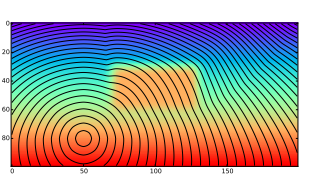
Adaptive construction of measure transport with application to Bayesian inference

D. Bigoni[†] (**dabi@mit.edu**), A. Spantini[†], Y.M. Marzouk[†]
[†]Massachusetts Institute of Technology

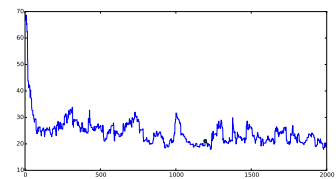
Milan – 16/06/2016

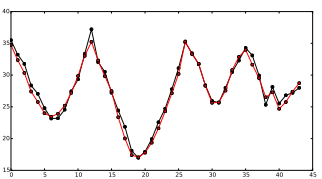
Inference – The Bayesian approach – Seismic tomography

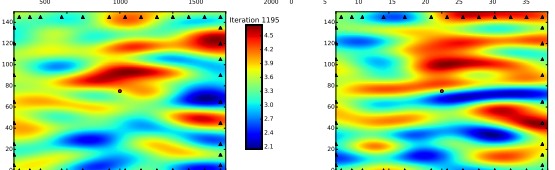
$$\Longrightarrow |\nabla u(\mathbf{x})| = \frac{1}{v(\mathbf{x})} \Longrightarrow$$



$$p(\theta|d) \propto \overbrace{p(d|\theta)}^{\text{Likelihood}} \; p(\theta)$$







"Knowing the arrival time and the hypocenter of a number of earthquakes, what is the probability distribution of the velocity model of the crust?"

The earthquake image is licensed under Public Domain via Wikimedia Commons. Work in collaboration with A. Zunino

Intractable distributions

1. We have **access to densities** but **sampling is challenging**.
   - Classical Bayesian inference problems [El Moselhy et al., 2012]

$$p(\theta|d) \propto p(d|\theta)p(\theta)$$

   - Filtering problems [see Spantini's presentation MS.42]

$$\begin{aligned} X_{t+1} &= \mathcal{F}(t, X_t, w_t) \qquad w_t \sim \mu \\ Y_t &= \mathcal{G}(t, X_t, \gamma_t) \qquad \gamma_t \sim \pi \end{aligned}$$

2. We have **access to samples** and we want to **characterize their density**. [Parno et al., 2015]

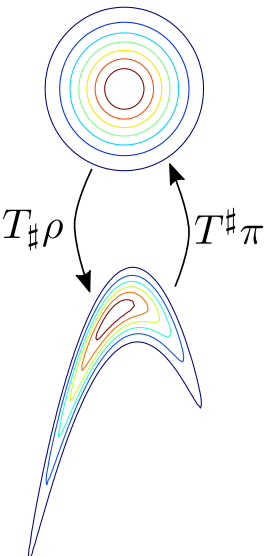
Measure transport – Pullbacks [PB] and Pushforwards [PF]

- $\rho : \mathbb{R}^d \to \mathbb{R}$ density of the distribution λ_1
- $\pi : \mathbb{R}^d \to \mathbb{R}$ density of the distribution λ_2
- Let $T : \mathbb{R}^d \to \mathbb{R}^d$ be the map such that either

PF $\quad \pi = T_\sharp \rho = \rho \circ T^{-1} |\nabla T^{-1}|$
PB $\quad \rho = T^\sharp \pi = \pi \circ T |\nabla T|$

This means T is such that for $X \sim \lambda_1$, either

PF $\quad T(X) = Y \sim \lambda_2$
PB $\quad T^{-1}(Y) = X \sim \lambda_1$



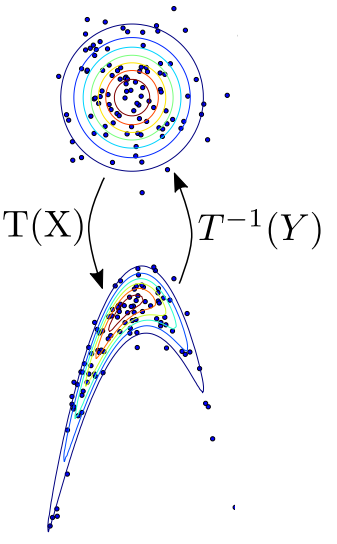
Measure transport – Pullbacks [PB] and Pushforwards [PF]

- $\rho : \mathbb{R}^d \to \mathbb{R}$ density of the distribution λ_1
- $\pi : \mathbb{R}^d \to \mathbb{R}$ density of the distribution λ_2
- Let $T : \mathbb{R}^d \to \mathbb{R}^d$ be the map such that either

PF $\quad \pi = T_\sharp \rho = \rho \circ T^{-1} |\nabla T^{-1}|$
PB $\quad \rho = T^\sharp \pi = \pi \circ T |\nabla T|$

This means T is such that for $X \sim \lambda_1$, either

PF $\quad T(X) = Y \sim \lambda_2$
PB $\quad T^{-1}(Y) = X \sim \lambda_1$



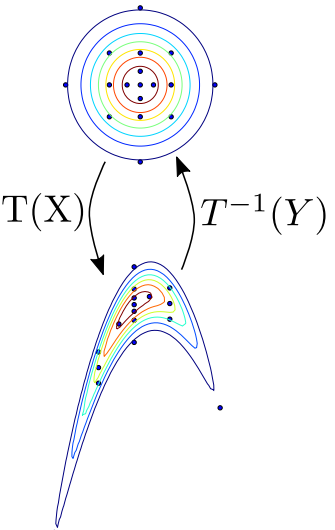
Measure transport – Pullbacks [PB] and Pushforwards [PF]

- $\rho : \mathbb{R}^d \to \mathbb{R}$ density of the distribution λ_1
- $\pi : \mathbb{R}^d \to \mathbb{R}$ density of the distribution λ_2
- Let $T : \mathbb{R}^d \to \mathbb{R}^d$ be the map such that either

