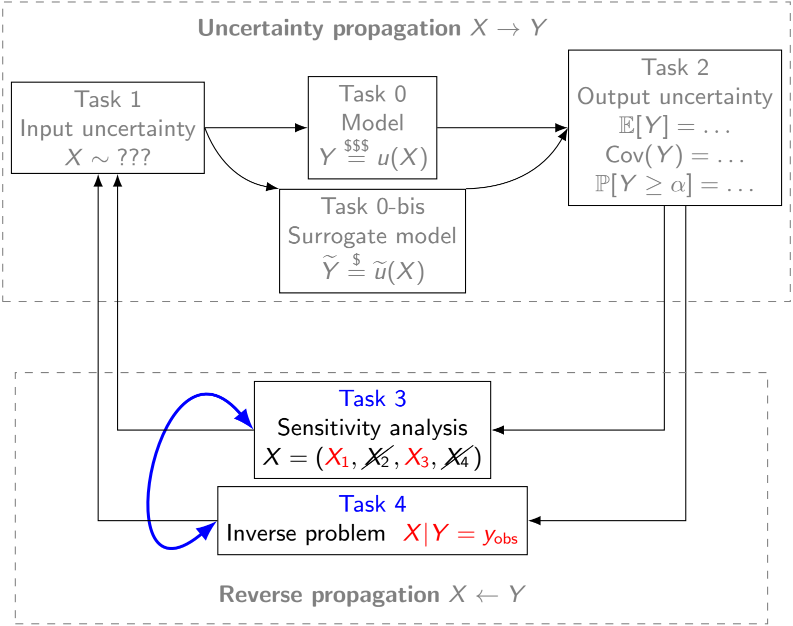
Poincaré inequalities for dimension reduction and efficient sampling

Part III: Dimension reduction for Bayesian inverse problems

Olivier Zahm
olivier.zahm@inria.fr

INRIA – Université Grenoble-Alpes
https://team.inria.fr/airsea/en/

ETICS
October 2025


Uncertainty propagation $X \to Y$
Task 1
Input uncertainty
$X \sim ???$
Task 0
Model
$Y \overset{\$\$\$}{=} u(X)$
Task 0-bis
Surrogate model
$\widetilde{Y} \overset{\$}{=} \widetilde{u}(X)$
Task 2
Output uncertainty
$\mathbb{E}[Y] = \dots$
$\mathrm{Cov}(Y) = \dots$
$\mathbb{P}[Y \geq \alpha] = \dots$
Task 3
Sensitivity analysis
$X = (X_1, \cancel{X_2}, X_3, \cancel{X_4})$
Task 4
Inverse problem $X | Y = y_{\text{obs}}$
Reverse propagation $X \leftarrow Y$

Inverse problem

Given an observation y_{obs} of

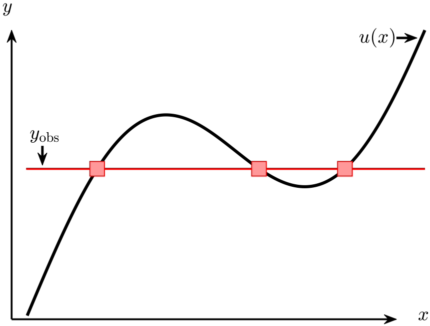
$$Y = u(X_1, \ldots, X_d) + \varepsilon,$$

where $\varepsilon \sim \mathcal{N}(0, \Sigma_{\text{obs}})$, how to identify the parameter X which is the most coherent with this observation?

The Bayesian perspective: update the "knowledge" on X after observing y_{obs}.

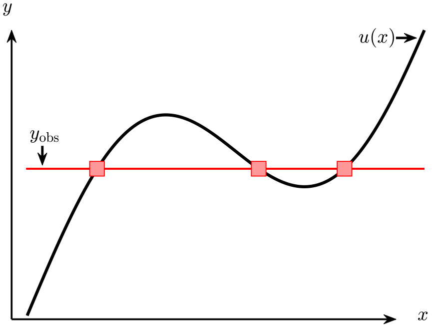
$\pi_X(x) \quad \longrightarrow$  $\longrightarrow \quad \pi_{X|Y}(x|y_{\text{obs}})$

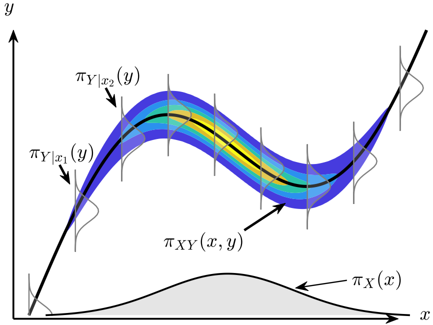
Variational



$$\min_{x} \underbrace{\tfrac{1}{2}\|y_{\mathrm{obs}} - u(x)\|^2_{\Sigma_{\mathrm{obs}}^{-1}}}_{\text{data mismatch}} + \underbrace{\lambda \mathcal{R}(x)}_{\text{regularization}}$$

Variational **V.S.** Bayesian

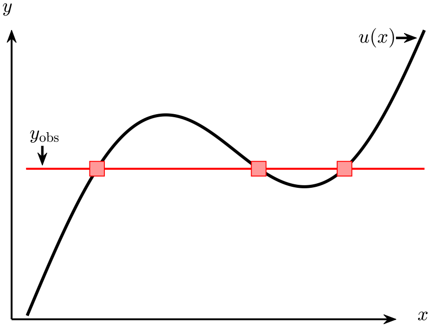


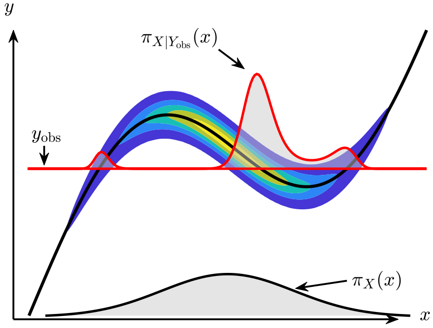


$$\min_{x} \underbrace{\frac{1}{2}\|y_{\text{obs}} - u(x)\|^2_{\Sigma_{\text{obs}}^{-1}}}_{\text{data mismatch}} + \underbrace{\lambda \mathcal{R}(x)}_{\text{regularization}}$$

$$\underbrace{\pi_{XY}(x,y)}_{\text{joint}} = \underbrace{\pi_{Y|X}(y|x)}_{\text{likelihood}} \underbrace{\pi_X(x)}_{\text{prior}}$$

Variational **V.S.** Bayesian



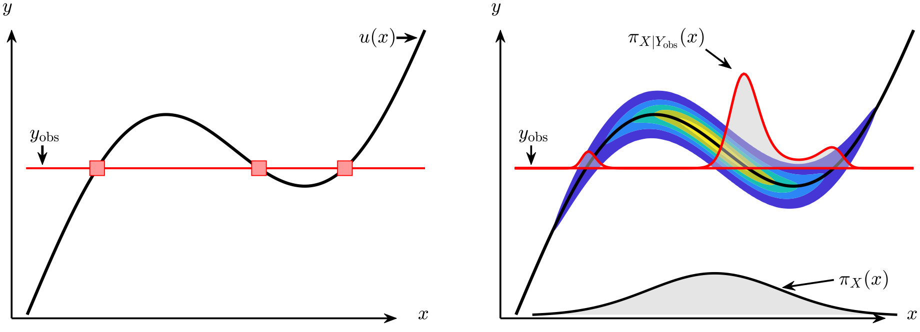


$$\min_x \; \underbrace{\tfrac{1}{2}\|y_{\text{obs}} - u(x)\|^2_{\Sigma_{\text{obs}}^{-1}}}_{\text{data mismatch}} + \underbrace{\lambda\mathcal{R}(x)}_{\text{regularization}}$$

$$\underbrace{\pi_{XY}(x,y)}_{\text{joint}} = \underbrace{\pi_{Y|X}(y|x)}_{\text{likelihood}}\underbrace{\pi_X(x)}_{\text{prior}}$$

Given y_{obs}, the **posterior** is given by $\boxed{\pi_{X|y_{\text{obs}}}(x) \propto \pi_{Y|X}(y_{\text{obs}}|x)\pi_X(x)}$

Variational **V.S.** Bayesian



$$\min_x \underbrace{\tfrac{1}{2}\|y_{\text{obs}} - u(x)\|^2_{\Sigma_{\text{obs}}^{-1}}}_{\text{data mismatch}} + \underbrace{\lambda \mathcal{R}(x)}_{\text{regularization}}$$

$$\underbrace{\pi_{XY}(x,y)}_{\text{joint}} = \underbrace{\pi_{Y|X}(y|x)}_{\text{likelihood}}\underbrace{\pi_X(x)}_{\text{prior}}$$

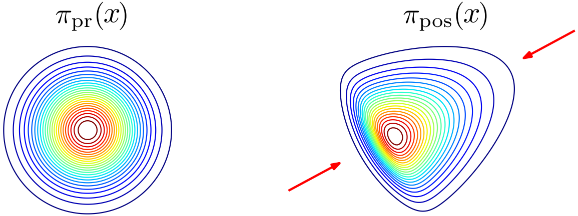
Given y_{obs}, the **posterior** is given by $\boxed{\pi_{X|y_{\text{obs}}}(x) \propto \pi_{Y|X}(y_{\text{obs}}|x)\pi_X(x)}$

Curse of dimensionality $x = (x_1, \ldots, x_d)$

Standard sampling algorithms (MCMC, Importance Sampling, Particle Filters) **struggle when** $d \gg 1$ (slow mixing, weight degeneracy, slow convergence...)

