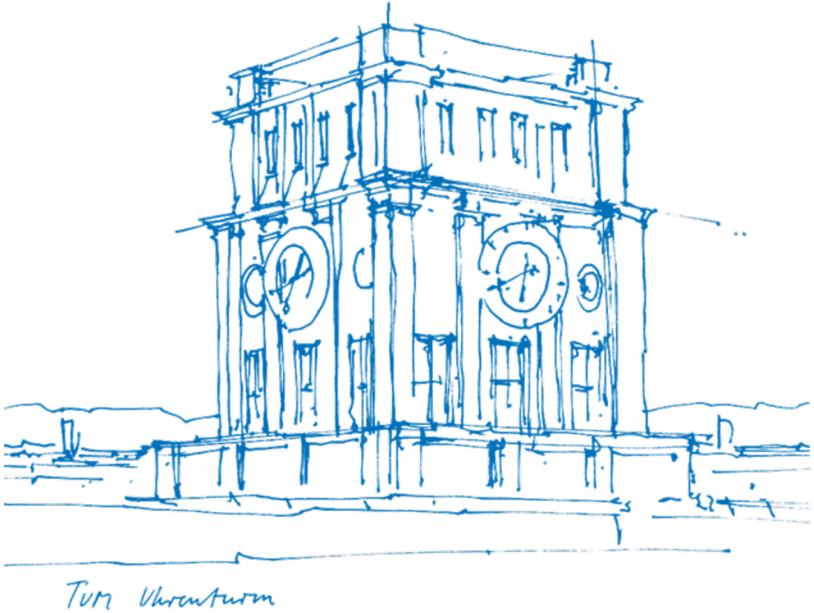
Lehrstuhl für Numerische Mathematik
Fakultät für Mathematik
Technische Universität München

Active Subspaces in Bayesian Inverse Problems

Mario Teixeira Parente

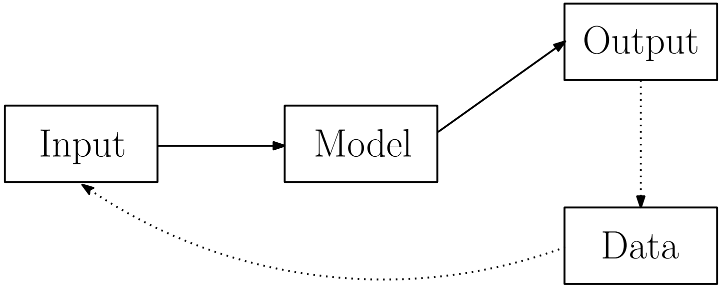
Thesis Defense

Garching, September 15, 2020



Motivation

Statistical inference of model inputs



Computational challenges

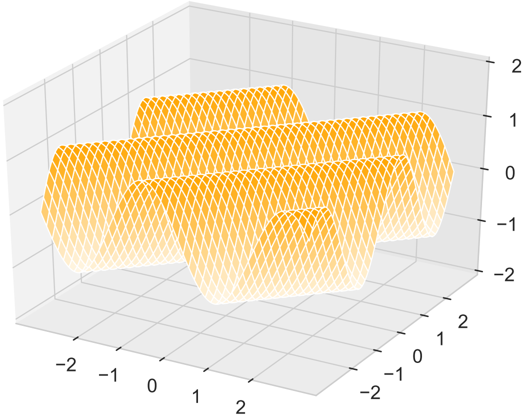
- Performance of a single model run $\Rightarrow$ efficient software, HPC, model reduction, ...
- Dimension of the input space ("curse of dimensionality") $\Rightarrow$ dimension reduction.

The **Active Subspace Method** is a gradient-based technique for subspace-based dimension reduction.

Outline

Outline

ASM – Motivation



Ridge functions $f(\mathbf{x}) = g(A^\top \mathbf{x})$, for $g : \mathbf{R}^k \to \mathbf{R}$, $k \leq n$, and $A \in \mathbf{R}^{n\times k}$, are constant along the null space of $A^\top$ since, for $\mathbf{v} \in \ker(A^\top)$,

$$f(\mathbf{x}+\mathbf{v}) = g(A^\top(\mathbf{x}+\mathbf{v})) = g(A^\top \mathbf{x}) = f(\mathbf{x}). \tag{1}$$

In the figure above, $f(\mathbf{x}) = \sin(-2x_1 + 2x_2)$.

ASM – Setup I (Constantine, Dow, and Wang, 2014)

- $(\Omega, \mathscr{A}, \mathbf{P})$ probability space.
- $\mathbf{X} \sim \mathbf{P}_\mathbf{X}$, $\mathbf{X} \in \mathbf{R}^n$, random vector (inputs, parameters).
- $\mathscr{X} := \operatorname{supp}(\mathbf{P}_\mathbf{X}) \subseteq \mathbf{R}^n$ continuity set (i. e., $\mathbf{P}_\mathbf{X}(\partial\mathscr{X}) = 0$). In particular, $\mathbf{P}_\mathbf{X}(\mathscr{X}) = 1$.
- $f : \mathscr{X} \to \mathbf{R}$ such that $\nabla f \in L^2(\mathscr{X}, \mathbf{P}_\mathbf{X})$.

Goal: Approximate f by a ridge function, i. e., find g and $A \in \mathbf{R}^{n\times k}$, $k \leq n$, such that

$$f(\mathbf{x}) \approx g(A^\top \mathbf{x}) \tag{2}$$

for each $\mathbf{x} \in \mathscr{X}$.

ASM – Setup II

Define

$$\begin{aligned} C &:= \mathbf{E}[\nabla f(\mathbf{X}) \nabla f(\mathbf{X})^\top] \\ &= \int_{\mathscr{X}} \nabla f(\mathbf{x}) \nabla f(\mathbf{x})^\top \rho_{\mathbf{X}}(\mathbf{x}) \, \mathrm{d}\mathbf{x} \\ &=: W \Lambda W^\top. \end{aligned} \tag{3}$$

Note that C is symmetric and positive semi-definite, i. e., we can choose

$$W = \begin{pmatrix} | & & | \\ \mathbf{w}_1 & \cdots & \mathbf{w}_n \\ | & & | \end{pmatrix} \in \mathbf{R}^{n \times n} \text{ to be orthogonal and } \Lambda = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} \in \mathbf{R}^{n \times n} \tag{4}$$

with

$$\lambda_1 \geq \cdots \geq \lambda_n \geq 0. \tag{5}$$

ASM – Setup III

Important:

$$\lambda_i = \mathbf{w}_i^\top C \mathbf{w}_i = \mathbf{E}[(\mathbf{w}_i^\top \nabla f(\mathbf{X}))^2]. \tag{6}$$

Question: What does λ_i small (or even zero) and λ_i large mean?

We hope that the eigenvalues λ_i are decaying quickly on a logarithmic scale. If so, for a certain $k \leq n$, split

$$W =: \begin{pmatrix} W_1 & W_2 \end{pmatrix} \tag{7}$$

for $W_1 \in \mathbf{R}^{n \times k}$ and $W_2 \in \mathbf{R}^{n \times (n-k)}$. The column space of W_1 (resp. W_2) is called the active (resp. inactive) subspace of f.

Define (random) variables for the subspaces by an orthogonal transformation, i. e.,

$$\begin{pmatrix} \mathbf{y} \\ \mathbf{z} \end{pmatrix} := W^\top \mathbf{x} = \begin{pmatrix} W_1^\top \mathbf{x} \\ W_2^\top \mathbf{x} \end{pmatrix}. \tag{8}$$

The variable $\mathbf{y} \in \mathbf{R}^k$ (resp. $\mathbf{z} \in \mathbf{R}^{n-k}$) is called the active (resp. inactive) variable.

