Covariance-modulated optimal transport and gradient flows

Matthias Erbar

joint with M. Burger, F. Hoffmann, D. Matthes, A. Schlichting

New Monge Problems and Applications

Paris, September 15, 2023

UNIVERSITÄT BIELEFELD

Fakultät für Mathematik

Outline

- Motivation: Bayesian inverse problems
- Covariance-modulated optimal transport
- Gradient flows induced by covariance-modulated transport
- Trend to equilibrium for modulated gradient flows

Inverse problems for parameter estimation

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

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

Inverse problems for parameter estimation

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

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

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

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

Inverse problems for parameter estimation

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

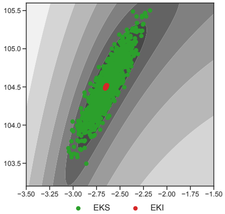
find parameter x : $\quad y = G(x) + \xi \quad$ for given model $G : \mathbb{R}^d \to \mathbb{R}^K$.

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

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

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 $\xi \in \mathbb{R}^d$

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

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

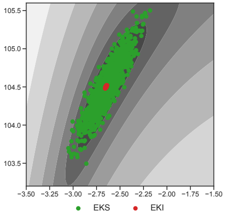
$$\pi(\mathrm{d}x) \propto \exp\left(-f(x)\right) \quad \text{with} \quad f(x) = \frac{1}{2}|y - G(x)|_\Gamma^2 + \frac{1}{2}|x|_\Sigma^2$$

Bayesian Inverse problem tasks

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

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

$$\dot{X}^{(j)} = - \quad \nabla f(X^{(j)}) + \sqrt{2 \quad} \dot{W}^{(j)},$$



Inverse problems for parameter estimation

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

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

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

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

Bayesian Inverse problem tasks

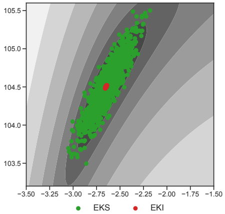
(1) Inversion: Find $x^* := \arg\max \pi(x)$ or **(2) Sampling** from π

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

$$\dot{X}^{(j)} = -\mathsf{C}(\rho^J)\nabla f(X^{(j)}) + \sqrt{2\mathsf{C}(\rho^J)}\,\dot{W}^{(j)},$$

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

$$\mathsf{C}(\rho^J) = \frac{1}{J}\sum_{k=1}^J \left(X^{(k)} - \overline{X}\right) \otimes \left(X^{(k)} - \overline{X}\right), \quad \overline{X} = \frac{1}{J}\sum_{k=1}^J X^{(k)}$$



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

Mean-field EKS and gradient flow

Question: Can the covariance-modulation improve convergence rates?

Mean-field EKS and gradient flow

Question: Can the covariance-modulation improve convergence rates?

Mean-field EKS and gradient flow

Question: Can the covariance-modulation improve convergence rates?

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

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

Mean-field EKS and gradient flow

Question: Can the covariance-modulation improve convergence rates?

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

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

- $\rho_t = \mathrm{law} X_t$ solves the non-linear Fokker-Planck equation

$$\partial_t \rho = \nabla \cdot \big(\mathsf{C}(\rho)(\nabla\rho + \rho\nabla f)\big) = \nabla \cdot \big(\rho\, \mathsf{C}(\rho)\nabla\mathcal{F}'(\rho)\big),$$

with energy $\mathcal{F}(\rho) = \int \log\rho \,\mathrm{d}\rho + \int f \,\mathrm{d}\rho$

covariance: $\mathsf{C}(\rho) = \displaystyle\int (x-\mathrm{M}(\rho))\otimes(x-\mathrm{M}(\rho))\,\mathrm{d}\rho,$ mean: $\mathrm{M}(\rho) = \displaystyle\int x\,\mathrm{d}\rho.$

Gradient flows

- Fokker-Planck equation

$$\partial_t \rho = \Delta \rho + \nabla \cdot (\rho \nabla f) = \nabla \cdot \big(\rho \nabla \mathcal{F}'(\rho)\big)$$

is the gradient flow of $\mathcal{F}$ in the L^2 Kantorovich-Wasserstein distance on $\mathcal{P}(\mathbb{R}^d)$ [JORDAN-KINDERLEHRER-OTTO '98]

$$W_2(\mu_0, \mu_1) = \inf \Big\{ \int_0^1 \int |V_t|^2 \mathrm{d}\mu_t \mathrm{d}t \ : \ \partial_t \mu + \nabla \cdot (\mu_t V_t) = 0 \\ \mu(0) = \mu_0, \mu(1) = \mu_1 \Big\}$$

Gradient flows

- Fokker-Planck equation

$$\partial_t \rho = \Delta\rho + \nabla\cdot(\rho\nabla f) = \nabla\cdot\big(\rho\nabla\mathcal{F}'(\rho)\big)$$

is the gradient flow of $\mathcal{F}$ in the L^2 Kantorovich-Wasserstein distance on $\mathcal{P}(\mathbb{R}^d)$
[JORDAN-KINDERLEHRER-OTTO '98]

$$W_2(\mu_0,\mu_1) = \inf\Big\{ \int_0^1 \int |V_t|^2 \mathrm{d}\mu_t \mathrm{d}t \ : \ \partial_t\mu + \nabla\cdot(\mu_t V_t) = 0$$
$$\mu(0)=\mu_0, \mu(1)=\mu_1 \Big\}$$