PF $\quad \pi = T_\sharp \rho = \rho \circ T^{-1} |\nabla T^{-1}|$
PB $\quad \rho = T^\sharp \pi = \pi \circ T |\nabla T|$

This means T is such that for $X \sim \lambda_1$, either

PF $\quad T(X) = Y \sim \lambda_2$
PB $\quad T^{-1}(Y) = X \sim \lambda_1$



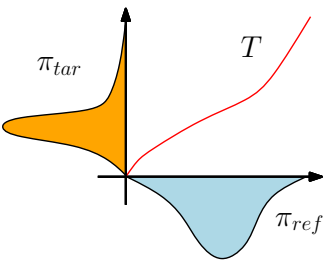
Measure transport – Pullbacks [PB] and Pushforwards [PF]

- $\rho : \mathbb{R}^d \to \mathbb{R}$ density of the distribution λ_1
- $\pi : \mathbb{R}^d \to \mathbb{R}$ density of the distribution λ_2
- Let $T : \mathbb{R}^d \to \mathbb{R}^d$ be the map such that either

PF $\quad \pi = T_\sharp \rho = \rho \circ T^{-1} |\nabla T^{-1}|$
PB $\quad \rho = T^\sharp \pi = \pi \circ T |\nabla T|$

This means T is such that for $X \sim \lambda_1$, either

PF $\quad T(X) = Y \sim \lambda_2$
PB $\quad T^{-1}(Y) = X \sim \lambda_1$



Note: T must preserve mass!

Knothe-Rosenblatt rearrangement

$\forall \lambda_1, \lambda_2$ absolutely continuous there exists a triangular monotone map s.t. $T(d\lambda_1) = d\lambda_2$

$$T(\mathbf{x}) = \begin{bmatrix} T^{(1)}(x_1) \\ T^{(2)}(x_1, x_2) \\ \vdots \\ T^{(d)}(x_1, \ldots, x_d) \end{bmatrix}$$

Measure transport

Practical setting

ρ is the density of $X \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ or some amenable distribution [**reference**]

π is the density of an intractable distribution [**target**]

We seek a map T^* such that $\quad T^*_\sharp \rho \approx \pi$.

Measure transport

Practical setting

ρ is the density of $X \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ or some amenable distribution [**reference**]

π is the density of an intractable distribution [**target**]

We seek a map T^* such that $\quad T^*_\sharp \rho \approx \pi$.

Minimization statement [El Moselhy et al., 2012]

For $\mathcal{H}^\Delta$ the space of lower-triangular maps, solve

$$\begin{aligned} T^* &= \arg\min_{T \in \mathcal{H}^\Delta} D_{\mathrm{KL}}(T_\sharp \rho \| \pi) = \arg\min_{T \in \mathcal{H}^\Delta} D_{\mathrm{KL}}(\rho \| T^\sharp \pi) \\ &= \arg\min_{T \in \mathcal{H}^\Delta} \mathbb{E}_\rho[\log \rho] - \mathbb{E}_\rho \left[\log \pi \circ T + \log |\nabla T|\right] \end{aligned}$$

such that $\partial_{x_i} T^{(i)} > 0$.

Measure transport

Practical setting

ρ is the density of $X \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ or some amenable distribution [**reference**]
π is the density of an intractable distribution [**target**]
We seek a map T^* such that $\quad T^*_\sharp \rho \approx \pi$.

Minimization statement [El Moselhy et al., 2012]

For $\mathcal{H}^\Delta$ the space of lower-triangular maps, solve

$$\begin{aligned} T^* &= \arg\min_{T \in \mathcal{H}^\Delta} D_{\mathrm{KL}}(T_\sharp \rho \| \pi) = \arg\min_{T \in \mathcal{H}^\Delta} D_{\mathrm{KL}}(\rho \| T^\sharp \pi) \\ &= \arg\min_{T \in \mathcal{H}^\Delta} \mathbb{E}_\rho[\log \rho] - \mathbb{E}_\rho \left[\log \pi \circ T + \log |\nabla T|\right] \end{aligned}$$

such that $\partial_{x_i} T^{(i)} > 0$.

1. We can use derivative based optimization if $\nabla_{\mathbf{x}} \log \pi$ or $\nabla^2_{\mathbf{x}} \log \pi$ are provided
2. We can explore π in parallel
3. We can generate i.i.d. samples from $T^*_\sharp \rho \approx \pi$ in parallel
4. We can estimate convergence!

Measure transport

Practical setting

ρ is the density of $X \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ or some amenable distribution [**reference**]
π is the density of an intractable distribution [**target**]
We seek a map T^* such that $\quad T^*_\sharp \rho \approx \pi$.

Minimization statement [El Moselhy et al., 2012]
For $\mathcal{H}^\Delta$ the space of lower-triangular maps, solve

$$T^* = \arg\min_{T \in \mathcal{H}^\Delta} D_{\mathrm{KL}}(T_\sharp \rho \| \pi) = \arg\min_{T \in \mathcal{H}^\Delta} D_{\mathrm{KL}}(\rho \| T^\sharp \pi)$$
$$= \arg\min_{T \in \mathcal{H}^\Delta} \mathbb{E}_\rho[\log \rho] - \mathbb{E}_\rho \left[\log \pi \circ T + \log |\nabla T|\right]$$

such that $\partial_{x_i} T^{(i)} > 0$.
Note: Approximation of d functions up to d-dimensional!

