Chair of Numerical Mathematics (M2)
Department of Mathematics
Technical University of Munich

Active subspaces for Bayesian inversion

Application to a methane hydrate model

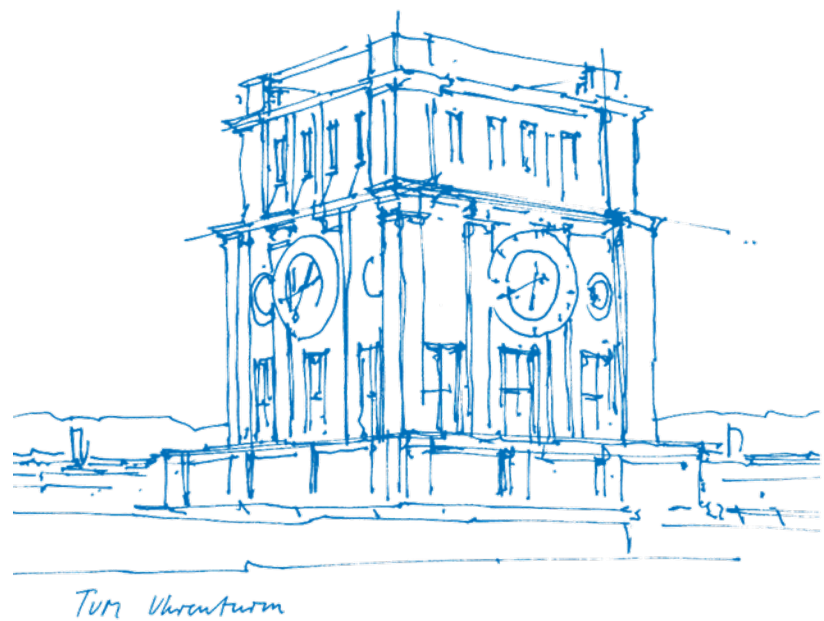
Mario Teixeira Parente

Chair of Numerical Mathematics (M2)

E-mail: `parente@ma.tum.de`

Web: `http://www.mateipa.de`

Garching, 02.03.2018



Outline

1. Bayesian inversion – Basics

2. Active subspaces – General

3. Active subspaces for Bayesian inversion

4. Application to a methane hydrate model

Chair of Numerical Mathematics (M2)
Department of Mathematics
Technical University of Munich

Outline

Bayesian inversion

Uncertainties are modeled probabilistically.

Deterministic inverse problems are in general ill-posed.
$\rightarrow$ Regularization by a Bayesian problem formulation.

Task: Get a probabilistic description of model parameters $\mathbf{x}$ incorporating data $\mathbf{d}$.

$$\mathbf{d} = \mathscr{G}(\mathbf{x}^{\star}) + \eta \tag{1}$$

- $\mathbf{d}$ Data
- $\mathscr{G}$ Forward/Observable map
- $\mathbf{x}^{\star}$ “True“ parameter
- η Noise

Chair of Numerical Mathematics (M2)
Department of Mathematics
Technical University of Munich

Using Bayes' theorem

Treat parameter **x** and data **d** as random variable. Assume $\eta \sim \mathscr{N}(0, \Gamma)$.

Task: "Get" posterior distribution on the parameter space.

$$\begin{aligned} \rho_{\text{post}}(\mathbf{x}|\mathbf{d}) &\propto \rho_{\text{like}}(\mathbf{d}|\mathbf{x})\rho_{\text{prior}}(\mathbf{x}) && (2) \\ &= \exp\left(-\frac{\|\mathbf{d}-\mathscr{G}(\mathbf{x})\|_\Gamma^2}{2}\right)\rho_{\text{prior}}(\mathbf{x}) && (3) \\ &= \exp\left(-f_{\mathbf{d}}(\mathbf{x})\right)\rho_{\text{prior}}(\mathbf{x}) && (4) \end{aligned}$$

- $\|\cdot\|_\Gamma := \|\Gamma^{-1/2}\cdot\|_2$
- ρ_{post} Posterior
- ρ_{like} Likelihood
- ρ_{prior} Prior
- $f_{\mathbf{d}}$ Data misfit function

Chair of Numerical Mathematics (M2)
Department of Mathematics
Technical University of Munich

Sampling from the posterior

One way: ***M**arkov **c**hain **M**onte **C**arlo* (MCMC)

Construct a Markov chain s.t. its stationary distribution is the posterior.
$\rightarrow$ Metropolis(-Hastings) algorithm

Algorithm 1 Metropolis algorithm for the posterior.

Assume a symmetric proposal density function τ, an initial point $\mathbf{x}_1$, and a prior ρ_{prior} is given.
For $i = 1, 2, \ldots, N-1$

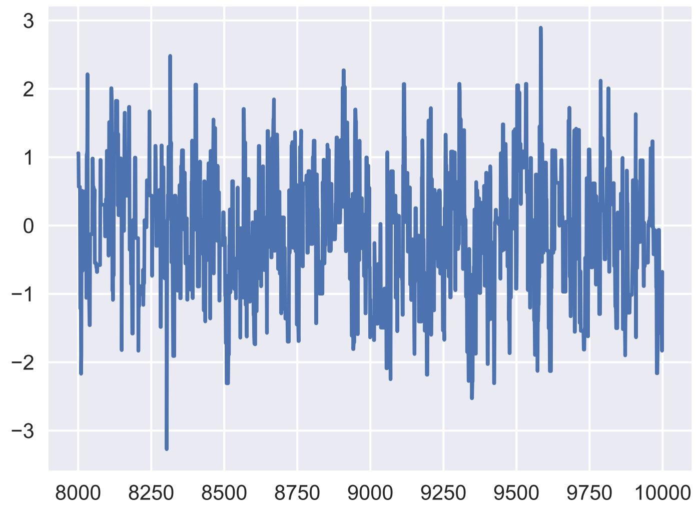
1. Draw a proposal $\tilde{\mathbf{x}}$ from τ centered at $\mathbf{x}_i$.
2. Calculate the acceptance ratio

$$\alpha(\tilde{\mathbf{x}}, \mathbf{x}_i) = \min\left(1, \frac{\exp(-f_{\mathbf{d}}(\tilde{\mathbf{x}}))\rho_{\text{prior}}(\tilde{\mathbf{x}})}{\exp(-f_{\mathbf{d}}(\mathbf{x}_i))\rho_{\text{prior}}(\mathbf{x}_i)}\right). \tag{5}$$

