathrm{C}(\rho_t)$ solve

$$\dot{A}_t = \tfrac{1}{2}\dot{C}_t A_t^{-T} \quad A_0 A_0^T = C_0$$

Non-symmetric!
Independent of choice of $\sqrt{C_0}$

Normalization and splitting

Setting: $f(x) = \frac{1}{2}|x - x_0|_B^2$

$$\partial_t \rho_t = \nabla \cdot \big(\mathrm{C}(\rho)\nabla\big(\rho + \rho B^{-1}(x - x_0)\big)\big).$$

Square-root: Given $C_t = \mathrm{C}(\rho_t)$ solve

$$\dot{A}_t = \tfrac{1}{2}\dot{C}_t A_t^{-T} \quad A_0 A_0^T = C_0$$

Non-symmetric!
Independent of choice of $\sqrt{C_0}$

Normalization map: For $m \in \mathbb{R}^d$, $A \in \mathbb{R}^{d\times d}$

$$T_{m,A}(x) = A^{-1}(x - m).$$

Normalization and splitting

Evolution of moments and shape

Moment equations: $m_t = \mathrm{M}(\rho_t)$, $C_t = \mathrm{C}(\rho_t)$

$$\begin{cases} \dot{m}_t = -C_t B^{-1}(m_t - x_0) \\ \dot{C}_t = 2C_t\big(\mathbb{1} - B^{-1}C_t\big) \end{cases}$$

Normalized shape evolution:

$$\eta_t = (T_{A_t,m_t})_\sharp \rho_t : \qquad \partial_t \eta_t = \Delta \eta_t + \nabla \cdot (\eta_t x)$$

Setting: $f(x) = \frac{1}{2}|x - x_0|_B^2$

$$\partial_t \rho_t = \nabla \cdot \Big(\mathrm{C}(\rho)\nabla\big(\rho + \rho B^{-1}(x - x_0)\big)\Big).$$

Square-root: Given $C_t = \mathrm{C}(\rho_t)$ solve

$$\dot{A}_t = \tfrac{1}{2}\dot{C}_t A_t^{-T} \qquad A_0 A_0^T = C_0$$

Non-symmetric!
Independent of choice of $\sqrt{C_0}$

Normalization map: For $m \in \mathbb{R}^d$, $A \in \mathbb{R}^{d\times d}$

$$T_{m,A}(x) = A^{-1}(x - m).$$

Normalization and splitting

Evolution of moments and shape

Moment equations: $m_t = \mathrm{M}(\rho_t)$, $C_t = \mathrm{C}(\rho_t)$

$$\dot m_t = -C_t B^{-1}(m_t - x_0)$$
$$\dot C_t = 2C_t\big(\mathbb{1} - B^{-1}C_t\big)$$

Normalized shape evolution:

$$\eta_t = (T_{A_t,m_t})_\sharp \rho_t : \qquad \partial_t \eta_t = \Delta \eta_t + \nabla\cdot(\eta_t x)$$

Ornstein-Uhlenbeck evolutions:

$$C_t = \Big(\big(1-e^{-2t}\big)B^{-1} + e^{-2t}C_0^{-1}\Big)^{-1}.$$

$\Rightarrow$ explicit sharp convergence rates possible!

Setting: $f(x) = \frac{1}{2}|x - x_0|_B^2$

$$\partial_t \rho_t = \nabla\cdot\Big(\mathrm{C}(\rho)\nabla\big(\rho + \rho B^{-1}(x - x_0)\big)\Big).$$