ASM – Setup IV

Recall: The goal was to find g and $A \in \mathbf{R}^{n \times k}$ such that $f(\mathbf{x}) \approx g(A^\top \mathbf{x})$ for each $\mathbf{x} \in \mathscr{X}$.

The map $\mathbf{x} \mapsto g(A^\top \mathbf{x})$ was said to be constant along the null space of $A^\top$.

$\Rightarrow$ We found $A = W_1$ since $\ker(W_1^\top) = \operatorname{ran}(W_2)$ (by $\ker(W_1^\top) \perp \operatorname{ran}(W_1)$ and orthogonality of W).

Next step: Find "best" function g.

It is well-known that, if $\mathbf{E}[f(\mathbf{X})^2] < \infty$, the conditional expectation of $f(\mathbf{X})$ given $\mathbf{Y} = W_1^\top \mathbf{X}$ minimizes the mean square error to f, i. e.,

$$\mathbf{E}[(f(\mathbf{X}) - \mathbf{E}[f(\mathbf{X}) \mid \mathbf{Y}])^2] \leq \mathbf{E}[(f(\mathbf{X}) - \mathbf{U})^2] \tag{9}$$

for any square-integrable random variable $\mathbf{U}$ which is measurable w.r.t. the σ-algebra generated by $\mathbf{Y}$.

ASM – Setup V

Notation: We denote

$$\mathbf{x} = WW^\top \mathbf{x} = W_1\mathbf{y} + W_2\mathbf{z} =: [\![\mathbf{y},\mathbf{z}]\!]_W = [\![\mathbf{y},\mathbf{z}]\!]. \tag{10}$$

Hence, we define

$$\begin{aligned} g(\mathbf{y}) &:= \mathbf{E}[f([\![\mathbf{Y},\mathbf{Z}]\!]) \,|\, \mathbf{Y} = \mathbf{y}] \\ &= \int_{\mathbf{R}^{n-k}} f([\![\mathbf{y},\mathbf{z}]\!])\, \rho_{\mathbf{Z}|\mathbf{Y}}(\mathbf{z}|\mathbf{y})\, \mathrm{d}\mathbf{z} \end{aligned} \tag{11}$$

and finally get the approximating ridge function

$$f_g(\mathbf{x}) := g(W_1^\top \mathbf{x}). \tag{12}$$

Question: How "good" is f_g in approximating f in terms of the neglected directions in W_2?

ASM – Common bounds

In Constantine et al. (2014), we have the following result.

Theorem

For each $\mathbf{P_X}$, there exists a Poincaré constant $C_P = C_P(\mathbf{P_X}) > 0$ such that

$$\mathbf{E}[(f(\mathbf{X}) - f_g(\mathbf{X}))^2] \leq C_P(\lambda_{k+1} + \cdots + \lambda_n). \tag{13}$$

Proof

Main ingredient: Probabilistic Poincaré inequality.
[For a random variable $\mathbf{U} \sim \mathbf{P_U}$ and a sufficiently regular function h, it holds that

$$\mathrm{Var}(h(\mathbf{U})) \leq C_\mathrm{P}\, \mathbf{E}[\|\nabla h(\mathbf{U})\|_2^2] \tag{14}$$

for some Poincaré constant $C_\mathrm{P} = C_\mathrm{P}(\mathbf{P_U}) > 0$.]
The main step is to compute

$$\begin{aligned} \mathbf{E}[(f(\mathbf{X}) - f_g(\mathbf{X}))^2] &= \mathbf{E}[(f([\![\mathbf{Y},\mathbf{Z}]\!]) - g(\mathbf{Y}))^2] \\ &\leq C_\mathrm{P}\, \mathbf{E}[\|\nabla^{\mathbf{z}} f(\mathbf{X})\|_2^2]. \end{aligned} \tag{15}$$

Note: The Poincaré constant C_P was taken w.r.t. $\mathbf{P_X}$ which is not correct in general. ∎

ASM – Generalized bounds I (Teixeira Parente, Wallin, and Wohlmuth, 2020)

A correct application of the probabilistic Poincaré inequality gives

$$\begin{aligned}\mathbf{E}[(f(\mathbf{X}) - f_g(\mathbf{X}))^2] &= \mathbf{E}[\mathbf{E}[(f([\![\mathbf{Y},\mathbf{Z}]\!]) - g(\mathbf{Y}))^2 \,|\, \mathbf{Y}]] \\ &\leq \mathbf{E}[C_{\mathbf{Y}} \cdot \mathbf{E}[\|\nabla^{\mathbf{z}} f([\![\mathbf{Y},\mathbf{Z}]\!])\|_2^2 \,|\, \mathbf{Y}]],\end{aligned} \tag{16}$$

where $C_{\mathbf{Y}} > 0$ is the Poincaré constant of $\mathbf{P}_{\mathbf{Z}|\mathbf{Y}}$ and hence randomly depending on $\mathbf{Y}$.

The constant $C_{\mathbf{Y}}$ is known to be uniform in $\mathbf{Y}$ for, e. g., compactly supported and so-called α-uniformly log-concave ($\alpha > 0$) distributions $\mathbf{P}_{\mathbf{X}}$.

Question: Can we find a counterexample, i. e., a distribution $\mathbf{P}_{\mathbf{X}}$ and a transformation W such that the corresponding $C_{\mathbf{Y}}$ is unbounded in $\mathbf{Y}$/not compactly supported?

ASM – Generalized bounds II

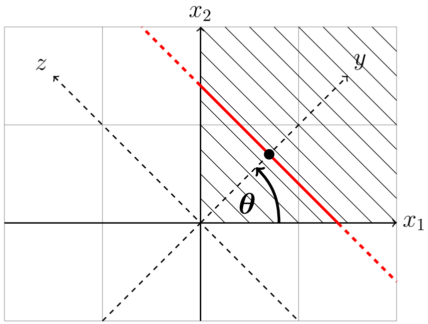
Answer: Yes.

Idea: We need to find a distribution that has heavier tails than any α-uniformly log-concave distribution but is still applicable for a Poincaré inequality. Look at the edge case when $\alpha \to 0$. $\Rightarrow$ General log-concave distributions.

Example: Exponential distribution with unit rates in two dimensions; rotation by 45°.

Use lower bound for $C_{\mathbf{Y}}$ of Bobkov (1999):

$$C_{\mathbf{y}} \geq \mathrm{Var}(|\mathbf{Z}| \mid \mathbf{Y} = \mathbf{y}) = \mathbf{y}^2/12. \tag{17}$$



ASM – Generalized bounds III

Question: Can we derive weaker bounds for general log-concave distributions?

Answer: Yes. Use Hölder's inequality with a weaker pair of conjugates to get

$$\mathbf{E}[(f(\mathbf{X})-f_g(\mathbf{X}))^2] \leq C_{\mathrm{P},\varepsilon,W}(\lambda_{k+1}+\cdots+\lambda_n)^{1/(1+\varepsilon)} \tag{18}$$

for $\varepsilon > 0$ and a constant $C_{\mathrm{P},\varepsilon,W} > 0$.

