

## LECTURES 2 & 3

### Covariance - Modulated Optimal Transport

- AGENDA:
1. Introduction and Definition
  2. Application to Bayesian Inverse Problems
  3. Formal  $W_c$  gradient flow structure
  4. Splitting in shape and moments
  5. Variance - modulated OT

Key Reference: "Covariance - modulated optimal transport and gradient flows"  
[Burger, Erbar, Hoffmann, Matthes, Schlichting (2024)]

## 1. Introduction and Definition

Let  $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^d)$ . Recall the following three formulations of the Wasserstein-2 metric  $W_2(\mu_0, \mu_1)$ :

### • Monge Formulation:

$$W_2^2(\mu_0, \mu_1) = \inf_T \left\{ \int |x - T(x)|^2 d\mu_0(x) : T_{\#}\mu_0 = \mu_1 \right\} \quad \text{minimizing over transport maps } T: \mathbb{R}^d \rightarrow \mathbb{R}^d$$

### • Kantorovich Formulation:

$$W_2^2(\mu_0, \mu_1) = \inf_{\gamma} \left\{ \iint |x-y|^2 d\gamma(x,y) : \begin{array}{l} \pi_x \# \gamma = \mu_0, \\ \pi_y \# \gamma = \mu_1 \end{array} \right\} \quad \text{minimizing over transport plans } \gamma \in \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d)$$

### • Benamou-Brenier Formulation:

$$W_2^2(\mu_0, \mu_1) = \inf_{\mu, V} \left\{ \int_0^1 \int_{\mathbb{R}^d} \|V_t(x)\|^2 d\mu_t(x) dt : \begin{array}{l} \partial_t \mu_t + \operatorname{div}(\mu_t V_t) = 0, \\ \mu|_{t=0} = \mu_0, \mu|_{t=1} = \mu_1 \end{array} \right\} \quad \text{minimizing over paths } \{\mu_t\}_{t \in [0,1]} \in \mathcal{P}_2(\mathbb{R}^d)$$

## Definition (Covariance-Modulated OT metric)

$$W_C(\mu_0, \mu_1)^2 = \inf_{\mu, V} \left\{ \int_0^1 \int_{\mathbb{R}^d} \|V_t(x)\|_{C(\mu_t)}^2 d\mu_t(x) dt : \begin{array}{l} \partial_t \mu_t + \operatorname{div}(\mu_t V_t) = 0, \\ \mu|_{t=0} = \mu_0, \mu|_{t=1} = \mu_1 \end{array} \right\}$$

where

$$M(\mu_t) = \int x d\mu_t(x), \quad C(\mu_t) = \int (x - M(\mu_t))(x - M(\mu_t))^T d\mu_t(x)$$

and

$$\|V_t\|_{C(\mu_t)}^2 = \langle V_t, C^{-1}(\mu_t) V_t \rangle$$

Remark: Main difference with  $W_2$  distance is a competition effect between spreading mass and minimizing path length.

- GOALS:**
- Understand the structure of minimizers (= geodesics)
  - Consider gradient flows in  $W_C$  metric: analyse dynamics and characterize convergence properties



Question: Why use  $W_c$  for gradient flows rather than  $W_2$ ?

$\Rightarrow$  Convergence speed up: let  $\pi = N(m_\infty, C_\infty)$  be a Gaussian target.

Define the energy  $E(\rho | \pi) = \int \rho \log \frac{\rho}{\pi}$  (KL divergence / relative entropy)

- Gradient flow for  $W_2$ :

$$\partial_t \rho = -\nabla_{W_2} E(\rho | \pi) = \operatorname{div}(\rho \nabla \delta_\rho E(\rho | \pi))$$

Then along solutions  $\rho_t$

$$E(\rho_t | \pi) \leq e^{-2\lambda t} E(\rho_0 | \pi) \quad \text{if } C_\infty^{-1} \geq \lambda I_d$$

[Stam, Gross, Bakry, Émery, Villani, Ledoux, ...] (see Course by Tuhin Ghosh in prep workshop)

- Gradient flow for  $W_c$ :  $\partial_t \rho = -\nabla_{W_c} E(\rho | \pi) = \operatorname{div}(\rho C(\rho) \nabla \delta_\rho E(\rho | \pi))$

Then along solutions  $\rho_t$

$$E(\rho_t | \pi) \leq k e^{-2t} E(\rho_0 | \pi) \quad \text{where } k = k(C(\rho_0), C_\infty)$$

Well-posedness of  $W_c$ : It is possible that  $W_c = \infty$ . This can happen when  $C(\mu_t)$  degenerates.

[Thm]  $W_c(\mu_0, \mu_1) < \infty$  if  $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^d)$ ,  $\operatorname{Im} C_0 = \operatorname{Im} C_1$ ,  $M(\mu_0) - M(\mu_1) \in \operatorname{Im} C_0$

$\rightarrow$  For simplicity, let us assume  $\operatorname{rank} C_0 = \operatorname{rank} C_1 = d$

(Else, can restrict analysis to lower-dimensional subspace)

Remarks:

- $W_c^2(\mu_0, \mu_1) = 0$  iff  $\mu_0 = \mu_1$

- $W_c^2(\mu, \delta_x) = +\infty$  for any  $x \in \mathbb{R}^d$  and  $\delta_x \neq \mu \in \mathcal{P}_2(\mathbb{R}^d)$

## 2. Application in Bayesian Inverse Problems

Recall the Ensemble Kalman Sampler (EKS) from Lecture 1 for solving Bayesian Inverse Problems. In the mean-field limit, and after linearizing the forward map, it corresponds to a preconditioned gradient decent

$$dX_t = -C(\text{Law}(X_t)) \nabla J(x) + \sqrt{2C(\text{Law}(X_t))} dW_t$$

Denote

$$\pi(x) = \frac{1}{Z} e^{-J(x)}, \quad Z = \int e^{-J(x)} dx$$

Then  $\rho_t = \text{Law}(X_t)$  solves

$$\begin{aligned} \partial_t \rho &= \text{div}(\rho C(\rho) \nabla J + C(\rho) \nabla \rho) \\ &= \text{div}(\rho C(\rho) [\nabla J + \nabla \log \rho]) \\ &= \text{div}(\rho C(\rho) \nabla \log \rho_\pi) = \text{div}(\rho C(\rho) \nabla \delta_\rho E(\rho | \pi)) \end{aligned}$$

$$\Rightarrow \boxed{\partial_t \rho = -\nabla_{W_c} E(\rho | \pi)} \quad (1)$$

$\Rightarrow$  To analyse EKS algorithm at the mean-field level, use Cov-modulated gradient flows: e.g. understand convexity of  $E(\cdot | \pi)$  along  $W_c$ -geodesics.

