



Technische Universität München
Ingenieur fakultät Bau Geo Umwelt
Professur für Risikoanalyse und Zuverlässigkeit

Bayesian analysis and rare event simulation of random fields

Felipe Uribe–Castillo

Vollständiger Abdruck der von der Ingenieur fakultät Bau Geo Umwelt der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktor-Ingenieurs (Dr.-Ing.)

genehmigten Dissertation.

Vorsitzender: Prof. Dr.-Ing. Habil. Fabian Duddeck

Prüfer der Dissertation:

1. Prof. Dr. Daniel Straub
2. Prof. Dr. Elisabeth Ullmann
3. Prof. Dr. Youssef M. Marzouk

Die Dissertation wurde am 09.03.2020 bei der Technischen Universität München eingereicht und durch die Ingenieur fakultät Bau Geo Umwelt am 04.12.2020 angenommen.

Titel in deutscher Sprache

Bayes'sche Inferenz und Zuverlässigkeitsanalyse von Zufallsfeldern.

Zusammenfassung

Numerische Modelle in Technik und Wissenschaft werden durch Parameter gesteuert, deren Werte in der Praxis oft nicht genau sind. Die damit verbundene Unsicherheit kann reduziert werden, indem das Modell mit Messungen der Systemantwort kombiniert wird. Ein solches inverses Problem kann durch Bayes'sche Inferenz gelöst werden, wobei Vorabinformationen mit Beobachtungen verknüpft werden. Das daraus resultierende probabilistische Modell kann verwendet werden, um Vorhersagen in Fällen zu treffen, in denen die Zuverlässigkeit des Systems das Hauptanliegen ist. Die Zuverlässigkeit kann anhand der Wahrscheinlichkeit von Versagensereignissen quantifiziert werden, für deren Schätzung spezielle Simulationsmethoden verwendet werden.

Die Komplexität dieser Aufgaben nimmt zu, wenn die Modellparameter räumlich variabel sind. Die mit räumlichen Variationen verbundene Unsicherheit wird durch Zufallsfelder modelliert, die durch eine endliche, aber in der Regel große Anzahl von Zufallsvariablen diskretisiert werden. Diese Arbeit trägt zu verschiedenen Aspekten der Bayes'schen Inferenz und Zuverlässigkeitsanalyse im Kontext von Zufallsfeldern bei.

Zunächst werden Parameterstudien durchgeführt, um die Auswirkungen verschiedener Auswahlmöglichkeiten im a-priori Zufallsfeldmodell auf die Lösung von Bayes'schen inversen Problemen zu analysieren. Der Schwerpunkt liegt auf der Karhunen–Loève (KL) Entwicklung von Zufallsfeldern und dem Einfluss der Anzahl der Erweiterungsterme auf die a-posteriori Zufallsfelddiskretisierung. Es wird gezeigt, dass für eine angemessene Genauigkeit der Darstellung des a-posteriori-Zufallsfelds eine höhere Trunkierungsordnung erforderlich ist als für das a-priori Zufallsfeld. Daher wird vorgeschlagen, im Rahmen der Bayes'schen Inferenz nicht nur die Expansionskoeffizienten selbst, sondern auch deren Anzahl zu lernen. Dies motiviert die Entwicklung eines sequentiellen Algorithmus namens *trans-dimensional BUS*. Der Ansatz erweitert die BUS Methode (*Bayesian Updating with Structural reliability methods*) auf Probleme mit variablen Dimensionen. Trans-dimensional BUS kann die a-posteriori Verteilung sowohl der Dimension in der KL-Diskretisierung als auch der zugehörigen Zufallsfelder effizient abschätzen.

Darüber hinaus führen wir einen Vergleich verschiedener Simulationsmethoden zur Schätzung von Versagenswahrscheinlichkeiten in Anwendungen mit Zufallsfeldern durch. Wenn der Eingabeparameterraum hochdimensional wird, ist es für diese Algorithmen im Allgemeinen schwierig, die Wahrscheinlichkeit seltener Ereignisse hinreichend genau abzuschätzen. Daher wird eine neuartige Technik vorgeschlagen, welche die intrinsische niedrigdimensionale Struktur von Zuverlässigkeitsproblemen ausnutzt. Unsere Methode erstellt einen sogenannten *failure-informed subspace* (FIS), um die Dimension des Zuverlässigkeitsproblems effektiv zu reduzieren. Dieser Ansatz eignet sich besonders für *importance sampling* mit der *cross-entropy* Methode, da aus dem FIS niedrigdimensionale Verteilungen abgeleitet werden können. Der Algorithmus ist in der Lage, Versagenswahrscheinlichkeiten in hohen Dimensionen mit erheblich geringerer Variabilität und Stichprobengröße im Vergleich zu Standardalgorithmen für die Zuverlässigkeitsanalyse abzuschätzen.

Abstract

Computational models of physical systems in engineering and science are controlled by parameters whose values are oftentimes known with imprecision. The associated uncertainty can be reduced by coupling the model with measurements of the system response. This represents an inverse problem that can be solved within the Bayesian statistical framework. The solution is given in terms of an updated probabilistic model, obtained from the combination of prior probability information with the computational model and the measurement data. The updated model can be utilized to make predictions about the physical system, which is particularly relevant when ensuring the reliability of the system is a main concern. The reliability is quantified in terms of the probability of undesirable (rare) failure events that are estimated by rare event simulation methods.

The complexity of the Bayesian inference and rare event simulation tasks increases when the model parameters are spatially variable. The uncertainty associated with spatial variation is modeled by random fields discretized into a finite, but oftentimes large, number of random variables. In this work, several aspects of Bayesian and rare event analysis of random fields are proposed and discussed.

Parameter studies are performed to investigate the implications of different choices in the prior random field model on the solution of Bayesian inverse problems. Since the prior random field is discretized with the Karhunen–Loève (KL) expansion, a major focus is the assessment of the number of expansion terms in the posterior random field representation. The results show that a larger truncation order is necessary to adequately represent the posterior field, compared to the order required for the prior field approximation. As a result, learning the set of expansion coefficients together with their dimension becomes a relevant task in the Bayesian analysis of random fields. This motivates the development of a sequential algorithm called *trans-dimensional BUS*. The method extends the BUS (Bayesian Updating with Structural reliability methods) approach to variable-dimensional problems. Trans-dimensional BUS efficiently estimates the posterior distribution of both, the truncation of the KL discretization and the associated random field parameters.

Furthermore, a comparison of different simulation methods for the estimation of failure probabilities in random field applications is carried out. In general, when the input parameter space becomes high-dimensional, estimating the probability of rare events *accurately* is challenging for the standard algorithms. Hence, a technique that exploits the intrinsic low-dimensional structure of reliability problems is proposed. The method constructs a so-called *failure-informed subspace* (FIS) to effectively reduce the dimension of the rare event simulation problem. This framework especially fits within importance sampling (IS) with the cross-entropy method since low-dimensional distribution models that operate on the FIS can be derived. The algorithm is able to estimate rare event probabilities in high dimensions with considerably smaller variability and sample size, compared to standard simulation algorithms.

Acknowledgements

I want to express my gratitude to “Professor” for giving me the opportunity of being part of the ERA group and participate in the BAYES project. This was a life-changing moment that allowed me to continue with the career path I chose since my bachelor studies. From our meetings, I always tried to learn from his decisions, experiences, and his way of leading a successful research group.

I would like to especially thank Iason for his careful supervision and constant support during my doctoral studies. Without his advice, ideas, and patience, I would have not been able to complete this work. I am also thankful for the many things he taught me about life with his way of being, and definitely for introducing me to yoga. I will miss our meetings and discussions.

I also want to thank Elisabeth Ullmann, who as a mentor was always eager to help and discuss about my PhD progress, and Youssef Marzouk from MIT for allowing me to do a research stay in his group. This proved to be one of the best experiences in my doctoral studies. Every meeting and discussion with him and members of his group had a great deal of information and was always an opportunity to learn new things.

I am grateful to my family and friends for their constant encouragement, support, and understanding through the years. Particularly to Elizabeth, who started the PhD with me and during this process became someone important to me. Sebastian, Max E., Max T. and Jonas, for which I find myself feeling especially appreciative for the memories we have; I will always value our friendship (thanks to Seb who carefully reviewed the German text in the abstract). I am also grateful to many other friends and colleagues: Wolfgang, Antonis, Hugo, Marco, Kilian, Ionuț, and Kenan (thanks to Wolfgang who acted as an extra supervisor in the initial part of my studies). Finally, I have to mention of course, Lukas, Michaela, and Peter, for making my stay in Munich especial, and Diego A. Alvarez for instilling in me the research spirit and his continued support during the years.

My doctoral research was funded by the *Deutsche Forschungsgemeinschaft* (DFG) through the TUM International Graduate School of Science and Engineering (IGSSE) within the project 10.02 BAYES. I thank these institutions for the financial support.

Munich, Germany.

February, 2020.

Publications by the Author

Parts of this thesis are extracts from articles published as part of Felipe Uribe’s doctoral research. Felipe is the main author of the journal articles:

- (5) F. Uribe, I. Papaioannou, Y. M. Marzouk, D. Straub. [Cross-entropy-based importance sampling with failure-informed dimension reduction for rare event simulation](#), *SIAM/ASA Journal on Uncertainty Quantification*, accepted.
- (4) F. Uribe, I. Papaioannou, J. Latz, W. Betz, E. Ullmann, D. Straub. [Bayesian inference with subset simulation in varying dimension applied to the Karhunen–Loève expansion](#), *International Journal of Numerical Methods in Engineering*, under review.
- (3) F. Uribe, I. Papaioannou, W. Betz, D. Straub (2020). [Bayesian inference of random fields represented with the Karhunen–Loève expansion](#), *Computer Methods in Applied Mechanics and Engineering*. 358, 112632.

and also of the conference articles:

- (2) F. Uribe, I. Papaioannou, W. Betz, E. Ullmann, D. Straub (2017). [Random fields in Bayesian inference: effects of the random field discretization](#). In: *12th International Conference on Structural Safety & Reliability (ICOSSAR)*. Ed. by C. Bucher, B. R. Ellingwood, and D. M. Frangopol. TU-Verlag Vienna, pp. 799–808 (peer-reviewed).
- (1) F. Uribe, I. Papaioannou, W. Betz, J. Latz, D. Straub (2017). [Bayesian model inference of random fields represented with the Karhunen–Loève expansion](#). In: *2nd International Conference on Uncertainty Quantification in Computational Sciences and Engineering (UNCECOMP)*. Ed. by M. Papadrakakis, V. Papadopoulos, and G. Stefanou. pp. 1–12 (no peer-reviewed).

The articles (2) and (3) appear in Chapter 4, (1) and (4) are present in Chapter 5, and (5) appears in Chapter 6.

Felipe worked on a further journal article during his doctoral research. Despite not being included in this thesis, it is listed below:

- (0) D. A. Alvarez, F. Uribe, J. E. Hurtado (2018). [Estimation of the lower and upper bounds on the probability of failure using subset simulation and random set theory](#). *Mechanical Systems and Signal Processing*. 100, pp. 782–801.

Contents

Abstract	v
Acknowledgements	vii
Publications by the Author	ix
1. Introduction	1
1.1. Background and motivation	1
1.2. Scope	6
1.3. Organization	7
2. Representing spatial variation with random fields	9
2.1. Definition of a random field	9
2.2. Fundamental random field properties	11
2.2.1. Statistical properties	11
2.2.2. Continuity, differentiability and smoothness	13
2.3. Gaussian and related fields	14
2.4. Covariance operators and kernels	16
2.4.1. Matérn covariance kernels	16
2.5. Representing random fields with the Karhunen–Loève expansion	18
2.5.1. General remarks on random field discretization	18
2.5.2. Karhunen–Loève expansion	18
2.6. Error measures for random field discretization	20
2.7. Numerical applications	21
2.7.1. Example 1: one-dimensional cantilever beam	21
2.7.2. Example 2: two-dimensional plate with a hole	28
3. Reliability analysis and rare event simulation	33
3.1. The rare event simulation problem	33
3.2. Simulation methods for the estimation of rare event probabilities	34
3.2.1. Monte Carlo and importance sampling	35
3.2.2. Sequential importance sampling	38
3.2.3. Subset simulation	40
3.2.4. Cross-entropy method	41
3.2.5. Improved cross-entropy method	43
3.3. Numerical applications	47
3.3.1. Example 1: quadratic LSF in varying dimensions	47
3.3.2. Example 2: one-dimensional diffusion problem	48
4. Bayesian analysis of engineering models	53
4.1. The Bayesian approach to inverse problems	53
4.2. General comments on Markov chain Monte Carlo	56
4.2.1. Metropolis–Hastings	57
4.2.2. Preconditioned Crank–Nicolson (conditional sampling)	58

4.3.	The BUS approach to Bayesian inversion	59
4.3.1.	BUS with subset simulation	61
4.4.	Updating of rare event probabilities	63
4.5.	Analytical example: one-dimensional cantilever beam	66
4.5.1.	Analytical posterior	67
4.5.2.	Approximation of the posterior	67
4.5.3.	Model evidence	73
4.6.	Numerical example: two-dimensional plate with a hole	77
4.6.1.	Approximated posterior	78
4.6.2.	Reliability updating	82
5.	Bayesian inference of random fields: dimension selection and averaging	85
5.1.	Mathematical framework	85
5.1.1.	Selection of the prior distribution for the dimension	87
5.2.	MCMC in varying dimensions	89
5.2.1.	Reversible jump MCMC	89
5.2.2.	Saturated space methods	92
5.3.	Trans-dimensional BUS formulation	96
5.3.1.	Trans-dimensional BUS with subset simulation	97
5.3.2.	Efficiency metrics for BUS-SuS and tBUS-SuS	99
5.4.	Numerical applications	100
5.4.1.	Example 1: one-dimensional cantilever beam	100
5.4.2.	Example 2: two-dimensional diffusion (groundwater flow)	109
6.	Rare event simulation of random fields: dimension reduction	115
6.1.	General comments on dimension reduction	115
6.1.1.	Likelihood-informed subspace	116
6.1.2.	Active subspace	117
6.1.3.	Certified dimension reduction	118
6.2.	Construction of the failure-informed subspace	118
6.2.1.	Rare event simulation as a Bayesian inversion	118
6.2.2.	Optimal profile function	119
6.2.3.	Optimal projector	120
6.3.	Failure-informed cross-entropy-based importance sampling	121
6.3.1.	Formulation	121
6.3.2.	Refinement step	125
6.4.	Numerical applications	126
6.4.1.	Example 1: linear LSF in varying dimensions	128
6.4.2.	Example 2: quadratic LSF in varying dimensions	130
6.4.3.	Example 3: one-dimensional diffusion	131
6.4.4.	Example 4: two-dimensional plate with a hole	132
7.	Conclusions and outlook	137
7.1.	Conclusions	137
7.2.	Outlook	141
A.	Fundamentals on probability theory	145
A.1.	Setting a probabilistic framework	145
A.2.	Distributions of random variables	146
A.3.	Conditioning	148

B. Finite element method for 2D elasticity problems	151
B.1. Fundamental equations	151
B.2. Finite element formulation	153
B.2.1. Discretization of the displacement field	154
B.2.2. Discretization of the deformation field	154
B.2.3. Discretization of the stress field	155
B.2.4. Stiffness matrix and force vector	156
B.2.5. The random field case: spatially variable Young's modulus	157
B.3. Adjoint differentiation for plane stress problems	158
B.3.1. Adjoint for the rare event simulation setting	158
B.3.2. Adjoint for the Bayesian inversion setting	160
 Bibliography	 176
 List of Figures	 177
 List of Tables	 181
 List of Abbreviations	 185
 Index	 185

“The important thing in science is not so much to obtain new facts as to discover new ways of thinking about them”.

William H. Bragg, 1862-1942

1.1. Background and motivation

In engineering and science most physical phenomena are mathematically modeled by partial differential equations (PDEs), subjected to initial and boundary conditions. Oftentimes, analytical solutions of these equations are cumbersome to derive, due to the complexity and particularity of engineering problems. Discretization methods, such as the finite element method [118], are utilized to approach the problem numerically. The advent of high-speed computers and easy-to-use software implementations have not only increased the range of applications, but also have allowed efficient and accurate estimation of PDE solutions (see for example [156]).

Numerical approximations to PDEs require a proper description of the *system inputs* described by initial and boundary conditions (e.g., restrictions, applied loads). The approximation also requires information about the *system parameters* (e.g., material and geometric properties). In practice, inputs and parameters are roughly defined (imperfect) or simply unknown. This phenomenon is referred to as *uncertainty*. In order to appropriately evaluate and predict the behavior of the system of interest, it is paramount to quantify such uncertainty.

Since the seminal work of Ghanem and Spanos [90], the area of uncertainty quantification (UQ) has attracted considerable interest in recent years, as a way to characterize and manage the randomness in engineering applications. The main purpose of UQ is to provide knowledge about the sensitivity of the system response, after introducing uncertain information about the system inputs and parameters into the numerical model. Depending on the type of application, one might encounter different types of uncertainty; in general one can differentiate two particular sources (after Helton [113]): *aleatory uncertainty* (or irreducible or inherent), related to the impossibility of a deterministic description of the intrinsic variability of the phenomenon being considered, and *epistemic uncertainty* (or reducible), related to the resulting lack of knowledge or data about the phenomenon being observed. A detailed classification is described in [135], which also discriminates the uncertainty in terms of the model parameters, the model inadequacy, observation errors, numerical implementation, among others.

There exist several mathematical theories to represent and quantify uncertainty (see, e.g., [7, 212]). Probability theory [137, 182] has been the classical approach to deal with aleatory uncertainty. This theory can also be employed to address epistemic uncertainty via the Bayesian formalism (see, e.g., [72]). Alternatively, non-probabilistic methods can be utilized to model epistemic uncertainty, including: fuzzy set theory [251], which enables representation of imprecise concepts in terms of linguistic terms; Dempster–Shafer evidence theory [63, 205], which represents epistemic knowledge on probability distributions propagated through multi-valued functions; and random sets [134, 162], which is an extension of a probability theory to set-valued rather

than point-valued functions and provides a generalization of many of the other methodologies [7, 8]. The distinction between aleatory and epistemic uncertainty, and specifically the methods that are used to quantify them is usually a matter of debate (much in the same way to the classical discussion between frequentist and Bayesian statistics). In general, the modeling choice will depend on the type of problem, on the quality and type of information that is at hand, and ultimately on the analyst preferences. In this work, we take a probabilistic view of all the sources of uncertainty.

In probabilistic UQ, we can distinguish two fundamental tasks [227]: (i) *forward uncertainty propagation*, dealing with the direct incorporation of the uncertain parameters through the system of interest. The main purpose is to characterize the uncertainty of the system response using for example stochastic finite elements [90]; and (ii) *inverse uncertainty quantification*, dealing with the incorporation of observational data into the computational model to infer or identify the uncertain parameters describing the system of interest. Specific modeling scenarios might target different UQ objectives involving the combination of the previous fundamental tasks. As a result, we usually encounter the following UQ sub-tasks:

- *Rare event simulation*, dealing with the estimation of failure/rare event probabilities of the system of interest [45]. This task can be approached from both, forward and inverse perspectives. In the inverse case, one is interested in updating the probability after the computational model assimilates the observational data [224].
- *Model surrogates*, in which the aim is to construct functions that approximate the underlying numerical model in an efficient manner. This area of UQ has connections with statistical learning and artificial intelligence approaches [120, 32]. Some well-known techniques are: polynomial chaos expansions [249, 90], pseudo-spectral approximations/sparse grids [52], low-rank tensor approximations [95], response surfaces [44], artificial neural networks and support vector machines [121, 120], Gaussian processes [190], deep neural networks [189, 188], among many others.
- *Model reduction*, including cases where dimension reduction techniques are applied to identify the most important variables controlling the system behavior. Some techniques include: sensitivity analysis [213, 200], projection and low-rank methods [57, 54], reduced order models [247], model selection [85, 245], and manifold interpolation [46].
- *Stochastic optimization*, dealing with optimization settings where uncertain parameters are part of the formulation, resulting in random objective functions and/or random constraints [31]. This task arises for example in decision-making under uncertainty [136], policy optimization problems [33], optimal experimental design [75, 81], along with others.

The complexity of UQ problems increases when the uncertain parameters fluctuate randomly in space. Common examples occur in computational mechanics where material parameters are spatially varying, such as Young's modulus in elasticity theory [153], conduction and convection coefficients in heat transfer problems [170], or permeability fields in hydraulic tomography applications [168]. The uncertainty related to spatially varying properties is generally represented by random fields which are understood as random variables that take values in a function space. Discretization methods are required to define tractable finite-dimensional representations of random fields. Several techniques have been proposed in the literature: spatial averaging [240, 239], shape function method [154], basis random variable method [143], midpoint method [65], Karhunen–Loève (KL) expansion [216], optimal linear estimation [145], orthogonal series expansion [254], spectral representation [207], and others.

Most of random fields used in practice are defined by their first two central moments, as well as their correlation kernel (so-called second-order fields). Monitoring and other observational data

can be used in combination with the PDE model to obtain information on these parameters. This is an inverse UQ task in which the interest is not only a point estimate of the parameters, but also a probabilistic characterization of their uncertainty. Such problems can be approached within the Bayesian statistical framework [131, 210, 193]. The idea is to combine prior information about the parameters (encoded into a probability distribution) with a likelihood function that accounts for the relation between the PDE model and the noisy data. The solution of the Bayesian inverse problem is given in terms of a so-called posterior probability distribution. Closed-form expressions of this posterior are only available in some particular cases; for instance, models involving conjugate priors [210, 193] or non-Gaussian priors given by nonlinear transformations of Gaussians [61]. Hence, Bayesian inverse problems are generally solved using: (i) sampling-based methods that rely on Markov chain Monte Carlo (MCMC), such as sequential Monte Carlo [62], sequential importance sampling [198], structural reliability methods [223]; or (ii) approximation methods, such as transport maps [73], variational inference [80], and likelihood-free methods [209]. Sampling-based techniques usually explore the posterior distribution on-the-fly by constructing a set of intermediate measures that converge to the full posterior. Particularly, the BUS (Bayesian Updating with Structural reliability methods) approach proposed in [223] reformulates a Bayesian inverse problem as a classical reliability analysis or rare event estimation problem. This construction allows one to employ algorithms for rare event computation to sample from the posterior. For an efficient and sequential solution of the inverse problem, BUS is typically combined with subset simulation [15], although other rare event estimation methods can also be employed (see, e.g., [224]).

The posterior solution is especially sensitive to the prior assumptions, when the uncertain parameters are spatially varying and modeled by random fields. A common approach is to impose a Gaussian process prior [190], whose second-order moment properties need to be defined. The choice of the prior remains a challenge even for homogeneous Gaussian random fields controlled only by the variance and correlation length. This is because the prior information about the covariance kernel is usually vague and the data might not be sufficient to clearly identify the spatial correlation structure. Therefore, the assumed prior has a large influence on the posterior random field solution and subsequent predictions. We analyze this influence by means of parameter studies on the prior random field model in [235]. Alternatively, hierarchical Bayesian inference can be applied to make the posterior solution more robust to the prior choices. The idea is to include hyperparameters in the specification of the covariance kernel of the field. For instance, the work [161] applies the KL expansion with a hierarchical Gaussian process prior using the mean and variance of the field as hyperparameters. The authors in [220] include the correlation length of the field as a hyperparameter; they propose a parametrized covariance function to reduce the computational cost associated to the repeated solution of the eigenvalue problem, implied by the sample-based inference process. The work in [142] introduces a Metropolis-within-Gibbs algorithm to infer jointly a parameterized Gaussian random field and its correlation length; they applied a reduced basis algorithm to decrease the computational cost of the simulation. Moreover, in [194], Cauchy and Gaussian hyperpriors are applied to the correlation length of the field using non-homogeneous Matérn covariance kernels.

A fundamental parameter in Bayesian inverse problems related to random fields is the dimension of the field discretization. As discussed before, series expansions are usually applied to approximate random fields. In practice, it is common to truncate the expansion after a finite number of terms based on some variance-representation criterion. This heuristic is generally valid in prior situations when no information or observations about the field are available. In the inversion case, the optimal number of terms in the series expansion is unknown and it is controlled by the data [235]. The task of choosing an appropriate truncation for posterior random fields is a problem of Bayesian model choice or selection, whereby *model* refers to the dimension induced by the random field discretization. In this case, the prior distribution is hierarchically

structured into a prior for the parameters conditioned on the dimension and a discrete prior for the dimension itself. The estimation of the resulting joint posterior requires the exploration of a discrete-continuous parameter space. This task can be performed by independently estimating the normalizing constants of the posteriors at each particular dimension; an approach referred to as *within-model simulation* (see, e.g., [85, 51]). For expensive PDE models this is computationally prohibitive. Alternatively, simultaneous inference on both, dimension and parameters, can be carried out using MCMC algorithms that explore the disjoint union space by moving between different dimensions; this approach is called *across-model simulation*. The standard algorithm for across-model simulation is the reversible jump MCMC algorithm [96, 97], which searches the parameter space using proposals that allow jumps between different dimensions. In addition to the reversible jump algorithm, the method proposed in [47] adapts the Gibbs sampler to deal with model choice problems. Also, the work [184] presents an iterative jump-diffusion algorithm to perform model comparison; the approach uses two steps to construct a Markov process that generates samples from the joint posterior distribution, one jump step that makes discrete transitions between models, and a diffusion step that draws parameter samples between those jumps. In [222], the collection of models is expressed as a marked point process with each point representing a model, afterwards an MCMC scheme implements a birth-death step to determine the stationary distribution of the process. Moreover, the authors in [24] modify the reversible jump MCMC, such that the parameter dimensions are fixed in a so-called product or saturated space defined over all possible models and parameters. In the same direction, the work [94] shows that the reversible jump MCMC can be derived as a classical Metropolis–Hastings algorithm applied to the saturated space. Most of the efficient algorithms for Bayesian inversion in fixed-dimensions proceed in a divide-and-conquer manner by splitting the inference process into a sequence of tasks whose evolution converges towards the posterior. This idea has also been translated to the variable-dimensional setting. In [9], the authors combine sequential importance sampling with the reversible jump MCMC. The work in [123] develops a population-based reversible jump MCMC that is able to estimate high- and trans-dimensional posterior distributions. We propose an extension of the BUS formulation to spaces of varying dimension for the KL expansion in [233].

Another essential task involving the treatment and modeling of uncertainty arises when the satisfactory operation of the system under consideration needs to be guaranteed. This belongs to the area of reliability analysis and rare event simulation. The objective is to estimate the probability of undesirable rare events associated to failure. The performance of the system is measured by a limit-state function (LSF) expressing the relation between extreme or maximum system responses and predefined admissible thresholds. The system response is given in terms of a quantity of interest (QoI), e.g., maximum stresses or displacements of a structural system. In rare event simulation, the probability of failure represents the evaluation of a potentially high-dimensional integral whose magnitude is very small. Specialized Monte Carlo algorithms are utilized for its computation, e.g., directional and line sampling [69, 139], importance sampling schemes [206, 76, 180, 198], and multilevel splitting methods [15, 48, 38, 232]. Observational data can also be integrated into the rare event analysis to update the reliability of the system. This corresponds to a combined inverse UQ and rare event simulation tasks, in which the aim is to infer the probability of failure using the model and data. The resulting updated probability of failure is used, for instance, in structural health monitoring, where continuous observational data is available. It is also applied to the re-analysis of existing structures to identify potentially unsafe systems. Several works have shown that the structural reliability computed before and after the data can differ significantly due to the information gained about the structural condition. We refer to [176, 127, 16] for some methods and structural health monitoring applications, and also to [224, 181, 88, 235] for problems involving random fields.

When the underlying system parameters are modeled as random fields, estimating failure

probabilities becomes a challenging task due to the presence of spatial correlations and the increase in dimensionality of the reliability problem. Even if one applies random field discretization techniques, the magnitude of the failure probability still depends on the discretization resolution. Hence, choosing an insufficient number of terms in the random field expansion leads to inaccurate probability estimates (underestimated spatial variability). The effects of spatial variation on the reliability of different systems have been reported in several studies. We highlight the work in [152], which discusses the influence of the finite element and random field discretization on the failure probability estimate; the authors derive a formula to guide the refinement of the random field mesh depending on the probability values. In [254], the impact of the truncation order of the random field expansion is evaluated; the authors show that in their particular problem the KL expansion converges from above (from unconservative estimates) to the target reliability index. In [151], an importance sampling methodology is developed for the estimation of rare event probabilities for elliptic PDEs with random field coefficients; asymptotic analysis of the probability of failure and the effect of the finite element mesh is also investigated. We follow and extend this idea in [235], where a study on the influence of the covariance kernel and the truncation of the KL expansion on prior and posterior failure probabilities is performed. Several other applications have been developed in geotechnical engineering; we mention the study [99], where the influence of the spatial variation on the probability of failure of slopes is analyzed; the authors show that the omission of the spatial variation leads to unconservative estimates of the failure probability.

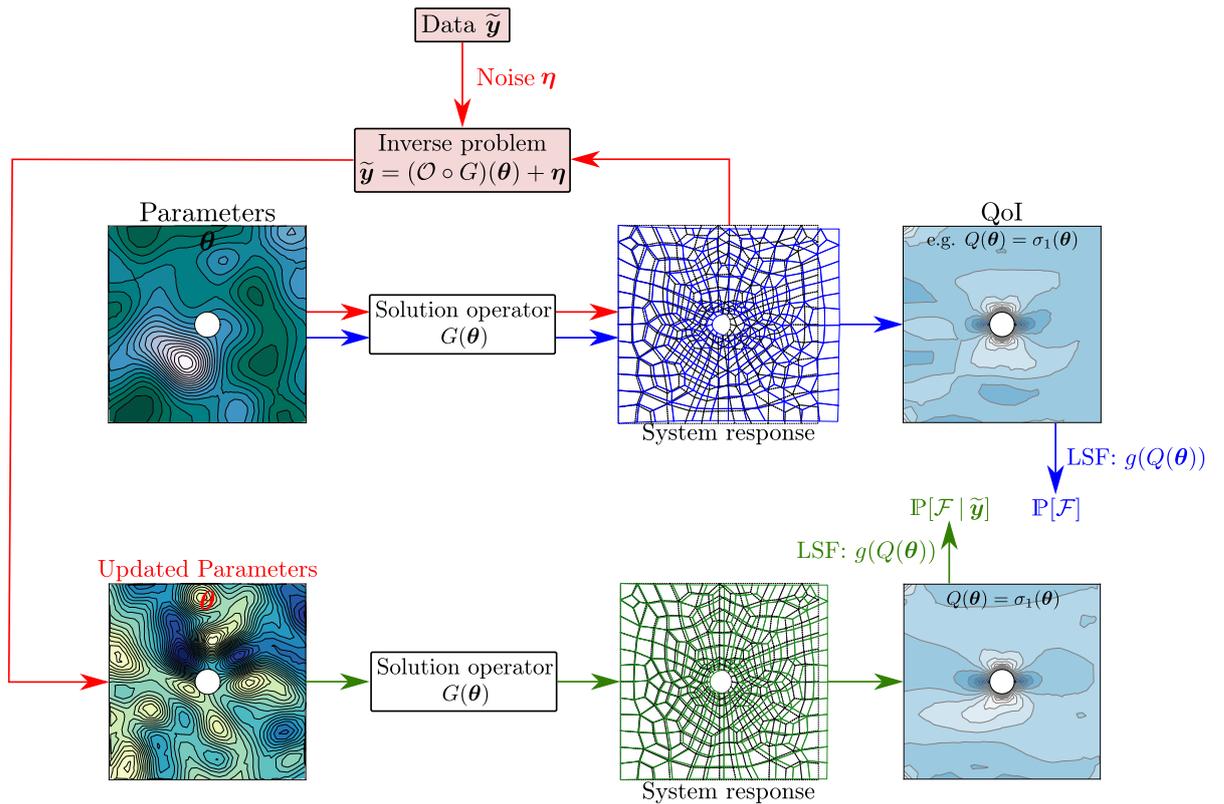


Figure 1.1: Schematic representation of inverse uncertainty quantification (red flow), forward rare event simulation (blue flow), and inverse rare event simulation (red and green flows) of an engineering model with random field parameters.

In addition to evaluating the influence of the spatial variation on the probability of failure, several efforts have been oriented towards developing efficient reliability estimation algorithms for random field applications. An approach is to combine the random field with the finite element model and to use polynomial chaos to expand the system response (so-called spectral stochastic

finite elements) [226, 5, 90]. The method in [18] uses a proper orthogonal decomposition to approximate the random field; the coefficients of this expansion are utilized to train a support vector machine that acts as a model surrogate during the failure probability estimation. In [128] the active subspace method is applied to fit Gaussian processes surrogates on a low-dimensional subspace; the surrogate is later employed in a Monte Carlo simulation to estimate the probability of failure. Moreover, we propose in [236] a method that identifies the intrinsic low-dimensional structure of reliability problems using dimension reduction techniques available in Bayesian inverse problems [57, 253]. A main source of such structure is the smoothing effect of the forward operator defining the LSF [253]. As a result, this function might vary predominantly along a few directions of the input parameter space, while being essentially constant in the remaining directions. The approach is developed in the context of the cross-entropy method [197, 179], and allows one to derive efficient importance sampling distribution models.

We conclude this section by summarizing the previous discussion in Figure 1.1. For inverse UQ, one aims at updating the random field after the inclusion of the observations (red flow). In forward rare event simulation, the objective is to estimate the probability of failure based on prior assumptions about the random field (blue flow). In inverse rare event simulation, the goal is to perform an inversion task to update the random field parameters, and subsequently estimate an updated probability of failure using the posterior random field (the combination of red and green flows).