ASM – Practical considerations (Constantine, 2015)

In practice, we approximate the matrix $C = \mathbf{E}[\nabla f(\mathbf{X}) \nabla f(\mathbf{X})^\top]$ by a finite Monte Carlo sum, i. e., by

$$\tilde{C} := \frac{1}{N_{\tilde{C}}} \sum_{j=1}^{N_{\tilde{C}}} \nabla f(\mathbf{X}_j) \nabla f(\mathbf{X}_j)^\top \tag{19}$$

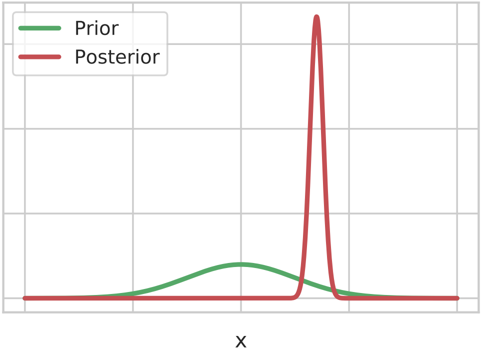
for $N_{\tilde{C}} > 0$ and $\mathbf{X}_j \overset{\text{i.i.d.}}{\sim} \mathbf{P}_{\mathbf{X}}$, $j = 1, \dots, N_{\tilde{C}}$.

It is well-known by an eigenvalue Bernstein inequality that $N_{\tilde{C}} = O(\log n)$ samples are enough to approximate "the first eigenvalues of C sufficiently accurate."

Outline

BIPs – Motivation

Bayesian inverse problems exploit observational data to update a prior distribution on model parameters to a posterior distribution.



In view of Uncertainty Quantification, the posterior distribution quantifies the updated, remaining uncertainty on model parameters.

BIPs – Setup I (Stuart, 2010; Dashti and Stuart, 2016)

Parameters are denoted by $\mathbf{X} \sim \mathbf{P}_{\mathbf{X}}$ or $\mathbf{x}$, both in $\mathbf{R}^n$.

The forward model $\mathscr{G} : \mathscr{X} \to \mathbf{R}^{n_\mathbf{d}}$ maps a parameter to a corresponding value for the Quantity of Interest (QoI).

The observational data $\mathbf{d} \in \mathbf{R}^{n_\mathbf{d}}$ are assumed to be noisy realizations of a model evaluation, i. e.,

$$\mathbf{D} = \mathscr{G}(\mathbf{X}) + \eta, \tag{20}$$

where noise $\eta \sim \mathscr{N}(0, \Gamma)$ is assumed to be Gaussian with mean zero and covariance matrix $\Gamma \in \mathbf{R}^{n_\mathbf{d} \times n_\mathbf{d}}$.

BIPs – Setup II

For observed data $\mathbf{d} \in \mathbf{R}^{n_\mathbf{d}}$, the posterior distribution $\mu^\mathbf{d}$ is the conditional distribution of parameters $\mathbf{X}$ given that $\mathbf{D} = \mathbf{d}$, i. e.,

$$\mu^\mathbf{d} := \mathbf{P}_{\mathbf{X}|\mathbf{D}}(\cdot|\mathbf{d}) = \mathbf{P}(\mathbf{X} \in \cdot \,|\, \mathbf{D} = \mathbf{d}). \tag{21}$$

The corresponding posterior density $\rho^\mathbf{d}$ is proportional to the likelihood and the prior density $\rho_0 := \rho_\mathbf{X}$, i. e.,

$$\rho^\mathbf{d}(\mathbf{x}) := \rho_{\mathbf{X}|\mathbf{D}}(\mathbf{x}|\mathbf{d}) \propto \exp(-f^\mathbf{d}(\mathbf{x})) \cdot \rho_0(\mathbf{x}) \tag{22}$$

for the data misfit function (or negative log-likelihood)

$$f^\mathbf{d}(\mathbf{x}) = \frac{1}{2}\|\mathbf{d} - \mathscr{G}(\mathbf{x})\|_\Gamma^2, \tag{23}$$

where $\|\cdot\|_\Gamma := \|\Gamma^{-1/2}\cdot\|_2$.

BIPs – Application of ASM (Constantine, Kent, and Bui-Thanh, 2016)

In the setting of ASM, we set

$$f = f^{\mathbf{d}} \tag{24}$$

and thus compute the active subspace of the data misfit function to obtain a ridge approximation

$$f^{\mathbf{d}}(\mathbf{x}) \approx g^{\mathbf{d}}(W_1^\top \mathbf{x}). \tag{25}$$

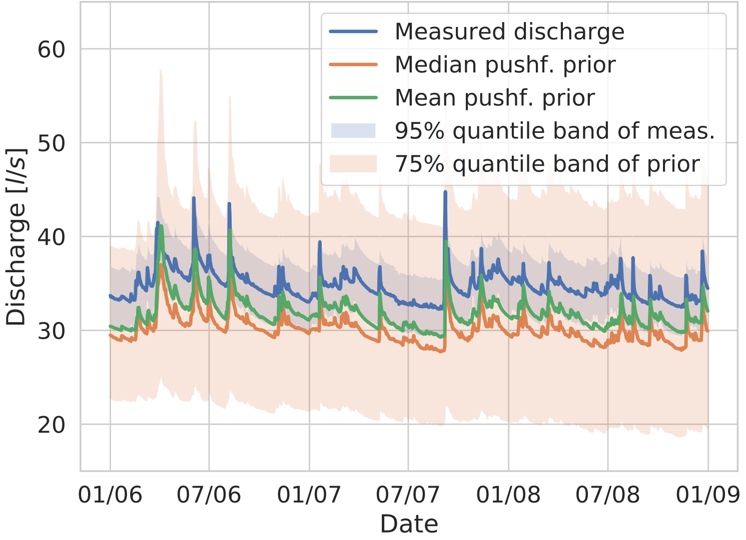
Consequently, we can find a low-dimensional approximation of the posterior density, i. e., for $\mathbf{x} = [\![\mathbf{y}, \mathbf{z}]\!]$,

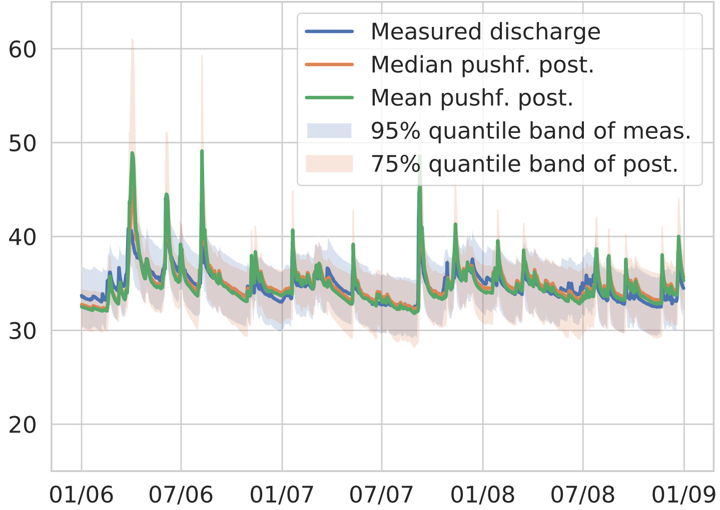
$$\begin{aligned} \rho^{\mathbf{d}}(\mathbf{x}) &\propto \exp(-f^{\mathbf{d}}(\mathbf{x})) \cdot \rho_0(\mathbf{x}) \\ &\approx \exp(-g^{\mathbf{d}}(W_1^\top \mathbf{x})) \cdot \rho_0(\mathbf{x}) \\ &= \underbrace{\exp(-g^{\mathbf{d}}(\mathbf{y})) \cdot \rho_{\mathbf{Y}}(\mathbf{y})}_{\propto: \rho^{\mathbf{d}}_{g,\mathbf{Y}}(\mathbf{y})} \cdot \rho_{\mathbf{Z}|\mathbf{Y}}(\mathbf{z}|\mathbf{y}). \end{aligned} \tag{26}$$

Outline

Case study (Teixeira Parente, Bittner, Mattis, Chiogna, and Wohlmuth, 2019)

Groundwater karst model for a discharge time series of the Kerschbaum spring recharge area in Waidhofen a.d. Ybbs (Austria).