Square-root: Given $C_t = \mathrm{C}(\rho_t)$ solve

$$\dot A_t = \tfrac{1}{2}\dot C_t A_t^{-T} \qquad A_0 A_0^T = C_0$$

Non-symmetric!
Independent of choice of $\sqrt{C_0}$

Normalization map: For $m \in \mathbb{R}^d$, $A \in \mathbb{R}^{d\times d}$

$$T_{m,A}(x) = A^{-1}(x - m).$$

EVI and consequences

Relative entropy:

$$\mathcal{E}(\rho|\rho_\infty) = \int \log \frac{\rho}{\rho_\infty} \,\mathrm{d}\rho$$

$$\rho(\mathrm{d}x)_\infty \propto e^{-\frac{1}{2}|x-x_0|_B^2} \,\mathrm{d}x$$

EVI and consequences

Theorem (Shape EVI)

For any $\nu \in \mathcal{P}_{0,\mathbb{1}}$ holds

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{W}_{0,\mathbb{1}}(\eta_t,\nu)^2 + \mathcal{W}_{0,\mathbb{1}}(\eta_t,\nu)^2 \le \mathcal{E}(\nu|\eta_\infty) - \mathcal{E}(\eta_t|\eta_\infty)$$

Relative entropy:

$$\mathcal{E}(\rho|\rho_\infty) = \int \log \frac{\rho}{\rho_\infty}\,\mathrm{d}\rho$$

$$\rho(\mathrm{d}x)_\infty \propto e^{-\frac{1}{2}|x-x_0|_B^2}\,\mathrm{d}x$$

EVI and consequences

Theorem (Shape EVI)

For any $\nu \in \mathcal{P}_{0,\mathbb{1}}$ holds

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{W}_{0,\mathbb{1}}(\eta_t,\nu)^2+\mathcal{W}_{0,\mathbb{1}}(\eta_t,\nu)^2 \leq \mathcal{E}(\nu|\eta_\infty)-\mathcal{E}(\eta_t|\eta_\infty)$$

Relative entropy:

$$\mathcal{E}(\rho|\rho_\infty) = \int \log\frac{\rho}{\rho_\infty}\,\mathrm{d}\rho$$

$$\rho(\mathrm{d}x)_\infty \propto e^{-\frac{1}{2}|x-x_0|_B^2}\,\mathrm{d}x$$

Exponential stability of shape:

$$\mathcal{W}_{0,\mathbb{1}}(\eta_t^1,\eta_t^2) \leq e^{-t}\mathcal{W}_{0,\mathbb{1}}(\eta_0^1,\eta_0^2).$$

Independent of quadratic potential B!

EVI and consequences

Theorem (Shape EVI)

For any $\nu \in \mathcal{P}_{0,\mathbb{1}}$ holds

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{W}_{0,\mathbb{1}}(\eta_t,\nu)^2+\mathcal{W}_{0,\mathbb{1}}(\eta_t,\nu)^2 \leq \mathcal{E}(\nu|\eta_\infty)-\mathcal{E}(\eta_t|\eta_\infty)$$

Corollary (HWI inequality)

For any $\nu \in \mathcal{P}_{0,\mathbb{1}}(\mathbb{R}^d)$

$$\mathcal{E}(\nu|\eta_\infty) \leq \sqrt{\mathcal{I}(\nu|\eta_\infty)}\mathcal{W}_{0,\mathbb{1}}(\nu,\eta_\infty) - \mathcal{W}_{0,\mathbb{1}}(\nu,\eta_\infty).$$

Relative entropy:

$$\mathcal{E}(\rho|\rho_\infty) = \int \log \frac{\rho}{\rho_\infty}\,\mathrm{d}\rho$$

$$\rho(\mathrm{d}x)_\infty \propto e^{-\frac{1}{2}|x-x_0|_B^2}\,\mathrm{d}x$$

Exponential stability of shape:

$$\mathcal{W}_{0,\mathbb{1}}(\eta_t^1,\eta_t^2) \leq e^{-t}\mathcal{W}_{0,\mathbb{1}}(\eta_0^1,\eta_0^2).$$

Independent of quadratic potential B!