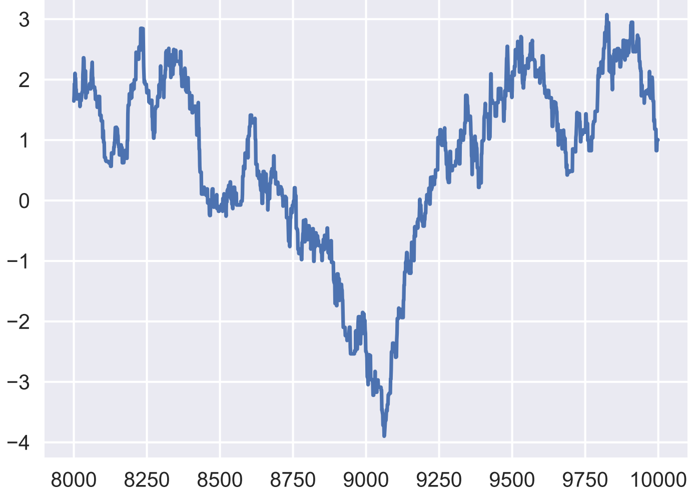
3. Draw $u \sim \mathscr{U}([0,1])$.
4. Set $\mathbf{x}_{i+1} = \tilde{\mathbf{x}}$ if $\alpha(\tilde{\mathbf{x}}, \mathbf{x}_i) \geq u$, otherwise set $\mathbf{x}_{i+1} = \mathbf{x}_i$.

Chair of Numerical Mathematics (M2)
Department of Mathematics
Technical University of Munich

Mixing behavior



(a) Mixing in two dimensions.



(b) Mixing in 100 dimensions.

Chair of Numerical Mathematics (M2)
Department of Mathematics
Technical University of Munich

Outline

Chair of Numerical Mathematics (M2)
Department of Mathematics
Technical University of Munich

Active subspaces[1]

Goal: Approximate a function $f : \mathbb{R}^n \to \mathbb{R}$ by a function $g : \mathbb{R}^k \to \mathbb{R}$ ($k < n$), i.e. find $\mathbf{A} \in \mathbb{R}^{n \times k}$ and g s.t.

$$f(\mathbf{x}) \approx g(\mathbf{A}^\top \mathbf{x}). \tag{6}$$

(Also called *ridge approximation*.)

Note: $g(\mathbf{A}^\top \mathbf{x})$ constant along $\mathscr{N}(\mathbf{A}^\top)$. $\left[\nabla(g(\mathbf{A}^\top \mathbf{x})) = \nabla g(\mathbf{A}^\top \mathbf{x})\,\mathbf{A}^\top\right]$

[1] [Constantine et al., 2014, Constantine, 2015]

Chair of Numerical Mathematics (M2)
Department of Mathematics
Technical University of Munich

Ridge approximation

Define

$$\mathbf{C} := \int_{\mathbb{R}^n} \nabla f \nabla f^\top \rho \, d\mathbf{x}. \tag{7}$$

$\mathbf{C}$ is symmetric and positive semi-definite.

$\rightarrow \mathbf{C} = \mathbf{W}\Lambda\mathbf{W}^\top$ for orthogonal $\mathbf{W} \in \mathbb{R}^{n\times n}$, $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$, $\lambda_i \geq 0$.

$$\lambda_i = \mathbf{w}_i^\top \mathbf{C} \mathbf{w}_i = \int_{\mathbb{R}^n} \left(\mathbf{w}_i^\top \nabla f\right)^2 \rho \, d\mathbf{x} \tag{8}$$

Ridge approximation

Define

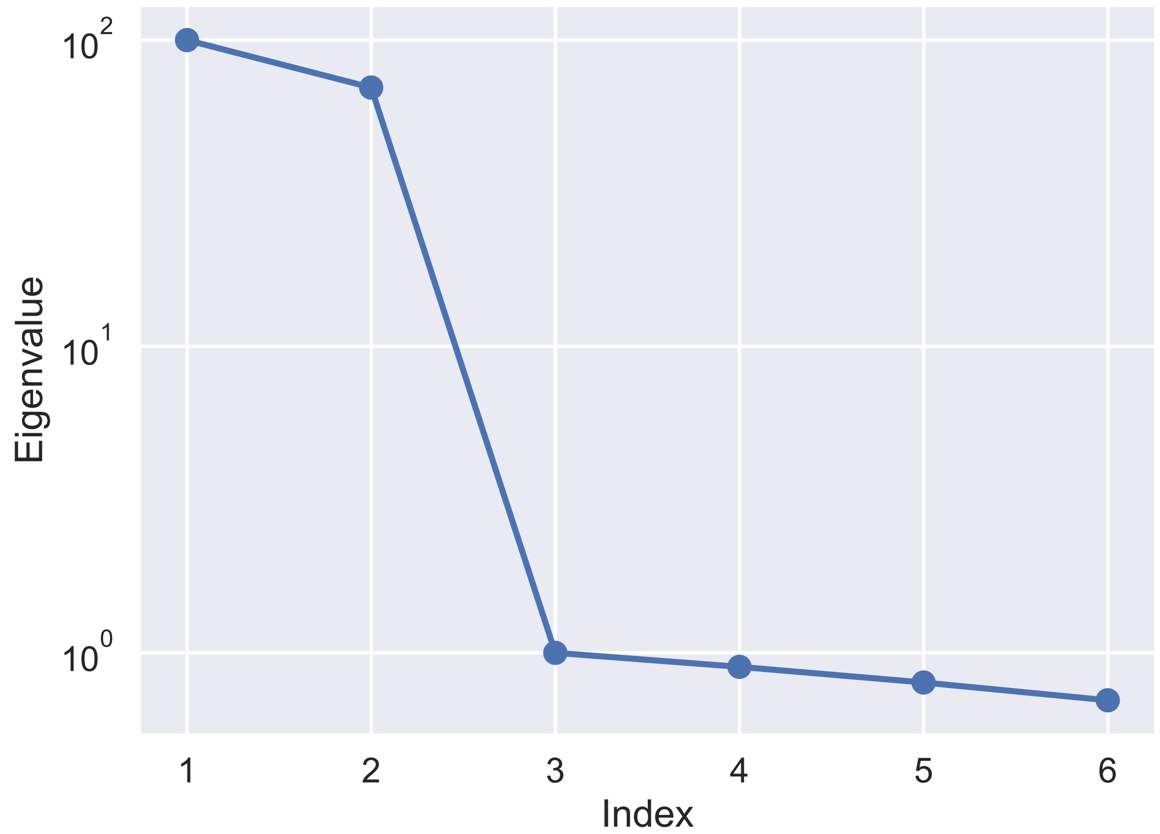
$$\mathbf{C} := \int_{\mathbb{R}^n} \nabla f \nabla f^\top \rho \, d\mathbf{x}. \tag{7}$$