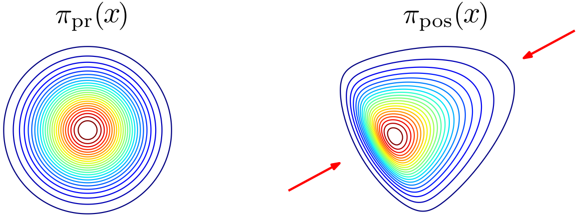
Low-effective dimension of Bayesian inverse problems

In many situations, the data are informative only on a **low-dimensional subspace**



Low-effective dimension of Bayesian inverse problems

In many situations, the data are informative only on a **low-dimensional subspace**



The **likelihood informed** components

$$X = \underbrace{U_m X_m}_{\text{informed}} + \underbrace{U_\perp X_\perp}_{\text{uninformed}}$$

Find a projection matrix $U_m \in \mathbb{R}^{d\times m}$ (**feature map**) and a reduced likelihood $\widetilde{\mathcal{L}}^y : \mathbb{R}^m \to \mathbb{R}_{\geq 0}$ (**profile function**) such that

$$\widetilde{\pi}_{X|Y}(x|y) \propto \widetilde{\mathcal{L}}^y(U_m^\top x)\pi_X(x)$$

yields a "good" posterior approximation to $\pi_{X|Y}(x|y) \propto \mathcal{L}^y(x)\pi_X(x)$

Subspace accelerated inference

[Vrugt 2009] Accelerating MCMC [...] with randomized **subspace sampling**, INT J NONLIN SCI NUM

[Marzouk Najm 2009] Dimensionality reduction and PC acceleration of Bayesian inference, JCP

[Cui Law Marzouk 2016] Dimension-independent **likelihood-informed** MCMC, JCP

[Constantine et.al. 2016] Accelerating MCMC with **active subspaces**, SIAM-SISC

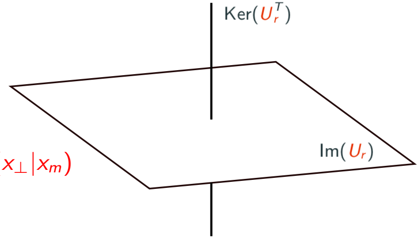
[Izmailov et.al. 2020] **Subspace inference** for Bayesian deep learning, PMLR

[Chen Ghattas 2020] **Projected** Stein variational gradient descent, NeurIPS

[Brennan et.al. 2020] Greedy inference with structure-exploiting **lazy** maps, NeurIPS

[.....]

$$X = \underbrace{U_m X_m}_{\in \mathrm{Im}(U_m)} + \underbrace{U_\perp X_\perp}_{\in \mathrm{Ker}(U_m^\top)}$$



Then

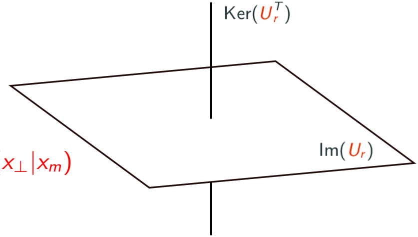
$$\widetilde{\pi}_{X|y_{\mathrm{obs}}}(x) = \underbrace{\left(\widetilde{\mathcal{L}}(x_m)\pi_{X_m}(x_m)\right)}_{\widetilde{\pi}_{X_m|y_{\mathrm{obs}}}(x_m)} \pi_{X_\perp|X_m}(x_\perp|x_m)$$

Concentrate the numerical effort on X_m

Subspace accelerated inference

- [Vrugt 2009] Accelerating MCMC [...] with randomized **subspace sampling**, INT J NONLIN SCI NUM
- [Marzouk Najm 2009] Dimensionality reduction and PC acceleration of Bayesian inference, JCP
- [Cui Law Marzouk 2016] Dimension-independent **likelihood-informed** MCMC, JCP
- [Constantine et.al. 2016] Accelerating MCMC with **active subspaces**, SIAM-SISC
- [Izmailov et.al. 2020] **Subspace inference** for Bayesian deep learning, PMLR
- [Chen Ghattas 2020] **Projected** Stein variational gradient descent, NeurIPS
- [Brennan et.al. 2020] Greedy inference with structure-exploiting **lazy** maps, NeurIPS
- [.....]

$$X = \underbrace{U_m X_m}_{\in \mathrm{Im}(U_m)} + \underbrace{U_\perp X_\perp}_{\in \mathrm{Ker}(U_m^\top)}$$



Then

$$\widetilde{\pi}_{X|y_{\mathrm{obs}}}(x) = \underbrace{\left(\widetilde{\mathcal{L}}(x_m)\pi_{X_m}(x_m)\right)}_{\widetilde{\pi}_{X_m|y_{\mathrm{obs}}}(x_m)} \pi_{X_\perp|X_m}(x_\perp|x_m)$$

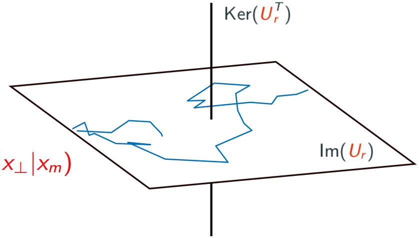
Concentrate the numerical effort on X_m

- Maximum a posteriori $\max_x \pi_{X|Y}(x|y)$ over $x \in \mathrm{Im}(U_m) + U_\perp x_\perp^0$

Subspace accelerated inference

- [Vrugt 2009] Accelerating MCMC [...] with randomized **subspace sampling**, INT J NONLIN SCI NUM
- [Marzouk Najm 2009] Dimensionality reduction and PC acceleration of Bayesian inference, JCP
- [Cui Law Marzouk 2016] Dimension-independent **likelihood-informed** MCMC, JCP
- [Constantine et.al. 2016] Accelerating MCMC with **active subspaces**, SIAM-SISC
- [Izmailov et.al. 2020] **Subspace inference** for Bayesian deep learning, PMLR
- [Chen Ghattas 2020] **Projected** Stein variational gradient descent, NeurIPS
- [Brennan et.al. 2020] Greedy inference with structure-exploiting **lazy** maps, NeurIPS
- [.....]

$$X = \underbrace{U_m X_m}_{\in \mathrm{Im}(U_m)} + \underbrace{U_\perp X_\perp}_{\in \mathrm{Ker}(U_m^\top)}$$



Then

$$\widetilde{\pi}_{X|y_{\mathrm{obs}}}(x) = \underbrace{\left(\widetilde{\mathcal{L}}(x_m)\pi_{X_m}(x_m)\right)}_{\widetilde{\pi}_{X_m|y_{\mathrm{obs}}}(x_m)} \pi_{X_\perp|X_m}(x_\perp|x_m)$$

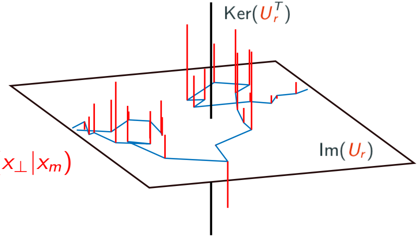
Concentrate the numerical effort on X_m

- Maximum a posteriori $\max_x \pi_{X|Y}(x|y)$ over $x \in \mathrm{Im}(U_m) + U_\perp x_\perp^0$
- MCMC to sample from $\widetilde{\pi}_{X|y}$
  1. **Subspace MCMC** to get samples $x_m^{(i)} \sim \widetilde{\pi}_{X_m|y_{\mathrm{obs}}}(x_m)$

Subspace accelerated inference

- [Vrugt 2009] Accelerating MCMC [...] with randomized **subspace sampling**, INT J NONLIN SCI NUM
- [Marzouk Najm 2009] Dimensionality reduction and PC acceleration of Bayesian inference, JCP
- [Cui Law Marzouk 2016] Dimension-independent **likelihood-informed** MCMC, JCP
- [Constantine et.al. 2016] Accelerating MCMC with **active subspaces**, SIAM-SISC
- [Izmailov et.al. 2020] **Subspace inference** for Bayesian deep learning, PMLR
- [Chen Ghattas 2020] **Projected** Stein variational gradient descent, NeurIPS
- [Brennan et.al. 2020] Greedy inference with structure-exploiting **lazy** maps, NeurIPS
- [.....]

$$X = \underbrace{U_m X_m}_{\in \mathrm{Im}(U_m)} + \underbrace{U_\perp X_\perp}_{\in \mathrm{Ker}(U_m^\top)}$$



Then

$$\widetilde{\pi}_{X|y_{\mathrm{obs}}}(x) = \underbrace{\left(\widetilde{\mathcal{L}}(x_m)\pi_{X_m}(x_m)\right)}_{\widetilde{\pi}_{X_m|y_{\mathrm{obs}}}(x_m)} \pi_{X_\perp|X_m}(x_\perp|x_m)$$

Concentrate the numerical effort on X_m

- Maximum a posteriori $\max_x \pi_{X|Y}(x|y)$ over $x \in \mathrm{Im}(U_m) + U_\perp x_\perp^0$
- MCMC to sample from $\widetilde{\pi}_{X|y}$
  1. **Subspace MCMC** to get samples $x_m^{(i)} \sim \widetilde{\pi}_{X_m|y_{\mathrm{obs}}}(x_m)$
  2. Draw samples from the **conditional prior** $x_\perp^{(i)} \sim \pi_{X_\perp|X_m}(x_\perp|x_m^{(i)})$
  3. Assemble $x^{(i)} = U_m x_m^{(i)} + U_\perp x_\perp^{(i)} \sim \widetilde{\pi}_{X|y}(x)$

Certified dimension reduction

Gaussian Linear problems: a well understood case

$$X \sim \mathcal{N}(\mu_{\text{pr}}, \Sigma_{\text{pr}})$$
$$Y = \underbrace{u(X)}_{AX} + \varepsilon,$$

where $\varepsilon \sim \mathcal{N}(0, \Sigma_{\text{obs}})$. Then

$$(X|Y = y) \ \sim \mathcal{N}(\mu_{\text{pos}}(y), \Sigma_{\text{pos}})$$

with $\mu_{\text{pos}}(y) =$ (blablabla) and

$$\boxed{\Sigma_{\text{pos}}^{-1} \ = \ \Sigma_{\text{pr}}^{-1} + \underbrace{A^{\top}\Sigma_{\text{obs}}^{-1}A}_{H}}$$

Gaussian Linear problems: a well understood case

$$X \sim \mathcal{N}(\mu_{\text{pr}}, \Sigma_{\text{pr}})$$
$$Y = \underbrace{u(X)}_{AX} + \varepsilon,$$

where $\varepsilon \sim \mathcal{N}(0, \Sigma_{\text{obs}})$. Then

$$(X|Y = y) \sim \mathcal{N}(\mu_{\text{pos}}(y), \Sigma_{\text{pos}})$$

with $\mu_{\text{pos}}(y) = $ (blablabla) and

$$\boxed{\Sigma_{\text{pos}}^{-1} \overset{(\star)}{=} \Sigma_{\text{pr}}^{-1} + \underbrace{A^\top \Sigma_{\text{obs}}^{-1} A}_{H}}$$

The prior-to-posterior **covariance contraction** is captured by H, and the most informed direction are given by

$$\min_{v \in \mathbb{R}^d} \frac{v^\top \Sigma_{\text{pos}} v}{v^\top \Sigma_{\text{pr}} v} \quad \overset{(\star)}{\Leftrightarrow} \quad \max_{v \in \mathbb{R}^d} \frac{v^\top H v}{v^\top \Sigma_{\text{pr}}^{-1} v}$$

[Flath et.al. 2011] Fast algorithms for Bayesian UQ in large-scale linear inverse problems [...], SIAM-SISC
[Bui-Thanh et.al. 2012] Extreme-scale UQ for bayesian inverse problems governed by pdes, IEEE
[Spantini et.al. 2015] Optimal low-rank approximations of Bayesian linear inverse problems, SIAM-SISC
[Polette 2025] Contributions to reduced basis methods for nonparametric Bayesian inference in geosciences, PhD
[.....]