EVI and consequences

WWU MÜNSTER

Theorem (Shape EVI)

For any $\nu \in \mathcal{P}_{0,\mathbb{1}}$ holds

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{W}_{0,\mathbb{1}}(\eta_t,\nu)^2 + \mathcal{W}_{0,\mathbb{1}}(\eta_t,\nu)^2 \le \mathcal{E}(\nu|\eta_\infty) - \mathcal{E}(\eta_t|\eta_\infty)$$

Corollary (HWI inequality)

For any $\nu \in \mathcal{P}_{0,\mathbb{1}}(\mathbb{R}^d)$

$$\mathcal{E}(\nu|\eta_\infty) \le \sqrt{\mathcal{I}(\nu|\eta_\infty)}\mathcal{W}_{0,\mathbb{1}}(\nu,\eta_\infty) - \mathcal{W}_{0,\mathbb{1}}(\nu,\eta_\infty).$$

Moment part: Explicitly estimated via ODEs!

Recall: $D((m_0, C_0), (m_1.C_1)) =$

$$\inf\left\{ \int_0^1 \dot m_t \cdot C_t^{-1}\dot m_t + \tfrac{1}{4}\operatorname{tr}\left(\dot C_t C_t^{-1}\dot C_t C_t^{-1}\right) \right\}$$

Relative entropy:

$$\mathcal{E}(\rho|\rho_\infty) = \int \log\frac{\rho}{\rho_\infty}\,\mathrm{d}\rho$$

$$\rho(\mathrm{d}x)_\infty \propto e^{-\frac{1}{2}|x-x_0|_B^2}\,\mathrm{d}x$$

Exponential stability of shape:

$$\mathcal{W}_{0,\mathbb{1}}(\eta_t^1,\eta_t^2) \le e^{-t}\mathcal{W}_{0,\mathbb{1}}(\eta_0^1,\eta_0^2).$$

Independent of quadratic potential B!

EVI and consequences

Theorem (Shape EVI)

For any $\nu \in \mathcal{P}_{0,\mathbb{1}}$ holds

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{W}_{0,\mathbb{1}}(\eta_t,\nu)^2+\mathcal{W}_{0,\mathbb{1}}(\eta_t,\nu)^2 \le \mathcal{E}(\nu|\eta_\infty)-\mathcal{E}(\eta_t|\eta_\infty)$$

Corollary (HWI inequality)

For any $\nu \in \mathcal{P}_{0,\mathbb{1}}(\mathbb{R}^d)$

$$\mathcal{E}(\nu|\eta_\infty) \le \sqrt{\mathcal{I}(\nu|\eta_\infty)}\mathcal{W}_{0,\mathbb{1}}(\nu,\eta_\infty) - \mathcal{W}_{0,\mathbb{1}}(\nu,\eta_\infty).$$

Moment part: Explicitly estimated via ODEs!
Recall: $D((m_0,C_0),(m_1.C_1)) =$

$$\inf\Big\{\int_0^1 \dot{m}_t\cdot C_t^{-1}\dot{m}_t + \tfrac{1}{4}\,\mathrm{tr}\big(\dot{C}_tC_t^{-1}\dot{C}_tC_t^{-1}\big)\Big\}$$

Relative entropy:

$$\mathcal{E}(\rho|\rho_\infty) = \int \log\frac{\rho}{\rho_\infty}\,\mathrm{d}\rho$$

$$\rho(\mathrm{d}x)_\infty \propto e^{-\frac{1}{2}|x-x_0|_B^2}\,\mathrm{d}x$$

Exponential stability of shape:

$$\mathcal{W}_{0,\mathbb{1}}(\eta_t^1,\eta_t^2) \le e^{-t}\mathcal{W}_{0,\mathbb{1}}(\eta_0^1,\eta_0^2).$$

Independent of quadratic potential B!