Source of low-dimensional structure

- Adaptivity
- Conditional independency
- Marginal independency
- Low-rank structure

Transport maps – Adaptivity

1. **Convergence criteria**
2. **Refinement criteria**

Transport maps – Adaptivity

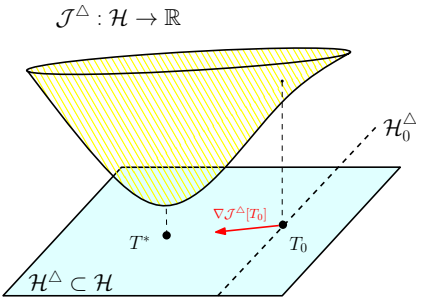
1. **Convergence criteria**
   At optimality $\Rightarrow D_{\mathrm{KL}}(T^*_\sharp \rho \| \pi) = 0$
   $\mathbb{V}\left[\log \frac{\rho}{T^{*\sharp}\pi}\right] \rightarrow \frac{1}{2} D_{\mathrm{KL}}(T^*_\sharp \rho \| \pi) \rightarrow 0$
2. **Refinement criteria**

Transport maps – Adaptivity

1. **Convergence criteria – Variance diagnostic** – $\mathbb{V}\left[\log \frac{\rho}{T^\sharp \pi}\right]$
2. **Refinement criteria**

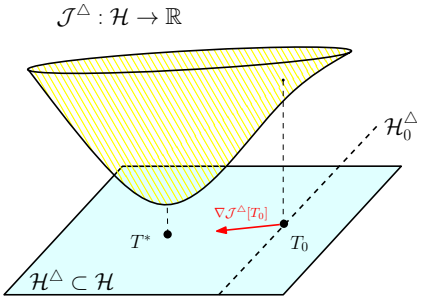
Transport maps – Adaptivity – Refinement criteria

- Minimize $\mathcal{J}^{\Delta}(T) = D_{KL}(T_\sharp \rho \| \pi)$, over $\mathcal{H}^{\Delta}$ $(\dim \mathcal{H}^{\Delta} = \infty)$



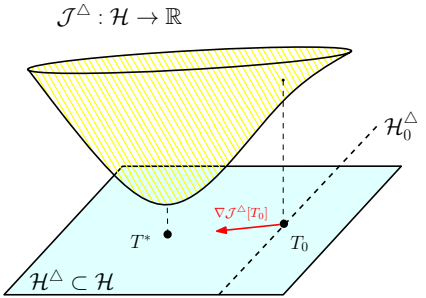
Transport maps – Adaptivity – Refinement criteria

- Minimize $\mathcal{J}^{\Delta}(T) = D_{KL}(T_\sharp \rho \| \pi)$, over $\mathcal{H}^{\Delta}$ $(\dim \mathcal{H}^{\Delta} = \infty)$
- For $\mathcal{H}_0^{\Delta} \subset \mathcal{H}^{\Delta}$ $(\dim \mathcal{H}_0^{\Delta} = n_0)$ $T_0^* = \arg\min_{T \in \mathcal{H}_0^{\Delta}} \mathcal{J}^{\Delta}(T)$



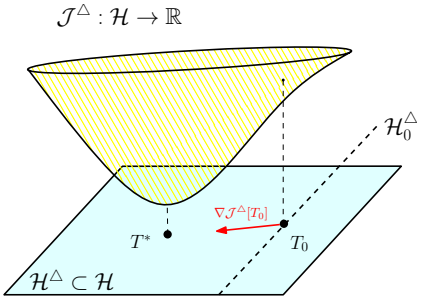
Transport maps – Adaptivity – Refinement criteria

- Minimize $\mathcal{J}^{\Delta}(T) = D_{KL}(T_\sharp \rho \| \pi)$, over $\mathcal{H}^{\Delta}$ $(\dim \mathcal{H}^{\Delta} = \infty)$
- For $\mathcal{H}_0^{\Delta} \subset \mathcal{H}^{\Delta}$ $(\dim \mathcal{H}_0^{\Delta} = n_0)$ $T_0^* = \arg\min_{T \in \mathcal{H}_0^{\Delta}} \mathcal{J}^{\Delta}(T)$
- Equivalent to: $\mathbf{a}_0^* = \arg\min_{\mathbf{a} \in \mathbb{R}^{n_0}} \mathcal{J}^{\Delta}(T[\mathbf{a}])$



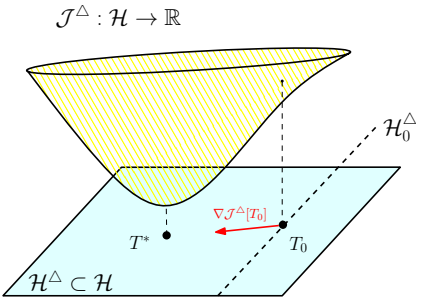
Transport maps – Adaptivity – Refinement criteria

- Minimize $\mathcal{J}^{\triangle}(T) = D_{KL}(T_\sharp \rho \| \pi)$, over $\mathcal{H}^{\triangle}$ $(\dim \mathcal{H}^{\triangle} = \infty)$
- For $\mathcal{H}_0^{\triangle} \subset \mathcal{H}^{\triangle}$ $(\dim \mathcal{H}_0^{\triangle} = n_0)$ $T_0^* = \arg\min_{T \in \mathcal{H}_0^{\triangle}} \mathcal{J}^{\triangle}(T)$
- Equivalent to: $\mathbf{a}_0^* = \arg\min_{\mathbf{a} \in \mathbb{R}^{n_0}} \mathcal{J}^{\triangle}(T[\mathbf{a}])$
- At optimality: $\nabla_{\mathbf{a}} \mathcal{J}^{\triangle}(T[\mathbf{a}_0^*]) = 0$

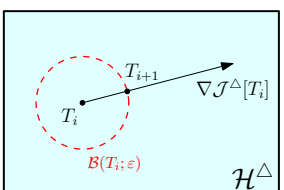


Transport maps – Adaptivity – Refinement criteria

- Minimize $\mathcal{J}^{\triangle}(T) = D_{KL}(T_\sharp \rho \| \pi)$, over $\mathcal{H}^{\triangle}$ $(\dim \mathcal{H}^{\triangle} = \infty)$
- For $\mathcal{H}_0^{\triangle} \subset \mathcal{H}^{\triangle}$ $(\dim \mathcal{H}_0^{\triangle} = n_0)$ $T_0^* = \arg\min_{T \in \mathcal{H}_0^{\triangle}} \mathcal{J}^{\triangle}(T)$
- Equivalent to: $\mathbf{a}_0^* = \arg\min_{\mathbf{a} \in \mathbb{R}^{n_0}} \mathcal{J}^{\triangle}(T[\mathbf{a}])$
- At optimality: $\nabla_{\mathbf{a}} \mathcal{J}^{\triangle}(T[\mathbf{a}_0^*]) = 0$



