Geodesic Density Tracking with Applications to Data Driven Modeling

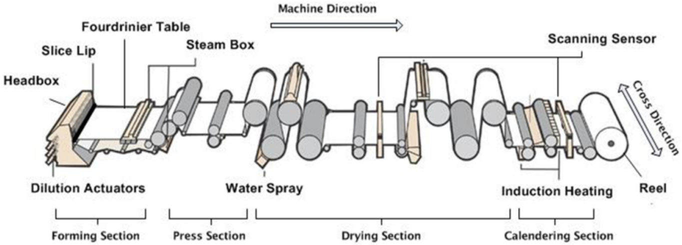
Abhishek Halder and Raktim Bhattacharya

Department of Aerospace Engineering
Texas A&M University
College Station, TX 77843-3141

American Control Conference
June 4, 2014

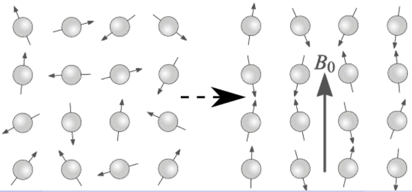
Motivating Applications

- **Process industry applications**



Source: Chu *et.al.* (2011), "Model Predictive Control and Optimization for Papermaking Process", doi: 10.5772/18535.

- **NMR spectroscopy and MRI applications**

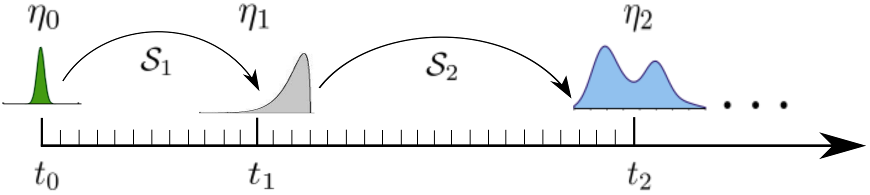


Prior Work on Density Based Modeling and Control

- **Covariance control:** R.E. Skelton *et. al.* (1985 - mid 1990s)
- **Asymptotic density control:** H. Wang *et. al.* (1999, 2001, 2005); Forbes, Forbes, Guay (2003)
- **Ensemble control:** Li and Khaneja (2007, 2009)

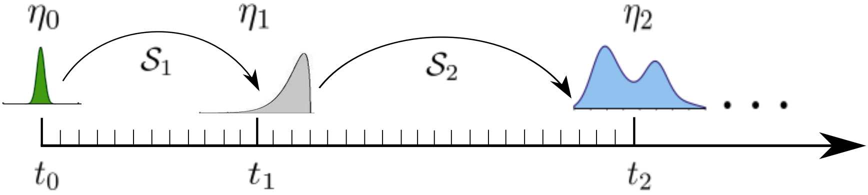
Our contribution: Optimal mass transport framework for finite time feedback density control, data-driven modeling and refinement

Template Problem Statement



- **Given:** a sequence of observed joint densities η_j for output vectors $y_j \in \mathbb{R}^d$ at times t_j, $j = 0, 1, \ldots, M$.
- **Find:** dynamical systems $\mathcal{S}_{j+1} : y_j \mapsto y_{j+1}$, over each horizon $[t_j, t_{j+1})$.

Template Problem Statement



- **Given:** a sequence of observed joint densities η_j for output vectors $y_j \in \mathbb{R}^d$ at times t_j, $j = 0, 1, \ldots, M$.
- **Find:** dynamical systems $\mathcal{S}_{j+1} : y_j \mapsto y_{j+1}$, over each horizon $[t_j, t_{j+1})$.
- **Minimum effort constraint:** $\mathcal{S}_{j+1}$ minimizes total transportation cost $\int_{\mathbb{R}^{2d}} \| y_{j+1} - y_j \|^2_{\ell_2(\mathbb{R}^d)} \rho(y_j, y_{j+1}) \, dy_j dy_{j+1}$ over all transportation policy $\rho(y_j, y_{j+1})$, such that $y_j \sim \eta_j$ and $y_{j+1} \sim \eta_{j+1}$.

