Data-free dimension reduction for Bayesian inverse problems

In collaboration with Tiangang Cui (Monash University)

Olivier Zahm
INRIA - AIRSEA project team

olivier.zahm@inria.fr
https://team.inria.fr/airsea/

SIAM-CSE,
Fort Worth, March 4, 2021.

Bayesian inverse problem

- Recover an **unknown parameter** x from the **noisy observation** of a forward model $G(x)$. For instance

$$y = G(x) + \xi \qquad \text{with} \qquad \xi \sim \mathcal{N}(0, \Gamma_{\text{obs}}).$$

- The distribution of $x|y$ is the **posterior distribution**

$$\underbrace{\pi_{\text{pos}}^{y}(x)}_{\text{posterior}} \propto \underbrace{\mathcal{L}_y(x)}_{\text{likelihood}} \underbrace{\pi_{\text{pr}}(x)}_{\text{prior}} \qquad \text{with} \qquad \mathcal{L}_y(x) = \mathbb{P}(y|x)$$

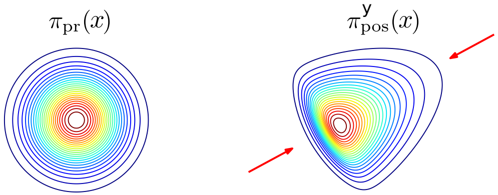
  - Draw samples $x \sim \pi_{\text{pos}}^{y}$
  - Find the MAP estimate $\max_x \pi_{\text{pos}}^{y}(x)$
  - Compute an expectation over posterior $\int h(x) \, \mathrm{d}\pi_{\text{pos}}^{y}(x)$

Curse of dimensionality

$$x = (x_1, \dots, x_d) \in \mathbb{R}^d$$

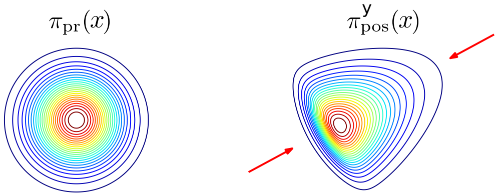
Standard algorithms suffer when $d \gg 1$ (slow convergence, complexity blows up...)

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 posterior distribution is *close* to

$$\widetilde{\pi}^{y}_{\mathrm{pos}}(x) \propto \widetilde{\mathcal{L}}(U_r^T x)\pi_{\mathrm{pr}}(x)$$

for some **positive function** $\widetilde{\mathcal{L}} = \widetilde{\mathcal{L}}_y : \mathbb{R}^r \to \mathbb{R}_+$ and some **matrix** $U_r \in \mathbb{R}^{d\times r}$ with rank r:

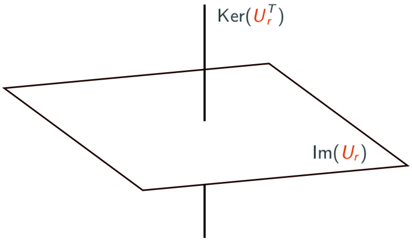
$$\mathbb{R}^d = \underbrace{\mathrm{Im}(U_r)}_{\pi^y_{\mathrm{pos}} \neq \pi_{\mathrm{pr}}} \oplus \underbrace{\mathrm{Ker}(U_r^T)}_{\pi^y_{\mathrm{pos}} \approx \pi_{\mathrm{pr}}}$$

How can this help us?

$$x = \underbrace{U_r x_r}_{\in \mathrm{Im}(U_r)} + \underbrace{U_\perp x_\perp}_{\in \mathrm{Ker}(U_r^T)}$$

Then

$$\widetilde{\pi}^y_{\mathrm{pos}}(x) = \underbrace{\left(\widetilde{\mathcal{L}}(x_r)\pi_{\mathrm{pr}}(x_r)\right)}_{\widetilde{\pi}^y_{\mathrm{pos}}(x_r)} \pi_{\mathrm{pr}}(x_\perp | x_r)$$

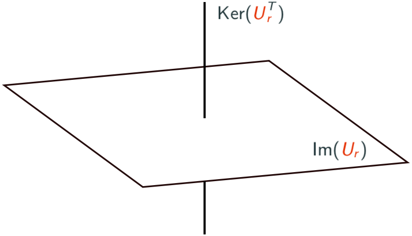


How can this help us?

$$x = \underbrace{U_r x_r}_{\in \mathrm{Im}(U_r)} + \underbrace{U_\perp x_\perp}_{\in \mathrm{Ker}(U_r^T)}$$

Then

$$\widetilde{\pi}^y_{\mathrm{pos}}(x) = \underbrace{\left(\widetilde{\mathcal{L}}(x_r)\pi_{\mathrm{pr}}(x_r)\right)}_{\widetilde{\pi}^y_{\mathrm{pos}}(x_r)} \pi_{\mathrm{pr}}(x_\perp | x_r)$$