By Itô's formula: if  $dX_t = b(t, X_t)dt + \sigma(t, X_t)dB_t$ ,  $b \in \mathbb{R}^d, \sigma \in \mathbb{R}^{d \times d}$  then  $\forall f \in C_c^\infty$ ,  $df(X_t) = L_t f(X_t)dt + \nabla f(X_t)^T \sigma(t, X_t)dB_t$  where the generator is given by  $L_t f(x) = b(t, x) \cdot \nabla f(x) + \frac{1}{2} \text{Tr}[\sigma \sigma^T(t, x) D^2 f(x)]$ . Taking  $\mathbb{E}$ , the  $dB_t$  term vanishes (it is a martingale). Then  $\frac{d}{dt} \mathbb{E}[f(X_t)] = \mathbb{E}[L_t f(X_t)] = \int L_t f(x) d\rho_t(x) \quad \forall f \in C_c^\infty$ . This is the weak formulation of the PDE  $\partial_t \rho_t = L_t^* \rho_t$ , where  $L_t^*$  = adjoint of  $L_t$ :  $L_t^* \rho = -\text{div}(\rho b) + \text{div}(\text{div}(\sigma \sigma^T \rho))$

### 3. Formal $W_c$ gradient flow structure

← formal Riemannian manifold picture

For simplicity, let's consider the subset

⚠ useful as computational heuristic

NOT rigorous as  $(\mathcal{P}_2, W_c)$  is NOT a manifold (lack of linear structure)

$$\mathcal{P}_+ := \{ \rho \in \mathcal{P}_2(\mathbb{R}^d) \cap L^1(\mathbb{R}^d) \cap C^\infty(\mathbb{R}^d) : \rho > 0 \text{ a.e.} \}$$

(same as Otto Calculus for  $(\mathcal{P}_2, W_2)$ )

For  $\rho \in \mathcal{P}_+$ , it holds  $C(\rho) > 0$ . This means we can apply the classical Otto calculus for  $(\mathcal{P}_2, W_2)$  on this set, adjusting the expressions with a  $C(\rho)$  weight.

### Tangent Space and Co-Tangent Space

The tangent space of  $(\mathcal{P}_+, W_c)$  at  $\rho \in \mathcal{P}_+$  is given by

$$T_\rho \mathcal{P}_+ = \{ \dot{\rho}|_{t=0} : \rho(t) = \text{curve in } \mathcal{P}_+, \rho(0) = \rho \} = \{ \sigma \in C^\infty(\mathbb{R}^d) : \int \sigma dx = 0 \}$$

Indeed, the mass constraint  $\int \sigma = 0$  follows since  $T_\rho \mathcal{P}_+$  is characterized by

$$\{ \sigma = \sigma_\phi := -\operatorname{div}(\rho C(\rho) \nabla \phi) : \phi \in C_c^\infty(\mathbb{R}^d) \}^{H_\rho^1}, \quad \|\phi\|_{H_\rho^1}^2 := \int \langle \nabla \phi, C(\rho) \nabla \phi \rangle d\rho$$

Note that

$$T_\rho \mathcal{P}_+ \simeq \overline{\{ \nabla \phi : \phi \in C_c^\infty(\mathbb{R}^d) \}}^{L_{\bar{\rho}}^2} \quad \text{where } \bar{\rho} := \bar{C}(\rho)(\cdot - m(\rho))_{\#} \rho$$

⇒ The tangent space has a Hilbert space structure.

The dual of  $T_\rho \mathcal{P}_+$  is the cotangent space  $T_\rho^* \mathcal{P}_+ := \{ \xi : T_\rho \mathcal{P}_+ \rightarrow \mathbb{R} \text{ linear and bounded} \}$  equipped with the dual norm  $\|\xi\|_{T_\rho^* \mathcal{P}_+} := \sup_{\substack{\sigma = -\Delta_\rho \phi \in T_\rho \mathcal{P}_+ \\ \|\phi\|_{H_\rho^1} \leq 1}} |\xi(\sigma)|$ , where  $\xi(\sigma) = \langle \xi, \sigma \rangle = \int \xi \sigma dx$  is the dual pairing.

$$\Rightarrow \|\xi\|_{T_\rho^* \mathcal{P}_+} = \|\xi\|_{H_\rho^1}$$

$$\begin{aligned} &= \int \xi (-\Delta_\rho \phi) dx = \int \langle \nabla \xi, C(\rho) \nabla \phi \rangle_\rho dx \\ &\leq \left( \int \langle \nabla \xi, C(\rho) \nabla \xi \rangle d\rho \right)^{1/2} \left( \int \langle \nabla \phi, C(\rho) \nabla \phi \rangle d\rho \right)^{1/2} \\ &= \|\xi\|_{H_\rho^1} \|\phi\|_{H_\rho^1} \end{aligned}$$