1.2. Scope

This thesis concentrates on inverse uncertainty quantification and rare event simulation in engineering models with random field parameters. For the random field discretization, we focus on series expansion methods, in particular on the KL expansion. The scope of this thesis evolves around the reciprocal connection between Bayesian inversion and rare event simulation. These links, which become clear in the upcoming chapters, are highlighted as follows:

$$\underbrace{\pi_{\mathcal{F}}(\boldsymbol{\theta}) = \frac{1}{p_{\mathcal{F}}} \mathbb{1}_{\mathcal{F}}(\boldsymbol{\theta}) \pi_{\text{pr}}(\boldsymbol{\theta})}_{\text{Posterior-failure density that characterizes rare event simulation}} \quad \underbrace{h(k, \boldsymbol{\theta}_k, v) = v - r \cdot \text{L}(k, \boldsymbol{\theta}_k; \tilde{\mathbf{y}})}_{\text{Limit-state function that characterizes Bayesian inversion in varying dimensions}}. \quad (1.1)$$

Bayesian inverse problems

The Bayesian approach to inverse problems occurring in PDEs representing engineering systems is discussed. We analyze analytically and numerically the effects of prior random field choices on the posterior solution and on the updating of rare event probabilities. This is performed via parameter studies on the truncation order of the KL expansion and the parameters defining the covariance kernel of the field, such as correlation length and smoothing parameter. The emphasis is on the dimension of the random field discretization, for which we derive an analytical expression of the model evidence. In this case, model refers to the dimension of the KL expansion coefficients.

When addressing the Bayesian inverse problem numerically, we utilize the BUS formulation which connects Bayesian inversion with rare event simulation. Particularly, we combine BUS with subset simulation, a method that performs well in high dimensions. An extension of the BUS approach to variable-dimensional parameter spaces is proposed. We discuss several trans-dimensional MCMC algorithms that can be combined with the proposed BUS extension, and derive a prior for the dimension parameter. The resulting trans-dimensional BUS method is able to identify not only the coefficients of the expansion, but also their dimension. This allows one to perform model (dimension) selection and averaging in order to identify an appropriate

truncation order of the KL expansion for the representation of the posterior random field.

Topics that are not in the scope are: inverse problems related to integral operators, e.g., image deblurring and reconstruction [108], non-Gaussian and non-additive noise models [36], and data assimilation problems, such as Bayesian filtering or smoothing [228, 191].

Rare event simulation

We focus on simulation methods for the estimation of the probability of rare events in high dimensions. A review of some advanced approaches is carried out, including: sequential importance sampling, subset simulation, cross-entropy method, improved cross-entropy method. The performance of the algorithms is tested with benchmark problems in high dimensions, and also examples involving random fields.

We emphasize the connection between rare event simulation and Bayesian inversion. The posterior in rare event simulation has a different interpretation, but it has a similar structure to the posterior in Bayesian inversion. As a result, we can adapt dimension reduction techniques from Bayesian analysis to find the low-dimensional structure of rare event simulation problems. The construction of the resulting low-dimensional subspace involves computations of the LSF gradient. Since the posterior is equivalent to the zero-variance biasing density in importance sampling, we can derive efficient and effectively low-dimensional biasing distribution families. This is exploited in the context of the cross-entropy method. The proposed methodology can estimate small failure probabilities in high dimensions with small sample size and low coefficient of variation, compared to standard simulation techniques.

Topics that are out of the scope include: the estimation of reliability indexes via approximation methods, such as the first and second order reliability methods [187], and system and time-variant reliability analysis [144].

1.3. Organization

Chapter 2 starts with introductory topics on the theory of spatial variation and random fields. We describe random fields in terms of finite-dimensional distributions and give some basic insights on its measure theoretic definition. Fundamental properties of random fields are also introduced including statistics, such as expectation and covariance, and structural properties such as homogeneity. Several regularity conditions are also described, including continuity, differentiability and smoothness. Later on, we introduce Gaussian random fields and the Matérn class of covariance kernels. We also present the KL expansion of random fields, which is a fundamental concept in the thesis. Several error measures used to quantify the quality of the random field approximation are also presented. The chapter concludes with two numerical applications in which we perform several parameter studies to assess the influence of the covariance operator on the random field discretization, and subsequent estimation of response quantities.

Chapter 3 discusses the rare event simulation problem and sampling methods for its solution (blue flow in Figure 1.1). The Bayesian interpretation of the problem is also highlighted. We begin with classic Monte Carlo simulation and importance sampling. Afterwards, advanced sequential simulation schemes are presented. These include sequential importance sampling, subset simulation, cross-entropy method, and improved cross-entropy method. We describe some connections between these methodologies and evaluate their performance for the solution of high-dimensional reliability problems, arising in random field applications.

Chapter 4 presents the Bayesian approach to inverse problems in the context of engineering models (red flow in Figure 1.1). Some well-known summary statistics of the posterior distribution are described. Afterwards, two MCMC methods are presented, the classical Metropolis–Hastings

and the preconditioned Crank–Nicolson algorithm. Later, the BUS approach to Bayesian inversion and its combination with subset simulation is introduced. Since Bayesian analysis not only involves retrieving the uncertain parameters, but also rare event simulation tasks, we present a short description on the updating of failure probabilities (red and green flows in Figure 1.1). The new results described in this chapter start from section 4.5 with an analytical study on the influence of the prior random field choices on the posterior random field and predictions. In there, we derive updating rules for the random fields and KL coefficients and provide a numerical error analysis. A major point here is the derivation of the model evidence for the truncation order of the KL expansion. Analogous studies are performed numerically in subsequent sections, where we also assess the influence of the prior assumption on the estimation of posterior rare event probabilities.

Chapter 5 discusses a more general type of inference where finding the model is also part of the objective. We begin with a description of the Bayesian model inference framework, and the two most common tasks: model selection and model averaging. In our case, models are equivalent to the dimension of the random field discretization, e.g., truncation order of the KL expansion. Thereafter, we propose a discrete prior for the dimension parameter. To address this type of inference problems, we employ across-model simulation algorithms that are able to jump between dimensions. Therefore, some MCMC methods that operate in spaces of varying dimension are described, including: the reversible jump MCMC and saturated space approaches (Metropolis-within-Gibbs and step-wise samplers). Later on, we extend the BUS formulation to deal with variable dimensional problems in a sequential manner. At the end of the chapter, we test our algorithm on two engineering problems with random field parameters.

Chapter 6 deals with dimension reduction techniques for rare event simulation problems (blue flow in Figure 1.1). First, a short review of the most common projection methods used to find the low-dimensional structure of nonlinear Bayesian inverse problems is described. These include the likelihood-informed subspace, active subspace and certified dimension reduction approaches. We employ the latter to construct a subspace that defines the intrinsic low-dimensionality of reliability problems. This is exploited in the context of importance sampling with the cross-entropy methods, where efficient low-dimensional biasing distribution models can be derived. The resulting method is assessed with four numerical examples: two standard high-dimensional reliability benchmark problems and two random field applications.

Chapter 7 presents some final considerations, a summary of the main contributions of the dissertation, and a pool of basic ideas that could be used for future research.

The thesis concludes with Appendix A, containing a short overview of measure theoretic probability theory, which is fundamental for a better understanding of probabilistic UQ; and with Appendix B, where introductory concepts of finite element analysis for 2D elasticity problems are presented. In particular, we derive an adjoint method for the estimation of the LSF gradient for the numerical example in subsection 6.4.4.

Representing spatial variation with random fields

“A model must be of a simple and tangible type if it is meant to convey a clear idea of what a realization looks like... Needless to say, a model must often be almost grotesquely oversimplified in comparison with the actual phenomenon studied.”

Bertil Matérn, 1917-2007 [164, p.27].

Spatial variation is a common characteristic of uncertain phenomena occurring in several areas of science; some examples include: the heterogeneous distribution of soils in an aquifer, the micro/macroscopic patterns of material properties, the geometric imperfections of structural elements after construction, the spatial structure of particles suspended in a fluid, the intensity and texture distribution of a picture in image compression and restoration.

Oftentimes, the patterns or realizations of uncertain quantities induced by spatial variation are significantly complex and one requires a probabilistic description for their analysis and interpretation. The theory of *stochastic processes* provides a tool for studying random phenomena that vary in time. This theory can be generalized to deal with models that vary randomly in space. In this case, elements of the so-called *index set* of a stochastic process are no longer related to time, but instead to spatial coordinates. This type of processes are referred to as *random fields*.

Random fields can be roughly understood as an arranged set of random variables. The arrangement is specified by the index set pointing to different spatial locations of some topological space, which by definition is infinite-dimensional. In practice, one seeks a suitable finite-dimensional representation of the random field. Most of the available techniques for random field discretization require not only the specification of a set of spatial points, but also information provided by the expectation and covariance. In general, the aim is that the random field discretization “succeeds in capturing the essential features of a complex random phenomenon in terms of a minimum number of physically meaningful and experimentally accessible parameters” [239]. One can argue with the last part of this statement since parameters that are not necessarily accessible can be retrieved within the framework of inverse problems (Chapters 4 and 5).

We start by reviewing important concepts on random field theory and the discretization method that is used in our studies. We refer to Appendix A for a short discussion on probability theory that supports the details presented in this chapter.

2.1. Definition of a random field

Consider the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, together with a bounded index set $D \subseteq \mathbb{R}^n$ representing a physical domain. Let $G^{(n,d)} := \{g : D \rightarrow \mathbb{R}^d; d, n \in \mathbb{N}_{>0}\}$ be the set of functions mapping from D to \mathbb{R}^d , and $\mathcal{G}^{(n,d)} := \{g \in G^{(n,d)} : g(\mathbf{x}_j) \in B_j(\mathbb{R}^d), \mathbf{x}_j \in D, j = 1, \dots, k\}$ be an associated σ -algebra, where $B_j(\mathbb{R}^d)$ are Borel sets on \mathbb{R}^d (e.g., the smallest σ -algebra containing products of half-open intervals in \mathbb{R}^d), and k is an arbitrary index [2]. A (n, d) -dimensional *random field* is a measurable function H from the space (Ω, \mathcal{F}) into the *realization space* $(G^{(n,d)}, \mathcal{G}^{(n,d)})$.

From this definition, a random field can be understood as a random variable that takes values in a function space.

Random fields are usually denoted by $H(\mathbf{x}, \omega)$, meaning the value that a function in $G^{(n,d)}$ takes at the location \mathbf{x} for a given $\omega \in \Omega$. For fixed ω_i , the \mathbb{R}^d -valued function $h_i(\mathbf{x}) = H(\mathbf{x}, \omega_i)$ is a *realization* or sample path of the random field. Conversely, for a fixed $\mathbf{x}_j \in D$, the function $H_j(\omega) = H(\mathbf{x}_j, \omega)$ is a random variable. This relation is explained graphically in Figure 2.1 for a $(1, d)$ -dimensional random field and in Figure 2.2 for a $(2, d)$ -dimensional one. As mentioned in [256], it is useful to think of \mathbf{x} as “spatial coordinate” and of each ω as a “particle” or “experiment”.

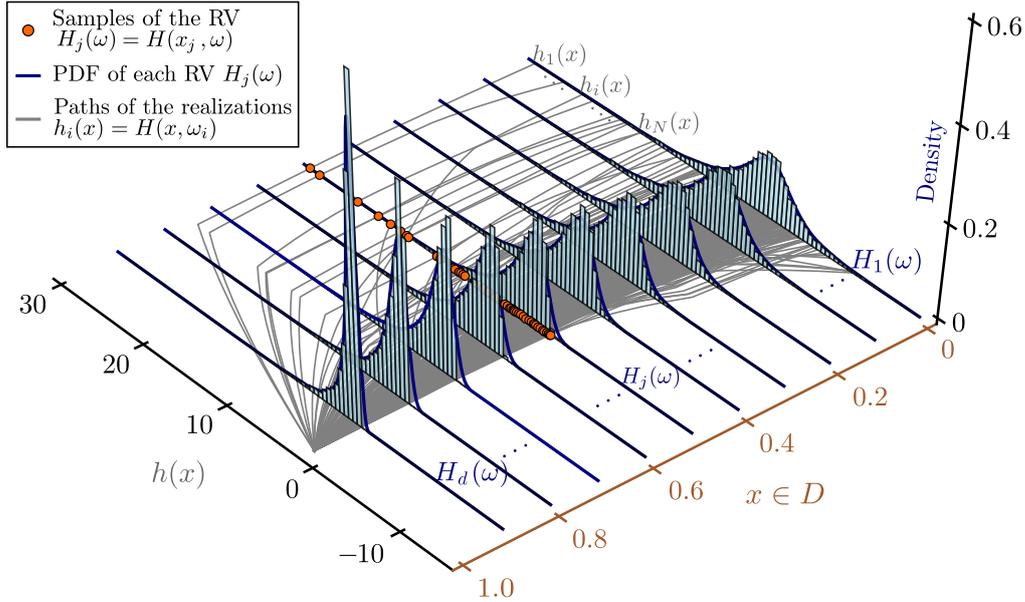


Figure 2.1: Evolution of marginal densities (represented as histograms) and realizations (gray lines) of a 1D random field: for a fixed element of the sample space $\omega_i \in \Omega$, the random field turns into a realization $h_i(x)$ across the spatial domain $D = [0, 1]$. For a fixed spatial coordinate $x_j \in D$, the random field turns into a random variable $H_j(\omega)$.

Given a probability measure \mathbb{P} on (Ω, \mathcal{F}) , one can define a corresponding law or distribution μ in the realization space $(G^{(n,d)}, \mathcal{G}^{(n,d)})$. This corresponds to $\mu(B_1 \times \cdots \times B_k) = \mathbb{P}[H(\mathbf{x}_1, \omega) \in B_1, \dots, H(\mathbf{x}_k, \omega) \in B_k]$, where each B denotes Borel sets in D . The associated collection of k cumulative distribution functions (CDF) given by

$$F_{H_1, \dots, H_k}(y_1, \dots, y_k) = \mathbb{P}[H(\mathbf{x}_1, \omega) \leq y_1, \dots, H(\mathbf{x}_k, \omega) \leq y_k] = \mathbb{P}[H_1(\omega) \leq y_1, \dots, H_k(\omega) \leq y_k], \quad (2.1)$$

with $k \leq d$, is known as the family of *finite-dimensional* (fi-di) distributions of the random field H [2]. For absolutely continuous fi-di CDFs, one can also define the corresponding set of probability density functions (PDF) as

$$\pi_{H_1, \dots, H_k}(y_1, \dots, y_k) = \frac{\partial^k F_{H_1, \dots, H_k}(y_1, \dots, y_k)}{\partial y_1 \cdots \partial y_k}. \quad (2.2)$$

Note that when $k = d$, one gets the full joint distribution of the field. Random fields can be directly defined in terms of the family of fi-di distributions (2.1) provided they exist and satisfy the consistency and symmetry conditions (see Kolmogorov’s extension theorem [256, p.11]):

- (i) *Consistency* (projection invariance): marginal distributions can be generated from higher dimensional ones, i.e., $F_{H_1, \dots, H_k, H_{k+1}, \dots, H_{k+m}}(y_1, \dots, y_k, \infty, \dots, \infty) = F_{H_1, \dots, H_k}(y_1, \dots, y_k)$ [215].

- (ii) *Symmetry* (permutation invariance): $F_{H_1, \dots, H_k}(y_1, \dots, y_k)$ is invariant under arbitrary permutation of the indices $j = 1, \dots, k$ (same permutation for both \mathbf{x}_j, y_j).

The fi-di distributions of the random field uniquely define the probability measure μ associated to the σ -algebra $\mathcal{G}^{(n,d)}$ in the realization space. This makes the measure-theoretic and fi-di-based definitions equivalent. Nevertheless, the characterization established by the fi-di distributions is the most used in practice since it is more intuitive to think of a random field as a collection of random variables representing uncertain values at each spatial coordinate in D (as it is seen in Figures 2.1 and 2.2).

Note that if $n = 1$ and instead of indexing spatial coordinates one indexes time, the random field reduces to a stochastic process. We also remark that the system of fi-di distributions does not fully characterize the random field realizations, in the sense that two random fields can share the same fi-di distributions but have completely different sample paths. This is because fi-di distributions do not determine the continuity properties of the random field (see, e.g., [30, p.482]).

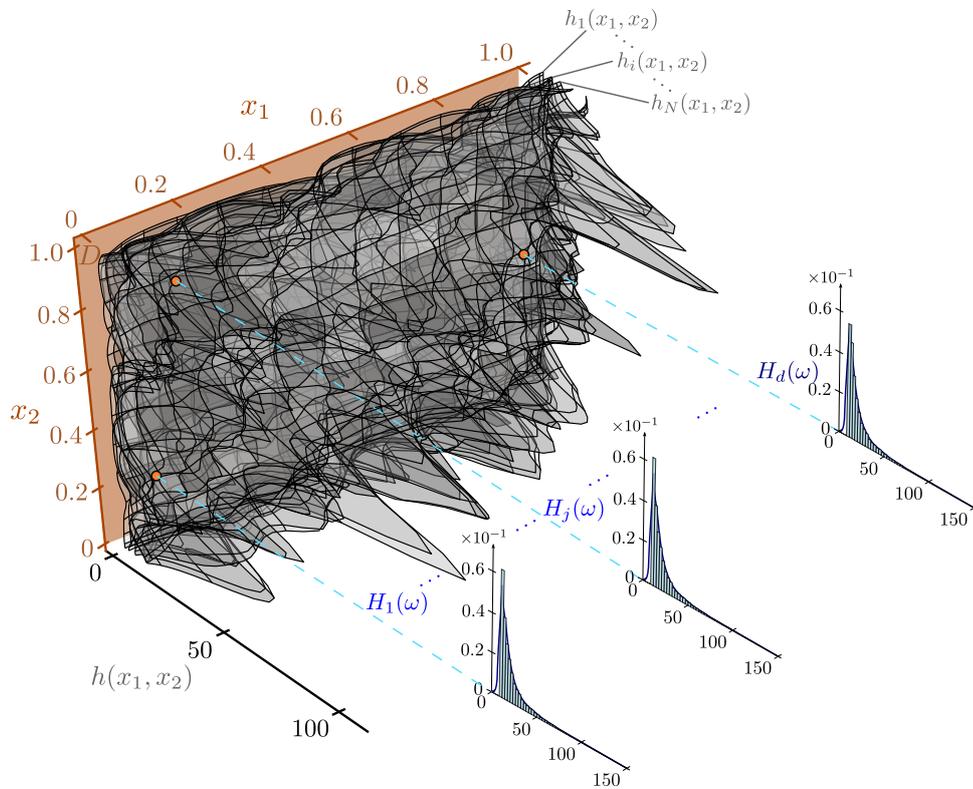


Figure 2.2: Some marginal densities (represented as histograms) and realizations (gray surfaces) of a 2D random field: as in Figure 2.1, for a fixed element of the sample space $\omega_i \in \Omega$, the random field turns into a realization $h_i(x_1, x_2)$ across the spatial domain $D = [0, 1] \times [0, 1]$. For a fixed spatial coordinate $\mathbf{x}_j = [x_1^{(j)}, x_2^{(j)}] \in D$, the random field turns into a random variable $H_j(\omega)$.

2.2. Fundamental random field properties

2.2.1. Statistical properties

Analogous to the analysis of random variables, some of the most important properties of random fields are characterized by their statistical moments. Consider a second-order random field expressed by the collection of random variables $H_{\mathbf{x}} = \{H(\mathbf{x}_1, \omega), \dots, H(\mathbf{x}_j, \omega), \dots, H(\mathbf{x}_d, \omega)\}$ with full joint fi-di distribution $F_{H_{\mathbf{x}}}$ and density $\pi_{H_{\mathbf{x}}}$, where $\mathbf{x} = \{\mathbf{x}_1, \dots, \mathbf{x}_j, \dots, \mathbf{x}_d : \mathbf{x}_j \in D\}$ is a countable set of spatial points.

The *expectation* or mean of a random field is the \mathbb{R}^d -valued function

$$\mu_H(\mathbf{x}) = \mathbb{E}[H(\mathbf{x}, \omega)] = \int_{\Omega} H(\mathbf{x}, \omega) \mathbb{P}(d\omega) = \int_{\mathbb{R}^n} \mathbf{y} dF_{H_{\mathbf{x}}}(\mathbf{y}) = \int_{\mathbb{R}^n} \mathbf{y} \pi_{H_{\mathbf{x}}}(\mathbf{y}) d\mathbf{y}, \quad (2.3)$$

where $\mathbf{y} = [y_1, \dots, y_d]$. The second moment or *correlation* of the random field $H(\mathbf{x}, \omega)$ with itself $H(\mathbf{x}', \omega)$ (i.e., the autocorrelation) is given by the $\mathbb{R}^{d \times d}$ -valued function

$$R_{HH}(\mathbf{x}, \mathbf{x}') = \mathbb{E}[H(\mathbf{x}, \omega)H(\mathbf{x}', \omega)] = \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \mathbf{y} \mathbf{y}' \pi_{H_{\mathbf{x}}, H_{\mathbf{x}'}}(\mathbf{y}, \mathbf{y}') d\mathbf{y} d\mathbf{y}', \quad (2.4)$$

where $\pi_{H_{\mathbf{x}}, H_{\mathbf{x}'}}(\mathbf{y}, \mathbf{y}')$ represents the joint density of the random field with itself [1]. Similarly, the *covariance* of the random field $H(\mathbf{x}, \omega)$ with itself $H(\mathbf{x}', \omega)$ (i.e., the autocovariance) is the $\mathbb{R}^{d \times d}$ -valued function

$$C_{HH}(\mathbf{x}, \mathbf{x}') = \text{Cov}[H(\mathbf{x}, \omega), H(\mathbf{x}', \omega)] = \mathbb{E}[H(\mathbf{x}, \omega)H(\mathbf{x}', \omega)] - \mu_H(\mathbf{x})\mu_H(\mathbf{x}') \quad (2.5a)$$

$$= R_{HH}(\mathbf{x}, \mathbf{x}') - \mu_H(\mathbf{x})\mu_H(\mathbf{x}'), \quad (2.5b)$$

note that $C_{HH}(\mathbf{x}, \mathbf{x}') = R_{HH}(\mathbf{x}, \mathbf{x}')$, if the random field is centered (the mean function is zero for all $\mathbf{x} \in D$). Moreover, the normalized covariance function

$$\rho_{HH}(\mathbf{x}, \mathbf{x}') = \frac{C_{HH}(\mathbf{x}, \mathbf{x}')}{\sigma_H(\mathbf{x})\sigma_H(\mathbf{x}')}, \quad (2.6)$$

is called the *correlation coefficient* function, where $\sigma_H(\mathbf{x})$ corresponds to the *standard deviation* of the field. If $\mathbf{x} = \mathbf{x}'$, the covariance becomes the variance function $\sigma_H^2(\mathbf{x}) = C_{HH}(\mathbf{x}, \mathbf{x}) = R_{HH}(\mathbf{x}, \mathbf{x}) - \mu_H^2(\mathbf{x})$, where $R_{HH}(\mathbf{x}, \mathbf{x}) = \mathbb{E}[H^2(\mathbf{x}, \omega)]$ is also known as the *mean-squared function*.

A random field is strictly homogeneous or just *homogeneous*, if the distribution functions (2.1) are invariant under arbitrary translations in space. That is, for all spatial coordinates $\mathbf{x}' \in D$

$$\mathbb{P}[H(\mathbf{x}_1, \omega) \leq y_1, \dots, H(\mathbf{x}_k, \omega) \leq y_k] = \mathbb{P}[H(\mathbf{x}_1 + \mathbf{x}', \omega) \leq y_1, \dots, H(\mathbf{x}_k + \mathbf{x}', \omega) \leq y_k], \quad (2.7)$$

where each $(\mathbf{x}_j + \mathbf{x}') \in D$ (with $j = 1, \dots, k$ and $k \leq d$). The random field is *weakly homogeneous*, if the mean function is constant across D and the correlation function only depends on the *separation vector* $\mathbf{h} = \mathbf{x} - \mathbf{x}'$, that is

$$\mu_H(\mathbf{x}) = \mu_H < \infty \quad \text{and} \quad R_{HH}(\mathbf{x}, \mathbf{x}') = R_{HH}(\mathbf{h}) = R_{HH}(-\mathbf{h}); \quad (2.8)$$

under (2.8), the covariance function (2.5) reduces to $C_{HH}(\mathbf{h}) = \sigma_H^2 \rho_{HH}(\mathbf{h})$, since the variance of a homogeneous field is also constant, $\sigma_H^2(\mathbf{x}) = \sigma_H^2 < \infty$.

Furthermore, if the homogeneous correlation/covariance function is also independent of the direction (rotations and reflections), the random field is called *isotropic*. In this case, the covariance is only a function of the norm on D . Since in most cases $D \subset \mathbb{R}^n$, we employ the Euclidean norm $\|\mathbf{h}\|_2 = h$ and we write $C_{HH}(\mathbf{h}) = C_{HH}(h)$. Therefore, isotropic covariance functions are a subclass of the homogeneous ones. One can also use a different norm to compute the distances between the spatial coordinates. This provides more flexibility in the application of homogeneous covariance functions. For instance, one can define an *anisotropic* random field by considering the non-Euclidean norm $h = \|\mathbf{h}^T \mathbf{A} \mathbf{h}\|_{\mathbf{A}} = \sqrt{\mathbf{h}^T \mathbf{A} \mathbf{h}}$, where the matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ must be positive semi-definite ($\mathbf{x}^T \mathbf{A} \mathbf{x} \geq 0$ for all $\mathbf{x} \in D$) in order to define a valid covariance function. The norm $\|\cdot\|_{\mathbf{A}}$ induces an ellipsoidal symmetry (at the iso-correlation surfaces) and

as a result, the correlation/covariance functions are termed *ellipsoidal* [164].

The dependence structure of physical phenomena is oftentimes spatially varying [84]. This requires the application of non-homogeneous random field models, which in some cases can be derived from homogeneous ones. For instance, consider an isotropic random field $H(\mathbf{x}, \omega)$, with $\mu_H(\mathbf{x}) = 0$ and $C_{HH}(\mathbf{x}, \mathbf{x}') = R_{HH}(h)$. We can now define another random field as $Y(\mathbf{x}, \omega) = \sigma_Y^2(\mathbf{x})H(\mathbf{x}, \omega)$, such that $\mu_Y(\mathbf{x}) = 0$ and $C_{YY}(\mathbf{x}, \mathbf{x}') = \sigma_Y(\mathbf{x})\sigma_Y(\mathbf{x}')R_{HH}(h)$. Since the variance is no longer constant, the random field $Y(\mathbf{x}, \omega)$ is non-homogeneous. Similarly, one can define a random field $Y(\mathbf{x}, \omega) = \mu_Y(\mathbf{x}) + H(\mathbf{x}, \omega)$ that is also non-homogeneous since the mean is spatially variable. Further combinations of isotropic/anisotropic random fields can also be applied to generate non-homogeneity [1].

Finally, one can also define whether two random fields are considered equivalent or not. The importance of this lies in the fact that even if two random fields share the same fi-di distributions, they do not necessarily generate the same realizations. First, we say that two random fields $H(\mathbf{x}, \omega), Y(\mathbf{x}, \omega)$ are *equivalent versions* of each other if $\mathbb{P}[\omega : H(\mathbf{x}, \omega) = Y(\mathbf{x}, \omega)] = 1$ for any $\mathbf{x} \in D$. If the random fields are equivalent, they share the same fi-di distributions [1]. However, as mentioned before, they are not necessarily identical in the sense that each field can have different sample paths (see classical example in [2, p.14]). In order to avoid such ‘‘pathological’’ cases and to employ fi-di distributions to also characterize the sample path properties of the field, one typically imposes a regularity property on the random fields, which is *separability*. Basically, this property considers that the index set D is a separable space¹. The idea behind the separability condition is to determine the properties of the random field based on a countable set of points of the domain D . In practice, separability poses no further restrictions since the discretized domain D itself satisfies the separability condition [70]. One always implicitly assumes that the fields are separable and as a result, the fi-di distributions completely characterize the random field.

2.2.2. Continuity, differentiability and smoothness

In practice, it is important that the realizations of a random field satisfy some regularity properties. Adler [2, Ch.3] derives a set of conditions on the fi-di distributions ensuring that the sample paths of random fields are equipped with these properties. The most fundamental regularity condition one requires in practice is continuous differentiability, which is directly related to the *smoothness* of the field. For example, Gaussian random fields are considered regular, if they have sample function continuity and differentiability. We shall now comment on some of these results, which must be taken into account when using random field theory to represent uncertainty.

Continuity of random fields is associated to the convergence of sequences of random variables $\{H(\mathbf{x}_1, \omega), \dots, H(\mathbf{x}_d, \omega)\}$ as $d \rightarrow \infty$. We say that a random field is *continuous* at $\mathbf{x} \in D$, if for every sequence of points $\{\mathbf{x}_d\}$ for which $\|\mathbf{x}_d - \mathbf{x}\| \rightarrow 0$ as $d \rightarrow \infty$, then

$$\mathbb{P}[\omega : |H(\mathbf{x}_d, \omega) - H(\mathbf{x}, \omega)| \rightarrow 0, d \rightarrow \infty] = 1. \quad (2.9)$$

Condition (2.9) is called almost-sure convergence, and it implies convergence in probability and in distribution (see, e.g., [249]). If (2.9) is applicable at each point $\mathbf{x} \in D$, the random field has sample function continuity [2]. Moreover, the random field is *mean-squared continuous* in D , if

$$\mathbb{E}[(H(\mathbf{x}_d, \omega) - H(\mathbf{x}, \omega))^2] \rightarrow 0 \quad \text{as } d \rightarrow \infty, \forall \mathbf{x} \in D. \quad (2.10)$$

¹A metric space D is *separable*, if it contains a countable everywhere dense subset D' . D' is *everywhere dense* in D , if $[D'] = D$, where $[D']$ denotes the *closure* of D' , meaning the set of all contact points of D' . A point $x \in D$ is a *contact point* of D' , if every neighborhood (open sphere) of x contains at least one point of D' [138].

Despite the fact that mean-squared continuity does not imply continuous sample paths, the condition (2.10) has an important relation with the covariance function. A random field is mean-squared continuous at $\mathbf{x} \in D$, if and only if $C_{HH}(\mathbf{x}, \mathbf{x}')$ is continuous at $\mathbf{x} = \mathbf{x}'$, i.e., continuous variance (see [1]). Particularly, for homogeneous random fields the mean-squared continuity holds, if and only if $C_{HH}(\mathbf{h})$ is continuous at $\mathbf{h} = \mathbf{0}$.

Differentiability of random fields is another important regularity property. The gradient of a random field is a space vector in $D \subset \mathbb{R}^n$, denoted $\nabla H(\mathbf{x}, \omega) = [\dot{H}_1(\mathbf{x}, \omega), \dots, \dot{H}_n(\mathbf{x}, \omega)]^\top$. For a fixed ω , each component is given by

$$\dot{H}_i(\mathbf{x}, \omega) = \frac{\partial H(\mathbf{x}, \omega)}{\partial x_i}. \quad (2.11)$$

In practice, we are mainly interested in mean-squared differentiability since it has a simple relation with the covariance function [1]. One says that a random field is *mean-squared differentiable*, if for every sequence of points $\{\mathbf{x}_d\}$ for which $\|\mathbf{x}_d - \mathbf{x}\| \rightarrow 0$ as $d \rightarrow \infty$, then

$$\mathbb{E} \left[\left(\dot{H}_i(\mathbf{x}_d, \omega) - \dot{H}_i(\mathbf{x}, \omega) \right)^2 \right] \rightarrow 0 \quad \text{as } d \rightarrow \infty, \text{ for } i = 1, \dots, n \text{ and } \forall \mathbf{x} \in D. \quad (2.12)$$

For homogeneous random fields with covariance function $C_{HH}(\mathbf{h})$, mean-squared differentiability (2.12) is specified by the covariance function. Specifically, if the derivative $\partial^2 C_{HH}(\mathbf{h}) / \partial h_i^2$ exists and is finite at $\mathbf{h} = \mathbf{0}$, for all components $i = 1, \dots, n$, then the random field is mean-squared differentiable [1]. This means that the smoothness of the covariance function defines the smoothness of the random field. Note that continuity and differentiability is directly related to the behavior of $C_{HH}(\mathbf{h})$ in the proximity to $\mathbf{0}$. For homogeneous Gaussian random fields, this property of the covariance function determines essentially the local behavior (continuity and differentiability) of the random field realizations [2].

2.3. Gaussian and related fields

A *Gaussian random field* $H(\mathbf{x}, \omega)$ is a random field whose fi-di distributions are all multivariate Gaussian. In this case, the field is determined by the mean (2.3) and covariance (2.5) functions. Gaussian fields are fundamental in the study of spatial variation, not only because they are reasonable models in some applications, but also (and mainly) because they have a simple construction that enables analytical tractability. The definition and properties of the multivariate Gaussian distribution are presented next.

A random vector $\mathbf{X} = [X_1, \dots, X_d]^\top$ is multivariate Gaussian distributed, if the \mathbb{R} -valued random variable $Y = \sum_{i=1}^d y_i X_i$ is Gaussian, for every $\{y_1, \dots, y_d\}$. The corresponding probability density of \mathbf{X} is

$$\pi_{\mathbf{X}}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^d \det(\boldsymbol{\Sigma}_{XX})}} \exp \left(-\frac{1}{2} \left[(\mathbf{x} - \boldsymbol{\mu}_X)^\top \boldsymbol{\Sigma}_{XX}^{-1} (\mathbf{x} - \boldsymbol{\mu}_X) \right] \right), \quad (2.13)$$

where the mean vector $\boldsymbol{\mu}_X \in \mathbb{R}^d$ has elements $\mathbb{E}[X_i] = \mu_X^{(i)}$, and the covariance matrix $\boldsymbol{\Sigma}_{XX} \in \mathbb{R}^{d \times d}$ has elements $\mathbb{E}[(X_i - \mu_i) - (X_j - \mu_j)] = \Sigma_{XX}^{(ij)}$.

Many important properties of multivariate Gaussian vectors relate to the individual distributions of their groupings. This is because any k -grouping of \mathbf{X} is also Gaussian distributed, with $k = \{1, \dots, d-1\}$ [239]. If one defines the partition $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2]$, such that $\mathbf{X}_1 = [X_1, \dots, X_k]$ and $\mathbf{X}_2 = [X_{k+1}, \dots, X_d]$, the mean vector and covariance matrix can be divided in terms of individual and crossed components associated to each grouping [239]

$$\boldsymbol{\mu}_X = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix} \quad \boldsymbol{\Sigma}_{XX} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21}^\top & \boldsymbol{\Sigma}_{22} \end{bmatrix}, \quad (2.14)$$

this representation allows us to compute related marginal and conditional distributions, which are paramount in Bayesian inference (cf. Chapter 4).

