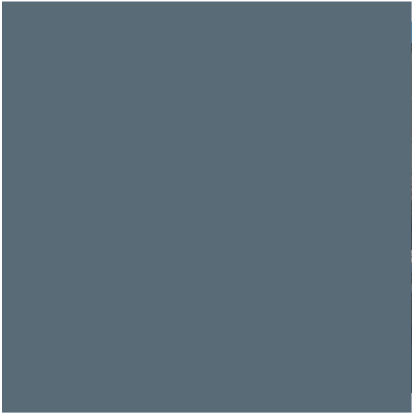
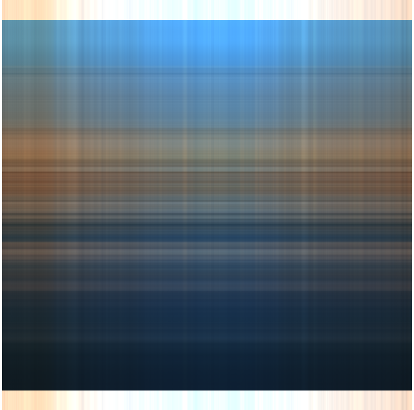
Hessian-based sampling in high dimensions for goal-oriented model order reduction

Peng Chen Omar Ghattas

Center for Computational Geosciences and Optimization
The Institute for Computational Engineering and Sciences
The University of Texas at Austin

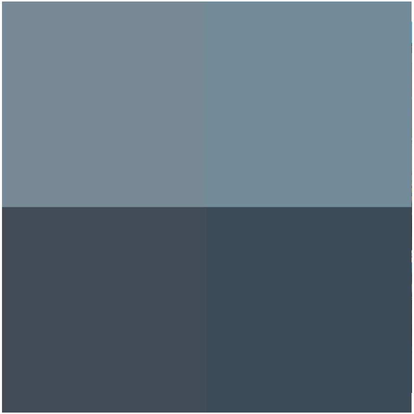
QUIET 2017 - Quantification of Uncertainty: Improving Efficiency and Technology
SISSA, International School for Advanced Studies, Trieste, 18-21 July 2017



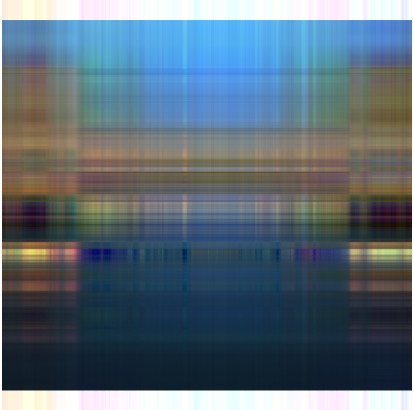


pixels $K = 1^2$

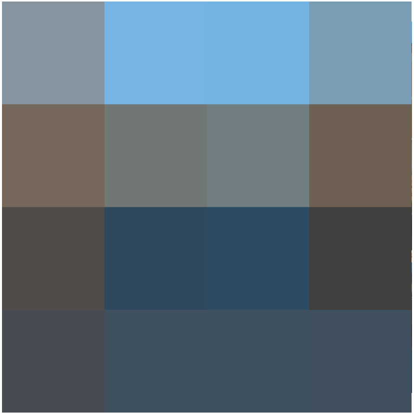
modes $K = 1$

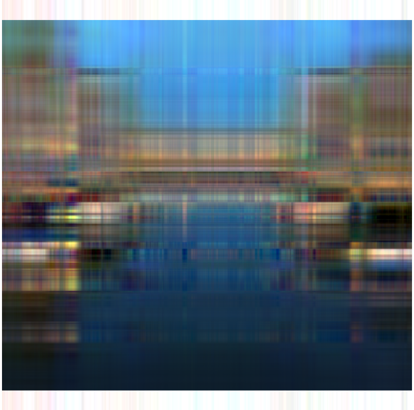


pixels $K = 2^2$



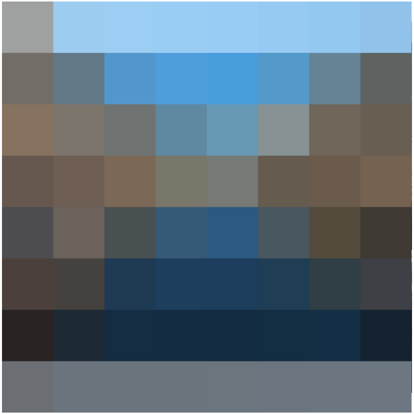
modes $K = 2$

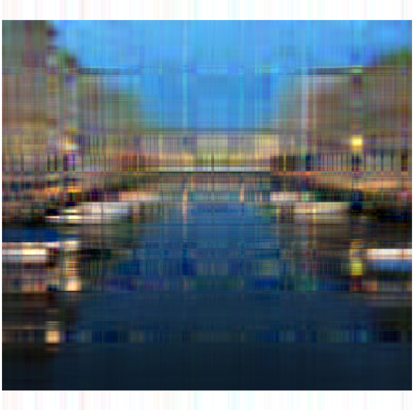




pixels $K = 4^2$

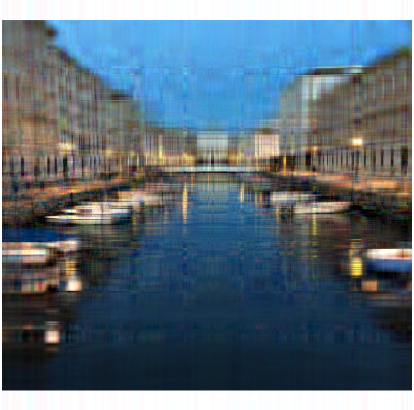
modes $K = 4$





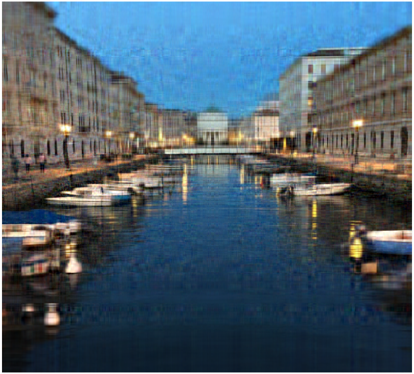
pixels $K = 8^2$

modes $K = 8$



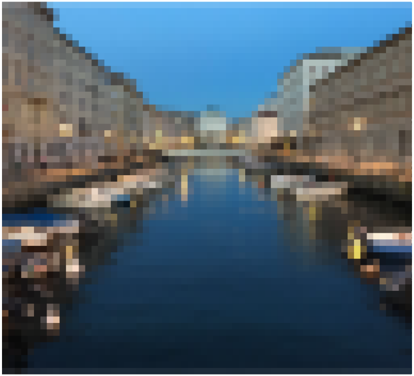
pixels $K = 16^2$

modes $K = 16$



pixels $K = 32^2$

modes $K = 32$



pixels $K = 64^2$

modes $K = 64$

$\#$ pixels $K = 128^2$

$\#$ modes $K = 128$

pixels $K = 256^2$

modes $K = 256$

pixels $K = 512^2$

modes $K = 512$

pixels $K = 1024^2$

modes $K = 1024$

Outline

Parameters

- Let $P \subset \mathbb{R}^K$ denote a K-dimensional parameter space, where $K \in \mathbb{N} \cup \infty$.

$$\boldsymbol{p} = (p_1, \dots, p_K) \in P.$$

- The parameter $\boldsymbol{p}$ lives in a box, w.l.o.g., $P = [-\sqrt{3}, \sqrt{3}]^K$, with uniform distribution

$$\boldsymbol{p} \sim \mu = \mathcal{U}([-\sqrt{3}, \sqrt{3}]^K),$$

with mean $\bar{\boldsymbol{p}} = \boldsymbol{0}$, and covariance $\mathbb{C} = \mathbb{I}$.

- The parameter $\boldsymbol{p}$ lives in the whole space, i.e., $P = \mathbb{R}^K$, with Gaussian distribution

$$\boldsymbol{p} \sim \mu = \mathcal{N}(\bar{\boldsymbol{p}}, \mathbb{C}),$$

with mean $\bar{\boldsymbol{p}}$, and covariance $\mathbb{C}$, s.p.d.

- Eg., $\mathbb{C}$ is discretized from a covariance operator $\mathcal{C}$, given by

$$\mathcal{C} = (-\delta\triangle + \gamma I)^{-\alpha},$$

which is self adjoint, positive, and of trace class.

Parameters

- Let $P \subset \mathbb{R}^K$ denote a K-dimensional parameter space, where $K \in \mathbb{N} \cup \infty$.

$$\boldsymbol{p} = (p_1, \dots, p_K) \in P.$$

- The parameter $\boldsymbol{p}$ lives in a box, w.l.o.g., $P = [-\sqrt{3}, \sqrt{3}]^K$, with uniform distribution

$$\boldsymbol{p} \sim \mu = \mathcal{U}([-\sqrt{3}, \sqrt{3}]^K),$$

with mean $\bar{\boldsymbol{p}} = \boldsymbol{0}$, and covariance $\mathbb{C} = \mathbb{I}$.

- The parameter $\boldsymbol{p}$ lives in the whole space, i.e., $P = \mathbb{R}^K$, with Gaussian distribution

$$\boldsymbol{p} \sim \mu = \mathcal{N}(\bar{\boldsymbol{p}}, \mathbb{C}),$$

with mean $\bar{\boldsymbol{p}}$, and covariance $\mathbb{C}$, s.p.d.

- Eg., $\mathbb{C}$ is discretized from a covariance operator $\mathcal{C}$, given by

$$\mathcal{C} = (-\delta\triangle + \gamma I)^{-\alpha},$$

which is self adjoint, positive, and of trace class.