### Onsager operator and Metric tensor.

Write the weighted elliptic operator  $\Delta_\rho \phi := \operatorname{div}(\rho C(\rho) \nabla \phi)$ . Then the Onsager operator  $\mathbb{K}_\rho$

is given by  $\mathbb{K}_\rho : T_\rho^* \mathcal{P}_+ \rightarrow T_\rho \mathcal{P}_+$  and its inverse  $\mathbb{G}_\rho = \mathbb{K}_\rho^{-1} : T_\rho \mathcal{P}_+ \rightarrow T_\rho^* \mathcal{P}_+$

$$\xi \mapsto -\Delta_\rho \xi$$

$$\sigma \mapsto \phi := (-\Delta_\rho)^{-1} \sigma$$

$\mathbb{G}_\rho$  is the metric tensor. It can also be expressed as  $g_\rho : T_\rho \mathcal{P}_+ \times T_\rho \mathcal{P}_+ \rightarrow \mathbb{R}$

$$\text{where } (\mathbb{G}_\rho \sigma_1)(\sigma_2) = \langle (-\Delta_\rho)^{-1} \sigma_1, \sigma_2 \rangle = \langle \phi_1, -\Delta_\rho \phi_2 \rangle \quad (\sigma_1, \sigma_2) \mapsto (\mathbb{G}_\rho \sigma_1)(\sigma_2)$$

$$= \int \langle \nabla \phi_1, C(\rho) \nabla \phi_2 \rangle d\rho = (\mathbb{G}_\rho \sigma_2)(\sigma_1)$$

⇒  $g_\rho$  or  $\mathbb{G}_\rho$  or  $\mathbb{K}_\rho$  equivalently

specify the dissipation mechanism.

### Rate functionals.

$$R(\rho, \sigma) = \frac{1}{2} \int \langle \nabla \phi, C(\rho) \nabla \phi \rangle d\rho = \frac{1}{2} \|\phi\|_{H_\rho^1}^2 \quad \text{for } \phi := (-\Delta_\rho)^{-1} \sigma = \mathbb{G}_\rho \sigma$$

$$\text{If we denote } V = C(\rho) \nabla \phi, \text{ then } R(\rho, \sigma) = \frac{1}{2} \int \langle V, \bar{C}(\rho)^{-1} V \rangle d\rho = \frac{1}{2} \int \|V\|_{\bar{C}(\rho)}^2 d\rho$$

$$R^*(\rho, \xi) = \frac{1}{2} \int \langle \nabla \xi, C(\rho) \nabla \xi \rangle d\rho = \frac{1}{2} \|\xi\|_{H_\rho^1}^2 = \sup_{\tilde{\sigma} \in T_\rho \mathcal{P}_+} [\langle \xi, \tilde{\sigma} \rangle - R(\rho, \tilde{\sigma})] \quad \text{Legendre transform of } R(\rho, \cdot)$$

Directional derivative of  $F: \mathcal{P}_+ \rightarrow \mathbb{R}$  at  $\rho \in \mathcal{P}_+$  is

$$DF[\rho]: T_\rho \mathcal{P}_+ \rightarrow \mathbb{R}$$

Then  $DF[\rho] \in T_\rho^* \mathcal{P}_+$

$$\sigma \mapsto \langle DF[\rho], \sigma \rangle = \left. \frac{d}{d\varepsilon} F(\rho + \varepsilon \sigma) \right|_{\varepsilon=0} = \langle \delta_\rho F[\rho], \sigma \rangle_{L^2}$$

can be identified with  $\delta_\rho F[\rho](\cdot)$  if  $\|\nabla \delta_\rho F\|_{H_p^1} < \infty$

$$\text{since } \langle \delta_\rho F, \sigma \rangle_{L^2} = \langle \delta_\rho F, -\Delta_\rho \phi \rangle_{L^2} \leq \|\nabla \delta_\rho F\|_{H_p^1} \|\nabla \phi\|_{H_p^1}$$

$$I(\rho) = \|\nabla \delta_\rho F\|_{T_\rho \mathcal{P}_+}^2 = \|\nabla \delta_\rho F\|_{H_p^1}^2 = R(\rho, \delta_\rho F) \text{ is called the Fisher information of } F.$$

← expression depends on the choice of metric

$\rho + \varepsilon \sigma \in \mathcal{P}_+$  since  $\int \sigma dx = 0$ ,  $\sigma = -\Delta_\rho \phi$   
first variation of  $F$

Energy-dissipation identity: Consider  $F(\rho) = E(\rho|\pi) = \int \rho \log \frac{\rho}{\pi} dx$  if  $\rho \ll \pi$  and  $+\infty$  else.

Differentiate along solutions  $\rho_t$  to (1):

$$\begin{aligned} \frac{d}{dt} E(\rho_t|\pi) &= \int \partial_t \rho \log \frac{\rho}{\pi} dx + \int \rho \partial_t \log \frac{\rho}{\pi} dx \\ &= \int \partial_t \rho \delta_\rho E(\rho|\pi) dx + \int \rho \frac{\pi}{\rho} \frac{\partial_t \rho}{\pi} dx \\ &= \int \operatorname{div}(\rho C(\rho) \nabla \delta_\rho E(\rho|\pi)) dx + \underbrace{\int \partial_t \rho dx}_{=0 \text{ as } \rho \text{ solves a continuity eqn on } \mathbb{R}^d} \\ &= - \int \langle \nabla \delta_\rho E(\rho|\pi), C(\rho) \nabla \delta_\rho E(\rho|\pi) \rangle \rho dx \quad \left. \begin{array}{l} \text{ } \\ \text{ } \end{array} \right\} V_t(x) := C(\rho_t) \nabla \delta_{\rho_t} E(\rho_t|\pi) \\ &= - \int \langle V, \bar{C}'(\rho) V \rangle \rho dx \\ &= -I(\rho_t|\pi) \quad \leftarrow \text{Fisher information of relative entropy} \end{aligned}$$