The fi-di distributions induced by the Gaussian assumption satisfy the conditions of consistency and symmetry (see, e.g., [1, Sec. 1.4]). Thus, the main task for the definition of Gaussian random fields is the specification of a proper correlation/covariance function. This is because the sample path continuity of Gaussian field is basically controlled by this function. For homogeneous Gaussian random fields, the sample path continuity can also be verified as [1]

$$\mathbb{E} \left[(H(\mathbf{x}, \omega) - H(\mathbf{x}', \omega))^2 \right] = 1 - \rho_{HH}(\|\mathbf{h}\|) \leq \frac{c}{|\ln(\|\mathbf{h}\|)|^{1+\epsilon}}, \quad (2.15)$$

for some finite $c > 0$ and $\epsilon > 0$. One can check that the bound (2.15) is an increasing function of $\|\mathbf{h}\|$ and the curves (for different ϵ) approach 0 with large slope. Since almost every continuous correlation function with an exponential decay complies with this condition, Abrahamsen [1, p.20] states that ‘‘Gaussian random fields with continuous mean and covariance functions will generate continuous sample paths’’.

Other non-Gaussian random fields can be generated from Gaussian ones. The process usually involves the application of nonlinear mappings that preserve probability in both underlying spaces, so-called isoprobabilistic transformations. The theory of *translation fields* is based on such transformations [102]. Consider a homogeneous Gaussian random field $H(\mathbf{x}, \omega)$ with mean zero and variance one, and the transformation $T = F_Y^{-1} \circ \Phi(\cdot)$, where $\Phi(\cdot)$ denotes the standard Gaussian CDF. A translation field $Y(\mathbf{x}, \omega)$ is defined as [101]

$$Y(\mathbf{x}, \omega) = T(H(\mathbf{x}, \omega)) = F_Y^{-1} \circ \Phi(H(\mathbf{x}, \omega)), \quad (2.16)$$

where the field $Y(\mathbf{x}, \omega)$ is also homogeneous and has marginal distributions F_Y . Depending on the distribution F_Y and the covariance function of the Gaussian field $H(\mathbf{x}, \omega)$, one can employ (2.16) to construct a non-Gaussian field. For instance, in [101], the idea is to define $Y(\mathbf{x}, \omega)$ using a polynomial chaos expansion of the Gaussian field in terms of Hermite polynomials. The covariance function of $Y(\mathbf{x}, \omega)$ is also derived from the polynomial chaos expansion. This scheme is referred to as the direct problem since one employs (2.16) to establish the relation between the Gaussian and non-Gaussian field. However, there are cases where translation model (2.16) does not exist. This is called the inverse problem since one no longer employs the relation (2.16), and only information about the marginals and covariance function is used. In this case, there are three ways to approach the problem [101]: (i) match the marginal distributions and approximate the covariance function, (ii) estimate the marginal distributions and match the covariance function, or (iii) approximate both the marginal distributions and the covariance functions. We refer to [100] for some examples and further discussion. We also point out that copula theory can be applied to model different types of non-Gaussian correlation structures (see, e.g., [243]).

A classical example of a non-Gaussian field is the χ^2 field with parameter k , which is defined as $Y(\mathbf{x}, \omega) = [H^{(1)}(\mathbf{x}, \omega)]^2 + \dots + [H^{(k)}(\mathbf{x}, \omega)]^2$, where each individual random field is homogeneous and Gaussian with mean zero, and they all share the same covariance function [2]. Another class of non-Gaussian random fields that has significant importance in engineering practice is the lognormal random field. This is because material properties are always positive quantities. Since the Gaussian and lognormal distributions are related to each other, one can define a translation model (2.16) from a Gaussian to a lognormal field that is simple to apply. Thus, a lognormal field realization $Y(\mathbf{x}, \omega)$ can be constructed by a transformation of a Gaussian

field realization $h(\mathbf{x}, \omega)$ as

$$Y(\mathbf{x}, \omega) = \exp(H(\mathbf{x}, \omega)). \quad (2.17)$$

In our applications, we employ homogeneous and non-homogeneous random field models that are either Gaussian or lognormal.

2.4. Covariance operators and kernels

Covariance kernels for random field modeling are empirical models used to define the particular spatial correlation of a random field. Three essential properties of covariance functions that follow from its definition are: (i) *symmetry*, $C_{HH}(\mathbf{x}, \mathbf{x}') = C_{HH}(\mathbf{x}', \mathbf{x})$, (ii) *positive semi-definiteness*, i.e., $\sum_{i=1}^d \sum_{j=1}^d c_i c_j C_{HH}(\mathbf{x}_i, \mathbf{x}_j) \geq 0$, for each $\{\mathbf{x}_1, \dots, \mathbf{x}_d\} \in D$ and $c_1, \dots, c_d \in \mathbb{R}$, and (iii) *continuity*. Moreover, covariance functions are close under addition, multiplication, limits and integration [1, Thm. 3.1].

In general, covariance functions can be understood as kernels. Consider $D \subset \mathbb{R}^n$ be a bounded domain. A function $k : D \times D \rightarrow \mathbb{R}$ is called a *Hilbert–Schmidt kernel* if

$$\int_D \int_D |k(x, y)|^2 dx dy < \infty; \quad (2.18)$$

the associated integral operator $K : L^2(D) \rightarrow L^2(D)$, defined as

$$(Ku)(x) = \int_D k(x, y)u(y)dy \quad \forall u \in L^2(D) \quad (2.19)$$

is called a *Hilbert–Schmidt operator*. Here, K is understood as a linear operator in an infinite-dimensional setting. In practice, a spatial discretization is imposed on D to define a finite representation of such operator, this process generates a so-called *Gram matrix*. In the context of covariance kernels, the resulting finite-dimensional operator is the *covariance matrix*.

Hilbert–Schmidt operators are in particular self-adjoint, this guarantees that covariance matrices are Hermitian, and thus, they have a complete set of eigenvectors and real eigenvalues [6] [138, p.248]. This is an important fact that allows one to invoke Mercer’s theorem [166] to define the spectral representation of the covariance kernel. This is because homogeneous correlation functions $R_{HH} : D \rightarrow \mathbb{R}$ are positive-semi-definite, if and only if they can be represented by the n -dimensional Fourier-Stieltjes integral:

$$R_{HH}(\mathbf{h}) = \int_D \exp(i\boldsymbol{\omega}\mathbf{h})d^n\Psi(\boldsymbol{\omega}), \quad (2.20)$$

where $\Psi(\boldsymbol{\omega})$ is a real, non-decreasing and bounded function, called the *spectral measure* or spectral distribution function. The representation (2.20) is due to Bochner’s theorem [35], and it basically specifies that all positive-semi-definite functions have an unique spectral representation.

Different parametric families of covariance functions are available in the literature, these include the Cauchy, Bessel and exponential families (several examples are reported in [1, 90, 219]). A well-known family in spatial statistics is the Matérn class.

2.4.1. Matérn covariance kernels

The Matérn kernel is defined by considering that for isotropic correlation functions the condition (2.20) can be replaced by a one-dimensional Bessel transform [250, p.353]. The original Matérn kernel is [164, p.18]

$$C_\nu(h) = \sigma_H^2 \frac{2}{\Gamma(\nu)} \left(\frac{bh}{2}\right)^\nu K_\nu(bh), \quad (2.21)$$

where $h = \|\mathbf{x} - \mathbf{x}'\|_2$, $\Gamma(\cdot)$ is the gamma function, $K_\nu(\cdot)$ is the modified Bessel function of the second kind, and the constants $b, \nu > 0$. For $\nu = 1$, the covariance function reduces to $C_1(h) = \sigma_H^2 \cdot (bhK_1(bh))$; this kernel is known as *elementary* and it represents the covariance of a Markov field [164]. A similar model was proposed by Whittle [246]:

$$C_\nu(h) = \sigma_H^2 \frac{1}{\Gamma(\nu + 1)} \left(\frac{h}{2b}\right)^\nu K_\nu(bh). \quad (2.22)$$

Due to the similarities between (2.21) and (2.22), the kernels are sometimes grouped together as the Whittle-Matérn class. However, Handcock and Stein [105] introduced the standard Matérn family formulation which re-parameterizes (2.21) in terms of a scale parameter $\ell \in \mathbb{R}_{>0}$ controlling the range of correlation (*correlation length*) and a *smoothing parameter* $\nu \in \mathbb{R}_{>0}$ controlling the smoothness of the random field. In this case, the Matérn kernel is expressed as

$$C_\nu(h) = \sigma_H^2 \frac{2^{1-\nu}}{\Gamma(\nu)} (bh)^\nu K_\nu(bh) \quad (2.23)$$

where $b = (\sqrt{2\nu}h)/\ell$. A slightly different version of (2.23) is proposed in [106], where $b = (2\sqrt{\nu}h)/\ell$ is used. In general, there exist several parametrizations of the Matérn kernel each of which have particular advantages in different fields (see for instance [221, p.31],[150, 83]). We employ the following type of Matérn covariance kernel [190]

$$C_\nu(h) = \sigma_H^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu}h}{\ell}\right)^\nu K_\nu\left(\frac{\sqrt{2\nu}h}{\ell}\right), \quad (2.24)$$

since we can define the special case $\nu = 1/2$ and limiting case $\nu \rightarrow \infty$:

$$C_{1/2}(h) = \sigma_H^2 \exp\left(-\frac{h}{\ell}\right) \quad \text{and} \quad C_\infty(h) = \sigma_H^2 \exp\left(-\frac{h^2}{2\ell^2}\right), \quad (2.25)$$

which correspond to the non-differentiable *exponential* and infinite-differentiable *squared exponential* (sometimes called Gaussian or radial basis function) covariance kernels, respectively. As discussed in subsection 2.2.2, we can verify mean-squared differentiability of the random field by evaluating the second derivative of (2.25)

$$\ddot{C}_{1/2}(h) = \sigma_H^2 \left[\frac{\exp(-|h|/\ell) \operatorname{sgn}^2(h)}{\ell^2} - \frac{2\delta(h)}{\ell} \right] \quad \text{and} \quad \ddot{C}_\infty(h) = \sigma_H^2 \frac{\exp(-h^2/(2\ell^2))(h^2 - \ell^2)}{\ell^4}, \quad (2.26)$$

at $h = 0$, where $\operatorname{sgn}(\cdot)$ and $\delta(\cdot)$ are the sign and Kronecker functions, respectively. In this case, we have that $\ddot{C}_{1/2}(0) = -\infty$ and $\ddot{C}_{1/2}(0) = -\sigma_H^2/\ell^2$. The random field realizations produced by an exponential covariance kernel are continuous but they are not mean-squared differentiable. A modification of the exponential kernel that fixes the mean-squared differentiability is discussed in [217]. Moreover, the squared exponential kernel produce ∞ -times mean-squared differentiable realizations as long as the field is of second-order (this can be verified in [1, p.24]).

The parameters of the Matérn model ℓ and ν can be fitted based on experimental measurements [105]. The value of ν determines the smoothness of the random field, and it is usually a fixed parameter since it is poorly identified in practical applications [150]; although in [109], a framework for its estimation is proposed. In general the recommendation is to choose a smooth kernel (as the squared exponential) for fitting smooth functions, while a non-differentiable kernel (as the exponential) is a better option for approximating continuous but non-differentiable functions.

2.5. Representing random fields with the Karhunen–Loève expansion

The definition of a random field implies the construction of a fi-di distribution family with $d \rightarrow \infty$. Since it is not feasible to collect sufficient data to verify the assumed probabilistic models, such a theoretical description is not used in practice. The process of representing a continuous-parameter random field in terms of a finite set of random variables requires the use of *stochastic discretization* algorithms [163]. We will briefly discuss a classification of such representation techniques and focus on the Karhunen–Loève expansion of random fields.

2.5.1. General remarks on random field discretization

The efficiency in the approximation of a continuous random field $H(\mathbf{x}, \omega)$ by a discretized $\hat{H}(\mathbf{x}, \omega)$ is quantified by its ability to accurately represent the original field with as few random variables as possible. Methods for random field representation can be classified into three main categories:

- (i) Point discretization methods, in which the random field is represented point-wise at each spatial location of the discretization; the standard algorithm is the midpoint method [65].
- (ii) Averaging discretization methods, in which the random field is expressed as weighted integrals over the domain. Some approaches are, the spatial averaging method [240], the shape function method [154] and the weighted integral method [64].
- (iii) Series expansion methods, where the random field is represented as a truncated series expansion of random variables and deterministic functions. Popular techniques are the perturbation method [239], the Karhunen–Loève expansion [216], the expansion optimal linear estimation [145], the orthogonal series expansion [254] and spectral representation [207].

In the following, we focus on the Karhunen–Loève expansion which is probably the most popular method for random field discretization in uncertainty quantification. One of the reasons for its popularity is that the Karhunen–Loève expansion is optimal in the mean-squared error sense as compared to any other spectral projection algorithm.

Remark 2.1. There exist alternative methods to generate Gaussian random field realizations directly. This is the case of circulant embedding [67], which employs fast Fourier transform to diagonalize a nested-block-circulant-matrix extension of a nested-block-Toeplitz covariance matrix. Optimization methods can also be utilized to sample discretized Gaussian random fields; in this case, sampling can be expressed as a randomized least squares problem [17]. Furthermore, the work [150] (based on [246]) shows that the solution of the stochastic PDE

$$\gamma \cdot (\kappa^2 - \Delta)H(\mathbf{x}) = W(\mathbf{x}) \quad \mathbf{x} \in D, \quad (2.27)$$

where W is a white noise process and Δ is the Laplacian operator, is a homogeneous Gaussian random field (under the Matérn kernel) whose correlation length and variance are controlled by the constants $\kappa > 0$ and $\gamma > 0$, respectively. Neumann boundary conditions are usually applied to solve (2.27) [150].

2.5.2. Karhunen–Loève expansion

The *Karhunen–Loève* (KL) expansion [155, 132] uses a linear combination of orthonormal functions chosen as the eigenfunctions resulting from the spectral decomposition of the covariance operator of the random field. The KL expansion is based on Mercer’s theorem [166], which provides a series representation of symmetric positive-definite functions. In this case, there exists

an orthonormal basis consisting of eigenfunctions $\{\phi_i\} : D \rightarrow L^2(D)$ of the covariance operator Σ together with a sequence of corresponding eigenvalues $\{\lambda_i\} \in [0, \infty)$. The covariance kernel $C_{HH}(\mathbf{x}, \mathbf{x}')$ has a representation of the form

$$C_{HH}(\mathbf{x}, \mathbf{x}') = \sum_{i=1}^{\infty} \lambda_i \phi_i(\mathbf{x}) \phi_i(\mathbf{x}'), \quad \forall \mathbf{x}, \mathbf{x}' \in D \quad (2.28)$$

where the eigenvalues are such that $\lambda_k \geq \lambda_{k+1}$ and $\lim_{k \rightarrow \infty} \lambda_k = 0$, and the eigenfunctions $\phi_k(\mathbf{x}) \in L^2(D)$. The series (2.28) is called a *Mercer expansion*.

Consider a square-integrable (second-order) random field $H(\mathbf{x}, \omega)$ defined over a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, indexed over D , and equipped with a continuous covariance function $C_{HH}(\mathbf{x}, \mathbf{x}')$. We can employ the orthonormal basis consisting of the eigenfunctions of the covariance operator together with the sequence of real and positive eigenvalues to represent a second-order random field $H(\mathbf{x}, \omega)$ as

$$H(\mathbf{x}, \omega) \approx \widehat{H}(\mathbf{x}; k, \boldsymbol{\theta}) := \mu_H(\mathbf{x}) + \sum_{i=1}^{\infty} \mathbf{1}(i \leq k) \sqrt{\lambda_i} \phi_i(\mathbf{x}) \theta_i(\omega), \quad (2.29)$$

where $\widehat{H}(\mathbf{x}; k, \boldsymbol{\theta})$ is the approximated field, $\mathbf{1}(\cdot)$ denotes the indicator function, k is the truncation order of the expansion, and $\boldsymbol{\theta} = \{\theta_i(\omega) : \Omega \rightarrow \mathbb{R}\}$ is a set of mutually uncorrelated random variables with mean zero and unit variance, that is, $\mathbb{E}[\theta_i] = 0$ and $\mathbb{E}[\theta_i \theta_j] = \delta_{ij}$, $\forall i, j \in \mathbb{N}$ (with δ the Kronecker function). The series (2.29) is called the *Karhunen–Loève expansion* of the random field. The KL expansion works as a dimension reduction technique, since instead of directly using a countable set of spatial points (as $d \rightarrow \infty$), we are now defining the uncertainty in terms of the KL coefficients $\boldsymbol{\theta} \subseteq \mathbb{R}^k$.

For Gaussian random fields, the variables $\theta_i(\omega)$ are independent and jointly Gaussian distributed. In the general case, the distribution of $\theta_i(\omega)$ is cumbersome to estimate since

$$\theta_i(\omega) = \frac{1}{\sqrt{\lambda_i}} \int_D H(\mathbf{x}, \omega) \phi_i(\mathbf{x}) d\mathbf{x} \quad i = 1, 2, \dots, \quad (2.30)$$

which requires the knowledge of random field structure for its computation. Non-Gaussian random fields can be represented with the KL expansion through the definition of different basis functions [115], or the application of a suitable isoprobabilistic transformation of an underlying Gaussian field (e.g., translation fields [61, 101]), as we briefly commented in section 2.3.

The set of eigenpairs $\{\lambda_i, \phi_i\}$ are obtained through the solution of a homogeneous Fredholm integral equation of the second kind [231, 90]

$$\int_D C_{HH}(\mathbf{x}, \mathbf{x}') \phi_i(\mathbf{x}') d\mathbf{x}' = \lambda_i \phi_i(\mathbf{x}), \quad (2.31)$$

whose analytical solution exist only for specific cases of covariance functions (some examples are provided in [90]). The main drawback of the KL expansion lies in the calculation of this integral eigenvalue problem. Most of the numerical techniques result in a considerable computational effort, since they often deal with the assembly of dense covariance matrices. For instance, classical approaches to (2.31) estimate the eigenfunctions as a linear combination of a set of complete basis functions, whose coefficients are obtained by minimizing the error arising from this representation. This is the idea behind projection methods such as collocation and Galerkin (see, e.g., [28]). Other techniques for computing (2.31) include, degenerate kernel methods [14], which approximate the target kernel by a separable one given by the sum of a finite number of products of functions; Nyström methods [186], which solve the integral equation using Gaussian quadrature rules; and wavelet-Galerkin techniques [185], which try to “sparsify”

the covariance matrix to improve efficiency. In general, one should be careful when solving (2.31) using a numerical scheme since the positive definiteness of the covariance matrix can be compromised. This is related to the estimation of the eigenvalues [186, Ch.19]. For instance, negative eigenvalues might occur when a very fine discretization of D is employed, especially when the random field is described by non-smooth covariance kernels or when it is weakly correlated.

Finally, note that the only requirement for the set $\{\phi_i(\mathbf{x})\}$ in (2.29) is to be complete and orthogonal. In principle, any other type of basis satisfying these conditions can be used in the expansion. However, the KL expansion is usually preferred over other alternatives due to the following properties [90]:

- (i) Error minimizing property: the KL approximation (2.29) minimizes the mean-squared error (MSE), such that [253]

$$\mathbb{E} \left[\left\| [H(\mathbf{x}, \omega) - \mu_H(\mathbf{x})] - [\hat{H}(\mathbf{x}; k, \boldsymbol{\theta}) - \mu_H(\mathbf{x})] \right\|_2^2 \right] = \sum_{i=k+1}^{\infty} \lambda_i^2. \quad (2.32)$$

- (ii) Uniqueness of the expansion: the random variables $\boldsymbol{\theta} = \{\theta_i(\omega) : \Omega \rightarrow \mathbb{R}\}$ in the expansion are orthonormal, if and only if the orthonormal functions $\{\phi_i\}$ and the constants $\{\lambda_i\}$ are the eigenpairs of the covariance matrix of the field [90].
- (iii) Spatial domain independence: the shape of the physical domain D of a random field does not alter the realizations computed by the KL expansion [145]. This means that the first- and second-order moments of a random field realization remain invariant to a change of the physical domain, despite the individual KL terms being affected. As a result, the original domain D can be replaced by a domain D' in which the numerical solution of (2.31) is easier and faster to obtain (offline cost). The process of changing the index set leads to a non-optimal representation, since the inner products and thus, the eigenvalues and eigenfunctions, are different for L^2 -functions on different spatial domains [133]. Even though the generated random fields will differ only by a small amount, more terms in the expansion will be required increasing the (online) computational cost.

2.6. Error measures for random field discretization

In the KL expansion, the number of terms to be included in the series is closely related to the magnitudes of the eigenvalues of the covariance operator, which in turn strongly depend on the correlation length and smoothness of the covariance function. Specifically, the quality of the discretization is quantified with respect to the level of accuracy in the estimation of the exact mean (bias) and variance (variability) functions of the random field.

One can define local point-wise error metrics for the mean and variance in terms of the relative difference between the exact and approximated random fields:

$$\epsilon_{\mu}(\mathbf{x}) = \left| \frac{\mathbb{E}[H(\mathbf{x}, \omega)] - \mathbb{E}[\hat{H}(\mathbf{x}; k, \boldsymbol{\theta})]}{\mathbb{E}[H(\mathbf{x}, \omega)]} \right| \quad \epsilon_{\sigma^2}(\mathbf{x}) = \left| \frac{\mathbb{V}[H(\mathbf{x}, \omega)] - \mathbb{V}[\hat{H}(\mathbf{x}; k, \boldsymbol{\theta})]}{\mathbb{V}[H(\mathbf{x}, \omega)]} \right|, \quad (2.33)$$

here ϵ_{μ} and ϵ_{σ^2} are the relative errors in the mean and variance, respectively.

Global error metrics can also be applied to quantify the overall quality of the random field representation. These are defined for the mean and variance, as their average values over the

index set D of the random field [28]

$$\bar{\epsilon}_\mu = \frac{1}{|D|} \int_D \epsilon_\mu(\mathbf{x}) d\mathbf{x} \quad \text{and} \quad \bar{\epsilon}_{\sigma^2} = \frac{1}{|D|} \int_D \epsilon_{\sigma^2}(\mathbf{x}) d\mathbf{x}, \quad (2.34)$$

where $|D| = \int_D d\mathbf{x}$.

For the KL expansion, the mean can be represented exactly and thus $\epsilon_\mu(\mathbf{x}) = 0$. However, there is an underlying variance error induced by the representation. We can compute the variance of the approximation as

$$\mathbb{V}[\hat{H}(\mathbf{x}; k, \boldsymbol{\theta})] = \sum_{i=1}^k \lambda_i \phi_i^2(\mathbf{x}) = \sum_{i=1}^k \lambda_i, \quad (2.35)$$

since the eigenfunctions are orthonormal. Under (2.35), the local variance error reduces to

$$\epsilon_{\sigma^2}(\mathbf{x}) = 1 - \frac{1}{\sigma_H^2} \sum_{i=1}^k \lambda_i; \quad (2.36)$$

this expression is commonly used to define the truncation order of the KL expansion.

2.7. Numerical applications

In this section, we analyze two engineering problems involving spatially varying properties. A parameter study on the Matérn class of covariance kernels and its influence on the random field representation is carried out. This also illustrates the KL expansion for the representation of random fields.

2.7.1. Example 1: one-dimensional cantilever beam

We consider the second example in [223] dealing with the updating of the spatially variable flexibility $F(x)$ of a cantilever beam. The beam has length $L = 5$ m (i.e., the domain is the interval $D = [0, L]$) and it is subjected to a deterministic point load $P = 20$ kN at the free end as shown in Figure 2.3.

The flexibility is described by a homogeneous Gaussian random field. The Matérn class (2.24) is considered as covariance function of the flexibility $C_{FF}(x, x')$. The mean of the field is $\mu_F = 1 \times 10^{-4}$ kN⁻¹m⁻² and the standard deviation is $\sigma_F = 0.35 \cdot \mu_F$ kN⁻¹m⁻². A parameter study on the correlation length ℓ and smoothing ν is performed.

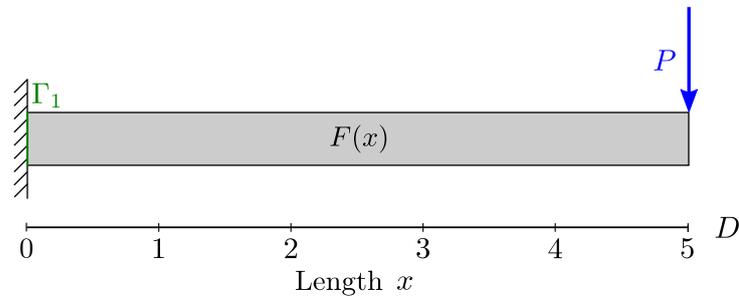


Figure 2.3: Cantilever beam configuration.

From Euler–Bernoulli theory, the bending moment in the beam $M(x)$ can be computed from the ordinary differential equation (ODE) [41]:

$$M(x) = -E(x)I \frac{d^2 w(x)}{dx^2} \implies -M(x)F(x) = \frac{d^2 w(x)}{dx^2}, \quad (2.37)$$

where $w(x)$ is the deflection, $E(x)$ is the elastic modulus, I is the moment of inertia, and $F(x) = (E(x)I)^{-1}$ is the flexibility of the beam (the inverse of the bending stiffness). Integrating twice (2.37) and noting that the bending moment of a cantilever beam can be calculated explicitly as $M(x) = (L - x)P$, the forward deflection response of the beam is obtained by solving the equation

$$w(x) = -P \int_0^x \int_0^s (L - t)F(t) dt ds. \quad (2.38)$$

We can express the forward operator in (2.38) in terms of the KL expansion of the flexibility field as

$$\hat{w}(x; k, \boldsymbol{\theta}) = -P \int_0^x \int_0^s (L - t) \left[\mu_F(t) + \sum_{i=1}^k \sqrt{\lambda_i} \phi_i(t) \theta_i \right] dt ds \quad (2.39a)$$

$$= -P \int_0^x \int_0^s (L - t) \mu_F(t) dt ds - P \int_0^x \int_0^s (L - t) \sum_{i=1}^k \sqrt{\lambda_i} \phi_i(t) \theta_i dt ds \quad (2.39b)$$

$$= \mu_w(x) - \sum_{i=1}^k \Phi_i(x) \sqrt{\lambda_i} \theta_i, \quad \text{where } \Phi_i(x) = P \int_0^x \int_0^s (L - t) \phi_i(t) dt ds. \quad (2.39c)$$

For a given domain discretization $\mathbf{x} = [0, \dots, x_j, \dots, x_d]$, (2.39c) can also be written in matrix form as

$$\hat{\mathbf{w}} = \boldsymbol{\mu}_w - \boldsymbol{\Phi} \boldsymbol{\Lambda} \boldsymbol{\theta} = \boldsymbol{\mu}_w - \mathbf{A} \boldsymbol{\theta}, \quad (2.40)$$

where $\boldsymbol{\mu}_w \in \mathbb{R}^d$ is the prior mean deflection vector computed from (2.41), and $\boldsymbol{\Phi} \boldsymbol{\Lambda} = \mathbf{A} \in \mathbb{R}^{d \times k}$. Here $\boldsymbol{\Phi} \in \mathbb{R}^{d \times k}$ is a matrix obtained by evaluating $\Phi_i(x)$ in (2.39c) (that is, $\Phi^{(j,i)} = \Phi_i(x_j)$, for $j = 1, \dots, d$), and $\boldsymbol{\Lambda} = \text{diag}(\sqrt{\boldsymbol{\lambda}}) \in \mathbb{R}^{k \times k}$ is a diagonal matrix with the square root of the eigenvalues.

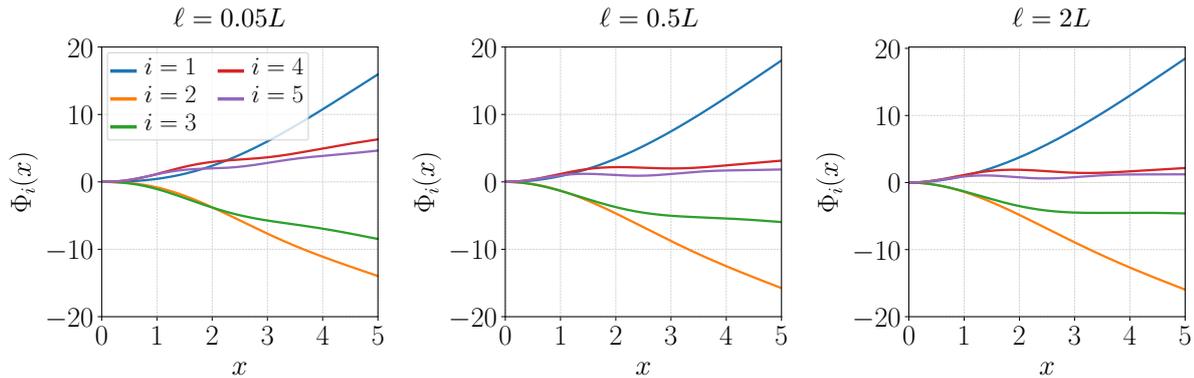


Figure 2.4: Function $\Phi_i(x)$ in (2.39c) using an exponential kernel, for different correlation lengths and first 5 terms in the KL expansion.

Observe that $\Phi_i(x)$ can be evaluated analytically given that the eigenpairs of the target covariance kernel are available. Figure 2.4 plots the values of this function for the exponential covariance kernel. The smoothing effect of the ODE is encoded in $\Phi_i(x)$ since it is integrating the eigenfunctions over the domain. This effect can be explained by the Riemann–Lebesgue lemma [107, p.8] since higher eigenfunctions (frequencies) are damped by the application of $\boldsymbol{\Phi}$ (cf., Figure 2.4 for $i = \{4, 5\}$), and as a result w will be smoother than F .

The mean and covariance functions of the deflection random field can be evaluated using the

prior information about the flexibility F and the forward ODE operator. Since F is a Gaussian random field and $w(x, F)$ is a linear function of F , the prior distribution of the deflection is also a Gaussian field. Therefore, an expression for the mean function of w can be obtained using μ_F in (2.38) as

$$\mu_w(x) = -P \int_0^x \int_0^s (L-t) \mu_F(t) dt ds = -\frac{P\mu_F}{6} x^2 (3L-x) \quad (2.41)$$

and similarly, the covariance function of w can be deduced for a given $C_{FF}(x, x')$ as

$$C_{ww}(x, x') = P^2 \int_0^{x'} \int_0^x \int_0^{s'} \int_0^s (L-t)(L-t') C_{FF}(t, t') dt dt' ds ds'. \quad (2.42)$$

The computation of (2.42) involves multiple integration of the flexibility covariance function. This task can be performed analytically or using a symbolic integration toolbox. However, we provide integration formulas in order to facilitate the numerical implementation. The covariance (2.42) can be estimated using a n_{GL} -point Gauss–Legendre (GL) quadrature as

$$C_{ww}(x, x') \approx P^2 \sigma_F^2 \left[\frac{xx'}{4} \sum_{k=1}^{n_{\text{GL}}} \sum_{l=1}^{n_{\text{GL}}} w_k w_l \frac{xx'(\xi_l + 1)(\xi_k + 1)}{16} \right. \\ \left. \sum_{j=1}^{n_{\text{GL}}} \sum_{i=1}^{n_{\text{GL}}} w_j w_i f_1 \left(\frac{x(\xi_l + 1)(\xi_i + 1)}{4}, \frac{x'(\xi_k + 1)(\xi_j + 1)}{4} \right) \right] \quad \forall x, x' \in D,$$

where $f_1(t, t') = (L-t)(L-t')R_{FF}(t, t')$, $R_{FF}(t, t')$ is the correlation function of the flexibility, w_k, w_l, w_j, w_i are the weights of the quadrature, and the points $\xi_k, \xi_l, \xi_j, \xi_i$ are the associated zeros of the Legendre polynomial.

The Matérn kernel class (2.24) is utilized as covariance function of the flexibility random field. This defines indirectly the covariance of the deflection field since it is propagated via (2.42). We now perform a parameter study on the correlation length and smoothness defining the covariance of the flexibility field (which in turn affect the deflection field). The smoothing parameters are chosen as $\nu = \{0.5, 2, \infty\}$ and we select the correlation lengths as a percentage of the domain length $\ell = \{0.05L, 0.5L, 2L\}$.

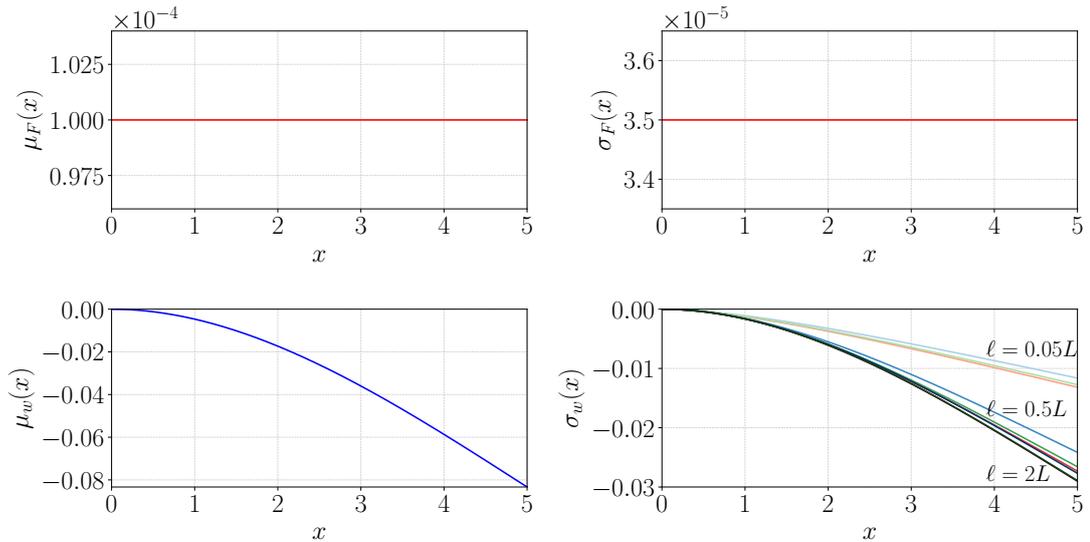


Figure 2.5: Mean and standard deviation of the flexibility (1st row) and deflection (2nd row) random fields. Different standard deviation functions are obtained for the deflection depending on the correlation length and smoothing parameters (bottom right panel, $\nu = 0.5$: blue, $\nu = 2$: green, and $\nu = \infty$: red) of the Matérn kernel used for the flexibility.