- mean field EKS

$$\partial_t \rho = \nabla\cdot\big(\mathsf{C}(\rho)(\nabla\rho + \rho\nabla f)\big) = \nabla\cdot\big(\rho\,\mathsf{C}(\rho)\nabla\mathcal{F}'(\rho)\big)$$

is gradient flow of $\mathcal{F}$ in a modified goemetry on $\mathcal{P}(\mathbb{R}^d)$
[GARBUNO-INIGO, HOFFMANN, LI STUART '20]

$$\mathcal{W}_{\mathsf{cov}}(\mu_0,\mu_1)^2 = \inf\Big\{ \int_0^1 \int \langle V_t, \mathsf{C}(\mu_t)^{-1} V_t\rangle \mathrm{d}\mu_t \mathrm{d}t \ : \ \partial_t\mu + \nabla\cdot(\mu_t V_t) = 0$$
$$\mu(0)=\mu_0, \mu(1)=\mu_1 \Big\}$$

Covariance modulated optimal transport

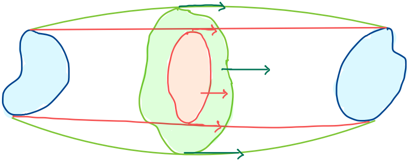
- For $\mu_0, \mu_1 \in \mathcal{P}(\mathbb{R}^d)$ consider dynamic transport problem

$$\mathcal{W}_{\mathsf{cov}}(\mu_0, \mu_1)^2 := \inf_{\mu, V} \int_0^1 \int_{\mathbb{R}^d} \|V_t\|^2_{\mathsf{C}(\mu_t)} \mathrm{d}\mu_t \mathrm{d}t,$$

minimizing over $(\mu_t, V_t)_{t\in[0,1]}$ solving $\partial_t \mu + \nabla \cdot (\mu V) = 0$ with

mean $\mathsf{M}(\mu) := \int x \mathrm{d}\mu(x)$, covariance $\mathsf{C}(\mu) := \int (x - \mathsf{M}(\mu))(x - \mathsf{M}(\mu))^T \mathrm{d}\mu(x)$

$$\|V\|^2_{\mathsf{C}(\mu)} := V \cdot \mathsf{C}(\mu)^{-1} V$$



Covariance modulated optimal transport

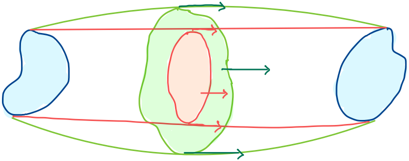
- For $\mu_0, \mu_1 \in \mathcal{P}(\mathbb{R}^d)$ consider dynamic transport problem

$$\mathcal{W}_{\mathsf{cov}}(\mu_0, \mu_1)^2 := \inf_{\mu, V} \int_0^1 \int_{\mathbb{R}^d} \|V_t\|^2_{\mathsf{C}(\mu_t)} \mathrm{d}\mu_t \mathrm{d}t,$$

minimizing over $(\mu_t, V_t)_{t\in[0,1]}$ solving $\partial_t \mu + \nabla \cdot (\mu V) = 0$ with

mean $\mathsf{M}(\mu) := \int x \mathrm{d}\mu(x)$, covariance $\mathsf{C}(\mu) := \int (x - \mathsf{M}(\mu))(x - \mathsf{M}(\mu))^T \mathrm{d}\mu(x)$

$$\|V\|^2_{\mathsf{C}(\mu)} := V \cdot \mathsf{C}(\mu)^{-1} V$$



- **Goals:**
  - structure of optimisers (competition: spreading vs. path length)
  - dynamics and convergence of gradient flows for $\mathcal{W}_{\mathsf{cov}}$

Modulated transport metric

- Recall: with $\|V\|_{\mathsf{C}}^2 = V \cdot \mathsf{C}^{-1} V$

$$\mathcal{W}_{\mathsf{cov}}(\mu_0, \mu_1)^2 := \inf \left\{ \int_0^1 \int_{\mathbb{R}^d} \|V_t\|^2_{\mathsf{C}(\mu_t)} \mathrm{d}\mu_t \mathrm{d}t \ , \quad \partial_t \mu + \nabla \cdot (\mu_t V_t) = 0 \right\}$$

- **Proposition:**
  $\mathcal{W}_{\mathsf{cov}}$ defines a (pseudo-) distance on $\mathcal{P}_2(\mathbb{R}^d)$ metrising weak convergence plus convergence of 2nd moments.

$$\mathcal{W}_{\mathsf{cov}}(\mu_0, \mu_1) < \infty \ \Leftrightarrow \ \Big(\ker \mathsf{C}(\mu_0) = \ker \mathsf{C}(\mu_1) \ \text{ and } \ \mathsf{M}(\mu_1) - \mathsf{M}(\mu_0) \perp \ker \mathsf{C}(\mu_0)\Big)$$

- mass cannot travel at finite cost in directions from $\ker \mathsf{C}(\mu_t)$;

$$\mathrm{supp}\,\mu_t \subset \mathrm{span}\big(\mathrm{supp}\,\mu_0\big) \text{ for all } t$$

- quantitatively:

$$e^{-2\sqrt{k_0 A}} \mathsf{C}(\mu_0) \le \mathsf{C}(\mu_t) \le e^{2\sqrt{k_0 A}} \mathsf{C}(\mu_0)$$

  for any curve with $A = \int_0^1 \int \|V_t\|^2_{\mathsf{C}(\mu_t)} \mathrm{d}\mu_t \mathrm{d}t < \infty$, $k_0 = \mathrm{rank}\, \mathsf{C}(\mu_0)$