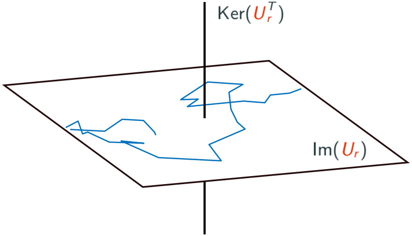


- MCMC to sample from $\widetilde{\pi}^y_{\mathrm{pos}}$
    1. **Subspace MCMC** to get samples $x_r^{(i)} \sim \widetilde{\pi}^y_{\mathrm{pos}}(x_r)$
    2. Draw samples from the **conditional prior** $x_\perp^{(i)} \sim \pi_{\mathrm{pr}}(x_\perp | x_r^{(i)})$
    3. Assemble $x^{(i)} = U_r x_r^{(i)} + U_\perp x_\perp^{(i)} \sim \widetilde{\pi}^y_{\mathrm{pos}}(x)$

$$x = \underbrace{U_r x_r}_{\in \mathrm{Im}(U_r)} + \underbrace{U_\perp x_\perp}_{\in \mathrm{Ker}(U_r^T)}$$

Then

$$\widetilde{\pi}^y_{\mathrm{pos}}(x) = \underbrace{\left(\widetilde{\mathcal{L}}(x_r)\pi_{\mathrm{pr}}(x_r)\right)}_{\widetilde{\pi}^y_{\mathrm{pos}}(x_r)} \pi_{\mathrm{pr}}(x_\perp | x_r)$$



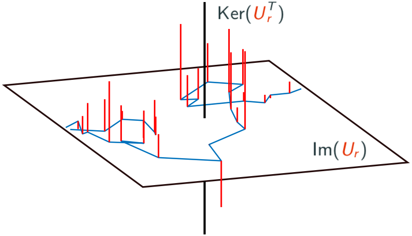
- MCMC to sample from $\widetilde{\pi}^y_{\mathrm{pos}}$
  1. **Subspace MCMC** to get samples $x_r^{(i)} \sim \widetilde{\pi}^y_{\mathrm{pos}}(x_r)$
  2. Draw samples from the **conditional prior** $x_\perp^{(i)} \sim \pi_{\mathrm{pr}}(x_\perp | x_r^{(i)})$
  3. Assemble $x^{(i)} = U_r x_r^{(i)} + U_\perp x_\perp^{(i)} \sim \widetilde{\pi}^y_{\mathrm{pos}}(x)$

How can this help us?

$$x = \underbrace{U_r x_r}_{\in \mathrm{Im}(U_r)} + \underbrace{U_\perp x_\perp}_{\in \mathrm{Ker}(U_r^T)}$$

Then

$$\widetilde{\pi}^y_{\mathrm{pos}}(x) = \underbrace{\left(\widetilde{\mathcal{L}}(x_r)\pi_{\mathrm{pr}}(x_r)\right)}_{\widetilde{\pi}^y_{\mathrm{pos}}(x_r)} \pi_{\mathrm{pr}}(x_\perp | x_r)$$

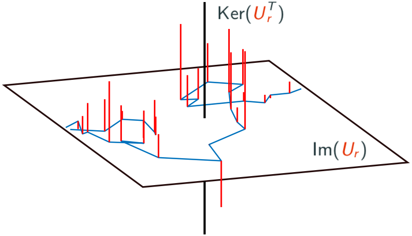


- MCMC to sample from $\widetilde{\pi}^y_{\mathrm{pos}}$
  1. **Subspace MCMC** to get samples $x_r^{(i)} \sim \widetilde{\pi}^y_{\mathrm{pos}}(x_r)$
  2. Draw samples from the **conditional prior** $x_\perp^{(i)} \sim \pi_{\mathrm{pr}}(x_\perp | x_r^{(i)})$
  3. Assemble $x^{(i)} = U_r x_r^{(i)} + U_\perp x_\perp^{(i)} \sim \widetilde{\pi}^y_{\mathrm{pos}}(x)$

$$x = \underbrace{U_r x_r}_{\in \mathrm{Im}(U_r)} + \underbrace{U_\perp x_\perp}_{\in \mathrm{Ker}(U_r^T)}$$

Then

$$\widetilde{\pi}^y_{\text{pos}}(x) = \underbrace{\left(\widetilde{\mathcal{L}}(x_r)\pi_{\text{pr}}(x_r)\right)}_{\widetilde{\pi}^y_{\text{pos}}(x_r)} \pi_{\text{pr}}(x_\perp | x_r)$$