Generalization to **nonlinear & nonGaussian** problems

$$\arg\max_{v \in \mathbb{R}^d} \frac{v^\top H v}{v^\top \Sigma_{\text{pr}}^{-1} v} \qquad \text{where } H = ???$$

Generalization to **nonlinear & nonGaussian** problems

$$\arg\max_{v\in\mathbb{R}^d} \frac{v^\top H v}{v^\top \Sigma_{\mathrm{pr}}^{-1} v} \qquad \text{where } H = ???$$

- [Cui et.al. 2016]: Replace A with $\nabla u(X)$ and take posterior expectation

$$H(y) = \int \nabla u(x) \Sigma_{\mathrm{obs}}^{-1} \nabla u(x)^\top \mathrm{d}\pi_{X|Y}(x|y)$$

Generalization to **nonlinear & nonGaussian** problems

$$\arg\max_{v\in\mathbb{R}^d} \frac{v^\top H v}{v^\top \Sigma_{\text{pr}}^{-1} v} \qquad \text{where } H = ???$$

- [Cui et.al. 2016]: Replace A with $\nabla u(X)$ and take posterior expectation

$$H(y) = \int \nabla u(x) \Sigma_{\text{obs}}^{-1} \nabla u(x)^\top \mathrm{d}\pi_{X|Y}(x|y)$$

- [Constantine et.al. 2016]: apply Active Subspace on $\log \mathcal{L}^y$ w.r.t. prior

$$H(y) = \int (\nabla \log \mathcal{L}^y)(\nabla \log \mathcal{L}^y)^\top \, \mathrm{d}\pi_X$$

Generalization to **nonlinear & nonGaussian** problems

$$\arg\max_{v\in\mathbb{R}^d} \frac{v^\top H v}{v^\top \Sigma_{\mathrm{pr}}^{-1} v} \qquad \text{where } H = ???$$

- [Cui et.al. 2016]: Replace A with $\nabla u(X)$ and take posterior expectation

$$H(y) = \int \nabla u(x) \Sigma_{\mathrm{obs}}^{-1} \nabla u(x)^\top \mathrm{d}\pi_{X|Y}(x|y)$$

- [Constantine et.al. 2016]: apply Active Subspace on $\log \mathcal{L}^y$ w.r.t. prior

$$H(y) = \int (\nabla \log \mathcal{L}^y)(\nabla \log \mathcal{L}^y)^\top \, \mathrm{d}\pi_X$$

- [Zahm et.al. 2022]: Use the **Fisher Information** matrix $\Rightarrow$ "certified" DR

$$\boxed{H(y) = \int (\nabla \log \mathcal{L}^y(x))(\nabla \log \mathcal{L}^y(x))^\top \, \mathrm{d}\pi_{X|Y}(x|y)}$$

Control the **Kullback-Leibler (KL)** [Zahm et al, 2022]

$$D_{\mathrm{KL}}(\pi_{X|y}||\widetilde{\pi}_{X|y}) = \int \log\left(\frac{\pi_{X|y}}{\widetilde{\pi}_{X|y}}\right) \mathrm{d}\pi_{X|y}, \quad \begin{cases} \pi_{X|y} \propto \mathcal{L}^y(x)\pi_X(x) \\ \widetilde{\pi}_{X|y} \propto \widetilde{\mathcal{L}}^y(U_m^\top x)\pi_X(x) \end{cases}$$

1

Control the **Kullback-Leibler (KL)** [Zahm et al, 2022]

$$D_{\mathrm{KL}}(\pi_{X|y}||\widetilde{\pi}_{X|y}) = \int \log\left(\frac{\pi_{X|y}}{\widetilde{\pi}_{X|y}}\right) \mathrm{d}\pi_{X|y}, \quad \begin{cases} \pi_{X|y} \propto \mathcal{L}^y(x)\pi_X(x) \\ \widetilde{\pi}_{X|y} \propto \widetilde{\mathcal{L}}^y(U_m^\top x)\pi_X(x) \end{cases}$$

- For any U_m, the KL-optimal profile is the **conditional expectation**

$$\boxed{\widetilde{\mathcal{L}}^y(x_m) = \mathbb{E}_{\pi_X}[\mathcal{L}^y(X)|U_m^\top X = x_m]}$$

and yields (exact marginal posterior)

$$\widetilde{\pi}_{X_m|Y}(x_m|y) = \pi_{X_m|Y}(x_m|y)$$

1

Control the **Kullback-Leibler (KL)** [Zahm et al, 2022]

$$D_{\mathrm{KL}}(\pi_{X|y}||\widetilde{\pi}_{X|y}) = \int \log\left(\frac{\pi_{X|y}}{\widetilde{\pi}_{X|y}}\right) \mathrm{d}\pi_{X|y}, \quad \begin{cases} \pi_{X|y} \propto \mathcal{L}^y(x)\pi_X(x) \\ \widetilde{\pi}_{X|y} \propto \widetilde{\mathcal{L}}^y(U_m^\top x)\pi_X(x) \end{cases}$$

- For any U_m, the KL-optimal profile is the **conditional expectation**

$$\boxed{\widetilde{\mathcal{L}}^y(x_m) = \mathbb{E}_{\pi_X}[\mathcal{L}^y(X)|U_m^\top X = x_m]}$$

  and yields (exact marginal posterior)

$$\widetilde{\pi}_{X_m|Y}(x_m|y) = \pi_{X_m|Y}(x_m|y)$$

- **Subspace LSI:** for any U_m, it holds

$$\boxed{D_{\mathrm{KL}}(\pi_{X|y}||\widetilde{\pi}_{X|y}) \leq \frac{\overline{\mathbb{C}}(\pi_X)}{2}\left(\mathbb{E}_{\pi_{X|y}}[\|\nabla \log \mathcal{L}^y\|^2] - \mathbb{E}_{\pi_{X|y}}[\|U_m^\top \nabla \log \mathcal{L}^y\|^2]\right)}$$

  where $\overline{\mathbb{C}}(\pi_X)$ is now the **subspace Logarithmic Sobolev constant**[1].

[1] Log-Sob inequality: $\mathbb{E}[u(X)^2 \log(u(X))] \leq \mathbb{C}(\pi_X)\mathbb{E}[\|u(X)\|^2]$ for all $u \geq 0$ s.t. $\mathbb{E}[u(X)] = 1$

Minimizing the RHS: PCA on $\nabla \log \mathcal{L}^y(X)$ with $X \sim \pi_{X|y}$

$$\boxed{D_{\mathrm{KL}}(\pi_{X|y}||\widetilde{\pi}_{X|y}) \leq \frac{\overline{\mathbb{C}}(\pi_X)}{2}\left(\mathbb{E}_{\pi_{X|y}}[\|\nabla \log \mathcal{L}^y\|^2] - \mathbb{E}_{\pi_{X|y}}[\|U_m^\top \nabla \log \mathcal{L}^y\|^2]\right)}$$

1. Compute the diagnostic matrix

$$H(y) = \int (\nabla \log \mathcal{L}^y)\,(\nabla \log \mathcal{L}^y)^\top \, \mathrm{d}\pi_{X|y}$$

2. Solve the eigenvalue problem $\boxed{H(y)v_i = \lambda_i v_i}$ with $\lambda_1 \geq \lambda_2 \geq \ldots \geq 0$ and assemble

$$\begin{aligned} U_m &= [v_1, \ldots, v_m] \\ U_\perp &= [v_{m+1}, \ldots, v_d] \end{aligned}$$

3. Finally we obtain

$$D_{\mathrm{KL}}(\pi_{X|y}||\widetilde{\pi}_{X|y}) \leq \frac{\overline{\mathbb{C}}(\pi_X)}{2} \sum_{i=m+1}^{d} \lambda_i$$

How to **control** $\overline{\mathbb{C}}(\pi_X)$?

- **Bakry–Émery** theorem for log-concave π_X

$$\mathrm{Hess}(-\log \pi_X(x)) \succeq \rho I_d \qquad \Rightarrow \qquad \overline{\mathbb{C}}(\pi_X) \leq 1/\rho$$

- **Holley–Stroock** perturbation lemma

$$\alpha\, \rho_X(x) \leq \pi_X(x) \leq \beta\, \rho_X(x) \qquad \Rightarrow \qquad \overline{\mathbb{C}}(\pi_X) \leq \frac{\beta}{\alpha} C(\rho_X)$$

- Combination of the two above.

$$\boxed{\{\text{these assumptions}\} \Rightarrow \{(\text{subspace})\text{LogSob}\ \} \Rightarrow \{(\text{subspace})\text{Poincaré}\}}$$

How to **control** $\overline{\mathbb{C}}(\pi_X)$?