Splitting in shape and moments

Theorem: For $\mu_0, \mu_1 \in \mathcal{P}_{2,+}(\mathbb{R}^d)$

$$\mathcal{W}_{\mathsf{cov}}(\mu_0, \mu_1)^2 = \inf_{\mathsf{R} \in SO(d)} \left\{ D_{\mathsf{R}}(\mu_0, \mu_1) + \mathcal{W}_{0,\mathbb{1}}(\bar{\mu}_0, \mathsf{R}_{\#}\bar{\mu}_1)^2 \right\}$$

with $\bar{\mu}_i = \mathsf{C}_i^{-1/2}(\cdot - \mathsf{m}_i)_* \mu_i$ normalisations of μ_i s.t. $\mathsf{M}(\bar{\mu}_i) = 0$, $\mathsf{C}(\bar{\mu}_i) = \mathbb{1}$ and

Splitting in shape and moments

Theorem: For $\mu_0, \mu_1 \in \mathcal{P}_{2,+}(\mathbb{R}^d)$

$$\mathcal{W}_{\mathsf{cov}}(\mu_0,\mu_1)^2 = \inf_{\mathsf{R}\in SO(d)} \left\{ D_{\mathsf{R}}(\mu_0,\mu_1) + \mathcal{W}_{0,\mathbb{1}}(\bar\mu_0, \mathsf{R}_\#\bar\mu_1)^2 \right\}$$

with $\bar\mu_i = \mathsf{C}_i^{-1/2}(\cdot - \mathsf{m}_i)_*\mu_i$ normalisations of μ_i s.t. $\mathsf{M}(\bar\mu_i) = 0$, $\mathsf{C}(\bar\mu_i) = \mathbb{1}$ and

- covariance constraint transport problem

$$\mathcal{W}_{0,\mathbb{1}}(\nu_0,\nu_1)^2 := \inf \Big\{ \int_0^1 \int_{\mathbb{R}^d} \|V_t\|^2 \mathrm{d}\mu_t \mathrm{d}t \ , \quad \partial_t \nu + \nabla \cdot (\mu V) = 0 \ ,$$
$$\mathsf{M}(\nu_t) = 0, \mathsf{C}(\nu_t) = \mathbb{1} \ \forall t \Big\}$$

Splitting in shape and moments

Theorem: For $\mu_0, \mu_1 \in \mathcal{P}_{2,+}(\mathbb{R}^d)$

$$\mathcal{W}_{\mathsf{cov}}(\mu_0, \mu_1)^2 = \inf_{\mathsf{R} \in SO(d)} \left\{ D_{\mathsf{R}}(\mu_0, \mu_1) + \mathcal{W}_{0,\mathbb{1}}(\bar{\mu}_0, \mathsf{R}_{\#}\bar{\mu}_1)^2 \right\}$$

with $\bar{\mu}_i = \mathsf{C}_i^{-1/2}(\cdot - \mathsf{m}_i)_* \mu_i$ normalisations of μ_i s.t. $\mathsf{M}(\bar{\mu}_i) = 0$, $\mathsf{C}(\bar{\mu}_i) = \mathbb{1}$ and

- covariance constraint transport problem

$$\mathcal{W}_{0,\mathbb{1}}(\nu_0, \nu_1)^2 := \inf \left\{ \int_0^1 \int_{\mathbb{R}^d} \|V_t\|^2 \mathrm{d}\mu_t \mathrm{d}t \,, \quad \partial_t \nu + \nabla \cdot (\mu V) = 0 \,, \right.$$
$$\left. \mathsf{M}(\nu_t) = 0, \mathsf{C}(\nu_t) = \mathbb{1} \ \forall t \right\}$$

- constraint moment problem

$$D_{\mathsf{R}}(\mu_0, \mu_1) := \inf_{\mathsf{m},\mathsf{C}} \left\{ \int_0^1 \dot{\mathsf{m}}_t \cdot (\mathsf{A}_t \mathsf{A}_t^T)^{-1} \dot{\mathsf{m}}_t + \mathsf{tr}\left[\dot{\mathsf{A}}_t \mathsf{A}_t^{-1} \dot{\mathsf{A}}_t \mathsf{A}_t^{-1}\right] \mathrm{d}t, \right.$$
$$\left. \mathsf{m}_i = \mathsf{M}(\mu_i), \mathsf{A}_0 = \mathsf{C}(\mu_0)^{\frac{1}{2}}, \mathsf{A}_1 \mathsf{R} = \mathsf{C}(\mu_1)^{\frac{1}{2}}, \mathsf{A}_t^{-1} \dot{\mathsf{A}}_t \text{ symmetric} \right\}$$

Splitting in shape and moments

Theorem: For $\mu_0, \mu_1 \in \mathcal{P}_{2,+}(\mathbb{R}^d)$

$$\mathcal{W}_{\mathsf{cov}}(\mu_0,\mu_1)^2 = \inf_{\mathsf{R}\in SO(d)} \Big\{ D_{\mathsf{R}}(\mu_0,\mu_1) + \mathcal{W}_{0,\mathbb{1}}(\bar\mu_0, \mathsf{R}_{\#}\bar\mu_1)^2 \Big\}$$

with $\bar\mu_i = \mathsf{C}_i^{-1/2}(\cdot - \mathsf{m}_i)_*\mu_i$ normalisations of μ_i s.t. $\mathsf{M}(\bar\mu_i) = 0$, $\mathsf{C}(\bar\mu_i) = \mathbb{1}$ and