Parameters

- Let $P \subset \mathbb{R}^K$ denote a K-dimensional parameter space, where $K \in \mathbb{N} \cup \infty$.

$$\boldsymbol{p} = (p_1, \dots, p_K) \in P.$$

- The parameter $\boldsymbol{p}$ lives in a box, w.l.o.g., $P = [-\sqrt{3}, \sqrt{3}]^K$, with uniform distribution

$$\boldsymbol{p} \sim \mu = \mathcal{U}([-\sqrt{3}, \sqrt{3}]^K),$$

with mean $\bar{\boldsymbol{p}} = \mathbf{0}$, and covariance $\mathbb{C} = \mathbb{I}$.

- The parameter $\boldsymbol{p}$ lives in the whole space, i.e., $P = \mathbb{R}^K$, with Gaussian distribution

$$\boldsymbol{p} \sim \mu = \mathcal{N}(\bar{\boldsymbol{p}}, \mathbb{C}),$$

with mean $\bar{\boldsymbol{p}}$, and covariance $\mathbb{C}$, s.p.d.

- Eg., $\mathbb{C}$ is discretized from a covariance operator $\mathcal{C}$, given by

$$\mathcal{C} = (-\delta \triangle + \gamma I)^{-\alpha},$$

which is self adjoint, positive, and of trace class.

Parameters

- Let $P \subset \mathbb{R}^K$ denote a K-dimensional parameter space, where $K \in \mathbb{N} \cup \infty$.

$$\boldsymbol{p} = (p_1, \dots, p_K) \in P.$$

- The parameter $\boldsymbol{p}$ lives in a box, w.l.o.g., $P = [-\sqrt{3}, \sqrt{3}]^K$, with uniform distribution

$$\boldsymbol{p} \sim \mu = \mathcal{U}([-\sqrt{3}, \sqrt{3}]^K),$$

$$\text{with mean } \bar{\boldsymbol{p}} = \boldsymbol{0}, \text{ and covariance } \mathbb{C} = \mathbb{I}.$$

- The parameter $\boldsymbol{p}$ lives in the whole space, i.e., $P = \mathbb{R}^K$, with Gaussian distribution

$$\boldsymbol{p} \sim \mu = \mathcal{N}(\bar{\boldsymbol{p}}, \mathbb{C}),$$

$$\text{with mean } \bar{\boldsymbol{p}}, \text{ and covariance } \mathbb{C}, \text{ s.p.d.}$$

- Eg., $\mathbb{C}$ is discretized from a covariance operator $\mathcal{C}$, given by

$$\mathcal{C} = (-\delta\triangle + \gamma I)^{-\alpha},$$

which is self adjoint, positive, and of trace class.

Parametric PDEs

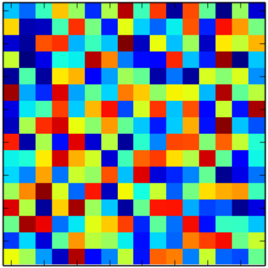
- Let V denote a Hilbert space with dual V'. Given $\boldsymbol{p} \in P$, μ-a.e., find $u \in V$ such that

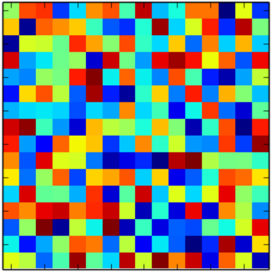
$$a(u, v; \boldsymbol{p}) = f(v) \quad \forall v \in V.$$

- $a(\cdot, \cdot; \boldsymbol{p}) : V \times V \to \mathbb{R}$ is a bilinear form, e.g.,

$$a(u, v; \boldsymbol{p}) = \int_D \kappa(\boldsymbol{p}) \nabla u \cdot \nabla v dx.$$

- $f(\cdot) \in V'$ is a linear functional.
- $s(\boldsymbol{p}) = s(u(\boldsymbol{p})) \in \mathbb{R}$ is a QoI.





$K = 16^2 = 256$

Ex 1. heat conduction in thermal blocks

$$\kappa(\boldsymbol{p}) = \sum_{k=1}^{K} k^{-\beta} \chi_{D_k}(x) p_k$$

$$\boldsymbol{p} \sim \mathcal{U}([-\sqrt{3}, \sqrt{3}]^K)$$

Parametric PDEs

- Let V denote a Hilbert space with dual V'.
  Given $\boldsymbol{p} \in P$, μ-a.e., find $u \in V$ such that

$$a(u, v; \boldsymbol{p}) = f(v) \quad \forall v \in V.$$

- $a(\cdot, \cdot; \boldsymbol{p}) : V \times V \to \mathbb{R}$ is a bilinear form, e.g.,

$$a(u, v; \boldsymbol{p}) = \int_D \kappa(\boldsymbol{p}) \nabla u \cdot \nabla v dx.$$

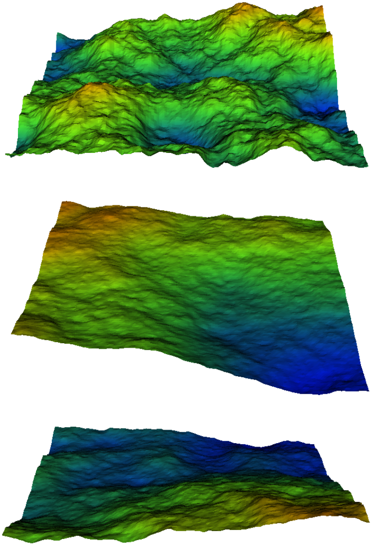
- $f(\cdot) \in V'$ is a linear functional.
- $s(\boldsymbol{p}) = s(u(\boldsymbol{p})) \in \mathbb{R}$ is a QoI.

Ex 2. subsurface flow in a porous medium

$$\kappa(\boldsymbol{p}) = e^{\boldsymbol{p}}$$

log-normal diffusion with

$$\boldsymbol{p} \in \mathcal{N}(\bar{\boldsymbol{p}}, \mathbb{C})$$



$K = 129^2 = 16,641$

Finite element approximation

Finite element space V_h,

$$\dim(V_h) = N_h$$

Given $\boldsymbol{p} \in P$, find $u_h \in V_h$ s.t.

$$a(u_h, v_h; \boldsymbol{p}) = f(v_h) \ \forall v_h \in V_h$$

The algebraic system is

$$\mathbb{A}_h(\boldsymbol{p})\boldsymbol{u}_h = \mathbf{f}_h$$

$$\mathbb{V} = [\boldsymbol{\psi}_1, \dots, \boldsymbol{\psi}_N]$$

$$\mathbb{V}^T \boldsymbol{u}_h = \boldsymbol{u}_N$$

$$\mathbb{V}^T \mathbb{A}_h(\boldsymbol{p})\mathbb{V} = \mathbb{A}_N(\boldsymbol{p})$$

$$\mathbb{V}^T \mathbf{f}_h = \mathbf{f}_N$$

Reduced basis approximation

Reduced basis space $V_N \subset V_h$,

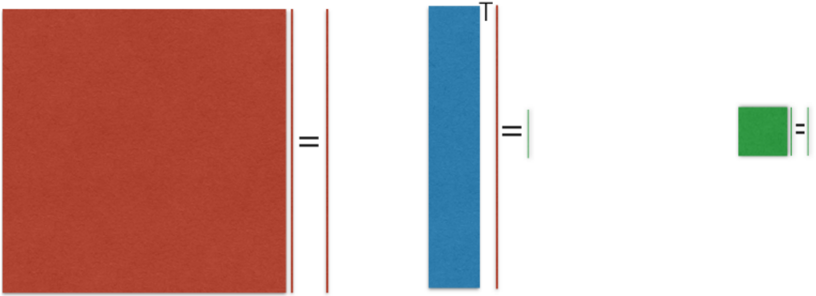
$$\dim(V_N) = N$$

Given $\boldsymbol{p} \in P$, find $u_N \in V_N$ s.t.

$$a(u_N, v_N; \boldsymbol{p}) = f(v_N) \ \forall v_N \in V_N$$

The algebraic system is

$$\mathbb{A}_N(\boldsymbol{p})\boldsymbol{u}_N = \mathbf{f}_N$$



Model order reduction – algorithms (Maday, Patera, Rozza, et. al.)

POD/SVD

Samples

$$\Xi_t = \{\boldsymbol{p}^n, n = 1, \dots, N_t\}$$

Compute snapshots

$$\mathbb{U} = [\boldsymbol{u}_h(\boldsymbol{p}^1), \dots, \boldsymbol{u}_h(\boldsymbol{p}^{N_t})]$$

Perform SVD

$$\mathbb{U} = \mathbb{V}\Sigma\mathbb{W}^T$$

Extract bases $\mathbb{V}[1:N,:]$

$$N = \operatorname{argmin}_n \mathcal{E}_n(\Sigma) \geq 1 - \varepsilon$$

Greedy algorithm

Samples

$$\Xi_t = \{\boldsymbol{p}^n, n = 1, \dots, N_t\}$$

Initialize V_N for $N = 1$ as

$$V_N = \text{span}\{u_h(\boldsymbol{p}^1)\}$$

Pick next sample such that

$$\boldsymbol{p}^{N+1} = \operatorname{argmax}_{\boldsymbol{p}\in\Xi_t} \Delta_N(\boldsymbol{p})$$

Update bases V_{N+1} as

$$V_N \oplus \text{span}\{u_h(\boldsymbol{p}^{N+1})\}$$

Offline-Online

Affine assumption/approx.

$$a = \sum_{q=1}^{Q} \theta_q(\boldsymbol{p}) a_q$$

Offline computation once

$$\mathbb{A}_N^q = \mathbb{V}^T \mathbb{A}_h^q \mathbb{V}, \mathbf{f}_N = \mathbb{V}^T \mathbf{f}_h$$

Online assemble

$$\mathbb{A}_N(\boldsymbol{p}) = \sum_{q=1}^{Q} \theta_q(\boldsymbol{p}) \mathbb{A}_N^q$$