Another way to see the same equation is via the kinetic relation: for any  $\xi \in T_\rho^* \mathcal{P}_+$ ,  $\sigma \in T_\rho \mathcal{P}_+$ ,

$$\begin{aligned} \langle \xi, \sigma \rangle &= \langle \xi, -\Delta_\rho \phi \rangle = \int \langle \nabla \xi, C(\rho) \nabla \phi \rangle d\rho \\ &\leq \left( \int \langle \nabla \xi, C(\rho) \nabla \xi \rangle d\rho \right)^{1/2} \left( \int \langle \nabla \phi, C(\rho) \nabla \phi \rangle d\rho \right)^{1/2} \\ &\leq \frac{1}{2} \|\xi\|_{H_p^1}^2 + \frac{1}{2} \|\phi\|_{H_p^1}^2 = R(\rho, \sigma) + R^*(\rho, \xi) \end{aligned}$$

Equality is reached if and only if

$$\sigma = K_\rho \xi$$

$$= \frac{1}{2} \langle \sigma, G_\rho \sigma \rangle + \frac{1}{2} \langle \xi, K_\rho \xi \rangle$$

Taking  $\xi = -DF[\rho]$  and  $\sigma = \dot{\rho}_t$ , we can characterize the gradient flow via the scalar relation

$$R(\rho, \dot{\rho}_t) + R^*(\rho, -DF[\rho]) = \langle -DF[\rho], \dot{\rho}_t \rangle = -\frac{d}{dt} F(\rho_t) \quad (2)$$

This is exactly the energy-dissipation identity above.

Gradient Flow Structure: Another way to write (2) is

$$g_{\rho}(\dot{\rho}_t, \sigma) = -\mathcal{D}F[\rho_t](\sigma) \quad \forall \sigma \in T_{\rho_t} \mathcal{P}_+$$

The gradient of  $F$  w.r.t.  $g$  is defined as

$$\nabla_g F(\rho) = \mathbb{K}_{\rho} \mathcal{D}F[\rho] = \mathbb{K}_{\rho} \delta_{\rho} F \in T_{\rho} \mathcal{P}_+$$

identify directional derivative with first variation

For  $W_c$ ,  $\nabla_{W_c} E(\rho|\pi) = -\operatorname{div}(\rho C(\rho) \nabla \delta_{\rho} E(\rho|\pi)) = \mathbb{K}_{\rho} \delta_{\rho} E(\rho|\pi)$

Then  $g_{\rho_t}(\dot{\rho}_t, \sigma) = -\mathcal{D}E[\rho_t](\sigma) = g_{\rho_t}(-\mathbb{K}_{\rho_t} \mathcal{D}E[\rho_t], \sigma) = g_{\rho_t}(-\nabla_{W_c} E(\rho_t), \sigma)$

$$\forall \sigma \in T_{\rho_t} \mathcal{P}_+$$

$$\Rightarrow \boxed{\dot{\rho}_t = -\nabla_{W_c} E(\rho_t|\pi)}$$

Remark  $\mathcal{D}E \in T_{\rho}^* \mathcal{P}_+$  is independent of the choice of metric, whereas  $\nabla_{W_c} E$  is not!

Refs For more details on Onsager operators

[Mielke] "Geodesic convexity of the relative entropy in reversible Markov chains"

For an application to the Fokker-Planck equation: [Eberle, Niethammer, Schlichting]

[Mielke] "An introduction to the analysis of gradient systems" [2023, Lecture Notes]

#### 4. Splitting in Shape and Moments

Key result (informal):

$$W_c(\mu_0, \mu_1)^2 = \inf_{R \in SO(d)} \left\{ \begin{array}{l} \text{Distance in shape } (R_{\#} \bar{\mu}_0, \bar{\mu}_1) \\ + \\ \text{Distance with } R \text{ constraint } ((m_0, C_0), (m_1, C_1)) \end{array} \right\}$$

$\bar{\mu}$  = normalization of  $\mu$

special orthogonal matrices (= rotations):  
 $SO(d) = \{R \in \mathbb{R}^{d \times d} : R^T = R^{-1}, \det R = 1, \text{ real entries}\}$

$(m_i, C_i) = (M(\mu_i), C(\mu_i))$

**Definition**  $A \in \mathbb{R}^{d \times d}$  (possibly non-symmetric) is the **left square root** of  $C \in S_+^d = \{M \in \mathbb{R}^{d \times d} : M^T = M, M \succ 0\}$  if  $AA^T = C$ .

← also defined for  $S_{\geq 0}^d$ , but then characterization of non-uniqueness via orthogonal transformations no longer holds ( $Q_1 \neq Q_2$  may have  $A = C^{1/2} Q_1 = C^{1/2} Q_2$ )

→ Given  $C$ , choice of  $A$  is not unique

→ Denote  $A = C^{1/2}$  the unique symmetric positive definite square root of  $C$ .

→ For any  $Q \in O(d) = \{\text{orthogonal matrices in } \mathbb{R}^{d \times d}\}$ ,  $A = C^{1/2} Q$  is a left square root of  $C$ .

**Definition**  $\hat{\mu}$  is a **normalization** of  $\mu$  if  $M(\hat{\mu}) = 0$ ,  $C(\hat{\mu}) = I_d$ .

→ Given  $\mu \in \mathcal{P}_2(\mathbb{R}^d)$  with  $C(\mu) > 0$ , let  $m = M(\mu)$  and  $A$  any left square root of  $C(\mu)$ . Define

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

Then  $\hat{\mu} := T_{m,A} \# \mu$  is a normalization of  $\mu$ . (All normalizations can be written this way.)

→ Normalizations are non-unique due to non-uniqueness of the choice of  $A$ .