The mean and standard deviation functions of the flexibility and deflection fields are shown in Figure 2.5. For the flexibility both functions are constant since the field is homogeneous. The deflection field is non-homogeneous as a result of applying the ODE operator to the flexibility field. The mean deflection is independent of the choice of the flexibility covariance and it is computed by (2.41). On the contrary, the standard deviations change with the choice of the Matérn kernel for the flexibility field, and one sees that as the correlation increases, the standard deviation also increases. This is also true for increasing values of the smoothing parameter, although its effect on the standard deviation is less significant compared to the correlation length (Figure 2.5, zoom-in bottom right panel, $\nu = 0.5$: blue, $\nu = 2$: green, and $\nu = \infty$: red).

The correlation functions of the prior flexibility random field are shown in Figure 2.6. As we decrease the correlation length, the resulting field becomes weakly correlated. In the limit $\ell \rightarrow 0$, the covariance matrix collapses to the diagonal and the spatial dependency vanishes (the resulting field can be seen as a white noise process). Conversely, as the correlation length increases, the field become strongly correlated; in the limit $\ell \rightarrow \infty$, the covariance values are all equal to the variance and the resulting field becomes a single random variable. Note that the effect of the correlation length is stronger than the action of the smoothing parameter.

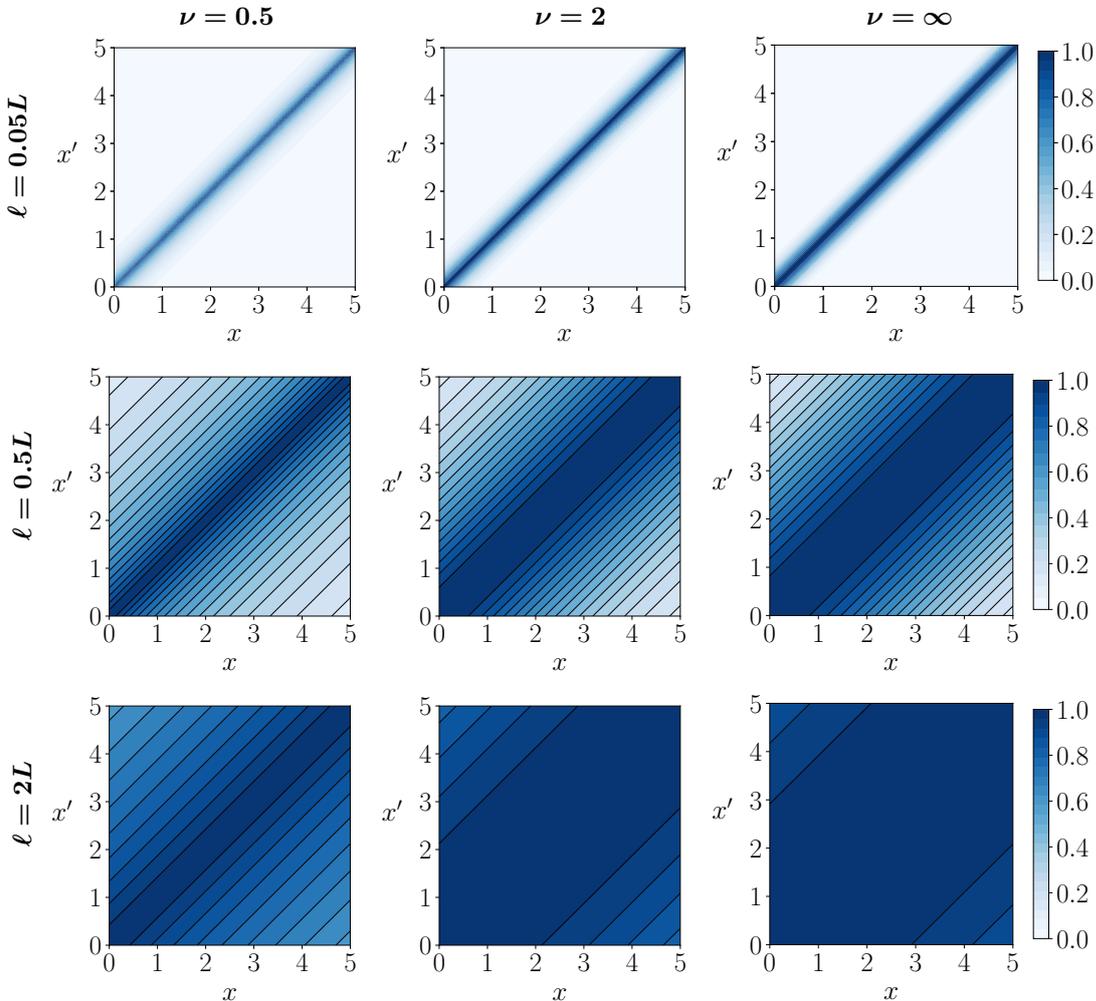


Figure 2.6: Correlation matrix of the flexibility random field associated to the Matérn kernel (2.24) for different correlation lengths (rows) and smoothing parameters (cols).

We show an analogous plot for the correlation functions for the deflection field in Figure 2.7. The correlation functions derive from the functional from induced by the Matérn kernel in Figure 2.6, due to the application of the forward operator (2.38). The associated covariance

operators are still well-defined. Despite the influence of the correlation length is similar to the flexibility field, there is an extra smoothing effect applied by the forward operator. As a result, the corresponding deflection covariances converge faster to the fully-correlated (random variable) case. The filtering/averaging influence of the forward response operator is usually neglected in the engineering practice. However, this effect is important when defining the QoIs that are relevant for the analysis of the structural system.

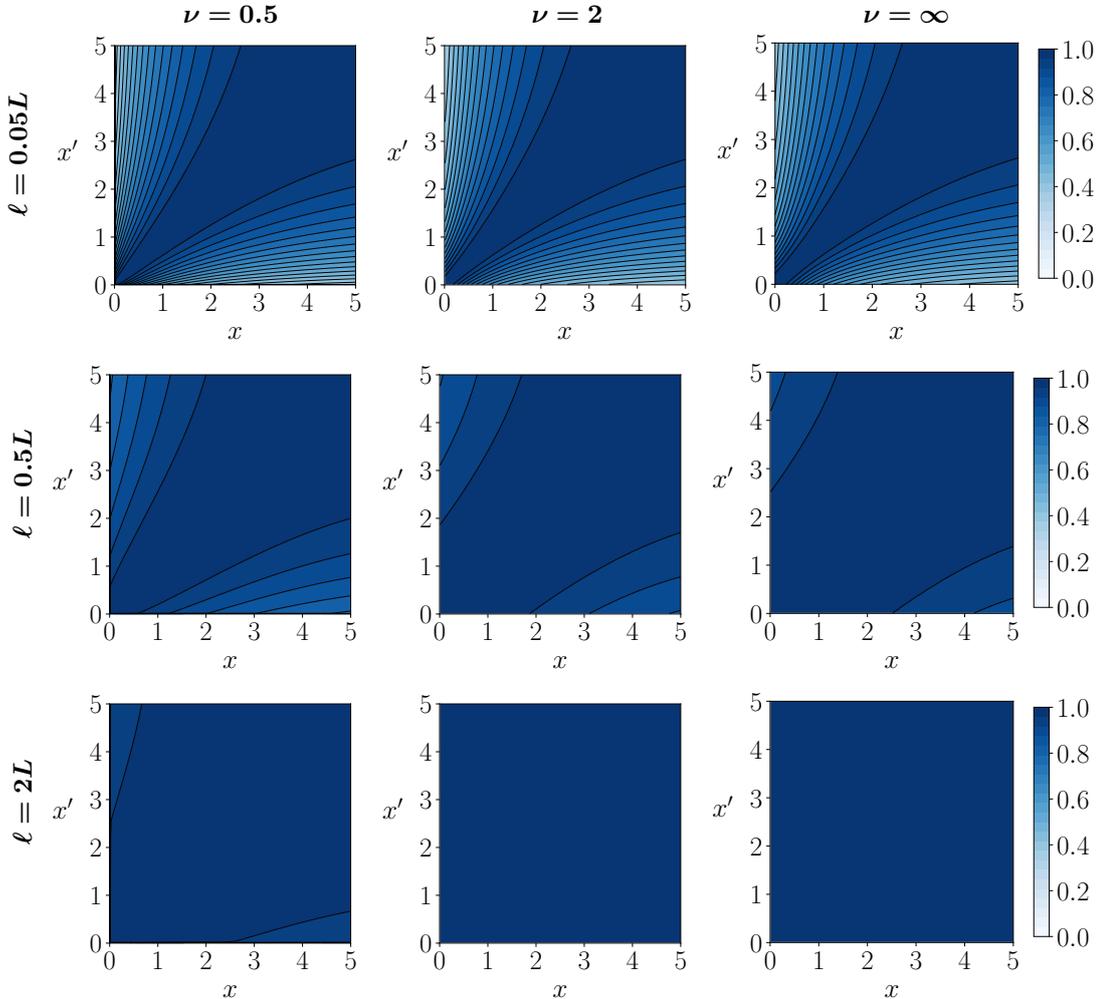


Figure 2.7: Correlation matrix of the deflection random field associated to the Matérn kernel by application of (2.42) for different correlation lengths (rows) and smoothing parameters (cols).

We now assess the error in the KL representation of the flexibility random field for the different parameter settings in the Matérn kernel. The Nyström method [186] is used to obtain the numerical solutions to the KL eigenvalue problem. The number of GL quadrature points in Nyström is selected from $n_{GL} = \{50, 100, 250, 500, 1000\}$. The reference global variance error is computed from the analytical solution of the KL eigenvalue problem when $\nu = 0.5$; for the other smoothing parameters the reference is computed with a Nyström method with $n_{GL} = 2000$.

Figure 2.8 shows the evolution of the global variance error (2.36) in the flexibility approximated by the KL expansion. This is shown for different parameters in the Matérn kernel and number of GL quadrature points in the Nyström method. The error decays faster when the random fields are smooth (increasing value of ν). In this case, Nyström also estimates the target error decay only with a few number of quadrature points. The behavior is completely different when the random fields are rougher (small values of ν), in those situations the Nyström method requires a higher number of points for an accurate estimation of the eigenvalue problem.

A similar study can be performed for the deflection random field. In this case, the rate of decay of the variance errors is faster in all parameter cases, due to the smoothing action of the forward operator. This can be intuitively explained by the shapes of the correlation functions in Figures 2.6 and 2.7. Hence, smaller truncation orders in the KL expansion are required to represent smooth QoIs, as in the case of those computed by application of the smoothing forward ODE equation.

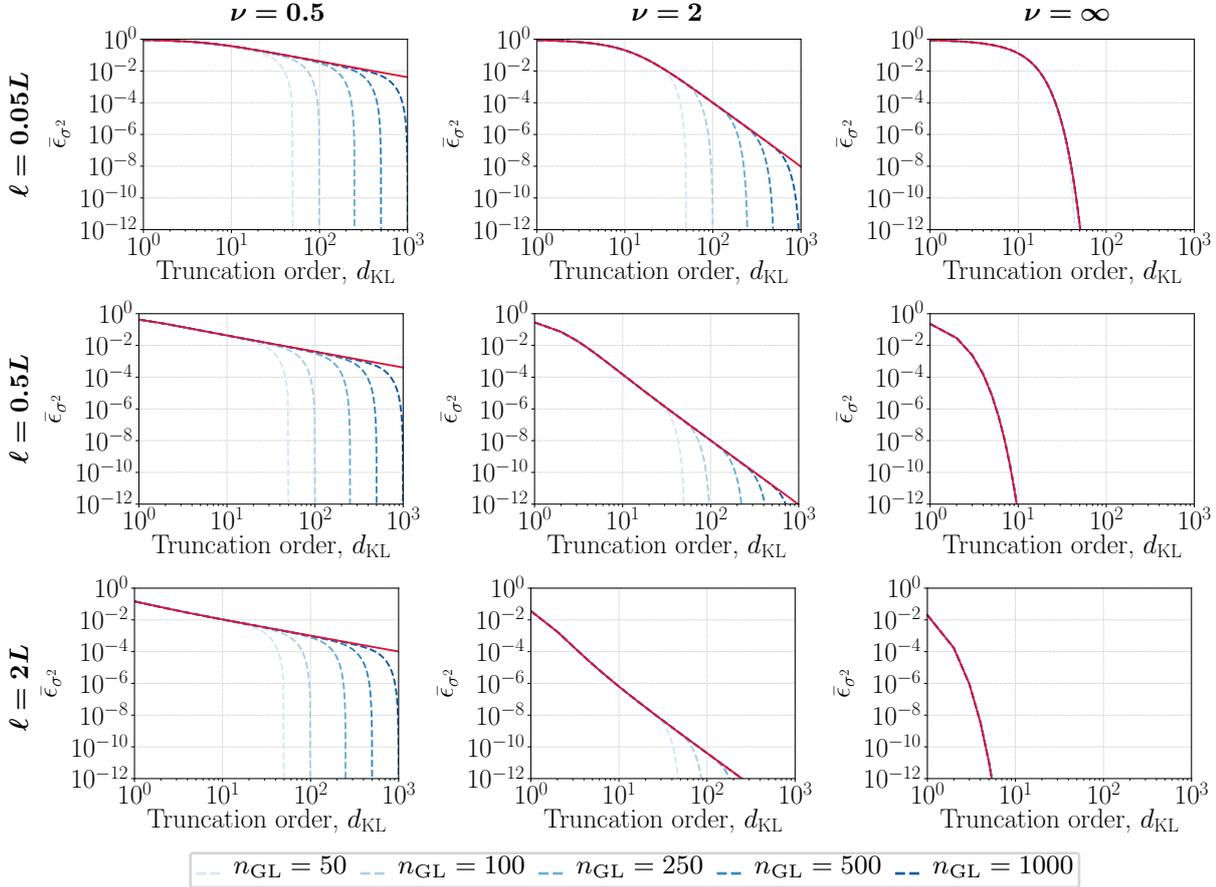


Figure 2.8: Evolution of the global variance error (2.36) for the flexibility random field with the KL terms $d_{\text{KL}} = k$: influence of the number of GL points on the solution of the KL eigenvalue problem for different correlation lengths and smoothing parameters of the Matérn kernel. The reference global variance error is shown in red.

We conclude the study on the KL expansion by showing some random field realizations for all the investigated parameter cases. The truncation order for each case is chosen as the number of KL terms that accounts for 95% of the flexibility random field variability (that is, by selecting the truncation that yields 5% reference error in Figure 2.8). These values are shown in Table 2.1 for the flexibility and deflection random fields.

Table 2.1.: Truncation order of the KL expansion k that complies with 95% of the random field variability. For the flexibility and deflection fields, shown as pairs (\cdot, \cdot) for each case.

		ν		
		0.5	2	∞
ℓ	0.05L	(81,10)	(18,4)	(12,3)
	0.5L	(9,3)	(2,1)	(1,1)
	2L	(3,2)	(1,1)	(1,1)

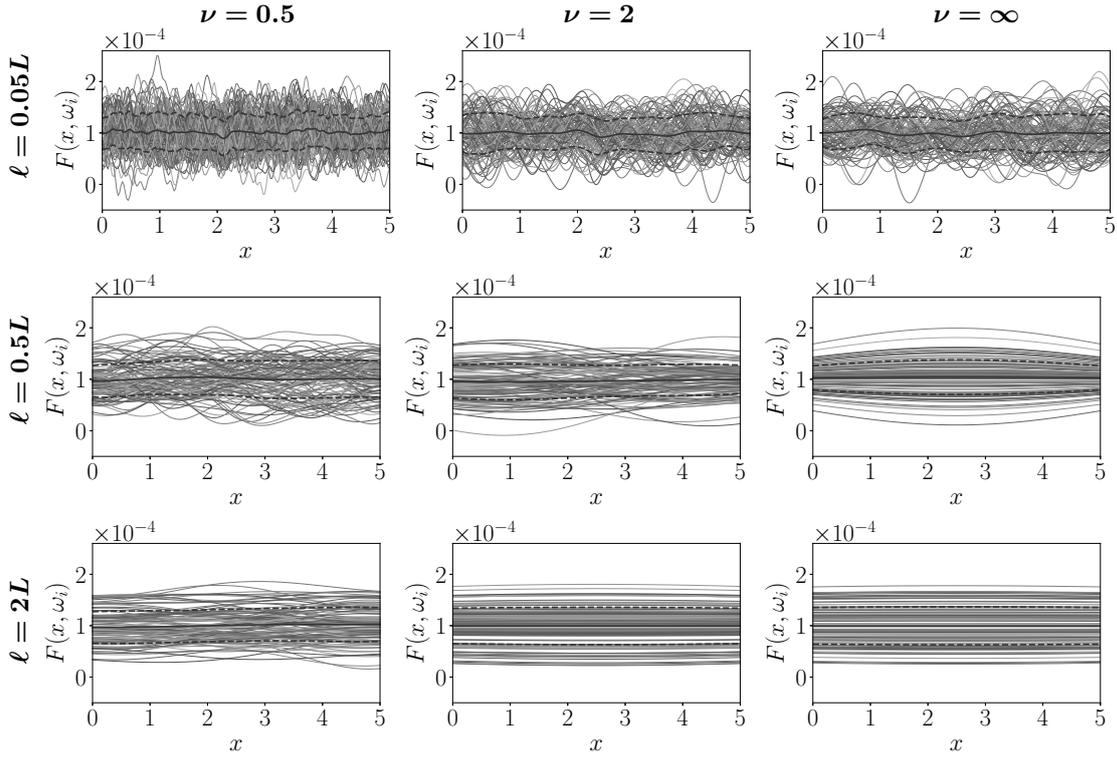


Figure 2.9: Flexibility random field realizations for different correlation lengths and smoothing parameters of the Matérn kernel modeling the flexibility covariance.

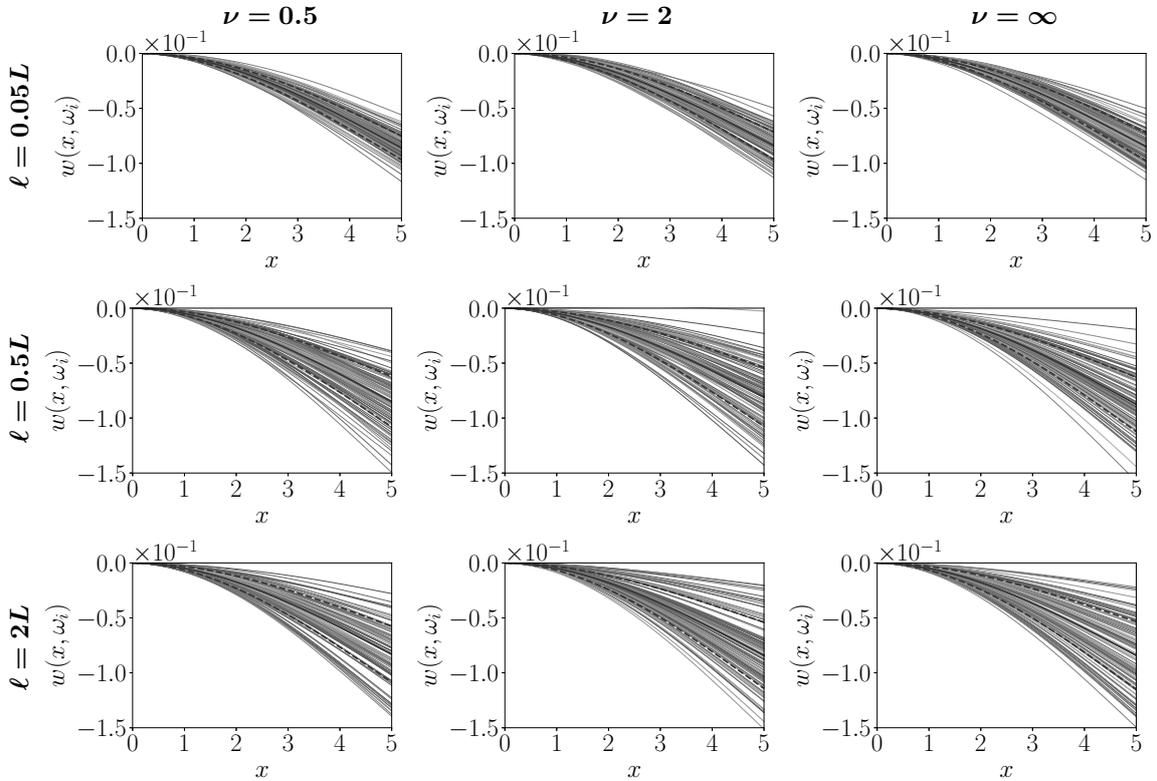


Figure 2.10: Deflection random field realizations for different correlation lengths and smoothing parameters of the Matérn kernel modeling the flexibility covariance.

As before, finer KL discretizations are required for the flexibility field as compared to the deflection. Figures 2.9 and 2.10 show 100 different realizations of the random fields for all the parameter cases. The mean and standard deviation bounds are also plotted. For the smooth flexibility random fields, the realizations are basically constant through the domain (random variable case). For the rough flexibility fields, the sample paths change at each spatial point in the domain (white noise case). Notice also that the smoother flexibility random fields yield deflection responses with larger variability, as seen in Figure 2.5.

2.7.2. Example 2: two-dimensional plate with a hole

We consider a steel plate model defined by a square domain $D \subseteq \mathbb{R}^2$ with length 0.32 m, thickness $t = 0.01$ m, and a hole of radius 0.02 m located at its center. Spatial coordinates are denoted by $\mathbf{x} = [x_1, x_2] \in D$. The displacement ($\mathbf{u}(x_1, x_2) = [u_{x_1}, u_{x_2}]^T$) field is computed using elasticity theory through a set of elliptic PDEs (see Cauchy–Navier equations in (B.4)) [129]. Due to the geometry of the plate, the PDEs can be simplified under the plane stress hypothesis as

$$G(\mathbf{x})\nabla^2\mathbf{u}(\mathbf{x}) + \frac{E(\mathbf{x})}{2(1-\nu)}\nabla(\nabla \cdot \mathbf{u}(\mathbf{x})) + \mathbf{b} = 0, \quad (2.43)$$

where $G(\mathbf{x}) := E(\mathbf{x})/(2(1+\nu))$ is the shear modulus, $\nu = 0.29$ is the Poisson ratio of the steel, and $\mathbf{b} = [0, -\rho g]^T$ is the vector of body forces acting on the plate, with g denoting the standard gravity and $\rho = 7850$ kg/m³ the density of the steel. We refer to Appendix B for further details on the model (2.43).

A Dirichlet boundary condition is imposed at the left edge of the plate, $\mathbf{u}(\mathbf{x}) = \mathbf{0}$ for $\mathbf{x} \in \Gamma_1$. Moreover, a random surface load q is applied at the right boundary Γ_2 . This action is modeled as a Gaussian random variable with mean $\mu_q = 60$ MPa and standard deviation $\sigma_q = 12$ MPa. Equation (2.43) is solved with the finite element method using 282 eight-node serendipity quadrilateral elements [175], as shown in Figure 2.11. The mesh is generated in the finite element mesh generator tool **Gmsh** [86].

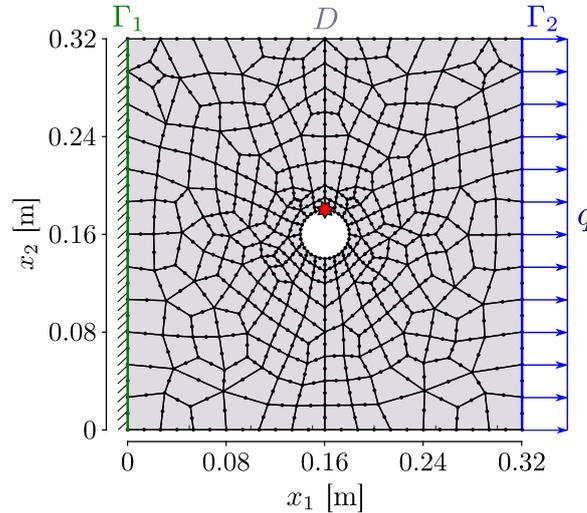


Figure 2.11: Plate with a hole. The red star marks a control point.

The Young's modulus $E(\mathbf{x})$ is assumed to be random and spatially variable. A lognormal random field with mean value $\mu_E = 2 \times 10^5$ MPa and standard deviation $\sigma_E = 3 \times 10^4$ MPa is used for its representation. We select the Matérn kernel (2.24) to model the covariance function of the underlying Gaussian field. The smoothing parameters are chosen as $\nu = \{0.5, 2, \infty\}$ and we select the correlation lengths as $\ell = \{0.02, 0.16, 0.32\}$ m. The truncated KL expansion (2.29) of the Young's modulus can be written as

$$\widehat{E}(\mathbf{x}; \boldsymbol{\theta}) := \exp \left[\mu_{E'} + \sum_{i=1}^k \sqrt{\lambda_i} \phi_i(\mathbf{x}) \theta_i \right], \quad (2.44)$$

where the underlying Gaussian parameters $\mu_{E'} = 26.011$ and $\sigma_{E'} = 0.149$, are computed from the mean and standard deviation of the lognormal random field. The eigenpairs are estimated with the Nyström method using 110 GL points in each direction. Realizations of the Young's modulus at the Gauss points of each element are evaluated using Nyström interpolation formula. We remark that for the computations we employ a different unit system to avoid possible numerical overflow (cm for length and ton/cm² for stiffness and stress). The following results are shown in this unit system.

Figure 2.12 plots the eigenvalues of the covariance operator for different values of the smoothing parameters and correlation lengths. The eigenvalue decay is faster for smooth covariance functions. This is analogous to the beam example where the eigenvalue decay can be seen in terms of the global variance error decay (cf. Figure 2.8). Some corresponding eigenfunctions are shown in Figure 2.13. Note that the hole modifies structure, especially for the first eigenfunction where there is a decrease in the function values at the vicinity of the hole.

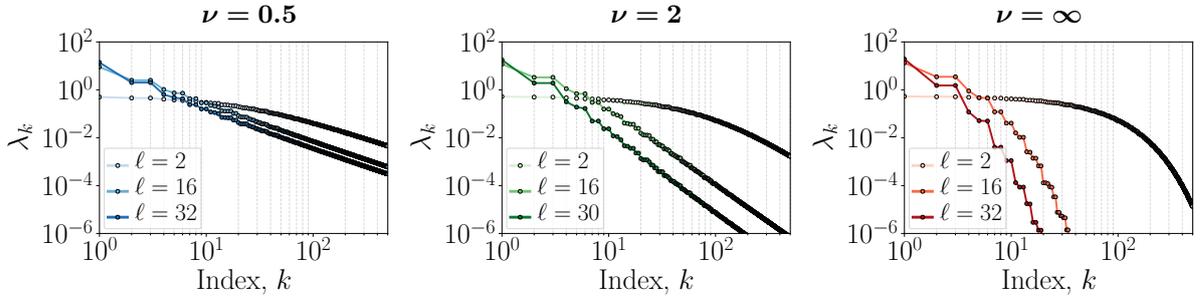


Figure 2.12: Plate example: evolution of the eigenvalues of the Matérn covariance operator with the truncation order of the KL expansion, for different smoothing parameters and correlation lengths.

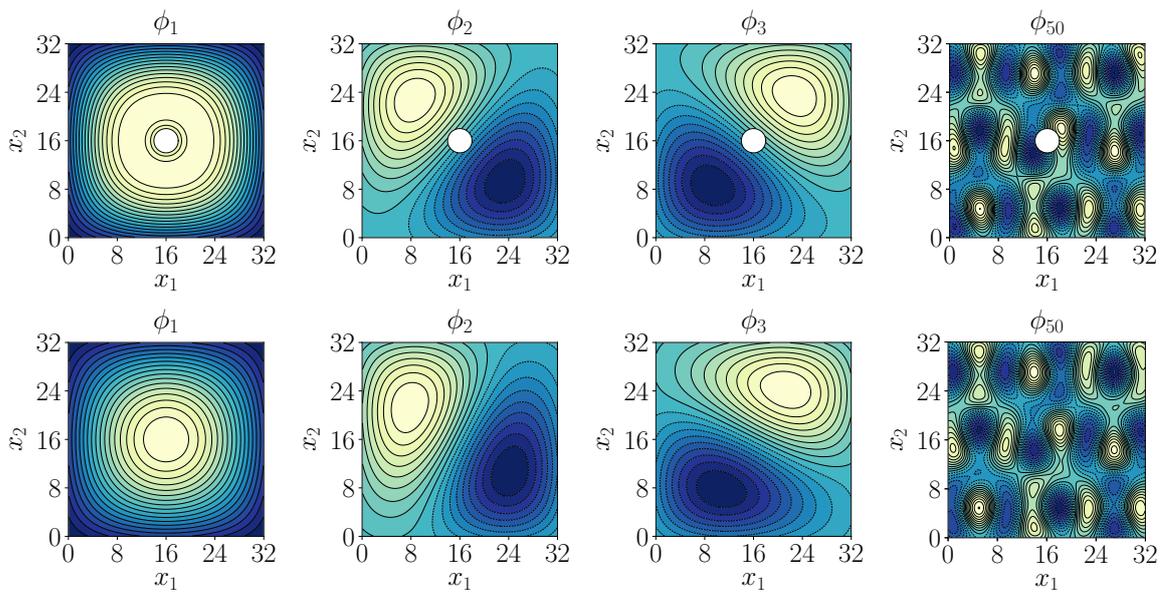


Figure 2.13: Plate example: selected eigenfunctions of the Matérn operator for $\nu = 0.5$ and $l = 2$. As a comparison, we also illustrate the eigenfunctions when there is no hole in the plate.

In the beam example, it is possible to derive the covariance of the forward quantity analytically, and thus, the truncation orders for the response field can be derived (Table 2.1). In this example such computation is cumbersome to perform. A Monte Carlo approach is applied to obtain samples (realizations) from the response field and estimate the local variance error from those samples. This requires fixing the truncation order of the KL expansion to a large value to define a reference variance. Thereafter, one can apply (2.33) for different truncation orders to obtain the corresponding evolution of the global variance error (2.34). We carry out this approach for the parameters cases $\ell = 2$ and $\nu = \{2, \infty\}$. The reference variances of the random fields are generated with a truncation order of $\{2300, 315\}$ for $\nu = \{2, \infty\}$, respectively. These values correspond to a 99.999% of the Young's modulus field variability. Moreover, we evaluate the global error not only for the Young's modulus, but also for the horizontal displacement u_{x_1} and the principal stress σ_1 fields (defined in (B.19a)). This is because we have seen that the spatial variation has different effects depending on whether the QoI is averaged by the forward operator.

Figure 2.14 plots the evolution of the global variance error for different truncation orders in the KL expansion. The error can be computed exactly for the Young's modulus field based on (2.36). For the σ_1 stress and displacement fields, we employ 5000 load samples and field realizations to approximate the variance. In this case, the computation is performed at fixed truncation orders $k = \{10, 25, 50, 75, 100, 150, 200, 300\}$. The error decays faster for the displacement field, since integration of the PDE operator is required for its computation. Such an averaging process reduces the effect of the local fluctuations of the spatially varying material parameter. The decay is less pronounced for the σ_1 stress, since this QoI is computed from the stress field which in turn is linearly dependent on the Young's modulus (see Hooke's law in Appendix B.2.3). Thus, the local random variations of the Young's modulus have a larger influence on the σ_1 stress. As discussed in the beam example, the smoothing effect is important when defining the QoIs that are relevant for the analysis. For instance, if the system performance depends on the displacements instead of the stresses, one can reduce the computational cost in the random field approximation by employing a smaller truncation order. This also shows that the smoothing effect of the forward operator typically induces a low-dimensional structure on the input parameter space. This fact will be exploited to perform dimension reduction in rare event simulation problems (cf. Chapter 6).

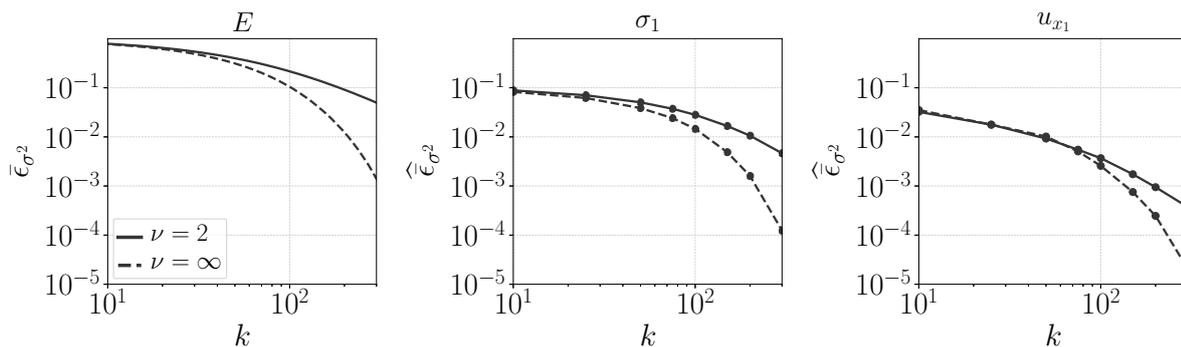


Figure 2.14: Plate example: evolution of the global variance error (2.36) with the number of terms in the KL expansion k . For the Young's modulus, σ_1 stress, and horizontal displacement fields. Using a Matérn kernel with $\ell = 2$ and $\nu = \{2, \infty\}$.

We conclude the random field representation studies by computing the mean and standard deviation functions of the Young's modulus and displacement fields. The smoothing parameter is fixed to $\nu = 0.5$ (exponential covariance) and we plot the function values for the investigated correlation lengths. The truncation order for each parameter case is chosen as the number of KL terms that accounts for 95% of the Young's modulus random field variability. These values

are shown in Table 2.2. As a result, we employ truncation orders $k = \{3561, 135, 36\}$ for each correlation length.

Table 2.2.: Truncation order of the KL expansion k that complies with 95% of the Young's modulus field variability.

		ν			
		0.5		2	∞
		2	3561	300	135
ℓ	16	135	7	4	
	32	36	2	2	

Figure 2.15 shows the mean and standard deviation functions. The values are estimated from 1000 independent load samples and field realizations. The mean and standard deviations of the Young's modulus are constant since the field is homogeneous (2039.43 and 305.91 in ton/cm² units, respectively). Nevertheless, the influence of correlation length is revealed by the patterns of the spatial variation, varying from a rough to smooth behavior, as ℓ increases. Similar to the beam example, the mean of the response quantity remains almost the same for all parameter cases, and the variability tends to increase as the random field becomes smoother (note the maximum values of the standard deviation for the displacement field).