C is symmetric and positive semi-definite.

$\rightarrow \mathbf{C} = \mathbf{W}\Lambda\mathbf{W}^\top$ for orthogonal $\mathbf{W} \in \mathbb{R}^{n\times n}$, $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$, $\lambda_i \geq 0$.

$$\lambda_i = \mathbf{w}_i^\top \mathbf{C} \mathbf{w}_i = \int_{\mathbb{R}^n} \left(\mathbf{w}_i^\top \nabla f\right)^2 \rho \, d\mathbf{x} \tag{8}$$

Eigenvalues



Subspaces

Decompose

$$\mathbf{W} = [\mathbf{W}_1 \;\; \mathbf{W}_2] \tag{9}$$

for $\mathbf{W}_1 \in \mathbb{R}^{n \times k}$, $\mathbf{W}_2 \in \mathbb{R}^{n \times (n-k)}$ and

$$\mathbf{x} = \mathbf{W}\mathbf{W}^\top \mathbf{x} = \mathbf{W}_1\mathbf{W}_1^\top \mathbf{x} + \mathbf{W}_2\mathbf{W}_2^\top \mathbf{x} \tag{10}$$

$$= \mathbf{W}_1 y + \mathbf{W}_2 z \tag{11}$$

for $y := \mathbf{W}_1^\top \mathbf{x}$, $z := \mathbf{W}_2^\top \mathbf{x}$.

$\mathbf{y}$ is called the *active*, $\mathbf{z}$ the *inactive* variable. $\mathscr{R}(\mathbf{W}_1)$ is the *active subspace*.

Approximation of f

(New) **Task**: If $\lambda_{k+1}, \dots, \lambda_n$ are small enough, approximate f by finding g s.t.

$$f(\mathbf{x}) \approx g(\mathbf{W}_1^\top \mathbf{x}). \tag{12}$$

Note: $\mathscr{N}(\mathbf{W}_1^\top) = \mathscr{R}(\mathbf{W}_2)$.

Algorithm 2 Computing the regression surface g in the active variable

---

Assume samples $\mathbf{x}_i$ $(i = 1, \dots, N)$ according to ρ and corresponding function values f_i $(i = 1, \dots, N)$ are given.

1. Compute samples $\mathbf{y}_i$ in the active subspace by

$$\mathbf{y}_i = \mathbf{W}_1^\top \mathbf{x}_i, \quad i = 1, \dots, N. \tag{13}$$

2. Find a regression surface g for pairs $(\mathbf{y}_i, f_i)$ such that

$$g(\mathbf{y}_i) \approx f_i, \quad i = 1, \dots, N. \tag{14}$$

Computing $\mathbf{C}^{[2]}$

Approximate

$$\mathbf{C} \approx \frac{1}{N} \sum_{i=1}^{N} \nabla f_i \nabla f_i^{\top} = \hat{\mathbf{W}} \hat{\Lambda} \hat{\mathbf{W}}^{\top} \tag{15}$$

with $\nabla f_i := \nabla f(\mathbf{x}_i)$ for samples $\mathbf{x}_i \sim \rho$.

Heuristic for N:

$$N = \alpha \times \ell \times \log(n) \tag{16}$$

- $\alpha \in [2, 10]$ oversampling factor
- ℓ number of eigenvalues to approximate
- n dimension of parameter space
- Basic tool: Eigenvalue Bernstein inequality for subexponential matrices

[2] [Constantine and Gleich, 2014]

Spectral gap – Theorem

Theorem

Let $\varepsilon > 0$ be "small enough" and choose $N = N(\varepsilon)$ "large enough". Then

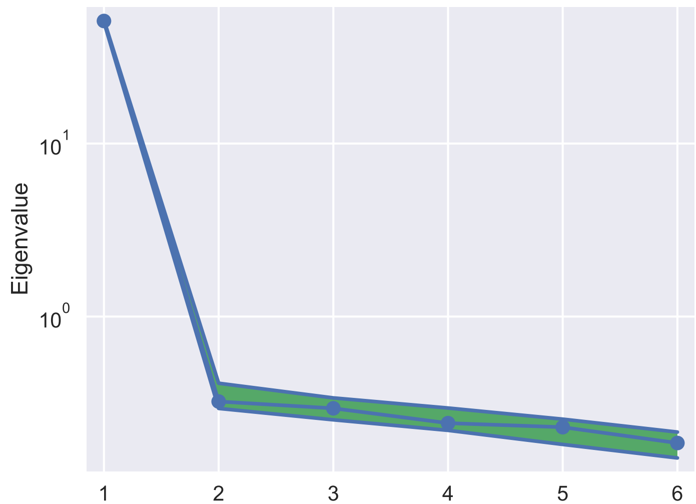
$$dist(\mathscr{R}(\mathbf{W}_1), \mathscr{R}(\hat{\mathbf{W}}_1)) \leq \frac{4\lambda_1\varepsilon}{\lambda_k - \lambda_{k+1}} \tag{17}$$

with "high probability".[3]