**Splitting of kinetic energy:** Rewrite  $W_c$  as

$$W_c(\mu_0, \mu_1)^2 = \inf_{\substack{(m_t, C_t)_{t \in [0,1]}}} \left\{ W_{m,C}(\mu_0, \mu_1)^2 \mid \begin{array}{l} m: [0,1] \rightarrow \mathbb{R}^d \\ C: [0,1] \rightarrow S_{\geq 0}^d \end{array} \text{ with } \begin{array}{l} (m(0), C(0)) = (M(\mu_0), C(\mu_0)) \\ (m(1), C(1)) = (M(\mu_1), C(\mu_1)) \end{array} \right\}$$

$$\text{where } W_{m,C}(\mu_0, \mu_1)^2 = \inf_{\mu, V} \left\{ \int_0^1 \int_{\mathbb{R}^d} \|V_t\|_{C_t}^2 d\mu_t dt, \begin{array}{l} \partial_t \mu + \operatorname{div}(\mu V) = 0, \mu|_{t=0} = \mu_0, \\ M(\mu_t) = m_t, C(\mu_t) = C_t, \mu|_{t=1} = \mu_1 \end{array} \right\}$$

Let  $\hat{\mu}_t := (\bar{A}_t^{-1}(\cdot - m_t))_{\#} \mu_t$  be a normalization of the path  $\mu_t$  solving  $\partial_t \mu_t + \text{div}(\mu_t V_t) = 0$ .

Then  $\hat{\mu}_t$  solves

$$\begin{aligned} \underbrace{\frac{d}{dt} \int \phi(x) d\hat{\mu}_t(x)}_{- \int \phi \text{div}(\hat{\mu}_t \hat{V}_t)} &= \frac{d}{dt} \int \phi(\bar{A}_t^{-1}(x - m_t)) d\mu_t(x) = \int -\nabla \phi(\bar{A}_t^{-1}(x - m_t)) \cdot [\bar{A}_t^{-1} \dot{\bar{A}}_t \bar{A}_t^{-1}(x - m_t) + \bar{A}_t^{-1} \dot{m}_t] d\mu_t - \int \phi(\bar{A}_t^{-1}(x - m_t)) \text{div}(\mu_t V_t) dx \\ &= \int \bar{A}_t^{-T} \nabla \phi(\bar{A}_t^{-1}(x - m_t)) \cdot V_t d\mu_t(x) \\ &= \int \nabla \phi(\bar{A}_t^{-1}(x - m_t)) \cdot \bar{A}_t^{-1} V_t d\mu_t(x) \\ \Rightarrow \hat{\mu}_t \text{ solves } \partial_t \hat{\mu}_t + \text{div}(\hat{\mu}_t \hat{V}_t) &= 0 \text{ with } \boxed{\hat{V}_t(x) = \bar{A}_t^{-1} [V_t(\bar{T}_t^{-1}(x)) - \dot{\bar{A}}_t x - \dot{m}_t]} \end{aligned}$$

splitting is exact!

**Lemma 1** The kinetic energy splits into

$$\int \int \|V_t\|_{C_t}^2 d\mu_t dt = \underbrace{\int \int \|\hat{V}_t\|^2 d\hat{\mu}_t dt}_{\text{shape}} + \underbrace{\int \left[ \dot{m}_t^T \bar{C}_t^{-1} \dot{m}_t + \frac{1}{4} \text{Tr} [\dot{\bar{C}}_t \bar{C}_t^{-1} \dot{\bar{C}}_t \bar{C}_t^{-1}] \right] dt}_{\text{moments}}$$

**Proof:** Define  $r_t(x) = \bar{A}_t^{-1} [\dot{\bar{A}}_t \bar{T}_t^{-1}(x) + \dot{m}_t]$ . Then

$$\begin{aligned} \int \|\hat{V}_t(x)\|^2 d\hat{\mu}_t(x) &= \int \|\hat{V}_t(\bar{T}_t^{-1}(x))\|^2 d\mu_t(x) = \int \|\bar{A}_t^{-1} [V_t(x) - \dot{\bar{A}}_t \bar{T}_t^{-1}(x) - \dot{m}_t]\|^2 d\mu_t(x) = \int \|\bar{A}_t^{-1} V_t(x) - r_t(x)\|^2 d\mu_t(x) \\ &= \int \|V_t\|_{C_t}^2 d\mu_t - \text{I} - \text{II} \quad \text{since } \bar{A}_t^T \bar{A}_t^{-1} = (\bar{A}_t \bar{A}_t^T)^{-1} = \bar{C}_t^{-1} \end{aligned}$$

with

$$\text{I} = \int \|r_t(x)\|^2 d\mu_t(x) \quad \text{and} \quad \text{II} = 2 \int \langle r_t(x), \bar{A}_t^{-1} V_t(x) - r_t(x) \rangle d\mu_t(x)$$

Dropping  $t$  for convenience:

$$\begin{aligned} \|r\|^2 &= \|\bar{A}^{-1} [\dot{\bar{A}} \bar{A}^{-1}(x - m) + \dot{m}]\|^2 = \langle \dot{m}, \bar{A}^T \bar{A}^{-1} \dot{m} \rangle + 2 \langle \bar{A}^{-1} \dot{\bar{A}} \bar{A}^{-1}(x - m), \bar{A}^{-1} \dot{m} \rangle + \langle x - m, \bar{A}^T \bar{A}^{-1} \bar{A}^T \bar{A}^{-1} \dot{\bar{A}} \bar{A}^{-1}(x - m) \rangle \\ &= \|\dot{m}\|_{\bar{C}}^2 + \langle \dot{m}, \bar{C} \dot{\bar{C}} \bar{C}^{-1}(x - m) \rangle + \frac{1}{4} \langle x - m, \bar{C} \dot{\bar{C}} \bar{C}^{-1} \dot{\bar{C}} \bar{C}^{-1}(x - m) \rangle \quad \text{Assuming } \boxed{\dot{\bar{A}} \bar{A}^{-1} = \frac{1}{2} \dot{\bar{C}} \bar{C}^{-1}} \quad (2) \end{aligned}$$

$$\begin{aligned} \Rightarrow \int \|r(x)\|^2 d\mu(x) &= \int \|\dot{m}\|_{\bar{C}}^2 d\mu + \langle \dot{m}, \bar{C} \dot{\bar{C}} \bar{C}^{-1} \int (x - m) d\mu(x) \rangle \\ &\quad + \frac{1}{4} \text{Tr} \left[ \underbrace{\int (x - m)(x - m)^T d\mu(x)}_{=C} \dot{\bar{C}} \bar{C} \bar{C}^{-1} \dot{\bar{C}} \bar{C}^{-1} \right] = 0 \\ \Rightarrow \text{I} &= \|\dot{m}_t\|_{\bar{C}_t}^2 + \frac{1}{4} \text{Tr} [\dot{\bar{C}} \bar{C} \bar{C}^{-1} \dot{\bar{C}} \bar{C}^{-1}] \end{aligned}$$