- **Bakry–Émery** theorem for log-concave π_X

$$\mathrm{Hess}(-\log \pi_X(x)) \succeq \rho I_d \qquad \Rightarrow \qquad \overline{\mathbb{C}}(\pi_X) \leq 1/\rho$$

- **Holley–Stroock** perturbation lemma

$$\alpha\, \rho_X(x) \leq \pi_X(x) \leq \beta\, \rho_X(x) \qquad \Rightarrow \qquad \overline{\mathbb{C}}(\pi_X) \leq \frac{\beta}{\alpha} C(\rho_X)$$

- Combination of the two above.

$$\boxed{\{\text{these assumptions}\} \Rightarrow \{(\text{subspace})\text{LogSob}\ \} \Rightarrow \{(\text{subspace})\text{Poincaré}\}}$$

The **ideal setting** remains $\pi_X = \mathcal{N}(0, I_d)$. As for the (subspace)Poincaré inequality, it is preferable to **whiten** the prior

$$\overline{X} = \Sigma_{\mathrm{pr}}^{-1/2}(X - \mu_{\mathrm{pr}})$$

This is equivalent to computing the **generalized** eigenvalue problem

$$\boxed{H(y)v_i = \lambda_i \Sigma_{\mathrm{pr}}^{-1} v_i}$$

Other functional inequalities bound other divergences

$$D_{\mathrm{KL}}(\pi_{X|y}||\widetilde{\pi}_{X|y}) \leq \underbrace{\frac{\overline{\mathbb{C}}(\pi_{\overline{X}})}{2}\left(\mathbb{E}_{\pi_{X|y}}[\|\Sigma_{\mathrm{pr}}^{1/2}\nabla \log \mathcal{L}^y\|^2] - \mathbb{E}_{\pi_{X|y}}[\|U_m^\top \Sigma_{\mathrm{pr}}^{1/2}\nabla \log \mathcal{L}^y\|^2]\right)}_{=(RHS)}$$

Other functional inequalities bound other divergences

$$D_{\mathrm{KL}}(\pi_{X|y}||\widetilde{\pi}_{X|y}) \leq \underbrace{\frac{\overline{\mathbb{C}}(\pi_{\overline{X}})}{2}\left(\mathbb{E}_{\pi_{X|y}}[\|\Sigma_{\mathrm{pr}}^{1/2}\nabla \log \mathcal{L}^y\|^2] - \mathbb{E}_{\pi_{X|y}}[\|U_m^\top \Sigma_{\mathrm{pr}}^{1/2}\nabla \log \mathcal{L}^y\|^2]\right)}_{=(RHS)}$$

- (subspace)**Poincaré inequality** bounds **Hellinger distance** [Cui Tong 2022]

$$D_{\mathrm{Hell}}(\pi_{X|y}||\widetilde{\pi}_{X|y}) \leq 2(RHS)$$

Usefull *e.g.* for Laplace priors $\pi_X(x) = \prod_{i=1}^d \exp(-\lambda|x_i|)$ which do not satisfy the (subspace)LSI...

[Flock et.al. 2024] Certified coordinate selection for BIP with **Laplace prior**, Inverse Problems

Other functional inequalities bound other divergences

$$D_{\mathrm{KL}}(\pi_{X|y}||\widetilde{\pi}_{X|y}) \leq \underbrace{\frac{\overline{\mathbb{C}}(\pi_{\overline{X}})}{2}\Big(\mathbb{E}_{\pi_{X|y}}[\|\Sigma_{\mathrm{pr}}^{1/2}\nabla \log \mathcal{L}^y\|^2] - \mathbb{E}_{\pi_{X|y}}[\|U_m^\top \Sigma_{\mathrm{pr}}^{1/2}\nabla \log \mathcal{L}^y\|^2]\Big)}_{=(RHS)}$$

- (subspace)**Poincaré inequality** bounds **Hellinger distance** [Cui Tong 2022]

$$D_{\mathrm{Hell}}(\pi_{X|y}||\widetilde{\pi}_{X|y}) \leq 2(RHS)$$

Usefull *e.g.* for Laplace priors $\pi_X(x) = \prod_{i=1}^d \exp(-\lambda|x_i|)$ which do not satisfy the (subspace)LSI...

[Flock et.al. 2024] Certified coordinate selection for BIP with **Laplace prior**, Inverse Problems

- (subspace)Φ-**Sobolev inequality** bounds the α-**divergence**

$$D_\alpha(\pi_{X|y}||\widetilde{\pi}_{X|y}) \leq \mathcal{J}_\alpha\big((RHS)\big)$$

This interpolates KL ($\alpha = 1$), Hellinger ($\alpha = 1/2$) and χ^2 ($\alpha = 2$)...

[Li Marzouk Zahm 2024] Principal feature detection via Φ-**Sobolev inequalities**, Bernouli.

Other functional inequalities bound other divergences

$$D_{\mathrm{KL}}(\pi_{X|y}||\widetilde{\pi}_{X|y}) \leq \underbrace{\frac{\overline{\mathbb{C}}(\pi_{\overline{X}})}{2}\left(\mathbb{E}_{\pi_{X|y}}[\|\Sigma_{\mathrm{pr}}^{1/2}\nabla\log\mathcal{L}^y\|^2] - \mathbb{E}_{\pi_{X|y}}[\|U_m^\top\Sigma_{\mathrm{pr}}^{1/2}\nabla\log\mathcal{L}^y\|^2]\right)}_{=(RHS)}$$

- (subspace)**Poincaré inequality** bounds **Hellinger distance** [Cui Tong 2022]

$$D_{\mathrm{Hell}}(\pi_{X|y}||\widetilde{\pi}_{X|y}) \leq 2(RHS)$$

  Usefull *e.g.* for Laplace priors $\pi_X(x) = \prod_{i=1}^d \exp(-\lambda|x_i|)$ which do not satisfy the (subspace)LSI...

  [Flock et.al. 2024] Certified coordinate selection for BIP with **Laplace prior**, Inverse Problems

- (subspace)Φ-**Sobolev inequality** bounds the α-**divergence**

$$D_\alpha(\pi_{X|y}||\widetilde{\pi}_{X|y}) \leq \mathcal{J}_\alpha((RHS))$$

  This interpolates KL ($\alpha = 1$), Hellinger ($\alpha = 1/2$) and χ^2 ($\alpha = 2$)...

  [Li Marzouk Zahm 2024] Principal feature detection via Φ-**Sobolev inequalities**, Bernouli.

- **Dimensional Logarithmic Sobolev inequality** improves the above bound

$$D_{\mathrm{KL}}(\pi_{X|y}||\widetilde{\pi}_{X|y}) \leq \text{``}\log(1+(RHS))\text{''}$$

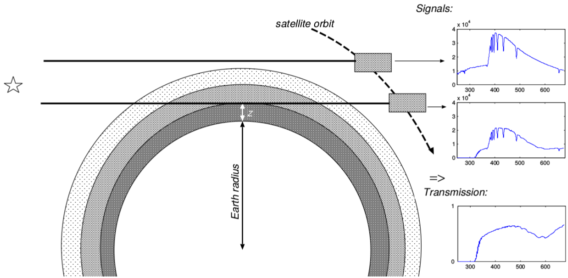
  [Li Cui Li Marzouk Zahm 2024] Sharp detection of LDS via **dimensional logSob inequalities**, arXiv.

In practice...

Benchmark Gomos: atmospheric remote sensing

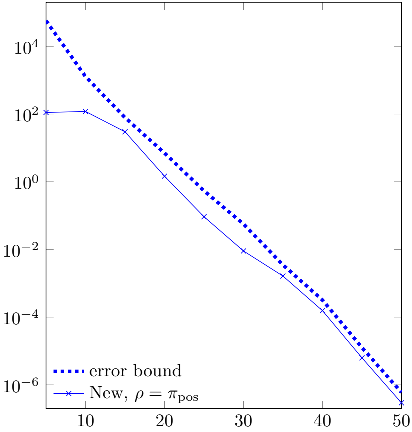
- [Haario Tamminen et al. 2004]
- Estimate gas densities $x = \rho^{\text{gas}}(z)$ from transmission spectra $y_\omega(z)$
- Beer's law:

$$y_\omega(z) = \exp\left(- \int_{\text{light path}} \sum_{\text{gas}} \alpha_\omega^{\text{gas}}(z(\zeta))\ \rho^{\text{gas}}(z(\zeta))\ \mathrm{d}\zeta \right) + \xi$$



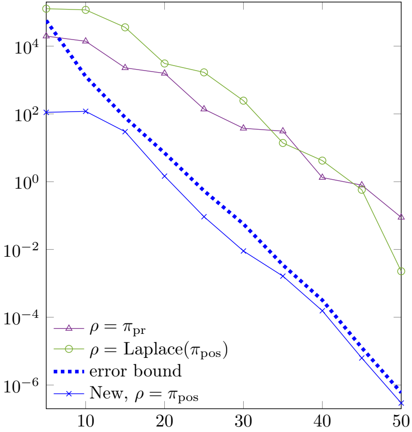
- Gaussian prior $\mathcal{N}(\mu_{\text{pr}}, \Sigma_{\text{pr}})$ with squared exponential kernel covariance
- After discretization of the atmosphere, $\dim(x) = 200$.

$$D_{\mathrm{KL}}(\pi_{X|y}||\widetilde{\pi}_{X|y}) = \mathit{function}(r)$$



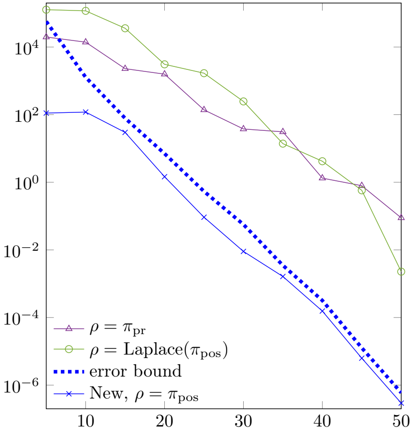
$$H(y) = \int (\nabla \log \mathcal{L}^y)(\nabla \log \mathcal{L}^y)^\top \, \mathrm{d}\pi_{X|y}$$

$$D_{\mathrm{KL}}(\pi_{X|y}||\widetilde{\pi}_{X|y}) = function(r)$$



$$H^{(\rho)}(y) = \int (\nabla \log \mathcal{L}^y)(\nabla \log \mathcal{L}^y)^\top \, \mathrm{d}\rho$$

