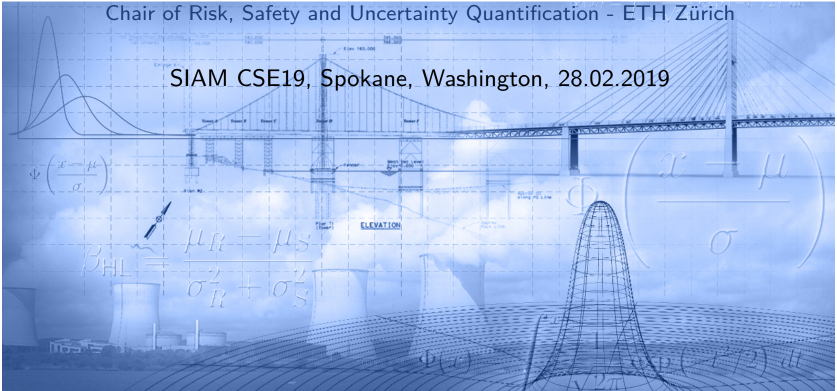
A new approach for Bayesian model calibration using stochastic spectral embedding

P.-R. Wagner, S. Marelli, C. Lataniotis, B. Sudret

Chair of Risk, Safety and Uncertainty Quantification - ETH Zürich

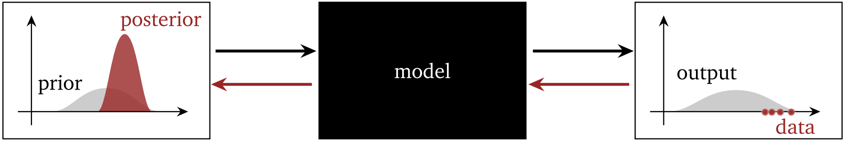
SIAM CSE19, Spokane, Washington, 28.02.2019



Motivation

Typical problem:

- Given computational model predicting observables $\mathcal{M} : \mathcal{D}_{\boldsymbol{X}} \to \mathbb{R}^{N_{\text{out}}}$ to be calibrated using data $\mathcal{Y}$
- The Bayesian model calibration framework allows computation of the distribution of the input parameters $\boldsymbol{X}$ conditioned on the data (posterior)

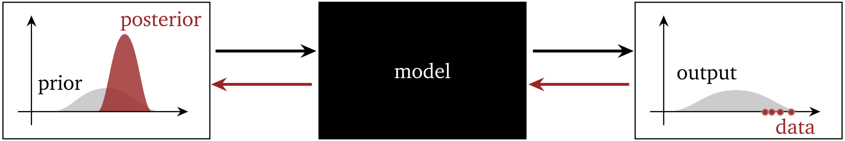


In practice the posterior computation is challenging

Motivation

Typical problem:

- Given computational model predicting observables $\mathcal{M} : \mathcal{D}_{\boldsymbol{X}} \to \mathbb{R}^{N_{\text{out}}}$ to be calibrated using data $\mathcal{Y}$
- The Bayesian model calibration framework allows computation of the distribution of the input parameters $\boldsymbol{X}$ conditioned on the data (posterior)



In practice the posterior computation is challenging

Outline

Framework

Given parameters $\boldsymbol{X} \sim \pi(\boldsymbol{x})$ and measurements $\mathcal{Y}$, the Bayesian inverse problem reads:

$$\pi(\boldsymbol{x}|\mathcal{Y}) = \frac{\mathcal{L}(\boldsymbol{x};\mathcal{Y})\pi(\boldsymbol{x})}{Z} \quad \text{where} \quad Z = \int_{\mathcal{D}_{\boldsymbol{X}}} \mathcal{L}(\boldsymbol{x};\mathcal{Y})\pi(\boldsymbol{x})\mathrm{d}\boldsymbol{x}$$

with:

- $\mathcal{L} : \mathcal{D}_{\boldsymbol{X}} \to \mathbb{R}^+$ likelihood function (measure of how well the model fits the data)
- $\pi(\boldsymbol{x}|\mathcal{Y})$ posterior density function

Quantities of Interest (QoI)

Often one is interested in expectations under $\pi(\boldsymbol{x}|\mathcal{Y})$ of QoI: $h(\boldsymbol{x}) : \mathcal{D}_{\boldsymbol{X}} \to \mathbb{R}$:

$$\mathbb{E}\left[h(\boldsymbol{X})|\mathcal{Y}\right] = \int_{\mathcal{D}_{\boldsymbol{X}}} h(\boldsymbol{x})\pi(\boldsymbol{x}|\mathcal{Y})\mathrm{d}\boldsymbol{x}$$

e.g. posterior moments.