Next, we claim that  $\int \text{II}(t) dt = 0$ :

$$\begin{aligned} \text{II} &= 2 \int \langle r, \bar{A}^{-1} V - r \rangle d\mu = 2 \int \langle r(\bar{T}^{-1}(x)), \bar{A}^{-1} V(\bar{T}^{-1}(x)) - r(\bar{T}^{-1}(x)) \rangle d\hat{\mu}(x) \\ &= 2 \int \langle \bar{A}^{-1} \dot{\bar{A}} x + \bar{A}^{-1} \dot{m}, \bar{A}^{-1} [V(\bar{T}^{-1}(x)) - \dot{\bar{A}} x - \dot{m}] \rangle d\hat{\mu}(x) = 2 \int \langle \bar{A}^{-1} \dot{\bar{A}} x + \bar{A}^{-1} \dot{m}, \hat{V}(x) \rangle d\hat{\mu}(x) \end{aligned}$$

Define  $\eta_t(x) = \frac{1}{2} \langle x, \bar{A}_t^{-1} \dot{\bar{A}}_t x \rangle + \langle x, \bar{A}_t^{-1} \dot{m}_t \rangle$ . By (2),  $\bar{A}_t^{-1} \dot{\bar{A}}_t = \frac{1}{2} \bar{A}_t^{-1} \dot{\bar{C}} \bar{C}_t^{-1} \bar{A}_t = \frac{1}{2} \bar{A}_t^{-1} \dot{\bar{C}} \bar{A}_t^{-T}$  is symmetric

Since  $H(\hat{\mu}_t) = 0$ ,  $C(\hat{\mu}_t) = \text{Id}$ , we have  $\int \eta_t(x) d\hat{\mu}_t(x) = \frac{1}{2} \text{Tr} [\bar{A}_t^{-1} \dot{\bar{A}}_t]$ , and so

$$\begin{aligned} \int \partial_t \eta_t(x) d\hat{\mu}_t(x) &= \int \left[ \frac{1}{2} \langle x, \partial_t (\bar{A}_t^{-1} \dot{\bar{A}}_t) x \rangle + \langle x, \partial_t (\bar{A}_t^{-1} \dot{m}_t) \rangle \right] d\hat{\mu}_t(x) = \frac{1}{2} \text{Tr} [\partial_t (\bar{A}_t^{-1} \dot{\bar{A}}_t)] \\ &= \frac{d}{dt} \frac{1}{2} \text{Tr} [\bar{A}_t^{-1} \dot{\bar{A}}_t] = \frac{d}{dt} \int \eta_t d\hat{\mu} \end{aligned}$$

Using  $\eta_t$  as a testfunction in the weak formulation of the continuity equation for  $(\hat{\mu}, \hat{V})$ :

$$\int_0^1 \int_{\mathbb{R}^d} \nabla \eta_t \cdot \hat{V}_t d\hat{\mu}_t dt = \int \eta_1 d\hat{\mu}_1 - \int \eta_0 d\hat{\mu}_0 - \int_0^1 \int \partial_t \eta_t d\hat{\mu}_t dt = \int \eta_1 d\hat{\mu}_1 - \int \eta_0 d\hat{\mu}_0 - \int_0^1 \frac{d}{dt} \int \eta_t d\hat{\mu}_t dt = 0$$

and  $\nabla \eta_t = \dot{A}_t^T \dot{A}_t^{-T} x + \dot{A}_t^T \dot{m}_t = \dot{A}_t^T \dot{A}_t^{-T} x + \dot{A}_t^T \dot{m}_t$  by symmetry, and so

$$\int_0^1 \mathbb{I}(t) dt = \int_0^1 \int \langle \dot{A}_t^T \dot{A}_t^{-T} x + \dot{A}_t^T \dot{m}_t, \hat{V}_t(x) \rangle d\hat{\mu}_t(x) dt = \text{LHS} = 0.$$

Putting everything together:

$$\int_0^1 \int \|V_t\|_{C_t}^2 d\mu_t dt = \int_0^1 \int \|\hat{V}_t\|^2 d\hat{\mu}_t dt + \int_0^1 \mathbb{I}(t) dt + \int_0^1 \mathbb{II}(t) dt$$

$$= \int_0^1 \int \|\hat{V}_t\|^2 d\hat{\mu}_t dt + \int_0^1 \left( \|\dot{m}_t\|_{C_t}^2 + \frac{1}{4} \text{Tr}[\dot{C} \dot{C}^{-1} \dot{C} \dot{C}^{-1}] \right) dt + 0.$$

□

Splitting of the metric:

Definition (Covariance-Constrained OT)

Let  $\mu_0, \mu_1 \in \mathcal{P}_{0,1}(\mathbb{R}^d) := \{ \rho \in \mathcal{P}_2(\mathbb{R}^d) : M(\rho) = 0, C(\rho) = I_d \}$ .

$$W_{0,1}(\mu_0, \mu_1)^2 = \inf_{(\mu, V)} \left\{ \int_0^1 \int \|V_t\|^2 d\mu_t dt : \begin{array}{l} \partial_t \mu_t + \text{div}(\mu_t V_t) = 0, \quad M(\mu_t) = 0, \\ \mu|_{t=0} = \mu_0, \mu|_{t=1} = \mu_1, \quad C(\mu_t) = I_d \end{array} \right\}$$

Definition (Rotation-constrained Moment Metric)