Intrinsic result! Comparison with W_2 will contain pre-factors [Carrillo-Vaes 2021]

Convergence to equilibrium

Splitting of entropy

$$\mathcal{E}(\rho|\mathsf{N}_{x_0,B}) = \mathcal{E}(\eta|\mathsf{N}_{0,\mathbb{1}}) + \mathcal{E}\big(\mathsf{N}_{\mathrm{M}(\rho),\mathrm{C}(\rho)}|\mathsf{N}_{x_0,B}\big).$$

Convergence to equilibrium

Splitting of entropy

$$\mathcal{E}(\rho|\mathsf{N}_{x_0,B}) = \mathcal{E}(\eta|\mathsf{N}_{0,\mathbb{1}}) + \mathcal{E}\big(\mathsf{N}_{\mathrm{M}(\rho),\mathrm{C}(\rho)}|\mathsf{N}_{x_0,B}\big).$$

Evolution of shape+moments:

$$\begin{cases} \dot{m}_t = -C_t B^{-1}(m_t - x_0) \\ \dot{C}_t = 2C_t\big(\mathbb{1} - B^{-1}C_t\big) \end{cases}$$

$$\partial_t \eta_t = \Delta\eta_t + \nabla\cdot(\eta_t x)$$

Convergence to equilibrium

Splitting of entropy

$$\mathcal{E}(\rho|\mathsf{N}_{x_0,B}) = \mathcal{E}(\eta|\mathsf{N}_{0,\mathbb{1}}) + \mathcal{E}\big(\mathsf{N}_{\mathrm{M}(\rho),\mathrm{C}(\rho)}|\mathsf{N}_{x_0,B}\big).$$

Evolution of shape+moments:

$$\begin{aligned} \dot m_t &= -C_t B^{-1}(m_t - x_0) \\ \dot C_t &= 2C_t\big(\mathbb{1} - B^{-1}C_t\big) \\ \partial_t\eta_t &= \Delta\eta_t + \nabla\cdot(\eta_t x) \end{aligned}$$

Comparison of norms:

$$|x|_B^2 \le \left\| C_t^{\frac12} B^{-1} C_t^{\frac12} \right\|_2 |x|_{C_t}^2$$

$$\left\| C_t^{\frac12} B^{-1} C_t^{\frac12} \right\|_2 \le 1 \vee \left\| B^{\frac12} C_0^{-1} B^{\frac12} \right\|_2$$

Convergence to equilibrium

Splitting of entropy

$$\mathcal{E}(\rho|\mathsf{N}_{x_0,B}) = \mathcal{E}(\eta|\mathsf{N}_{0,\mathbb{1}}) + \mathcal{E}\big(\mathsf{N}_{\mathrm{M}(\rho),\mathrm{C}(\rho)}|\mathsf{N}_{x_0,B}\big).$$

Theorem (Convergence of shape+moments)

$$\mathcal{E}(\eta_t|\mathsf{N}_{0,\mathbb{1}}) \le e^{-2t}\mathcal{E}(\eta_0|\mathsf{N}_{0,\mathbb{1}})$$

$$\mathcal{E}(\mathsf{N}_{m_t,C_t}|\mathsf{N}_{x_0,B}) \le \kappa(B,C_0)e^{-2t}\mathcal{E}(\mathsf{N}_{m_0,C_0}|\mathsf{N}_{x_0,B})$$

with

$$\kappa(B,C_0) = \Big(1 \vee \|B^{\frac{1}{2}}C_0^{-1}B^{\frac{1}{2}}\|_2\Big)\Big(1 \vee \|B^{-\frac{1}{2}}C_0B^{-\frac{1}{2}}\|_2\Big)$$

Evolution of shape+moments:

$$\begin{aligned}\dot{m}_t &= -C_tB^{-1}(m_t - x_0)\\ \dot{C}_t &= 2C_t\big(\mathbb{1} - B^{-1}C_t\big)\\ \partial_t\eta_t &= \Delta\eta_t + \nabla\cdot(\eta_t x)\end{aligned}$$

