Robust and adaptive construction of measure transports for Bayesian inference

---

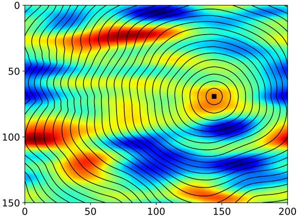
D. Bigoni (**dabi@mit.edu**), A. Spantini, Y.M. Marzouk

Massachusetts Institute of Technology

Past and present contributors:
Tarek El Moselhy, Matthew Parno, Xun Huan, Rebecca Morrison,
Ricardo M. Batista, Benjamin Zhang, Zheng Wang

SIAM CSE – Atlanta – 02/27/2017

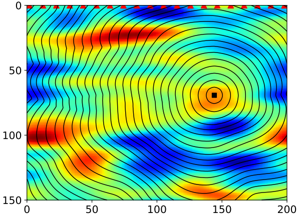
Bayesian inference – an example

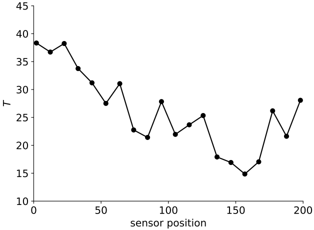


Physical model

$$|\nabla \overbrace{G(\mathbf{x})}^{\text{travel time}}| = \underbrace{v(\mathbf{x})}_{\text{velocity field}}{}^{-1}$$

Bayesian inference – an example





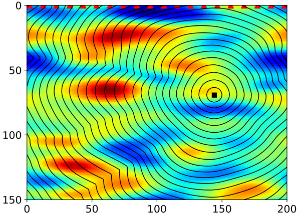
Physical model

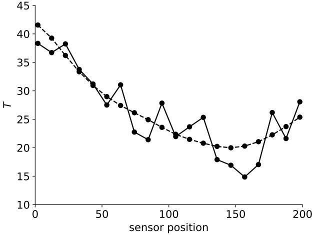
$$|\nabla \overbrace{G(\mathbf{x})}^{\text{travel time}}| = \underbrace{v(\mathbf{x})}_{\text{velocity field}}{}^{-1}$$

Data model

$$\overbrace{\mathbf{d}}^{\text{data}} = \mathbf{G}(\mathbf{v}) + \underbrace{\varepsilon}_{\text{noise}}$$

Bayesian inference – an example





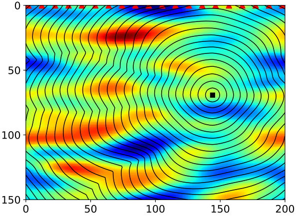
Physical model

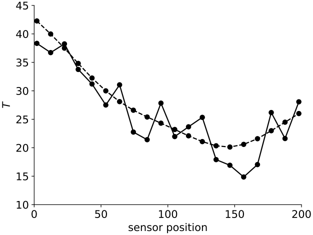
$$|\nabla \overbrace{G(\mathbf{x})}^{\text{travel time}}| = \underbrace{v(\mathbf{x})}_{\text{velocity field}}{}^{-1}$$

Data model

$$\overbrace{\mathbf{d}}^{\text{data}} = \mathbf{G}(\mathbf{v}) + \underbrace{\varepsilon}_{\text{noise}}$$

Bayesian inference – an example





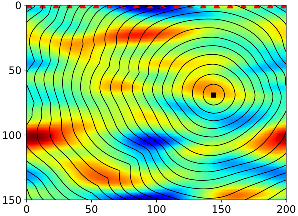
Physical model

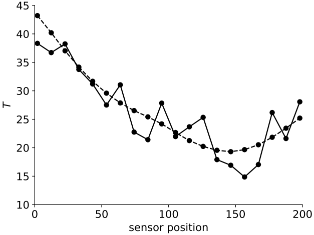
$$\left|\nabla \overbrace{G(\mathbf{x})}^{\text{travel time}}\right| = \underbrace{v(\mathbf{x})}_{\text{velocity field}}{}^{-1}$$

Data model

$$\overbrace{\mathbf{d}}^{\text{data}} = \mathbf{G}(\mathbf{v}) + \underbrace{\varepsilon}_{\text{noise}}$$

Bayesian inference – an example





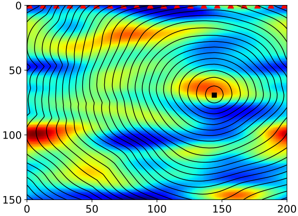
Physical model

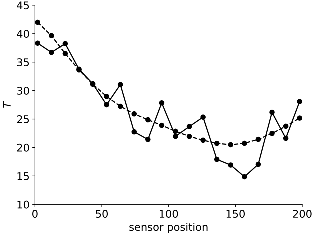
$$|\nabla \overbrace{G(\mathbf{x})}^{\text{travel time}}| = \underbrace{v(\mathbf{x})}_{\text{velocity field}}{}^{-1}$$

Data model

$$\overbrace{\mathbf{d}}^{\text{data}} = \mathbf{G}(\mathbf{v}) + \underbrace{\varepsilon}_{\text{noise}}$$

Bayesian inference – an example





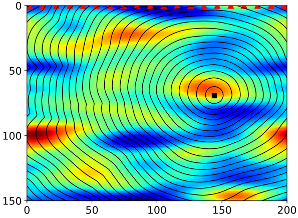
Physical model

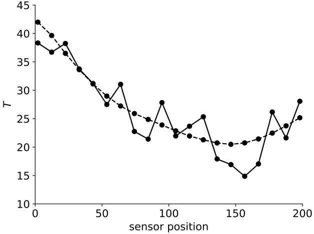
$$|\nabla \overbrace{G(\mathbf{x})}^{\text{travel time}}| = \underbrace{v(\mathbf{x})}_{\text{velocity field}}{}^{-1}$$

Data model

$$\overbrace{\mathbf{d}}^{\text{data}} = \mathbf{G}(\mathbf{v}) + \underbrace{\varepsilon}_{\text{noise}}$$

Bayesian inference – an example





Physical model

$$|\nabla \overbrace{G(\mathbf{x})}^{\text{travel time}}| = \underbrace{v(\mathbf{x})}_{\text{velocity field}}{}^{-1}$$

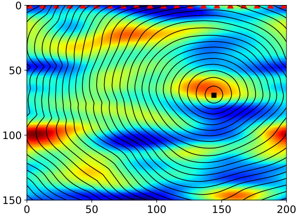
Data model

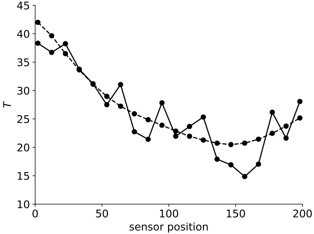
$$\overbrace{\mathbf{d}}^{\text{data}} = \mathbf{G}(\mathbf{v}) + \underbrace{\boldsymbol{\varepsilon}}_{\text{noise}}$$