- covariance constraint transport problem

$$\mathcal{W}_{0,\mathbb{1}}(\nu_0,\nu_1)^2 := \inf\Big\{ \int_0^1 \int_{\mathbb{R}^d} \|V_t\|^2 \mathrm{d}\mu_t \mathrm{d}t \ , \quad \partial_t \nu + \nabla\cdot(\mu V) = 0 \ , \\ \mathsf{M}(\nu_t) = 0, \mathsf{C}(\nu_t) = \mathbb{1} \ \forall t \Big\}$$

- constraint moment problem

$$D_{\mathsf{R}}(\mu_0,\mu_1) := \inf_{\mathsf{m},\mathsf{C}} \Big\{ \int_0^1 \dot{\mathsf{m}}_t \cdot (\mathsf{A}_t\mathsf{A}_t^T)^{-1}\dot{\mathsf{m}}_t + \mathsf{tr}\big[\dot{\mathsf{A}}_t\mathsf{A}_t^{-1}\dot{\mathsf{A}}_t\mathsf{A}_t^{-1}\big]\mathrm{d}t, \\ \mathsf{m}_i = \mathsf{M}(\mu_i), \mathsf{A}_0 = \mathsf{C}(\mu_0)^{\frac{1}{2}}, \mathsf{A}_1\mathsf{R} = \mathsf{C}(\mu_1)^{\frac{1}{2}}, \mathsf{A}_t^{-1}\dot{\mathsf{A}}_t \text{ symmetric} \Big\}$$

Note: any (non-symmetric) root A of C sat. $\mathsf{AR} = \mathsf{C}^{1/2}$ for suitable $\mathsf{R} \in O(d)$

Proof sketch

- separate minimisation of transport and evolution of moments

$$\mathcal{W}_{\mathsf{cov}}(\mu_0,\mu_1) = \inf\left\{\mathcal{W}_{\mathsf{m},\mathsf{C}}(\mu_0,\mu_1) \;:\; \mathsf{m}:[0,1]\to\mathbb{R}^d\,, \mathsf{C}:[0,1]\to S_+^{d\times d}\right\}$$

$$\mathcal{W}_{\mathsf{m},\mathsf{C}}(\mu_0,\mu_1)^2 := \inf\left\{\int_0^1\int \|V_t\|_{\mathsf{C}_t}^2 \mathrm{d}\mu_t \mathrm{d}t \;:\; \partial_t\mu_t + \nabla\cdot(\mu_t V_t) = 0, \right.$$

$$\left. \mathsf{M}(\mu_t) = \mathsf{m}_t, \mathsf{C}(\mu)_t = \mathsf{C}_t \;\forall t\right\}$$

Proof sketch

- separate minimisation of transport and evolution of moments

$$\mathcal{W}_{\mathsf{cov}}(\mu_0,\mu_1) = \inf\left\{\mathcal{W}_{\mathsf{m},\mathsf{C}}(\mu_0,\mu_1) \ : \ \mathsf{m} : [0,1] \to \mathbb{R}^d \ , \mathsf{C} : [0,1] \to S_+^{d\times d}\right\}$$

$$\mathcal{W}_{\mathsf{m},\mathsf{C}}(\mu_0,\mu_1)^2 := \inf\Big\{ \int_0^1 \int \|V_t\|_{\mathsf{C}_t}^2 \mathrm{d}\mu_t \mathrm{d}t \ : \ \partial_t\mu_t + \nabla\cdot(\mu_t V_t) = 0, \quad \mathsf{M}(\mu_t) = \mathsf{m}_t, \mathsf{C}(\mu)_t = \mathsf{C}_t \ \forall t\Big\}$$

- for (non-symmetric) root A_t of C_t and $\tilde{\mu}_t := \left(A_t^{-1}(\cdot - \mathsf{m}_t)\right)_{\#}\mu_t$ with $\mathsf{M}(\tilde{\mu}_t) = 0$, $\mathsf{C}(\tilde{\mu}_t) = \mathbb{1}$ and suitable $\tilde{V}_t$:

$$\partial_t\tilde{\mu}_t + \nabla\cdot(\tilde{\mu}_t\tilde{V}_t) = 0\ , \quad \int \|V_t\|_{\mathsf{C}_t}^2 \mathrm{d}\mu_t = \int \|\tilde{V}_t\|^2 \mathrm{d}\tilde{\mu}_t + \dot{\mathsf{m}}\cdot(\mathsf{A}\mathsf{A}^T)^{-1}\dot{\mathsf{m}} + \mathrm{tr}\big[\dot{\mathsf{A}}\mathsf{A}^{-1}\dot{\mathsf{A}}\mathsf{A}^{-}$$

Proof sketch

- separate minimisation of transport and evolution of moments

$$\mathcal{W}_{\mathsf{cov}}(\mu_0,\mu_1) = \inf\left\{\mathcal{W}_{\mathsf{m},\mathsf{C}}(\mu_0,\mu_1) \ : \ \mathsf{m} : [0,1]\to\mathbb{R}^d \ , \mathsf{C} : [0,1]\to S_+^{d\times d}\right\}$$

$$\mathcal{W}_{\mathsf{m},\mathsf{C}}(\mu_0,\mu_1)^2 := \inf\left\{\int_0^1\int \|V_t\|^2_{\mathsf{C}_t}\mathrm{d}\mu_t\mathrm{d}t \ : \ \partial_t\mu_t + \nabla\cdot(\mu_t V_t) = 0, \right.$$
$$\left. \mathsf{M}(\mu_t) = \mathsf{m}_t, \mathsf{C}(\mu)_t = \mathsf{C}_t \ \forall t\right\}$$