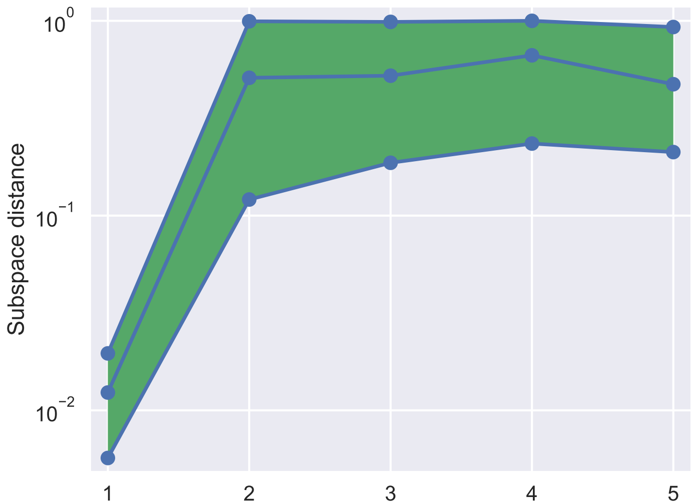
$$\left[\text{dist}(\mathscr{R}(\mathbf{W}_1), \mathscr{R}(\hat{\mathbf{W}}_1)) = \|\mathbf{W}_1\mathbf{W}_1^\top - \hat{\mathbf{W}}_1\hat{\mathbf{W}}_1^\top\|_2 = \|\mathbf{W}_1^\top\hat{\mathbf{W}}_2\|_2\right]$$

[3] [Constantine and Gleich, 2014]

Bootstrap intervals and subspace distances[4]



(a) Eigenvalues with bootstrap intervals.



(b) Subspace distances and bootstrap intervals.

- **Bootstrap**: Reuse samples to get a variance-like quantity
- Compute **subspace distances** and also bootstrap intervals

[4] [Constantine and Gleich, 2014]

Chair of Numerical Mathematics (M2)
Department of Mathematics
Technical University of Munich

Outline

Chair of Numerical Mathematics (M2)
Department of Mathematics
Technical University of Munich

Active subspaces for Bayesian inversion[5]

In the context of Bayesian inversion, we set

$$f(\mathbf{x}) = f_{\mathbf{d}}(\mathbf{x}) := \frac{\|\mathbf{d} - \mathscr{G}(\mathbf{x})\|_{\Gamma}^2}{2}. \tag{18}$$

Suppose we found a lower dimensional map $g_{\mathbf{d}}$ s.t. $f_{\mathbf{d}}(\mathbf{x}) \approx g_{\mathbf{d}}(\mathbf{W}_1^\top \mathbf{x}) = g_{\mathbf{d}}(\mathbf{y})$. Then

$$\begin{aligned}
\rho_{\text{post}}(\mathbf{x}|\mathbf{d}) &\propto \rho_{\text{like}}(\mathbf{d}|\mathbf{x})\rho_{\text{prior}}(\mathbf{x}) && (19)\\
&\propto \exp(-f_{\mathbf{d}}(\mathbf{x}))\rho_{\text{prior}}(\mathbf{x}) && (20)\\
&\approx \underbrace{\exp(-g_{\mathbf{d}}(\mathbf{y}))\rho_{\text{prior}}(\mathbf{y})}_{\propto \rho_{\text{post}}(\mathbf{y}|\mathbf{d})}\rho_{\text{prior}}(\mathbf{z}|\mathbf{y}) && (21)
\end{aligned}$$

$\rightarrow$ MCMC in the active subspace.

[5] [Constantine et al., 2016]

MCMC in the active subspace

Algorithm 3 Metropolis algorithm for the posterior in the active subspace

Assume a symmetric proposal density function τ, an initial point $\mathbf{y}_1$, and a kernel density estimate $\hat{\rho}$ for the prior on the active variable $\rho_{\text{prior}}(\mathbf{y})$ is given.

For $i = 1, 2, \ldots, N-1$

1. Draw a proposal $\tilde{\mathbf{y}}$ from τ centered at $\mathbf{y}_i$.
2. Calculate the acceptance ratio

$$\alpha(\tilde{\mathbf{y}}, \mathbf{y}_i) = \min\left(1, \frac{\exp(-g_{\mathbf{d}}(\tilde{\mathbf{y}}))\hat{\rho}(\tilde{\mathbf{y}})}{\exp(-g_{\mathbf{d}}(\mathbf{y}_i))\hat{\rho}(\mathbf{y}_i)}\right). \tag{22}$$

3. Draw $u \sim \mathscr{U}([0,1])$.
4. Set $\mathbf{y}_{i+1} = \tilde{\mathbf{y}}$ if $\alpha(\tilde{\mathbf{y}}, \mathbf{y}_i) \geq u$, otherwise set $\mathbf{y}_{i+1} = \mathbf{y}_i$.

MCMC in the active subspace gives *active samples* $\mathbf{y}$. Sample additionally from $\rho_{\text{prior}}(\mathbf{z}|\mathbf{y})$ to get

$$\mathbf{x} = \mathbf{W}_1\mathbf{y} + \mathbf{W}_2\mathbf{z}. \tag{23}$$

Chair of Numerical Mathematics (M2)
Department of Mathematics
Technical University of Munich

Outline

Chair of Numerical Mathematics (M2)
Department of Mathematics
Technical University of Munich

Methane hydrates

- Crystalline solids formed when water molecules enclatharate methane molecules.
- Stable at low temperatures and high pressures.
- Occur in permafrost regions and marine off-shores.
- If warmed or depressurized, they destabilize and dissociate into water and gas.

  $\Rightarrow$ Energy content estimated to exceed the *combined* energy content of all other conventional fossil fuels[6].