Online solve and evaluate

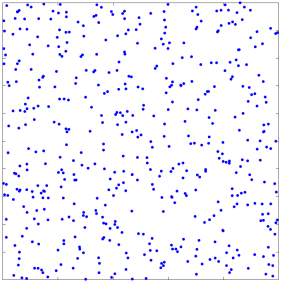
$$\mathbb{A}_N(\boldsymbol{p})\boldsymbol{u}_N = \mathbf{f}_N, s(\boldsymbol{p}) = \mathbf{s}_N^T \boldsymbol{u}_N$$

Goal-oriented a-posteriori error estimate $\Delta_N(\boldsymbol{p})$ – dual weighted residual

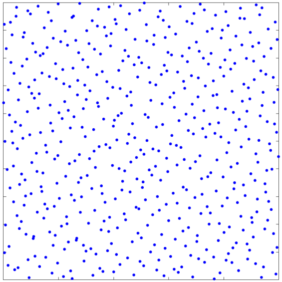
$\Delta_N(\boldsymbol{p}) = f(\varphi_N) - a(u_N, \varphi_N; \boldsymbol{p})$, where dual Prob.: $a(w_N, \varphi_N; \boldsymbol{p}) = s(w_N)\ \forall w_N \in W_N$.

$$\Delta_N(\boldsymbol{p}) = \bar{\mathbf{f}}_N^T \varphi_N - \sum_{q=1}^{Q} \theta_q(\boldsymbol{p}) \varphi_N^T \bar{\mathbb{A}}_N^q \boldsymbol{u}_N, \text{ where } \bar{\mathbf{f}}_N = \mathbb{W}^T \mathbf{f}_h, \text{ and } \bar{\mathbb{A}}_N^q = \mathbb{W}^T \mathbb{A}_h^q \mathbb{V}$$

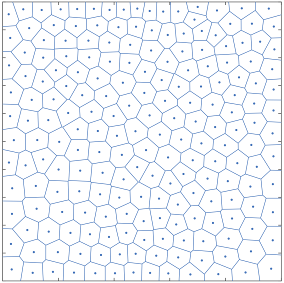
Model order reduction – samples



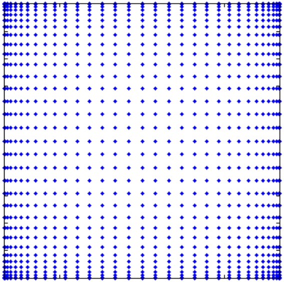
random



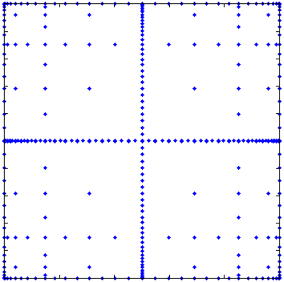
quasi-random



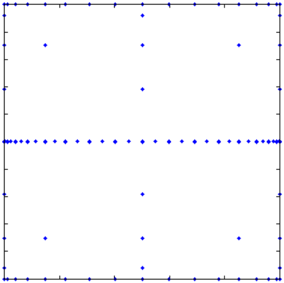
centroidal Voronoi tessellation
(Du, Gunzburger, et. al.)



tensor grid

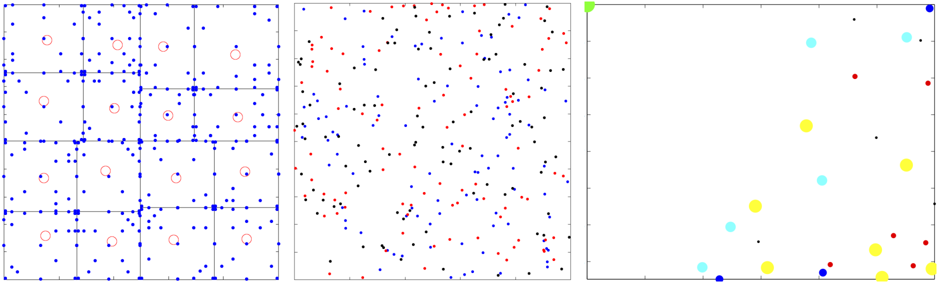


sparse grid
(Liao, Elman, et. al.)



anisotropic sparse grid
(C., Schwab, et. al.)

Model order reduction – samples



hp-adaptive-rb (Eftang, Patera, et. al.)

adaptive-add-remove (Hesthaven, Stamm, et. al.)

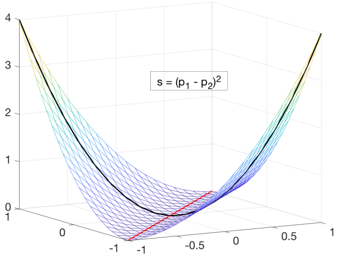
hybrid goal-oriented adaptive (C., Quarteroni, et. al.)

Outline

Hessian-based sampling – Hessian

- Hessian $\mathbb{H} \in \mathbb{R}^{K \times K}$, the second-order partial derivatives of s with respect to $\boldsymbol{p}$, i.e.,

$$\mathbb{H}_{kl} = \frac{\partial^2 s}{\partial p_k \partial p_l}, \quad k, l \in 1, \dots, K.$$



- The eigendirections corresponding to the **leading eigenvalues** of the Hessian are the directions along which the s changes the most in the parameter space.

$$\mathbb{H} = \begin{pmatrix} 2 & -2 \\ -2 & 2 \end{pmatrix}, \lambda_1 = 1, \lambda_2 = 0$$

- Thus, **sampling in the subspace** of leading eigendirections presumably provide the most representative samples that capture the variation of the s, at least locally.

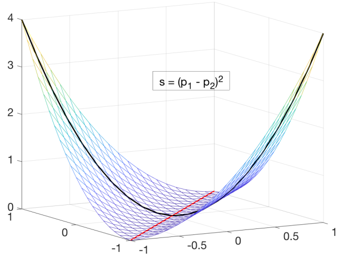
- It has been widely used in large-scale computation for

solving nonlinear problems,
control/optimization, **parameter estimation**, **data assimilation**

Hessian-based sampling – Hessian

- Hessian $\mathbb{H} \in \mathbb{R}^{K \times K}$, the second-order partial derivatives of s with respect to $\boldsymbol{p}$, i.e.,

$$\mathbb{H}_{kl} = \frac{\partial^2 s}{\partial p_k \partial p_l}, \quad k, l \in 1, \dots, K.$$



$$\mathbb{H} = \begin{pmatrix} 2 & -2 \\ -2 & 2 \end{pmatrix}, \lambda_1 = 1, \lambda_2 = 0$$

- The eigendirections corresponding to the **leading eigenvalues** of the Hessian are the directions along which the s changes the most in the parameter space.

- Thus, **sampling in the subspace** of leading eigendirections presumably provide the most representative samples that capture the variation of the s, at least locally.

- It has been widely used in large-scale computation for

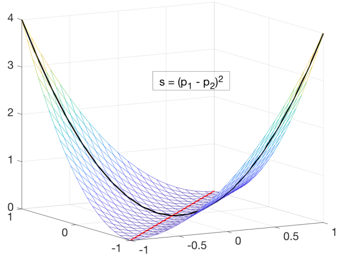
solving nonlinear problems,
control/optimization, **parameter estimation**, **data assimilation**

Hessian-based sampling – Hessian

- Hessian $\mathbb{H} \in \mathbb{R}^{K \times K}$, the second-order partial derivatives of s with respect to $\boldsymbol{p}$, i.e.,

$$\mathbb{H}_{kl} = \frac{\partial^2 s}{\partial p_k \partial p_l}, \quad k, l \in 1, \dots, K.$$

- The eigendirections corresponding to the **leading eigenvalues** of the Hessian are the directions along which the s changes the most in the parameter space.



$$\mathbb{H} = \begin{pmatrix} 2 & -2 \\ -2 & 2 \end{pmatrix}, \lambda_1 = 1, \lambda_2 = 0$$

- Thus, **sampling in the subspace** of leading eigendirections presumably provide the most representative samples that capture the variation of the s, at least locally.

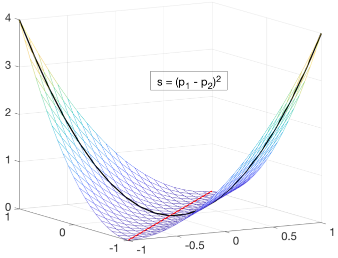
- It has been widely used in large-scale computation for

  solving nonlinear problems,
  control/optimization, parameter estimation, data assimilation

Hessian-based sampling – Hessian

- Hessian $\mathbb{H} \in \mathbb{R}^{K\times K}$, the second-order partial derivatives of s with respect to $\boldsymbol{p}$, i.e.,

$$\mathbb{H}_{kl} = \frac{\partial^2 s}{\partial p_k \partial p_l}, \quad k, l \in 1, \ldots, K.$$



- The eigendirections corresponding to the **leading eigenvalues** of the Hessian are the directions along which the s changes the most in the parameter space.

$$\mathbb{H} = \begin{pmatrix} 2 & -2 \\ -2 & 2 \end{pmatrix}, \lambda_1 = 1, \lambda_2 = 0$$

- Thus, **sampling in the subspace** of leading eigendirections presumably provide the most representative samples that capture the variation of the s, at least locally.