Fix  $R \in SO(d)$ . For a given curve  $C_t \in AC([0,1], S_{\geq 0}^d)$ , define  $A_t$  as the unique solution to

$$\dot{A}_t = \frac{1}{2} \dot{C}_t A_t^{-T}, \quad A_0 = C_0^{1/2} \quad (3)$$

Then define

$$\mathcal{D}_R(\mu_0, \mu_1)^2 = \inf_{(m, C)} \left\{ \int_0^1 \|\dot{m}_t\|_{C_t}^2 + \frac{1}{4} \text{Tr}[\dot{C}_t \dot{C}_t^{-1} \dot{C}_t \dot{C}_t^{-1}] dt : \begin{array}{l} (m_0, C_0) = (M(\mu_0), C(\mu_0)), \\ (m_1, C_1) = (M(\mu_1), C(\mu_1)), \end{array} \quad \bar{C}_t^{1/2} A_t = R \right\}$$

Denote by  $\bar{\mu} = \bar{T}_\# \mu$ ,  $\bar{T} = \bar{C}_t^{1/2} (x - M(\mu))$ .

Theorem 2 (Splitting of  $W_C$ ) For  $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^d)$  with  $C(\mu_i) > 0$ ,

$$W_C(\mu_0, \mu_1)^2 = \inf_{R \in SO(d)} \left\{ W_{0,1}(R_\# \bar{\mu}_0, \bar{\mu}_1)^2 + \mathcal{D}_R(\mu_0, \mu_1) \right\}$$



Remark If either  $\bar{\mu}_0$  or  $\bar{\mu}_1$  is spherically symmetric, then

$$W_c(\mu_0, \mu_1) = W_{0, Id}(\bar{\mu}_0, \bar{\mu}_1) + D(\mu_0, \mu_1)$$

where 
$$D(\mu_0, \mu_1) := \inf_{R \in SO(d)} D_R(\mu_0, \mu_1) = \inf_{(m, C)} \left\{ \int_0^1 \left( \| \dot{m}_t \|^2_{C_t} + \frac{1}{4} \text{Tr} [\dot{C}_t^{-1} \dot{C}_t^{-1}] \right) dt : \begin{matrix} (m_0, C_0) = (M(\mu_0), C(\mu_0)), \\ (m_1, C_1) = (M(\mu_1), C(\mu_1)). \end{matrix} \right\}$$

→ Splitting is exact in this case.

Remark (Choice of left square root)

- Given  $C_t$ , there is a unique  $A_t$  solving (3), define curve of co-rotations  $R: AC([0,1], S_{\geq 0}^d) \rightarrow AC([0,1], SO(d))$

Then  $R_0 = Id$ , and  $R_t \in SO(d)$ .

$$C_t \mapsto C_t^{-1/2} A_t.$$

(Pf Sketch: ODE for  $R_t$  is such that solutions remain in  $SO(d)$  and  $R_0 = Id \in SO(d)$ .)

- Solution  $A_t$  to (3) satisfies (2) (hence  $\dot{A}_t^T \dot{A}_t$  is symmetric), and is indeed a left square root of  $C_t$ :

$$\frac{d}{dt} (A_t A_t^T) = \dot{A}_t A_t^T + A_t \dot{A}_t^T = \frac{1}{2} \dot{C}_t + \frac{1}{2} \dot{C}_t = \dot{C}_t \quad \text{and} \quad A_0 A_0^T = C_0.$$

Both is true for any choice of  $A_0$  s.t.  $A_0 A_0^T = C_0$  as initial condition in (3).

Degree of freedom: fixing  $A_0$  equivalent to fixing  $R_1 = C_1^{-1/2} A_1$ .

⇒ choice of  $R$  in splitting of  $W_c$  equivalent to choice of  $A_0$  when normalizing  $\mu_t$ .

- In Lemma 1, we used normalization  $\hat{\mu}_t := \dot{A}_t^T (-m_t) \# \mu_t$  with  $A_t$  satisfying (2).

In Theorem 2, we used normalization  $\bar{\mu}_t := C_t^{-1/2} (-m_t) \# \mu_t$  with symmetric square root  $C_t^{1/2}$ .

To relate  $\hat{\mu}_t$  to  $\bar{\mu}_t$ , write  $A_t = C_t^{1/2} R_t$ . Then

$$\bar{\mu}_t = R_t \# \hat{\mu}_t$$

$$\left( \text{Pf: } \int \phi d\hat{\mu}_t = \int \phi (\dot{A}_t^T (x-m)) d\mu_t = \int \phi (\dot{R}_t^T C_t^{-1/2} (x-m)) d\mu_t = \int \phi (\dot{R}_t^T z) d\bar{\mu}_t(z) \Rightarrow \hat{\mu}_t = \dot{R}_t^T \# \bar{\mu}_t \right)$$

Hence, with  $R = C_t^{1/2} A_t$ , we have  $(\hat{\mu}_0, \hat{\mu}_1) = (\bar{\mu}_0, R_1^T \bar{\mu}_1)$  if we chose the initial condition in (3) as  $A_0 = C_0^{1/2}$ .

Note that  $W_{0, Id}$  is invariant by rotation: for any  $\tilde{R} \in SO(d)$ ,  $W_{0, Id}(\tilde{R} \# \bar{\mu}_0, \tilde{R} \# \bar{\mu}_1) = W_{0, Id}(\bar{\mu}_0, \bar{\mu}_1)$

⇒ In Theorem 2, we have  $W_{0, Id}(\hat{\mu}_0, \hat{\mu}_1) = W_{0, Id}(\bar{\mu}_0, R_1^T \bar{\mu}_1) = W_{0, Id}(R_1 \# \bar{\mu}_0, \bar{\mu}_1)$

Corollary 3 (Gaussians) If either  $\mu_0$  or  $\mu_1$  is a Gaussian, then the splitting in Theorem 2 is exact.