Outline

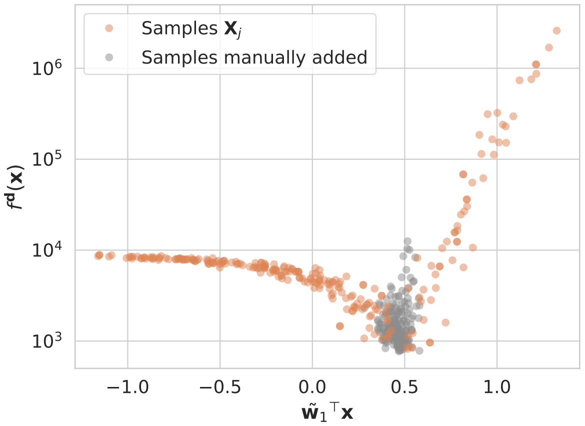
Iterative ASM – Idea I (Section 7 in dissertation)

Problem: Since sensitivities are computed w.r.t. the prior, i. e.,

$$C_0 = \int_{\mathcal{X}} \nabla f^{\mathbf{d}}(\mathbf{x}) \nabla f^{\mathbf{d}}(\mathbf{x})^\top \rho_0(\mathbf{x})\, \mathrm{d}\mathbf{x}, \tag{27}$$

a "bad prior" may cause a misleading active subspace.

Example:

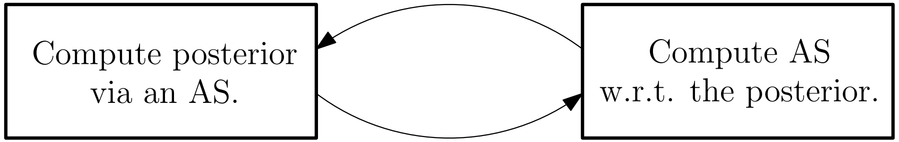


Iterative ASM – Idea II

Actual goal: Compute sensitivities w.r.t. the posterior, i. e.,

$$C^{\mathbf{d}} := \int_{\mathcal{X}} \nabla f^{\mathbf{d}}(\mathbf{x}) \nabla f^{\mathbf{d}}(\mathbf{x})^{\top} \rho^{\mathbf{d}}(\mathbf{x}) \, d\mathbf{x}. \tag{28}$$

Problem:



$\Rightarrow$ Iterative scheme.

Iterative ASM – Idea III

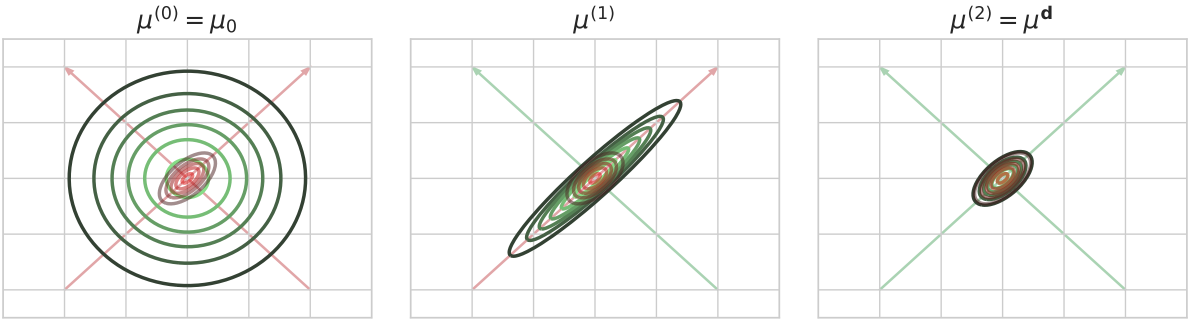
Compute a sequence of distributions $\mu^{(\ell)}$ approaching $\mu^{\mathbf{d}}$ starting from the prior $\mu^{(0)} = \mu_0$.

Ideal algorithm: Set $\mu^{(0)} := \mu_0$. In the ℓ-th step, $\mathbf{X}^{(\ell)} \sim \mu^{(\ell)}$ and the main steps are:

- Compute
$$C^{(\ell)} := \mathbf{E}[\nabla f^{\mathbf{d}}(\mathbf{X}^{(\ell)}) \nabla f^{\mathbf{d}}(\mathbf{X}^{(\ell)})^\top] = W^{(\ell)} \Lambda^{(\ell)} W^{(\ell)\top}. \tag{29}$$
- Decide for an active subspace, i. e., for $W_1^{(\ell)}$, and compute
$$g^{(\ell)}(\mathbf{y}) := \mathbf{E}[f^{\mathbf{d}}([\![\mathbf{Y}^{(\ell)}, \mathbf{Z}^{(\ell)}]\!]_{W^{(\ell)}}) \mid \mathbf{Y}^{(\ell)} = \mathbf{y}]. \tag{30}$$
- Compute samples $\mathbf{X}^{(\ell+1)} \sim \mu^{(\ell+1)}$ using $g^{(\ell)}$.

Iterative ASM – Idea IV

Illustrative example:

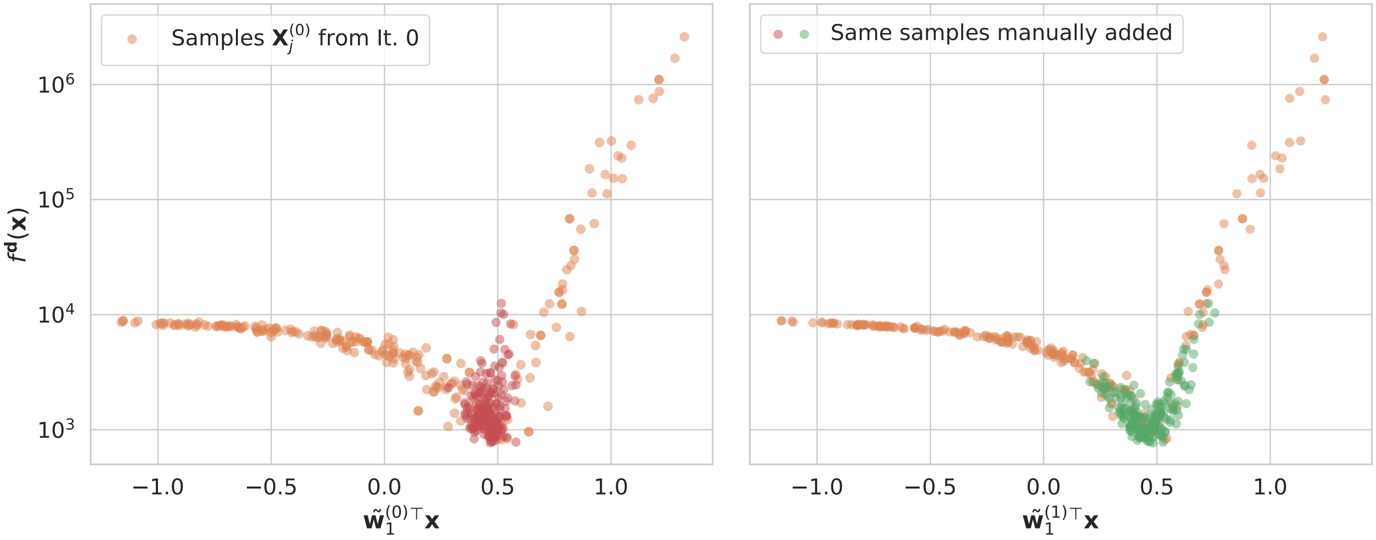


Remark I: Convergence and consistency of the ideal algorithm are formally well-understood for a linear Gaussian BIP.