- It has been widely used in large-scale computation for

solving nonlinear problems,
control/optimization, **parameter estimation**, **data assimilation**

Hessian-based sampling – $\mathbb{C}$-preconditioned Hessian

- Let s_{quad} denote the **quadratic/Taylor** approximation of s given by

$$s_{\text{quad}}(\boldsymbol{p}) = s(\bar{\boldsymbol{p}}) + \boldsymbol{g}_{\bar{\boldsymbol{p}}}^T(\boldsymbol{p} - \bar{\boldsymbol{p}}) + \frac{1}{2}(\boldsymbol{p} - \bar{\boldsymbol{p}})^T \mathbb{H}_{\bar{\boldsymbol{p}}}(\boldsymbol{p} - \bar{\boldsymbol{p}}), \tag{1}$$

  where $\boldsymbol{g}_{\bar{\boldsymbol{p}}}$ and $\mathbb{H}_{\bar{\boldsymbol{p}}}$ represent the **gradient** and the **Hessian** of s at $\bar{\boldsymbol{p}}$.
- The expectation of s_{quad} can be computed as

$$\mathbb{E}[s_{\text{quad}}] = s(\bar{\boldsymbol{p}}) + \frac{1}{2}\text{tr}(\tilde{\mathbb{H}}_{\bar{\boldsymbol{p}}}), \tag{2}$$

  $\text{tr}(\tilde{\mathbb{H}}_{\bar{\boldsymbol{p}}})$: trace of the **covariance preconditioned Hessian** $\tilde{\mathbb{H}}_{\bar{\boldsymbol{p}}} = \mathbb{C}\mathbb{H}_{\bar{\boldsymbol{p}}}$ at the mean $\bar{\boldsymbol{p}}$.
- It is equivalent to the sum of all the eigenvalues, i.e.,

$$\text{tr}(\tilde{\mathbb{H}}_{\bar{\boldsymbol{p}}}) = \sum_{k=1}^{K} \lambda_k(\tilde{\mathbb{H}}_{\bar{\boldsymbol{p}}}). \tag{3}$$

- If λ_k **decay fast**, sampling in a **low-dimensional** subspace of eigenvectors:

$$\boldsymbol{p}_L = \sum_{l=1}^{L} (\boldsymbol{p}, \varphi_l)_2 \varphi_l. \tag{4}$$

Hessian-based sampling – $\mathbb{C}$-preconditioned Hessian

- Let s_{quad} denote the **quadratic/Taylor** approximation of s given by

$$s_{\text{quad}}(\boldsymbol{p}) = s(\bar{\boldsymbol{p}}) + \boldsymbol{g}_{\bar{\boldsymbol{p}}}^T(\boldsymbol{p} - \bar{\boldsymbol{p}}) + \frac{1}{2}(\boldsymbol{p} - \bar{\boldsymbol{p}})^T \mathbb{H}_{\bar{\boldsymbol{p}}}(\boldsymbol{p} - \bar{\boldsymbol{p}}), \tag{1}$$

  where $\boldsymbol{g}_{\bar{\boldsymbol{p}}}$ and $\mathbb{H}_{\bar{\boldsymbol{p}}}$ represent the **gradient** and the **Hessian** of s at $\bar{\boldsymbol{p}}$.
- The expectation of s_{quad} can be computed as

$$\mathbb{E}[s_{\text{quad}}] = s(\bar{\boldsymbol{p}}) + \frac{1}{2}\text{tr}(\tilde{\mathbb{H}}_{\bar{\boldsymbol{p}}}), \tag{2}$$

  $\text{tr}(\tilde{\mathbb{H}}_{\bar{\boldsymbol{p}}})$: trace of the **covariance preconditioned Hessian** $\tilde{\mathbb{H}}_{\bar{\boldsymbol{p}}} = \mathbb{C}\mathbb{H}_{\bar{\boldsymbol{p}}}$ at the mean $\bar{\boldsymbol{p}}$.
- It is equivalent to the sum of all the eigenvalues, i.e.,

$$\text{tr}(\tilde{\mathbb{H}}_{\bar{\boldsymbol{p}}}) = \sum_{k=1}^{K} \lambda_k(\tilde{\mathbb{H}}_{\bar{\boldsymbol{p}}}). \tag{3}$$

- If λ_k **decay fast**, sampling in a **low-dimensional** subspace of eigenvectors:

$$\boldsymbol{p}_L = \sum_{l=1}^{L} (\boldsymbol{p}, \varphi_l)_2 \varphi_l. \tag{4}$$

Hessian-based sampling – $\mathbb{C}$-preconditioned Hessian

- Let s_{quad} denote the **quadratic/Taylor** approximation of s given by

$$s_{\text{quad}}(\boldsymbol{p}) = s(\bar{\boldsymbol{p}}) + \boldsymbol{g}_{\bar{\boldsymbol{p}}}^T(\boldsymbol{p} - \bar{\boldsymbol{p}}) + \frac{1}{2}(\boldsymbol{p} - \bar{\boldsymbol{p}})^T \mathbb{H}_{\bar{\boldsymbol{p}}}(\boldsymbol{p} - \bar{\boldsymbol{p}}), \quad (1)$$

  where $\boldsymbol{g}_{\bar{\boldsymbol{p}}}$ and $\mathbb{H}_{\bar{\boldsymbol{p}}}$ represent the **gradient** and the **Hessian** of s at $\bar{\boldsymbol{p}}$.
- The expectation of s_{quad} can be computed as

$$\mathbb{E}[s_{\text{quad}}] = s(\bar{\boldsymbol{p}}) + \frac{1}{2}\text{tr}(\tilde{\mathbb{H}}_{\bar{\boldsymbol{p}}}), \quad (2)$$

  $\text{tr}(\tilde{\mathbb{H}}_{\bar{\boldsymbol{p}}})$: trace of the **covariance preconditioned Hessian** $\tilde{\mathbb{H}}_{\bar{\boldsymbol{p}}} = \mathbb{C}\mathbb{H}_{\bar{\boldsymbol{p}}}$ at the mean $\bar{\boldsymbol{p}}$.
- It is equivalent to the sum of all the eigenvalues, i.e.,

$$\text{tr}(\tilde{\mathbb{H}}_{\bar{\boldsymbol{p}}}) = \sum_{k=1}^{K} \lambda_k(\tilde{\mathbb{H}}_{\bar{\boldsymbol{p}}}). \quad (3)$$

- If λ_k **decay fast**, sampling in a **low-dimensional** subspace of eigenvectors:

$$\boldsymbol{p}_L = \sum_{l=1}^{L} (\boldsymbol{p}, \varphi_l)_2 \varphi_l. \quad (4)$$

Hessian-based sampling – $\mathbb{C}$-preconditioned Hessian

- Let s_{quad} denote the **quadratic/Taylor** approximation of s given by

$$s_{\text{quad}}(\boldsymbol{p}) = s(\bar{\boldsymbol{p}}) + \boldsymbol{g}_{\bar{\boldsymbol{p}}}^T(\boldsymbol{p} - \bar{\boldsymbol{p}}) + \frac{1}{2}(\boldsymbol{p} - \bar{\boldsymbol{p}})^T \mathbb{H}_{\bar{\boldsymbol{p}}}(\boldsymbol{p} - \bar{\boldsymbol{p}}), \tag{1}$$

  where $\boldsymbol{g}_{\bar{\boldsymbol{p}}}$ and $\mathbb{H}_{\bar{\boldsymbol{p}}}$ represent the **gradient** and the **Hessian** of s at $\bar{\boldsymbol{p}}$.
- The expectation of s_{quad} can be computed as

$$\mathbb{E}[s_{\text{quad}}] = s(\bar{\boldsymbol{p}}) + \frac{1}{2}\text{tr}(\tilde{\mathbb{H}}_{\bar{\boldsymbol{p}}}), \tag{2}$$

  $\text{tr}(\tilde{\mathbb{H}}_{\bar{\boldsymbol{p}}})$: trace of the **covariance preconditioned Hessian** $\tilde{\mathbb{H}}_{\bar{\boldsymbol{p}}} = \mathbb{C}\mathbb{H}_{\bar{\boldsymbol{p}}}$ at the mean $\bar{\boldsymbol{p}}$.
- It is equivalent to the sum of all the eigenvalues, i.e.,

$$\text{tr}(\tilde{\mathbb{H}}_{\bar{\boldsymbol{p}}}) = \sum_{k=1}^{K} \lambda_k(\tilde{\mathbb{H}}_{\bar{\boldsymbol{p}}}). \tag{3}$$

- If λ_k **decay fast**, sampling in a **low-dimensional** subspace of eigenvectors:

$$\boldsymbol{p}_L = \sum_{l=1}^{L} (\boldsymbol{p}, \boldsymbol{\varphi}_l)_2 \boldsymbol{\varphi}_l. \tag{4}$$

Hessian-based sampling – from local to global Hessian

- **Hessian at the mean**: let $(\lambda_k, \varphi_k)_{k=1}^K$ denote the eigenpairs of $\tilde{\mathbb{H}}_{\bar{p}} = \mathbb{C}\mathbb{H}_{\bar{p}}$, or equivalently the generalized eigenpairs of $(\mathbb{H}_{\bar{p}}, \mathbb{C}^{-1})$ for computational efficiency

$$\mathbb{H}_{\bar{p}}\varphi_k = \lambda_k \mathbb{C}^{-1}\varphi_k. \quad (5)$$

- **Averaged Hessian**: we can replace the Hessian at the mean by

$$\mathbb{H} = \int_P \mathbb{H}_p d\mu(p) \approx \frac{1}{M}\sum_{m=1}^{M} \mathbb{H}_{p^m}, \quad (6)$$

with p^m sampled according to its probability distribution μ.