Proof: For a Gaussian  $N(m, C)$  and rotation  $R \in SO(d)$ , observe that

$$R \# N(m, C) = N(Rm, RCR^T)$$

Hence,

$$R \# \overline{N(m, C)} = R \# N(0, Id) = N(0, Id) = \overline{N(m, C)} \quad \square$$

## 5. Variance - Modulated Optimal Transport

Definition (Variance-Modulated OT) For  $\text{var}(\mu_t) = \text{Tr } C_t$ , define

$$W_{\text{var}}(\mu_0, \mu_1)^2 = \inf_{(\mu, V)} \left\{ \int_0^1 \frac{1}{\text{var}(\mu_t)} \int \|V_t\|^2 d\mu_t dt, \begin{array}{l} \partial_t \mu + \text{div}(\mu V) = 0, \\ \mu|_{t=0} = \mu_0, \mu|_{t=1} = \mu_1 \end{array} \right\}$$

Remark Similarly as for  $W_c$ ,  $W_{\text{var}}(\mu_0, \mu_1) = 0$  iff  $\mu_0 = \mu_1$ .

Further,  $W_{\text{var}}(\mu_0, \mu_1) = +\infty$  if either  $\text{var}(\mu_0) = 0$  or  $\text{var}(\mu_1) = 0$ .

In particular, the distance to any Dirac is infinite

For any  $\mu \in \mathcal{P}_2(\mathbb{R}^d)$  with  $M(\mu) = m$  and  $\text{var}(\mu) = \sigma^2 > 0$ ,

let  $\tilde{\mu}$  be a normalization of  $\mu$  such that  $M(\tilde{\mu}) = 0$  and  $\text{var}(\tilde{\mu}) = 1$ .

Then

$$\tilde{\mu} = (\sigma^{-1}(\cdot - m))_{\#} \mu$$

Theorem 4 | (Splitting of  $W_{\text{var}}$ )  $W_{\text{var}}(\mu_0, \mu_1)^2 = W_{0,1}(\tilde{\mu}_0, \tilde{\mu}_1)^2 + D_{\text{var}}(\mu_0, \mu_1)^2$

$$\text{where } W_{0,1}^2(\tilde{\mu}_0, \tilde{\mu}_1) := \inf_{(\mu, V)} \left\{ \int_0^1 \int \|V_t\|^2 d\mu_t dt : \begin{array}{l} \partial_t \mu + \text{div}(\mu V) = 0, \quad M(\mu_t) = 0, \\ \mu|_{t=0} = \tilde{\mu}_0, \mu|_{t=1} = \tilde{\mu}_1, \quad \text{var}(\mu_t) = 1 \end{array} \right\}$$

with normalization  $\tilde{\mu}$  s.t.  $M(\tilde{\mu}) = 0$ ,  $\text{var}(\tilde{\mu}) = 1$ , and where

$$D_{\text{var}}(\mu_0, \mu_1)^2 = \inf_{(m, \sigma)} \left\{ \int_0^1 \frac{\|\dot{m}_t\|^2 + |\dot{\sigma}_t|^2}{|\sigma_t|^2} dt : \begin{array}{l} (m_0, \sigma_0) = (M(\mu_0), \text{var}(\mu_0)), \\ (m_1, \sigma_1) = (M(\mu_1), \text{var}(\mu_1)). \end{array} \right\}$$

Remark The constrained minimization problem  $W_{0,1}$  has been studied in detail by [Carlen, Gangbo 2003]. Optimal curves in  $W_{0,1}$  are explicitly computable: Let  $\rho_t$  be the optimal  $W_2$  curve connecting  $\tilde{\mu}_0$  and  $\tilde{\mu}_1$ . Then the optimal  $W_{0,1}$  curve between  $\tilde{\mu}_0$  and  $\tilde{\mu}_1$  is the normalization of  $\rho_t(t)$  for a time reparametrization  $\tau(t)$ . ← so that after normalizing it's still constant speed  
Note that  $M(\rho_t) = 0$  and  $\text{var}(\rho_t) = 1 - t(1-t) W_2(\tilde{\mu}_0, \tilde{\mu}_1)$  which exists and is unique

Optimal curves for the three optimization problems  $W_{\text{var}}$ ,  $W_{0,1}$ ,  $D_{\text{var}}$  exist and are given explicitly.

### Theorem 5 | (Optimal Curves)

- The optimal curve  $\mu_t$  for  $W_{\text{var}}$  exists and is obtained by

$$\mu_t = (\sigma_t x + m_t)_{\#} \tilde{\mu}_t, \text{ where } \tilde{\mu}_t = \text{optimal curve for } W_{0,1}$$

$$(m_t, \sigma_t) = \text{optimal curve for } D_{\text{var}}$$

- The optimal curve  $\tilde{\mu}_t$  for  $W_{0,1}$  is given in [Carlen, Gangbo 2003] (see Remark above)

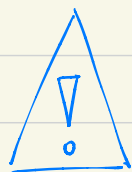
- The optimality conditions for  $D_{\text{var}}$  are given by

and for  $n := |m_1 - m_0|$ ,

$$\frac{d}{dt} \left( \frac{\dot{m}}{\sigma^2} \right) = 0, \quad \frac{\ddot{\sigma}}{\sigma} - \frac{(\dot{\sigma})^2}{\sigma^2} = - \frac{(\dot{m})^2}{\sigma^2}$$

$$D_{\text{var}}(\mu_0, \mu_1)^2 = \frac{1}{2} \left| \log \left( \frac{n^2 + \sigma_0^2 + \sigma_1^2 - \sqrt{(n^2 + \sigma_0^2 + \sigma_1^2)^2 - 4\sigma_0^2\sigma_1^2}}{2\sigma_0\sigma_1} \right) \right|^2$$

Explicit expressions for  $(m_t, \sigma_t)$  can be computed.



Existence of geodesics is a much harder problem for Covariance-modulated OT. For instance, geodesics for  $W_{0,1d}$  cannot be obtained from normalizing and rescaling  $W_2$  geodesics between  $\bar{\mu}_0$  and  $\bar{\mu}_1$ .