$\mathcal{J}^{\triangle}(T_{i+1}) < \mathcal{J}^{\triangle}(T_i)$



- The first variation $\nabla \mathcal{J}^{\triangle}\left(T[\mathbf{a}_0^*]\right) \neq 0$, which means

There exists $\varepsilon > 0$ such that

$$\mathcal{J}^{\triangle}\left(T[\mathbf{a}_0^*] - \varepsilon \nabla \mathcal{J}^{\triangle}\left(T[\mathbf{a}_0^*]\right)\right) < \mathcal{J}^{\triangle}\left(T[\mathbf{a}_0^*]\right)$$

Transport maps – Adaptivity – Refinement criteria

- The first variation $\nabla \mathcal{J}^{\Delta}\left(T[\mathbf{a}_0^*]\right) : \mathbb{R}^d \to \mathbb{R}^d$ is a map!

$$\nabla \mathcal{J}^{\Delta}\left(T[\mathbf{a}_0^*]\right) = \left(\nabla_{\mathbf{x}} T\right)^{-1} \left(\nabla_{\mathbf{x}} \log \frac{\rho}{T[\mathbf{a}_0^*]^{\sharp}\pi} \right)$$

Transport maps – Adaptivity – Refinement criteria

- The first variation $\nabla \mathcal{J}^{\Delta}\left(T[\mathbf{a}_0^*]\right) : \mathbb{R}^d \to \mathbb{R}^d$ is a map!

$$\nabla \mathcal{J}^{\Delta}\left(T[\mathbf{a}_0^*]\right) = \left(\nabla_{\mathbf{x}} T\right)^{-1} \left(\nabla_{\mathbf{x}} \log \frac{\rho}{T[\mathbf{a}_0^*]^{\sharp}\pi} \right)$$

- No new evaluation of $\nabla_{\mathbf{x}} \log \pi$ is required at the sample points $\{\mathbf{x}_k\}$

Transport maps – Adaptivity – Refinement criteria

- The first variation $\nabla \mathcal{J}^{\Delta}\left(T[\mathbf{a}_0^*]\right) : \mathbb{R}^d \to \mathbb{R}^d$ is a map!

$$\nabla \mathcal{J}^{\Delta}\left(T[\mathbf{a}_0^*]\right) = \left(\nabla_{\mathbf{x}} T\right)^{-1} \left(\nabla_{\mathbf{x}} \log \frac{\rho}{T[\mathbf{a}_0^*]^{\sharp}\pi} \right)$$

- No new evaluation of $\nabla_{\mathbf{x}} \log \pi$ is required at the sample points $\{\mathbf{x}_k\}$
- For $\mathcal{H}^{\Delta} \supset \mathcal{H}_1^{\Delta} \supset \mathcal{H}_0^{\Delta}$ $\left(\dim \mathcal{H}_1^{\Delta} = n_1\right)$, solve

$$\mathbf{b}_1^* = \underset{\mathbf{b} \in \mathbb{R}^{n_1}}{\arg\min} \left\| U[\mathbf{b}] - \nabla \mathcal{J}^{\Delta}\left(T[\mathbf{a}_0^*]\right) \right\|_{L^2_\rho}$$

Transport maps – Adaptivity – Refinement criteria

- The first variation $\nabla \mathcal{J}^{\Delta}\left(T[\mathbf{a}_0^*]\right) : \mathbb{R}^d \to \mathbb{R}^d$ is a map!

$$\nabla \mathcal{J}^{\Delta}\left(T[\mathbf{a}_0^*]\right) = \left(\nabla_{\mathbf{x}} T\right)^{-1} \left(\nabla_{\mathbf{x}} \log \frac{\rho}{T[\mathbf{a}_0^*]^{\sharp}\pi} \right)$$

- No new evaluation of $\nabla_{\mathbf{x}} \log \pi$ is required at the sample points $\{\mathbf{x}_k\}$
- For $\mathcal{H}^{\Delta} \supset \mathcal{H}_1^{\Delta} \supset \mathcal{H}_0^{\Delta}$ $\left(\dim \mathcal{H}_1^{\Delta} = n_1\right)$, solve

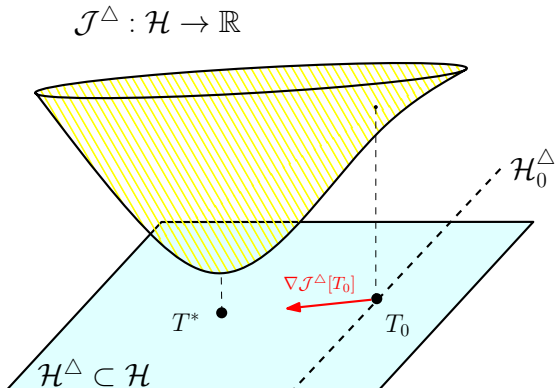
$$\mathbf{b}_1^* = \underset{\mathbf{b} \in \mathbb{R}^{n_1}}{\arg\min} \left\| U[\mathbf{b}] - \nabla \mathcal{J}^{\Delta}\left(T[\mathbf{a}_0^*]\right) \right\|_{L^2_\rho}$$

$U[\mathbf{b}_1^*]$ informs about:

- **active variables** to be included in the components $T^{(i)}$
- **active coefficients** to be included in the parametrization of $T^{(i)}$

Transport maps – Adaptivity

1. **Convergence criteria – Variance diagnostic** – $\mathbb{V}\left[\log \frac{\rho}{T^\sharp \pi}\right]$
2. **Refinement criteria – First variation** – $\nabla \mathcal{J}^\triangle \left(T[\mathbf{a}^*]\right)$