Three Variants of the Template Problem

- **Finite horizon feedback control of densities:**

  Find state or output feedback $u(\cdot)$ with *pre-specified control structure* on $\mathcal{S}_{j+1}$

- **Data driven d^{th} order modeling:**

  Find $\mathcal{S}_{j+1}$ with no *a priori* knowledge

- **Density based model refinement:**

  Think of "source density" η_j as *nominal model prediction*, and "target density" η_{j+1} as *true observation*, at the same physical time. Find refined model from the nominal/baseline model.

Template Problem $\rightsquigarrow$ Optimal Mass Transport

- **Gaspard Monge (1781), Leonid Kantorovich (1942)**: move a pile of soil from an excavation to another site through minimum work

- Defines **Wasserstein distance** W, a metric on the space of densities

$$\begin{aligned} W^2 &= \text{optimal transport cost} \\ &= \inf_{\varrho \in \mathcal{P}_2(\rho, \widehat{\rho})} \int_{\mathbb{R}^{2d}} \parallel y - \widehat{y} \parallel^2_{\ell_2(\mathbb{R}^d)} \varrho\left(y, \widehat{y}\right) \, dy d\widehat{y} \\ &= \inf_{\varrho \in \mathcal{P}_2(\rho, \widehat{\rho})} \underbrace{\mathbb{E}\left[\parallel y - \widehat{y} \parallel^2_{\ell_2(\mathbb{R}^d)}\right]}_{J_1(\varrho)} \end{aligned}$$

- Equivalently, $W^2 = \inf_{\beta(\cdot)} \underbrace{\int_{\widehat{\mathcal{Y}}} \parallel \beta\left(\widehat{y}\right) - \widehat{y} \parallel^2_{\ell_2(\mathbb{R}^d)} \; \widehat{\rho}\left(\widehat{y}\right) d\widehat{y}}_{J_2(\beta)}$, subject to

$c\left(\beta\right) = \left|\det\left(\nabla\beta\right)\right| \rho \circ \beta\left(\widehat{y}\right) - \widehat{\rho}\left(\widehat{y}\right) = 0.$

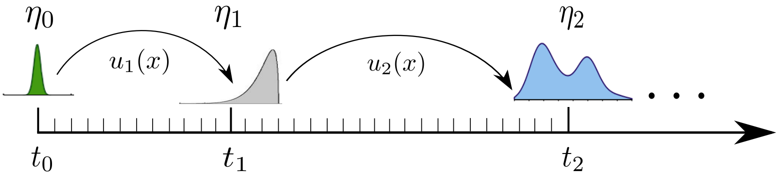
Optimal Mass Transport Background

- **Brenier (1991):** optimal $\beta^\star(\cdot)$ exists and is unique. Further, $\beta^\star(\cdot) = \nabla\psi$. Here $\psi : \mathbb{R}^d \mapsto \mathbb{R}$, and is convex.

- **Benamou & Brenier (2001):** Consider the space-time variational formulation $T \inf\limits_{(\varphi,v)} \underbrace{\int_{\mathbb{R}^d}\int_0^T \varphi(\widehat{y},s) \parallel v(\widehat{y},s) \parallel^2_{\ell_2(\mathbb{R}^d)} \, d\widehat{y}ds}_{J_3(\varphi,v)}$ subject to $\dfrac{\partial\varphi}{\partial s} + \nabla\cdot(\varphi v) = 0$, $\varphi(\cdot,0) = \widehat{\eta}$, $\varphi(\cdot,T) = \eta$. Then $J_3^\star = W^2$ and $v^\star$ is gradient flow.

- $W^2 = \underbrace{\inf\limits_{\varrho\in\mathcal{P}_2(\rho,\widehat{\rho})} J_1(\varrho)}_{\text{infinite dimensional LP}} = \underbrace{\inf\limits_{\beta:c(\beta)=0} J_2(\beta)}_{\text{Nonlinear nonconvex optimization}} = \underbrace{T \inf\limits_{(\varphi,v)} J_3(\varphi,v)}_{\text{Nonsmooth convex optimization}}$