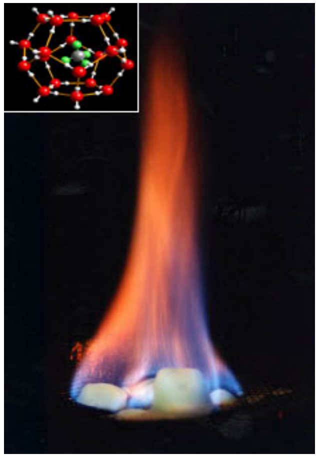


Figure : Burning hydrate (source: *United States Geological Survey*)

[6] [Piñero et al., 2013]

Methane hydrates model

- Interested in hydrates' reaction under forces/loading
  $\rightarrow$ Elastoplasticity model
- Linear relationship of *stress* σ and *elastic strain* ε^e through Hooke's law

$$\sigma = \mathbf{C} : \varepsilon^e,$$

  where $\mathbf{C}$ is a symmetric, positive definite, fourth order elastic stiffness tensor depending on the material (Lamé coefficients, Young's modulus, Poisson's ratio, ...).
- More details in e.g. [Gupta et al., 2015]

Chair of Numerical Mathematics (M2)
Department of Mathematics
Technical University of Munich

Inference for model parameters[8]

Goal: Match *stress-strain curves* of the elastoplasticity model for methane hydrates and quantify uncertainties in parameters.

- Data for *gas-saturated* sediments from laboratory experiments[7]
- 8 parameters controlling evolution of curves (peak stress, hardening and softening slopes)
- Uniform prior on $[-1, 1]^8$
- 2% noise level (Γ is diagonal)

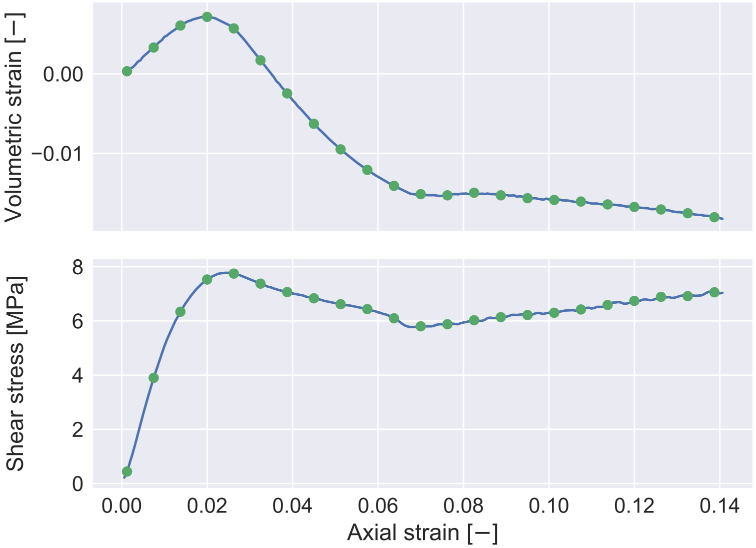


Figure : $n_{\mathbf{d}} = 46$ data points on stress and strain curves.

[7] [Deusner et al., 2012]

[8] [Teixeira Parente et al., 2018]

Chair of Numerical Mathematics (M2)
Department of Mathematics
Technical University of Munich

Computational costs

Costs lie in computing the gradient

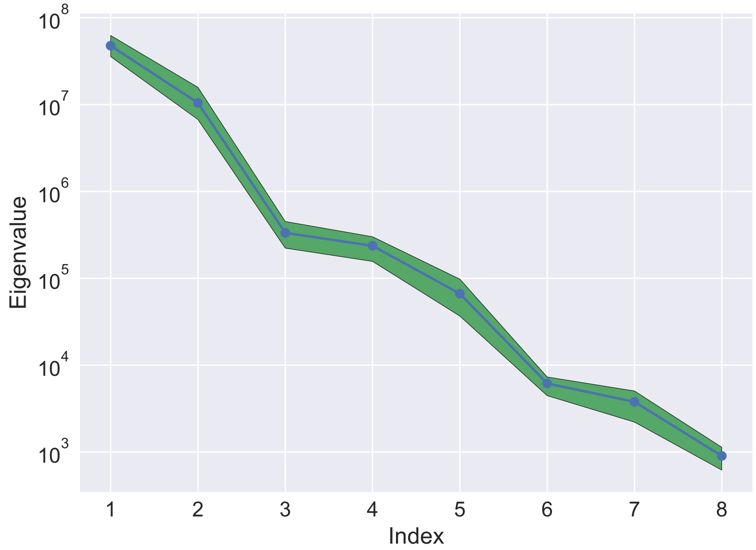
$$\nabla f_{\mathbf{d}}(\mathbf{x}) = \nabla \mathscr{G}(\mathbf{x})^\top \Gamma^{-1}(\mathscr{G}(\mathbf{x}) - \mathbf{d}) \tag{24}$$

with $\nabla \mathscr{G} : \mathbb{R}^n \to \mathbb{R}^{n_{\mathbf{d}} \times n}$, $\mathbf{d} \in \mathbb{R}^{n_{\mathbf{d}}}$.

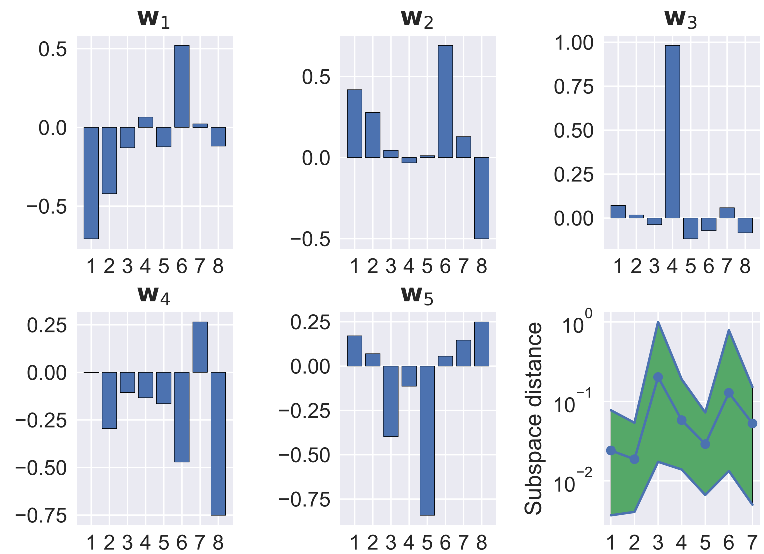
We apply (central) finite differences. $\to$ 17 forward runs

Number of gradient samples was $N = 250$ (trivially parallelizable).