Example – The banana distribution

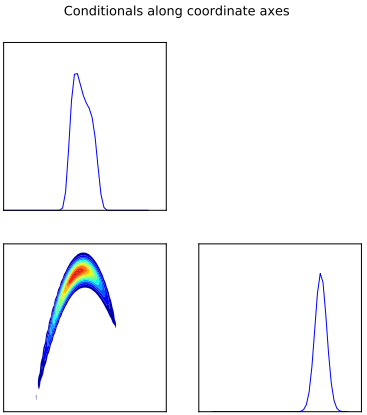
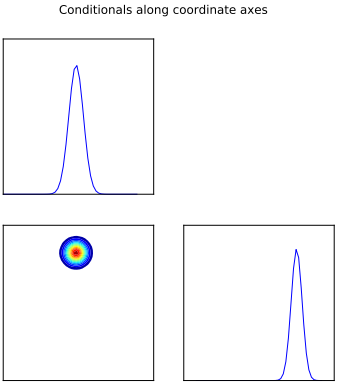
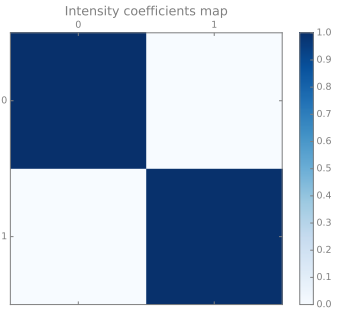