- MCMC to sample from $\widetilde{\pi}^y_{\text{pos}}$
  1. **Subspace MCMC** to get samples $x_r^{(i)} \sim \widetilde{\pi}^y_{\text{pos}}(x_r)$
  2. Draw samples from the **conditional prior** $x_\perp^{(i)} \sim \pi_{\text{pr}}(x_\perp | x_r^{(i)})$
  3. Assemble $x^{(i)} = U_r x_r^{(i)} + U_\perp x_\perp^{(i)} \sim \widetilde{\pi}^y_{\text{pos}}(x)$

- Get samples from the **exact** posterior π^y_{pos} by correcting $x^{(i)}$ via importance weights or a Metropolis scheme [Cui, Zahm 2021], [Cui, Law, Marzouk 2016],...

Objective of the talk

Controlled approximation problem

Given $\varepsilon > 0$, build an approximation of π_{pos}^y under the form of

$$\widetilde{\pi}_{\text{pos}}^y(x) \propto \widetilde{\mathcal{L}}(U_r^T x)\pi_{\text{pr}}(x) \qquad \text{with} \left\{ \begin{array}{l} \widetilde{\mathcal{L}} : \mathbb{R}^r \to \mathbb{R}_{\geq 0} \\ U_r \in \mathbb{R}^{d\times r} \end{array} \right.$$

with $r = r(\varepsilon) \ll d$ such that

$$\boxed{D_{\text{KL}}(\pi_{\text{pos}}^y || \widetilde{\pi}_{\text{pos}}^y) \leq \varepsilon}$$

Road map:

1. Constructing $U_r = U_r(y)$ using gradients of the likelihood
2. Data-free dimension reduction $U_r = \cancel{U_r(y)}$
3. A sampling strategy
4. Conclusion

Constructing $U_r = U_r(y)$ using gradients of the likelihood

$$\widetilde{\pi}^y_{\text{pos}}(x) \propto \widetilde{\mathcal{L}}(U_r^T x)\pi_{\text{pr}}(x)$$

Optimal profile function $\widetilde{\mathcal{L}}$ given U_r [Banerjee, Guo, Wang 2005]

For any (fixed) U_r, the function $\widetilde{\mathcal{L}}$ which minimizes $D_{\text{KL}}\left(\pi^y_{\text{pos}} \middle|\middle| \widetilde{\pi}^y_{\text{pos}}\right)$ is

$$\widetilde{\mathcal{L}}(x_r) = \mathbb{E}_{X\sim\pi_{\text{pr}}}\left(\mathcal{L}_y(X) \middle| U_r^T X = x_r\right)$$

As a consequence, $\widetilde{\pi}^y_{\text{pos}}(x)$ writes

$$\widetilde{\pi}^y_{\text{pos}}(x) \propto \pi^y_{\text{pos}}(x_r)\pi_{\text{pr}}(x_\perp|x_r)$$

$$\widetilde{\pi}^y_{\text{pos}}(x) \propto \widetilde{\mathcal{L}}(U_r^T x)\pi_{\text{pr}}(x)$$

Optimal profile function $\widetilde{\mathcal{L}}$ given U_r [Banerjee, Guo, Wang 2005]

For any (fixed) U_r, the function $\widetilde{\mathcal{L}}$ which minimizes $D_{\text{KL}}\left(\pi^y_{\text{pos}} \middle|\middle| \widetilde{\pi}^y_{\text{pos}}\right)$ is

$$\widetilde{\mathcal{L}}(x_r) = \mathbb{E}_{X\sim\pi_{\text{pr}}}\left(\mathcal{L}_y(X) \middle| U_r^T X = x_r\right)$$

As a consequence, $\widetilde{\pi}^y_{\text{pos}}(x)$ writes

$$\widetilde{\pi}^y_{\text{pos}}(x) \propto \pi^y_{\text{pos}}(x_r)\pi_{\text{pr}}(x_\perp | x_r)$$

Build U_r by minimizing a certified error bound [Zahm et al. 2018]

Assume $\pi_{\text{pr}} = \mathcal{N}(m_{\text{pr}}, \Sigma_{\text{pr}})$ and let $\widetilde{\mathcal{L}}$ be as above. Then we have

$$D_{\text{KL}}\left(\pi^y_{\text{pos}} \middle|\middle| \widetilde{\pi}^y_{\text{pos}}\right) \leq \frac{1}{2}\int \|(I_d - U_r U_r^T)\nabla \log \mathcal{L}_y\|^2_{\Sigma_{\text{pr}}} \, d\pi^y_{\text{pos}}$$

- proof relies on logarithmic Sobolev inequalities [Gross 1975]
- can be extended to more general (non-Gaussian) priors