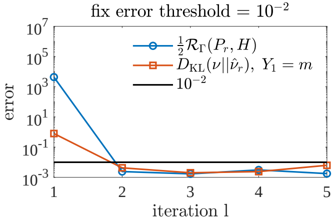
$$D_{\mathrm{KL}}(\pi_{X|y} || \widetilde{\pi}_{X|y}) = \mathit{function}(r)$$

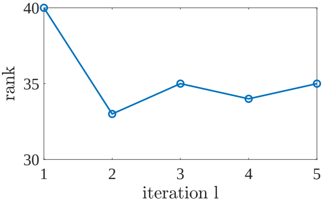


$$H^{(\rho)}(y) = \int (\nabla \log \mathcal{L}^y)(\nabla \log \mathcal{L}^y)^\top \, \mathrm{d}\rho$$

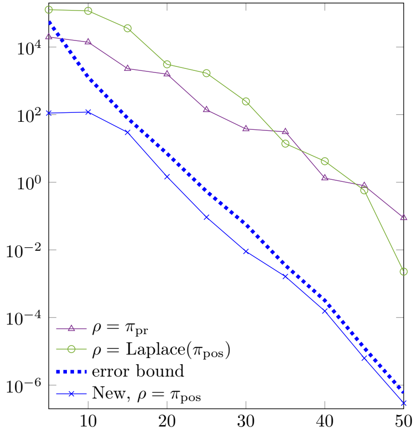
Iterative procedure:

$$\pi_X \to H^{(\pi_X)} \to \widetilde{\pi}_{X|y} \to H^{(\widetilde{\pi}_{X|y})} \to \dots$$





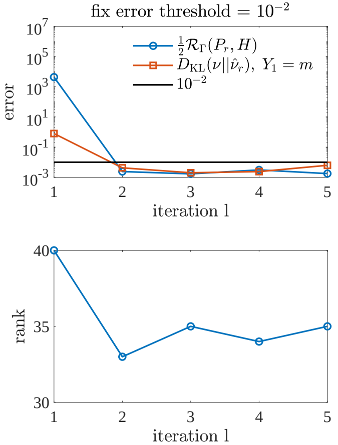
$$D_{\mathrm{KL}}(\pi_{X|y}||\widetilde{\pi}_{X|y}) = \mathit{function}(r)$$



$$H^{(\rho)}(y) = \int (\nabla \log \mathcal{L}^y)(\nabla \log \mathcal{L}^y)^\top \, \mathrm{d}\rho$$

Iterative procedure:

$$\pi_X \to H^{(\pi_X)} \to \widetilde{\pi}_{X|y} \to H^{(\widetilde{\pi}_{X|y})} \to \ldots$$



We have to re-run the whole procedure for every new piece of observed data y...

Data-free dimension reduction $U_m = \cancel{U_m(y)}$

Problem statement

Recall that, in the Bayesian perspective, the observed data y is a **realization** of a random variable

$$Y \sim \pi_{\text{data}}$$

Objective

Find a $U_m = \cancel{U_m(Y)}$ such that

$$D_{\text{KL}}(\pi_{X|Y} || \widetilde{\pi}_{X|Y}) \leq tol \tag{1}$$

in high probability (w.r.t. Y).

By Markov inequality,

$$\mathbb{E}_Y \left[D_{\text{KL}}(\pi_{X|Y} || \widetilde{\pi}_{X|Y}) \right] \leq \varepsilon$$

is sufficient to ensure (1) with probability greater than $1 - \varepsilon / tol$.

Problem statement

Recall that, in the Bayesian perspective, the observed data y is a **realization** of a random variable

$$Y \sim \pi_{\text{data}}$$

Objective

Find a $U_m = \cancel{U_m(Y)}$ such that

$$D_{\text{KL}}(\pi_{X|Y} || \widetilde{\pi}_{X|Y}) \leq tol \tag{1}$$

in high probability (w.r.t. Y).

By Markov inequality,

$$\mathbb{E}_Y \left[D_{\text{KL}}(\pi_{X|Y} || \widetilde{\pi}_{X|Y}) \right] \leq \varepsilon$$

is sufficient to ensure (1) with probability greater than $1 - \varepsilon / tol$.

Remark: this corresponds to controlling the **Conditional Mutual Information**

$$CMI(Y, X_\perp | X_r) \leq \varepsilon$$

Online / Offline procedure

Offline

1. Compute
$$H = \mathbb{E}\big(H(Y)\big), \quad Y \sim \pi_Y$$
2. Solve the generalized eigenvalue problem $H u_i = \lambda_i \Sigma_{\mathrm{pr}}^{-1} u_i$ and let
$$U_m = [u_1, \dots, u_m] \in \mathbb{R}^{d\times m}$$

Online

3. Receive a realization y of Y,
4. Compute the optimal function $\widetilde{\mathcal{L}}^y(x_m) = \mathbb{E}_{\pi_X}\big(\mathcal{L}^y(X) \big| U_m^\top X = x_m\big)$
5. Assemble the posterior approximation $\widetilde{\pi}_{X|y}(x) \propto \widetilde{\mathcal{L}}^y(U_m^T x)\pi_X(x)$

Proposition [Cui et.al. 2021]

Assume $\overline{\mathbb{C}}(\pi_X) < \infty$. The above procedure yields

$$\mathbb{E}\Big[D_{\mathrm{KL}}(\pi_{X|Y} || \widetilde{\pi}_{X|Y})\Big] \leq \frac{\overline{\mathbb{C}}(\pi_{\overline{X}})}{2} \sum_{i=m+1}^{d} \lambda_i$$

Online / Offline procedure

Offline

1. Compute

$$H = \mathbb{E}\big(H(Y)\big), \quad Y \sim \pi_Y$$

2. Solve the generalized eigenvalue problem $H u_i = \lambda_i \Sigma_{\text{pr}}^{-1} u_i$ and let

$$U_m = [u_1, \ldots, u_m] \in \mathbb{R}^{d\times m}$$

Online

3. Receive a realization y of Y,
4. Compute the optimal function $\widetilde{\mathcal{L}}^y(x_m) = \mathbb{E}_{\pi_X}\left(\mathcal{L}^y(X) \middle| U_m^\top X = x_m\right)$
5. Assemble the posterior approximation $\widetilde{\pi}_{X|y}(x) \propto \widetilde{\mathcal{L}}^y(U_m^T x)\pi_X(x)$

Proposition [Cui et.al. 2021]

Assume $\overline{\mathbb{C}}(\pi_X) < \infty$. The above procedure yields

$$\mathbb{E}\Big[D_{\text{KL}}(\pi_{X|Y} || \widetilde{\pi}_{X|Y})\Big] \leq \frac{\overline{\mathbb{C}}(\pi_{\overline{X}})}{2} \sum_{i=m+1}^{d} \lambda_i$$

If **Gaussian prior**, "dimensional log-Sob" improvement [Li et.al. 2024]

$$\mathbb{E}\Big[D_{\text{KL}}(\pi_{X|Y} || \widetilde{\pi}_{X|Y})\Big] \leq \frac{1}{2} \sum_{i=m+1}^{d} \log(1 + \lambda_i)$$

How to compute $H = \mathbb{E}\big[H(Y)\big]$?

Proposition

$$H = \int \mathcal{I}(x)\,\mathrm{d}\pi_X(x)$$

where $\mathcal{I}(x)$ is the **Fisher information matrix** of the likelihood $\mathcal{L}^y(x) \propto \pi(y|x)$ defined by

$$\mathcal{I}(x) = \int \nabla \log \mathcal{L}^y(x) \nabla \log \mathcal{L}^y(x)^T \,\pi(y|x)\mathrm{d}y$$

How to compute $H = \mathbb{E}[H(Y)]$?

Proposition

$$H = \int \mathcal{I}(x)\, \mathrm{d}\pi_X(x)$$

where $\mathcal{I}(x)$ is the **Fisher information matrix** of the likelihood $\mathcal{L}^y(x) \propto \pi(y|x)$ defined by

$$\mathcal{I}(x) = \int \nabla \log \mathcal{L}^y(x) \nabla \log \mathcal{L}^y(x)^T \pi(y|x)\mathrm{d}y$$

- **Gaussian likelihood:** $\mathcal{L}^y(x) = \exp(-\frac{1}{2}\|u(x) - y\|^2_{\Sigma_{\mathrm{obs}}^{-1}})$

$$\boxed{H = \int \nabla u(x)^T \Sigma_{\mathrm{obs}}^{-1} \nabla u(x)\, \mathrm{d}\pi_X(x)}$$

This corresponds to the **Active Subspace** on the forward model u!

How to compute $H = \mathbb{E}\big[H(Y)\big]$?

Proposition

$$H = \int \mathcal{I}(x)\,\mathrm{d}\pi_X(x)$$

where $\mathcal{I}(x)$ is the **Fisher information matrix** of the likelihood $\mathcal{L}^y(x) \propto \pi(y|x)$ defined by

$$\mathcal{I}(x) = \int \nabla \log \mathcal{L}^y(x) \nabla \log \mathcal{L}^y(x)^T \, \pi(y|x)\mathrm{d}y$$

- **Gaussian likelihood:** $\mathcal{L}^y(x) = \exp(-\frac{1}{2}\|u(x) - y\|^2_{\Sigma_{\text{obs}}^{-1}})$

$$\boxed{H = \int \nabla u(x)^T \Sigma_{\text{obs}}^{-1} \nabla u(x)\,\mathrm{d}\pi_X(x)}$$

  This corresponds to the **Active Subspace** on the forward model u!