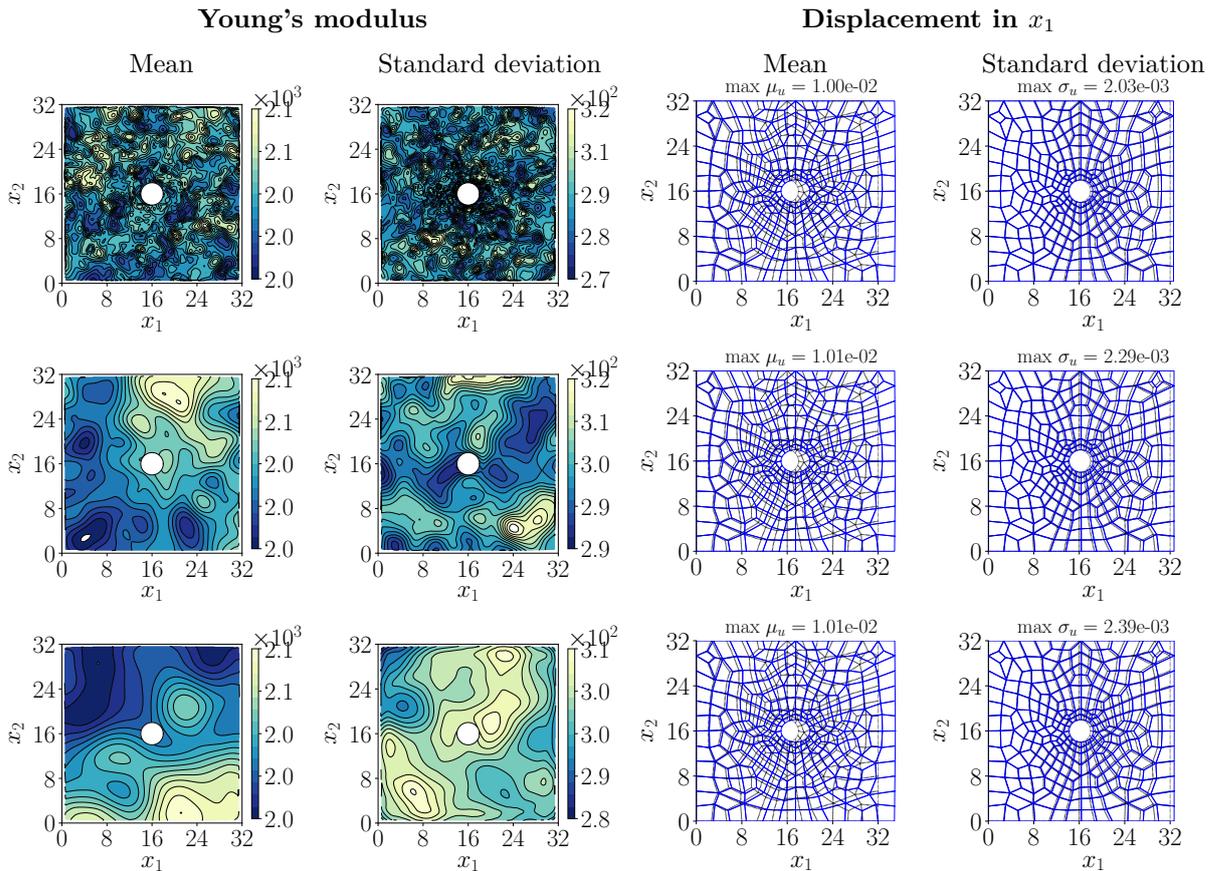


Figure 2.15: Plate example: mean and standard deviation fields estimated from 1000 realizations. Using a Matérn covariance with $\nu = 0.5$ and $\ell = 2$ (1st row), $\ell = 16$ (2nd row), $\ell = 32$ (3rd row).

Reliability analysis and rare event simulation

“If uncertainty arises solely from the random nature of the phenomena involved, it can be dealt rationally by probabilistic reasoning; any conceivable condition of a structure is associated with a numerical measure of the probability of its occurrence”.

Alfred M. Freudenthal, 1906-1977 [82, p.1].

The treatment and modeling of uncertainty is a fundamental task for the analysis and design of physical systems. This is particularly relevant when ensuring satisfactory operation of the system during its lifetime is the main concern. In such cases, the system performance is evaluated in terms of the probability of system failure. The performance of the system is described by a limit-state function specifying whether the system has failed or not. This is based on comparing a critical predefined threshold with the system response generated by application of a forward response operator (e.g., a PDE) on the uncertain input parameters.

A special challenge in structural reliability and rare event simulation involves the analysis of extreme or rare events for which the associated probabilities are very small [45]. In this chapter, we discuss advanced simulation methods for the estimation of failure probabilities arising in engineering applications with random field parameters.

3.1. The rare event simulation problem

Consider the canonical probability space $(\Omega, \mathcal{F}, \mathbb{P}) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mathbb{P})$, with $\mathcal{B}(\mathbb{R}^d)$ denoting Borel sets on \mathbb{R}^d and \mathbb{P} a probability measure [12]. The components of the uncertain parameter vector are modeled as random variables, such that $\boldsymbol{\theta} = \boldsymbol{\theta}(\omega) = \omega$ (identity map, cf. Appendix A), with $\omega \in \Omega$ and $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^d$. The distribution of $\boldsymbol{\theta}$ is assumed to have a density $\pi_{\text{pr}}(\boldsymbol{\theta}) = d\mathbb{P}/d\lambda$ with respect to the Lebesgue measure λ on \mathbb{R}^d ; we call this the *prior* or nominal probability density.

Different modes of failure can be grouped in a so-called *limit-state function* (LSF) $g : \Theta \rightarrow \mathbb{R}$, usually defined as $g(\boldsymbol{\theta}) = \beta - \mathcal{Q}(\boldsymbol{\theta})$. Here, β is a predefined maximum allowed threshold, and $\mathcal{Q}(\boldsymbol{\theta}) : \Theta \rightarrow \mathbb{R}$ is a forward response operator that maps the parameter $\boldsymbol{\theta}$ to a quantity of interest (QoI) describing the performance of the system. The failure hypersurface defined by $g(\boldsymbol{\theta}) = 0$ splits the parameter space into the *safe set* $\mathcal{S} = \{\boldsymbol{\theta} \in \Theta : g(\boldsymbol{\theta}) > 0\}$ and the *failure set* $\mathcal{F} = \{\boldsymbol{\theta} \in \Theta : g(\boldsymbol{\theta}) \leq 0\}$. The probability of \mathcal{F} under the prior distribution, also known as the *probability of failure* $p_{\mathcal{F}}$, is defined as

$$p_{\mathcal{F}} = \mathbb{P}[\mathcal{F}] = \int_{\Theta} \mathbb{1}_{\mathcal{F}}(\boldsymbol{\theta}) \pi_{\text{pr}}(\boldsymbol{\theta}) \, d\boldsymbol{\theta} = \mathbb{E}_{\pi_{\text{pr}}}[\mathbb{1}_{\mathcal{F}}(\boldsymbol{\theta})], \quad (3.1)$$

where $\mathbb{1}_{\mathcal{F}} : \mathbb{R}^d \rightarrow \{0, 1\}$ stands for the indicator function, taking values $\mathbb{1}_{\mathcal{F}}(\boldsymbol{\theta}) = 1$ when $\boldsymbol{\theta} \in \mathcal{F}$, and $\mathbb{1}_{\mathcal{F}}(\boldsymbol{\theta}) = 0$ otherwise. In rare event simulation, the probability of failure (3.1) represents the solution of a potentially high-dimensional integral with a very small value (typically in the

range $10^{-3} - 10^{-10}$).

From (3.1), one sees that the probability of failure is obtained by integrating the product of the indicator function and the prior density over the input parameter space Θ . As it will be seen in Chapter 4, this is analogous to the model evidence Z in the context of Bayes' theorem (4.2). The constant Z is computed similarly by integrating the product of a likelihood function and the prior density. We can exploit this connection to formulate the rare event simulation problem as a Bayesian inference task:

$$\pi_{\mathcal{F}}(\boldsymbol{\theta}) = \frac{1}{p_{\mathcal{F}}} \mathbb{1}_{\mathcal{F}}(\boldsymbol{\theta}) \pi_{\text{pr}}(\boldsymbol{\theta}), \quad (3.2)$$

where the indicator function acts as a likelihood function, and $\pi_{\mathcal{F}}$ can be interpreted as a “posterior-failure” density of the parameters given the failure event \mathcal{F} . Note that $\pi_{\mathcal{F}}$ is not a posterior density in the Bayesian inference sense, since there is no observational data entering into the likelihood function (in [241] this density is referred to as “fictitious” posterior). In this setting, $\pi_{\mathcal{F}}$ can be seen as a density conditional on the failure domain with its normalizing constant equal to the target probability of failure (see also [198, p.259]). We remark that the formulation (3.2) is valid, if the integral (3.1) is finite and $\mathbb{1}_{\mathcal{F}}(\boldsymbol{\theta})$ is measurable, which is true if and only if the failure set \mathcal{F} is measurable (see, e.g., [201]).

Remark 3.1. Classical approaches for estimating (3.1) operate in the standard Gaussian space; meaning that \mathbb{P} is assumed to be a standard Gaussian measure on \mathbb{R}^d . Several isoprobabilistic transformations, such as the Knothe–Rosenblatt and Nataf constructions exist to perform this “whitening” task (see, e.g., [144, Ch.4]). Therefore, we assume that the uncertain parameters are distributed as $\boldsymbol{\theta} \sim \pi_{\text{pr}} = \mathcal{N}(\boldsymbol{\mu}_{\text{pr}}, \boldsymbol{\Sigma}_{\text{pr}})$, with $\boldsymbol{\mu}_{\text{pr}} = \mathbf{0}$ and $\boldsymbol{\Sigma}_{\text{pr}} = \mathbf{I}_d$, where $\mathbf{I}_d \in \mathbb{R}^{d \times d}$ denotes the identity matrix.

Remark 3.2. Several smooth approximations of the indicator function $f(\boldsymbol{\theta}; s) \approx \mathbb{1}_{\mathcal{F}}(\boldsymbol{\theta})$ are available in the literature. Such functions become relevant in the context of importance sampling, since they allow one to approach the rare event simulation problem sequentially. A comparison of different approximations is reported in [140] in the context of sensitivity analysis. For instance, we consider the approximations:

$$f^{\text{log}}(\boldsymbol{\theta}; s) = \frac{1}{2} \left[1 + \tanh \left(-\frac{g(\boldsymbol{\theta})}{s} \right) \right] \quad \text{and} \quad f^{\text{erf}}(\boldsymbol{\theta}; s) = \Phi \left(-\frac{g(\boldsymbol{\theta})}{s} \right), \quad (3.3)$$

which correspond to the standard logistic function and the standard Gaussian CDF (error function), respectively. In the limit, when the *smoothing parameter* $s \rightarrow 0$, both functions converge to the indicator function.

3.2. Simulation methods for the estimation of rare event probabilities

The estimation of failure probabilities essentially involves the exploration of tails of probability distributions. In practice, a way to approach this problem is via (analytical) approximation methods, such as the first-order [187] and second-order [79] reliability methods (FORM and SORM). The idea in both algorithms is to represent the LSF using a truncated Taylor series expansion around the point on the failure hypersurface that is closest to the origin of an underlying standard Gaussian space. This poses an optimization task in which the objective is to find the so-called *design point*, defining the minimum distance to the origin called the *reliability index* and denoted β . In FORM, β is used to fully characterize the probability of failure via $p_{\mathcal{F}} \approx \Phi(-\beta)$. Particularly, FORM requires information on the value of the LSF function and its gradient, while SORM needs an additional knowledge of second-order derivatives of the LSF. A drawback of these approximations is that the accuracy decreases with increasing dimension of

the parameter space and the nonlinearity of the failure hypersurface [114, 238].

Monte Carlo methods provide another way to estimate the rare event simulation problem. The process involves the computation of statistical averages of response quantities, which are formulated as probability integrals over the parameter space [174, 198]. In this section, we focus on Monte Carlo-based algorithms for the estimation of $p_{\mathcal{F}}$.

3.2.1. Monte Carlo and importance sampling

Monte Carlo (MC) is a simulation procedure in which the quantity we want to compute, typically an integral of a given response variable $g(\boldsymbol{\theta})$, is expressed as an expected value over the distribution of $\boldsymbol{\theta}$. The idea is to directly replicate the underlying physical process by generating independent and identically distributed (i.i.d.) samples $\{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_N\}$ from the distribution of the uncertain parameter $\boldsymbol{\theta}$, and compute their average. In an abuse of notation, the samples are denoted as the variable itself; as a distinction, we oftentimes use an indexing subscript to specifically refer to samples.

In the context of rare event simulation, the interest is in computing the expectation (3.1). Hence, applying MC simulation amounts to

$$p_{\mathcal{F}} \approx \hat{p}_{\mathcal{F}}^{\text{MC}} = \frac{1}{N} \sum_{i=1}^N \mathbf{1}_{\mathcal{F}}(\boldsymbol{\theta}_i), \quad (3.4)$$

where $\{\boldsymbol{\theta}_i\}_{i=1}^N \stackrel{\text{i.i.d.}}{\sim} \pi_{\text{pr}}$. Two fundamental questions arise when using MC simulation: (i) how do we sample from the distribution of $\boldsymbol{\theta}$?, and (ii) how close is the MC average to the actual value of the expectation?

The first question is addressed by the theory of random variable generation. Classical methods used to compute samples from probability distributions include: inverse transform, rejection sampling [172], Box-Muller transform [42], polar method [158], Latin hypercube sampling [165] and the ziggurat algorithm [159]. In turn these algorithms are based on pseudo-random number generation schemes (uniform draws between 0 and 1), such as the linear congruential generator, Mersenne twister, Xorshift, among others. We refer to [174] for details on these approaches.

For the second question, one needs to justify and characterize properties of the MC estimator. This essentially requires application of the laws of large numbers [174]. Using the *strong law of large numbers* (SLLN), one can show that the MC estimator converges almost surely to the exact expectation, and also that the estimator is *unbiased*. We assume that the random variable $\mathbf{1}_{\mathcal{F}}(\boldsymbol{\theta})$ has finite mean $\mu < \infty$ and variance $\sigma^2 < \infty$. Formally, for the mean of the MC estimator we write

$$\text{SLLN} : \hat{p}_{\mathcal{F}}^{\text{MC}} \xrightarrow{\text{a.s.}} p_{\mathcal{F}} \quad \text{and} \quad \mathbb{E}[\hat{p}_{\mathcal{F}}^{\text{MC}}] = \frac{1}{N} \sum_{i=1}^N \mathbb{E}[\mathbf{1}_{\mathcal{F}}(\boldsymbol{\theta}_i)] = p_{\mathcal{F}}. \quad (3.5)$$

Note that the expected value of $\mathbf{1}_{\mathcal{F}}(\boldsymbol{\theta})$ is $\mu = p_{\mathcal{F}}$ and the variance is given by $\sigma^2 = p_{\mathcal{F}}(1 - p_{\mathcal{F}})$. By the *central limit theorem* (CLT), we characterize the variance of the MC estimator as

$$\text{CLT} : \frac{\hat{p}_{\mathcal{F}}^{\text{MC}} - p_{\mathcal{F}}}{\sigma/\sqrt{N}} \xrightarrow{\text{d}} \mathcal{N}(0, 1) \quad \text{and} \quad \mathbb{V}[\hat{p}_{\mathcal{F}}^{\text{MC}}] = \frac{1}{N^2} \sum_{i=1}^N \mathbb{V}[\mathbf{1}_{\mathcal{F}}(\boldsymbol{\theta}_i)] = \frac{p_{\mathcal{F}}(1 - p_{\mathcal{F}})}{N}; \quad (3.6)$$

a useful measure of the accuracy for the MC estimator is given by the coefficient of variation

$$\text{cv}(\hat{p}_{\mathcal{F}}^{\text{MC}}) = \frac{\sqrt{\mathbb{V}[\hat{p}_{\mathcal{F}}^{\text{MC}}]}}{\mathbb{E}[\hat{p}_{\mathcal{F}}^{\text{MC}}]} = \sqrt{\frac{1 - p_{\mathcal{F}}}{p_{\mathcal{F}}N}}; \quad (3.7)$$

oftentimes, the root-mean-squared error (RMSE) is also employed to characterize the failure

probability estimate:

$$\text{RMSE}(\hat{p}_{\mathcal{F}}) = \sqrt{\mathbb{E}[(\hat{p}_{\mathcal{F}} - p_{\mathcal{F}})^2]} = \frac{1}{p_{\mathcal{F}}} \sqrt{(\hat{p}_{\mathcal{F}} - p_{\mathcal{F}})^2 + (\hat{p}_{\mathcal{F}} \cdot \text{cv}(\hat{p}_{\mathcal{F}}))^2}. \quad (3.8)$$

The variance (3.6) can also be calculated by noting that $\mathbf{1}_{\mathcal{F}}(\boldsymbol{\theta})$ is distributed according to a Bernoulli random variable with probability $p_{\mathcal{F}}$. For small $p_{\mathcal{F}}$, the variance of the MC estimator can be approximated as $\text{V}[\hat{p}_{\mathcal{F}}^{\text{MC}}] \approx p_{\mathcal{F}}/N$. In practice, we usually require that the coefficient of variation of the MC estimator is proportional to a user-defined value $\delta > 0$:

$$\text{cv}(\hat{p}_{\mathcal{F}}^{\text{MC}}) \approx \frac{\sqrt{p_{\mathcal{F}}/N}}{p_{\mathcal{F}}} = \frac{1}{\sqrt{p_{\mathcal{F}}N}} = \delta \quad \text{then} \quad N \approx \frac{1}{\delta^2 p_{\mathcal{F}}}, \quad (3.9)$$

as a result, the number of samples required to obtain a coefficient of variation equal to δ is inversely proportional to the target failure probability. Therefore, in reliability problems involving the computation of rare event probabilities (small $p_{\mathcal{F}}$), and/or cases where one can only afford a small number of model evaluations (small N), the variability of $\hat{p}_{\mathcal{F}}^{\text{MC}}$ becomes very large.

Consequently, despite being robust to the type and dimension of the problem, the estimation of (3.1) via (3.4) requires a large number of samples from the prior distribution to achieve a suitable accuracy on $\hat{p}_{\mathcal{F}}^{\text{MC}}$. Variance reduction techniques can be applied to obtain MC estimators with smaller variability (see [174] for an overview), some methods include: common random numbers, antithetic variates, stratified sampling, control variates (including multilevel MC), and importance sampling. We focus on the latter since it provides the foundations of many efficient sequential algorithms utilized in rare event simulation.

The idea of *importance sampling* (IS) [130, 206] is to employ an auxiliary distribution that allows one to obtain (rare) failure samples more frequently. Consider the following modified version of (3.1)

$$p_{\mathcal{F}} = \int_{\boldsymbol{\Theta}} \frac{\mathbf{1}_{\mathcal{F}}(\boldsymbol{\theta})\pi_{\text{pr}}(\boldsymbol{\theta})}{\pi_{\text{bias}}(\boldsymbol{\theta})} \pi_{\text{bias}}(\boldsymbol{\theta}) d\boldsymbol{\theta} = \mathbb{E}_{\pi_{\text{bias}}} \left[\frac{\mathbf{1}_{\mathcal{F}}(\boldsymbol{\theta})\pi_{\text{pr}}(\boldsymbol{\theta})}{\pi_{\text{bias}}(\boldsymbol{\theta})} \right], \quad (3.10)$$

where π_{bias} is the importance or *biasing density*, satisfying the relation $\text{supp}(\mathbf{1}_{\mathcal{F}}(\boldsymbol{\theta})\pi_{\text{pr}}(\boldsymbol{\theta})) \subseteq \text{supp}(\mathbf{1}_{\mathcal{F}}(\boldsymbol{\theta})\pi_{\text{bias}}(\boldsymbol{\theta}))$ (this is required to satisfy the SLLN [45]). The purpose of the biasing density is to make the occurrence of the rare event \mathcal{F} more likely. Based on (3.10), the IS estimate of the probability of failure (3.1) is

$$\hat{p}_{\mathcal{F}}^{\text{IS}} = \frac{1}{N} \sum_{i=1}^N \mathbf{1}_{\mathcal{F}}(\boldsymbol{\theta}_i) w(\boldsymbol{\theta}_i) \quad \text{with} \quad w(\boldsymbol{\theta}_i) = \frac{\pi_{\text{pr}}(\boldsymbol{\theta}_i)}{\pi_{\text{bias}}(\boldsymbol{\theta}_i)}, \quad (3.11)$$

where $\{\boldsymbol{\theta}_i\}_{i=1}^N \stackrel{\text{i.i.d.}}{\sim} \pi_{\text{bias}}$, and each value $w(\boldsymbol{\theta}_i)$ represents a *weight* (or likelihood ratio) that corrects for the use of the biasing density and ensures that the IS estimator remains unbiased. Indeed, we can see that $\mathbb{E}_{\pi_{\text{bias}}}[\hat{p}_{\mathcal{F}}^{\text{IS}}] = \mathbb{E}_{\pi_{\text{pr}}}[\mathbf{1}_{\mathcal{F}}(\boldsymbol{\theta})] = p_{\mathcal{F}}$. Moreover, the variance of the IS estimator is

$$\begin{aligned} \text{V}_{\pi_{\text{bias}}}[\hat{p}_{\mathcal{F}}^{\text{IS}}] &= \frac{1}{N} \left(\mathbb{E}_{\pi_{\text{bias}}} \left[(\mathbf{1}_{\mathcal{F}}(\boldsymbol{\theta})w(\boldsymbol{\theta}) - p_{\mathcal{F}})^2 \right] \right) \\ &= \frac{1}{N} \left(\int_{\boldsymbol{\Theta}} \frac{\mathbf{1}_{\mathcal{F}}(\boldsymbol{\theta})\pi_{\text{pr}}^2(\boldsymbol{\theta})}{\pi_{\text{bias}}(\boldsymbol{\theta})} - 2p_{\mathcal{F}}\mathbf{1}_{\mathcal{F}}(\boldsymbol{\theta})\pi_{\text{pr}}(\boldsymbol{\theta}) + p_{\mathcal{F}}^2\pi_{\text{bias}}(\boldsymbol{\theta}) d\boldsymbol{\theta} \right) \\ &= \frac{1}{N} \left(\underbrace{\int_{\boldsymbol{\Theta}} \frac{\mathbf{1}_{\mathcal{F}}(\boldsymbol{\theta})\pi_{\text{pr}}^2(\boldsymbol{\theta})}{\pi_{\text{bias}}(\boldsymbol{\theta})} d\boldsymbol{\theta}}_{\mathbb{E}_{\pi_{\text{bias}}}[(\mathbf{1}_{\mathcal{F}}(\boldsymbol{\theta})w(\boldsymbol{\theta}))^2]} - p_{\mathcal{F}}^2 \right). \end{aligned} \quad (3.12)$$

Since the aim is to reduce the variance (3.12), we require a biasing density π_{bias} that minimizes the term $\mathbb{E}_{\pi_{\text{bias}}}[(\mathbf{1}_{\mathcal{F}}(\boldsymbol{\theta})w(\boldsymbol{\theta}))^2]$, that is

$$\pi_{\text{bias}}^*(\boldsymbol{\theta}) = \arg \min_{\pi_{\text{bias}}} \mathbb{E}_{\pi_{\text{bias}}}[(\mathbf{1}_{\mathcal{F}}(\boldsymbol{\theta})w(\boldsymbol{\theta}))^2], \quad (3.13)$$

where π_{bias}^* is the so-called *optimal biasing* density. The minimizer of (3.13) can be found by applying Jensen's inequality (under the assumption that $\mathbf{1}_{\mathcal{F}}(\boldsymbol{\theta})w(\boldsymbol{\theta})$ is convex):

$$\mathbb{E}_{\pi_{\text{bias}}}[(\mathbf{1}_{\mathcal{F}}(\boldsymbol{\theta})w(\boldsymbol{\theta}))^2] \geq (\mathbb{E}_{\pi_{\text{bias}}}[\mathbf{1}_{\mathcal{F}}(\boldsymbol{\theta})w(\boldsymbol{\theta})])^2. \quad (3.14)$$

The relation in (3.14) is strict if $\mathbf{1}_{\mathcal{F}}(\boldsymbol{\theta})w(\boldsymbol{\theta})$ is almost everywhere constant [45]. Hence, the optimal biasing density, generating a zero-variance estimate, is given by

$$\begin{aligned} \pi_{\text{bias}}^*(\boldsymbol{\theta}) &\propto \mathbb{E}_{\pi_{\text{bias}}}[\mathbf{1}_{\mathcal{F}}(\boldsymbol{\theta})w(\boldsymbol{\theta})] = \mathbf{1}_{\mathcal{F}}(\boldsymbol{\theta})\pi_{\text{pr}}(\boldsymbol{\theta}) \\ &= \frac{1}{p_{\mathcal{F}}} \mathbf{1}_{\mathcal{F}}(\boldsymbol{\theta})\pi_{\text{pr}}(\boldsymbol{\theta}) = \pi_{\mathcal{F}}(\boldsymbol{\theta}). \end{aligned} \quad (3.15)$$

Note that the optimal biasing density is equivalent to the posterior-failure density defined in (3.2) and its construction requires the knowledge of the target failure probability in advance. However, even if π_{bias}^* in (3.15) is not usable in practice, it still provides a guideline on how to build useful IS schemes. This is exploited, for instance, in sequential importance sampling and the cross-entropy method. We point out that large deviation techniques can be employed to derive asymptotic expansions of the optimal biasing distribution, such that they describe the most probable path leading to the rare event (see, e.g., [13, 45]).

In some situations, one of the weights $w(\boldsymbol{\theta}_i)$ can be considerably larger than all the others. In this case, one has effectively only one sample contributing to the estimator. It is then convenient to have a diagnostic that reveals whether the weights are showing this behavior or not. A common metric is the *effective sample size* (N_{ESS}):

$$N_{\text{ESS}} = \frac{(\sum_{i=1}^N w(\boldsymbol{\theta}_i))^2}{\sum_{i=1}^N (w(\boldsymbol{\theta}_i))^2} = \frac{(\sum_{i=1}^N w_i)^2}{\sum_{i=1}^N w_i^2} = \frac{(\widetilde{W})^2}{\widetilde{W^2}}, \quad (3.16)$$

where \widetilde{W} denotes the sum of the weights, $\widetilde{W^2}$ the sum of the squared weights, and $1 \leq N_{\text{ESS}} \leq N$ (in an abuse of notation, we denote each weight as $w_i = w(\boldsymbol{\theta}_i)$). The weights are all the same when $N_{\text{ESS}} = N$. Conversely, if the weights are very unequal, the IS estimator is averaging only with $N_{\text{ESS}} \ll N$ samples and thus it is less accurate. Another way to express the N_{ESS} is via the coefficient of variation of the weights [174]

$$N_{\text{ESS}} = \frac{N}{1 + (\text{cv}(\mathbf{w}))^2}, \quad (3.17)$$

where $\mathbf{w} = \{w_i\}_{i=1}^N$ is the set of weights. The effective sample size is not always an optimal diagnostic and different modifications have been proposed to improve it. A common approach is to define N_{ESS} with respect to $\mathbf{1}_{\mathcal{F}}(\boldsymbol{\theta})$, since the effectiveness of IS depends also on this function (see for example [174, Sec. 9.3]).

Finally, we remark that other specialized simulation algorithms are employed for the estimation of (3.1), including: directional and line sampling [69, 114, 139], advanced IS schemes such as, sequential importance sampling [180] and the cross-entropy method [197], and multilevel splitting methods [48, 38, 242, 232], which include subset simulation [15].

We describe in the following sections some of these algorithms and compare them in the context of random fields.

3.2.2. Sequential importance sampling

Sequential importance sampling (SIS) [180] defines a parametrization of the optimal IS density (3.15), such that it can be estimated in a sequential manner. Since the construction of (3.15) is unfeasible, one can consider a sequence of intermediate biasing densities of the form

$$\pi_{\text{bias}}(\boldsymbol{\theta}; s_j) = \frac{1}{p_j} f(\boldsymbol{\theta}; s_j) \pi_{\text{pr}}(\boldsymbol{\theta}), \quad (3.18)$$

where p_j is an intermediate failure probability or normalizing constant and $j = 0, \dots, n_{\text{lv}}$ is a level index, with n_{lv} denoting the total number of levels. The function $f(\boldsymbol{\theta}; s_j)$ is a smooth approximation of the failure indicator function and it can be defined from (3.3). The intermediate biasing densities are then parameterized via the smoothing parameters $\infty = s_0 > s_1 > \dots > s_j > \dots > s_{n_{\text{lv}}} > 0$. In the context of sequential MC [62], the elements of the set $\{s_j\}_{i=0}^{n_{\text{lv}}}$ are called temperatures and n_{lv} is the number of tempering steps. To estimate the probability of failure, SIS requires samples from the final biasing density $\pi_{\text{bias}}(\boldsymbol{\theta}; s_{n_{\text{lv}}})$ and the value of its normalizing constant $p_{n_{\text{lv}}}$.

Starting at level $j = 0$ from an a biasing density $\pi_{\text{bias}}(\boldsymbol{\theta}; s_0) = \pi_{\text{pr}}(\boldsymbol{\theta})$, we can estimate the intermediate probabilities at subsequent levels as

$$\begin{aligned} p_{j+1} &= \int_{\Theta} f(\boldsymbol{\theta}; s_{j+1}) \pi_{\text{pr}}(\boldsymbol{\theta}) d\boldsymbol{\theta} = \int_{\Theta} f(\boldsymbol{\theta}; s_{j+1}) \pi_{\text{pr}}(\boldsymbol{\theta}) \frac{\pi_{\text{bias}}(\boldsymbol{\theta}; s_j)}{\pi_{\text{bias}}(\boldsymbol{\theta}; s_j)} d\boldsymbol{\theta} \\ &= p_j \int_{\Theta} \frac{f(\boldsymbol{\theta}; s_{j+1}) \pi_{\text{pr}}(\boldsymbol{\theta})}{f(\boldsymbol{\theta}; s_j) \pi_{\text{pr}}(\boldsymbol{\theta})} \pi_{\text{bias}}(\boldsymbol{\theta}; s_j) d\boldsymbol{\theta} \\ &= p_j \int_{\Theta} \underbrace{\frac{f(\boldsymbol{\theta}; s_{j+1})}{f(\boldsymbol{\theta}; s_j)}}_{w(\boldsymbol{\theta})} \pi_{\text{bias}}(\boldsymbol{\theta}; s_j) d\boldsymbol{\theta} = p_j \mathbb{E}_{\pi_{\text{bias},j}}[w(\boldsymbol{\theta})], \end{aligned} \quad (3.19)$$

with the notation $\pi_{\text{bias},j} = \pi_{\text{bias}}(\boldsymbol{\theta}; s_j)$. Notice in (3.19) that the weights are given by the ratio of subsequent smooth approximations of the indicator function. From (3.19), we can define an IS estimator for the ratio of normalizing constants

$$S_j = \frac{p_{j+1}}{p_j} = \mathbb{E}_{\pi_{\text{bias},j}}[w(\boldsymbol{\theta})] \approx \frac{1}{N} \sum_{i=1}^N w(\boldsymbol{\theta}_i^{(j)}) = \widehat{S}_j, \quad (3.20)$$

where $\{\boldsymbol{\theta}_i^{(j)}\}_{i=1}^N \stackrel{\text{i.i.d.}}{\sim} \pi_{\text{bias},j}$, and we include a superscript to denote the set of samples at the j -th level.

Given samples from $\pi_{\text{bias},j}$, we can obtain next level samples from $\pi_{\text{bias},j+1}$ by applying a *resampling-moving* scheme [62]: (i) use a resampling method to remove the samples with small weights since they barely contribute to the estimation of (3.20). This is performed by selecting randomly with replacement from the samples $\{\boldsymbol{\theta}_i^{(j)}\}_{i=1}^N$, with probability assigned to each sample proportional to their weights $\{w(\boldsymbol{\theta}_i^{(j)})\}_{i=1}^N$; (ii) move the position of the resulting samples to regions of high probability of the next intermediate biasing $\pi_{\text{bias},j}$. This is performed by a MCMC algorithm that has stationary distribution $\pi_{\text{bias},j}$. We repeat this strategy until the last level $j = n_{\text{lv}}$ is reached; at this stage the normalizing constant of the final biasing density is estimated as

$$\widehat{p}_{n_{\text{lv}}} = \prod_{j=0}^{n_{\text{lv}}} \widehat{S}_j. \quad (3.21)$$

In order to obtain an accurate estimate of \widehat{S}_j in (3.20), one needs to ensure that each pair of consecutive biasing densities do not change significantly. A way to guarantee a small variation

is by controlling the variance of the weights. The idea is to match the effective sampling size to a target predefined value, as it is typically carried out in sequential MC approaches (e.g., [62]). Such task can be equivalently performed by requiring that the sample coefficient of variation of the weights at each level is equal to an user-defined target value δ^* [180]. Therefore, the smoothing parameters are updated at each level via the optimization problem

$$s_{j+1} = \arg \min_{s \in (0, s_j)} \left(\text{cv} \left(f(\boldsymbol{\theta}^{(j)}, s) / f(\boldsymbol{\theta}^{(j)}, s_j) \right) - \delta^* \right)^2, \quad (3.22)$$

which is solved without further evaluations of the LSF. A value of $\delta^* = 1.5$ is suggested in [180]; in this case, the effective coefficient of variation of the failure probability estimate is $\delta^* / \sqrt{N} \approx 0.047$ using $N = 10^3$ samples per level.

Algorithm 1 SIS method.