Building $U_r \in \mathbb{R}^{d\times r}$ by minimizing the bound

$$D_{\mathrm{KL}}\left(\pi^y_{\mathrm{pos}} \middle|\middle| \widetilde{\pi}^y_{\mathrm{pos}}\right) \leq \frac{1}{2}\int \|(I_d - U_r U_r^T)\nabla \log \mathcal{L}_y\|^2_{\Sigma_{\mathrm{pr}}} \, \mathrm{d}\pi^y_{\mathrm{pos}}$$

Principal Component Analysis of $\nabla \log \mathcal{L}_y$:

1. Compute
$$\mathrm{H}(y) = \int \left(\nabla \log \mathcal{L}_y\right)\left(\nabla \log \mathcal{L}_y\right)^T \, \mathrm{d}\pi^y_{\mathrm{pos}}$$
2. Solve the generalized eigenvalue problem
$$\mathrm{H}(y) u_i^y = \lambda_i^y \Sigma_{\mathrm{pr}}^{-1} u_i^y$$
3. Assemble
$$U_r = [u_1^y, \ldots, u_r^y] \in \mathbb{R}^{d\times r}$$

In the end we get

$$D_{\mathrm{KL}}\left(\pi^y_{\mathrm{pos}} \middle|\middle| \widetilde{\pi}^y_{\mathrm{pos}}\right) \leq \frac{1}{2}(\lambda^y_{r+1} + \cdots + \lambda^y_d)$$

Building $U_r \in \mathbb{R}^{d\times r}$ by minimizing the bound

$$D_{\text{KL}}\left(\pi^y_{\text{pos}} \middle|\middle| \widetilde{\pi}^y_{\text{pos}}\right) \leq \frac{1}{2}\int \|(I_d - U_r U_r^T)\nabla \log \mathcal{L}_y\|^2_{\Sigma_{\text{pr}}} \; \mathrm{d}\pi^y_{\text{pos}}$$

Principal Component Analysis of $\nabla \log \mathcal{L}_y$:

1. Compute

$$\mathrm{H}(y) = \int (\nabla \log \mathcal{L}_y)(\nabla \log \mathcal{L}_y)^T \; \mathrm{d}\pi^y_{\text{pos}}$$

2. Solve the generalized eigenvalue problem

$$\mathrm{H}(y) u^y_i = \lambda^y_i \Sigma^{-1}_{\text{pr}} u^y_i$$

3. Assemble

$$U_r = [u^y_1, \ldots, u^y_r] \in \mathbb{R}^{d\times r}$$

In the end we get

$$D_{\text{KL}}\left(\pi^y_{\text{pos}} \middle|\middle| \widetilde{\pi}^y_{\text{pos}}\right) \leq \frac{1}{2}(\lambda^y_{r+1} + \cdots + \lambda^y_d)$$

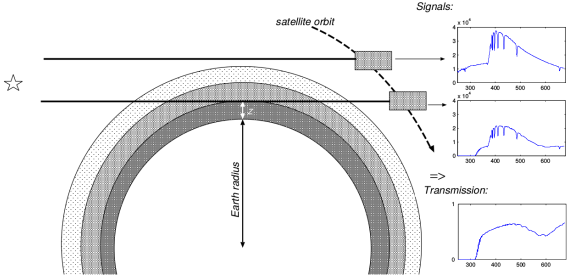
Alternative: coordinate selection $U_r x = x_\tau$ for some $\tau \subset \{1, \ldots, d\}$

$$D_{\text{KL}}\left(\pi^y_{\text{pos}} \middle|\middle| \widetilde{\pi}^y_{\text{pos}}\right) \leq \frac{1}{2}\sum_{i \notin \tau} H(y)_{i,i}(\Sigma_{\text{pr}})_{i,i}$$

Benchmark Gomos: atmospheric remote sensing (eg [Haario Tamminen et al. 2004])

- Estimate gas densities $x = \rho^{\text{gas}}(z)$ from transmission spectra $y_\omega(z)$
- Beer's law:

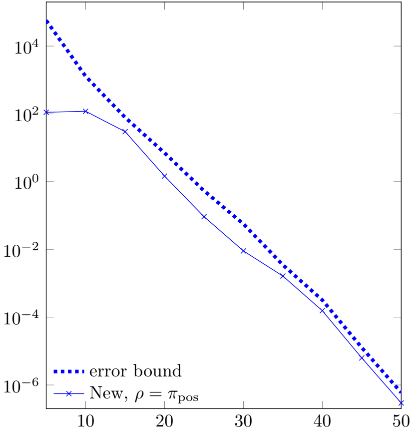
$$y_\omega(z) = \exp\left(- \int_{\text{light path}} \sum_{\text{gas}} \alpha_\omega^{\text{gas}}(z(\zeta))\ \rho^{\text{gas}}(z(\zeta))\ \mathrm{d}\zeta \right) + \xi$$