Inference model

$$\underbrace{\pi_{\text{pos}}(\mathbf{v}|\mathbf{d})}_{\text{posterior}} \propto \overbrace{\mathcal{L}_{\mathbf{d}}(\mathbf{v})}^{\text{likelihood}} \underbrace{\pi_{\text{pr}}(\mathbf{v})}_{\text{prior}} = \pi_{\boldsymbol{\varepsilon}}(\mathbf{d} - \mathbf{G}(\mathbf{v}))\pi_{\text{pr}}(\mathbf{v})$$

Bayesian inference – an example





Inference model

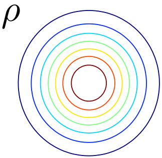
$$\underbrace{\pi_{\mathrm{pos}}(\mathbf{v}|\mathbf{d})}_{\text{posterior}} \propto \overbrace{\mathcal{L}_{\mathbf{d}}(\mathbf{v})}^{\text{likelihood}} \underbrace{\pi_{\mathrm{pr}}(\mathbf{v})}_{\text{prior}} = \pi_{\boldsymbol{\varepsilon}}(\mathbf{d} - \mathbf{G}(\mathbf{v}))\pi_{\mathrm{pr}}(\mathbf{v})$$

Decisions under uncertainty

$$\min_{\delta} \int L(\mathbf{v}, \delta)\pi_{\mathrm{pos}}(\mathbf{v}|\mathbf{d})d\mathbf{v}$$

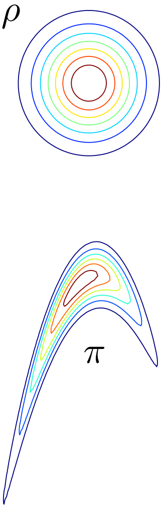
Measure transport – Pullbacks [PB] and Pushforwards [PF]

- Distribution $\boldsymbol{\nu}_\rho$ with density $\rho : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$



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

- Distribution $\boldsymbol{\nu}_\rho$ with density $\rho : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$
- Distribution $\boldsymbol{\nu}_\pi$ with density $\pi : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$

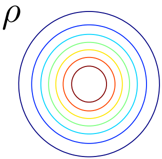


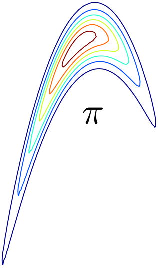
Measure transport – Pullbacks [PB] and Pushforwards [PF]

- Distribution $\boldsymbol{\nu}_\rho$ with density $\rho : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$
- Distribution $\boldsymbol{\nu}_\pi$ with density $\pi : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$
- For $T : \mathbb{R}^d \to \mathbb{R}^d$ we define

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

PB $\quad T^\sharp \pi = \pi \circ T |\nabla T|$





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

- Distribution $\boldsymbol{\nu}_\rho$ with density $\rho : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$
- Distribution $\boldsymbol{\nu}_\pi$ with density $\pi : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$
- For $T : \mathbb{R}^d \to \mathbb{R}^d$ we define

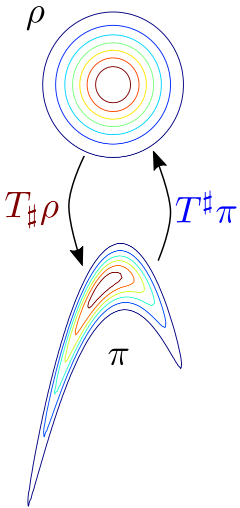
PF $T_\sharp \rho = \rho \circ T^{-1} |\nabla T^{-1}|$

PB $T^\sharp \pi = \pi \circ T |\nabla T|$

- We want T such that

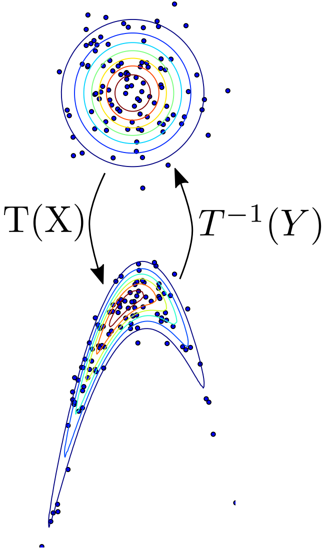
PF $T_\sharp \rho = \pi$

PB $T^\sharp \pi = \rho$



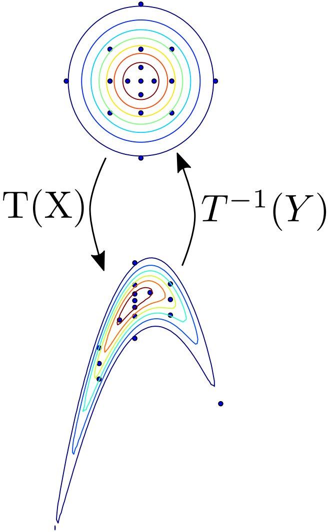
Measure transport – Pullbacks [PB] and Pushforwards [PF]

- Distribution $\boldsymbol{\nu}_\rho$ with density $\rho : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$
- Distribution $\boldsymbol{\nu}_\pi$ with density $\pi : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$
- For $T : \mathbb{R}^d \to \mathbb{R}^d$ we define
  - **PF** $T_\sharp \rho = \rho \circ T^{-1} |\nabla T^{-1}|$
  - **PB** $T^\sharp \pi = \pi \circ T |\nabla T|$
- We want T such that
  - **PF** For $X \sim \boldsymbol{\nu}_\rho$, $T(X) \sim \boldsymbol{\nu}_\pi$
  - **PB** For $Y \sim \boldsymbol{\nu}_\pi$, $T^{-1}(Y) \sim \boldsymbol{\nu}_\rho$



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

- Distribution $\boldsymbol{\nu}_\rho$ with density $\rho : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$
- Distribution $\boldsymbol{\nu}_\pi$ with density $\pi : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$
- For $T : \mathbb{R}^d \to \mathbb{R}^d$ we define
  - **PF** $T_\sharp \rho = \rho \circ T^{-1} |\nabla T^{-1}|$
  - **PB** $T^\sharp \pi = \pi \circ T |\nabla T|$
- We want T such that
  - **PF** For $X \sim \boldsymbol{\nu}_\rho$, $T(X) \sim \boldsymbol{\nu}_\pi$
  - **PB** For $Y \sim \boldsymbol{\nu}_\pi$, $T^{-1}(Y) \sim \boldsymbol{\nu}_\rho$