Plots of eigenpairs



(a) Eigenvalues with spectral gaps after λ_2 and λ_5.



(b) First 5 eigenvectors and subspaces distances.

$\rightarrow$ Comparison between $2D$ and $5D$ subspace.

Chair of Numerical Mathematics (M2)
Department of Mathematics
Technical University of Munich

2D summary plot of $f_{\mathbf{d}}$

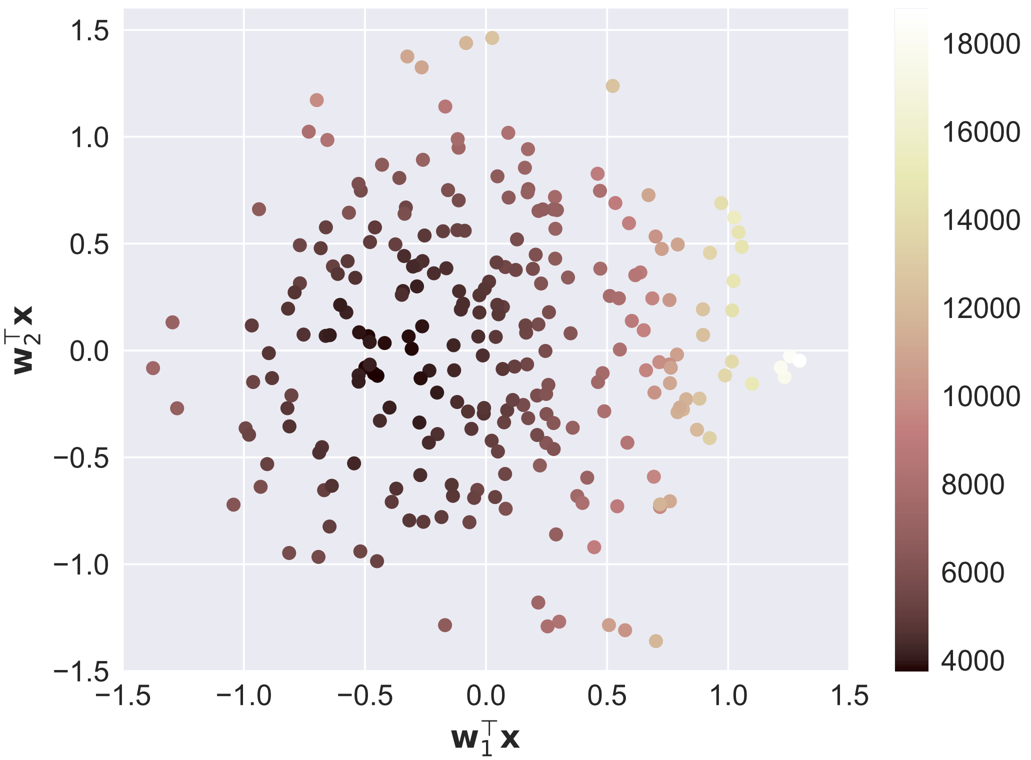


Figure : r^2 scores: 0.9776 ($2D$), 0.9824 ($5D$)

Chair of Numerical Mathematics (M2)
Department of Mathematics
Technical University of Munich

Prior/posterior on the active variable

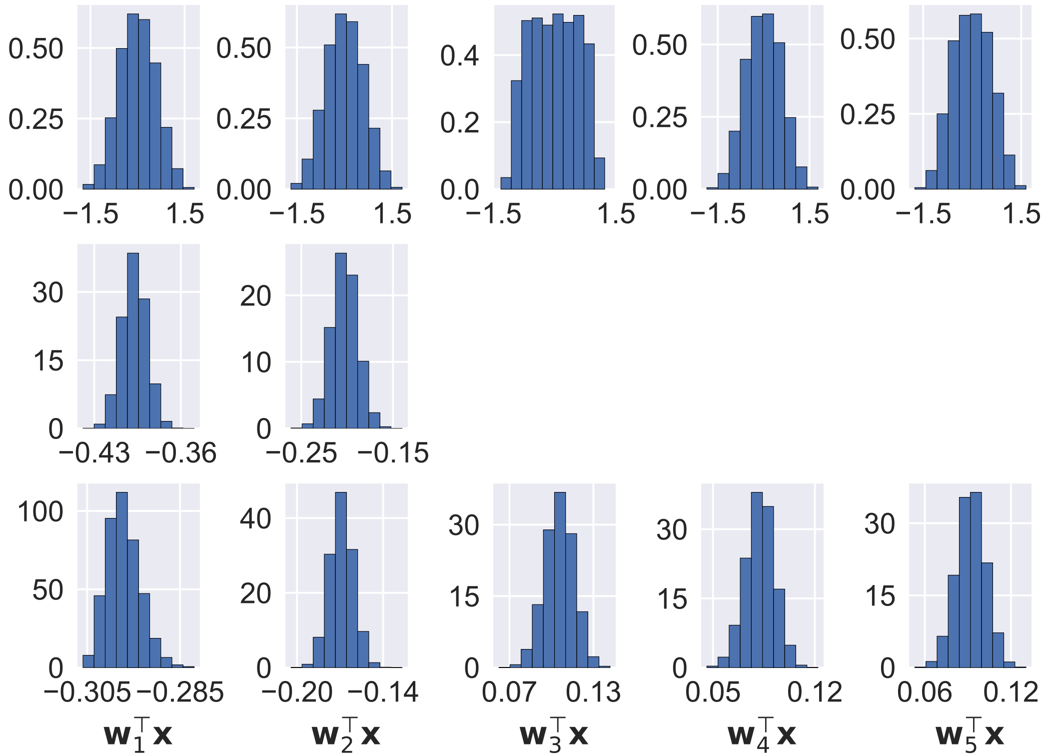


Figure : Top row: prior on the active variable(s); middle row: posterior on the $2D$ active variable; bottom row: posterior on the $5D$ active variable