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

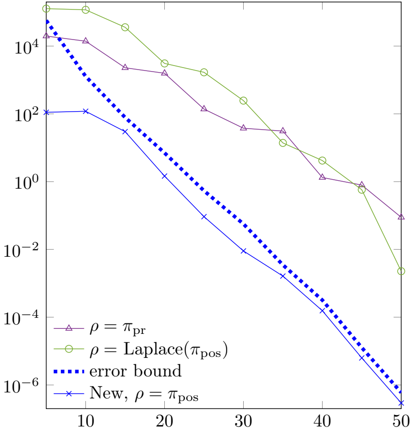
Results [Zahm, Cui, Law, Spantini, Marzouk 2018]

$$D_{\mathrm{KL}}(\pi^y_{\mathrm{pos}}||\widetilde{\pi}^y_{\mathrm{pos}}) = \mathit{function}(r)$$



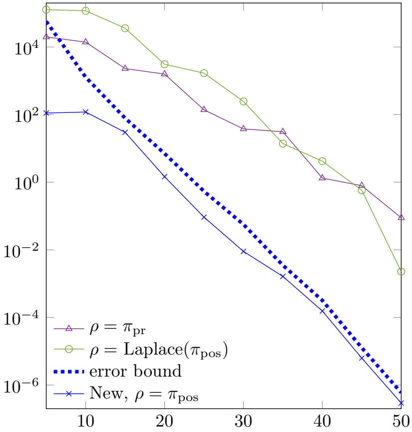
$$\mathrm{H}(y) = \int (\nabla \log \mathcal{L}_y)(\nabla \log \mathcal{L}_y)^T \, \mathrm{d}\pi^y_{\mathrm{pos}}$$

$$D_{KL}(\pi^y_{pos} || \widetilde{\pi}^y_{pos}) = function(r)$$



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

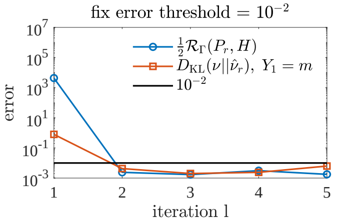
$D_{\text{KL}}(\pi_{\text{pos}}^y || \widetilde{\pi}_{\text{pos}}^y) = function(r)$

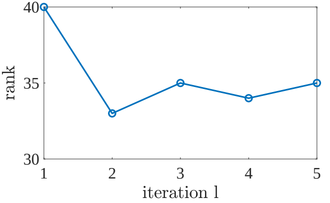


$$\mathrm{H}^{(\rho)}(y) = \int (\nabla \log \mathcal{L}_y)(\nabla \log \mathcal{L}_y)^T \, \mathrm{d}\rho$$

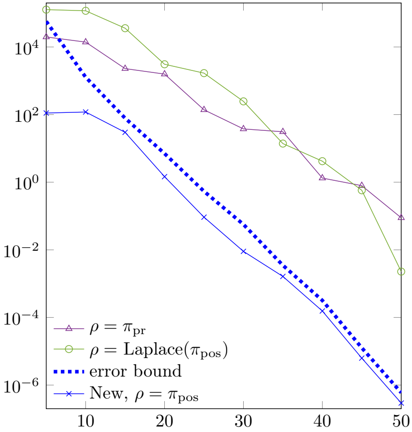
Iterative procedure:

$$\pi_{\text{pr}} \to H^{(\pi_{\text{pr}})} \to \widetilde{\pi}_{\text{pos}}^y \to H^{(\widetilde{\pi}_{\text{pos}}^y)} \to \dots$$





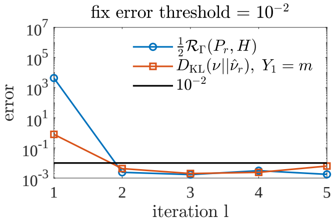
$D_{\text{KL}}(\pi_{\text{pos}}^y || \widetilde{\pi}_{\text{pos}}^y) = \textit{function}(r)$

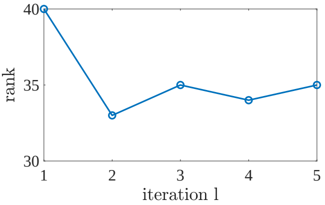


$$\mathrm{H}^{(\rho)}(y) = \int (\nabla \log \mathcal{L}_y)(\nabla \log \mathcal{L}_y)^T \, \mathrm{d}\rho$$