Finite Horizon Feedback Control of Densities



- Theorem: Consider tracking Gaussians $\eta_j = \mathcal{N}(\mu_j, \Sigma_j)$, under LTI structure $x_{j+1} = Ax_j + Bu_j$. Let

$$\theta_j = \mu_{j+1} - \mu_j, \quad \Theta_j = \Sigma_{j+1}^{\frac{1}{2}} \left(\Sigma_{j+1}^{\frac{1}{2}} \Sigma_j \Sigma_{j+1}^{\frac{1}{2}} \right)^{-\frac{1}{2}} \Sigma_{j+1}^{\frac{1}{2}}.$$

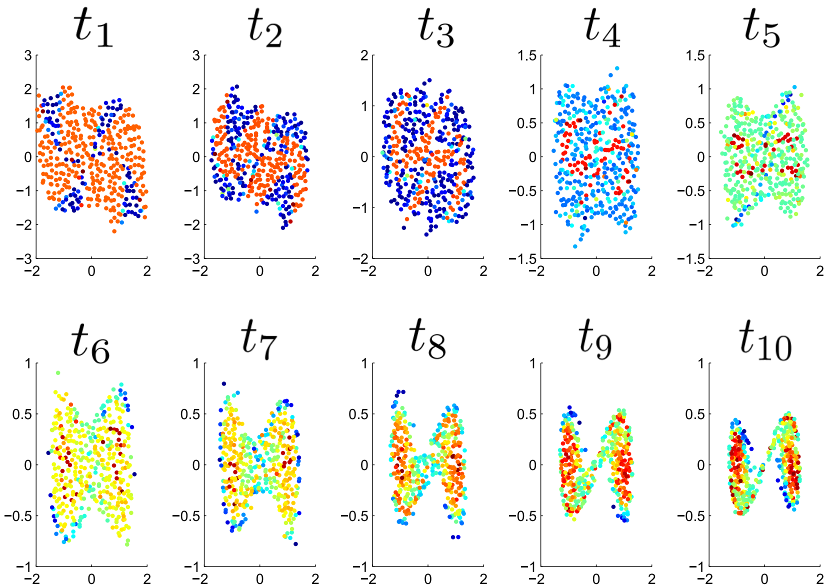
The state feedback $u_j^\star \triangleq u^\star(x_j)$ guaranteeing optimal transport

1. exists iff $(\Theta_j - A), \theta_j \in \ker\left(I - BB^\dagger\right)$
2. if exists, then must be affine form $u_j^\star = K_j x_j + \kappa_j$, where $K_j = B^\dagger(\Theta_j - A) - \left(I - BB^\dagger\right)R$, and $\kappa_j = B^\dagger \theta_j - \left(I - BB^\dagger\right) r$
3. is unique, if B is full rank.

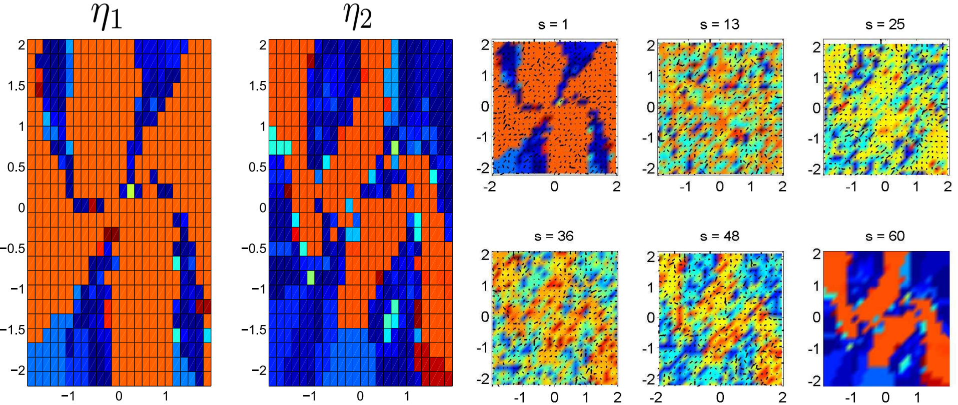
Data Driven d^{th} Order Modeling

- Duffing vector field (unknown to modeler) to generate data: $\dot{x}_1 = x_2, \quad \dot{x}_2 = -\alpha x_1^3 - \beta x_1 - \delta x_2$, $y = \{x_1, x_2\}^\top$, $\alpha = 1$, $\beta = -1$, $\delta = 0.5$
- Density propagation with 500 samples from initial density $\xi_0 = \mathcal{U}\left([-2, 2]^2\right)$
- 10 snapshot data $\{t_j, \eta_j\}_{j=1}^{10}$
- Subdivided each of the 10 intervals $[t_j, t_{j+1})$, $j = 0, \ldots, 9$ into 60 sub-intervals.
- Want to compute optimal transport vector field $v_{j \to j+1}$ for each of those intervals