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

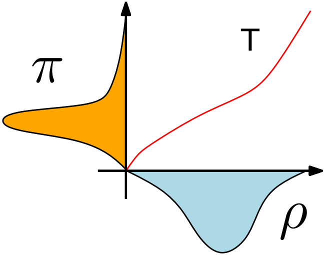
- Distribution $\boldsymbol{\nu}_\rho$ with density $\rho : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$
- Distribution $\boldsymbol{\nu}_\pi$ with density $\pi : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$
- For $T : \mathbb{R}^d \to \mathbb{R}^d$ we define

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

  PB $T^\sharp \pi = \pi \circ T |\nabla T|$
- We want T such that

  PF For $X \sim \boldsymbol{\nu}_\rho$, $T(X) \sim \boldsymbol{\nu}_\pi$

  PB For $Y \sim \boldsymbol{\nu}_\pi$, $T^{-1}(Y) \sim \boldsymbol{\nu}_\rho$



Knothe-Rosenblatt rearrangement

$\forall \boldsymbol{\nu}_\rho, \boldsymbol{\nu}_\pi$ absolutely continuous there exists a **triangular monotone** map s.t. $T(d\boldsymbol{\nu}_\rho) = d\boldsymbol{\nu}_\pi$

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

Triangular monotone maps

$$\mathcal{T}_\triangle = \left\{ T : \mathbb{R}^d \to \mathbb{R}^d : \overbrace{[T(\mathbf{x})]_k = T^{(k)}(x_1, \dots, x_k)}^{\text{triangular}} \text{ and } \overbrace{\partial_{x_k} T^{(k)} > 0}^{\text{monotone}} \right\}$$

Triangular monotone maps

$$\mathcal{T}_\triangle = \left\{ T : \mathbb{R}^d \to \mathbb{R}^d : \overbrace{[T(\mathbf{x})]_k = T^{(k)}(x_1, \dots, x_k)}^{\text{triangular}} \text{ and } \overbrace{\partial_{x_k} T^{(k)} > 0}^{\text{monotone}} \right\}$$

Integrated squared representation

$$T^{(k)}(x_{1:k}) = c_k(x_{1:k-1}) + \int_0^{x_k} \Big(h_k(x_{1:k-1}, t) + \varepsilon \Big)^2 dt$$

Triangular monotone maps

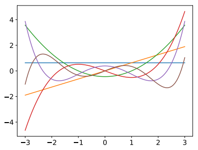
$$\mathcal{T}_\triangle = \left\{ T : \mathbb{R}^d \to \mathbb{R}^d : \overbrace{[T(\mathbf{x})]_k = T^{(k)}(x_1, \ldots, x_k)}^{\text{triangular}} \text{ and } \overbrace{\partial_{x_k} T^{(k)} > 0}^{\text{monotone}} \right\}$$

Integrated squared representation

$$T^{(k)}(x_{1:k}) = \boxed{c_k(x_{1:k-1})} + \int_0^{x_k} \left(\boxed{h_k(x_{1:k-1}, t)} + \varepsilon \right)^2 dt$$

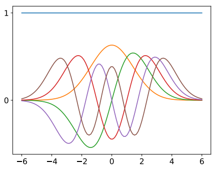
Constant part – $\mathcal{T}^0_\triangle$

$$c_k(x_{1:k-1}) = \sum_{\mathbf{i} \in \mathcal{I}_k} \mathbf{a}_\mathbf{i} \Phi_\mathbf{i}(x_{1:k-1})$$



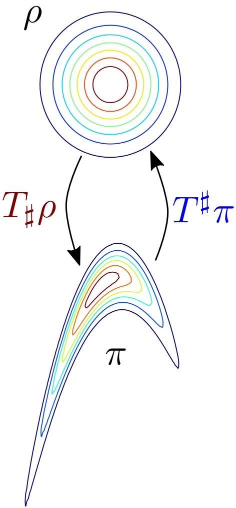
Squared part – $\mathcal{T}^0_\triangle$

$$h_k(x_{1:k-1}, t) = \sum_{\mathbf{j} \in \mathcal{J}_k} \mathbf{b}_\mathbf{j} \Psi_\mathbf{j}(x_{1:k-1}, t)$$



Knothe-Rosenblatt rearrangement

$\forall \boldsymbol{\nu}_\rho, \boldsymbol{\nu}_\pi$ absolutely continuous there exists a **triangular monotone** map s.t. $T(d\boldsymbol{\nu}_\rho) = d\boldsymbol{\nu}_\pi$



**How to find the map $T \in \mathcal{T}_\triangle$
such that $T_\sharp \rho = \pi$?**

Minimize KL-divergence to find optimal map

$$T^* = \underset{T \in \mathcal{T}_\triangle}{\arg\min}\, D_{\mathrm{KL}}(T_\sharp \rho \| \pi) = \underset{T \in \mathcal{T}_\triangle}{\arg\min}\, \mathbb{E}_\rho \left[\log \frac{\rho}{T^\sharp \pi} \right]$$

Minimize KL-divergence to find optimal map

$$T^* = \arg\min_{T \in \mathcal{T}_\triangle} D_{\mathrm{KL}}(T_\sharp \rho \| \pi) = \arg\min_{T \in \mathcal{T}_\triangle} \mathbb{E}_\rho \left[\log \frac{\rho}{T^\sharp \pi} \right]$$

- + **Derivative based unconstrained optimization** if gradients are available
- + We can **explore π in parallel**
- + We can **generate i.i.d. samples** from $T_\sharp^* \rho \propto \pi$ **in parallel**

Minimize KL-divergence to find optimal map

$$T^* = \arg\min_{T\in\mathcal{T}_\triangle} D_{\mathrm{KL}}(T_\sharp\rho\|\pi) = \arg\min_{T\in\mathcal{T}_\triangle} \mathbb{E}_\rho\left[\log\frac{\rho}{T^\sharp\pi}\right]$$

+ **Derivative based unconstrained optimization** if gradients are available
+ We can **explore π in parallel**
+ We can **generate i.i.d. samples** from $T_\sharp^*\rho \propto \pi$ **in parallel**

− We need to **approximate d functions up to d-dimensional!**

Minimize KL-divergence to find optimal map

$$T^* = \arg\min_{T \in \mathcal{T}_\triangle} D_{\mathrm{KL}}(T_\sharp \rho \| \pi) = \arg\min_{T \in \mathcal{T}_\triangle} \mathbb{E}_\rho \left[\log \frac{\rho}{T^\sharp \pi} \right]$$

- \+ **Derivative based unconstrained optimization** if gradients are available
- \+ We can **explore π in parallel**
- \+ We can **generate i.i.d. samples** from $T_\sharp^* \rho \propto \pi$ **in parallel**
- − We need to **approximate d functions up to d-dimensional!**

Sources of low-dimensional structure

- Smoothness
- Marginal independence
- Conditional independence
- Low-rank structure