Framework

Given parameters $\boldsymbol{X} \sim \pi(\boldsymbol{x})$ and measurements $\mathcal{Y}$, the Bayesian inverse problem reads:

$$\pi(\boldsymbol{x}|\mathcal{Y}) = \frac{\mathcal{L}(\boldsymbol{x};\mathcal{Y})\pi(\boldsymbol{x})}{Z} \quad \text{where} \quad Z = \int_{\mathcal{D}_{\boldsymbol{X}}} \mathcal{L}(\boldsymbol{x};\mathcal{Y})\pi(\boldsymbol{x})\mathrm{d}\boldsymbol{x}$$

with:

- $\mathcal{L} : \mathcal{D}_{\boldsymbol{X}} \to \mathbb{R}^+$ likelihood function (measure of how well the model fits the data)
- $\pi(\boldsymbol{x}|\mathcal{Y})$ posterior density function

Quantities of Interest (QoI)

Often one is interested in expectations under $\pi(\boldsymbol{x}|\mathcal{Y})$ of QoI: $h(\boldsymbol{x}) : \mathcal{D}_{\boldsymbol{X}} \to \mathbb{R}$:

$$\mathbb{E}\left[h(\boldsymbol{X})|\mathcal{Y}\right] = \int_{\mathcal{D}_{\boldsymbol{X}}} h(\boldsymbol{x})\pi(\boldsymbol{x}|\mathcal{Y})\mathrm{d}\boldsymbol{x}$$

e.g. posterior moments.

Outline

Solution

No analytical solution (exception: conjugate distributions). Typically MCMC based approaches to sample from:

$$\boldsymbol{X}|\mathcal{Y} \sim \pi(\boldsymbol{x}|\mathcal{Y})$$

- Metropolis-Hastings Hastings (1970)
- Hamiltonian Monte Carlo Neal (2012)
- Affine invariant ensemble sampler Goodman and Weare (2010)
- ...

Combined with simulation-based forward propagation to estimate QoI ($\mathbb{E}\left[h(\boldsymbol{X})|\mathcal{Y}\right]$).

Problems

- Require tuning & post-processing
- No clear convergence criterion
- Multimodality in the posterior

Solution

No analytical solution (exception: conjugate distributions). Typically MCMC based approaches to sample from:

$$\boldsymbol{X}|\mathcal{Y} \sim \pi(\boldsymbol{x}|\mathcal{Y})$$

- Metropolis-Hastings Hastings (1970)
- Hamiltonian Monte Carlo Neal (2012)
- Affine invariant ensemble sampler Goodman and Weare (2010)
- ...

Combined with simulation-based forward propagation to estimate QoI ($\mathbb{E}\left[h(\boldsymbol{X})|\mathcal{Y}\right]$).

Problems

- Require tuning & post-processing
- No clear convergence criterion
- Multimodality in the posterior

Outline

Spectral likelihood expansions (SLE)

Principle

Assuming $\boldsymbol{X} \sim \pi(\boldsymbol{x}) = \prod_{i=1}^{M} \pi(x_i)$ (independent), SLE approximates $\mathcal{L}(\boldsymbol{X})$ (for fixed observations $\mathcal{Y}$) with a finite sum of orthonormal polynomials:

$$\mathcal{L}(\boldsymbol{X}) = \sum_{\boldsymbol{\alpha} \in \mathbb{N}^M} a_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(\boldsymbol{X}) \approx \sum_{\boldsymbol{\alpha} \in \mathcal{A}} a_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(\boldsymbol{X})$$

where $\Psi_{\boldsymbol{\alpha}} \stackrel{\text{def}}{=} \prod_{i=1}^{M} \psi_{\alpha_i}^{(i)}(x_i)$ and $\int_{\mathcal{D}_{X_i}} \psi_j^{(i)}(x_i) \psi_k^{(i)}(x_i) \pi(x_i) \mathrm{d}x_i = \delta_{jk}$