Figure: Target π

Example – The banana distribution

Iteration 1 – Pushforward $T_\sharp \rho$

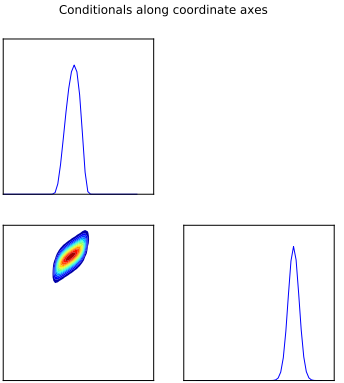


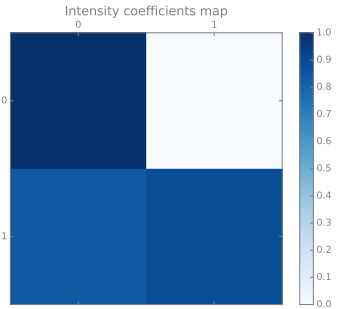


$\nabla_{\mathbf{x}} T$

Example – The banana distribution

Iteration 2 – Pushforward $T_\sharp \rho$

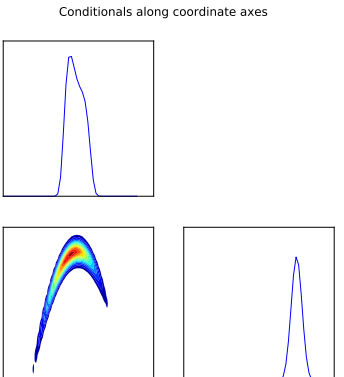


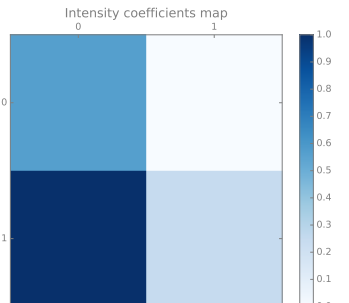


$\nabla_{\mathbf{x}} T$

Example – The banana distribution

Iteration 3 – Pushforward $T_\sharp \rho$

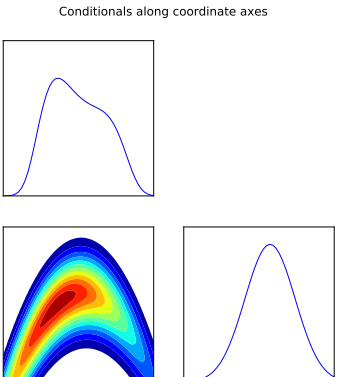


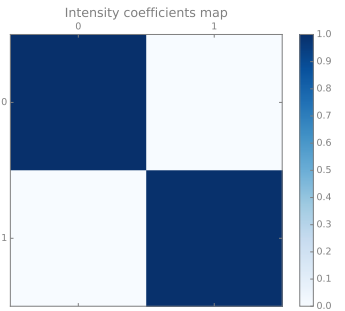


$\nabla_{\mathbf{x}} T$

Example – The banana distribution

Iteration 1 – Pullback $T^\sharp\pi$

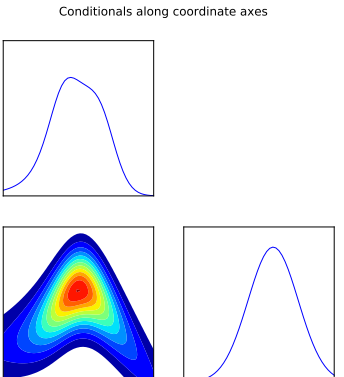


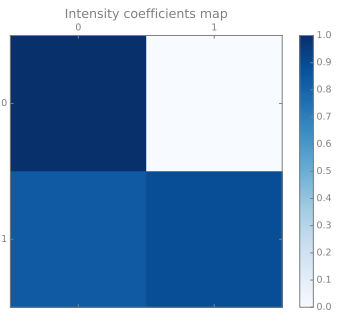


$\nabla_{\mathbf{x}} T$

Example – The banana distribution

Iteration 2 – Pullback $T^\sharp\pi$

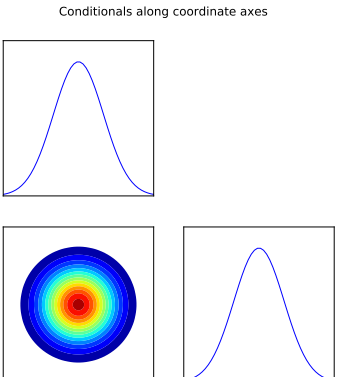


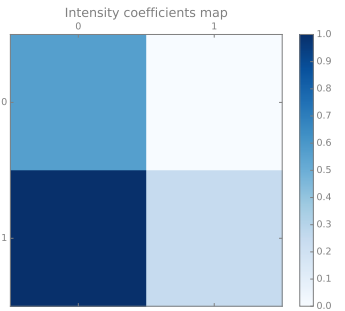


$\nabla_{\mathbf{x}} T$

Example – The banana distribution

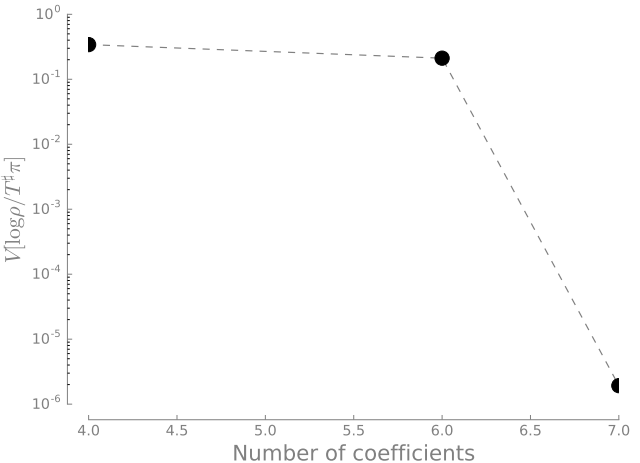
Iteration 3 – Pullback $T^\sharp\pi$





$\nabla_{\mathbf{x}} T$

Example – The banana distribution



Variance diagnostic

Example – Stochastic volatility

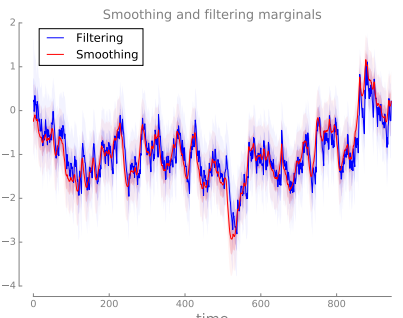
- Latent log-volatilities modeled with an AR(1) process for $t = 1, \ldots, N$ $(N = 30)$

$$X_{t+1} = \mu + \phi(X_t - \mu) + \eta_t \ , \quad \eta_t \sim \mathcal{N}(0,1) \ , \quad X_1 \sim \mathcal{N}\left(0, 1/\left(1-\phi^2\right)\right)$$

- Observe the mean return for holding the asset at time t

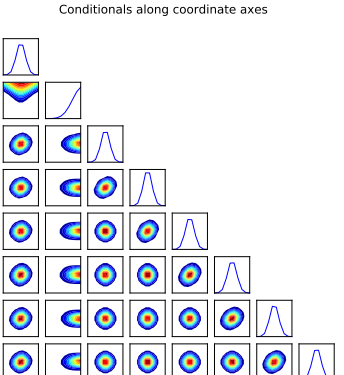
$$Y_t = \varepsilon_t \exp(X_t/2) \ , \quad \varepsilon_t \sim \mathcal{N}(0,1)$$

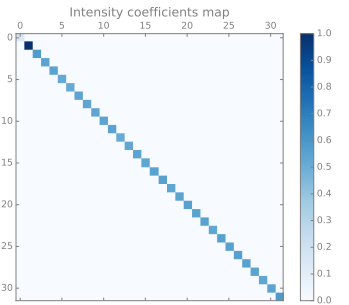
- We want to characterize $\pi \sim \mu, \phi, \mathbf{X}_{1:N} | \mathbf{Y}_{1:N}$