Comparison of norms:

$$|x|_B^2 \le \Big\|C_t^{\frac{1}{2}}B^{-1}C_t^{\frac{1}{2}}\Big\|_2 |x|_{C_t}^2$$

$$\Big\|C_t^{\frac{1}{2}}B^{-1}C_t^{\frac{1}{2}}\Big\|_2 \le 1 \vee \Big\|B^{\frac{1}{2}}C_0^{-1}B^{\frac{1}{2}}\Big\|_2$$

Convergence to equilibrium

Splitting of entropy

$$\mathcal{E}(\rho|\mathsf{N}_{x_0,B}) = \mathcal{E}(\eta|\mathsf{N}_{0,\mathbb{1}}) + \mathcal{E}\big(\mathsf{N}_{\mathrm{M}(\rho),\mathrm{C}(\rho)}|\mathsf{N}_{x_0,B}\big).$$

Theorem (Convergence of shape+moments)

$$\mathcal{E}(\eta_t|\mathsf{N}_{0,\mathbb{1}}) \le e^{-2t}\mathcal{E}(\eta_0|\mathsf{N}_{0,\mathbb{1}})$$

$$\mathcal{E}(\mathsf{N}_{m_t,C_t}|\mathsf{N}_{x_0,B}) \le \kappa(B,C_0)e^{-2t}\mathcal{E}(\mathsf{N}_{m_0,C_0}|\mathsf{N}_{x_0,B})$$

with

$$\kappa(B,C_0) = \Big(1 \vee \|B^{\frac{1}{2}}C_0^{-1}B^{\frac{1}{2}}\|_2\Big)\Big(1 \vee \|B^{-\frac{1}{2}}C_0B^{-\frac{1}{2}}\|_2\Big)$$

- κ improves if $m_0 = x_0$
- Similar esimates for Fisher information
  $\Rightarrow$ exponential smoothing of gradients

Evolution of shape+moments:

$$\dot{m}_t = -C_tB^{-1}(m_t - x_0)$$
$$\dot{C}_t = 2C_t\big(\mathbb{1} - B^{-1}C_t\big)$$
$$\partial_t\eta_t = \Delta\eta_t + \nabla\cdot(\eta_t x)$$

Comparison of norms:

$$|x|_B^2 \le \Big\|C_t^{\frac{1}{2}}B^{-1}C_t^{\frac{1}{2}}\Big\|_2|x|_{C_t}^2$$

$$\Big\|C_t^{\frac{1}{2}}B^{-1}C_t^{\frac{1}{2}}\Big\|_2 \le 1 \vee \Big\|B^{\frac{1}{2}}C_0^{-1}B^{\frac{1}{2}}\Big\|_2$$

Comparison to var-modulated gradient flows

Let $B \in \mathbb{R}^{d\times d}_{\mathrm{sym},+}$ be fixed and consider energy

$$\mathcal{E}(\rho) = \int \log \rho \,\mathrm{d}\rho + \frac{1}{2}\int \langle x, Bx\rangle \,\mathrm{d}\rho.$$

Variance-modulated GF

$$\partial_t \rho_t = \mathrm{var}(\rho_t)\nabla \cdot (\rho_t \nabla \mathcal{E}'(\rho_t)).$$

Covariance-modulated GF

$$\partial_t \rho_t = \nabla \cdot (\rho_t \,\mathrm{C}(\rho_t)\nabla \mathcal{E}'(\rho_t)).$$

Comparison to var-modulated gradient flows

Let $B \in \mathbb{R}^{d\times d}_{\mathrm{sym},+}$ be fixed and consider energy

$$\mathcal{E}(\rho) = \int \log\rho \,\mathrm{d}\rho + \frac{1}{2}\int \langle x, Bx\rangle \,\mathrm{d}\rho.$$