- **Combined Hessian**: we compute the eigenvectors of Hessian at different samples

$$\mathbb{H}_{p^m}\varphi_k^m = \lambda_k^m \mathbb{C}^{-1}\varphi_k^m, \quad m = 1, \dots, M. \quad (7)$$

Then we combine them with weights (e.g. $w_k^m = \sqrt{\lambda_k^m}$) and compress them by SVD

$$\Phi = (w_1^1\varphi_1^1, \dots, w_{L_1}^1\varphi_{L_1}^1, \dots, w_1^M\varphi_1^M, \dots, w_{L_M}^M\varphi_{L_M}^M). \quad (8)$$

- **Hessian at the mean**: let $(\lambda_k, \boldsymbol{\varphi}_k)_{k=1}^K$ denote the eigenpairs of $\tilde{\mathbb{H}}_{\bar{\boldsymbol{p}}} = \mathbb{C}\mathbb{H}_{\bar{\boldsymbol{p}}}$, or equivalently the generalized eigenpairs of $(\mathbb{H}_{\bar{\boldsymbol{p}}}, \mathbb{C}^{-1})$ for computational efficiency

$$\mathbb{H}_{\bar{\boldsymbol{p}}}\boldsymbol{\varphi}_k = \lambda_k \mathbb{C}^{-1}\boldsymbol{\varphi}_k. \tag{5}$$

- **Averaged Hessian**: we can replace the Hessian at the mean by

$$\mathbb{H} = \int_P \mathbb{H}_{\boldsymbol{p}} d\mu(\boldsymbol{p}) \approx \frac{1}{M}\sum_{m=1}^{M} \mathbb{H}_{\boldsymbol{p}^m}, \tag{6}$$

with $\boldsymbol{p}^m$ sampled according to its probability distribution μ.

- **Combined Hessian**: we compute the eigenvectors of Hessian at different samples

$$\mathbb{H}_{\boldsymbol{p}^m}\varphi_k^m = \lambda_k^m \mathbb{C}^{-1}\varphi_k^m, \quad m = 1, \dots, M. \tag{7}$$

Then we combine them with weights (e.g. $w_k^m = \sqrt{\lambda_k^m}$) and compress them by SVD

$$\Phi = (w_1^1\varphi_1^1, \dots, w_{L_1}^1\varphi_{L_1}^1, \dots, w_1^M\varphi_1^M, \dots, w_{L_M}^M\varphi_{L_M}^M). \tag{8}$$

Hessian-based sampling – from local to global Hessian

- **Hessian at the mean**: let $(\lambda_k, \boldsymbol{\varphi}_k)_{k=1}^K$ denote the eigenpairs of $\tilde{\mathbb{H}}_{\bar{\boldsymbol{p}}} = \mathbb{C}\mathbb{H}_{\bar{\boldsymbol{p}}}$, or equivalently the generalized eigenpairs of $(\mathbb{H}_{\bar{\boldsymbol{p}}}, \mathbb{C}^{-1})$ for computational efficiency

$$\mathbb{H}_{\bar{\boldsymbol{p}}}\boldsymbol{\varphi}_k = \lambda_k \mathbb{C}^{-1}\boldsymbol{\varphi}_k. \tag{5}$$

- **Averaged Hessian**: we can replace the Hessian at the mean by

$$\mathbb{H} = \int_P \mathbb{H}_{\boldsymbol{p}} d\mu(\boldsymbol{p}) \approx \frac{1}{M}\sum_{m=1}^{M} \mathbb{H}_{\boldsymbol{p}^m}, \tag{6}$$

  with $\boldsymbol{p}^m$ sampled according to its probability distribution μ.

- **Combined Hessian**: we compute the eigenvectors of Hessian at different samples

$$\mathbb{H}_{\boldsymbol{p}^m}\boldsymbol{\varphi}_k^m = \lambda_k^m \mathbb{C}^{-1}\boldsymbol{\varphi}_k^m, \quad m = 1, \dots, M. \tag{7}$$

  Then we combine them with weights (e.g. $w_k^m = \sqrt{\lambda_k^m}$) and compress them by SVD

$$\Phi = (w_1^1\boldsymbol{\varphi}_1^1, \dots, w_{L_1}^1\boldsymbol{\varphi}_{L_1}^1, \dots, w_1^M\boldsymbol{\varphi}_1^M, \dots, w_{L_M}^M\boldsymbol{\varphi}_{L_M}^M). \tag{8}$$

Hessian-based sampling – Hessian action

- We employ a Lagrange multiplier method to compute the action of Hessian:

$$\mathcal{L}(u, v, \boldsymbol{p}) = s(u) + f(v) - a(u, v; \boldsymbol{p}), \tag{9}$$

where v is the adjoint variable or the Lagrange multiplier.

- With first order variation, we obtain the adjoint problem: find $v \in V$ such that

$$a(w, v; \boldsymbol{p}) = s(w) \quad \forall w \in V. \tag{10}$$

- Given $(u, v, \boldsymbol{p})$, we compute the Hessian action in $\hat{\boldsymbol{p}}$ by the second order variation

$$\begin{pmatrix} \partial_{uu}\mathcal{L} & \partial_{uv}\mathcal{L} & \partial_{up}\mathcal{L} \\ \partial_{vu}\mathcal{L} & \partial_{vv}\mathcal{L} & \partial_{vp}\mathcal{L} \\ \partial_{pu}\mathcal{L} & \partial_{pv}\mathcal{L} & \partial_{pp}\mathcal{L} \end{pmatrix} \begin{pmatrix} \hat{u} \\ \hat{v} \\ \hat{\boldsymbol{p}} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \mathbb{H}_{\boldsymbol{p}}\hat{\boldsymbol{p}} \end{pmatrix}, \tag{11}$$

the **incremental adjoint problem**: find $\hat{v} \in V$ such that

$$a(\tilde{u}, \hat{v}; \boldsymbol{p}) = -\partial_{\boldsymbol{p}} a(\tilde{u}, v; \boldsymbol{p})\hat{\boldsymbol{p}} \quad \forall \tilde{u} \in V, \tag{12}$$

the **incremental state problem**: find $\hat{u} \in V$ such that

$$a(\hat{u}, \tilde{v}; \boldsymbol{p}) = -\partial_{\boldsymbol{p}} a(u, \tilde{v}; \boldsymbol{p})\hat{\boldsymbol{p}} \quad \forall \tilde{v} \in V, \tag{13}$$

and the **Hessian action** in direction $\hat{\boldsymbol{p}}$ as

$$\mathbb{H}_{\boldsymbol{p}}\hat{\boldsymbol{p}} = -\partial_{\boldsymbol{p}} a(\hat{u}, v; \boldsymbol{p}) - \partial_{\boldsymbol{p}} a(u, \hat{v}; \boldsymbol{p}) - \partial_{\boldsymbol{p}\boldsymbol{p}} a(u, v; \boldsymbol{p})\hat{\boldsymbol{p}}. \tag{14}$$

Hessian-based sampling – Hessian action

- We employ a Lagrange multiplier method to compute the action of Hessian:

$$\mathcal{L}(u, v, \boldsymbol{p}) = s(u) + f(v) - a(u, v; \boldsymbol{p}), \quad (9)$$

  where v is the adjoint variable or the Lagrange multiplier.
- With first order variation, we obtain the adjoint problem: find $v \in V$ such that

$$a(w, v; \boldsymbol{p}) = s(w) \quad \forall w \in V. \quad (10)$$

- Given $(u, v, \boldsymbol{p})$, we compute the Hessian action in $\hat{\boldsymbol{p}}$ by the second order variation

$$\begin{pmatrix} \partial_{uu}\mathcal{L} & \partial_{uv}\mathcal{L} & \partial_{up}\mathcal{L} \\ \partial_{vu}\mathcal{L} & \partial_{vv}\mathcal{L} & \partial_{vp}\mathcal{L} \\ \partial_{pu}\mathcal{L} & \partial_{pv}\mathcal{L} & \partial_{pp}\mathcal{L} \end{pmatrix} \begin{pmatrix} \hat{u} \\ \hat{v} \\ \hat{\boldsymbol{p}} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \mathbb{H}_p \hat{\boldsymbol{p}} \end{pmatrix}, \quad (11)$$

  the **incremental adjoint problem**: find $\hat{v} \in V$ such that

$$a(\tilde{u}, \hat{v}; \boldsymbol{p}) = -\partial_p a(\tilde{u}, v; \boldsymbol{p})\hat{\boldsymbol{p}} \quad \forall \tilde{u} \in V, \quad (12)$$

  the **incremental state problem**: find $\hat{u} \in V$ such that

$$a(\hat{u}, \tilde{v}; \boldsymbol{p}) = -\partial_p a(u, \tilde{v}; \boldsymbol{p})\hat{\boldsymbol{p}} \quad \forall \tilde{v} \in V, \quad (13)$$

  and the **Hessian action** in direction $\hat{\boldsymbol{p}}$ as

$$\mathbb{H}_p \hat{\boldsymbol{p}} = -\partial_p a(\hat{u}, v; \boldsymbol{p}) - \partial_p a(u, \hat{v}; \boldsymbol{p}) - \partial_{pp} a(u, v; \boldsymbol{p})\hat{\boldsymbol{p}}. \quad (14)$$

Hessian-based sampling – Hessian action

- We employ a Lagrange multiplier method to compute the action of Hessian:

$$\mathcal{L}(u, v, \boldsymbol{p}) = s(u) + f(v) - a(u, v; \boldsymbol{p}), \tag{9}$$

  where v is the adjoint variable or the Lagrange multiplier.