- for (non-symmetric) root A_t of C_t and $\tilde{\mu}_t := \left(A_t^{-1}(\cdot - \mathsf{m}_t)\right)_{\#}\mu_t$ with $\mathsf{M}(\tilde{\mu}_t) = 0$, $\mathsf{C}(\tilde{\mu}_t) = \mathbb{1}$ and suitable $\tilde{V}_t$:

$$\partial_t\tilde{\mu}_t + \nabla\cdot(\tilde{\mu}_t\tilde{V}_t) = 0 \ , \quad \int\|V_t\|^2_{\mathsf{C}_t}\mathrm{d}\mu_t = \int\|\tilde{V}_t\|^2\mathrm{d}\tilde{\mu}_t + \dot{\mathsf{m}}\cdot(\mathsf{A}\mathsf{A}^T)^{-1}\dot{\mathsf{m}} + \mathrm{tr}\big[\dot{\mathsf{A}}\mathsf{A}^{-1}\dot{\mathsf{A}}\mathsf{A}^{-}$$

- if V is optimal, then exists $\bar{\phi}$ s.t. $\bar{V} = \nabla\bar{\phi}$ provided $A_t^{-1}\dot{A}_t$ is symmetric

- $\tilde{\mu}_1 = \mathsf{R}_{\#}\bar{\mu}_1$ for $\mathsf{R} = \mathsf{A}_1^{-1}\mathsf{C}_1^{1/2}$ and $\bar{\mu}_1 = \mathsf{C}_{\#}^{-1/2}\mu_1$

Competition of shape and moments

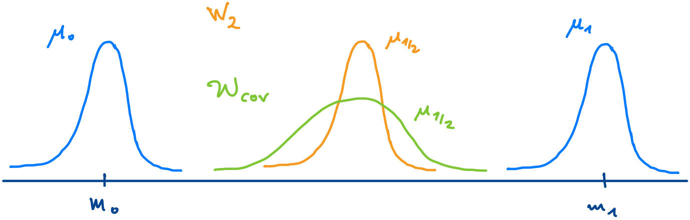
Example: consider Gaussians $\gamma_{m_0,C}, \gamma_{m_1,C}$ with $\delta = m_1 - m_0$

- classical transport: translation by δ

$$W_2\big(\gamma_{m_0,C}, \gamma_{m_1,C}\big) = \|\delta\|$$

- modulated transport: inflation+translation to increase covariance / decrease cost

$$\mathcal{W}_{\text{cov}}\big(\gamma_{m_0,C}, \gamma_{m_1,C}\big) \sim \log\|\delta\| \qquad \text{for } \|\delta\| \gg 1$$



Variance modulated/constraint optimal transport

- variance modulated transport:

$$\mathcal{W}_{\mathsf{var}}(\mu_0,\mu_1)^2 := \inf\left\{ \int_0^1 \frac{1}{\mathrm{var}(\mu_t)} \int_{\mathbb{R}^d} \|V_t\|^2 \mathrm{d}\mu_t \mathrm{d}t \ , \quad \partial_t \mu + \nabla\cdot(\mu V) = 0 \ , \right\}$$

Variance modulated/constraint optimal transport

- variance modulated transport:

$$\mathcal{W}_{\mathsf{var}}(\mu_0,\mu_1)^2 := \inf\left\{ \int_0^1 \frac{1}{\mathrm{var}(\mu_t)} \int_{\mathbb{R}^d} \|V_t\|^2 \mathrm{d}\mu_t \mathrm{d}t\ , \quad \partial_t\mu + \nabla\cdot(\mu V) = 0\ , \right\}$$

- Splitting in shape and moment:

$$\mathcal{W}_{\mathsf{var}}(\mu_0,\mu_1)^2 = \mathcal{W}_{0,1}(\bar\mu_0,\bar\mu_1)^2 + D_{\mathsf{var}}\big(\mu_0,\mu_1\big)^2$$

with

$$\mathcal{W}_{0,1}(\mu_0,\mu_1)^2 := \inf\Big\{ \int_0^1 \int_{\mathbb{R}^d} \|V_t\|^2 \mathrm{d}\mu_t \mathrm{d}t\ , \quad \partial_t\mu + \nabla\cdot(\mu V) = 0\ ,$$
$$\mathsf{M}(\mu_t) = 0, \mathrm{var}(\mu_t) = 1\ \forall t \Big\}$$

$$D_{\mathrm{var}}\big(\mu_0,\mu_1\big)^2 := \inf\Big\{ \int_0^1 \frac{\|\dot{\mathsf{m}}_t\|^2 + |\dot\sigma_t|^2}{\sigma_t^2} \mathrm{d}t,$$
$$\mathsf{m}(i) = \mathsf{M}(\mu_i), \sigma(i)^2 = \mathrm{var}(\mu_i), i = 0,1 \Big\}$$