Remark II: In practice, we do not aim to converge to the posterior since the exact quantities used in the algorithm are not available. But we can use the algorithm with approximate quantities as a "preconditioner" for BIPs with bad prior distributions.

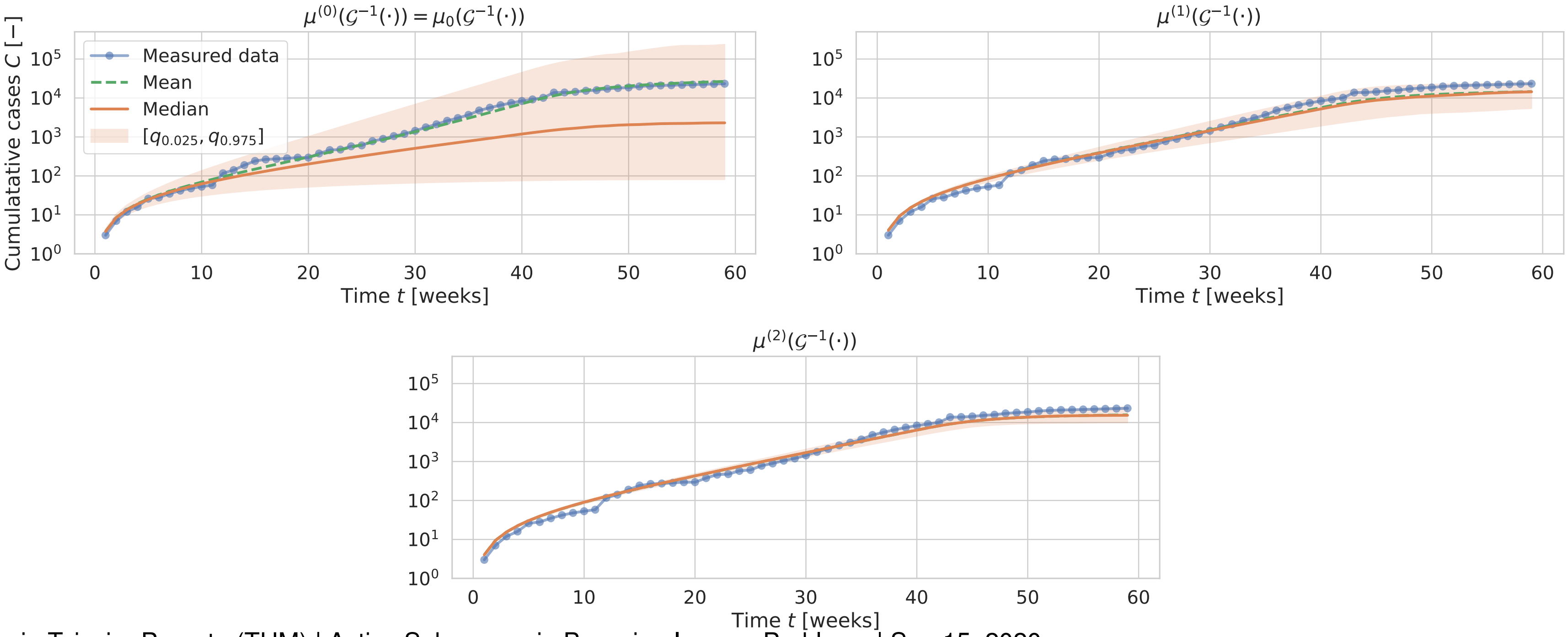
Iterative ASM – Case study I

Compartmental model for 2014 Ebola outbreak in West Africa (Barbarossa et al., 2015).



Iterative ASM – Case study II

The following figures show the evolution of the push-forward distribution of $\mu^{(\ell)}$, $\ell = 0, 1, 2$.



Outline

Summary

We presented a counterexample to existing Active Subspace theory and provided a solution by generalized bounds.

A case study involving a complex high-dimensional hydrological model demonstrated that ASM can substantially reduce the dimension and computational expenses of Bayesian inverse problems.

For BIPs with "bad priors", we suggested an iterative scheme to find a better active subspace by computing sensitivities in regions of low data misfit. This was demonstrated on a model for the 2014 Ebola outbreak in West Africa.

Thank you!

I would like to **sincerely thank**

- the **audience** for their attention,
- my **supervisor** Prof. Barbara Wohlmuth for stimulating discussions during the last 4 years,
- my **collaborators** (sorted by surname):

  GEOMAR Helmholtz Centre for Ocean Research Kiel: Christian Deusner, Shubhangi Gupta

  Lund University, Department of Statistics: Prof. Krzysztof Podgórski, Jonas Wallin

  TUM Chair of Hydrology and River Basin Management: Daniel Bittner, Prof. Gabriele Chiogna
- the **TUM M2 unit** for their support,
- the **TUM International Graduate School for Science and Engineering** (IGSSE) for financial support, GSC 81,
- and the **chair of the examination board** Prof. Christina Kuttler for organizing my defense.

Publications

Teixeira Parente, M., Wallin, J., & Wohlmuth, B., Generalized Bounds for Active Subspaces, *Electronic Journal of Statistics*, 14(1):917–943, 2020

Bittner, D., **Teixeira Parente, M.**, Mattis, S., Wohlmuth, B., & Chiogna, G., Identifying Relevant Hydrological and Catchment Properties in Active Subspaces: An Inference Study of a Lumped Karst Aquifer Model. *Advances in Water Resources*, 135, 103472, 2020.

Teixeira Parente, M., Bittner, D., Mattis, S., Chiogna, G., & Wohlmuth, B., Bayesian Calibration and Sensitivity Analysis for a Karst Aquifer Model using Active Subspaces, *Water Resources Research*, 55(8):7086–7107, 2019

Teixeira Parente, M., Mattis, S., Gupta, S., Deusner, C., & Wohlmuth, B., Efficient Parameter Estimation for a Methane Hydrate Model with Active Subspaces, *Computational Geosciences*, 23(2):355–372, 2019

References

M. V. Barbarossa, A. Dénes, G. Kiss, Y. Nakata, G. Röst, and Z. Vizi. Transmission Dynamics and Final Epidemic Size of Ebola Virus Disease Outbreaks with Varying Interventions. *PLoS ONE*, 10(7), 2015.

S. G. Bobkov. Isoperimetric and Analytic Inequalities for Log-Concave Probability Measures. *The Annals of Probability*, 27(4):1903–1921, 1999.

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

P. G. Constantine, E. Dow, and Q. Wang. Active Subspace Methods in Theory and Practice: Applications to Kriging Surfaces. *SIAM Journal on Scientific Computing*, 36(4):A1500–A1524, 2014.