Nagel & Sudret, J. Comput. Phys. (2016)

Practical implementation

There exist numerous techniques to compute polynomial chaos expansions (PCEs). Here we use least angle regression with adaptive maximum polynomial degrees.

Blatman & Sudret, J. Comput. Phys. (2011)

Spectral likelihood expansions (SLE)

Principle

Assuming $\boldsymbol{X} \sim \pi(\boldsymbol{x}) = \prod_{i=1}^{M} \pi(x_i)$ (independent), SLE approximates $\mathcal{L}(\boldsymbol{X})$ (for fixed observations $\mathcal{Y}$) with a finite sum of orthonormal polynomials:

$$\mathcal{L}(\boldsymbol{X}) = \sum_{\boldsymbol{\alpha} \in \mathbb{N}^M} a_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(\boldsymbol{X}) \approx \sum_{\boldsymbol{\alpha} \in \mathcal{A}} a_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(\boldsymbol{X})$$

where $\Psi_{\boldsymbol{\alpha}} \overset{\text{def}}{=} \prod_{i=1}^{M} \psi_{\alpha_i}^{(i)}(x_i)$ and $\int_{\mathcal{D}_{X_i}} \psi_j^{(i)}(x_i) \psi_k^{(i)}(x_i) \pi(x_i) \mathrm{d}x_i = \delta_{jk}$

Nagel & Sudret, J. Comput. Phys. (2016)

Practical implementation

There exist numerous techniques to compute polynomial chaos expansions (PCEs). Here we use least angle regression with adaptive maximum polynomial degrees.

Blatman & Sudret, J. Comput. Phys. (2011)

Spectral likelihood expansions (SLE)

Analytical posterior

Following the computation of the coefficients $a_{\boldsymbol{\alpha}}$, the full posterior distribution or QoI can be computed analytically:

Post-processing $a_{\boldsymbol{\alpha}}$

$$Z \approx a_{\mathbf{0}}$$

$$\pi(\boldsymbol{x}|\mathcal{Y}) \approx \frac{\pi(\boldsymbol{x})}{Z} \sum_{\boldsymbol{\alpha}\in\mathcal{A}} a_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(\boldsymbol{x})$$

$$\mathbb{E}\left[h(\boldsymbol{X})|\mathcal{Y}\right] \approx \frac{1}{a_{\mathbf{0}}} \sum_{\boldsymbol{\alpha}\in\mathcal{A}} a_{\boldsymbol{\alpha}} b_{\boldsymbol{\alpha}} \quad \text{after} \quad h(\boldsymbol{X}) \approx \sum_{\boldsymbol{\alpha}\in\mathcal{A}} b_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(\boldsymbol{X})$$

Spectral likelihood expansions (SLE)

Analytical posterior

Following the computation of the coefficients $a_{\boldsymbol{\alpha}}$, the full posterior distribution or QoI can be computed analytically:

Post-processing $a_{\boldsymbol{\alpha}}$

$$Z \approx a_{\mathbf{0}}$$

$$\pi(\boldsymbol{x}|\mathcal{Y}) \approx \frac{\pi(\boldsymbol{x})}{Z} \sum_{\boldsymbol{\alpha}\in\mathcal{A}} a_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(\boldsymbol{x})$$

$$\mathbb{E}\left[h(\boldsymbol{X})|\mathcal{Y}\right] \approx \frac{1}{a_{\mathbf{0}}} \sum_{\boldsymbol{\alpha}\in\mathcal{A}} a_{\boldsymbol{\alpha}} b_{\boldsymbol{\alpha}} \quad \text{after} \quad h(\boldsymbol{X}) \approx \sum_{\boldsymbol{\alpha}\in\mathcal{A}} b_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(\boldsymbol{X})$$

However, for a reasonable approximation this requires very high degree polynomials!