- With first order variation, we obtain the adjoint problem: find $v \in V$ such that

$$a(w, v; \boldsymbol{p}) = s(w) \quad \forall w \in V. \tag{10}$$

- Given $(u, v, \boldsymbol{p})$, we compute the Hessian action in $\hat{\boldsymbol{p}}$ by the second order variation

$$\begin{pmatrix} \partial_{uu}\mathcal{L} & \partial_{uv}\mathcal{L} & \partial_{u\boldsymbol{p}}\mathcal{L} \\ \partial_{vu}\mathcal{L} & \partial_{vv}\mathcal{L} & \partial_{v\boldsymbol{p}}\mathcal{L} \\ \partial_{\boldsymbol{p}u}\mathcal{L} & \partial_{\boldsymbol{p}v}\mathcal{L} & \partial_{\boldsymbol{p}\boldsymbol{p}}\mathcal{L} \end{pmatrix} \begin{pmatrix} \hat{u} \\ \hat{v} \\ \hat{\boldsymbol{p}} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \mathbb{H}_{\boldsymbol{p}}\hat{\boldsymbol{p}} \end{pmatrix}, \tag{11}$$

  the **incremental adjoint problem**: find $\hat{v} \in V$ such that

$$a(\tilde{u}, \hat{v}; \boldsymbol{p}) = -\partial_{\boldsymbol{p}} a(\tilde{u}, v; \boldsymbol{p})\hat{\boldsymbol{p}} \quad \forall \tilde{u} \in V, \tag{12}$$

  the **incremental state problem**: find $\hat{u} \in V$ such that

$$a(\hat{u}, \tilde{v}; \boldsymbol{p}) = -\partial_{\boldsymbol{p}} a(u, \tilde{v}; \boldsymbol{p})\hat{\boldsymbol{p}} \quad \forall \tilde{v} \in V, \tag{13}$$

  and the **Hessian action** in direction $\hat{\boldsymbol{p}}$ as

$$\mathbb{H}_{\boldsymbol{p}}\hat{\boldsymbol{p}} = -\partial_{\boldsymbol{p}} a(\hat{u}, v; \boldsymbol{p}) - \partial_{\boldsymbol{p}} a(u, \hat{v}; \boldsymbol{p}) - \partial_{\boldsymbol{p}\boldsymbol{p}} a(u, v; \boldsymbol{p})\hat{\boldsymbol{p}}. \tag{14}$$

Hessian-based sampling – Randomized SVD

Algorithm 1 Randomized SVD for generalized eigenvalue problem $(\mathbb{H}_{\boldsymbol{p}}, \mathbb{C}^{-1})$

Input: $\mathbb{H}_{\boldsymbol{p}}, \mathbb{C}^{-1}$, the number of eigenpairs L, an oversampling factor $l = 5 \sim 10$.
Output: eigenpairs (Λ_L, Ψ_L) with $\Lambda_L = \text{diag}(\lambda_1, \dots, \lambda_L)$ and $\Psi_L = (\psi_1, \dots, \psi_L)$.

1. Draw a Gaussian random matrix $\Omega \in \mathbb{R}^{K\times(L+l)}$.
2. Compute $Y = \mathbb{C}(\mathbb{H}_{\boldsymbol{p}}\Omega)$.
3. Compute QR-factorization $Y = QR$ such that $Q^\top \mathbb{C}^{-1} Q = I_{L+l}$.
4. Form $T = Q^\top \mathbb{H}_{\boldsymbol{p}} Q$ and compute eigendecomposition $T = S\Lambda S^\top$.
5. Extract $\Lambda_L = \Lambda(1 : L, 1 : L)$ and $\Psi_L = QS_L$ with $S_L = S(:, 1 : L)$.

- **Computational cost** is dominated by two (so-called **double pass** algorithm)

$$\mathbb{H}_{\boldsymbol{p}}\Omega, \quad \mathbb{H}_{\boldsymbol{p}}Q$$

  which needs $4(L+l)$ linearized PDE solves, i.e., incremental problems.
- **Approximation error** is bounded by remaining eigenvalues $\leq C(L,l)(\sum_{l\geq L}\lambda_l^2)^{1/2}$.

Nathan Halko, Per-Gunnar Martinsson, and Joel A. Tropp. "Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions." SIAM review 53.2 (2011): 217-288.

Algorithm 2 Randomized SVD for generalized eigenvalue problem $(\mathbb{H}_{\boldsymbol{p}}, \mathbb{C}^{-1})$

Input: $\mathbb{H}_{\boldsymbol{p}}, \mathbb{C}^{-1}$, the number of eigenpairs L, an oversampling factor $l = 5 \sim 10$.
Output: eigenpairs (Λ_L, Ψ_L) with $\Lambda_L = \text{diag}(\lambda_1, \dots, \lambda_L)$ and $\Psi_L = (\psi_1, \dots, \psi_L)$.

1. Draw a Gaussian random matrix $\Omega \in \mathbb{R}^{K \times (L+l)}$.
2. Compute $Y = \mathbb{C}(\mathbb{H}_{\boldsymbol{p}}\Omega)$.
3. Compute QR-factorization $Y = QR$ such that $Q^\top \mathbb{C}^{-1} Q = I_{L+l}$.
4. Form $T = Q^\top \mathbb{H}_{\boldsymbol{p}} Q$ and compute eigendecomposition $T = S \Lambda S^\top$.
5. Extract $\Lambda_L = \Lambda(1:L, 1:L)$ and $\Psi_L = QS_L$ with $S_L = S(:, 1:L)$.

- **Computational cost** is dominated by two (so-called **double pass** algorithm)

$$\mathbb{H}_{\boldsymbol{p}}\Omega, \quad \mathbb{H}_{\boldsymbol{p}}Q$$

  which needs $4(L+l)$ linearized PDE solves, i.e., incremental problems.
- **Approximation error** is bounded by remaining eigenvalues $\leq C(L,l)(\sum_{l \geq L} \lambda_l^2)^{1/2}$.

Nathan Halko, Per-Gunnar Martinsson, and Joel A. Tropp. "Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions." SIAM review 53.2 (2011): 217-288.

Hessian-based sampling – Randomized SVD

Algorithm 3 Randomized SVD for generalized eigenvalue problem $(\mathbb{H}_{\boldsymbol{p}}, \mathbb{C}^{-1})$

Input: $\mathbb{H}_{\boldsymbol{p}}, \mathbb{C}^{-1}$, the number of eigenpairs L, an oversampling factor $l = 5 \sim 10$.
Output: eigenpairs (Λ_L, Ψ_L) with $\Lambda_L = \text{diag}(\lambda_1, \dots, \lambda_L)$ and $\Psi_L = (\psi_1, \dots, \psi_L)$.

1. Draw a Gaussian random matrix $\Omega \in \mathbb{R}^{K \times (L+l)}$.
2. Compute $Y = \mathbb{C}(\mathbb{H}_{\boldsymbol{p}}\Omega)$.
3. Compute QR-factorization $Y = QR$ such that $Q^\top \mathbb{C}^{-1} Q = I_{L+l}$.
4. Form $T = Q^\top \mathbb{H}_{\boldsymbol{p}} Q$ and compute eigendecomposition $T = S\Lambda S^\top$.
5. Extract $\Lambda_L = \Lambda(1:L, 1:L)$ and $\Psi_L = QS_L$ with $S_L = S(:, 1:L)$.

- **Computational cost** is dominated by two (so-called **double pass** algorithm)

$$\mathbb{H}_{\boldsymbol{p}}\Omega, \quad \mathbb{H}_{\boldsymbol{p}}Q$$

  which needs $4(L+l)$ linearized PDE solves, i.e., incremental problems.
- **Approximation error** is bounded by remaining eigenvalues $\leq C(L,l)(\sum_{l \geq L} \lambda_l^2)^{1/2}$.

Nathan Halko, Per-Gunnar Martinsson, and Joel A. Tropp. "Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions." SIAM review 53.2 (2011): 217-288.

Outline

- Let V denote a Hilbert space with dual V'. Given $\boldsymbol{p} \in P$, μ-a.e., find $u \in V$ such that

$$a(u, v; \boldsymbol{p}) = f(v) \quad \forall v \in V.$$

- $a(\cdot, \cdot; \boldsymbol{p}) : V \times V \to \mathbb{R}$ is a bilinear form, e.g.,

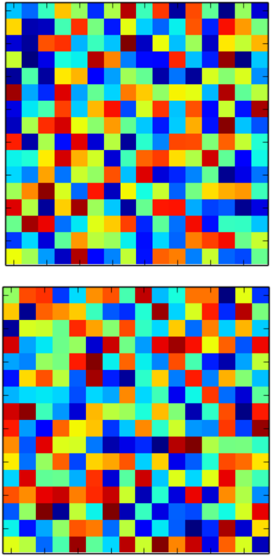
$$a(u, v; \boldsymbol{p}) = \int_D \kappa(\boldsymbol{p}) \nabla u \cdot \nabla v dx.$$

- $f = 0$, $u = 1$ on top, $u = 0$ on bottom.
- The QoI $s(u(\boldsymbol{p})) = 100 \int_{[0,0.1]^2} u(\boldsymbol{p}) dx$.
- $N_t = 1000$, uniform mesh of 65×65.

Ex 1. heat conduction in thermal blocks

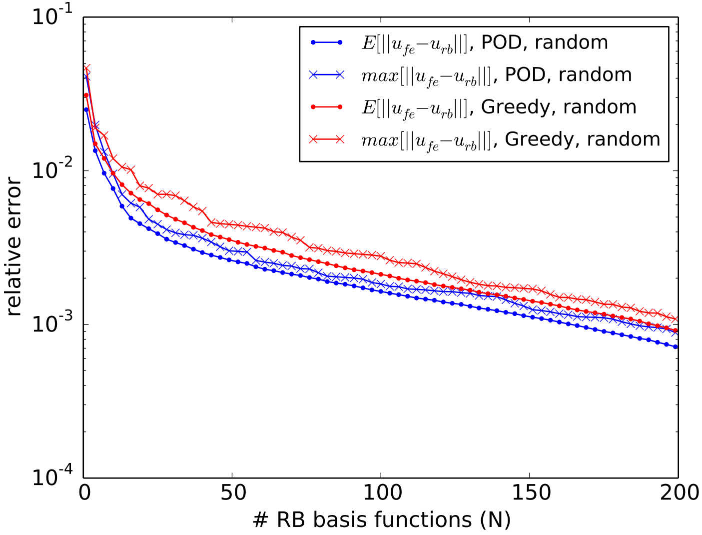
$$\kappa(\boldsymbol{p}) = \sum_{k=1}^{K} k^{-\beta} \chi_{D_k}(x) p_k, \quad \beta = 1$$