Data Driven d^{th} Order Modeling

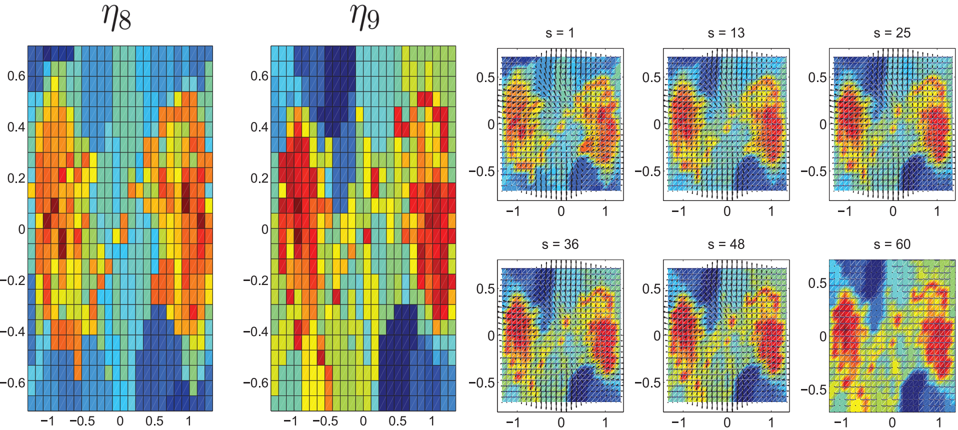


Data Driven d^{th} Order Modeling of $v_{1\to 2}$



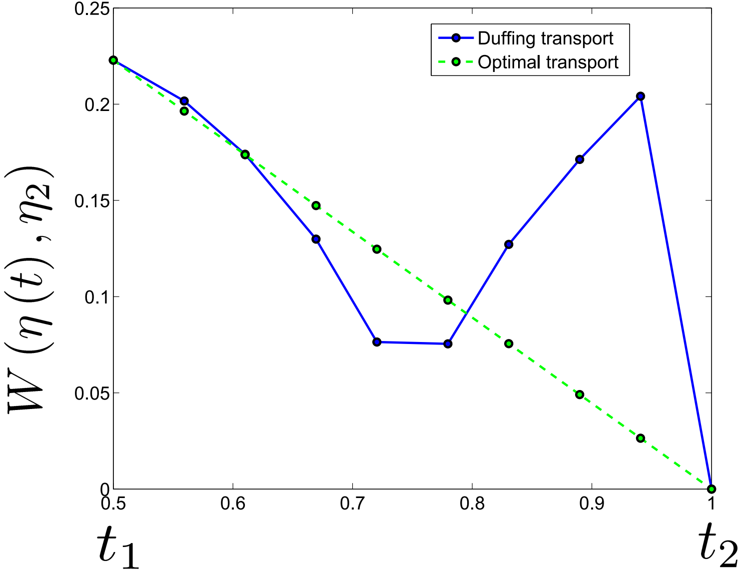
(a) (b)