Convergence criterion

$$T^* = \arg\min_{T \in \mathcal{T}_\triangle} D_{\mathrm{KL}}(T_\sharp \rho \| \pi) = \arg\min_{T \in \mathcal{T}_\triangle} \mathbb{E}_\rho \left[\log \frac{\rho}{T^\sharp \pi} \right]$$

Optimal $T^* \in \mathcal{T}_\triangle$ and $\int \pi = 1 \quad \Rightarrow \quad D_{\mathrm{KL}}(T^*_\sharp \rho \| \pi) = 0$

But, optimal $\tilde{T}^* \in \mathcal{T}^0_\triangle$ or $\int \pi \neq 1 \quad \Rightarrow \quad D_{\mathrm{KL}}(\tilde{T}^*_\sharp \rho \| \pi) \neq 0$

Convergence criterion

$$T^* = \underset{T \in \mathcal{T}_\triangle}{\arg\min}\, D_{\mathrm{KL}}(T_\sharp \rho \| \pi) = \underset{T \in \mathcal{T}_\triangle}{\arg\min}\, \mathbb{E}_\rho \left[\log \frac{\rho}{T^\sharp \pi} \right]$$

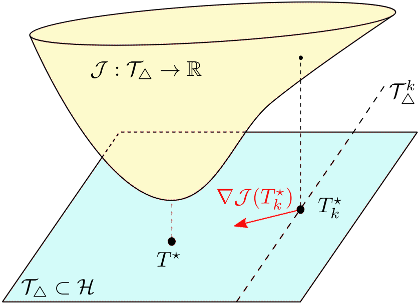
Optimal $T^* \in \mathcal{T}_\triangle$ and $\int \pi = 1 \quad \Rightarrow \quad D_{\mathrm{KL}}(T^*_\sharp \rho \| \pi) = 0$

But, optimal $\tilde{T}^* \in \mathcal{T}^0_\triangle$ or $\int \pi \neq 1 \quad \Rightarrow \quad D_{\mathrm{KL}}(\tilde{T}^*_\sharp \rho \| \pi) \neq 0$

$\mathbb{V}\left[\log \frac{\rho}{T^\sharp \pi}\right] \xrightarrow{T \to T^*} 0$ as $\frac{1}{2} D_{\mathrm{KL}}(T^*_\sharp \rho \| \pi) \xrightarrow{T \to T^*} -\log \int \pi$

Refinement criterion

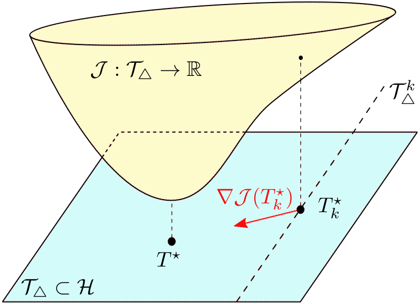
$$\boxed{T^* = \arg\min_{T \in \mathcal{T}_\triangle} \underbrace{D_{KL}(T_\sharp \rho \| \pi)}_{\mathcal{J}(T)}}$$



Refinement criterion

$$T^* = \arg\min_{T\in\mathcal{T}_\triangle} \underbrace{D_{KL}(T_\sharp\rho\|\pi)}_{\mathcal{J}(T)}$$

$$T_0^* = \arg\min_{T\in\mathcal{T}_\triangle^0} \mathcal{J}(T)$$

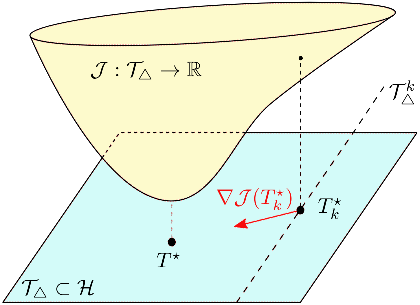


Refinement criterion

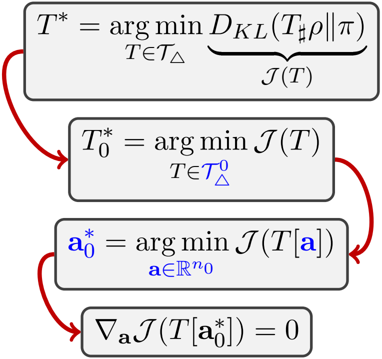
$$T^* = \arg\min_{T\in\mathcal{T}_\triangle} \underbrace{D_{KL}(T_\sharp\rho\|\pi)}_{\mathcal{J}(T)}$$

$$T_0^* = \arg\min_{T\in\mathcal{T}_\triangle^0} \mathcal{J}(T)$$

$$\mathbf{a}_0^* = \arg\min_{\mathbf{a}\in\mathbb{R}^{n_0}} \mathcal{J}(T[\mathbf{a}])$$



Refinement criterion

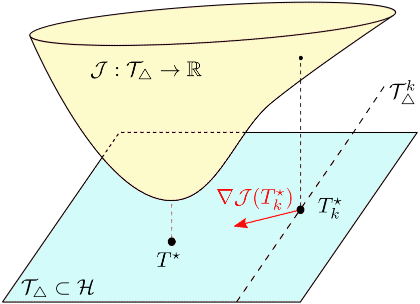


$$T^* = \arg\min_{T\in\mathcal{T}_\triangle} \underbrace{D_{KL}(T_\sharp\rho\|\pi)}_{\mathcal{J}(T)}$$

$$T_0^* = \arg\min_{T\in\mathcal{T}_\triangle^0} \mathcal{J}(T)$$

$$\mathbf{a}_0^* = \arg\min_{\mathbf{a}\in\mathbb{R}^{n_0}} \mathcal{J}(T[\mathbf{a}])$$

$$\nabla_{\mathbf{a}}\mathcal{J}(T[\mathbf{a}_0^*]) = 0$$



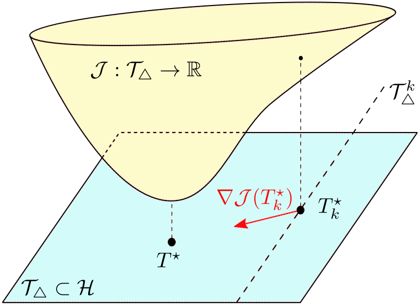
Refinement criterion

$$T^* = \arg\min_{T\in\mathcal{T}_\triangle} \underbrace{D_{KL}(T_\sharp\rho\|\pi)}_{\mathcal{J}(T)}$$