Stochastic spectral embedding (SSE)

Sequentially approximate likelihood with sum of K low-degree PCEs $f_{PC}^{k}(\boldsymbol{X}) = \sum_{\boldsymbol{\alpha}\in\mathcal{A}^k} a_{\boldsymbol{\alpha}}^{k}\Psi_{\boldsymbol{\alpha}}^{k}(\boldsymbol{X})$ on multiple levels and subdomains $\mathcal{D}_{\boldsymbol{X}}^{k}$ using the residual of the previous PCEs

$$\mathcal{L}(\boldsymbol{X}) \approx \sum_{k=1}^{K} f_{PC}^{k}(\boldsymbol{X})\mathbf{1}_{\mathcal{D}_{\boldsymbol{X}}^{k}}(\boldsymbol{X})$$

Stochastic spectral embedding (SSE)

Sequentially approximate likelihood with sum of K low-degree PCEs $f^k_{PC}(\boldsymbol{X}) = \sum_{\boldsymbol{\alpha}\in\mathcal{A}^k} a^k_{\boldsymbol{\alpha}}\Psi^k_{\boldsymbol{\alpha}}(\boldsymbol{X})$ on multiple levels and subdomains $\mathcal{D}^k_{\boldsymbol{X}}$ using the residual of the previous PCEs

$$\mathcal{L}(\boldsymbol{X}) \approx \sum_{k=1}^{K} f^k_{PC}(\boldsymbol{X})\mathbf{1}_{\mathcal{D}^k_{\boldsymbol{X}}}(\boldsymbol{X})$$

Post-processing $a^k_{\boldsymbol{\alpha}}$

Preserves the post-processing properties of SLE:

$$Z \approx \sum_{k=1}^{K} c_k a^k_{\mathbf{0}}, \quad \text{where} \quad c_k = \int_{\mathcal{D}^k_{\boldsymbol{X}}} \pi(\boldsymbol{x})\mathrm{d}\boldsymbol{x}$$

$$\pi(\boldsymbol{x}|\mathcal{Y}) \approx \frac{\pi(\boldsymbol{x})}{Z}\sum_{k=1}^{K} f^k_{PC}(\boldsymbol{x})\mathbf{1}_{\mathcal{D}^k_{\boldsymbol{X}}}(\boldsymbol{x})$$

$$\mathbb{E}\left[h(\boldsymbol{X})|\mathcal{Y}\right] \approx \frac{1}{Z}\sum_{k=1}^{K} c_k \cdot \sum_{\boldsymbol{\alpha}\in\mathcal{A}^k} a^k_{\boldsymbol{\alpha}} b^k_{\boldsymbol{\alpha}} \quad \text{after} \quad h(\boldsymbol{x}) \approx \sum_{\boldsymbol{\alpha}\in\mathcal{A}^k} b^k_{\boldsymbol{\alpha}}\Psi^k_{\boldsymbol{\alpha}}(\boldsymbol{x})$$

Stochastic spectral embedding (SSE)

Sequentially approximate likelihood with sum of K low-degree PCEs $f_{PC}^{k}(\boldsymbol{X}) = \sum_{\boldsymbol{\alpha}\in\mathcal{A}^k} a_{\boldsymbol{\alpha}}^{k}\Psi_{\boldsymbol{\alpha}}^{k}(\boldsymbol{X})$ on multiple levels and subdomains $\mathcal{D}_{\boldsymbol{X}}^{k}$ using the residual of the previous PCEs

$$\mathcal{L}(\boldsymbol{X}) \approx \sum_{k=1}^{K} f_{PC}^{k}(\boldsymbol{X})\mathbf{1}_{\mathcal{D}_{\boldsymbol{X}}^{k}}(\boldsymbol{X})$$

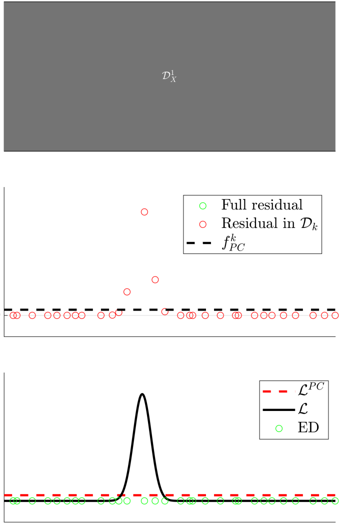
How to sequentially build embedded PCEs?