Data Driven d^{th} Order Modeling of $v_{8\to 9}$

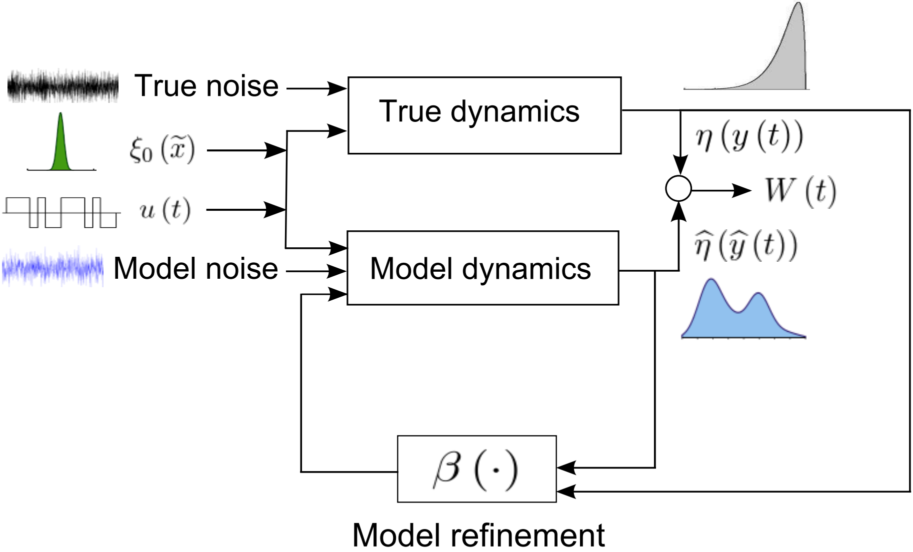


(a) (b)

Data Driven d^{th} Order Modeling: Duffing transport vs. Optimal Transport for $[t_1, t_2)$



Density Based Model Refinement



Density Based Model Refinement: Formulation

- Startegy: Only refine the output model (why?)
- For example, consider proposed model $\dot{\widehat{x}} = \widehat{f}\left(\widehat{x}\right)$, $\widehat{y} = \widehat{h}\left(\widehat{x}\right)$
- Call $\widehat{y_j^-} \triangleq \widehat{y}\left(t_j\right)$. We know η_j and $\widehat{\eta}_j$.
- We seek $\beta_j : \mathbb{R}^{n_o} \mapsto \mathbb{R}^{n_o}$, so that $\widehat{y_j^+} = \beta_j\left(\widehat{y_j^-}\right)$ satisying $\widehat{y_j^+} \sim \eta_j$ and $\widehat{y_j^-} \sim \widehat{\eta}_j$
- Then the refined model is: $\dot{\widehat{x}} = \widehat{f}\left(\widehat{x}\right)$, $\widehat{y_j^+} = \beta_j \circ \widehat{h}\left(\widehat{x}\right)$
- Seek optimal push-forward: $\displaystyle\inf_{\beta(\cdot)} \underbrace{\int_{\widehat{\mathcal{Y}}} \parallel \beta_j\left(\widehat{y_j^-}\right) - \widehat{y_j^-} \parallel^2_{\ell_2(\mathbb{R}^{n_o})} \; \widehat{\eta}_j d\widehat{y_j^-}}_{J_2(\beta)}$,

  subject to $\eta_j = \beta_j \sharp \widehat{\eta}_j$.

Linear Gaussian Model Refinement

- Theorem: Consider discrete-time deterministic LTI pairs: (A, C), $\left(\widehat{A}, \widehat{C}\right)$, sarting with $\xi_0 = \mathcal{N}(\mu_0, \Sigma_0)$. Then refined model is:
  $\widehat{x}_{j+1} = \widehat{A}\widehat{x}_j$, $\widehat{y}_j^+ = \Theta_j \widehat{C}\widehat{x}_j + \theta_j$.
  $\Theta_j = \Sigma_j^{1/2}\left(\Sigma_j^{1/2}\widehat{\Sigma}_j\Sigma_j^{1/2}\right)^{-1/2}\Sigma_j^{1/2}$, and $\theta_j = \mu_j - \widehat{\mu}_j$.

  The s^{th} synthetic time PDF at j^{th} physical time is:
  $\mathcal{N}\left(\mu_{\widehat{y}\to y}(s), \Sigma_{\widehat{y}\to y}(s)\right)$, where

$$\mu_{\widehat{y}\to y}(s) = \left[(1-s)\ \widehat{C}\widehat{A}^j + s\, CA^j\right]\mu_0,$$

$$\Sigma_{\widehat{y}\to y}(s) = \left[(1-s)\ I + s\,\Theta(j)\right]\left(\left(\widehat{C}\widehat{A}^j\right)\Sigma_0\left(\widehat{C}\widehat{A}^j\right)^\top\right)\left[(1-s)\ I + s\,\Theta(j)\right].$$