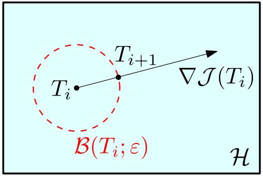
$$T_0^* = \arg\min_{T\in\mathcal{T}_\triangle^0} \mathcal{J}(T)$$

$$\mathbf{a}_0^* = \arg\min_{\mathbf{a}\in\mathbb{R}^{n_0}} \mathcal{J}(T[\mathbf{a}])$$

$$\nabla_{\mathbf{a}}\mathcal{J}(T[\mathbf{a}_0^*]) = 0$$



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



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

There exists $\varepsilon > 0$ such that

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

Use the first variation to enrich the approximation space

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

Use the first variation to enrich the approximation space

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

- $\nabla \mathcal{J}\left(T[\mathbf{a}_0^*]\right) : \mathbb{R}^d \to \mathbb{R}^d$ **is a map** in $\mathcal{H} \supset \mathcal{T}_\triangle$

Use the first variation to enrich the approximation space

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

- $\nabla \mathcal{J}\left(T[\mathbf{a}_0^*]\right) : \mathbb{R}^d \to \mathbb{R}^d$ **is a map** in $\mathcal{H} \supset \mathcal{T}_\triangle$

Projection on $\mathcal{T}_\triangle^1 \supset \mathcal{T}_\triangle^0$

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

Use the first variation to enrich the approximation space

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

- $\nabla \mathcal{J}\left(T[\mathbf{a}_0^*]\right) : \mathbb{R}^d \to \mathbb{R}^d$ **is a map** in $\mathcal{H} \supset \mathcal{T}_\triangle$

Projection on $\mathcal{T}_\triangle^1 \supset \mathcal{T}_\triangle^0$

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

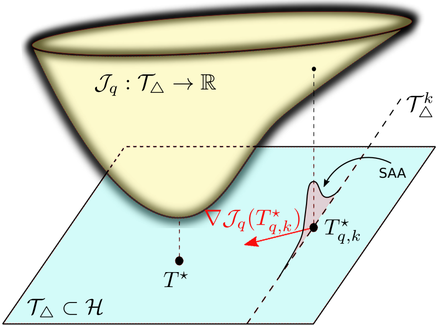
- **No new evaluation** of $\nabla_{\mathbf{x}} \log \pi$ is required
- $U[\mathbf{b}_1^*]$ informs about **active variables** to be included
- $U[\mathbf{b}_1^*]$ informs about **active coefficients** to be included

Controlling the sample average accuracy

$$T_k^* = \underset{T \in \mathcal{T}_\triangle^k}{\arg\min} -\mathbb{E}_\rho \left[\log T^\sharp \pi \right] \approx \underset{T \in \mathcal{T}_\triangle^k}{\arg\min} \overbrace{- \sum_{1 \leq i \leq q} \log T^\sharp \pi(\mathbf{x}_i)}^{\mathcal{J}_q(T)} =: T_{q,k}^*$$

Controlling the sample average accuracy

$$T_k^* = \arg\min_{T\in\mathcal{T}_\triangle^k} -\mathbb{E}_\rho\left[\log T^\sharp\pi\right] \approx \arg\min_{T\in\mathcal{T}_\triangle^k} \overbrace{-\sum_{1\le i\le q}\log T^\sharp\pi(\mathbf{x}_i)}^{\mathcal{J}_q(T)} =: T^*_{q,k}$$

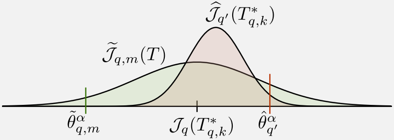


Controlling the sample average accuracy

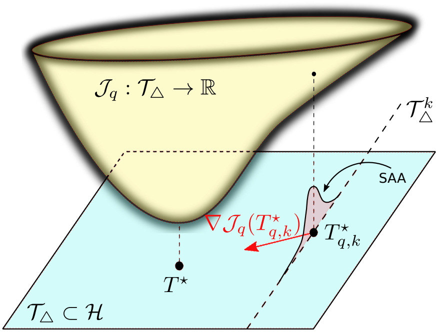
$$T^*_k = \underset{T\in\mathcal{T}^k_\triangle}{\arg\min} -\mathbb{E}_\rho\left[\log T^\sharp\pi\right] \approx \underset{T\in\mathcal{T}^k_\triangle}{\arg\min} - \overbrace{\sum_{1\leq i\leq q} \log T^\sharp\pi(\mathbf{x}_i)}^{\mathcal{J}_q(T)} =: T^*_{q,k}$$

Sample average approximation

$$\tilde{\theta}_{q,m} \leq \mathcal{J}(T^*_k) \leq \hat{\theta}_{q'}$$

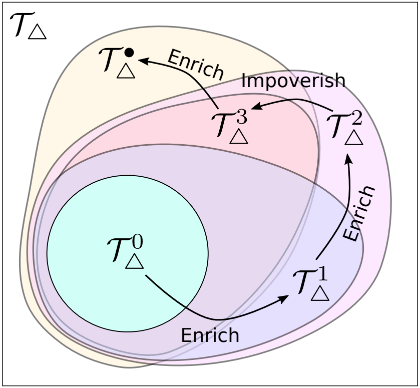


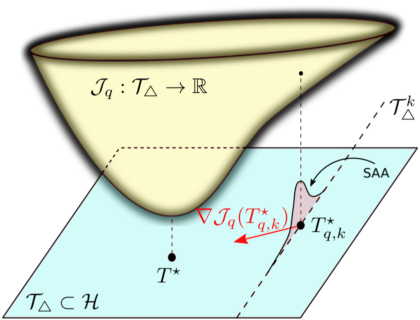
$$\widetilde{J}_{q,m}(T) = \frac{1}{m}\sum_{i=1}^{m} \min_{T\in\mathcal{T}^k_\triangle} \mathcal{J}_q(T)$$



Adaptivity ingredients

- **Convergence criterion – Variance diagnostic** : $\mathbb{V}\left[\log \frac{\rho}{T^\sharp \pi}\right]$
- **Refinement criterion – First variation** : $\nabla \mathcal{J}\left(T[\mathbf{a}^*]\right)$
- **Stability criterion – Sample average approximation** : $\tilde{\theta}_{q,m} \leq \mathcal{J}(T_k^*) \leq \hat{\theta}_{q'}$





The banana distribution – $d = 2$

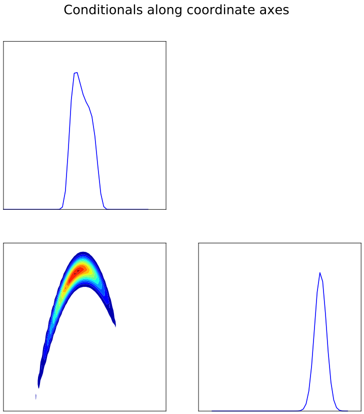
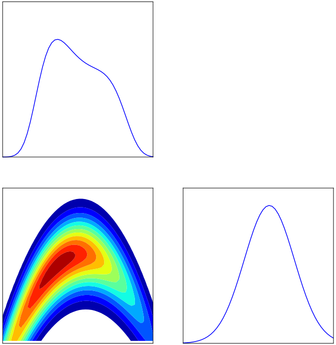