Construction of SSE - 1D example

Algorithm

Initialize $\mathcal{D}_{\boldsymbol{X}}^1$ with initial error $\varepsilon^1 = \infty$ and an initial experimental design (ED), repeat:

1. Select $\mathcal{D}_{\boldsymbol{X}}^k$ with maximum error (among terminal domains)
2. *Optional:* Enrich ED
3. Construct PCE f_{PC}^k using the residual of $\mathcal{L}$ on ED
4. Split $\mathcal{D}_{\boldsymbol{X}}^k$ into 2 subdomains by cutting $\mathcal{D}_{\boldsymbol{X}}^k$ in half
5. Estimate error ε^i in each subdomain.

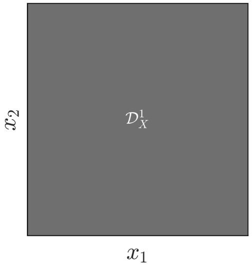


Construction of SSE - higher dimensions

Algorithm

Initialize $\mathcal{D}_{\boldsymbol{X}}^1$ with initial error $\varepsilon^1 = \infty$, repeat:

1. Select $\mathcal{D}_{\boldsymbol{X}}^k$ with maximum error (among terminal domains);
2. *Optional:* Enrich experimental design (ED)
3. Construct PCE f_{PC}^k using the residual of $\mathcal{L}$ on ED
4. Split $\mathcal{D}_{\boldsymbol{X}}^k$ into $2M$ candidate subdomains by cutting $\mathcal{D}_{\boldsymbol{X}}^k$ in half along each dimension
5. Estimate error ε_{cand}^i in each candidate subdomain. Keep subdomain pair with maximum error

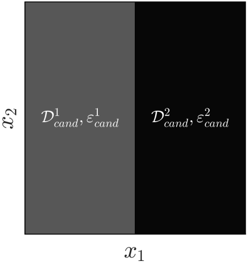


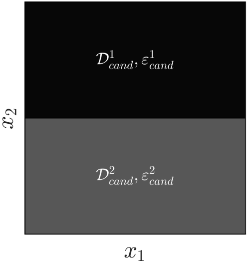
Construction of SSE - higher dimensions

Algorithm

Initialize $\mathcal{D}^1_{\boldsymbol{X}}$ with initial error $\varepsilon^1 = \infty$, repeat:

1. Select $\mathcal{D}^k_{\boldsymbol{X}}$ with maximum error (among terminal domains);
2. *Optional:* Enrich experimental design (ED)
3. Construct PCE f^k_{PC} using the residual of $\mathcal{L}$ on ED
4. Split $\mathcal{D}^k_{\boldsymbol{X}}$ into $2M$ candidate subdomains by cutting $\mathcal{D}^k_{\boldsymbol{X}}$ in half along each dimension
5. Estimate error ε^i_{cand} in each candidate subdomain. Keep subdomain pair with maximum error



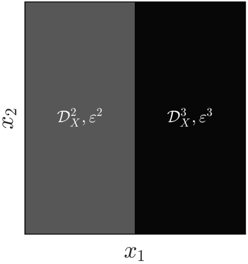


Construction of SSE - higher dimensions

Algorithm

Initialize $\mathcal{D}_{\boldsymbol{X}}^1$ with initial error $\varepsilon^1 = \infty$, repeat:

1. Select $\mathcal{D}_{\boldsymbol{X}}^k$ with maximum error (among terminal domains);
2. *Optional:* Enrich experimental design (ED)
3. Construct PCE f_{PC}^k using the residual of $\mathcal{L}$ on ED
4. Split $\mathcal{D}_{\boldsymbol{X}}^k$ into $2M$ candidate subdomains by cutting $\mathcal{D}_{\boldsymbol{X}}^k$ in half along each dimension
5. Estimate error ε_{cand}^i in each candidate subdomain. Keep subdomain pair with maximum error