Iterative procedure:

$$\pi_{\text{pr}} \to H^{(\pi_{\text{pr}})} \to \widetilde{\pi}_{\text{pos}}^y \to H^{(\widetilde{\pi}_{\text{pos}}^y)} \to \dots$$





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

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

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 $U_r = \cancel{U_r(Y)}$ such that

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

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

By Markov inequality,

$$\mathbb{E}\Big(D_{\text{KL}}(\pi_{\text{pos}}^Y || \widetilde{\pi}_{\text{pos}}^Y)\Big) \leq \varepsilon$$

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

Offline

1. Compute
$$\mathrm{H} = \mathbb{E}(\mathrm{H}(Y))$$
2. Solve the generalized eigenvalue problem $\mathrm{H}u_i = \lambda_i \Sigma_{\mathrm{pr}} u_i$ and let
$$U_r = [u_1, \ldots, u_r] \in \mathbb{R}^{d \times r}$$

Online

3. Receive a realization y of Y,
4. Compute the optimal function $\widetilde{\mathcal{L}}_y = \mathbb{E}_{\pi_{\mathrm{pr}}}(\mathcal{L}_y | \sigma(U_r^T))$
5. Assemble the posterior approximation $\widetilde{\pi}^y_{\mathrm{pos}} \propto \widetilde{\mathcal{L}}_y(U_r^T x)\pi_{\mathrm{pr}}(x)$

Proposition

Assume $\pi_{\mathrm{pr}} = \mathcal{N}(m_{\mathrm{pr}}, \Sigma_{\mathrm{pr}})$. The above procedure yields
$$\mathbb{E}\Big(D_{\mathrm{KL}}(\pi^Y_{\mathrm{pos}} || \widetilde{\pi}^Y_{\mathrm{pos}})\Big) \leq \frac{1}{2}(\lambda_{r+1} + \cdots + \lambda_d)$$

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

Proposition [Cui, Zahm 2021]

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

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

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

Proposition [Cui, Zahm 2021]

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

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

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

Explicit expression on the Fisher information matrix when:

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

$$\boxed{H = \int \nabla G(x)^T \Gamma_{\mathrm{obs}}^{-1} \nabla G(x)^T \, \mathrm{d}\pi_{\mathrm{pr}}(x)}$$

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

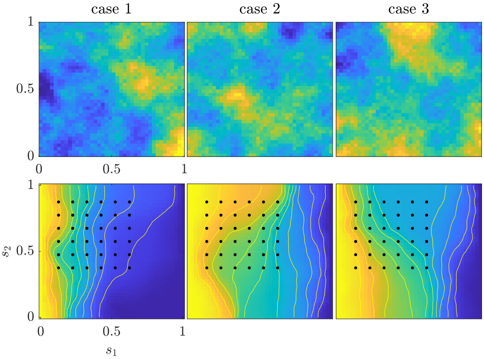
$$\boxed{H = \int \nabla G(x)^T \mathrm{diag}\Big(G_1(x), \ldots, G_m(x)\Big)^{-1} \nabla G(x)^T \, \mathrm{d}\pi_{\mathrm{pr}}(x)}$$

- ...

A numerical example: elliptic PDE

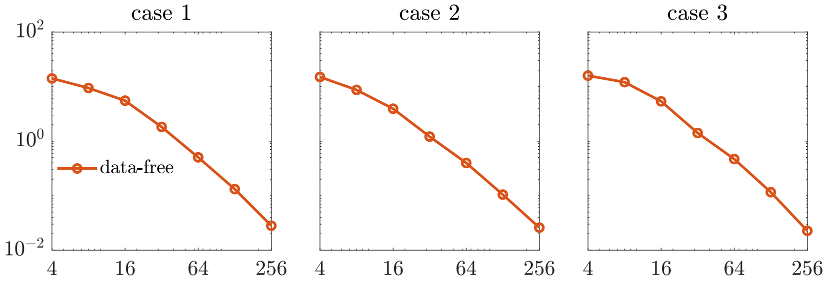
$$-\nabla \cdot \Big(\kappa(s)\nabla p(s)\Big) = f(s), \quad s \in [0,1]^2$$

- 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 $Y^{(1)}, Y^{(2)}, Y^{(3)}$

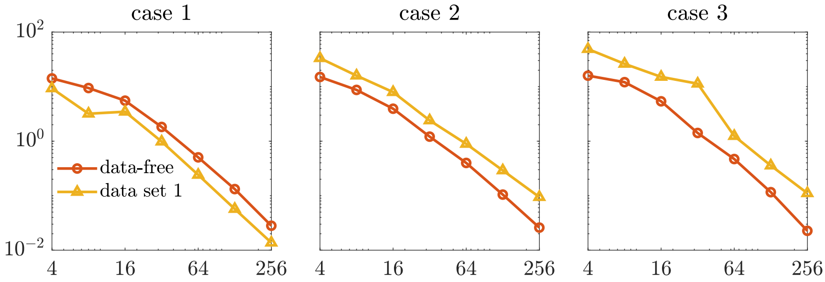
$$D_{KL}(\pi_{\text{pos}}^{Y^{(i)}} || \widetilde{\pi}_{\text{pos}}^{Y^{(i)}}) = function(r)$$