Input : number of samples per level N , number of MCMC chains N_c , target coefficient of variation δ^* , LSF $g(\boldsymbol{\theta})$, smooth indicator $f(\boldsymbol{\theta}; s)$, prior density $\pi_{\text{pr}}(\boldsymbol{\theta})$, maximum iterations m

```

1 Generate  $N$  samples from the prior  $\boldsymbol{\theta}^{(0)} \sim \pi_{\text{pr}}(\cdot)$ 
2 Evaluate LSF  $\mathbf{g}_{\text{eval}} \leftarrow g(\boldsymbol{\theta}^{(0)})$ , and indicator function  $\mathbf{d}_{\text{eval}} \leftarrow \mathbf{g}_{\text{eval}} \leq 0 = \mathbf{1}_{\mathcal{F}}(\boldsymbol{\theta}^{(0)})$ 
3 Set initial smoothing parameter  $s_0 \leftarrow \infty$ 
4 Initial level  $j \leftarrow 0$ 
5 while True do
6     Evaluate smooth indicator function  $\mathbf{f}_{\text{eval}}^{(j)} \leftarrow f(\boldsymbol{\theta}^{(j)}; s_j)$ 
7     Compute the coefficient of variation of the ratio between the indicator function and its smooth approximation
            $\widehat{\text{cv}}^{(j)} \leftarrow \text{cv}(\mathbf{d}_{\text{eval}} / \mathbf{f}_{\text{eval}}^{(j)})$ 
8     if  $(\widehat{\text{cv}}^{(j)} \leq \delta^*)$  or  $(j > m)$  then
9         | Break
10    end
11    /* Weighting: */
12    Use the level samples  $\boldsymbol{\theta}^{(j)}$  to update the smoothing parameter  $s_{j+1}$  as per (3.22)
13    if  $j = 0$  then
14        | Compute the weights  $\mathbf{w}^{(j)} \leftarrow \mathbf{f}_{\text{eval}}^{(j)}$ 
15    else
16        | Compute the weights  $\mathbf{w}^{(j)} \leftarrow \mathbf{f}_{\text{eval}}^{(j)} / \mathbf{f}_{\text{eval}}^{(j-1)}$ 
17    end
18    Estimate the ratio of normalizing constants  $\widehat{S}_j \leftarrow \frac{1}{N} \sum_{i=1}^N w_i^{(j)}$ 
19    Normalize the weights  $\widetilde{\mathbf{w}}^{(j)}$ 
20    /* Resampling: */
21    Generate a set of samples  $\boldsymbol{\theta}_{\text{seed}}$  consisting of  $N_c$  particles randomly sampled with replacement from
            $\boldsymbol{\theta}^{(j)}$ , where the probabilities associated with each entry correspond to the normalized weights  $\widetilde{\mathbf{w}}^{(j)}$ 
22    /* Moving: */
23    Use  $\boldsymbol{\theta}_{\text{seed}}$  as seeds of for a MCMC algorithm to generate a new set of samples  $\boldsymbol{\theta}^{(j+1)}$ ; this process
           also returns the associated LSF evaluations  $\mathbf{g}_{\text{eval}} \leftarrow g(\boldsymbol{\theta}^{(j+1)})$  and indicator  $\mathbf{d}_{\text{eval}} \leftarrow \mathbf{g}_{\text{eval}} \leq 0$ 
24     $j \leftarrow j + 1$ 
25 end
26  $n_{\text{lv}} \leftarrow j$ 
27 Compute  $\widehat{p}_{n_{\text{lv}}} \leftarrow \prod_{i=0}^{n_{\text{lv}}} \widehat{S}_i$ 
28 Compute the IS estimator  $\widehat{p}_{\mathcal{F}} \leftarrow \frac{\widehat{p}_{n_{\text{lv}}}}{N} \sum_{i=1}^N d_{\text{eval}}(i) / f_{\text{eval}}^{(j)}(i)$ 

```

Output : $\widehat{p}_{\mathcal{F}}$

The SIS algorithm proceeds until the coefficient of variation of the ratio between the indicator function and its smooth approximation is smaller than the target δ^* . At the final level $j = n_{\text{lv}}$, the N samples are used to approximate the probability of failure with an IS estimator

$$\hat{p}_{\mathcal{F}}^{\text{SIS}} = \frac{\hat{p}_{n_{\text{lv}}}}{N} \sum_{i=1}^N \frac{\mathbb{1}_{\mathcal{F}}(\boldsymbol{\theta}_i)}{f(\boldsymbol{\theta}_i, s_{n_{\text{lv}}})}. \quad (3.23)$$

Algorithm 1 summarizes the SIS method. We refer to section 4.2 for some MCMC methods that can be applied in the moving step within SIS.

3.2.3. Subset simulation

Subset simulation (SuS) [15] is especially designed for the estimation of rare event probabilities. This method transforms a rare event simulation problem into a sequence of problems involving more frequent events.

The process is carried out by expressing a very small probability of failure as the product of larger conditional probabilities in the following way: the parameter space is divided into a decreasing sequence of nested subsets (called *intermediate failure domains*) starting from the whole space and narrowing down to the target failure set, i.e., $\Theta = \mathcal{F}_0 \supset \mathcal{F}_1 \supset \dots \supset \mathcal{F}_{n_{\text{lv}}} = \mathcal{F}$, such that $\mathcal{F} = \bigcap_{j=0}^{n_{\text{lv}}} \mathcal{F}_j$. By the general product rule (or chain rule) of probability, the probability of failure can be expressed as [15]

$$p_{\mathcal{F}} = \mathbb{P}\left[\bigcap_{j=0}^{n_{\text{lv}}} \mathcal{F}_j\right] = \mathbb{P}\left[\mathcal{F}_{n_{\text{lv}}} \mid \bigcap_{j=0}^{n_{\text{lv}}-1} \mathcal{F}_j\right] \mathbb{P}\left[\bigcap_{j=0}^{n_{\text{lv}}-1} \mathcal{F}_j\right] \quad (3.24)$$

$$= \mathbb{P}[\mathcal{F}_{n_{\text{lv}}} \mid \mathcal{F}_{n_{\text{lv}}-1}] \mathbb{P}\left[\bigcap_{j=0}^{n_{\text{lv}}-1} \mathcal{F}_j\right] \quad (3.25)$$

$$= \prod_{j=1}^{n_{\text{lv}}} \mathbb{P}[\mathcal{F}_j \mid \mathcal{F}_{j-1}], \quad (3.26)$$

where $\mathbb{P}[\mathcal{F}_j \mid \mathcal{F}_{j-1}]$ represents the conditional probability at the $(j-1)$ -th level. Each intermediate failure domain is defined as the set $\mathcal{F}_j = \{\boldsymbol{\theta} \in \Theta : g(\boldsymbol{\theta}) \leq \tau_j\}$, where $\tau_1 > \dots > \tau_j > \dots > \tau_{n_{\text{lv}}} = 0$ forms a decreasing sequence of threshold levels. Note that the conditional probabilities can be estimated by MC as,

$$\mathbb{P}[\mathcal{F}_j \mid \mathcal{F}_{j-1}] \approx \hat{p}_j = \frac{1}{N} \sum_{i=1}^N \mathbb{1}_{\mathcal{F}_j}(\boldsymbol{\theta}_i^{(j-1)}), \quad (3.27)$$

where $\{\boldsymbol{\theta}_i^{(j-1)}\}_{i=1}^N$ are samples distributed according to the conditional densities $\pi(\cdot \mid \mathcal{F}_{j-1})$. The initial samples $\{\boldsymbol{\theta}_i^{(0)}\}_{i=1}^N$ are i.i.d., according to the prior. However, for the subsequent intermediate levels, the calculation of $\mathbb{P}[\mathcal{F}_j \mid \mathcal{F}_{j-1}]$, $j = 1, \dots, n_{\text{lv}}$ is more complicated since it is necessary to draw samples from the conditional densities. This task is performed by generating Markov chains that sequentially populate the target failure domain \mathcal{F} .

The intermediate threshold values are cumbersome to define in advance. One alternative is to select them adaptively, such that the conditional failure probability of each level corresponds to a common fixed value $\mathbb{P}[\mathcal{F}_j \mid \mathcal{F}_{j-1}] = p_0$. The recommended intermediate probability value is $p_0 \in [0.1, 0.3]$ [255]. The adaptive selection implies that each threshold value τ_j is defined as the $N \cdot (1 - p_0)$ -th order statistic (p_0 -quantile) of the set of LSF values $g(\boldsymbol{\theta}^{(j-1)})$.

Provided that a suitable MCMC method is used for sampling the involved intermediate conditional distributions, the SuS method proceeds as follows: at the initial MC stage, the first intermediate level τ_1 is selected as the p_0 -percentile of the LSF values $g(\boldsymbol{\theta}^{(0)})$, by doing so, N_s samples from $\boldsymbol{\theta}^{(0)}$ that lie in \mathcal{F}_1 will be distributed according to an intermediate density $\pi(\cdot \mid \mathcal{F}_1)$. Those samples provide seeds to simulate Markov chains using a MCMC algorithm that has $\pi(\cdot \mid \mathcal{F}_1)$ as stationary distribution; this process generates new samples $\{\boldsymbol{\theta}_i^{(1)}\}_{i=1}^N$. We repeat this simulation strategy until the failure event \mathcal{F} is reached. The failure probability estimate of

the SuS method is given by,

$$\hat{p}_{\mathcal{F}}^{\text{SuS}} = \prod_{j=1}^{n_{\text{lv}}} \hat{p}_j = p_0^{n_{\text{lv}}-1} \underbrace{\frac{1}{N} \sum_{i=1}^N \mathbb{1}_{\mathcal{F}}(\boldsymbol{\theta}_i^{n_{\text{lv}}-1})}_{n_{\mathcal{F}}/N}, \quad (3.28)$$

where the second term corresponds to the intermediate conditional probability at the last level $\tau_{n_{\text{lv}}} = 0$ computed as the ratio between the number of samples that lie in \mathcal{F} , denoted $n_{\mathcal{F}}$, and the total number of samples per level N .

Remark 3.3. SuS is a splitting method [48] in which the intermediate levels are fixed *a priori*. Adaptive strategies are also utilized in splitting algorithms to estimate the levels on-the-fly. In this case, they proceed similarly to SuS by fixing the intermediate probabilities to a predefined value and estimating the thresholds based on order-statistics; see, e.g., [38, 59].

A general procedure for the implementation of SuS is shown Algorithm 2. We refer to [177] for a variety of MCMC methods that are used to sample the intermediate levels within SuS.

Algorithm 2 SuS method.

Input : number of samples per level N , prescribed intermediate probability p_0 , LSF $g(\boldsymbol{\theta})$, maximum iterations m .

```

/* Initial MC */
1 Generate  $N$  samples from the prior  $\boldsymbol{\theta}^{(0)} \sim \pi_{\text{pr}}(\cdot)$ 
2 Evaluate LSF  $\mathbf{g}_{\text{eval}} \leftarrow g(\boldsymbol{\theta}^{(0)})$ 
3 Initial level  $j \leftarrow 0$  and threshold  $\tau_j \leftarrow \infty$ 
/* SuS steps */
4 while  $\tau_j > 0$  and  $j < m$  do
5      $j \leftarrow j + 1$ 
    /* Update threshold */
6     Compute the indices resulting from the ascending sorting of LSF values,  $\text{idx} \leftarrow \text{sort}(\mathbf{g}_{\text{eval}})$ 
7     Sort the samples according to the sort-key  $\text{idx}$ ,  $\boldsymbol{\theta}_{\text{sort}} \leftarrow \boldsymbol{\theta}^{(j-1)}(\text{idx})$ 
8     Set the intermediate threshold level  $\tau_j$  as the  $p_0$ -quantile of the LSF values in  $\mathbf{g}_{\text{eval}}$ 
9     Compute the number of failure samples in the level,  $n_{\mathcal{F}} \leftarrow \sum_{i=1}^N (\mathbf{g}_{\text{eval}}(i) \leq \max(\tau_j, 0))$ 
10    if  $(\tau_j \leq 0)$  then
11         $\tau_j \leftarrow 0$ 
12         $\hat{p}_j \leftarrow n_{\mathcal{F}}/N$ 
13    else
14         $\hat{p}_j \leftarrow p_0$ 
15    end
16    Select  $n_{\mathcal{F}}$  seeds,  $\boldsymbol{\theta}_{\text{seed}} \leftarrow \boldsymbol{\theta}_{\text{sort}}(0 : n_{\mathcal{F}})$ 
17    Randomize the ordering of the seeds  $\boldsymbol{\theta}_{\text{seed}}$ 
    /* MCMC step */
18    Use  $\boldsymbol{\theta}_{\text{seed}}$  as seeds for an MCMC algorithm to generate a new set of samples  $\boldsymbol{\theta}^{(j)}$ ; this process also
        returns the associated LSF evaluations  $\mathbf{g}_{\text{eval}} \leftarrow g(\boldsymbol{\theta}^{(j)})$ 
19 end
20  $n_{\text{lv}} \leftarrow j$ 
21 Compute  $\hat{p}_{\mathcal{F}} \leftarrow \prod_{i=1}^{n_{\text{lv}}} \hat{p}_i$ 
Output :  $\hat{p}_{\mathcal{F}}$ 

```

3.2.4. Cross-entropy method

The standard *cross-entropy* (CE) method [197] considers the problem of approximating $\pi_{\text{bias}}^*(\boldsymbol{\theta})$ by a *parametric biasing* density $\pi_{\text{bias}}(\boldsymbol{\theta}; \boldsymbol{v})$, with reference parameters \boldsymbol{v} . The approximation is selected from a family of densities $\Pi = \{\pi_{\text{bias}}(\boldsymbol{\theta}; \boldsymbol{v}) \mid \boldsymbol{v} \in \Upsilon\}$ designed to be of simpler form than

π_{bias}^* . Thereafter, the objective is to find $\mathbf{v}^* \in \Upsilon$ such that the distance between the optimal and approximated biasing densities is minimal. The dissimilarity between these distributions is measured by the cross-entropy or Kullback–Leibler divergence (KLD) [91]

$$\begin{aligned} D_{\text{KL}}(\pi_{\text{bias}}^* || \pi_{\text{bias}}) &= \int_{\Theta} \ln \left(\frac{\pi_{\text{bias}}^*(\boldsymbol{\theta})}{\pi_{\text{bias}}(\boldsymbol{\theta}; \mathbf{v})} \right) \pi_{\text{bias}}^*(\boldsymbol{\theta}) d\boldsymbol{\theta} \\ &= \int_{\Theta} \ln \pi_{\text{bias}}^*(\boldsymbol{\theta}) \pi_{\text{bias}}^*(\boldsymbol{\theta}) d\boldsymbol{\theta} - \int_{\Theta} \ln \pi_{\text{bias}}(\boldsymbol{\theta}; \mathbf{v}) \pi_{\text{bias}}^*(\boldsymbol{\theta}) d\boldsymbol{\theta}. \end{aligned} \quad (3.29)$$

The first term in (3.29) is invariant with respect to any choice of π_{bias} and the problem reduces to the optimization task:

$$\mathbf{v}^* = \arg \max_{\mathbf{v} \in \Upsilon} \mathbb{E}_{\pi_{\text{bias}}^*} [\ln \pi_{\text{bias}}(\boldsymbol{\theta}; \mathbf{v})], \quad (3.30)$$

where \mathbf{v}^* denotes the optimal reference parameters. We can substitute the optimal biasing density (3.15) into (3.30) to express the optimization program as

$$\mathbf{v}^* = \arg \max_{\mathbf{v} \in \Upsilon} \mathbb{E}_{\pi_{\text{pr}}} [\ln \pi_{\text{bias}}(\boldsymbol{\theta}; \mathbf{v}) \mathbb{1}_{\mathcal{F}}(\boldsymbol{\theta})]. \quad (3.31)$$

The expectation (3.31) can be estimated by MC using samples from the prior distribution. However, this is impractical if \mathcal{F} defines a rare event. In order to efficiently evaluate (3.31), we apply IS with biasing distribution $\pi_{\text{bias}}(\boldsymbol{\theta}; \mathbf{v}') \in \Pi$ (for reference parameters $\mathbf{v}' \in \Upsilon$):

$$\mathbf{v}^* = \arg \max_{\mathbf{v} \in \Upsilon} \mathbb{E}_{\pi_{\text{bias}}(\cdot; \mathbf{v}')} [\ln \pi_{\text{bias}}(\boldsymbol{\theta}; \mathbf{v}) \mathbb{1}_{\mathcal{F}}(\boldsymbol{\theta}) w(\boldsymbol{\theta}; \mathbf{v}')] \quad \text{with} \quad w(\boldsymbol{\theta}; \mathbf{v}') = \frac{\pi_{\text{pr}}(\boldsymbol{\theta})}{\pi_{\text{bias}}(\boldsymbol{\theta}; \mathbf{v}')}. \quad (3.32)$$

We can further employ the IS estimator of the expectation (3.32) to define the stochastic optimization problem:

$$\mathbf{v}^* \approx \hat{\mathbf{v}}^* = \arg \max_{\mathbf{v} \in \Upsilon} \mathcal{J}(\mathbf{v}) \quad \text{with} \quad \mathcal{J}(\mathbf{v}) = \frac{1}{N} \sum_{i=1}^N \ln \pi_{\text{bias}}(\boldsymbol{\theta}_i; \mathbf{v}) \mathbb{1}_{\mathcal{F}}(\boldsymbol{\theta}_i) w(\boldsymbol{\theta}_i; \mathbf{v}'), \quad (3.33)$$

where $\{\boldsymbol{\theta}_i\}_{i=1}^N \stackrel{\text{i.i.d.}}{\sim} \pi_{\text{bias}}(\cdot; \mathbf{v}')$. Under the assumption that $\mathcal{J}(\mathbf{v})$ is convex and differentiable with respect to \mathbf{v} , the solution to (3.33) can be computed by $\nabla_{\mathbf{v}} \mathcal{J}(\mathbf{v}) = 0$ [197]. Moreover, if the biasing distribution belongs to the natural exponential family, the solution of the stochastic optimization problem can be computed analytically. For instance, if Π is a collection of Gaussian densities, the parameter \mathbf{v} is selected from the space Υ containing mean vectors and covariance matrices. In this case, the reference parameter estimator $\hat{\mathbf{v}}^*$ has an explicit updating rule (see, for instance [89]).

In principle, estimating the optimal reference parameters using $\pi_{\text{bias}}(\boldsymbol{\theta}; \mathbf{v}')$ in (3.32) yields better efficiency than using $\pi_{\text{pr}}(\boldsymbol{\theta})$ in (3.31). However, one still requires a good initial choice of \mathbf{v}' , such that a substantial number of samples from $\pi_{\text{bias}}(\boldsymbol{\theta}; \mathbf{v}')$ lie in the failure domain. This is addressed in the CE method by gradually approaching the target failure event. The idea is to construct a sequence of intermediate sets $\mathcal{F}_j = \{\boldsymbol{\theta} \in \Theta : g(\boldsymbol{\theta}) \leq \gamma_j\}$, with intermediate thresholds $\gamma_j \geq 0$.

The CE optimization (3.33) is now solved at each level with respect to an intermediate optimal biasing density $\pi_{\text{bias},j}^*(\boldsymbol{\theta}) \propto \mathbb{1}_{\mathcal{F}_j}(\boldsymbol{\theta}) \pi_{\text{pr}}(\boldsymbol{\theta})$ associated to a failure threshold γ_j . Starting from an initial reference parameter estimate $\hat{\mathbf{v}}_0$, the sequential stochastic CE program reads

$$\hat{\mathbf{v}}_{j+1} = \arg \max_{\mathbf{v} \in \Upsilon} \frac{1}{N} \sum_{i=1}^N \ln \pi_{\text{bias}}(\boldsymbol{\theta}_i; \mathbf{v}) \tilde{w}_i^{(j)} \quad \text{with} \quad \tilde{w}_i^{(j)} = \mathbb{1}_{\mathcal{F}_j}(\boldsymbol{\theta}_i) \frac{\pi_{\text{pr}}(\boldsymbol{\theta}_i)}{\pi_{\text{bias}}(\boldsymbol{\theta}_i; \hat{\mathbf{v}}_j)}, \quad (3.34)$$

where $\{\boldsymbol{\theta}_i\}_{i=1}^N \stackrel{\text{i.i.d.}}{\sim} \pi_{\text{bias}}(\cdot; \hat{\boldsymbol{v}}_j)$, and the vector $\tilde{\boldsymbol{w}}^{(j)}$ contains the j -th level weights computed with respect to the failure indicator function. Each failure threshold γ_j is defined as the $N \cdot (1 - \rho)$ -th order statistic (ρ -quantile) of the sequence of LSF values $\{g_i = g(\boldsymbol{\theta}_i)\}_{i=1}^N$. The quantile value ρ is chosen to ensure that a good portion of the samples from $\pi_{\text{bias}}(\cdot; \hat{\boldsymbol{v}}_j)$ fall in the failure set \mathcal{F}_j , usually $\rho \in [0.01, 0.1]$ [198].

The CE algorithm proceeds until an intermediate threshold is such that $\gamma_j \leq 0$, for which at least $N \cdot \rho$ samples lie in the target failure set \mathcal{F} . These final “elite” samples are used to estimate the reference parameter of the approximated optimal biasing density, i.e., $\hat{\boldsymbol{v}}_{n_{\text{lv}}} = \hat{\boldsymbol{v}}^*$, where n_{lv} denotes the total number of levels or iterations. Samples from the latter are then used to compute the probability of failure via (3.11).

Note that if the prior and the biasing densities belong to the same parametric family, the initial estimate of the reference parameters is typically selected as the parameters defining the prior (e.g., $\hat{\boldsymbol{v}}_0 = [\boldsymbol{\mu}_{\text{pr}}, \boldsymbol{\Sigma}_{\text{pr}}]$). Algorithm 3 summarizes the CE method.

Remark 3.4. The CE method can also be defined in a nonparametric way [196], where the order in the KLD is exchanged, i.e., $D_{\text{KL}}(\pi_{\text{bias}} \parallel \pi_{\text{bias}}^*)$. Hence, the aim is to minimize the KLD to the optimal biasing distribution from all possible distributions satisfying certain moment-matching constraints; this is also the idea of the generalized CE method [39].

Algorithm 3 CE method.

Input : number of samples per level N , quantile value ρ , LSF $g(\boldsymbol{\theta})$, prior density $\pi_{\text{pr}}(\boldsymbol{\theta})$, maximum iterations m

```

1 Set the parametric family of biasing distributions  $\pi_{\text{bias}}(\boldsymbol{\theta}; \boldsymbol{v})$ 
2 Set initial reference parameters  $\hat{\boldsymbol{v}}_0$ 
3 Initial level  $j \leftarrow 0$ 
4 while True do
    /* Moving */
5     Generate  $N$  samples from the biasing distribution  $\boldsymbol{\theta}^{(j)} \sim \pi_{\text{bias}}(\cdot; \hat{\boldsymbol{v}}_j)$ 
6     Evaluate LSF  $\boldsymbol{g}_{\text{eval}} \leftarrow g(\boldsymbol{\theta}^{(j)})$ 
7     Set the intermediate threshold level  $\gamma_j$  as the  $\rho$ -quantile of the LSF values in  $\boldsymbol{g}_{\text{eval}}$ 
8     Evaluate the indicator function at the current level  $\boldsymbol{d}_{\text{eval}} \leftarrow \mathbf{1}_{\mathcal{F}_j}(\boldsymbol{\theta}^{(j)}) = \boldsymbol{g}_{\text{eval}} \leq \gamma_j$ 
9     if  $(\gamma_j \leq 0)$  or  $(j > m)$  then
10         |  $\gamma_j \leftarrow 0$ 
11         | Break
12     end
    /* Weigthing */
13     Compute the weights  $\boldsymbol{w}^{(j)} \leftarrow \exp(\ln \pi_{\text{pr}}(\boldsymbol{\theta}^{(j)}) - \ln \pi_{\text{bias}}(\boldsymbol{\theta}^{(j)}; \hat{\boldsymbol{v}}_j))$ 
14     Compute the weights associated to the indicator function  $\tilde{\boldsymbol{w}}^{(j)} \leftarrow \boldsymbol{d}_{\text{eval}} \cdot \boldsymbol{w}^{(j)}$ 
    /* Optimization */
15     Find the next reference parameters  $\hat{\boldsymbol{v}}_{j+1}$  by solving the stochastic optimization problem (3.34) or
        using a closed-form update if available.
16      $j \leftarrow j + 1$ 
17 end

```

Output : $\hat{p}_{\mathcal{F}}$

3.2.5. Improved cross-entropy method

In the standard CE method, only $N \cdot \rho$ samples drawn from the intermediate biasing densities contribute to the estimation of each $\hat{\boldsymbol{v}}_{j+1}$ in (3.34). This is due to the selection of the failure thresholds γ_j and the shape of the intermediate optimal biasing densities. These are conditional distributions that are defined at each intermediate failure event using an indicator function.

Since the remaining $N \cdot (1 - \rho)$ samples can potentially support the estimation of each reference parameter $\hat{\boldsymbol{v}}_{j+1}$, it is ideal to use all samples.

The *improved cross-entropy* (iCE) method [179] re-defines the intermediate optimal biasing densities using an approximation of the indicator function that guarantees a smooth transition towards the optimal biasing density. This modification allows one to employ all the N samples in the estimation of the reference parameters at each intermediate level.

After using any of the smooth approximations (3.3), the sequential CE optimization problem (3.34) is now defined by the sequence $\{s_j > 0\}_{j=0}^{n_{iv}}$, instead of the failure thresholds γ_j . The new intermediate optimal biasing density associated with a smoothing parameter s_j is defined as $\pi_{\text{bias},j}^*(\boldsymbol{\theta}; s_j) = 1/p_j(f(\boldsymbol{\theta}; s_j)\pi_{\text{pr}}(\boldsymbol{\theta}))$, where p_j is the normalizing constant. Therefore, we can re-formulate the sequential CE stochastic optimization problem (3.34) as

$$\hat{\boldsymbol{v}}_{j+1} = \arg \max_{\boldsymbol{v} \in \Upsilon} \frac{1}{N} \sum_{i=1}^N \ln \pi_{\text{bias}}(\boldsymbol{\theta}_i; \boldsymbol{v}) \tilde{w}_i^{(j+1)} \quad \text{with} \quad \tilde{w}_i^{(j+1)} = f(\boldsymbol{\theta}_i; s_{j+1}) \frac{\pi_{\text{pr}}(\boldsymbol{\theta}_i)}{\pi_{\text{bias}}(\boldsymbol{\theta}_i; \hat{\boldsymbol{v}}_j)}, \quad (3.35)$$

where $\{\boldsymbol{\theta}_i\}_{i=1}^N \stackrel{\text{i.i.d.}}{\sim} \pi_{\text{bias}}(\cdot; \hat{\boldsymbol{v}}_j)$, and the vector $\tilde{\boldsymbol{w}}^{(j+1)}$ contains the $(j+1)$ -th level weights computed with respect to the smooth approximation of the indicator function at s_{j+1} . The objective function (3.35) is analogous to the CE stochastic optimization (3.34), and thus both problems are solved identically.

Algorithm 4 iCE method.

Input : number of samples per level N , target coefficient of variation δ^* , LSF $g(\boldsymbol{\theta})$, smooth indicator $f(\boldsymbol{\theta}; s)$, prior density $\pi_{\text{pr}}(\boldsymbol{\theta})$, maximum iterations m

- 1 Set the parametric family of the biasing distributions $\pi_{\text{bias}}(\boldsymbol{\theta}; \boldsymbol{v})$
 - 2 Set initial reference parameters $\hat{\boldsymbol{v}}_0$ and smoothing parameter $s_0 \leftarrow \infty$
 - 3 Initial level $j \leftarrow 0$
 - 4 **while** *True* **do**
 - /* Moving */
 - 5 Generate N samples from the biasing distribution $\boldsymbol{\theta}^{(j)} \sim \pi_{\text{bias}}(\cdot; \hat{\boldsymbol{v}}_j)$
 - 6 Evaluate LSF $\boldsymbol{g}_{\text{eval}} \leftarrow g(\boldsymbol{\theta}^{(j)})$, and indicator function $\boldsymbol{d}_{\text{eval}} \leftarrow \boldsymbol{g}_{\text{eval}} \leq 0 = \mathbb{1}_{\mathcal{F}}(\boldsymbol{\theta}^{(j)})$
 - 7 Evaluate smooth indicator function $\boldsymbol{f}_{\text{eval}} \leftarrow f(\boldsymbol{\theta}^{(j)}; s_j)$
 - 8 Compute the coefficient of variation of the ratio between the indicator function and its smooth approximation
 - $\hat{\text{cv}}^{(j)} \leftarrow \text{cv}(\boldsymbol{d}_{\text{eval}}/\boldsymbol{f}_{\text{eval}})$
 - 9 **if** $(\hat{\text{cv}}^{(j)} \leq \delta^*)$ or $(j > m)$ **then**
 - 10 | Break
 - 11 **end**
 - /* Weighting */
 - 12 Compute the weights $\boldsymbol{w}^{(j)} \leftarrow \exp(\ln \pi_{\text{pr}}(\boldsymbol{\theta}^{(j)}) - \ln \pi_{\text{bias}}(\boldsymbol{\theta}^{(j)}; \hat{\boldsymbol{v}}_j))$
 - 13 Use the level samples $\boldsymbol{\theta}^{(j)}$ to estimate the smoothing parameter s_{j+1} as per (3.36)
 - 14 Update the smooth indicator values $\boldsymbol{f}_{\text{eval}} \leftarrow f(\boldsymbol{\theta}^{(j)}; s_{j+1})$
 - 15 Compute the weights associated to the smooth indicator function $\tilde{\boldsymbol{w}}^{(j+1)} \leftarrow \boldsymbol{w}^{(j)} \boldsymbol{f}_{\text{eval}}$
 - /* Optimization */
 - 16 Find the next reference parameters $\hat{\boldsymbol{v}}_{j+1}$ by solving the stochastic optimization problem (3.35) or using a closed-form update if available.
 - 17 $j \leftarrow j + 1$
 - 18 **end**
 - 19 Compute the IS estimator $\hat{p}_{\mathcal{F}} \leftarrow \frac{1}{N} \sum_{i=1}^N d_{\text{eval}}(i) \cdot w_i^{(j)}$
- Output** : $\hat{p}_{\mathcal{F}}$
-

Similarly to SIS, the smoothing parameters are chosen adaptively to ensure that each consecutive pair of intermediate optimal biasing distributions do not differ significantly from one another. The idea is to match the sample coefficient of variation of the weights $\text{cv}(\tilde{\boldsymbol{w}}^{(j)})$ at

each level with a target value δ^* . Therefore, the smoothing parameters are estimated via the optimization

$$s_{j+1} = \arg \min_{s \in (0, s_j)} \left(\text{cv}(f(\boldsymbol{\theta}; s) \mathbf{w}^{(j)}) - \delta^* \right)^2, \quad (3.36)$$

where $\mathbf{w}^{(j)} = \{\pi_{\text{pr}}(\boldsymbol{\theta}_i) / \pi_{\text{bias}}(\boldsymbol{\theta}_i; \hat{\mathbf{v}}_j)\}_{i=1}^N$. Notice that (3.36) is solved without further evaluations of the LSF. The iCE algorithm proceeds until the cv of the ratio between the indicator function and its smooth approximation is smaller than the target δ^* . The final N samples are used to approximate the probability of failure with the IS estimator (3.11). Algorithm 4 summarizes the iCE method.

Remark 3.5. The reference parameters of the biasing distribution family in the CE and iCE methods are usually computed by a closed-form expression. This is because the optimization program is equivalent to fitting the parametric family to the set of weighted samples. In this case, maximum likelihood estimation is employed to obtain the updating formulas of the reference parameters. In this work, we employ the single Gaussian family, whose updating rule can be found in [89], and the von Mises–Fisher–Nakagami family, whose updating formulae is derived in [179].

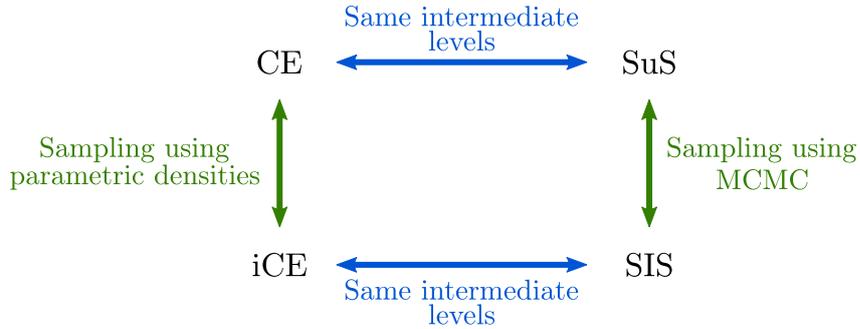


Figure 3.1: Relationships between the sequential rare event simulation methods discussed in this section.

The previous simulation methods share similar properties between one and the other. In particular, we highlight the following relations which are summarized in Figure 3.1:

- (i) The intermediate distributions used within SuS can be regarded as optimal IS distributions at each level. Both densities have the form $\pi_{\text{bias},j}^*(\boldsymbol{\theta}) \propto \mathbb{1}_{\mathcal{F}_j}(\boldsymbol{\theta})\pi_{\text{pr}}(\boldsymbol{\theta}) = \pi(\boldsymbol{\theta} \mid \mathcal{F}_j)$. Therefore, SuS uses comparable optimal IS densities as the CE method. Even the intermediate threshold levels are computed in a similar manner using order statistics of the LSF values at a given level. However, to populate the next level, SuS uses equally weighted seeds in a MCMC algorithm with stationary distribution $\pi(\boldsymbol{\theta} \mid \mathcal{F}_j)$, while CE uses weighted elite samples to fit a parametric density that performs the sampling.
- (ii) In contrast to SIS, SuS does not require resampling since all the sample seeds are defined to have the same weights. However, the moving step in both algorithms is based on MCMC. Traditionally for SuS, there is no burn-in period required since the seeds belong to the domain of the intermediate distribution, and the MCMC method is modified to always reject samples that do not belong in that domain. This is not the case for SIS, where the seeds are selected from the samples with higher weights and standard MCMC methods are applied. The differences in the moving steps for both algorithms also has to do with the fact that the intermediate densities have a different functional form and parametrization; one is based on the failure indicator function and the other one on its smooth approximation.

- (iii) SIS and iCE employ the same intermediate optimal biasing densities, which are parameterized in terms of smoothing parameters defining the approximation of the failure indicator function.

Figure 3.2 illustrates the evolution of the simulation methods (by rows) presented in this section at the first three intermediate levels (by columns). Python scripts of these algorithms are available in a [GitHub repository](#).