- moment interpolation problem explicitely solvable

Variance constraint optimal transport

- **Theorem** [Carlen-Gangbo '03]:

$$\mathcal{W}_{\mathsf{m},\sigma}(\mu_0, \mu_1) = 2\sigma \arcsin\left(\frac{W_2(\mu_0, \mu_1)}{2\sigma}\right)$$

Variance constraint optimal transport

- **Theorem** [Carlen-Gangbo '03]:

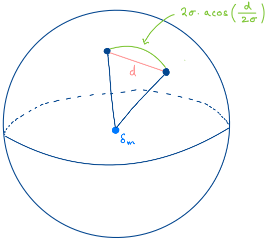
$$\mathcal{W}_{\mathsf{m},\sigma}(\mu_0, \mu_1) = 2\sigma \arcsin\left(\frac{W_2(\mu_0, \mu_1)}{2\sigma}\right)$$

- $\mathcal{W}_{\mathsf{m},\sigma}$ distance induced by W_2 on

$$\mathcal{E}_{\mathsf{m},\sigma} := \left\{\mu \in \mathcal{P}_2(\mathbb{R}^d) \ : \ \mathsf{M}(\mu) = \mathsf{m} \, , \mathrm{var}(\mu) = \sigma^2\right\}$$

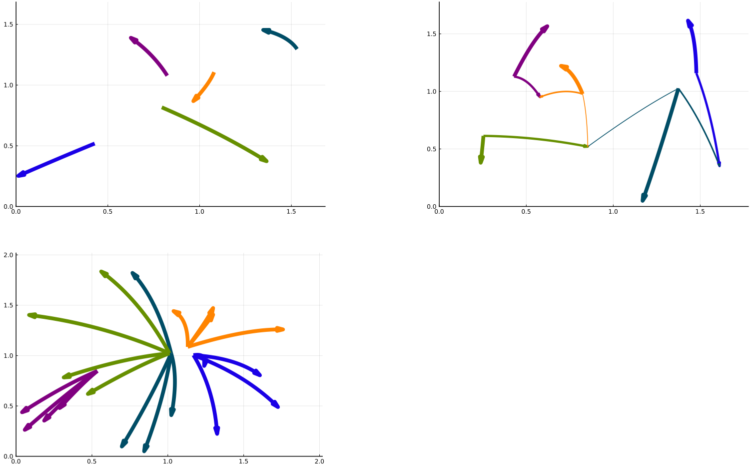
'sphere' of radius σ in $\left(\mathcal{P}_2(\mathbb{R}^d), W_2\right)$ centered at δ_{m}

Geodesics are obtained by translation/scaling of Wasserstein geodesics



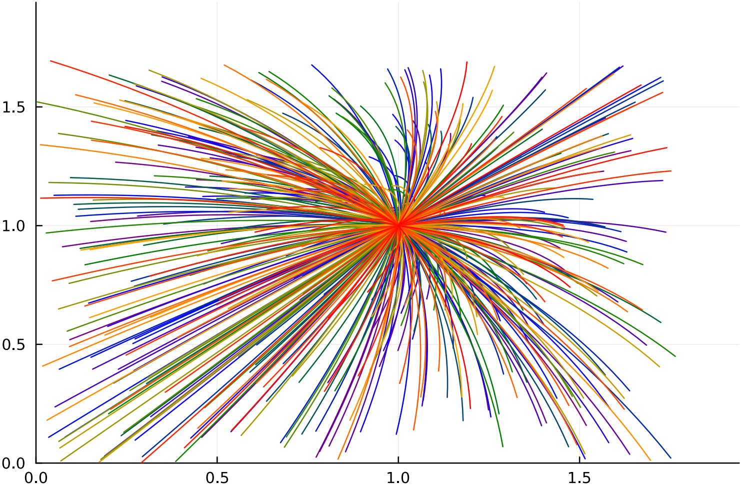
Structure of shape optimisers for $\mathcal{W}_{\text{cov}}$

- normalisation of Wasserstein geodesic not optimal for covariance modulation!
- some numerical results



Structure of shape optimisers for $\mathcal{W}_{\text{cov}}$

- normalisation of Wasserstein geodesic not optimal for covariance modulation!
- some numerical results



Structure of shape optimisers for $\mathcal{W}_{\text{cov}}$

- dual transport problem:

$$\frac{1}{2}\mathcal{W}_{0,\mathbb{1}}(\mu_0,\mu_1)^2 = \sup_{\phi,\Theta}\Big\{ \int \phi_1 \mathrm{d}\mu_1 - \int \phi_0 \mathrm{d}\mu_0 - \int_0^T \mathrm{tr}[\Theta_t]\mathrm{d}t \ : \ \partial_t\phi(x) + \frac{1}{2}|\nabla\phi_t|^2(x) + x\cdot\Theta_t x \le 0 \Big\}$$

Structure of shape optimisers for $\mathcal{W}_{\text{cov}}$

- dual transport problem:

$$\frac{1}{2}\mathcal{W}_{0,\mathbb{1}}(\mu_0,\mu_1)^2 = \sup_{\phi,\Theta}\Big\{\int \phi_1 \mathrm{d}\mu_1 - \int \phi_0 \mathrm{d}\mu_0 - \int_0^T \mathrm{tr}[\Theta_t]\mathrm{d}t \;:\; \partial_t\phi(x) + \frac{1}{2}|\nabla\phi_t|^2(x) + x\cdot\Theta_t x \le 0\Big\}$$