- **Poisson likelihood:** $\mathcal{L}^y(x) = \prod_{i=1}^m \frac{G_i(x)^{y_i}\exp(-G_i(x))}{y_i!}$

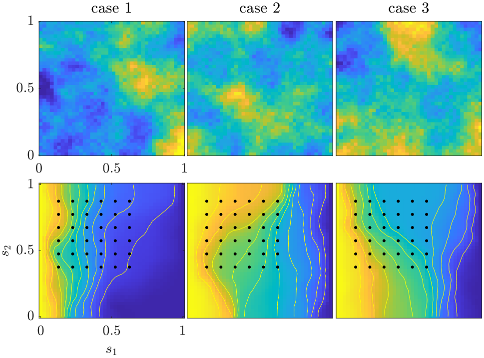
$$\boxed{H = \int \nabla u(x)^T \mathrm{diag}\Big(u_1(x), \ldots, u_m(x)\Big)^{-1} \nabla u(x)^T \,\mathrm{d}\pi_X(x)}$$

- ...

A numerical example: elliptic PDE

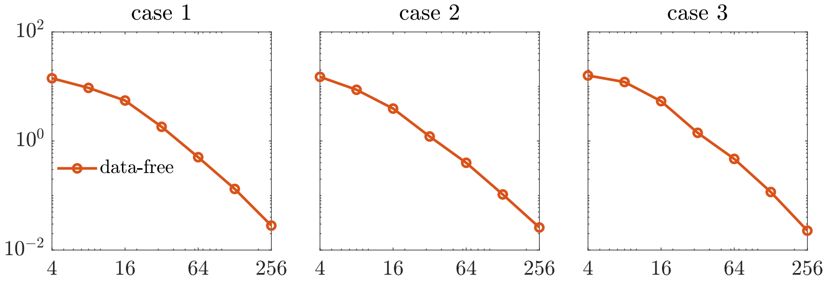
$$-\nabla \cdot \Big(\kappa(s)\nabla p(s)\Big) = f(s), \quad s \in [0,1]^2$$

- Parameter: $x = \log \kappa$
- Data: $y = (p(s_1), \dots, p(s_m)) + \mathcal{N}(0, \Gamma_{\text{obs}})$ (Gaussian likelihood)
- Gaussian prior: $-\Delta x + \gamma x = \mathcal{W}$ with $\mathcal{W}$ = white noise and $\gamma = 10$



Elliptic PDE: results for three different observations $Y^{(1)}, Y^{(2)}, Y^{(3)}$

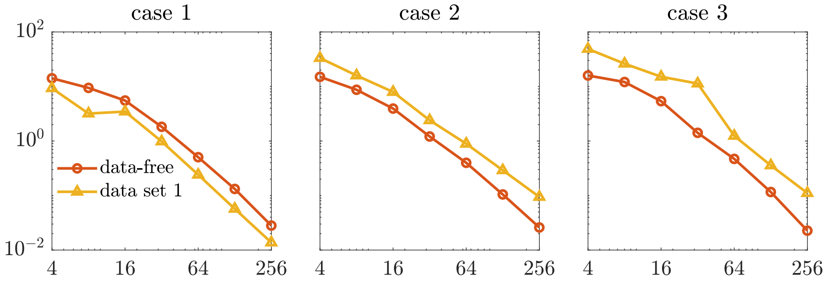
$$D_{\mathrm{KL}}(\pi_{X|Y^{(i)}} || \widetilde{\pi}_{X|Y^{(i)}}) = \mathit{function}(m)$$



- data-free: U_m computed via $H = \mathbb{E}(H(Y))$

Elliptic PDE: results for three different observations $Y^{(1)}, Y^{(2)}, Y^{(3)}$

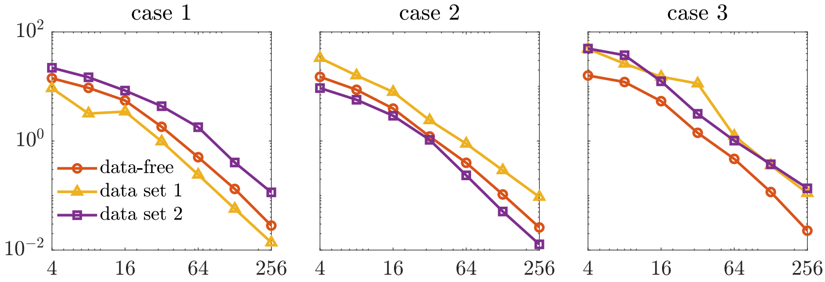
$$D_{\mathrm{KL}}(\pi_{X|Y^{(i)}}||\widetilde{\pi}_{X|Y^{(i)}}) = \mathit{function}(m)$$



- data-free: U_m computed via $H = \mathbb{E}(H(Y))$
- data set 1: U_m computed via $H(Y^{(1)})$

Elliptic PDE: results for three different observations $Y^{(1)}, Y^{(2)}, Y^{(3)}$

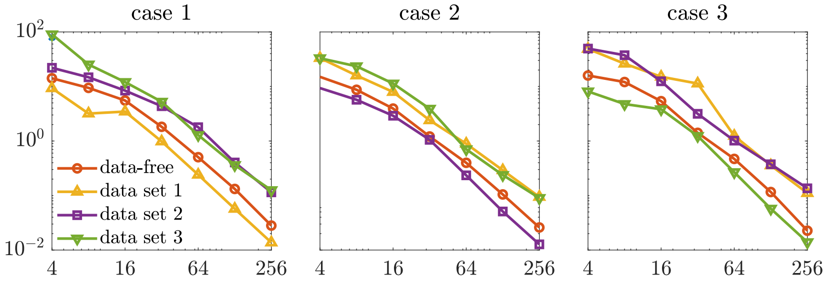
$$D_{\mathrm{KL}}(\pi_{X|Y^{(i)}} || \widetilde{\pi}_{X|Y^{(i)}}) = \mathit{function}(m)$$



- data-free: U_m computed via $H = \mathbb{E}(H(Y))$
- data set 1: U_m computed via $H(Y^{(1)})$
- data set 2: U_m computed via $H(Y^{(2)})$

Elliptic PDE: results for three different observations $Y^{(1)}, Y^{(2)}, Y^{(3)}$

$$D_{\mathrm{KL}}(\pi_{X|Y^{(i)}} || \widetilde{\pi}_{X|Y^{(i)}}) = \mathit{function}(m)$$

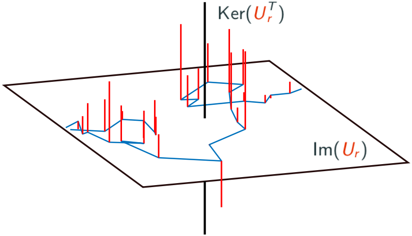


- data-free: U_m computed via $H = \mathbb{E}(H(Y))$
- data set 1: U_m computed via $H(Y^{(1)})$
- data set 2: U_m computed via $H(Y^{(2)})$
- data set 3: U_m computed via $H(Y^{(3)})$

Sample from the approximate posterior

$$\widetilde{\pi}_{X|y}(x) = \pi_{X_m|y}(x_m)\ \pi_X(x_\perp|x_m)$$

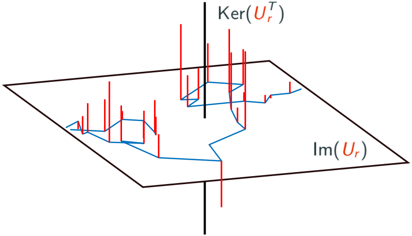
1. Marginal posterior $x_m^{(i)} \sim \pi_{X_m|y}(x_m)$
2. Conditional prior $x_\perp^{(i)} \sim \pi_X(x_\perp|x_m^{(i)})$
3. Assemble $U_m x_m^{(i)} + U_\perp x_\perp^{(i)} \sim \widetilde{\pi}_{X|y}(x)$



Sample from the approximate posterior

$$\widetilde{\pi}_{X|y}(x) = \pi_{X_m|y}(x_m)\ \pi_X(x_\perp|x_m)$$

1. Marginal posterior $x_m^{(i)} \sim \pi_{X_m|y}(x_m)$
2. Conditional prior $x_\perp^{(i)} \sim \pi_X(x_\perp|x_m^{(i)})$
3. Assemble $U_m x_m^{(i)} + U_\perp x_\perp^{(i)} \sim \widetilde{\pi}_{X|y}(x)$



Pseudo-Marginal MCMC [Andrieu, Roberts 2009] to sample $\pi_{X_m|y}(x_m)$

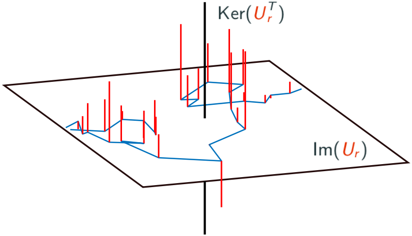
$$\pi_{X_m|y}(x_m) = \int \frac{\pi_{X|y}(x_m + \widetilde{x}_\perp)}{\pi_X(\widetilde{x}_\perp|x_m)} \pi_X(\widetilde{x}_\perp|x_m)\mathrm{d}\widetilde{x}_\perp \quad \approx \quad \frac{1}{N}\sum_{i=1}^{N} \frac{\pi_{X|y}(x_m + \widetilde{x}_\perp^{(i)})}{\pi_X(\widetilde{x}_\perp^{(i)}|x_m)}$$

- **Low-variance** estimator by construction of U_m (N can be small)
- Pseudo-Marginal trick: redraw $\widetilde{x}_\perp^{(i)} \sim \pi_X(x_\perp|x_m)$ **at every iteration**.