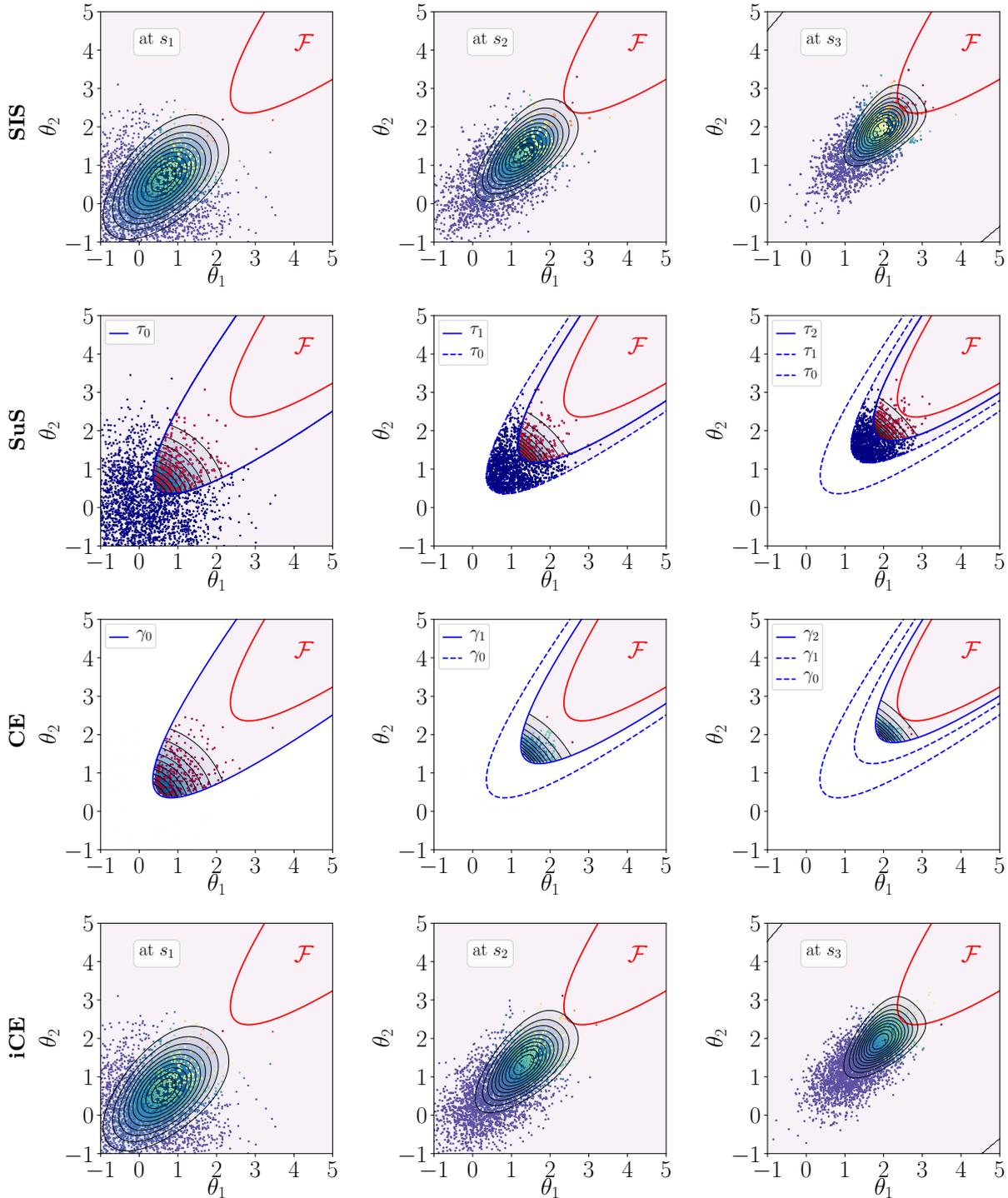


Figure 3.2: Samples and contours of the intermediate optimal biasing densities at three initial simulation levels for SIS, SuS, CE and iCE methods. The red solid line marks the target failure level (defined by LSF (3.37) with $d = 2$) and the color in the samples accounts for the weight values.

3.3. Numerical applications

We illustrate the performance of the algorithms on two examples. One involving a quadratic LSF that is independent of the dimension of the input parameter space and a second one dealing with a diffusion equation with random field coefficient.

Since the problems are moderate to high-dimensional, the CE and iCE methods are implemented with the von Mises–Fisher–Nakagami mixture, and the SIS and SuS algorithms employ an adaptive preconditioned Crank–Nicolson as MCMC sampler (cf., section 4.2). For SIS, a burn-in period of $0.1N$ is employed. When required, the target coefficient of variation of the weights is $\delta^* = 1.5$.

3.3.1. Example 1: quadratic LSF in varying dimensions

We define the quadratic LSF [177]

$$g(\boldsymbol{\theta}) = \beta + \frac{\kappa}{4} (\theta_1 - \theta_2)^2 - \frac{1}{\sqrt{d}} \sum_{i=1}^d \theta_i, \quad (3.37)$$

where $\beta = 3.5$ and the parameter $\kappa = 3$ defines the curvature of the LSF at the point in the parameter space with the largest probability density (the larger the curvature, the smaller $p_{\mathcal{F}}$). The reference probability of failure is independent of the dimension and it can be computed analytically as [68]

$$p_{\mathcal{F}} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi(-u + v^2(\kappa/2))\phi(v) \, dudv, \quad (3.38)$$

where ϕ denotes the standard Gaussian PDF. In this case, the target failure probability is $p_{\mathcal{F}} \approx 6.518 \times 10^{-5}$, and different dimensions $d = \{2, 10, 25, 50, 100, 200, 500\}$ are used to define the LSF (3.37).

We assess the accuracy of the estimate by each method for different sample sizes $N = \{500, 1000, 5000\}$. For each N , we perform $N_{\text{sim}} = 100$ independent runs of each algorithm. The results are shown in Figure 3.3. As the sample size increases, the estimates concentrate around the exact value of the failure probability. SIS and SuS can estimate the $p_{\mathcal{F}}$ for all the dimensions. CE and iCE methods can only estimate the failure probability in low dimensions when the sample size is small. However, their performance improve as the sample size increases. Note that iCE performs better than standard CE with respect to increasing dimension of the parameter space; when $N = 5000$, CE works in dimensions up to 200, while iCE is able to estimate $p_{\mathcal{F}}$ in dimensions up to 500 (possibly more, since 500 is the maximum dimension used in the study).

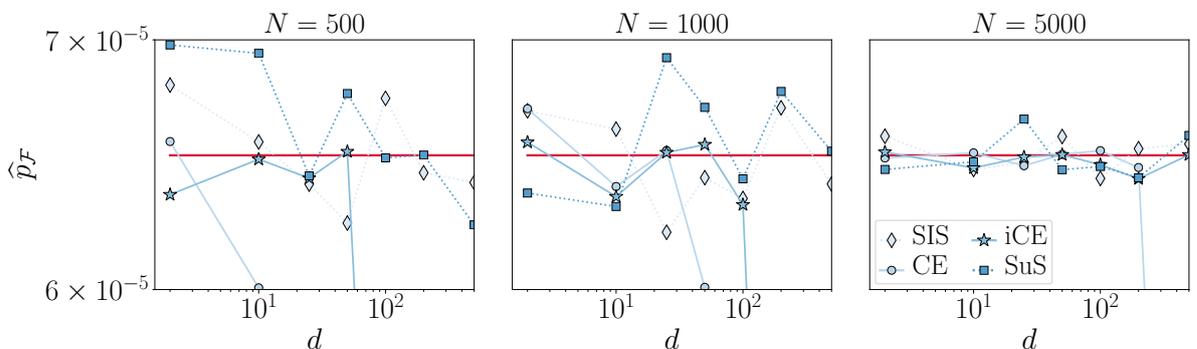


Figure 3.3: Failure probability estimate obtained by each method: for increasing dimension of the parameter space and number of samples per level. Estimated as an average of $N_{\text{sim}} = 100$ independent simulations. The solid red line marks the exact probability of failure.

We also show the values of the RMSE (3.8) in Table 3.1 (values larger than 0.6 are not included). The RMSE is controlled by the coefficient of variation of the failure probability estimate since the bias is small (or zero) for all the methods. Note that the CE and iCE methods yield smaller error when an appropriate sample size is employed. The difficulty for those methods lies in the dimension of the parameter space. Despite we are employing a suitable parametric family for sampling in high dimensions, a large number of samples per level is still required to perform a good fitting as the dimension increases. Moreover, SIS and SuS generate larger error values since they are based on MCMC sampling, which induces some sample correlation especially for SuS. The error generated by SIS is smaller than the one produced by SuS at the same sample sizes. This can be related to the extra burn-in period that the MCMC sampling within SIS requires, and also to the different parametrization of the biasing densities at the intermediate levels. As also seen for iCE, the application of the smooth failure indicator, which allows one to use information from all samples, gives better performance. In this numerical experiment, we found that the most efficient (in computational time) methods are CE and iCE, while the one that requires more running time is SIS.

Table 3.1.: RMSE of failure probability estimates obtained by each method (average of $N_{\text{sim}} = 100$). For SIS, extra $0.1N$ samples are used due to burn-in.

	d	N		
		500	1000	5000
SIS	50	0.26	0.24	0.10
	100	0.31	0.19	0.09
	200	0.32	0.24	0.10
	500	0.33	0.25	0.09
CE	50	–	0.31	0.06
	100	–	–	0.06
	200	–	–	0.12
	500	–	–	–
iCE	50	0.25	0.17	0.06
	100	0.59	0.17	0.06
	200	–	–	0.07
	500	–	–	0.09
SuS	50	0.51	0.39	0.16
	100	0.49	0.33	0.14
	200	0.46	0.40	0.18
	500	0.52	0.37	0.17

3.3.2. Example 2: one-dimensional diffusion problem

Consider the stochastic linear elliptic equation on a one-dimensional spatial domain $D = [0, 1]$:

$$\frac{d}{dx} \left(\kappa(x, \omega) \frac{du(x, \omega)}{dx} \right) = -J(x), \quad (3.39)$$

with Dirichlet boundary condition at $u(1, \omega) = 1$ and a random Neumann boundary condition at $x = 0$:

$$\kappa(x, \omega) \frac{du(x, \omega)}{dx} \Big|_{x=0} = -F(\omega), \quad (3.40)$$

where the flux F is Gaussian distributed with mean $\mu_F = -2$ and variance $\sigma_F^2 = 0.5$. This example problem is taken from [160].

The source term is defined as the superposition of four weighted Gaussian plumes with standard width $\sigma_J = 0.05$. The plumes have equal strengths $w = 0.8$ and they are centered at locations $\boldsymbol{\mu}_J = [0.2, 0.4, 0.6, 0.8]$:

$$J(x) = w \sum_{i=1}^4 \mathcal{N}(x; \mu_J(i), \sigma_J^2). \quad (3.41)$$

The hydraulic conductivity is represented with a lognormal random field, where the underlying Gaussian field κ' has mean $\mu_{\kappa'} = 1$ and variance $\sigma_{\kappa'}^2 = 0.3$. The covariance kernel of the Gaussian field is exponential with correlation lengths $\ell = \{0.1, 0.9\}$. We employ the KL expansion to represent the conductivity field using truncation orders $k = \{5, 50, 100, 200\}$. The source term, integrated source term, and some random field realizations are shown in Figure 3.4.

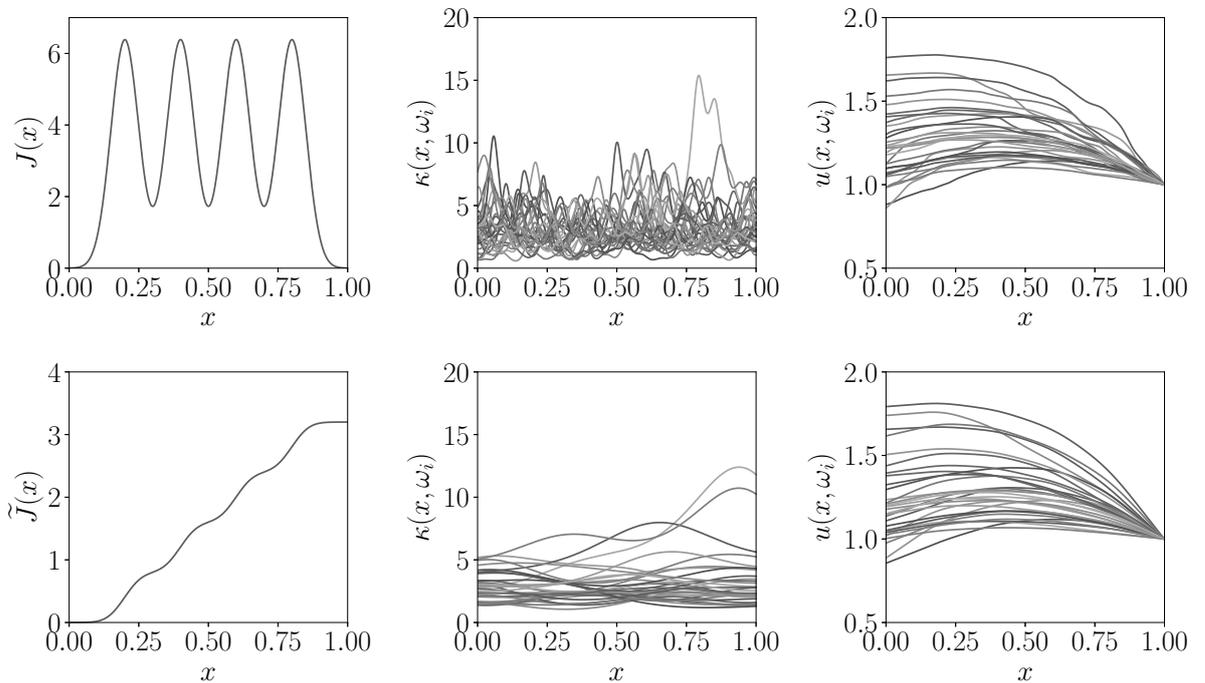


Figure 3.4: Source term, integrated source term, and some realizations of the hydraulic conductivity and pressure head fields using an exponential kernel with $\ell = 0.1$ (1st row) and $\ell = 0.9$ (2nd row).

The boundary value problem (3.39) is solved semi-analytically as follows

$$\frac{d}{dx} \left(\kappa(x) \frac{du(x)}{dx} \right) = -J(x) \quad (3.42a)$$

$$\kappa(x) \frac{du(x)}{dx} = - \int_0^x J(t) dt + c_1 = -\tilde{J}(x) + c_1 \quad (3.42b)$$

$$u(x) = \int_0^x \left[\frac{-\tilde{J}(t) + c_1}{\kappa(t)} \right] dt + c_2. \quad (3.42c)$$

Applying the Neumann boundary condition (at $x = 0$) to (3.42b), we get $c_1 = -F$. And applying the Dirichlet boundary condition (at $x = 1$) to (3.42c), we obtain $c_2 = 1 - \int_0^1 (-\tilde{J}(t) - F)/\kappa(t) dt$, which can be solved numerically for a given realization of $\kappa(t)$; here $\tilde{J}(t)$ can also be a numerical integration of (3.41). Hence, the following equation is used to compute the response

pressure head:

$$u(x, \omega) = \int_0^x \left[\frac{-\tilde{J}(t) - F}{\kappa(t, \omega)} \right] dt + 1 - \int_0^1 \left[\frac{-\tilde{J}(t) - F}{\kappa(t, \omega)} \right] dt. \quad (3.43)$$

We define the LSF by considering that the failure of the system occurs when the maximum pressure head exceeds a given admissible threshold $u^* = 2.7$:

$$g(\boldsymbol{\theta}) = u^* - \max(u(x; \boldsymbol{\theta})), \quad (3.44)$$

where the uncertain parameter vector $\boldsymbol{\theta} = [\boldsymbol{\theta}^{(F)}, \boldsymbol{\theta}^{(KL)}] \in \mathbb{R}^{k+1}$, accounts for the flux random variable and the KL coefficients used for the conductivity random field representation.

The forward equation (3.43) is evaluated on the domain D , which is discretized into 501 equally-spaced points. The accuracy of the probability estimator by each simulation method is evaluated at sample sizes $N = \{1000, 5000\}$. For each N , we perform $N_{\text{sim}} = 50$ independent runs of each algorithm. As in the previous example, extra $0.1N$ samples are used for SIS due to burn-in.

The failure probability estimates and associated coefficients of variation are shown in Table 3.2, for the random field represented with the correlation length $\ell = 0.1$. In this case, the LSF (3.44) depends on the dimension of the input parameter space. The probability of failure increases with the truncation order of the KL expansion. After representing sufficient variability of the random field (around 98% with $k = 101$ terms), the rate at which the failure probability increases is less pronounced. The performance of the methods is similar to the one obtained in the example of subsection 3.3.1.

Table 3.2.: Statistics of the failure probability estimates using $\ell = 0.1$ obtained by each method.

	k	$\hat{p}_{\mathcal{F}}$		cv	
		$N = 1000$	$N = 5000$	$N = 1000$	$N = 5000$
SIS	5	1.82×10^{-5}	1.71×10^{-5}	0.21	0.10
	50	3.13×10^{-5}	3.28×10^{-5}	0.25	0.11
	100	3.38×10^{-5}	3.41×10^{-5}	0.29	0.09
	200	3.38×10^{-5}	3.36×10^{-5}	0.32	0.09
CE	5	1.61×10^{-5}	1.68×10^{-5}	0.16	0.05
	50	3.03×10^{-5}	3.24×10^{-5}	0.22	0.07
	100	–	3.34×10^{-5}	–	0.07
	200	–	3.44×10^{-5}	–	0.10
iCE	5	1.71×10^{-5}	1.71×10^{-5}	0.07	0.04
	50	3.21×10^{-5}	3.22×10^{-5}	0.12	0.03
	100	3.36×10^{-5}	3.32×10^{-5}	0.11	0.05
	200	3.33×10^{-5}	3.41×10^{-5}	0.21	0.05
SuS	5	1.64×10^{-5}	1.67×10^{-5}	0.33	0.14
	50	3.09×10^{-5}	3.23×10^{-5}	0.30	0.13
	100	3.69×10^{-5}	3.31×10^{-5}	0.39	0.15
	200	3.29×10^{-5}	3.23×10^{-5}	0.31	0.13

Similar results are shown in Table 3.3 for the random field represented with a correlation length $\ell = 0.9$. Again, the probability of failure increases with the truncation order of the KL expansion. However, the increment scales down faster than the case with small correlation length. This is because less terms are required to fully represent the random field (98% with $k = 11$ terms). Note the effect of the correlation length on the magnitude of the failure probabilities. As the correlation length increases the probability of failure becomes smaller; the order of magnitude

changes from 10^{-5} to 10^{-3} . In this case, the increase in spatial correlation drives most of the prior probability mass towards the failure domain (i.e., the failure event is less rare).

Table 3.3.: Statistics of the failure probability estimates using $\ell = 0.9$ obtained by each method.

	k	$\hat{p}_{\mathcal{F}}$		cv	
		$N = 1000$	$N = 5000$	$N = 1000$	$N = 5000$
SIS	5	1.67×10^{-3}	1.66×10^{-3}	0.126	0.079
	50	1.69×10^{-3}	1.74×10^{-3}	0.169	0.086
	100	1.74×10^{-3}	1.76×10^{-3}	0.202	0.077
	200	1.73×10^{-3}	1.72×10^{-3}	0.176	0.077
CE	5	1.63×10^{-3}	1.65×10^{-3}	0.069	0.024
	50	1.74×10^{-3}	1.73×10^{-3}	0.133	0.038
	100	1.72×10^{-3}	1.74×10^{-3}	0.191	0.045
	200	–	1.74×10^{-3}	–	0.054
iCE	5	1.64×10^{-3}	1.63×10^{-3}	0.047	0.022
	50	1.72×10^{-3}	1.72×10^{-3}	0.061	0.024
	100	1.73×10^{-3}	1.73×10^{-3}	0.081	0.030
	200	1.74×10^{-3}	1.74×10^{-3}	0.079	0.027
SuS	5	1.62×10^{-3}	1.66×10^{-3}	0.194	0.102
	50	1.66×10^{-3}	1.71×10^{-3}	0.211	0.099
	100	1.70×10^{-3}	1.73×10^{-3}	0.177	0.099
	200	1.60×10^{-3}	1.73×10^{-3}	0.203	0.094

We conclude by pointing out that the probability of failure can change significantly, if the spatial correlation takes negative values (see for example [98]). Although it is possible to employ kernel functions that allow for both, positive and negative correlation, in spatial variation it is common to consider positive correlation. Such assumption is usually employed for homogeneous material properties since there is an underlying uniformity in which the random field realizations cannot be completely different at two spatial locations. For random heterogeneous media, negative correlation models can be allowed, and in such situations the effect on the reliability of the system can be certainly different. Another example where the correlation values change, occurs when observational data is used to identify the random field. This is because the updating process naturally modifies the correlation structure of the field, and as a result the associated failure probabilities change. This type of analysis is called *reliability updating* and will be discussed in Chapter 4.

Bayesian analysis of engineering models

“Bayes’ theorem is to the theory of probability what Pythagoras’ theorem is to geometry”.

Harold Jeffreys, 1891-1989 [125, p.19].

Computational models of engineering systems require specification of both, system input (e.g., boundary conditions and applied external actions) and system parameters (e.g., material or geometric properties). These quantities are in practice inherently random or known with imprecision. Hence, physical systems are typically subjected to *uncertainty*.

The integration of data into engineering models is key to reducing the uncertainty and performing accurate predictions. The combination of the observations with the computational model can provide information about the uncertain inputs and parameters. This represents an *inverse problem* in which the idea is to discover from a set of observations, the causal elements or parameters that generated them. The stability of the inverse problem is mainly controlled by the dimension of the parameter space, the structure of the mathematical model and the observations, which are in most cases limited and noisy. As a result, inverse problems are prone to be ill-posed (according to Hadamard [103]). This means that different parameter values are consistent with the data (uniqueness), or that the parameters cannot be identified at all (existence), or that the solution is highly sensitive to small perturbations in the data (stability).

Classical regularization methods, such as Tikhonov [230], truncated singular value decomposition [107], or total variation [122] are oftentimes used to solve inverse problems. These methods are deterministic and incorporate regularizers or penalty functions as stabilization procedure. In several cases, the solution is computed by optimization methods, or in more standard scenarios the problem is expressed by a least-squares task. Another approach to inverse problems is Bayesian statistical methods. They provide a tool to regularize the inverse problem by incorporating a probabilistic description of the model parameters that combines prior information with observations [20, 131]. One of the main advantages is that the potential uncertainty in the parameter values can also be quantified.

We adopt the Bayesian approach for the solution of inverse problems and start this chapter by summarizing fundamental concepts on this theory. Since a Bayesian analysis not only involves retrieving the uncertain parameters, but also rare event simulation tasks, we also discuss the updating of failure probabilities. Thereafter, we present some numerical methods for the solution of Bayesian inverse problems. We wrap up with some numerical examples involving PDE models with random field coefficients.

4.1. The Bayesian approach to inverse problems

Suppose that the goal is to perform inference about a random quantity in order to improve the predictive capacity of a numerical model. For instance, the model can be a finite element solution of a PDE representing the response of a structural system, and the associated random parameters could be the Poisson ratio or the Young’s modulus of a constitutive material. We want our model to learn the parameters using data or observations characterizing the system

behavior. The task of updating the model using observation data represents an inverse problem.

We begin by considering the forward problem $\mathbf{y} = G(\boldsymbol{\theta})$, where G is a *solution operator* expressing the relationship between the uncertain input parameters $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^d$ and the model response \mathbf{y} . Moreover, let the *forward response operator* be a function from the parameter space Θ to the *data space* $\mathcal{Y} \subseteq \mathbb{R}^m$, i.e., $\mathcal{G} = \mathcal{O} \circ G : \Theta \rightarrow \mathcal{Y}$. The operator \mathcal{G} is also called parameter-to-observation map since it is defined as the composition of the solution operator G with an *observation operator* \mathcal{O} . Note that \mathcal{G} provides a link between the parameter and data spaces, and it describes how the observational data can be computed given the parameters $\boldsymbol{\theta}$.

The observed data $\tilde{\mathbf{y}} \in \mathcal{Y}$ frequently contains noise. In classical inverse problems, this is usually modeled as additive and mutually independent from the uncertain parameters. Therefore, the data can be represented as a perturbed evaluation of the forward response operator

$$\tilde{\mathbf{y}} := \mathcal{G}(\boldsymbol{\theta}_{\text{true}}) + \boldsymbol{\eta}, \quad (4.1)$$

where $\boldsymbol{\theta}_{\text{true}} \in \Theta$ is the unknown and deterministic *true parameter* vector, $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\text{obs}})$ is the observation noise assumed to be Gaussian distributed with mean zero and non-singular covariance matrix $\boldsymbol{\Sigma}_{\text{obs}} \in \mathbb{R}^{m \times m}$. We remark that other types of noise models also exist in practice; for example, a Poisson multiplicative noise is considered in image reconstruction [107].

As briefly mentioned in the chapter introduction, the inverse problem of finding the parameter $\boldsymbol{\theta}_{\text{true}}$ given the noisy relation (4.1) is generally ill-posed. Bayesian statistical methods offer a framework that integrates the observations with prior information, providing a mechanism of stabilization [131]. In this case, instead of inferring a deterministic $\boldsymbol{\theta}_{\text{true}}$, we assume that the parameter is a random vector $\boldsymbol{\theta}$ distributed according to an initial/prior probability distribution. Thereafter, the idea is to update the (*prior*) probability distribution of the parameters by including information about the numerical model and observed data (*likelihood*). Solving the Bayesian inverse problem then amounts to estimate or characterize the updated (*posterior*) distribution.

In engineering, we are typically interested in models where G implies the solution of a stochastic PDE with random field coefficients. In principle, the solution operator is defined as $G : L^2(D) \rightarrow L^2(D)$ since both, input and output, are required to be square-integrable realizations of two different random fields defined on a physical domain D . The Bayes' theorem in this functional setting is interpreted as the Radon–Nikodym derivative of the posterior probability measure with respect to the prior probability measure [225, 60]. Alternatively, we can parameterize the random field using the KL expansion (2.29), such that the domain of the solution operator is replaced by the finite-dimensional parameter space $\Theta \subseteq \mathbb{R}^k$. In this case, the solution operator is a map $G : \Theta \rightarrow L^2(D)$, and the observation operator is given as $\mathcal{O} : L^2(D) \rightarrow \mathcal{Y}$.

As in Chapter 3, we consider a canonical probability space $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k), \mathbb{P})$, such that $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^k$. The random vector $\boldsymbol{\theta}$ is assumed to be distributed according to an initial *prior* probability density $\pi_{\text{pr}}(\boldsymbol{\theta}) = d\mathbb{P}/d\lambda$ with respect to the Lebesgue measure λ on \mathbb{R}^k . Intuitively, the role of the prior (as regularization) is to allocate lower probability to parameter values that are not “expected”, and higher probability to parameters that are “expected”. It has been shown that the prior must have sufficiently heavy tails to achieve stability in the Bayesian inverse problem; the conditions are discussed in [115]. Moreover, consider a distribution on the data space \mathcal{Y} with well-defined probability density $\pi(\cdot | \boldsymbol{\theta})$, conditioned on a realization $\boldsymbol{\theta}$ of the uncertain parameters. Such density follows from the statistical model of the observation noise assumed in (4.1), linking the forward operator and data. The *likelihood function* is defined from this conditional density with fixed argument equal to the measurement data, i.e., $L(\boldsymbol{\theta}; \tilde{\mathbf{y}}) := \pi(\tilde{\mathbf{y}} | \boldsymbol{\theta})$. The updated belief about $\boldsymbol{\theta}$ is represented by the *posterior* probability density $\pi_{\text{pos}}(\boldsymbol{\theta} | \tilde{\mathbf{y}})$. Through Bayes' theorem this conditional density is defined as [19, 141]

$$\pi_{\text{pos}}(\boldsymbol{\theta} \mid \tilde{\mathbf{y}}) = \frac{1}{Z_{\tilde{\mathbf{y}}}} \mathbf{L}(\boldsymbol{\theta}; \tilde{\mathbf{y}}) \pi_{\text{pr}}(\boldsymbol{\theta}), \quad (4.2)$$

where $Z_{\tilde{\mathbf{y}}} = \int_{\Theta} \mathbf{L}(\boldsymbol{\theta}; \tilde{\mathbf{y}}) \pi_{\text{pr}}(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta}$ is the normalizing constant of $\pi_{\text{pos}}(\boldsymbol{\theta} \mid \tilde{\mathbf{y}})$, called the marginal likelihood or *model evidence* and it gives information about the plausibility of the assumed model.

Since we are dealing with Gaussian or related random fields, the formulation considers multivariate Gaussian priors $\boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\mu}_{\text{pr}}, \boldsymbol{\Sigma}_{\text{pr}})$, where $\boldsymbol{\mu}_{\text{pr}}$ and $\boldsymbol{\Sigma}_{\text{pr}}$ are the prior mean and covariance. Hence, the posterior (4.2) can be written as

$$\pi_{\text{pos}}(\boldsymbol{\theta} \mid \tilde{\mathbf{y}}) \propto \exp\left(-\frac{1}{2} \underbrace{\left\| \boldsymbol{\Sigma}_{\text{obs}}^{-1/2} (\tilde{\mathbf{y}} - \mathcal{G}(\boldsymbol{\theta})) \right\|_{\mathbf{y}}^2}_{\Phi(\boldsymbol{\theta}; \tilde{\mathbf{y}})} - \frac{1}{2} \left\| \boldsymbol{\Sigma}_{\text{pr}}^{-1/2} (\boldsymbol{\theta} - \boldsymbol{\mu}_{\text{pr}}) \right\|_{\Theta}^2\right), \quad (4.3)$$

where $\Phi(\boldsymbol{\theta}; \tilde{\mathbf{y}}) = -\ln \mathbf{L}(\boldsymbol{\theta}; \tilde{\mathbf{y}})$ is the so-called *potential* (also data-misfit function or negative log-likelihood), and it is often employed for numerical stability when performing the Bayesian inversion.

Remark 4.1. We can use the KL expansion to further reduce (4.3), since the prior distribution has mean and covariance are given by $\boldsymbol{\mu}_{\text{pr}} = \mathbf{0}$ and $\boldsymbol{\Sigma}_{\text{pr}} = \mathbf{I}_k$ ($\mathbf{I}_k \in \mathbb{R}^{k \times k}$ denotes the identity matrix); the dimension k is in this case the truncation order of the expansion. Note that the information about the second-order properties of the random field (mean and covariance structure) enters directly in the definition of the likelihood function via the forward response operator.

Closed-form expressions for the posterior (4.2) are only available in some particular cases. For instance, models involving *conjugate priors* [210, 193], or non-Gaussian priors given by nonlinear transformations of Gaussians [61]. This class of priors can be considered as *informative*, and oftentimes arise in engineering applications where there is potential information about the parameters beforehand. Other *non-conjugate models* include priors that are non-informative such as, Laplace, Jeffrey, and matching priors (see [193] for a discussion). Furthermore, extra layers of prior models arise in Bayesian hierarchical applications. An advantage of this strategy is that the posterior distribution becomes less sensitive to the prior choices as the parameter hierarchies increase [193]. This is especially relevant in the context of random fields, where several parameters control the field representation. Alternatively, one can also explore the effects of the prior choices on the posterior solution by performing parameter studies. This allows one to identify practical guidelines and relations that support prior specification for random fields.

The computation of the posterior (4.2) generally requires MC-based sampling procedures. Obtaining samples from the posterior using standard MC techniques can be difficult, if the normalizing factor $Z_{\tilde{\mathbf{y}}}$ is cumbersome to compute. MCMC is another collection of methods that do not require the knowledge of this normalizing constant. In this case, it is only necessary to know the shape of the posterior distribution. In the next section, we describe two important algorithms that are oftentimes the core of many sequential approaches used to perform a sample characterization of the posterior.

We conclude this section by defining well-known summary statistics utilized in Bayesian computations. The posterior predictive distribution is also included due to its relevance when using the updated model to estimate future system states:

- The *posterior mean* (PM) is the optimal estimator of $\boldsymbol{\theta}_{\text{true}}$ in the MSE sense [193]

$$\boldsymbol{\theta}_{\text{PM}}(\tilde{\mathbf{y}}) = \int_{\Theta} \boldsymbol{\theta} \pi_{\text{pos}}(\boldsymbol{\theta} \mid \tilde{\mathbf{y}}) \mathrm{d}\boldsymbol{\theta}. \quad (4.4)$$

- The *maximum a posteriori estimator* (MAP) is the Bayesian analogue of the frequentist maximum likelihood estimator (MLE) [193]

$$\boldsymbol{\theta}_{\text{MAP}}(\tilde{\mathbf{y}}) = \arg \max_{\boldsymbol{\theta} \in \Theta} \pi_{\text{pos}}(\boldsymbol{\theta} \mid \tilde{\mathbf{y}}); \quad (4.5)$$

note that for (4.3), the MAP estimator is given by

$$\boldsymbol{\theta}_{\text{MAP}}(\tilde{\mathbf{y}}) = \arg \min_{\boldsymbol{\theta} \in \Theta} \Phi(\boldsymbol{\theta}; \tilde{\mathbf{y}}) + \frac{1}{2} \left\| \boldsymbol{\Sigma}_{\text{pr}}^{-1/2} (\boldsymbol{\theta} - \boldsymbol{\mu}_{\text{pr}}) \right\|_{\Theta}^2, \quad (4.6)$$

which can be seen as a classical regularization of the inverse problem (generalized Tikhonov) with stabilization term given by the prior.

- *Credible intervals*, denoted $\text{CI}(\tilde{\mathbf{y}})$, where the unobserved parameter values belong to with probability $\mathbb{P}[\boldsymbol{\theta} \in \text{CI}(\tilde{\mathbf{y}}) \mid \tilde{\mathbf{y}}] = 1 - \alpha$. That is, the probability that the CI bounds contain the true parameter is $1 - \alpha$, where α is a user-defined threshold. Credible intervals support statements, such as “based on the data, there is a $1 - \alpha$ probability that the true parameter value is in $\text{CI}(\tilde{\mathbf{y}})$ ”.
- The *posterior predictive density* accounts for possible unobserved values \mathbf{y}_{new} conditional on the observed data $\tilde{\mathbf{y}}$,

$$\pi_{\text{pred}}(\mathbf{y}_{\text{new}} \mid \tilde{\mathbf{y}}) = \int_{\Theta} \pi(\mathbf{y}_{\text{new}} \mid \boldsymbol{\theta}, \tilde{\mathbf{y}}) \pi_{\text{pos}}(\boldsymbol{\theta} \mid \tilde{\mathbf{y}}) d\boldsymbol{\theta}; \quad (4.7)$$

under the assumption that the observed and unobserved data are conditional independent given $\boldsymbol{\theta}$, the posterior predictive distribution can be further simplified as

$$\pi_{\text{pred}}(\mathbf{y}_{\text{new}} \mid \tilde{\mathbf{y}}) = \int_{\Theta} L(\boldsymbol{\theta}; \mathbf{y}_{\text{new}}) \pi_{\text{pos}}(\boldsymbol{\theta} \mid \tilde{\mathbf{y}}) d\boldsymbol{\theta}. \quad (4.8)$$

4.2. General comments on Markov chain Monte Carlo

Markov chain Monte Carlo (MCMC) is essentially MC integration using Markov chains. MC draws samples for the required distribution and then forms sample averages to approximate expectations. MCMC draws these samples using a Markov chain that has the target posterior distribution as its stationary distribution.