- Lagrangian formulation:

$$\mathcal{W}_{0,\mathbb{1}}(\mu_0,\mu_1)^2 = \inf_{\Pi}\Big\{\mathbb{E}_\Pi\Big[\int_0^1 |\dot\gamma_t|^2\mathrm{d}t\Big] : \gamma_{0,1}\sim\mu_{0,1}, \mathbb{E}_\Pi[\gamma_t]=0, \mathbb{E}_\Pi[\gamma_t\gamma_t^T]=\mathbb{1}\Big\}$$

optimal Π concentrated on solutions to

$$\ddot\gamma_t + \Lambda_t\gamma_t = 0$$

with Λ_t Langrange multiplier for global covariance constraint

$\Rightarrow$ interaction of particle trajectories

Existence of covariance modulated/constraint geodesics

Theorem:

- Every $\mu_0, \mu_1 \in \mathcal{P}_{0,\mathbb{1}}(\mathbb{R}^d)$ with $\mathcal{W}_{0,\mathbb{1}}(\mu_0, \mu_1)^2 < \frac{1}{8}$ are connected by a $\mathcal{W}_{0,\mathbb{1}}$ geodesic.
- Every $\mu_0, \mu_1 \in \mathcal{P}_{2,+}(\mathbb{R}^d)$ with $\mathcal{W}_{\mathsf{cov}}(\mu_0, \mu_1)^2 < \frac{1}{8} + D(\mu_0, \mu_1)^2$ are connected by a $\mathcal{W}_{\mathsf{cov}}$ geodesic.

Problem: minimising sequences might lose constraint in the limit

Existence of covariance modulated/constraint geodesics

Theorem:

- Every $\mu_0, \mu_1 \in \mathcal{P}_{0,\mathbb{1}}(\mathbb{R}^d)$ with $\mathcal{W}_{0,\mathbb{1}}(\mu_0, \mu_1)^2 < \frac{1}{8}$ are connected by a $\mathcal{W}_{0,\mathbb{1}}$ geodesic.
- Every $\mu_0, \mu_1 \in \mathcal{P}_{2,+}(\mathbb{R}^d)$ with $\mathcal{W}_{\text{cov}}(\mu_0, \mu_1)^2 < \frac{1}{8} + D(\mu_0, \mu_1)^2$ are connected by a $\mathcal{W}_{\text{cov}}$ geodesic.

Problem: minimising sequences might lose constraint in the limit

Theorem:

Let $\mu_0, \mu_1 \in \mathcal{P}_{0,\mathbb{1}}(\mathbb{R}^d)$ be symmetric in all d directions. Then there exists a $\mathcal{W}_{0,\mathbb{1}}$ geodesic connecting them.

There are a coupling π of μ_0, μ_1 and parameters $\omega_1, \ldots, \omega_d$ s.t. the optimal path measure is given by superposition of the curves

$$\gamma_k(s) = \frac{\sin \omega_k (1-s)}{\sin \omega_k} x_k + \frac{\sin \omega_k s}{\sin \omega_k} y_k$$

weighted by $\pi(\mathrm{d}x, \mathrm{d}y)$.

Convexity of energy (formal)

- **Theorem:** ([McCann '97])

  Let $U : [0,\infty) \to [0,\infty)$ convex s.t. $\lambda \mapsto \lambda^d U(\lambda^{-d})$ convex non-inc.
  Then the internal energy

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

  is convex along W_2 geodesics, i.e. $\frac{\mathrm{d}^2}{\mathrm{d}t^2}\mathcal{U}(\rho_t)\Big|_{t=0} \geq 0$.

Convexity of energy (formal)

- **Theorem: ([McCann '97])**

  Let $U : [0,\infty) \to [0,\infty)$ convex s.t. $\lambda \mapsto \lambda^d U(\lambda^{-d})$ convex non-inc.
  Then the internal energy

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

  is convex along W_2 geodesics, i.e. $\frac{\mathrm{d}^2}{\mathrm{d}t^2}\mathcal{U}(\rho_t)\Big|_{t=0} \geq 0$.

- **Theorem: (constraint improves convexity)**

  Along any $W_{0,\mathbb{1}}$-geodesic we have

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

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

  In particular, entropy $U(r) = r\log r$ ($P(r) = r$) is 1-convex.

Convexity of energy (formal)

- **Theorem: ([McCann '97])**

  Let $U : [0,\infty) \to [0,\infty)$ convex s.t. $\lambda \mapsto \lambda^d U(\lambda^{-d})$ convex non-inc.
  Then the internal energy
  $$\mathcal{U}(\rho) = \int U(\rho)\,\mathrm{d}x$$
  is convex along W_2 geodesics, i.e. $\frac{\mathrm{d}^2}{\mathrm{d}t^2}\mathcal{U}(\rho_t)\Big|_{t=0} \geq 0.$

- **Theorem: (constraint improves convexity)**

  Along any $W_{0,\mathbb{1}}$-geodesic we have
  $$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\mathcal{U}(\rho_t)\Big|_{t=0} \geq |\dot\rho_0|_{\mathcal{W}_{0,\mathbf{1}}} \int P(\rho) \geq 0,$$
  with the *pressure* $P(r) = rU'(r) - U(r)$.

  In particular, entropy $U(r) = r\log r$ ($P(r) = r$) is 1-convex.