Example – Stochastic volatility

Iteration 1 – Pullback $T^\sharp \pi$



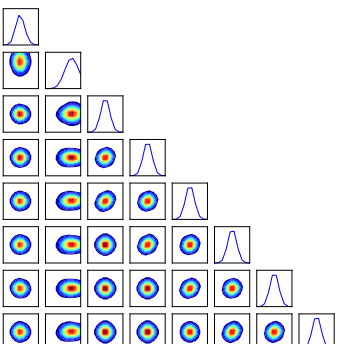


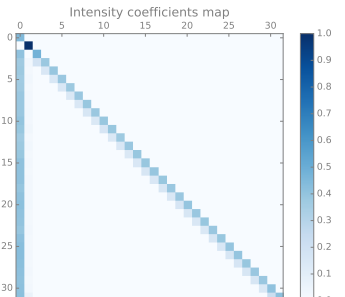
$\nabla_{\mathbf{x}} T$

Example – Stochastic volatility

Iteration 2 – Pullback $T^\sharp\pi$

Conditionals along coordinate axes



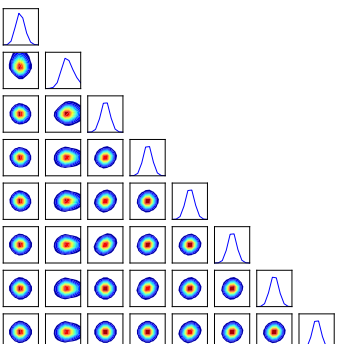


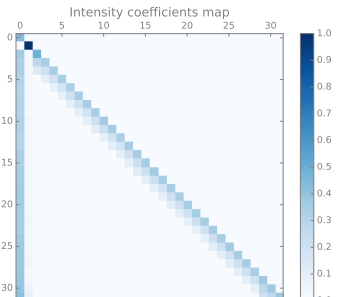
$\nabla_{\mathbf{x}} T$

Example – Stochastic volatility

Iteration 3 – Pullback $T^\sharp\pi$

Conditionals along coordinate axes

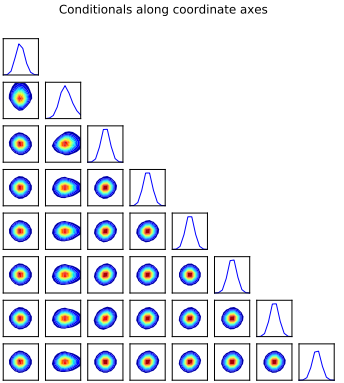


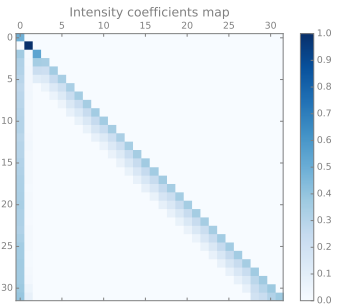


$\nabla_{\mathbf{x}} T$

Example – Stochastic volatility

Iteration 4 – Pullback $T^\sharp\pi$

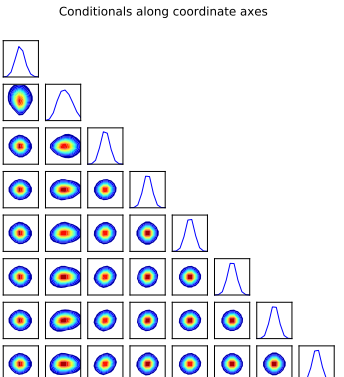


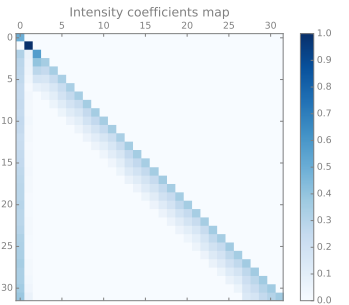


$\nabla_{\mathbf{x}} T$

Example – Stochastic volatility

Iteration 5 – Pullback $T^\sharp \pi$

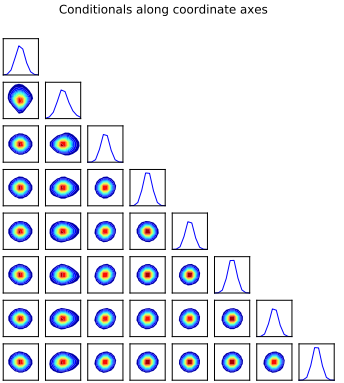


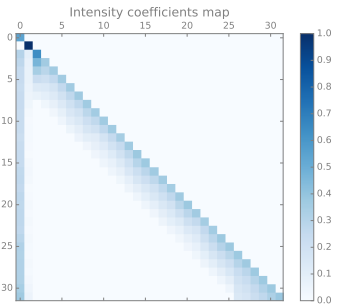


$\nabla_{\mathbf{x}} T$

Example – Stochastic volatility

Iteration 6 – Pullback $T^\sharp\pi$

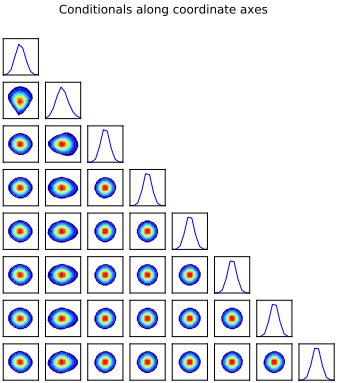


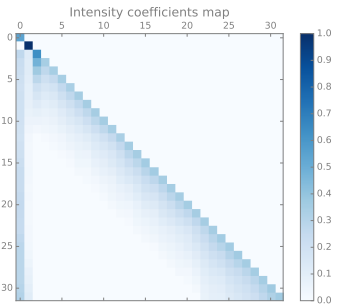


$\nabla_{\mathbf{x}} T$

Example – Stochastic volatility

Iteration 7 – Pullback $T^\sharp\pi$

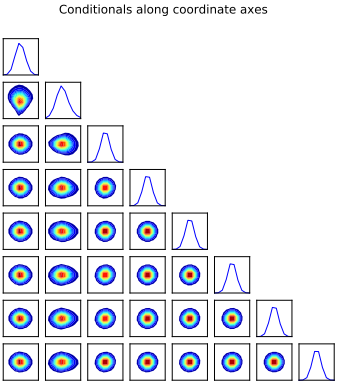


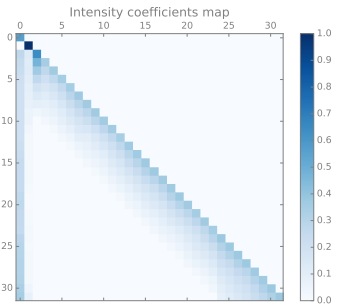


$\nabla_{\mathbf{x}} T$

Example – Stochastic volatility

Iteration 7 – Pullback $T^\sharp\pi$



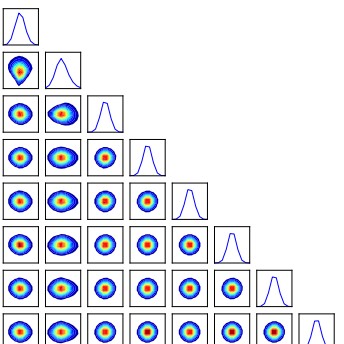


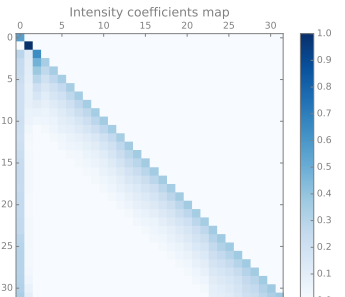
$\nabla_{\mathbf{x}} T$

Example – Stochastic volatility

Iteration 9 – Pullback $T^\sharp \pi$

Conditionals along coordinate axes



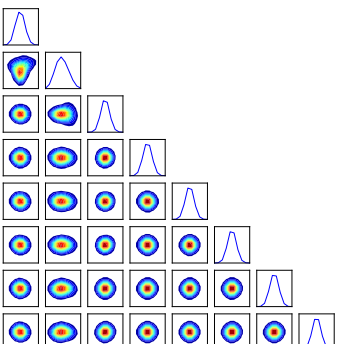


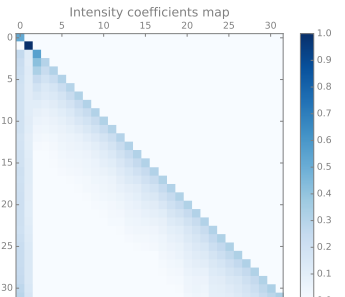
$\nabla_{\mathbf{x}} T$

Example – Stochastic volatility

Iteration 10 – Pullback $T^\sharp \pi$

Conditionals along coordinate axes



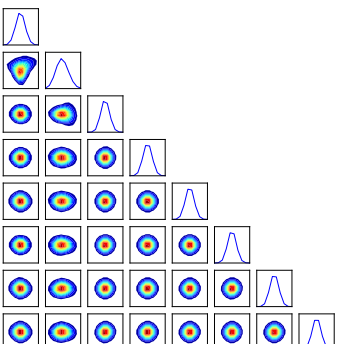


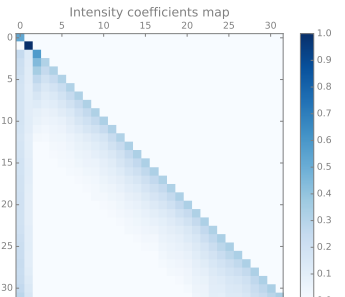
$\nabla_{\mathbf{x}} T$

Example – Stochastic volatility

Iteration 11 – Pullback $T^\sharp \pi$

Conditionals along coordinate axes