$$\boldsymbol{p} \sim \mathcal{U}([-\sqrt{3}, \sqrt{3}]^K)$$

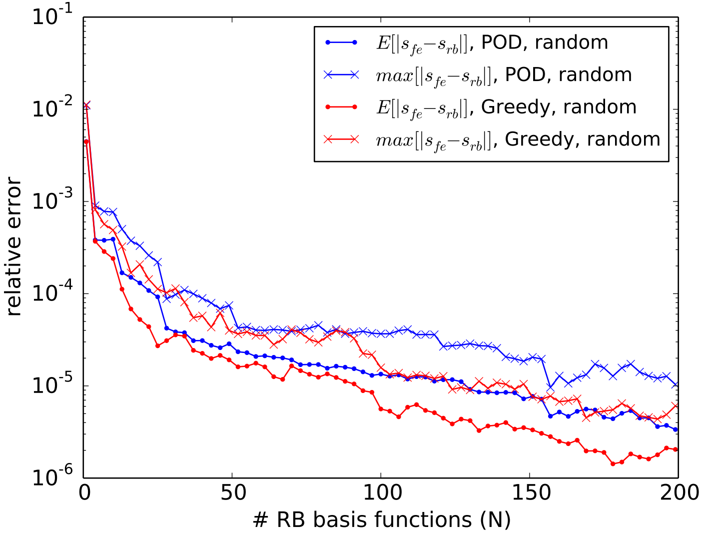


$K = 16^2 = 256$

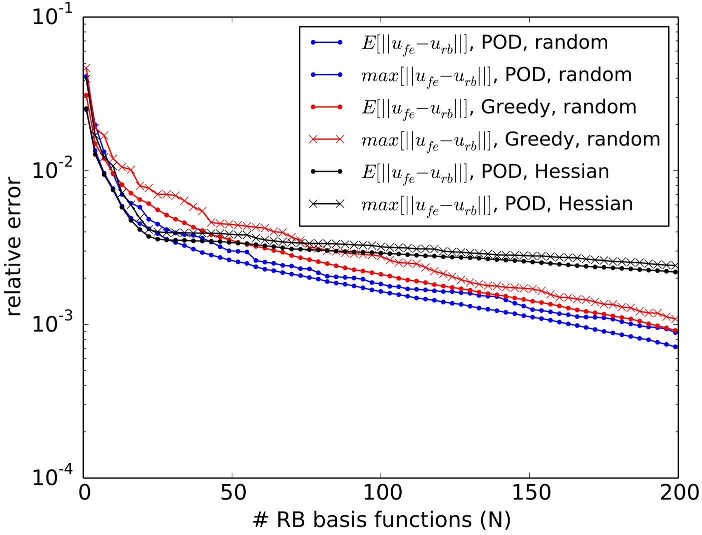
RB errors POD vs Greedy for solution



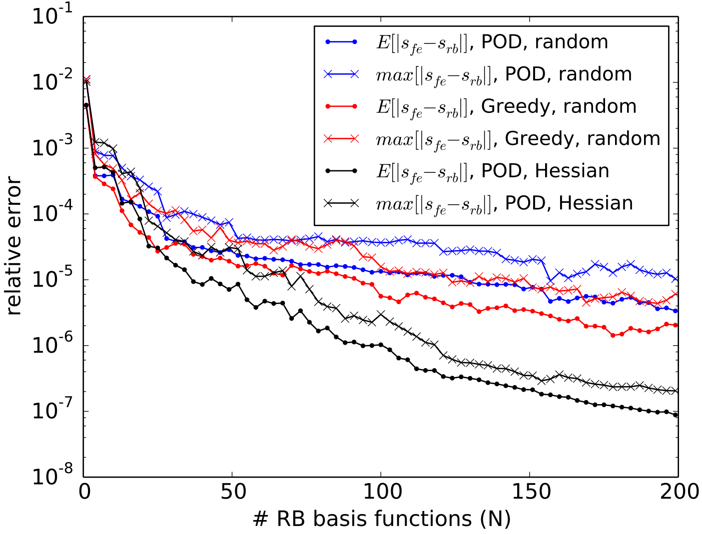
RB errors POD vs Greedy for QoI



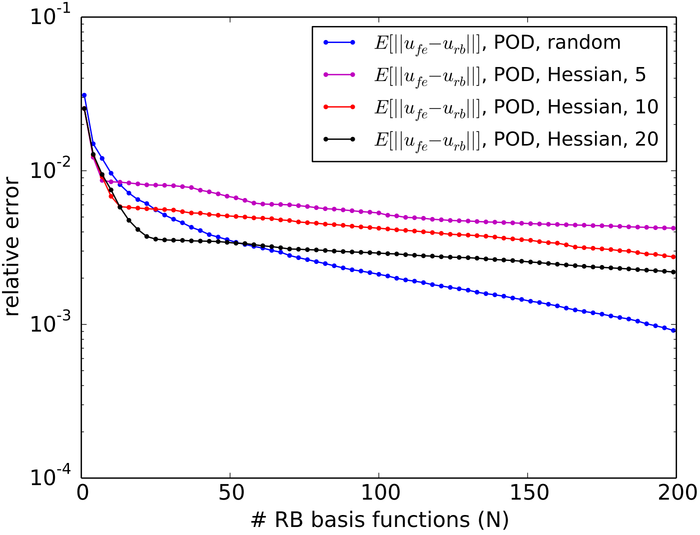
RB errors random vs Hessian for solution



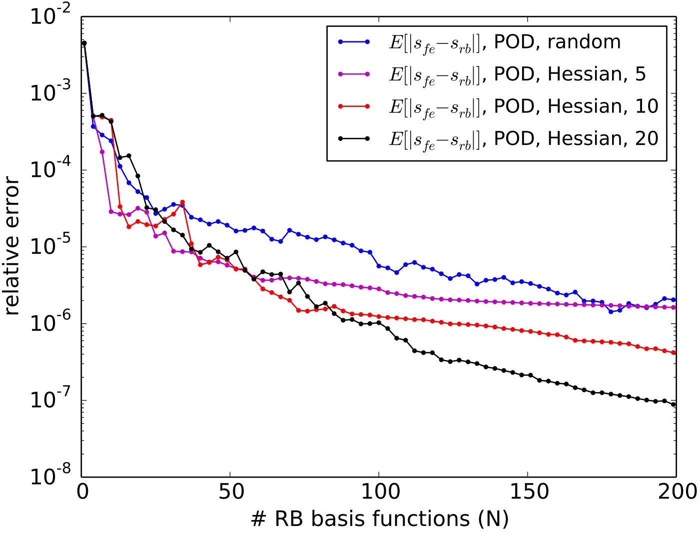
RB errors random vs Hessian for QoI



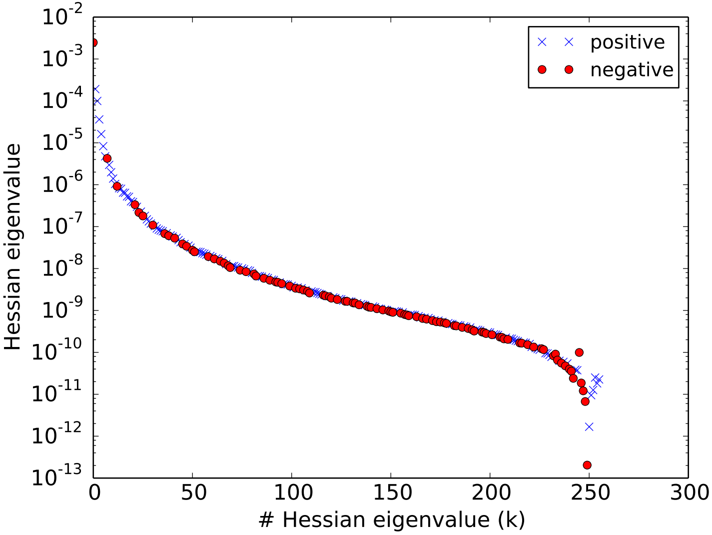
RB errors with different # Hessian modes for solution



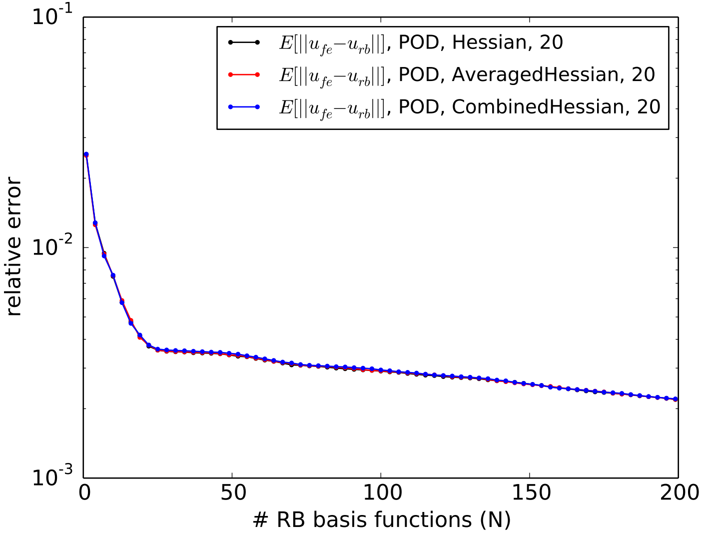
RB errors with different # Hessian modes for QoI