Posterior in $2D$

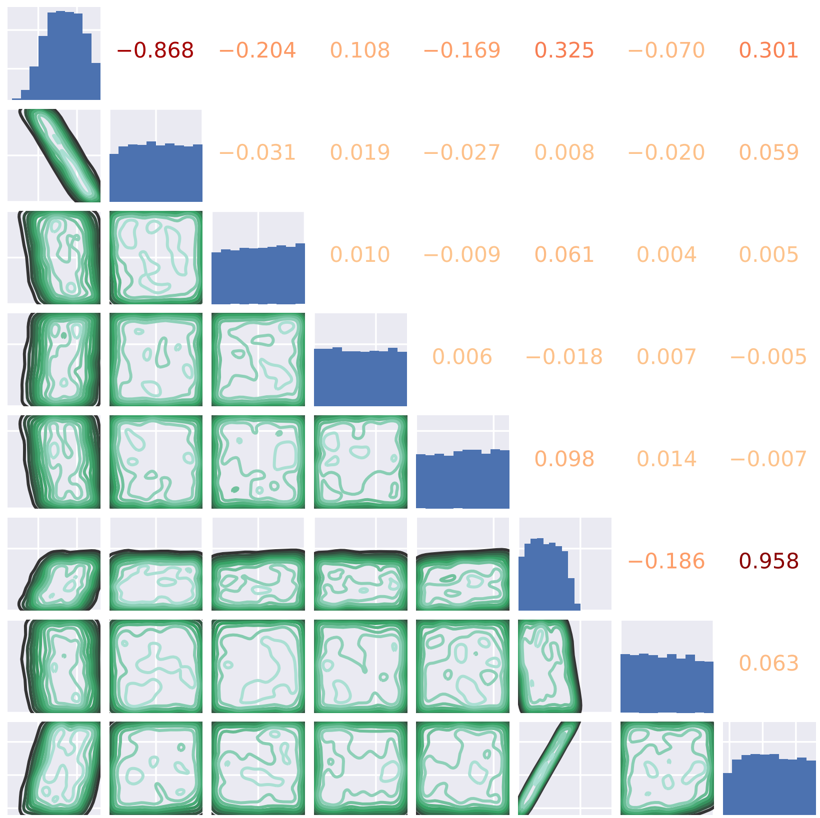


Figure : Uni- and bivariate marginals of $2D$ posterior. Colored numbers display the absolute value of correlation coefficients.

Chair of Numerical Mathematics (M2)
Department of Mathematics
Technical University of Munich

Posterior in $5D$

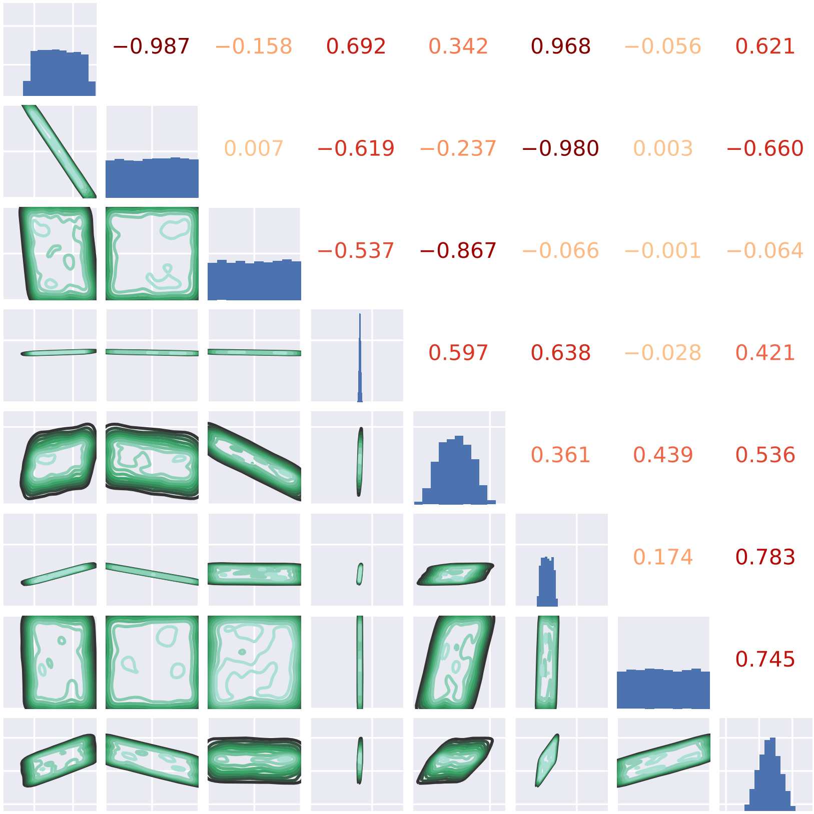
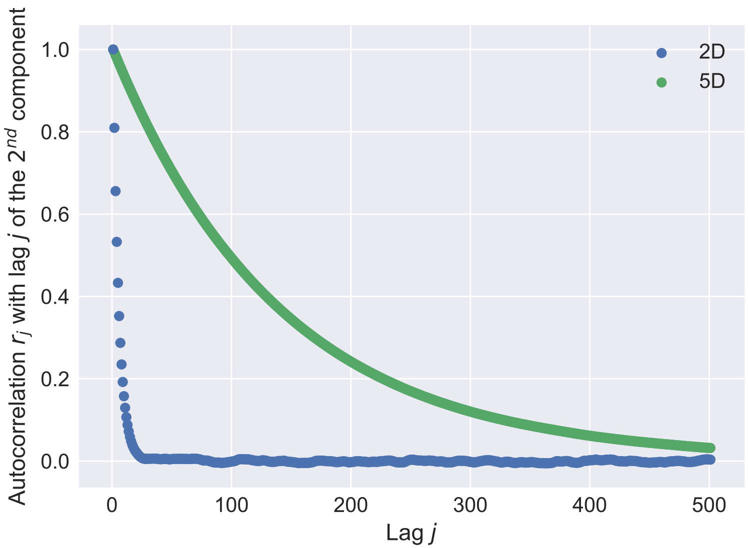
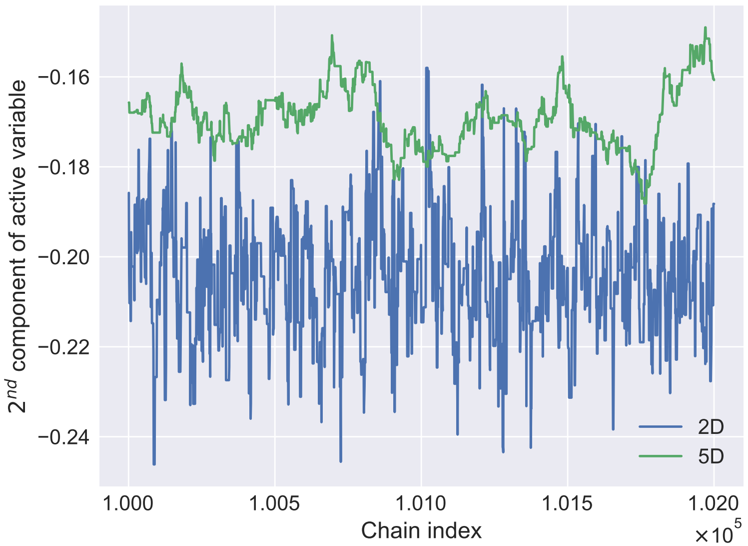