Figure: Target π

The banana distribution – $d = 2$

Iteration 1 – Pullback $T^\sharp \pi$


Conditionals along coordinate axes

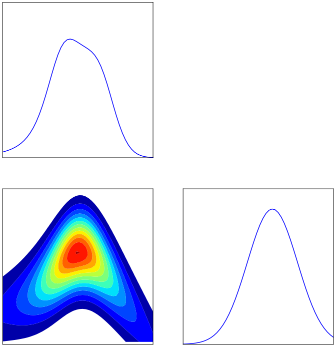

Intensity coefficients map

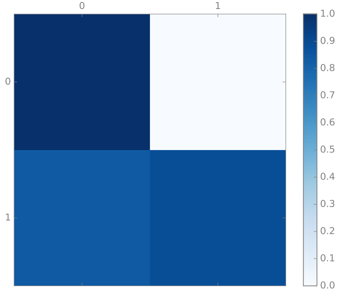
$\nabla_{\mathbf{x}} T$

Reminder: $T^\sharp \pi \approx \rho$, where ρ is the density of $\mathcal{N}(0, \mathbf{I})$

The banana distribution – $d = 2$

Iteration 2 – Pullback $T^\sharp \pi$


Conditionals along coordinate axes

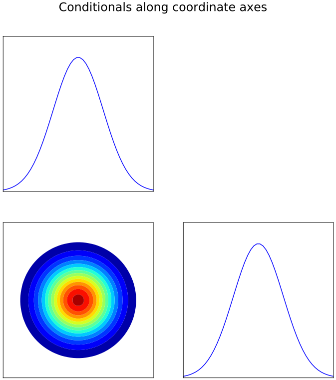

Intensity coefficients map

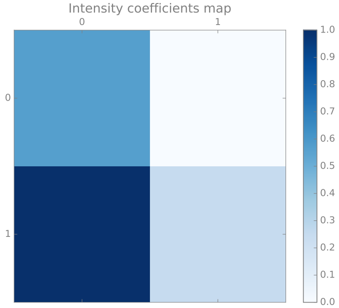
$$\nabla_{\mathbf{x}} T$$

Reminder: $T^\sharp \pi \approx \rho$, where ρ is the density of $\mathcal{N}(0, \mathbf{I})$

The banana distribution – $d = 2$

Iteration 3 – Pullback $T^\sharp \pi$

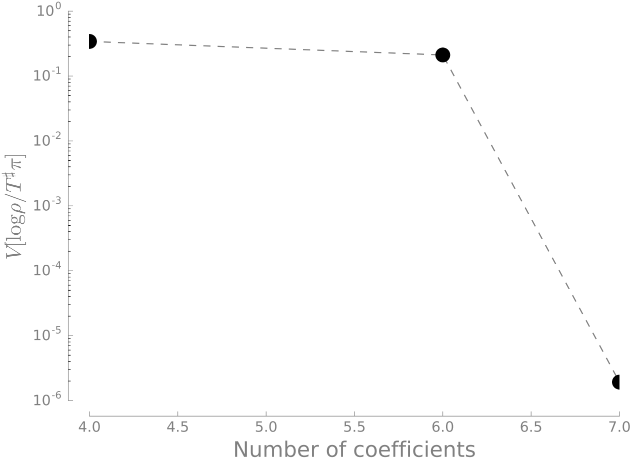




$\nabla_{\mathbf{x}} T$

Reminder: $T^\sharp \pi \approx \rho$, where ρ is the density of $\mathcal{N}(0, \mathbf{I})$

The banana distribution – $d = 2$



Variance diagnostic

Stochastic volatility of financial assets – $d = 32$

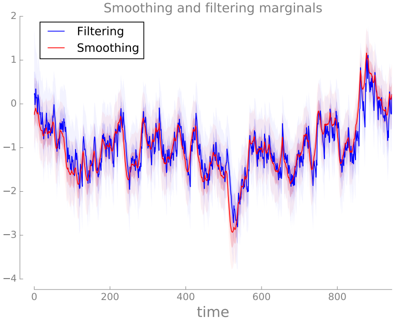
- Latent log-volatilities modeled with an AR(1) process for $t = 1, \ldots, N$ $(N = 30)$

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

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

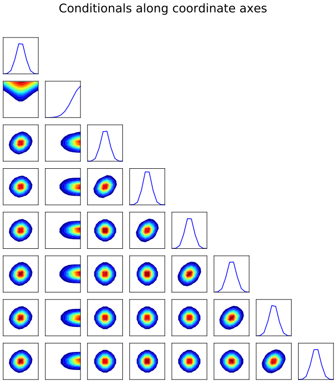
$$Y_t = \varepsilon_t \exp(X_t/2) \ , \quad \varepsilon_t \sim \mathcal{N}(0,1)$$

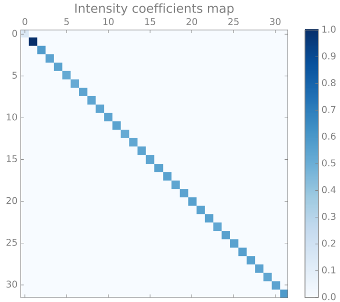
- We want to characterize $\pi \sim \mu, \phi, \mathbf{X}_{1:N} | \mathbf{Y}_{1:N}$



Stochastic volatility of financial assets – $d = 32$

Iteration 1 – Pullback $T^\sharp \pi$



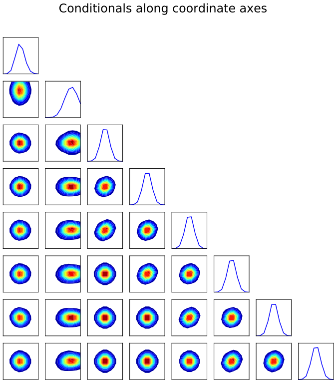


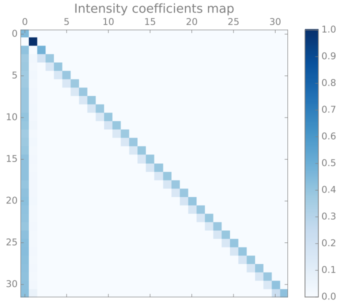
$\nabla_{\mathbf{x}} T$

Reminder: $T^\sharp \pi \approx \rho$, where ρ is the density of $\mathcal{N}(0, \mathbf{I})$