Linear Gaussian Model Refinement

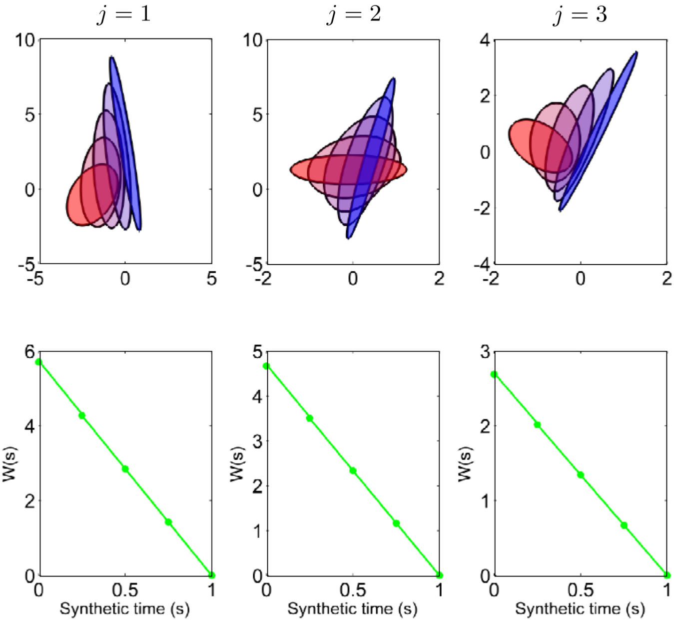
- Theorem: Consider discrete-time deterministic LTI pairs: (A, C), $\left(\widehat{A}, \widehat{C}\right)$, sarting with $\xi_0 = \mathcal{N}(\mu_0, \Sigma_0)$. Then refined model is:
  $\widehat{x}_{j+1} = \widehat{A}\widehat{x}_j$, $\widehat{y}_j^+ = \Theta_j \widehat{C}\widehat{x}_j + \theta_j$.
  $\Theta_j = \Sigma_j^{1/2}\left(\Sigma_j^{1/2}\widehat{\Sigma}_j\Sigma_j^{1/2}\right)^{-1/2}\Sigma_j^{1/2}$, and $\theta_j = \mu_j - \widehat{\mu}_j$.

  The s^{th} synthetic time PDF at j^{th} physical time is:
  $\mathcal{N}\left(\mu_{\widehat{y}\to y}(s), \Sigma_{\widehat{y}\to y}(s)\right)$, where

$$\begin{aligned}
\mu_{\widehat{y}\to y}(s) &= \left[(1-s)\ \widehat{C}\widehat{A}^j + s\, C A^j\right]\mu_0, \\
\Sigma_{\widehat{y}\to y}(s) &= \left[(1-s)\ I + s\,\Theta(j)\right]\left(\left(\widehat{C}\widehat{A}^j\right)\Sigma_0\left(\widehat{C}\widehat{A}^j\right)^\top\right)\left[(1-s)\ I + s\,\Theta(j)\right].
\end{aligned}$$

- Example: $A = \begin{bmatrix} 0.4 & -0.1 \\ 2 & 0.6 \end{bmatrix}$, $\widehat{A} = \begin{bmatrix} 0.2 & -0.7 \\ -0.7 & 0.1 \end{bmatrix}$,
  $C = \begin{bmatrix} -1 & 0.03 \\ -0.2 & 0.8 \end{bmatrix}$, $\widehat{C} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. $\mu_0 = \{1,\ 3\}^\top, \Sigma_0 = \begin{bmatrix} 10 & 6 \\ 6 & 7 \end{bmatrix}$

Linear Gaussian Model Refinement: Example



Summary

- Optimal transport framework for systems with density level observation
- Closed form feedback control for minimum effort linear Gaussian tracking
- Convex optimization framework for data driven modeling
- Closed form solution for linear Gaussian model refinement

Summary

- Optimal transport framework for systems with density level observation
- Closed form feedback control for minimum effort linear Gaussian tracking
- Convex optimization framework for data driven modeling
- Closed form solution for linear Gaussian model refinement

- Funding support: NSF CSR Award # 1016299

Thank you.