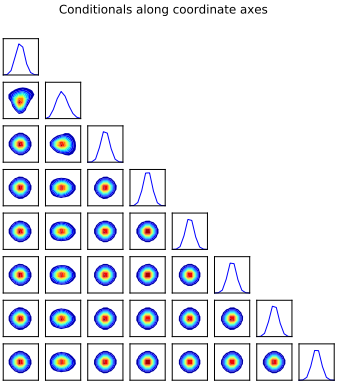


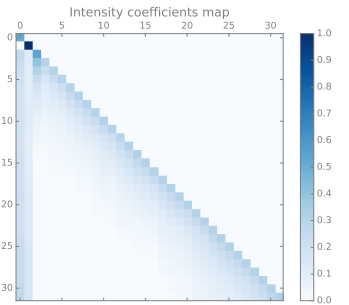


$\nabla_{\mathbf{x}} T$

Example – Stochastic volatility

Iteration 12 – Pullback $T^\sharp \pi$



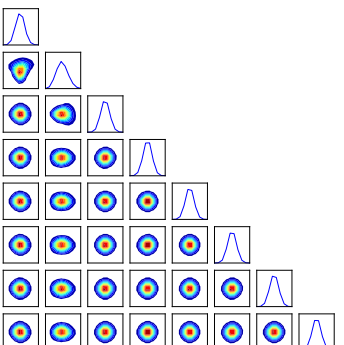


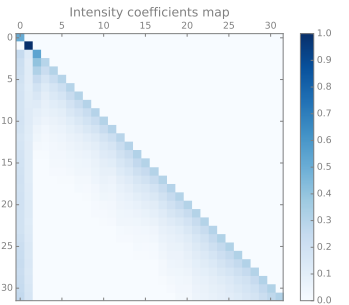
$\nabla_{\mathbf{x}} T$

Example – Stochastic volatility

Iteration 13 – Pullback $T^\sharp \pi$

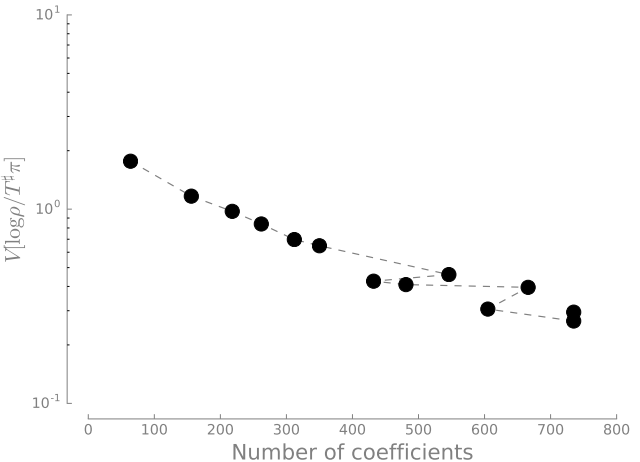
Conditionals along coordinate axes





$\nabla_{\mathbf{x}} T$

Example – Stochastic volatility



Variance diagnostic

Conclusions

Transport maps potential advantages

- Turns a Bayesian inference problem into an optimization problem
- We can use derivative based optimization if $\nabla_{\mathbf{x}} \log \pi$ or $\nabla^2_{\mathbf{x}} \log \pi$ are provided
- We can explore π in parallel
- We can generate i.i.d. samples from $T^*_\sharp \rho \approx \pi$ in parallel
- We can estimate convergence!

Key contributions

- First variation informed adaptivity
  - No need for additional evaluation of π
  - Exploits marginal independence present in the problem
- Numerical robust algorithm to perform adaptivity

Thanks to:

U.S. DEPARTMENT OF ENERGY

Parametrization of transport maps

Requirements:

- Lower-triangular structure

$$T : \mathbb{R}^d \to \mathbb{R}^d$$

$$\mathbf{x} \mapsto \begin{bmatrix} T^{(1)}(x_1) \\ T^{(2)}(x_1, x2) \\ \vdots \\ T^{(d)}(x_1, \ldots, x_d) \end{bmatrix}$$

Note: $\log |\nabla T| = \sum_{i=1}^{d} \log \partial_{x_i} T^{(i)}$

Parametrization of transport maps

Requirements:

- Lower-triangular structure
- Monotonicity (mass preservation), i.e. $|\nabla T| = \prod_{i=1}^{d} \partial_{x_i} T^{(i)} > 0$

$$T^{(i)}(x_1, \dots, x_i) = c_i(x_1, \dots, x_{i-1}) + \int_0^{x_i} \exp\left(h_i(x_1, \dots, x_{i-1}, t)\, dt\right.$$

$$c_i(x_1, \dots, x_{i-1}) = \sum_{j=1}^{N} a_j^{(i)} \Phi_j(x_1, \dots, x_{i-1})$$

$$h_i(x_1, \dots, x_{i-1}, t) = \sum_{j=1}^{N} b_j^{(i)} \Psi_j(x_1, \dots, x_{i-1}, t)$$

1. $\{\Phi_j\}$ Hermite polynomials, $\{\Psi_j\}$ Hermite functions
2. $\{\Phi_j\}$ and $\{\Psi_j\}$ radial basis functions