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

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

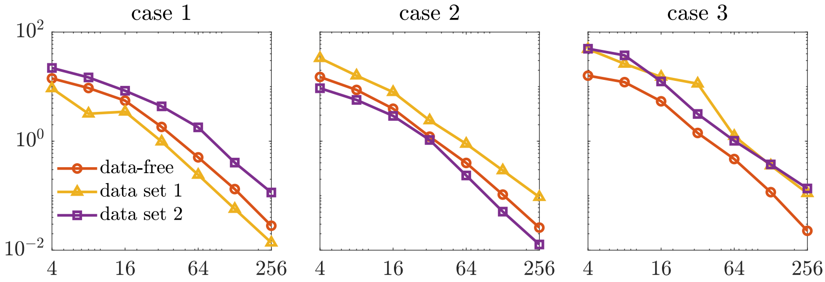
$$D_{\mathrm{KL}}(\pi_{\mathrm{pos}}^{Y^{(i)}} || \widetilde{\pi}_{\mathrm{pos}}^{Y^{(i)}}) = function(r)$$



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

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

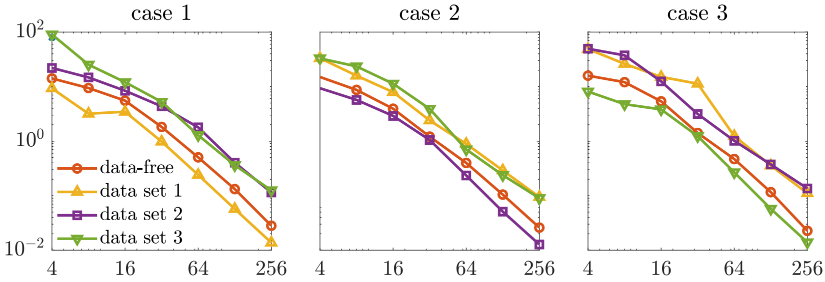
$$D_{KL}(\pi_{pos}^{Y^{(i)}} || \widetilde{\pi}_{pos}^{Y^{(i)}}) = function(r)$$



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

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

$$D_{KL}(\pi_{pos}^{Y^{(i)}} || \widetilde{\pi}_{pos}^{Y^{(i)}}) = \mathit{function}(r)$$



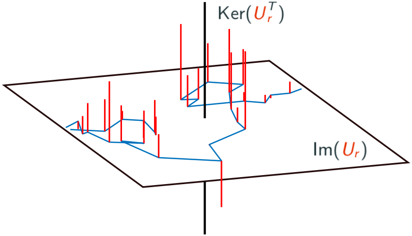
- data-free: U_r computed via $\mathrm{H} = \mathbb{E}(\mathrm{H}(Y))$
- data set 1: U_r computed via $\mathrm{H}(Y^{(1)})$
- data set 2: U_r computed via $\mathrm{H}(Y^{(2)})$
- data set 3: U_r computed via $\mathrm{H}(Y^{(3)})$

A sampling strategy

Sample from the approximate posterior

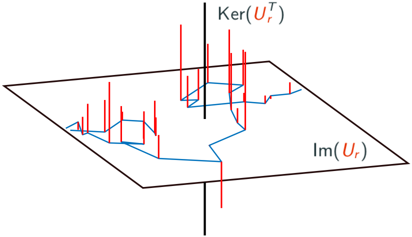
$$\widetilde{\pi}^y_{\text{pos}}(x) = \pi^y_{\text{pos}}(x_r)\ \pi_{\text{pr}}(x_\perp | x_r)$$

1. Marginal posterior $x_r^{(i)} \sim \pi^y_{\text{pos}}(x_r)$
2. Conditional prior $x_\perp^{(i)} \sim \pi_{\text{pr}}(x_\perp | x_r^{(i)})$
3. Assemble $U_r x_r^{(i)} + U_\perp x_\perp^{(i)} \sim \widetilde{\pi}^y_{\text{pos}}(x)$



$$\widetilde{\pi}^y_{\text{pos}}(x) = \pi^y_{\text{pos}}(x_r)\ \pi_{\text{pr}}(x_\perp | x_r)$$