P. G. Constantine, C. Kent, and T. Bui-Thanh. Accelerating Markov Chain Monte Carlo with Active Subspaces. *SIAM Journal on Scientific Computing*, 38 (5):A2779–A2805, 2016.

M. Dashti and A. M. Stuart. The Bayesian Approach to Inverse Problems. *Handbook of Uncertainty Quantification*, pages 1–118, 2016.

A. M. Stuart. Inverse Problems: A Bayesian Perspective. *Acta Numerica*, 19:451–559, 2010.

M. Teixeira Parente, J. Wallin, and B. Wohlmuth. Generalized Bounds for Active Subspaces. *Electronic Journal of Statistics*, 14(1):917–943, 2020.

M. Teixeira Parente, D. Bittner, S. A. Mattis, G. Chiogna, and B. Wohlmuth. Bayesian Calibration and Sensitivity Analysis for a Karst Aquifer Model using Active Subspaces. *Water Resources Research*, 55(8):7086–7107, 2019.

Supplementary Material

ASM – Generalized bounds IV

Theorem

Let $\varepsilon > 0$. If $\|\nabla f(\mathbf{X})\|_2^2 \leq L$ $\mathbf{P}$-a.s. for some constant $L > 0$, then

$$\mathbf{E}[(f(\mathbf{X}) - f_g(\mathbf{X}))^2] \leq C_{P,\varepsilon,W}\,(\lambda_{k+1} + \cdots + \lambda_n)^{1/(1+\varepsilon)}, \tag{31}$$

where

$$C_{P,\varepsilon,W} = C_{P,\varepsilon,W}(\varepsilon, n, k, L, W, \mathbf{P_X}) := L^{\varepsilon/(1+\varepsilon)}\mathbf{E}[C_{\mathbf{Y}}^{(1+\varepsilon)/\varepsilon}]^{\varepsilon/(1+\varepsilon)}. \tag{32}$$

Proof

Main ingredient: Hölder's inequality for $\mathbf{E}[C_{\mathbf{Y}} \cdot \mathbf{E}[\|\nabla^{\mathbf{z}} f([\![\mathbf{Y}, \mathbf{Z}]\!])\|_2^2 \,|\, \mathbf{Y}]]$ with a weaker pair of conjugates

$$(p, q) = ((1+\varepsilon)/\varepsilon, 1+\varepsilon) \tag{33}$$

instead of $(p, q) = (+\infty, 1)$. ■

Remark: For exponential distributions, the constant $C_{P,\varepsilon,W}$ can be bounded uniformly in W by an analytical expression.

ASM – Practical considerations II

Approximation of eigenvalues of C

Theorem (Constantine (2015, Thm. 3.3))

Assume that $\|\nabla f(\mathbf{x})\|_2 \leq L$ *for some* $L > 0$ *and all* $\mathbf{x} \in \mathscr{X}$. *For* $\varepsilon \in (0,1]$, *it holds that*

$$\mathbf{P}\left(\tilde{\lambda}_\ell \geq (1+\varepsilon)\lambda_\ell\right) \leq (n-\ell+1)\exp\left(-\frac{N_{\tilde{C}}\lambda_\ell\varepsilon^2}{4L^2}\right) \tag{34}$$

and

$$\mathbf{P}\left(\tilde{\lambda}_\ell \leq (1-\varepsilon)\lambda_\ell\right) \leq \ell\exp\left(-\frac{N_{\tilde{C}}\lambda_\ell^2\varepsilon^2}{4\lambda_1 L^2}\right) \tag{35}$$

for $\ell = 1, \ldots, n$.

Proof

Main ingredient: Eigenvalue Bernstein inequality for a finite sum of random, independent, and symmetric matrices satisfying a subexponential growth condition. ∎

ASM – Practical considerations III

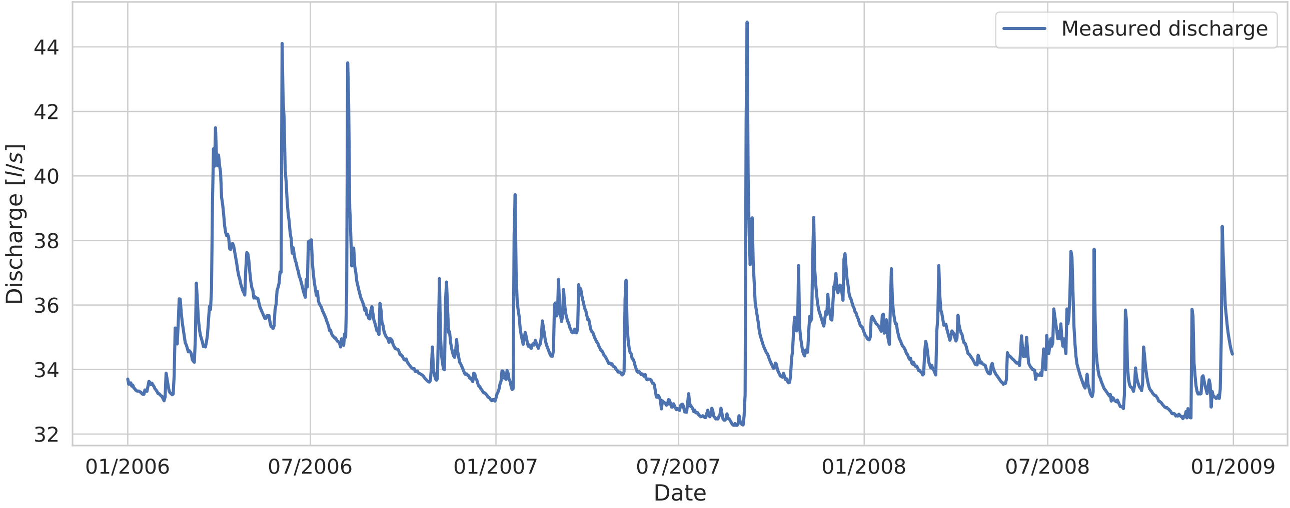
The function $g(\mathbf{y}) = \mathbf{E}[f([\![\mathbf{Y},\mathbf{Z}]\!])\,|\,\mathbf{Y}=\mathbf{y}]$ is also approximated by a finite Monte Carlo sum, i. e., by

$$g_N(\mathbf{y}) := \frac{1}{N}\sum_{j=1}^{N} f([\![\mathbf{y},\mathbf{Z}_j^{\mathbf{y}}]\!]) \tag{36}$$

for $N > 0$ and $\mathbf{Z}_j^{\mathbf{y}} \overset{\text{i.i.d.}}{\sim} \mathbf{P}_{\mathbf{Z}|\mathbf{Y}}(\cdot|\mathbf{y})$, $j = 1,\dots,N$.