Sample from the ~~approximate~~ exact posterior

$$\widetilde{\pi}_{X|y}(x) = \pi_{X_m|y}(x_m)\ \pi_X(x_\perp|x_m)$$

1. Marginal posterior $x_m^{(i)} \sim \pi_{X_m|y}(x_m)$
2. ~~Conditional prior~~ $\cancel{x_\perp^{(i)} \sim \pi_X(x_\perp|x_m^{(i)})}$
3. Assemble $U_m x_m^{(i)} + U_\perp x_\perp^{(i)} \sim \widetilde{\pi}_{X|y}(x)$



Pseudo-Marginal MCMC [Andrieu, Roberts 2009] to sample $\pi_{X_m|y}(x_m)$

$$\pi_{X_m|y}(x_m) = \int \frac{\pi_{X|y}(x_m + \widetilde{x}_\perp)}{\pi_X(\widetilde{x}_\perp|x_m)} \pi_X(\widetilde{x}_\perp|x_m)\mathrm{d}\widetilde{x}_\perp \quad \approx \quad \frac{1}{N}\sum_{i=1}^{N} \frac{\pi_{X|y}(x_m + \widetilde{x}_\perp^{(i)})}{\pi_X(\widetilde{x}_\perp^{(i)}|x_m)}$$

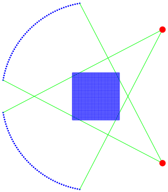
- **Low-variance** estimator by construction of U_m (N can be small)
- Pseudo-Marginal trick: redraw $\widetilde{x}_\perp^{(i)} \sim \pi_X(x_\perp|x_m)$ **at every iteration**.

Recycle $\widetilde{x}_\perp^{(i)}$ to sample the exact full posterior $\pi_{X|y}(x)$ [Cui Zahm 2021]

Instead of drawing $x_\perp^{(i)} \sim \pi_X(x_\perp|x_m^{(i)})$, pick $x_\perp^{(i)} \in \{\widetilde{x}_\perp^{(1)}, \ldots, \widetilde{x}_\perp^{(N)}\}$ at random.

A numerical example: X-ray tomography with Poisson data

Identify the density of a material in a domain of interest (blue square) using two X-ray sources (red points) and $m = 100$ sensors (blue points)



- Data: $Y \in \mathbb{N}^m$ integer-valued vector (number of incident photons)
- **Poisson likelihood** of the form

$$\mathcal{L}^y(x) = \prod_{i=1}^{m} \frac{G_i(x)^{y_i} \exp(-G_i(x))}{y_i!}$$

  where the forward model $G(x)$ stems from Beer's law.
- **Laplace prior**

$$\pi_X(x) \propto \prod_{i=1}^{d=64^2} \exp(-\lambda |x_i|)$$

- We use **coordinate selection** to reduce the dimension.

MCMC using H-MALA[2] as a baseline to sample the exact posterior

We use **Integrated Auto Correlation Time (IACT)** to measure the mixing performances of the MCMC.

		IACT	$\sqrt{\mathrm{var}}[\log \widetilde{\mathcal{L}}_N^y]$
$N=2$	$r=16$	85.1±2.7	1.54±0.02
	$r=32$	54.1±3.1	0.61±.007
	$r=48$	49.4±2.6	0.45±.002
$N=5$	$r=16$	60.0±6.2	0.93±.006
	$r=32$	47.6±2.5	0.39±.004
	$r=48$	46.5±1.4	0.29±.001

IACT of the full-dimensional H-MALA: 95.9 ± 3.3

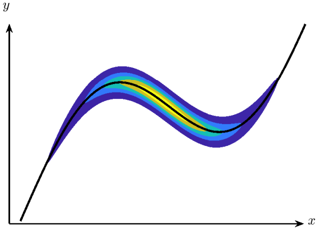
[2] Hessian-preconditioned Metropolis-Adjusted Langevin Algorithm

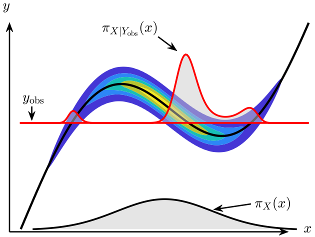
Joint density estimation for fast Bayesian inversion

Objective

Build an approximation of the **joint density** $\widetilde{\pi}_{XY} \approx \pi_{XY}$ which **allows fast conditioning** on $Y = y_{obs}$

$$\widetilde{\pi}_{X|y_{obs}}(x) \propto \widetilde{\pi}_{XY}(x, y_{obs})$$





[Cui, Dolgov, Zahm 2023] Scalable conditional deep inverse Rosenblatt transports using tensor-trains and **gradient-based dimension reduction**, JCP

Approximation of π_{XY} using **low-rank tensor formats**

$$\widetilde{\pi}_{XY}(x,y) \propto \qquad g^{Y}(y) g^{X}(x)$$

Approximation of π_{XY} using **low-rank tensor formats**

$$\widetilde{\pi}_{XY}(x,y) \propto \quad \sum_{i=1}^{r} g_i^Y(y) g_i^X(x)$$

Approximation of π_{XY} using **low-rank tensor formats**

$$\widetilde{\pi}_{XY}(x,y) \propto \left(\sum_{i=1}^{r} g_i^{Y}(y) g_i^{X}(x) \right)^2 \geq 0$$

Approximation of π_{XY} using **low-rank tensor formats**

$$\widetilde{\pi}_{XY}(x,y) \propto \left(\sum_{i=1}^{r} g_i^Y(y) g_i^X(x) \right)^2 \geq 0$$

Generalization to higher dimensions $x = (x_1, \ldots, x_d)$ and $y = (y_1, \ldots, y_p)$:

- The **canonical** tensor format

$$\widetilde{\pi}_{XY}(x,y) \propto \left(\sum_{i=1}^{r} g_i^{Y_p}(y_p) \ldots g_i^{Y_1}(y_1)\; g_i^{X_1}(x_1) \ldots g_i^{X_d}(x_d) \right)^2$$

Approximation of π_{XY} using **low-rank tensor formats**

$$\widetilde{\pi}_{XY}(x,y) \propto \left(\sum_{i=1}^{r} g_i^Y(y) g_i^X(x) \right)^2 \geq 0$$

Generalization to higher dimensions $x = (x_1, \dots, x_d)$ and $y = (y_1, \dots, y_p)$:

- The **canonical** tensor format

$$\widetilde{\pi}_{XY}(x,y) \propto \left(\sum_{i=1}^{r} g_i^{Y_p}(y_p) \dots g_i^{Y_1}(y_1) \; g_i^{X_1}(x_1) \dots g_i^{X_d}(x_d) \right)^2$$

- The **tensor train (TT)** format [Oseledets2011]

$$\widetilde{\pi}_{XY}(x,y) \propto \left(\sum_{i_1, i_2, \dots, i_{d-1}=1}^{r_1, r_2, \dots, r_{d-1}} g_{i_1}^{Y_p}(y_p) g_{i_1, i_2}^{Y_{p-1}}(y_{p-1}) \dots g_{i_{d-2}, i_{d-1}}^{X_{d-1}}(x_{d-1}) g_{i_{d-1}}^{X_d}(x_d) \right)^2$$

where the g's are **matrix valued functions**. Better approximation properties compared to the canonical format, but it is **sensitive to variable ordering**.

Tensor Train approximation

The **Hellinger distance** satisfies

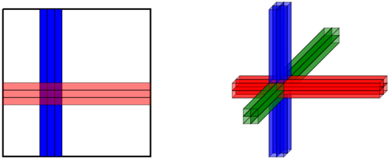
$$D_{\mathrm{Hell}}(\pi_{XY}, \widetilde{\pi}_{XY})^2 := \frac{1}{2}\int (\sqrt{\pi_{XY}} - \sqrt{\widetilde{\pi}_{XY}})^2 \mathrm{d}xy$$

$$\leq 2 \int \left(\sqrt{\pi_{XY}} - \sum_{i_1,\ldots,i_{d-1}=1}^{r_1,\ldots,r_{d-1}} g_{i_1}^{Y_m} g_{i_1,i_2}^{Y_{m-1}} \ldots g_{i_{d-2},i_{d-1}}^{X_{d-1}} g_{i_{d-1}}^{X_d} \right)^2 \mathrm{d}xy$$

Minimizing the RHS requires solving a standard **least-square problem** to approximate $\sqrt{\pi_{XY}}$.

Tensor Train approximation

- Construct the g's using **adaptive cross approximation** algorithms [Dolgov et al, 2014]. This requires evaluating *fibers* of π_{XY}:



- The ranks $r = (r_1, r_2, \ldots, r_{d-1})$ can be **automatically adapted** to reach a prescribed tolerance ε
- Complexity of $\mathcal{O}((d+p)\, r_{\max}^2)$ memory and $\mathcal{O}((d+p)\, r_{\max}^3)$ flops: **linear in the dimension**, but **cubic in the ranks**.
- The efficiency of tensor trains (*i.e.* how large is $r = r(\varepsilon)$) is sensitive to the **order of the variables**.

Parameter & data **rotation**: $X \leftarrow U^\top X$ and $Y \leftarrow V^\top Y$

$$\boxed{X = \underbrace{U_m X_m}_{\text{informed}} + U_\perp X_\perp \qquad Y = \underbrace{V_s Y_s}_{\text{informative}} + V_\perp Y_\perp}$$

Parameter & data **rotation**: $X \leftarrow U^\top X$ and $Y \leftarrow V^\top Y$

$$\boxed{X = \underbrace{U_m X_m}_{\text{informed}} + U_\perp X_\perp \qquad Y = \underbrace{V_s Y_s}_{\text{informative}} + V_\perp Y_\perp}$$

- **Idea:** approximate $\pi_{XY}(x,y) \propto \exp\left(-\frac{1}{2}\|u(x)-y\|^2\right)\pi_X(x)$ with

$$\widetilde{\pi}_{XY}(x,y) = \underbrace{\pi_{Y_\perp|Y_s X}(y_\perp|y_s, x_0)}_{\text{conditional likelihood}} \underbrace{\pi_{X_m Y_s}(x_m, y_s)}_{\text{marginal joint}} \underbrace{\pi_{X_\perp|X_m}(x_\perp|x_m)}_{\text{conditional prior}}$$

Applying the **subspace Poincaré inequalities**, we get

$$\boxed{D_{\mathrm{Hell}}(\pi_{XY}||\widetilde{\pi}_{XY}) \leq \overline{\mathbb{C}}(\pi_X)\left(\mathbb{E}_{\pi_X}[\|U_\perp^\top \nabla u(X)^\top\|_F^2] + \mathbb{E}_{\pi_X}[\|V_\perp^\top(u(X)-u(x_0))\|^2]\right)}$$