Stochastic volatility of financial assets – $d = 32$

Iteration 2 – Pullback $T^{\sharp}\pi$



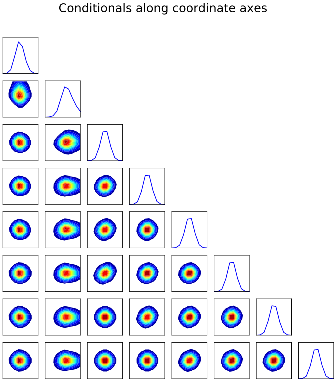


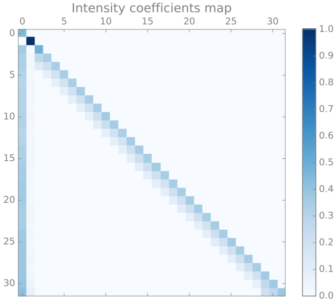
$\nabla_{\mathbf{x}} T$

Reminder: $T^{\sharp}\pi \approx \rho$, where ρ is the density of $\mathcal{N}(0, \mathbf{I})$

Stochastic volatility of financial assets – $d = 32$

Iteration 3 – Pullback $T^{\sharp}\pi$



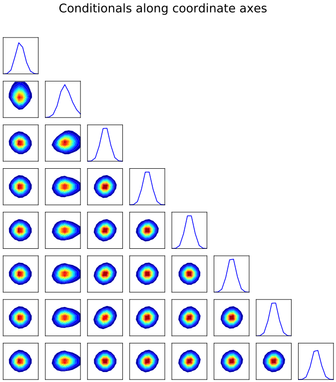


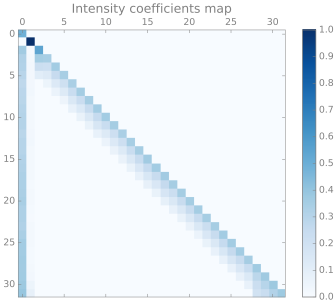
$\nabla_{\mathbf{x}} T$

Reminder: $T^{\sharp}\pi \approx \rho$, where ρ is the density of $\mathcal{N}(0, \mathbf{I})$

Stochastic volatility of financial assets – $d = 32$

Iteration 4 – Pullback $T^\sharp \pi$



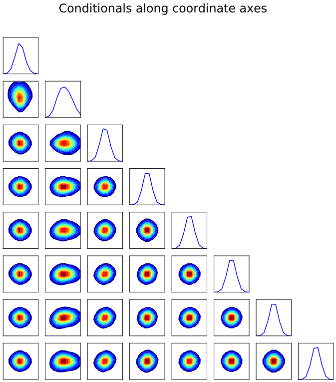


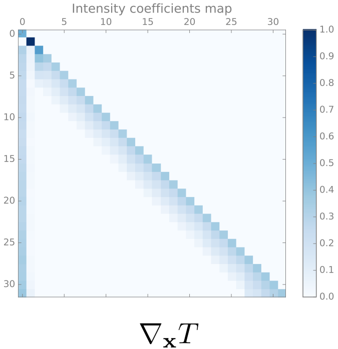
$\nabla_{\mathbf{x}} T$

Reminder: $T^\sharp \pi \approx \rho$, where ρ is the density of $\mathcal{N}(0, \mathbf{I})$

Stochastic volatility of financial assets – $d = 32$

Iteration 5 – Pullback $T^\sharp \pi$

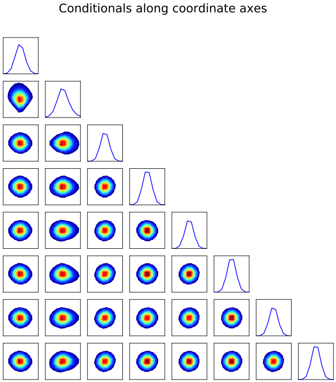


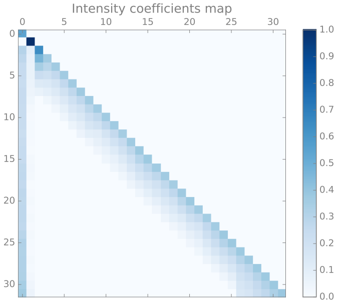


Reminder: $T^\sharp \pi \approx \rho$, where ρ is the density of $\mathcal{N}(0, \mathbf{I})$

Stochastic volatility of financial assets – $d = 32$

Iteration 6 – Pullback $T^\sharp\pi$



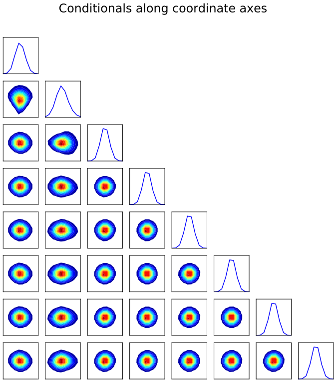


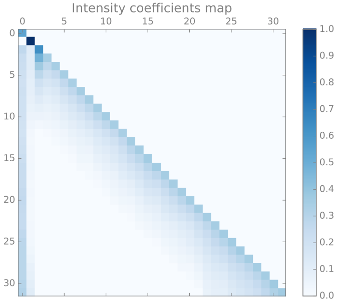
$\nabla_{\mathbf{x}} T$

Reminder: $T^\sharp\pi \approx \rho$, where ρ is the density of $\mathcal{N}(0, \mathbf{I})$

Stochastic volatility of financial assets – $d = 32$

Iteration 7 – Pullback $T^\sharp\pi$



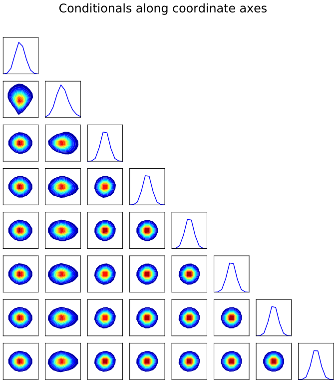


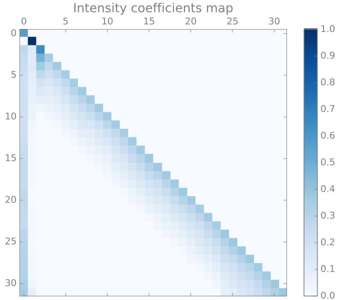
$\nabla_{\mathbf{x}} T$

Reminder: $T^\sharp\pi \approx \rho$, where ρ is the density of $\mathcal{N}(0, \mathbf{I})$