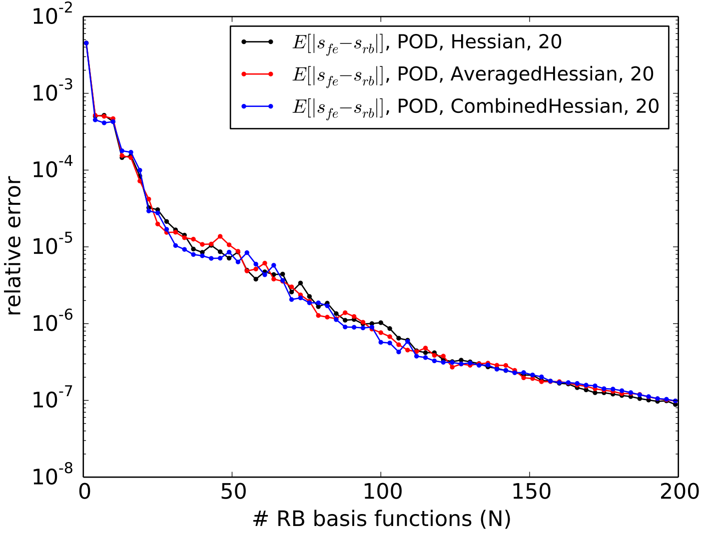


Decay of the eigenvalues of Hessian at mean



Comparison of different Hessians for solution




$E[|s_{fe}-s_{rb}|]$, POD, Hessian, 20
$E[|s_{fe}-s_{rb}|]$, POD, AveragedHessian, 20
$E[|s_{fe}-s_{rb}|]$, POD, CombinedHessian, 20
relative error
10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{-6}
10^{-7}
10^{-8}
0
50
100
150
200
RB basis functions (N)

Subsurface flow in porous medium

- Let V denote a Hilbert space with dual V'. Given $\boldsymbol{p} \in P$, μ-a.e., find $u \in V$ such that

$$a(u, v; \boldsymbol{p}) = f(v) \quad \forall v \in V.$$

- $a(\cdot, \cdot; \boldsymbol{p}) : V \times V \to \mathbb{R}$ is a bilinear form, e.g.,

$$a(u, v; \boldsymbol{p}) = \int_D \kappa(\boldsymbol{p}) \nabla u \cdot \nabla v dx.$$

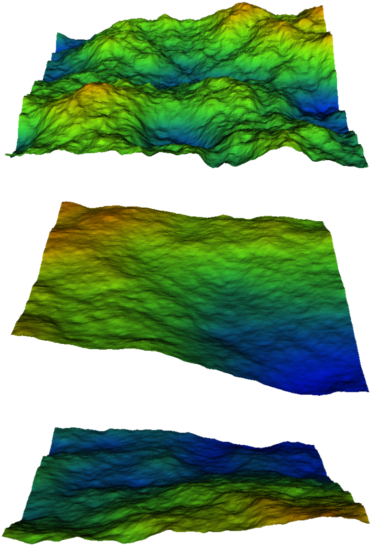
- $f = 0$, $u = 1$ on top, $u = 0$ on bottom.
- The QoI $s(u(\boldsymbol{p})) = 100 \int_{[0,0.1]^2} u(\boldsymbol{p}) dx$.
- $N_t = 1000$, uniform mesh of 129×129.

Ex 2. subsurface flow in a porous medium

$$\kappa(\boldsymbol{p}) = e^{\boldsymbol{p}}$$

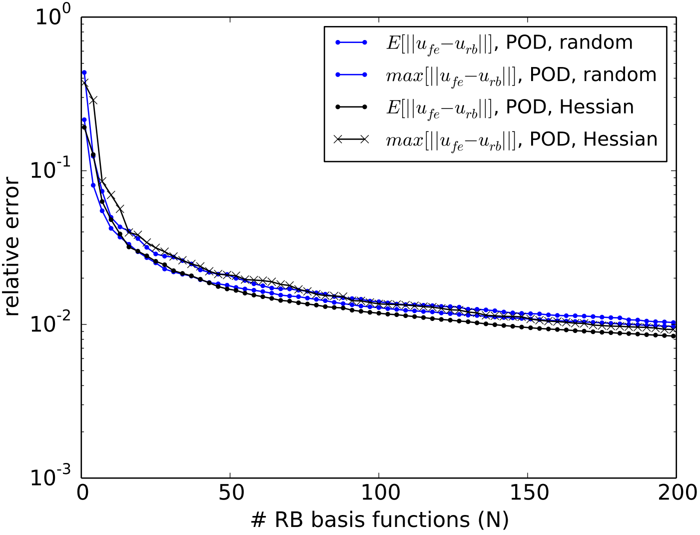
log-normal diffusion with

$$\boldsymbol{p} \in \mathcal{N}(\bar{\boldsymbol{p}}, \mathbb{C}), \quad \mathcal{C} = (-\Delta + 0.5I)^{-2}$$

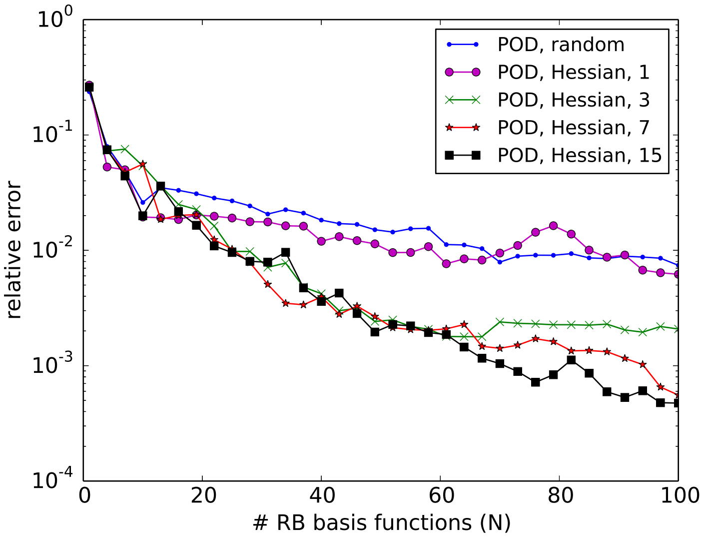


$K = 129^2 = 16,641$

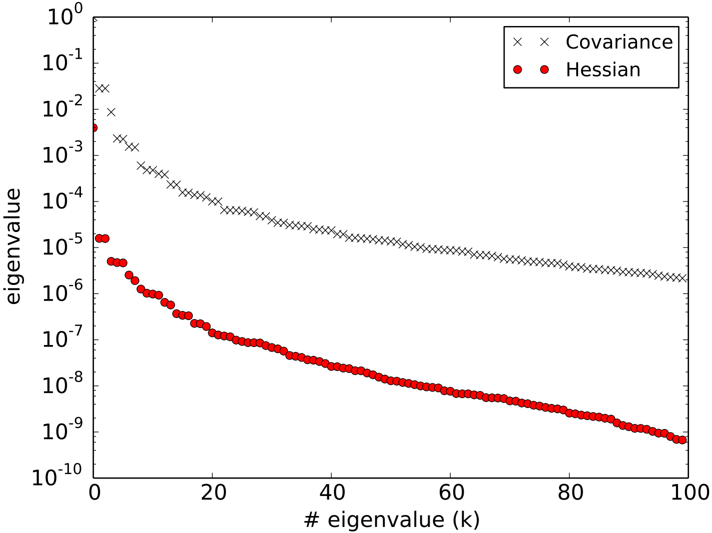
RB errors random vs Hessian for solution



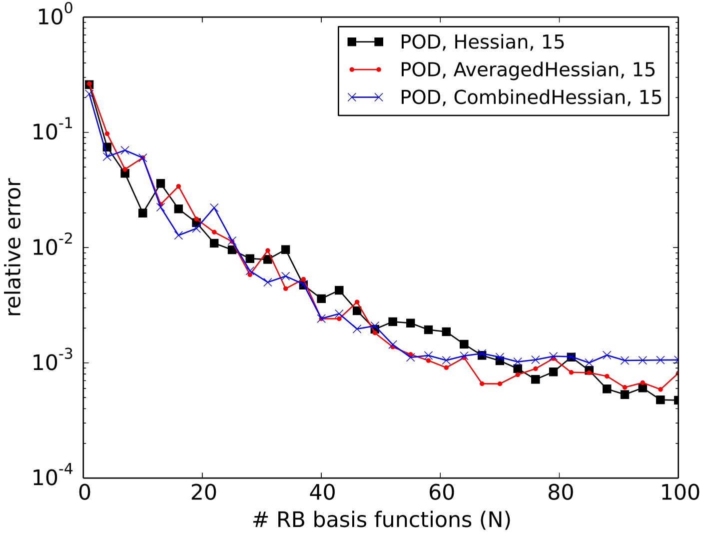
RB errors random vs Hessian for QoI



Decay of eigenvalues of Hessian $\mathbb{H}_{\bar{p}}$ and covariance $\mathbb{C}$



Comparison of different Hessians for QoI



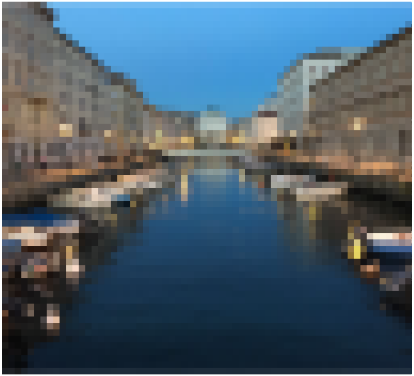
Summary

- Even the solution manifold is high-dimensional, the manifold of the QoI may be **low-dimensional**, which can be detected by the **eigenvalues of the Hessian**.
- A **scalable Hessian-based sampling algorithm** is developed, whose cost is independent of the nominal dimensions but only **intrinsic dimensions** for QoI.
- Further investigation on **adaptive Hessian sampling**, **local–global sampling**, **empirical interpolation**, **nonlinear problems**, **properties of different QoI**.
- **Rigorous analysis** of goal-oriented error estimate for Hessian-based sampling.

P. Chen, and O. Ghattas. Hessian-based sampling in high-dimensional parameter space for goal-oriented model order reduction, preprint, 2017.

P. Chen, U. Villa, and O. Ghattas. Hessian-based adaptive sparse quadrature for infinite-dimensional Bayesian inverse problems, preprint, 2017.

P. Chen, U. Villa, and O. Ghattas. Taylor approximation and variance reduction for PDE-constrained optimal control problems under uncertainty, preprint, 2017.



pixels $K = 64^2$

modes $K = 64$