Parameter & data **rotation**: $X \leftarrow U^\top X$ and $Y \leftarrow V^\top Y$

$$\boxed{X = \underbrace{U_m X_m}_{\text{informed}} + U_\perp X_\perp \qquad Y = \underbrace{V_s Y_s}_{\text{informative}} + V_\perp Y_\perp}$$

- **Idea:** approximate $\pi_{XY}(x,y) \propto \exp\left(-\frac{1}{2}\|u(x)-y\|^2\right)\pi_X(x)$ with

$$\widetilde{\pi}_{XY}(x,y) = \underbrace{\pi_{Y_\perp|Y_s X}(y_\perp|y_s,x_0)}_{\text{conditional likelihood}}\underbrace{\pi_{X_m Y_s}(x_m,y_s)}_{\text{marginal joint}}\underbrace{\pi_{X_\perp|X_m}(x_\perp|x_m)}_{\text{conditional prior}}$$

  Applying the **subspace Poincaré inequalities**, we get

$$\boxed{D_{\text{Hell}}(\pi_{XY}||\widetilde{\pi}_{XY}) \leq \overline{\mathbb{C}}(\pi_X)\left(\mathbb{E}_{\pi_X}[\|U_\perp^\top \nabla u(X)^\top\|_F^2] + \mathbb{E}_{\pi_X}[\|V_\perp^\top(u(X)-u(x_0))\|^2]\right)}$$

- **Minimizing the bound** yields U, V containing the first eigenvectors of

$$H_X = \mathbb{E}_{\pi_X}\left[\nabla u(X)^\top \nabla u(X)\right] \qquad \text{(active subspace)}$$

$$H_Y = \mathbb{E}_{\pi_X}\left[u(X)u(X)^\top\right] + \text{rank-1} \qquad \text{(PCA on } u(X)\text{)}$$

- We propose to **rotate** the parameter & data as follow

$$X \leftarrow U^\top X \qquad Y \leftarrow V^\top Y$$

Illustration: elliptic PDE with Besov prior

$$-\mathrm{div}(\kappa \nabla u) = f \qquad \text{on } \Omega = [0,1]^2$$

Parameter: $\log \kappa = \sum_{i=1}^{d} x_i \psi_i$ with Besov prior $(d = 22)$

$$\begin{cases} (\psi_1, \ldots, \psi_d) : \text{wavelet basis} \\ x \sim \pi_X(x) \propto \exp(-\gamma \|x\|_1) \end{cases}$$

Data: noisy observations $(m = 16)$

$$y = u(\text{16 black points}) + \mathcal{N}(0, \Sigma_{\text{obs}})$$

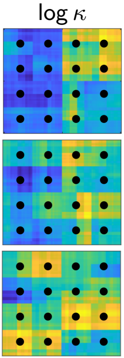


Illustration: elliptic PDE with Besov prior

$$-\mathrm{div}(\kappa \nabla u) = f \quad \text{on } \Omega = [0,1]^2$$

Parameter: $\log \kappa = \sum_{i=1}^{d} x_i \psi_i$ with Besov prior $\qquad (d = 22)$

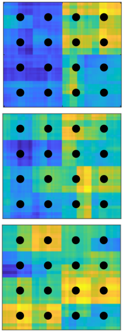
$$\begin{cases} (\psi_1, \ldots, \psi_d) : \text{wavelet basis} \\ x \sim \pi_X(x) \propto \exp(-\gamma \|x\|_1) \end{cases}$$

Data: noisy observations $\qquad (m = 16)$

$$y = u(\text{16 black points}) + \mathcal{N}(0, \Sigma_{\text{obs}})$$

Tensor Train approximation of π_{XY} using adaptive rank (fixed precision)

log κ



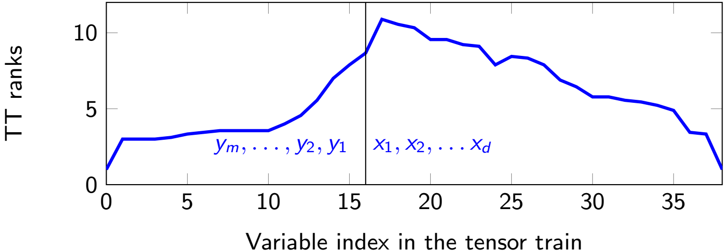


Illustration: elliptic PDE with Besov prior

$$-\mathrm{div}(\kappa \nabla u) = f \quad \text{on } \Omega = [0,1]^2$$

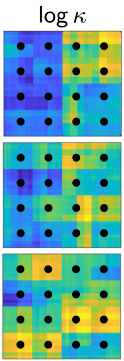
Parameter: $\log \kappa = \sum_{i=1}^{d} x_i \psi_i$ with Besov prior $(d = 22)$

$$\begin{cases} (\psi_1, \ldots, \psi_d) : \text{wavelet basis} \\ x \sim \pi_X(x) \propto \exp(-\gamma \|x\|_1) \end{cases}$$

Data: noisy observations $(m = 16)$

$$y = u(\text{16 black points}) + \mathcal{N}(0, \Sigma_{\text{obs}})$$

Tensor Train approximation of π_{XY} using adaptive rank (fixed precision)



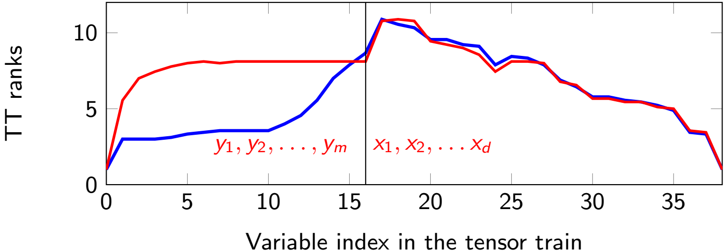


Illustration: elliptic PDE with Besov prior

$$-\mathrm{div}(\kappa \nabla u) = f \quad \text{on } \Omega = [0,1]^2$$

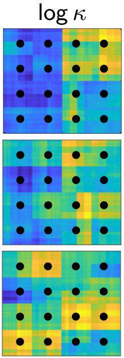
Parameter: $\log \kappa = \sum_{i=1}^{d} x_i \psi_i$ with Besov prior $\qquad (d = 22)$

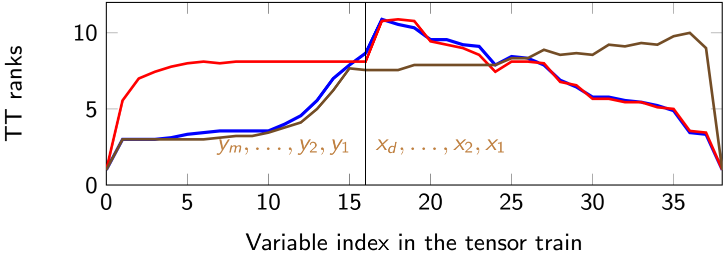
$$\begin{cases} (\psi_1, \ldots, \psi_d) : \text{wavelet basis} \\ x \sim \pi_X(x) \propto \exp(-\gamma \|x\|_1) \end{cases}$$

Data: noisy observations $\qquad (m = 16)$

$$y = u(\text{16 black points}) + \mathcal{N}(0, \Sigma_{\text{obs}})$$

Tensor Train approximation of π_{XY} using adaptive rank (fixed precision)





Conclusion

Certified likelihood-informed subspace

$$\widetilde{\pi}_{X|Y}(x|y) \propto \widetilde{\mathcal{L}}^y(U_m^\top x)\pi_X(x)$$

Gradient-based methods: (1) use functional inequalities to bound the error $\widetilde{\pi}_{X|Y} \approx \pi_{X|Y}$ and (2) minimize that bound over U_m using the diagnostic matrix

$$H(y) = \int (\nabla \log \mathcal{L}^y)(\nabla \log \mathcal{L}^y)^\top \, d\pi_{X|y}$$

Many extensions:

- **Data-Free:** reduce X before observing the data
- **Parameter & Data reduction:** informed parameter & informative data
- **"Heavy" tail (Laplace) priors:** can we still use Log-Sob/Poincaré?
- **Rare event estimation:** nonsmooth likelihood: $\pi_X^\alpha(x) \propto 1_{u(x)\geq\alpha}\pi_X(x)$

[Zahm et.al. 2022] **Certified** dimension reduction in nonlinear Bayesian inverse problems, Math.Comp.
[Cui Zahm 2021] **Data-Free** Likelihood-Informed DR of Bayesian Inverse Problems, Inverse Problems.
[Cui Tong 2022] A unified performance analysis of likelihood informed subspace methods, Bernouli.
[Li Marzouk Zahm 2024] Principal feature detection via **Φ-Sobolev inequalities**, Bernouli.
[Li Cui Li Marzouk Zahm 2024] Sharp detection of LDS via **dimensional logSob inequalities**, IMA-II.
[Flock et.al. 2024] Certified coordinate selection for BIP with **Laplace prior**, Inverse Problems

If you didn't get enough today...

- Extention to **Conditional Mutual Independence (CMI)** to detect independence between X and Y given Z

$$\mathrm{CMI}(X, Y|Z) \leq \overline{C}(\pi_{XYZ})^2 \mathbb{E}[\|\nabla_X \nabla_Y \log \pi_{XYZ}(XYZ)\|_F^2]$$

[Baptista et.al. 2022] Gradient-based data and parameter dimension reduction, arXiv
[Baptista et.al. 2024] Learning Non-Gaussian Graphical Models [...], JMLR

- Bound the **Expectation Information Gain (EIG)** for goal-oriented sensor placement

$$\begin{aligned} \mathrm{EIG}(Y_\tau|X_r) &= \mathbb{E}_Y[\mathrm{D_{KL}}(\pi_{X_r|Y_\tau}||\pi_{X_r})] \\ &\geq \mathbb{E}_Y[\mathrm{D_{KL}}(\pi_{X|Y}||\pi_X)] - \overline{C}(\pi_X)\mathbb{E}[\|V_\tau^\top \nabla u(X) U_r\|_F^2] \end{aligned}$$

[Chen et.al. 2025] Coupled **input-output** dimension reduction [...], SIAM-SISC

Questions?