Stochastic volatility of financial assets – $d = 32$

Iteration 7 – Pullback $T^\sharp \pi$



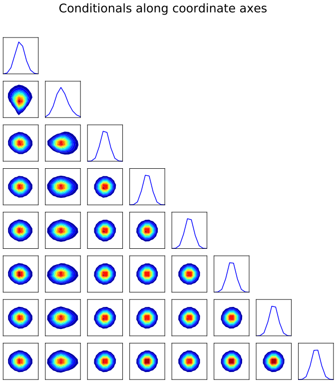


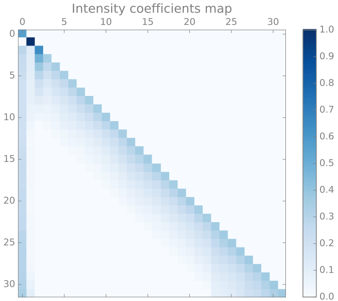
$\nabla_{\mathbf{x}} T$

Reminder: $T^\sharp \pi \approx \rho$, where ρ is the density of $\mathcal{N}(0, \mathbf{I})$

Stochastic volatility of financial assets – $d = 32$

Iteration 9 – Pullback $T^\sharp \pi$



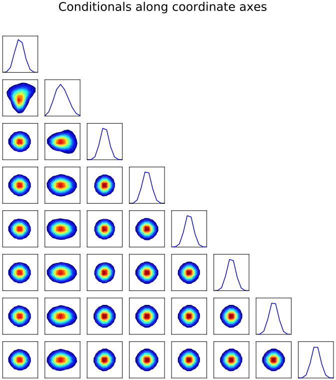


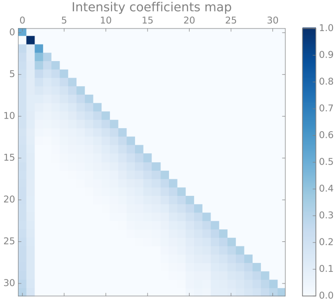
$\nabla_{\mathbf{x}} T$

Reminder: $T^\sharp \pi \approx \rho$, where ρ is the density of $\mathcal{N}(0, \mathbf{I})$

Stochastic volatility of financial assets – $d = 32$

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



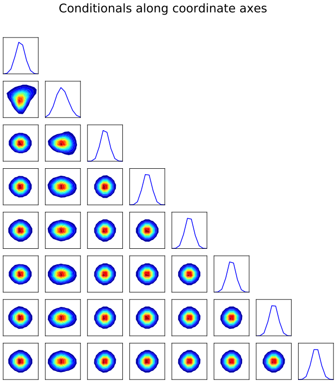


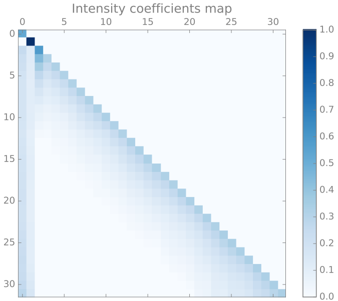
$\nabla_{\mathbf{x}} T$

Reminder: $T^\sharp \pi \approx \rho$, where ρ is the density of $\mathcal{N}(0, \mathbf{I})$

Stochastic volatility of financial assets – $d = 32$

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



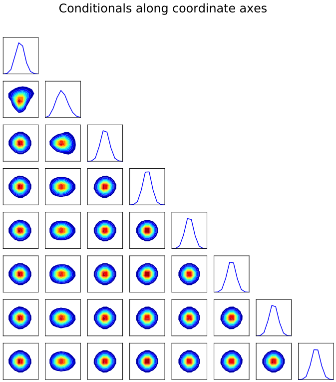


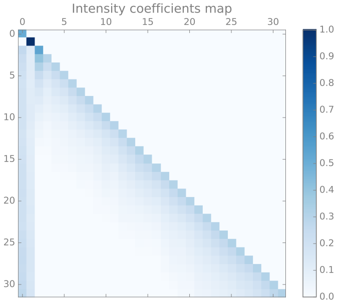
$$\nabla_{\mathbf{x}} T$$

Reminder: $T^\sharp \pi \approx \rho$, where ρ is the density of $\mathcal{N}(0, \mathbf{I})$

Stochastic volatility of financial assets – $d = 32$

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



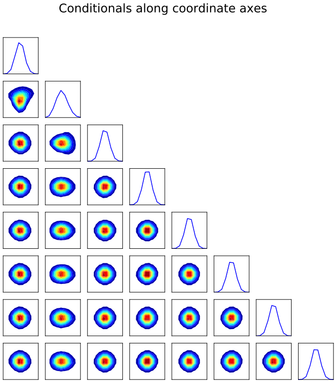


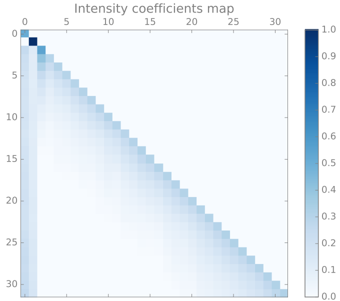
$\nabla_{\mathbf{x}} T$

Reminder: $T^\sharp \pi \approx \rho$, where ρ is the density of $\mathcal{N}(0, \mathbf{I})$

Stochastic volatility of financial assets – $d = 32$

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

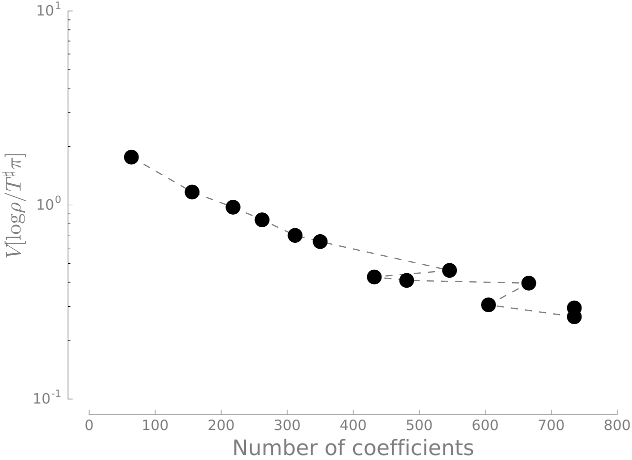




$\nabla_{\mathbf{x}} T$

Reminder: $T^\sharp \pi \approx \rho$, where ρ is the density of $\mathcal{N}(0, \mathbf{I})$

Stochastic volatility of financial assets – $d = 32$



Variance diagnostic

Key contributions

Robust adaptive algorithm for characterizing probability measures via **deterministic couplings** and **optimization**, exploiting **smoothness** and **marginal independence**

Contacts:

Daniele Bigoni – **dabi@mit.edu**
Alessio Spantini – **spantini@mit.edu**
Youssef Marzouk – **ymarz@mit.edu**

Thanks to: U.S. DEPARTMENT OF ENERGY

References:

El Moselhy et al. “Bayesian inference with optimal maps”
Parno et al. “Transport map accelerated Markov chain Monte Carlo”
Marzouk et al. “An introduction to sampling via measure transport”
Spantini et al. “Inference via low-dimensional couplings”