Case study II

Measured discharge data



Case study III

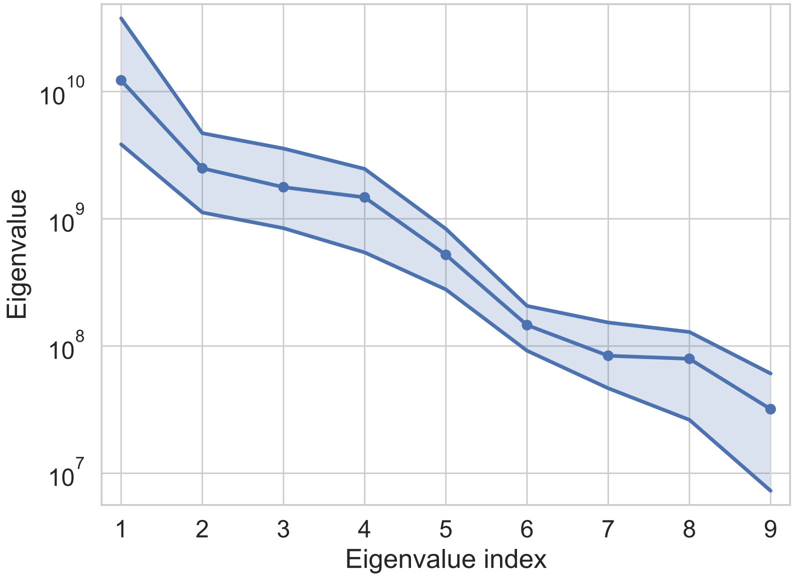
Problem characteristics:

- Parameter space is $n = 21$-dimensional.
- Data consists of $n_{\mathbf{d}} = 1096$ discharge values.
- We assume a 5% noise level on the measured data.
- The prior distribution is chosen to be uniform on intervals predefined by hydrologists.
- A single model run needs about 2.5 seconds.

Implementation: The computation of $N_{\tilde{C}} = 1000$ gradients needed about 4.3 hours using 7 CPU cores in parallel.

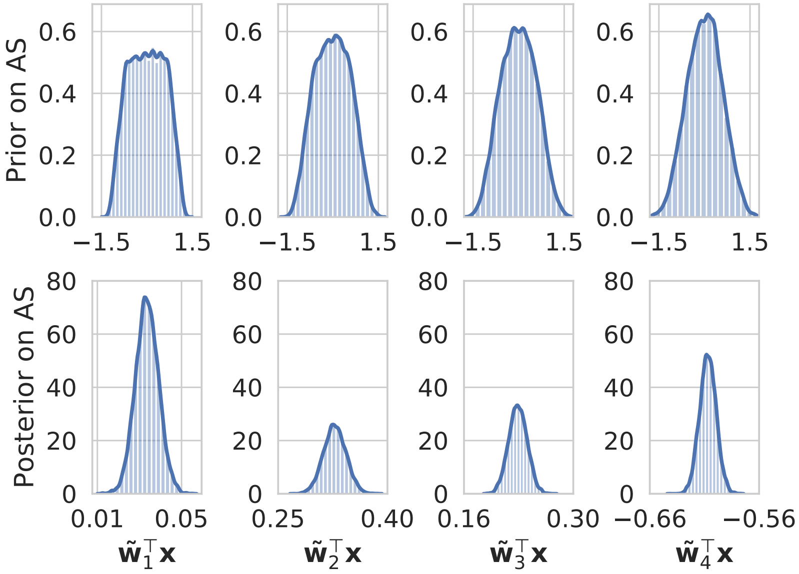
Case study IV

The following figure shows the approximated eigenvalues with an "uncertainty band" to reflect their random nature.



Case study V

The following figure shows the prior (top) and posterior (bottom) distribution on the active subspace.



Iterative ASM – Analysis

Proposition (Dissertation, Prop. 7.3.1)

Let $\mu_0 = \mathscr{N}(\mathbf{m}_0, I)$ *with* $\mathbf{m}_0 \in \mathbf{R}^n$*. Suppose that* $\mathscr{G}(\mathbf{x}) := A\mathbf{x}$ *for* $\mathbf{x} \in \mathscr{X}$ *with* $A \in \mathbf{R}^{n_\mathbf{d} \times n}$*, and* $\eta \sim \mathscr{N}\left(0, \gamma^2 I\right)$ *for* $\gamma > 0$*.*
Furthermore, assume that $\mathbf{d} = \mathscr{G}(\mathbf{m}_0)$*.*
Considering the ideal algorithm, set $\Lambda := \Lambda^{(0)}$ *and* $W := W^{(0)}$*. Then, for every iteration* $\ell \in \mathbf{N}_0$*, it holds that*

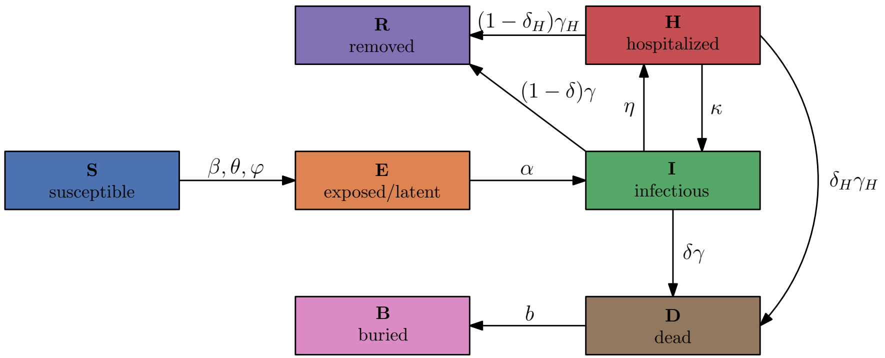
$$\mathbf{X}^{(\ell)} \sim \mathscr{N}\left(\mathbf{m}_0, \Sigma^{(\ell)}\right) \tag{37}$$

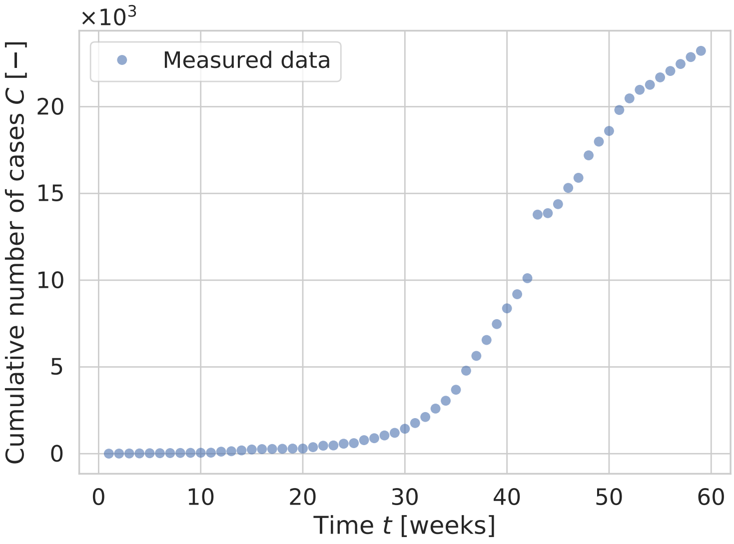
with

$$\Sigma^{(\ell)} = W \begin{pmatrix} \left(I + \Lambda^{1/2}_{1:K^{(\ell)}}\right)^{-1} & 0 \\ 0 & I \end{pmatrix} W^\top \tag{38}$$