1. Marginal posterior $x_r^{(i)} \sim \pi^y_{\text{pos}}(x_r)$
2. Conditional prior $x_\perp^{(i)} \sim \pi_{\text{pr}}(x_\perp | x_r^{(i)})$
3. Assemble $U_r x_r^{(i)} + U_\perp x_\perp^{(i)} \sim \widetilde{\pi}^y_{\text{pos}}(x)$



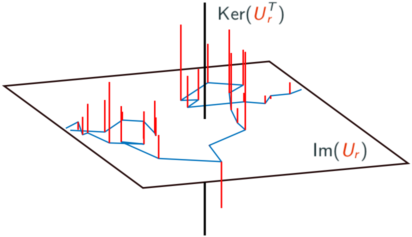
Pseudo-Marginal MCMC [Andrieu, Roberts 2009] to sample from $\pi^y_{\text{pos}}(x_r)$

$$\pi^y_{\text{pos}}(x_r) = \int \frac{\pi^y_{\text{pos}}(x_r + \widetilde{x}_\perp)}{\pi_{\text{pr}}(\widetilde{x}_\perp | x_r)} \pi_{\text{pr}}(\widetilde{x}_\perp | x_r) \mathrm{d}\widetilde{x}_\perp \quad \approx \quad \frac{1}{N} \sum_{i=1}^{N} \frac{\pi^y_{\text{pos}}(x_r + \widetilde{x}_\perp^{(i)})}{\pi_{\text{pr}}(\widetilde{x}_\perp^{(i)} | x_r)}$$

- **Low-variance** estimator by construction of U_r (N can be small)
- Pseudo-Marginal trick: redraw $\widetilde{x}_\perp^{(i)} \sim \pi_{\text{pr}}(x_\perp | x_r)$ at each MCMC iteration.

$$\widetilde{\pi}^y_{\text{pos}}(x) = \pi^y_{\text{pos}}(x_r)\ \pi_{\text{pr}}(x_\perp | x_r)$$

1. Marginal posterior $x_r^{(i)} \sim \pi^y_{\text{pos}}(x_r)$
2. ~~Conditional prior $x_\perp^{(i)} \sim \pi_{\text{pr}}(x_\perp | x_r^{(i)})$~~
3. Assemble $U_r x_r^{(i)} + U_\perp x_\perp^{(i)} \sim \widetilde{\pi}^y_{\text{pos}}(x)$



Pseudo-Marginal MCMC [Andrieu, Roberts 2009] to sample from $\pi^y_{\text{pos}}(x_r)$

$$\pi^y_{\text{pos}}(x_r) = \int \frac{\pi^y_{\text{pos}}(x_r + \widetilde{x}_\perp)}{\pi_{\text{pr}}(\widetilde{x}_\perp | x_r)} \pi_{\text{pr}}(\widetilde{x}_\perp | x_r) \mathrm{d}\widetilde{x}_\perp \quad \approx \quad \frac{1}{N} \sum_{i=1}^{N} \frac{\pi^y_{\text{pos}}(x_r + \widetilde{x}_\perp^{(i)})}{\pi_{\text{pr}}(\widetilde{x}_\perp^{(i)} | x_r)}$$

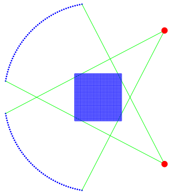
- **Low-variance** estimator by construction of U_r (N can be small)
- Pseudo-Marginal trick: redraw $\widetilde{x}_\perp^{(i)} \sim \pi_{\text{pr}}(x_\perp | x_r)$ at each MCMC iteration.

Recycle $\widetilde{x}_\perp^{(i)}$ to sample from the exact full posterior $\pi^y_{\text{pos}}(x)$ [Cui, Zahm 2021]

Instead of drawing $x_\perp^{(i)} \sim \pi_{\text{pr}}(x_\perp | x_r^{(i)})$, pick $x_\perp^{(i)} \in \{\widetilde{x}_\perp^{(1)}, \dots, \widetilde{x}_\perp^{(N)}\}$ 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**

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

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

MCMC using H-MALA[1] 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

[1] Hessian-preconditioned Metropolis-Adjusted Langevin Algorithm

Conclusion

Conclusion

- Detect the low effective dimensionality of Bayesian inverse problems by:
  - deriving an upper bound on the error (KL-divergence)
  - minimizing the bound ($\equiv$ PCA on $\nabla \log \mathcal{L}_y$)
  - fundamentally gradient-based: need access to Fisher information matrix
- Extension to data-free:
  - find the directions which *will be* informed by the data
  - provides bound on KL-divergence in expectation
  - speedup in MCMC computations

[Cui, Zahm 2021] *Data-free likelihood-informed dimension reduction for bayesian inverse problems*, Inverse Problems 2021.