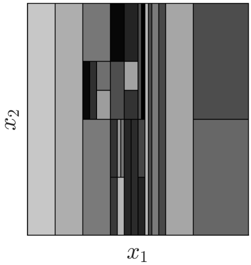


Construction of SSE - higher dimensions

Algorithm

Initialize $\mathcal{D}_{\boldsymbol{X}}^1$ with initial error $\varepsilon^1 = \infty$, repeat:

1. Select $\mathcal{D}_{\boldsymbol{X}}^k$ with maximum error (among terminal domains);
2. *Optional:* Enrich experimental design (ED)
3. Construct PCE f_{PC}^k using the residual of $\mathcal{L}$ on ED
4. Split $\mathcal{D}_{\boldsymbol{X}}^k$ into $2M$ candidate subdomains by cutting $\mathcal{D}_{\boldsymbol{X}}^k$ in half along each dimension
5. Estimate error ε_{cand}^i in each candidate subdomain. Keep subdomain pair with maximum error

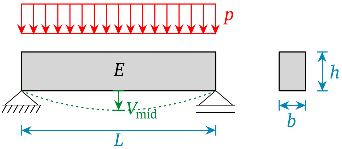


Outline

Simply supported beam

- Mid-span deflection data
  $\mathcal{Y} = \{y_1, \dots, y_N\}$
- Computational forward model:

$$\mathcal{M}(E,p) = \frac{5pL^4}{32Ebh^3}$$



- Likelihood with $\boldsymbol{x} \stackrel{\text{def}}{=} (E, p, \sigma^2)$:

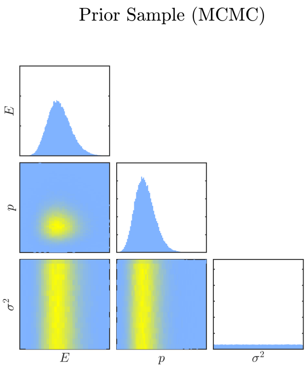
$$\mathcal{L}(\boldsymbol{x};\mathcal{Y}) = \prod_{i=1}^{N} \mathcal{N}(y_i | \mathcal{M}(E,p), \sigma^2)$$

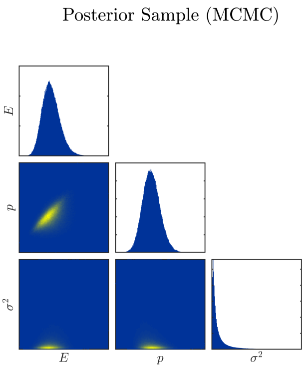
- Prior distributions:

x_i	$\pi(x_i)$
E	lognormal
p	lognormal
σ^2	uniform

Simply supported beam

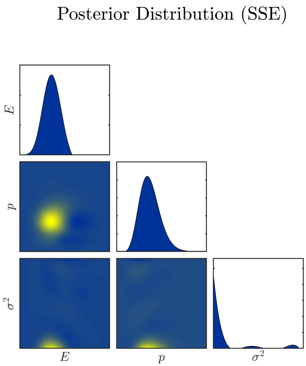
The problem is solved with MCMC as a reference solution (10^5 sample points):

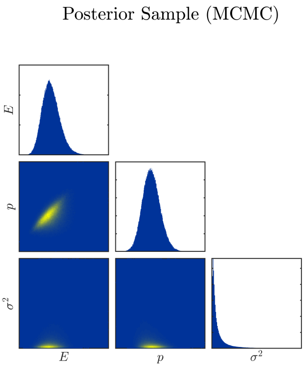




Simply supported beam

The problem is now solved using the SSE method:





Conjugate prior

- Data $\mathcal{Y} = \{\boldsymbol{y}_1, \dots, \boldsymbol{y}_N\}$ with $\boldsymbol{y}_i \in \mathbb{R}^4$
- Computational forward model:

$$\mathcal{M}(\boldsymbol{x}) = \boldsymbol{x}$$

- Likelihood:

$$\mathcal{L}(\boldsymbol{x}; \mathcal{Y}) = \prod_{i=1}^{N} \mathcal{N}(\boldsymbol{y}_i | \mathcal{M}(\boldsymbol{x}), \boldsymbol{\Sigma})$$

- Prior distribution:

$$\pi(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{\mu}_{\text{prior}}, \boldsymbol{\Sigma}_{\text{prior}})$$

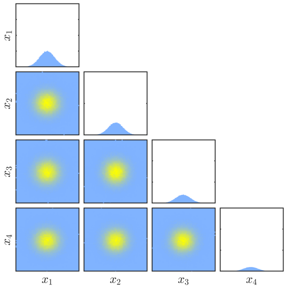
- Posterior distribution is given by multivariate normal distribution ($\boldsymbol{\mu}_{\text{post}}$ and $\boldsymbol{\Sigma}_{\text{post}}$ can be computed analytically):

$$\mathcal{N}(\boldsymbol{\mu}_{\text{post}}, \boldsymbol{\Sigma}_{\text{post}})$$

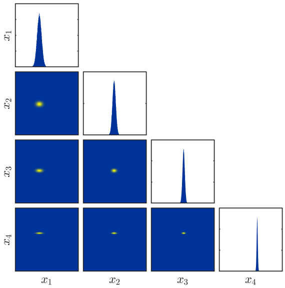
Conjugate prior

The analytical solution is available:

Prior Sample (MCMC)

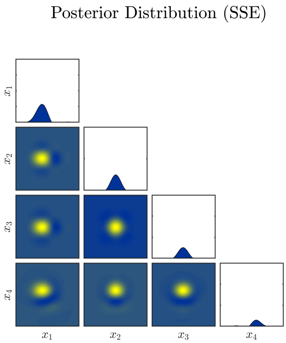


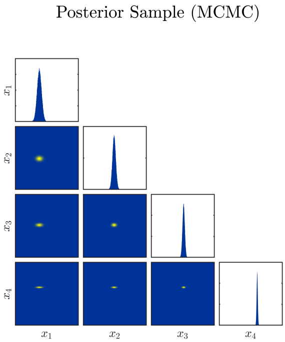
Posterior Sample (MCMC)



Conjugate prior

The problem is now solved using the SSE method:



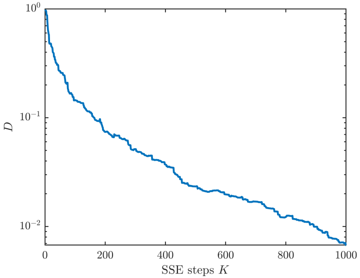


Conjugate prior

Analyze convergence by:

$$D \stackrel{\mathsf{def}}{=} \frac{\mathbb{E}\left[\left(\pi(\boldsymbol{X}|\mathcal{Y}) - \pi^{SSE}(\boldsymbol{X}|\mathcal{Y})\right)^2\right]}{\mathrm{Var}\left[\pi(\boldsymbol{X}|\mathcal{Y})\right]}$$

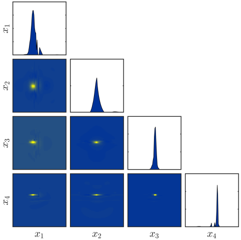
with $\boldsymbol{X} \sim \pi(\boldsymbol{x})$



Outline

Conclusion

Posterior Distribution (SSE)



- Bayesian model calibration is a powerful tool for model calibration with many applications
- MCMC-based sampling approaches for computing posterior characteristics are often inpracticable in real-world problems
- Spectral likelihood expansion is a alternative approach. However, accurate likelihood approximation requires prohibitively high polynomial degrees
- Stochastic spectral embedding maintains the post-processing capabilities of SLE while remaining computationally feasible

Outlook

- Improve subdomain selection by improved local error estimate ε_i
- Develop reliable convergence diagnostics
- Investigate higher dimensions

www.uqlab.com