for some natural number $0 \leq K^{(\ell)} \leq n$*.*

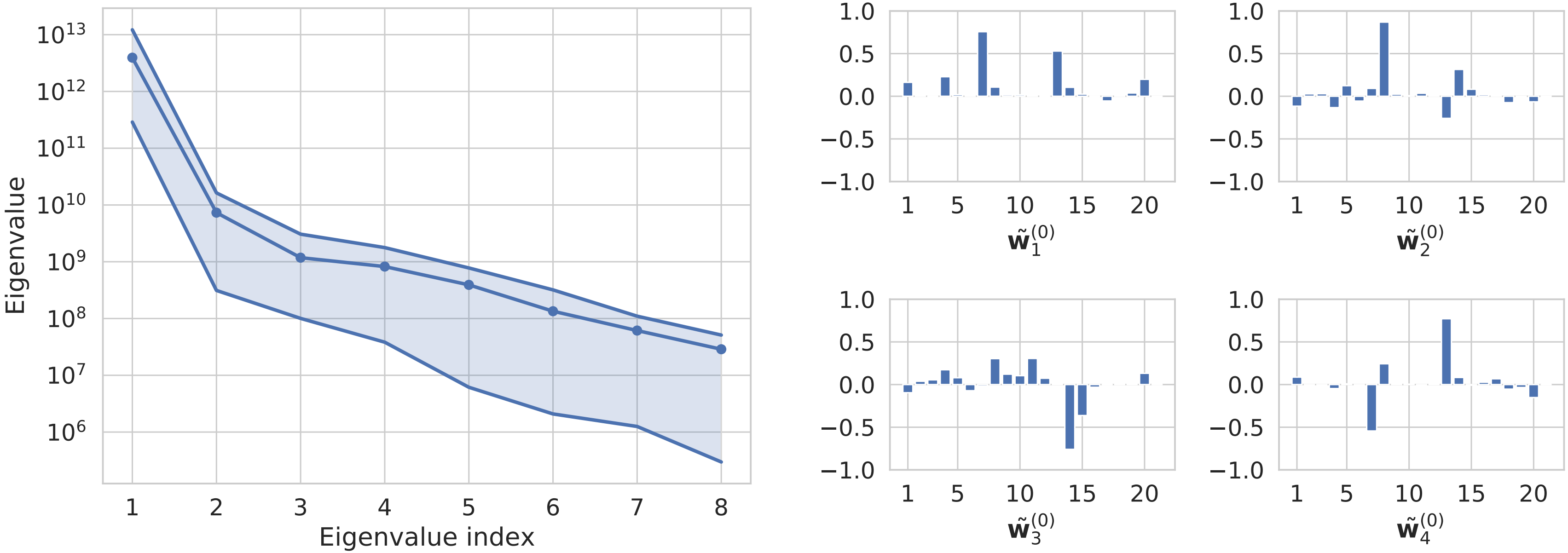
Iterative ASM – Case study, Fig. I





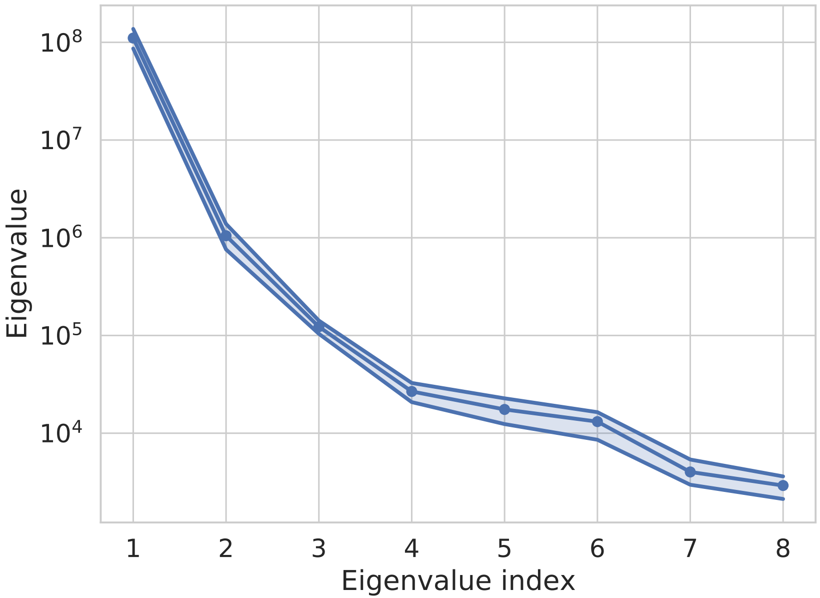
Iterative ASM – Case study, Fig. II

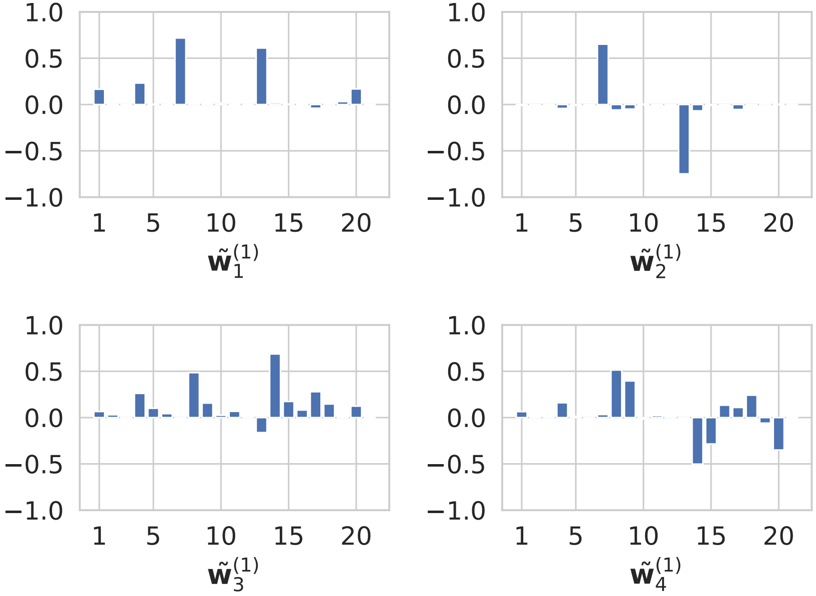
Eigendecomposition It. 0



Iterative ASM – Case study, Fig. III

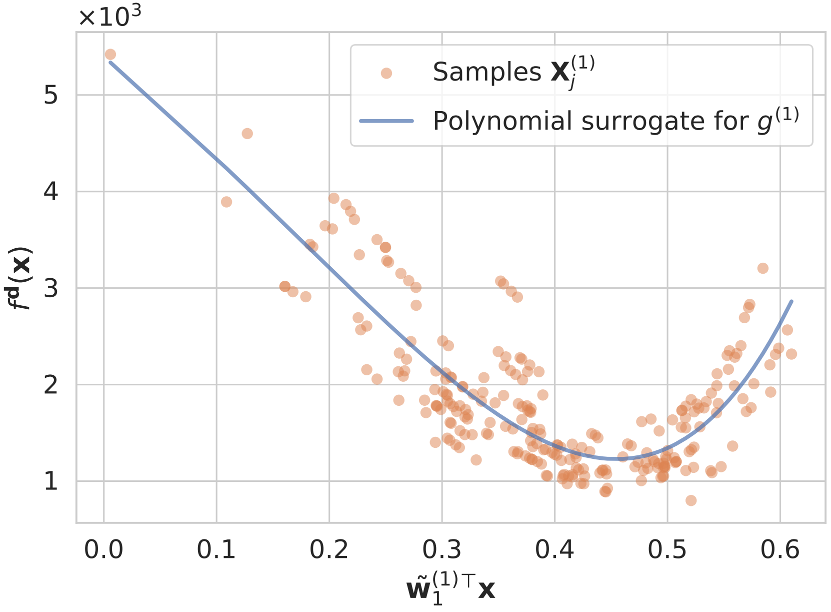
Eigendecomposition It. 1





Iterative ASM – Case study, Fig. IV

Surrogate for $g^{(1)}$



Iterative ASM – Case study, Fig. V

Subspace distances It. 0 $\leftrightarrow$ It. 1 $\leftrightarrow$ It. 2