Figure : Uni- and bivariate marginals of $5D$ posterior. Colored numbers display the absolute value of correlation coefficients.

Autocorrelations + mixing



(a) Autocorrelations are much less in $2D$.



(b) Mixing is better in $2D$.

Chair of Numerical Mathematics (M2)
Department of Mathematics
Technical University of Munich

Posterior means

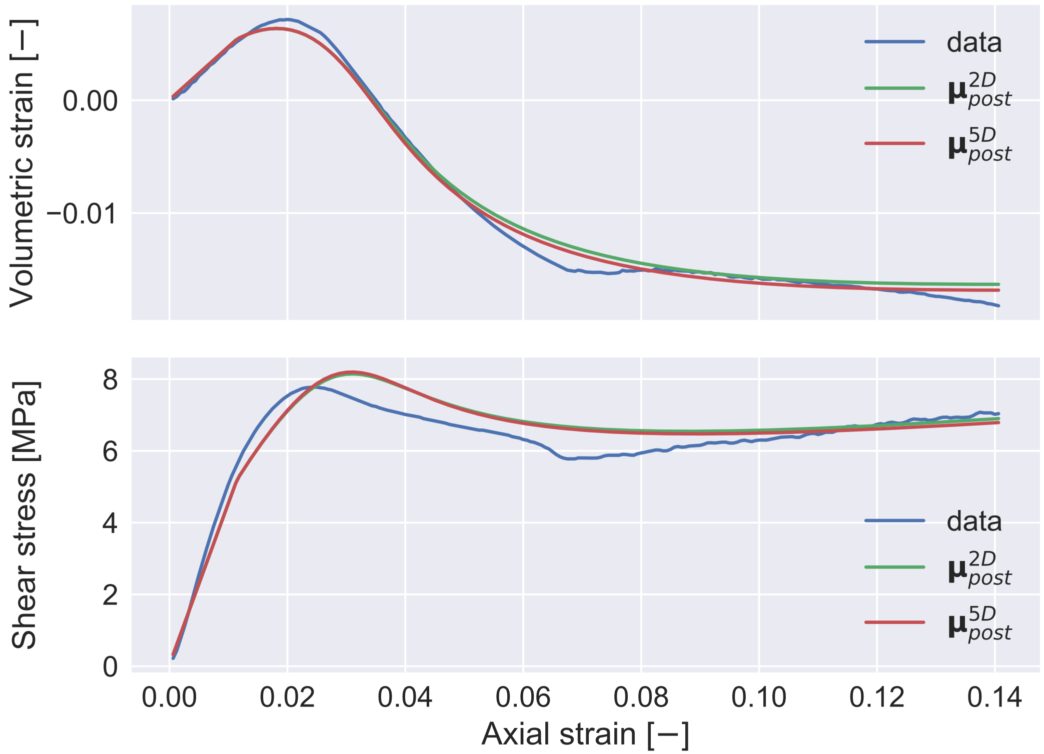


Figure : Forward evaluations with $2D$ and $5D$ posterior mean values.

References I

- Constantine, P. and Gleich, D. (2014).
  Computing active subspaces with Monte Carlo.
  arXiv preprint arXiv:1408.0545.

- Constantine, P. G. (2015).
  Active Subspaces, volume 2 of *SIAM Spotlights*.
  Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.
  Emerging Ideas for Dimension Reduction in Parameter Studies.

- Constantine, P. G., Dow, E., and Wang, Q. (2014).
  Active subspace methods in theory and practice: applications to kriging surfaces.
  SIAM J. Sci. Comput., 36(4):A1500–A1524.

- Constantine, P. G., Kent, C., and Bui-Thanh, T. (2016).
  Accelerating Markov chain Monte Carlo with Active Subspaces.
  SIAM J. Sci. Comput., 38(5):A2779–A2805.

References II

Deusner, C., Bigalke, N., Kossel, E., and Haeckel, M. (2012).
Methane production from gas hydrate deposits through injection of supercritical CO2.
Energies, 5(7):2112.

Gupta, S., Helmig, R., and Wohlmuth, B. (2015).
Non-isothermal, multi-phase, multi-component flows through deformable methane hydrate reservoirs.
Computational Geosciences, 19(5):1063–1088.

Piñero, E., Marquardt, M., Hensen, C., Haeckel, M., and Wallmann, K. (2013).
Estimation of the global inventory of methane hydrates in marine sediments using transfer functions.
Biogeosciences, 10(2):959–975.

Teixeira Parente, M., Mattis, S., Gupta, S., Deusner, C., and Wohlmuth, B. (2018).
Efficient parameter estimation for a methane hydrate model with active subspaces.
arXiv preprint arXiv:1801.09499.