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

Splitting of the gradient flow

- EKS gradient flow for posterior $\pi(\mathrm{d}x) \propto \exp\left(-f(x)\right)$ with $f(x) = \frac{1}{2}|x - x_0|_B^2$

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

Splitting of the gradient flow

- EKS gradient flow for posterior $\pi(\mathrm{d}x) \propto \exp\left(-f(x)\right)$ with $f(x) = \frac{1}{2}|x - x_0|_B^2$

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

- evolution of moments $\mathsf{m}_t = \mathsf{M}(\rho_t)$, $\mathsf{C}_t = \mathsf{C}(\rho_t)$:

$$\dot{\mathsf{m}}_t = -\mathsf{C}_t B^{-1}(\mathsf{m}_t - x_0)\ , \quad \dot{\mathsf{C}}_t = 2\mathsf{C}_t\left(\mathbf{1} - B^{-1}\mathsf{C}_t\right)$$

$$\Rightarrow \mathsf{C}_t = \left(\left(1 - e^{-2t}\right)B^{-1} + e^{-2t}\mathsf{C}_0^{-1}\right)^{-1}$$

Splitting of the gradient flow

- EKS gradient flow for posterior $\pi(\mathrm{d}x) \propto \exp\big(-f(x)\big)$ with $f(x) = \frac{1}{2}|x - x_0|_B^2$

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

- evolution of moments $\mathsf{m}_t = \mathsf{M}(\rho_t)$, $\mathsf{C}_t = \mathsf{C}(\rho_t)$:

$$\dot{\mathsf{m}}_t = -\mathsf{C}_t B^{-1}(\mathsf{m}_t - x_0)\ , \quad \dot{\mathsf{C}}_t = 2\mathsf{C}_t\big(\mathbf{1} - B^{-1}\mathsf{C}_t\big)$$
$$\Rightarrow \mathsf{C}_t = \Big(\big(1 - e^{-2t}\big)B^{-1} + e^{-2t}\mathsf{C}_0^{-1}\Big)^{-1}$$

- evolution of normalized shape: solve

$$\dot{\mathsf{A}}_t = \tfrac{1}{2}\dot{\mathsf{C}}_t\mathsf{A}_t^{-T} \quad \mathsf{A}_0\mathsf{A}_0^T = \mathsf{C}_0 \quad (\text{then } A_t^{-1}\dot{A}_t \text{ symmetric, } \mathsf{A}_t\mathsf{A}_t^T = \mathsf{C}_t)$$

then $\eta_t = \mathsf{A}_t^{-1}(\cdot - \mathsf{m}_t)_*\rho_t$ solves

$$\partial_t \eta = \Delta\eta - \nabla\cdot(\eta x)$$

EVI and consequences

- **relative entropy:**

$$\mathcal{E}(\eta|\mathsf{N}_{0,\mathbf{1}}) = \int \log \eta \mathrm{d}\eta + \frac{1}{2}\int |x|^2 \mathrm{d}\eta \ ,$$

- **Theorem: (shape EVI)** For any $\nu \in \mathcal{P}_{0,\mathbb{1}}$

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

EVI and consequences

- **relative entropy:**

$$\mathcal{E}(\eta|\mathsf{N}_{0,\mathbf{1}}) = \int \log \eta \mathrm{d}\eta + \frac{1}{2}\int |x|^2 \mathrm{d}\eta \ ,$$

- **Theorem: (shape EVI)** For any $\nu \in \mathcal{P}_{0,\mathbb{1}}$

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

- **Corollary: (exponential stability of shape)**

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

Independent of quadratic potential B!

- **Wasserstein stability**

$$W_2(\rho_t, N_{x_0,B}) \leq e^{-t}\kappa(B,C_0)\Big(\inf_{R\in SO(d)} W_2\big(R_\sharp\bar\rho_0, N_{0,\mathbb{1}}\big)^2 + |m_0 - x_0|^2_{C_0} + \Big\|\Big(B^{\frac12}C_0^{-1}B^{\frac12}\Big)^{\frac12} - \mathbb{1}\Big\|^2_{\mathsf{HS}}\Big)^{\frac12}$$

where $\kappa(B,C_0) := \|B\|_2 \max\Big\{1, \Big\|B^{-\frac12}C_0B^{-\frac12}\Big\|_2\Big\}$

Note: compare with [Carrillo-Vaes '21]

Convergence to equilibrium

- **splitting of entropy**

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

- **Theorem: (convergence of shape+moments)**

$$\mathcal{E}(\eta_t|\mathsf{N}_{0,\mathbf{1}}) \le e^{-2t}\mathcal{E}(\eta_0|\mathsf{N}_{0,\mathbf{1}})$$
$$\mathcal{E}(\mathsf{N}_{m_t,C_t}|\mathsf{N}_{x_0,B}) \le \kappa e^{-2t}\mathcal{E}(\mathsf{N}_{m_0,C_0}|\mathsf{N}_{x_0,B})$$

with

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

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

Comparison variance vs covariance modulation

Let $B \in \mathbb{R}^{d\times d}_{\text{sym},+}$ 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

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

Note: Exponential rate depends on EVs of B!

covariance-modulated GF

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

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

with

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

Note: Only prefactors depend on EVs of B!

Summary and open questions

Summary:

- covariance-modulated transport distance:
  - splitting in shape and moments
  - improved convexity properties
- EKS is gradient flow in covariance-modulated geometry
- uniform exponential convergence rates for quadratic potentials

Open questions:

- unconditional existence of geodesics for covariance-modulated OT?
- explicit structure of optimal curves?
- non-quadratic potentials?
- different modulations (Hessian of ρ)?

Thank you for your attention!