Variance-modulated GF

$$\partial_t \rho_t = \mathrm{var}(\rho_t)\nabla\cdot(\rho_t \nabla \mathcal{E}'(\rho_t)).$$

$$\mathcal{E}(\rho_t|\rho_\infty) \le e^{-2t\lambda}\mathcal{E}(\rho_t|\rho_\infty),$$

with

$$\lambda = \min\Big\{\frac{d}{\|B\|_2\|B^{-1}\|_2}, \frac{d}{\|B\|_2\|C_0^{-1}\|}\Big\}.$$

Covariance-modulated GF

$$\partial_t \rho_t = \nabla\cdot(\rho_t\,\mathrm{C}(\rho_t)\nabla \mathcal{E}'(\rho_t)).$$

$$\mathcal{E}(\rho_t|\rho_\infty) \le \kappa\; e^{-2t}\mathcal{E}(\rho_t|\rho_\infty),$$

with

$$\kappa = \Big(1\vee \|B^{\frac{1}{2}}C_0^{-1}B^{\frac{1}{2}}\|_2\Big)\Big(1\vee\|B^{-\frac{1}{2}}C_0B^{-\frac{1}{2}}\|_2\Big).$$

Comparison to var-modulated gradient flows

Let $B \in \mathbb{R}^{d\times d}_{\mathrm{sym},+}$ be fixed and consider energy

$$\mathcal{E}(\rho) = \int \log \rho \,\mathrm{d}\rho + \frac{1}{2}\int \langle x, Bx\rangle \,\mathrm{d}\rho.$$

Variance-modulated GF

$$\partial_t \rho_t = \mathrm{var}(\rho_t)\nabla\cdot(\rho_t \nabla \mathcal{E}'(\rho_t)).$$

$$\mathcal{E}(\rho_t|\rho_\infty) \leq e^{-2t\lambda}\mathcal{E}(\rho_t|\rho_\infty),$$

with

$$\lambda = \min\Big\{\frac{d}{\|B\|_2\|B^{-1}\|_2}, \frac{d}{\|B\|_2\|C_0^{-1}\|}\Big\}.$$

Exponential rate depends on EVs of B!

Covariance-modulated GF

$$\partial_t \rho_t = \nabla\cdot(\rho_t\,\mathrm{C}(\rho_t)\nabla \mathcal{E}'(\rho_t)).$$

$$\mathcal{E}(\rho_t|\rho_\infty) \leq \kappa\, e^{-2t}\mathcal{E}(\rho_t|\rho_\infty),$$

with

$$\kappa = \Big(1 \vee \|B^{\frac12}C_0^{-1}B^{\frac12}\|_2\Big)\Big(1\vee \|B^{-\frac12}C_0B^{-\frac12}\|_2\Big).$$

Only prefactors depend on EVs of B!

Summary and open questions

Summary:

- Covariance-weighted transport distance:
  - splitting of the distance in shape and moments
  - local existance of geodesics
  - improved convexity properties
- EKS is gradient flow of Covariance-modulated metric
- Uniform exponential convergence rates for quadratic potentials

Summary and open questions

WWU MÜNSTER

Summary:

- Covariance-weighted transport distance:
  - splitting of the distance in shape and moments
  - local existance of geodesics
  - improved convexity properties
- EKS is gradient flow of Covariance-modulated metric
- Uniform exponential convergence rates for quadratic potentials

Open questions:

- full existence result for geodesics of Covariance-modulated OT?
- Non-quadratic potentials?
- Different modulations (Hessian of ρ)?
- Pathwise formulations?

Summary and open questions

Summary:

- Covariance-weighted transport distance:
  - splitting of the distance in shape and moments
  - local existance of geodesics
  - improved convexity properties
- EKS is gradient flow of Covariance-modulated metric
- Uniform exponential convergence rates for quadratic potentials

Open questions:

- full existence result for geodesics of Covariance-modulated OT?
- Non-quadratic potentials?
- Different modulations (Hessian of ρ)?
- Pathwise formulations?

Thank you for your attention!