Let $\pi_{\text{pos}}(\boldsymbol{\theta} \mid \tilde{\mathbf{y}})$ be the target distribution of interest. The purpose of a MCMC algorithm is to generate a collection of states according to $\pi_{\text{pos}}(\boldsymbol{\theta} \mid \tilde{\mathbf{y}})$. When the computation of the model evidence is intractable, the posterior density is only known up to its scaling constant, that is $\pi_{\text{pos}}(\boldsymbol{\theta} \mid \tilde{\mathbf{y}}) \propto \pi_{\text{pr}}(\boldsymbol{\theta}) L(\boldsymbol{\theta}; \tilde{\mathbf{y}}) = \pi(\boldsymbol{\theta})$. We refer to $\pi(\boldsymbol{\theta})$ as the *target function*. MCMC generates a Markov process that asymptotically reaches a unique stationary distribution $\pi(\boldsymbol{\theta}) \propto \pi_{\text{pos}}(\boldsymbol{\theta} \mid \tilde{\mathbf{y}})$.

A Markov process has a unique stationary (invariant) distribution $\pi(\boldsymbol{\theta})$ when is reversible (equilibrium in probability/detailed balance) and ergodic (aperiodic and irreducible). Furthermore, Markov processes are uniquely defined by the transition kernel modeling the probability of transitioning from any given state $\boldsymbol{\theta}$ to any other $\boldsymbol{\theta}'$. Classical MCMC algorithms must satisfy the detailed balance condition [87]

$$\int_A \pi(\boldsymbol{\theta}) P(\boldsymbol{\theta}, \boldsymbol{\theta}') d\boldsymbol{\theta} = \int_B \pi(\boldsymbol{\theta}') P(\boldsymbol{\theta}', \boldsymbol{\theta}) d\boldsymbol{\theta}', \quad (4.9)$$

where $\boldsymbol{\theta} \in A$, $\boldsymbol{\theta}' \in B$, and $A, B \subset \Theta$, and $P(\cdot, \cdot)$ denotes a Markov transition kernel. Different approaches are proposed to generate kernels P that ensure (4.9). We discuss the standard

Metropolis–Hastings and the preconditioned Crank–Nicolson algorithms.

Remark 4.2. There is an increasing interest in MCMC algorithms that are not reversible, so-called *non-reversible MCMC*. The main idea is to design Markov chains that avoid transitions that fall back to the state from which the chain just came, by incorporating a sense of direction. As a result, these algorithms have better performance in terms of mixing behavior, convergence rate and reduction of the asymptotic variance. These include for example the non-reversible Metropolis–Hastings [29] and the Bouncy Particle Sampler [40].

4.2.1. Metropolis–Hastings

In Metropolis–Hastings (MH), the condition (4.9) is satisfied by separating the transition into two stages: the proposal and the acceptance/rejection steps. These are represented respectively by the proposal distribution $q(\boldsymbol{\theta} \mid \boldsymbol{\theta}')$ which accounts for the conditional probability of $\boldsymbol{\theta}$ given the proposed state $\boldsymbol{\theta}'$, and the acceptance distribution $\alpha(\boldsymbol{\theta}, \boldsymbol{\theta}')$ which expresses the conditional probability of accepting the proposed state $\boldsymbol{\theta}'$. The transition kernel can be written as a product of them [87]

$$P(\boldsymbol{\theta} \mid \boldsymbol{\theta}') = \underbrace{\int_B q(\boldsymbol{\theta} \mid \boldsymbol{\theta}') \alpha(\boldsymbol{\theta}, \boldsymbol{\theta}') \, d\boldsymbol{\theta}'}_{\text{acceptance}} + \underbrace{\mathbb{1}(\boldsymbol{\theta} \in B) \int_{\Theta} q(\boldsymbol{\theta} \mid \boldsymbol{\theta}') (1 - \alpha(\boldsymbol{\theta}, \boldsymbol{\theta}')) \, d\boldsymbol{\theta}'}_{\text{rejection}}. \quad (4.10)$$

Substituting (4.10) into (4.9), yields

$$\int_A \pi(\boldsymbol{\theta}) \int_B q(\boldsymbol{\theta} \mid \boldsymbol{\theta}') \alpha(\boldsymbol{\theta}, \boldsymbol{\theta}') \, d\boldsymbol{\theta}' \, d\boldsymbol{\theta} = \int_B \pi(\boldsymbol{\theta}') \int_A q(\boldsymbol{\theta}' \mid \boldsymbol{\theta}) \alpha(\boldsymbol{\theta}', \boldsymbol{\theta}) \, d\boldsymbol{\theta} \, d\boldsymbol{\theta}', \quad (4.11)$$

which can be written in compact form as

$$\int_{A \times B} \pi(\boldsymbol{\theta}) q(\boldsymbol{\theta} \mid \boldsymbol{\theta}') \alpha(\boldsymbol{\theta}, \boldsymbol{\theta}') \, d\boldsymbol{\theta} = \int_{A \times B} \pi(\boldsymbol{\theta}') q(\boldsymbol{\theta}' \mid \boldsymbol{\theta}) \alpha(\boldsymbol{\theta}', \boldsymbol{\theta}) \, d\boldsymbol{\theta}'. \quad (4.12)$$

The equality (4.12) holds if

$$\pi(\boldsymbol{\theta}) q(\boldsymbol{\theta} \mid \boldsymbol{\theta}') \alpha(\boldsymbol{\theta}, \boldsymbol{\theta}') = \pi(\boldsymbol{\theta}') q(\boldsymbol{\theta}' \mid \boldsymbol{\theta}) \alpha(\boldsymbol{\theta}', \boldsymbol{\theta}) \quad (4.13a)$$

$$\frac{q(\boldsymbol{\theta} \mid \boldsymbol{\theta}') \alpha(\boldsymbol{\theta}, \boldsymbol{\theta}')}{q(\boldsymbol{\theta}' \mid \boldsymbol{\theta}) \alpha(\boldsymbol{\theta}', \boldsymbol{\theta})} = \frac{\pi(\boldsymbol{\theta}')}{\pi(\boldsymbol{\theta})} \quad (4.13b)$$

$$\frac{\alpha(\boldsymbol{\theta}, \boldsymbol{\theta}')}{\alpha(\boldsymbol{\theta}', \boldsymbol{\theta})} = \frac{\pi(\boldsymbol{\theta}') q(\boldsymbol{\theta}' \mid \boldsymbol{\theta})}{\pi(\boldsymbol{\theta}) q(\boldsymbol{\theta} \mid \boldsymbol{\theta}')}. \quad (4.13c)$$

An acceptance mechanism that fulfills the condition in (4.13c) is proposed by Hastings [112] (based on Metropolis et al. [167], which uses a symmetric proposal distribution)

$$\alpha(\boldsymbol{\theta}, \boldsymbol{\theta}') = \min \left(1, \frac{\pi(\boldsymbol{\theta}') q(\boldsymbol{\theta}' \mid \boldsymbol{\theta})}{\pi(\boldsymbol{\theta}) q(\boldsymbol{\theta} \mid \boldsymbol{\theta}')} \right). \quad (4.14)$$

At each iteration t of the MH algorithm, the next state of the Markov chain $\boldsymbol{\theta}_{t+1}$ is chosen by sampling a candidate point $\boldsymbol{\theta}'$ from the proposal distribution $q(\cdot \mid \boldsymbol{\theta}_t)$. The candidate point $\boldsymbol{\theta}'$ is accepted when a standard uniform random number is less than or equal to the acceptance (4.14), otherwise the candidate point is rejected. If the candidate is accepted, the next state becomes $\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}'$. If the candidate is rejected, the chain does not move and $\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t$. Applying this procedure, the entire collection of points form a Markov chain that represents a set of samples from the target function π .

One of the difficulties with MH is the potentially strong dependence in the Markov chain

states (samples). A standard way to reduce this effect is by running the Markov chain for many iterations until the initial state value is forgotten. This requires discarding a considerable amount of initial samples; the number of discarded samples is known as *burn-in*. Moreover, the performance of MH deteriorates with increasing dimension of the parameter space. In this case, the acceptance probability at each step becomes very small as the dimension increases, resulting in slow convergence rates and a large number of repeated samples. Several algorithms have been designed to improve the performance of MCMC in high dimensions; they are usually referred to as *dimension-independent* MCMC algorithms. The standard technique is the preconditioned Crank–Nicolson method [56, 199], which can also be seen as the conditional sampling algorithm proposed in [177].

4.2.2. Preconditioned Crank–Nicolson (conditional sampling)

A main requirement for a MCMC algorithm to be dimension-independent is that of being well-defined in the function spaces. Sampling from posterior distributions in high dimensions can be achieved through path reconstructions of stochastic differential equations. These are designed to be invariant with respect to reference (prior) or target (posterior) densities. Given a target distribution proportional to the posterior $\pi(\boldsymbol{\theta}) \propto \pi_{\text{pos}}(\boldsymbol{\theta} \mid \tilde{\mathbf{y}}) = \exp(-\Psi(\boldsymbol{\theta}; \tilde{\mathbf{y}}))$, the overdamped Langevin dynamic equation [104, 71],

$$\frac{d\boldsymbol{\theta}_t}{dt} = -\nabla\Psi(\boldsymbol{\theta}; \tilde{\mathbf{y}}) + \sqrt{2} \frac{dW_t}{dt} \quad \text{with initial state } \boldsymbol{\theta}_0, \quad (4.15)$$

has invariant distribution $\pi(\boldsymbol{\theta})$, i.e., if $\boldsymbol{\theta}_0 \sim \pi(\cdot)$ then $\boldsymbol{\theta}_t \sim \pi(\cdot)$ for all $t > 0$ [60, p.382]. Furthermore, for a symmetric and positive semi-definite preconditioner \mathbf{P} , (4.15) can be expressed as

$$\frac{d\boldsymbol{\theta}_t}{dt} = -\mathbf{P}\nabla\Psi(\boldsymbol{\theta}; \tilde{\mathbf{y}}) + \sqrt{2\mathbf{P}} \frac{dW_t}{dt}, \quad (4.16)$$

where W_t is a standard Brownian motion process and $\Psi(\boldsymbol{\theta}; \tilde{\mathbf{y}})$ is the logarithm of the posterior density:

$$\Psi(\boldsymbol{\theta}; \tilde{\mathbf{y}}) = \frac{1}{2} \left\| \boldsymbol{\Sigma}_{\text{pr}}^{-1/2} \boldsymbol{\theta} \right\|_{\boldsymbol{\Theta}}^2 + \Phi(\boldsymbol{\theta}; \tilde{\mathbf{y}}) + \ln Z_{\tilde{\mathbf{y}}}, \quad (4.17a)$$

$$\text{with gradient } \nabla\Psi(\boldsymbol{\theta}; \tilde{\mathbf{y}}) = \boldsymbol{\Sigma}_{\text{pr}}^{-1} \boldsymbol{\theta} + \nabla\Phi(\boldsymbol{\theta}; \tilde{\mathbf{y}}). \quad (4.17b)$$

Replacing the gradient in (4.17b) into (4.16) yields,

$$\frac{d\boldsymbol{\theta}_t}{dt} = -\mathbf{P} \left(\boldsymbol{\Sigma}_{\text{pr}}^{-1} \boldsymbol{\theta} + \nabla\Phi(\boldsymbol{\theta}; \tilde{\mathbf{y}}) \right) + \sqrt{2\mathbf{P}} \frac{dW_t}{dt}. \quad (4.18)$$

Although Brownian motion paths are not differentiable, the derivative is interpreted in the sense of distributions; in this case, the derivative of the process is a stationary white noise, denoted ξ_0 [102]. The preconditioned stochastic differential equation in (4.18) can be solved using a numerical discretization algorithm. For instance, the Crank–Nicolson scheme [186] computes the next location $\boldsymbol{\theta}'$ of the process after time $\Delta > 0$ through the formula,

$$\boldsymbol{\theta}' - \boldsymbol{\theta} = \left(-\mathbf{P}\boldsymbol{\Sigma}_{\text{pr}}^{-1} \left(\frac{\boldsymbol{\theta}' + \boldsymbol{\theta}}{2} \right) - \gamma\mathbf{P}\nabla\Phi(\boldsymbol{\theta}; \tilde{\mathbf{y}}) \right) \Delta + \sqrt{2\mathbf{P}\Delta} \xi_0 \quad \text{with } \xi_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d), \quad (4.19)$$

here the Crank–Nicolson approximation is applied on the linear part of the drift, and the tuning parameter $\gamma \in [0, 1]$ is included in the nonlinear part to distinguish between random walk and Langevin proposals [25]. Despite there is no effect of the preconditioner on the invariance property of the simulated process, it influences the convergence to the target density [60]. Therefore, by considering $\gamma = 0$ in the case when the preconditioner is equal to the prior covariance $\mathbf{P} = \boldsymbol{\Sigma}_{\text{pr}}$,

the following proposal arise [56]:

$$\left(\mathbf{I}_d + \frac{\Delta}{2}\mathbf{I}_d\right)\boldsymbol{\theta}' = \left(\mathbf{I}_d - \frac{\Delta}{2}\mathbf{I}_d\right)\boldsymbol{\theta} + \sqrt{2\boldsymbol{\Sigma}_{\text{pr}}\Delta}\xi_0, \quad (4.20)$$

and if we define $\xi \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\text{pr}})$, the proposal step can be expressed as

$$\boldsymbol{\theta}' = \frac{(2 - \Delta)}{(2 + \Delta)}\boldsymbol{\theta} + \frac{\sqrt{8\Delta}}{(2 + \Delta)}\xi \quad \longrightarrow \quad \boldsymbol{\theta}' = \sqrt{1 - \beta^2}\boldsymbol{\theta} + \beta\xi, \quad (4.21)$$

where $\beta = \sqrt{8\Delta}/(2 + \Delta)$. A common choice for the scaling is $\beta \in (0, 1]$ for discretization steps $\Delta \in (0, 2]$ (in general $\beta \rightarrow 0$ as $\Delta \rightarrow \infty$). The acceptance probability of the preconditioned Crank–Nicolson (pCN) algorithm only requires the potential function and is given by [60]

$$\alpha(\boldsymbol{\theta}, \boldsymbol{\theta}') = \min\{1, \exp(\Phi(\boldsymbol{\theta}; \tilde{\mathbf{y}}) - \Phi(\boldsymbol{\theta}'; \tilde{\mathbf{y}}))\}. \quad (4.22)$$

The pCN algorithm proceeds in a similar manner to MH. Note that pCN is motivated for Gaussian random functions. For non-Gaussian parameter settings, an isoprobabilistic mapping is required to express them as a Gaussian. Finally, we remark that the pCN proposal coincides with the conditional sampling proposal $q(\boldsymbol{\theta}' | \boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\theta}'; \rho \cdot \boldsymbol{\theta}, 1 - \rho^2)$ introduced in [177], where the parametrization is in terms of a correlation coefficient parameter $\rho = \sqrt{1 - \beta^2}$.

MCMC algorithms are usually incorporated within sequential schemes to draw samples from the posterior more efficiently. Some approaches include sequential MC, sequential IS [62], nested sampling [211], or structural reliability methods [223]. The idea is to explore the posterior on-the-fly by constructing a set of intermediate measures that converge to the full posterior. We focus on the BUS approach [223], which allows expressing a Bayesian inverse problem as an equivalent rare event simulation task.

4.3. The BUS approach to Bayesian inversion

Consider the target function $\pi(\boldsymbol{\theta}) \propto \pi_{\text{pos}}(\boldsymbol{\theta} | \tilde{\mathbf{y}})$. The posterior density can be characterized by drawing samples from this unnormalized density. Particularly, the *rejection sampling algorithm* generates samples from $\pi(\boldsymbol{\theta})$ using a proposal density $q(\boldsymbol{\theta})$. The proposal is selected such that it dominates the target function; this means that $q(\boldsymbol{\theta})$ must have equal or heavier tails than those of $\pi(\boldsymbol{\theta})$. Therefore, the proposal satisfies the relation

$$\sup_{\boldsymbol{\theta} \in \Theta} \left(\frac{\pi(\boldsymbol{\theta})}{q(\boldsymbol{\theta})} \right) \leq \bar{c} < \infty \quad \text{for some covering constant } \bar{c} \in \mathbb{R}_{>0} \quad (4.23)$$

and $\text{supp}(\pi(\boldsymbol{\theta})) \subseteq \text{supp}(q(\boldsymbol{\theta}))$. Thereafter, samples drawn from $q(\boldsymbol{\theta})$ are rejected strategically to make the resulting accepted samples distributed according to $\pi(\boldsymbol{\theta})$.

A simple choice for the proposal density is the prior distribution $\pi_{\text{pr}}(\boldsymbol{\theta})$. The acceptance probability α in rejection sampling [174] becomes

$$\alpha = \frac{\pi(\boldsymbol{\theta})}{\bar{c} \cdot q(\boldsymbol{\theta})} = \frac{\pi_{\text{pr}}(\boldsymbol{\theta}) L(\boldsymbol{\theta}; \tilde{\mathbf{y}})}{\bar{c} \cdot \pi_{\text{pr}}(\boldsymbol{\theta})} = c \cdot L(\boldsymbol{\theta}; \tilde{\mathbf{y}}), \quad (4.24)$$

where $c = 1/\bar{c} \in \mathbb{R}_{>0}$ and the covering constant is selected such that $\bar{c} \geq L_{\text{max}} = \max(L(\boldsymbol{\theta}; \tilde{\mathbf{y}}))$. In this case, rejection sampling amounts to: (i) drawing a standard uniform random number $v \sim \text{Unif}[0, 1]$, (ii) sampling a candidate from the prior $\boldsymbol{\theta} \sim \pi_{\text{pr}}(\boldsymbol{\theta})$, and (iii) accepting the candidate if $v \leq \alpha = c \cdot L(\boldsymbol{\theta}; \tilde{\mathbf{y}})$.

This particular acceptance-rejection mechanism allows us to generate the space

$$\mathcal{H} = \{(\boldsymbol{\theta}, v) \in \mathcal{X} : h(\boldsymbol{\theta}, v) \leq 0\}, \quad \text{where} \quad h(\boldsymbol{\theta}, v) = v - c \cdot L(\boldsymbol{\theta}; \tilde{\mathbf{y}}) \quad (4.25)$$

and $\mathcal{X} = [\boldsymbol{\Theta}, \Upsilon]$ is an augmented parameter space with $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and $v \in \Upsilon \subseteq [0, 1]$. Figure 4.1 illustrates the relation between rejection sampling and BUS.

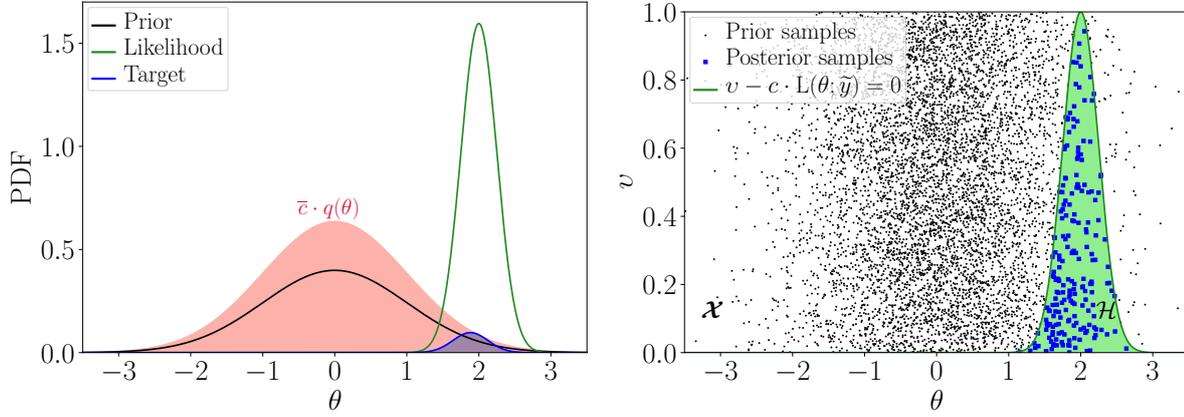


Figure 4.1: BUS as a rejection sampling algorithm. Left: standard rejection sampling with proposal distribution based on the prior. Right: BUS incorporates an auxiliary standard uniform random variable v to express the Bayesian inverse problem as a rare event simulation task.

In the context of reliability analysis and rare event simulation, the space \mathcal{H} defines a failure domain with LSF $h(\boldsymbol{\theta}, v)$. If one is able to generate samples $\{(\boldsymbol{\theta}, v)\} \in \mathcal{X}$ that fall into \mathcal{H} , the resulting parameter samples $\{\boldsymbol{\theta}\} \in \boldsymbol{\Theta}$ will be distributed according to the posterior $\pi_{\text{pos}}(\boldsymbol{\theta} | \tilde{\mathbf{y}})$. The $\{v\}$ samples are auxiliary. This connection is the foundation of the BUS approach, since one can employ existing methods from rare event simulation to perform Bayesian inference. Indeed, the previous rejection sampling algorithm corresponds to applying standard MC simulation for the solution of a rare event estimation problem defined by the LSF $h(\boldsymbol{\theta}, v)$.

Specifically, the posterior distribution in (4.2) is expressed in BUS by conditioning the joint distribution of $(\boldsymbol{\theta}, v) \in \mathcal{X}$ to the space \mathcal{H} , and marginalizing over the auxiliary uniform random variable v , that is

$$\pi_{\text{pos}}(\boldsymbol{\theta} | \tilde{\mathbf{y}}) = \frac{\pi_{\text{pr}}(\boldsymbol{\theta})}{Z_{\boldsymbol{\theta}}} \int_0^1 \mathbb{1}_{\mathcal{H}}(\boldsymbol{\theta}, v) dv \quad (4.26a)$$

$$\text{with} \quad Z_{\boldsymbol{\theta}} = \int_{\boldsymbol{\Theta}} \int_0^1 \mathbb{1}_{\mathcal{H}}(\boldsymbol{\theta}, v) \pi_{\text{pr}}(\boldsymbol{\theta}) dv d\boldsymbol{\theta}, \quad (4.26b)$$

where the indicator function is such that $\mathbb{1}_{\mathcal{H}}(\boldsymbol{\theta}, v) = 1$ when $(\boldsymbol{\theta}, v) \in \mathcal{H}$, and $\mathbb{1}_{\mathcal{H}}(\boldsymbol{\theta}, v) = 0$ otherwise.

Note that the application of BUS requires the knowledge of the constant $c = 1/\bar{c}$. From (4.24), it is seen that the covering constant is optimally chosen as the maximum of the likelihood function $\bar{c} = L_{\text{max}}$. If $\bar{c} < L_{\text{max}}$, the resulting samples will be distributed according to a truncated posterior distribution [27]. Conversely, if $\bar{c} > L_{\text{max}}$, the efficiency of BUS decreases because the value of $p_{\mathcal{H}}$ will be small and more samples are required for its estimation. Since in many cases L_{max} is not known in advance and its computation poses an additional cost, we can employ the adaptive strategy introduced in [27], for which the constant c is computed at each step of the simulation.

Recall from Chapter 3 that the main objective in rare event simulation is the estimation of the probability of failure. When employing the BUS framework, the estimator of the failure

probability $\hat{p}_{\mathcal{H}}$ is related to the probability that the samples belong to the domain \mathcal{H} , i.e., $p_{\mathcal{H}} = \mathbb{P}[\mathcal{H}] = \mathbb{P}[h(\boldsymbol{\theta}, v) \leq 0]$. This probability, which is obtained as a by-product of BUS, is used to approximate the model evidence as [223]

$$Z_{\tilde{\mathbf{y}}} \approx \hat{Z}_{\tilde{\mathbf{y}}} = \frac{\hat{p}_{\mathcal{H}}}{c} = \bar{c} \cdot \hat{p}_{\mathcal{H}}. \quad (4.27)$$

4.3.1. BUS with subset simulation

In order to efficiently compute samples from the posterior distribution, BUS is often combined with subset simulation (subsection 3.2.3). The combination of BUS with SuS (termed BUS-SuS), performs Bayesian inversion sequentially. This is because SuS transforms the task of estimating the rare event $\{h(\boldsymbol{\theta}, v) \leq 0\}$ into a sequence of problems involving more frequent events.

In BUS-SuS, the parameter space \mathcal{X} is divided into a decreasing sequence of nested subsets or intermediate levels, starting from the whole space and narrowing down to the target posterior space, i.e., $\mathcal{X} = \mathcal{H}_0 \supset \mathcal{H}_1 \supset \dots \supset \mathcal{H}_{n_{\text{lv}}} = \mathcal{H}$, such that $\mathcal{H} = \bigcap_{j=0}^{n_{\text{lv}}} \mathcal{H}_j$, where n_{lv} is the number of intermediate levels. The probability that the prior samples fall into the posterior space, $p_{\mathcal{H}}$ is given by

$$p_{\mathcal{H}} = \prod_{j=1}^{n_{\text{lv}}} \mathbb{P}[\mathcal{H}_j \mid \mathcal{H}_{j-1}], \quad (4.28)$$

where $\mathbb{P}[\mathcal{H}_j \mid \mathcal{H}_{j-1}]$ represents the conditional probability at level $(j - 1)$. Each intermediate level is defined as the set $\mathcal{H}_j = \{(\boldsymbol{\theta}, v) \in \mathcal{X} : h(\boldsymbol{\theta}, v) \leq \tau_j\}$, where $\infty = \tau_0 > \tau_1 > \dots > \tau_j > \dots > \tau_{n_{\text{lv}}} = 0$, is a decreasing sequence of threshold levels (cf. Figure 4.1).

Recall from SuS (subsection 3.2.3) that the probability $p_{\mathcal{H}}$ is computed as $\hat{p}_{\mathcal{H}} = p_0^{n_{\text{lv}}-1} \cdot \hat{p}_{n_{\text{lv}}}$, where $\hat{p}_{n_{\text{lv}}}$ is the conditional probability at the last level, estimated by MC as the ratio of the number of samples that lie in \mathcal{H} and the number of samples per level N . The probability estimate $\hat{p}_{\mathcal{H}}$ can be used to compute the model evidence via (4.27). At the last simulation level of BUS-SuS, the $\hat{p}_{n_{\text{lv}}} \cdot N$ samples that lie in \mathcal{H} are used as seeds to generate the final batch of N samples conditional on \mathcal{H} . The resulting samples are uniformly weighted but correlated samples of the posterior distribution.

Remark 4.3. Reliability problems are generally solved in the standard Gaussian space (cf. Remark 3.1). Due to the BUS-SuS formulation, this also translates to the Bayesian inversion. Hence, a new standard Gaussian parameter vector $\boldsymbol{\vartheta} = [\boldsymbol{\theta}, \theta_v]^\top \in \mathbb{R}^{k+1}$ can be defined by combining the parameter vector $\boldsymbol{\theta}$ and the transformed auxiliary variable $\theta_v = \Phi^{-1}(v)$, where $\Phi(\cdot)$ denotes the standard Gaussian CDF.

Remark 4.4. In order to guarantee a smooth transition between the intermediate levels, as well as for numerical stability, the LSF (4.25) is expressed in terms of the log-likelihood. Applying the natural logarithm to each term of $h(\boldsymbol{\theta}, v)$ in (4.25) yields [66]

$$h_{\ln}(\boldsymbol{\vartheta}) = \ln(\Phi(\theta_v)) - \ln(c \cdot L(\boldsymbol{\theta}; \tilde{\mathbf{y}})) = \ln(\Phi(\theta_v)) + \ln(\bar{c}) - \ln L(\boldsymbol{\theta}; \tilde{\mathbf{y}}). \quad (4.29)$$

Remark 4.5. BUS-SuS can be seen as a sequential auxiliary variable MCMC method [149, Ch.4], in which one augments the parameter space by adding the standard uniform random variable v . As a result, BUS-SuS shares some similarities with the slice sampler [171], and especially, with the nested sampling algorithm [211], which is a popular method in statistical physics for the estimation of normalizing constants. The connections of BUS-SuS with nested sampling are highlighted in [26, p.191].

The BUS-SuS method in the standard Gaussian space is shown in Algorithm 5, which follows from the SuS Algorithm 2. For consistency, we assume that the log-likelihood function includes the transformation from the standard Gaussian parameters to the original space. Moreover, we

refer to [177] for a variety of MCMC methods that can be used for BUS-SuS, which require some modifications of the algorithms described in section 4.2.

Algorithm 5 BUS-SuS in the standard Gaussian space.

Input : number of samples per level N , conditional probability p_0 , covering constant \bar{c} , log-likelihood function $\ln L(\cdot; \tilde{\mathbf{y}})$, maximum iterations m

```

/* Initial MC */
1 Draw  $N$  samples from the standard Gaussian,  $\boldsymbol{\vartheta}_0 = [\boldsymbol{\theta}_0, \theta_{v,0}] \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{d+1})$ 
2 Compute initial log-likelihood function values,  $L_{\text{eval}} \leftarrow \ln L(\boldsymbol{\theta}_0; \tilde{\mathbf{y}})$ 
3 Set  $j \leftarrow 0$  and  $\tau_0 \leftarrow \infty$ 
/* BUS-SuS steps */
4 while  $\tau_j > 0$  and  $j < m$  do
5   Increase intermediate level counter,  $j \leftarrow j + 1$ 
   /* Update threshold */
6   Compute LSF values,  $h_{\text{eval}} \leftarrow \ln(\Phi(\theta_{v,j-1})) + \ln(\bar{c}) - L_{\text{eval}}$ 
7   Sort  $h_{\text{eval}}$  in ascending order and create a vector  $\text{idx}$  to store the indices of this sorting
8   Create  $\boldsymbol{\vartheta}_{\text{sort}}$  as the parameter samples  $\boldsymbol{\vartheta}_{j-1}$  sorted according to  $\text{idx}$ 
9   Set the intermediate threshold level  $\tau_j$  as the  $p_0$ -percentile of the values in  $h_{\text{eval}}$ 
10  Compute the number of samples in the  $j$ -th intermediate level,  $N_{\mathcal{H}_j} \leftarrow \sum_{i=1}^N (h_{\text{eval}}^{(i)} \leq \max(0, \tau_j))$ 
11  if  $\tau_j > 0$  then
12    |  $\hat{p}_{j-1} \leftarrow p_0$ 
13  else
14    |  $\tau_j \leftarrow 0$  and  $\hat{p}_{j-1} \leftarrow N_{\mathcal{H}_j}/N$ 
15  end
16  Select seeds for the MCMC step,  $\boldsymbol{\vartheta}_{\text{seed}} \leftarrow \{\boldsymbol{\vartheta}_{\text{sort}}^{(i)}\}_{i=1}^{N_{\mathcal{H}_j}}$ 
17  Randomize the ordering of the seeds
   /* MCMC step */
18  Generate next level values  $\{\boldsymbol{\vartheta}_j^{(i)}, L_{\text{eval}}^{(i)}\}_{i=1}^N$  from the randomized seeds ( $\boldsymbol{\vartheta}_{\text{seed}}$ ) and intermediate level
    $\tau_j$  using an MCMC algorithm. Here, each seed is used to construct a chain with  $N_c = \text{floor}(N/N_s)$ 
   states, where  $N_s = N_{\mathcal{H}_j}$  is the number of seeds
19 end
20  $n_{\text{lv}} \leftarrow j$ 
21 Estimate  $\hat{p}_{\mathcal{H}} = \prod_{i=0}^{n_{\text{lv}}} \hat{p}_i$ 
22 Estimate the model evidence  $\hat{Z}_{\tilde{\mathbf{y}}} = \hat{p}_{\mathcal{H}} \cdot \bar{c}$ 
Output: posterior parameter samples  $\{\boldsymbol{\vartheta}_{n_{\text{lv}}}^{(i)}\}_{i=1}^N$  and model evidence  $\hat{Z}_{\tilde{\mathbf{y}}}$ .

```

Illustration 1. Consider the 1D diffusion example in subsection 3.3.2, we are interested in finding the hydraulic conductivity using $m = 10$ noisy pressure observations. The noise is Gaussian with mean zero and standard deviation $\sigma_{\text{obs}} = 0.01$. This setting is adapted from [160].

Recall that the hydraulic conductivity is represented with a lognormal random field. In this case, the covariance kernel is exponential with correlation length $\ell = 0.1$ and the truncation of the KL expansion is fixed to $k = 5$. The BUS-SuS algorithm is applied to solve this inverse problem. The dimension of the parameter space is $d = 7$, including the flux and standard uniform random variables; the parameter vector is grouped as $\boldsymbol{\theta} = [F, \boldsymbol{\theta}^{(\text{KL})}, \nu]$. We use $N = 1 \times 10^4$ samples per level and $N_{\text{sim}} = 50$ independent simulation runs; the conditional probability is set to $p_0 = 0.1$. The covering constant is found by optimization and is equal to $\ln(\bar{c}) \approx 33.29$; we utilize the LSF (4.29). An adaptive pCN algorithm is used as the MCMC within BUS-SuS to keep a constant acceptance rate; this algorithm follows from [10, 177].

The BUS-SuS solution takes in average 5 simulation levels to reach the posterior. The pCN parameter β is set to 0.6 at the initial level and scales down to 0.035 (in average) at the final level, after performing the adaptation steps. The posterior mean and standard deviation, computed from the complete sample pool, are $\boldsymbol{\mu}_{\text{pos}} = [-0.806, -0.480, -0.149, -0.047, -1.463, 0.487, 0.120]$ and $\boldsymbol{\sigma}_{\text{pos}} = [0.101, 0.409, 0.573, 0.504, 0.569, 0.638, 0.101]$, respectively. Moreover,

the estimated model evidence (average of the 50 simulations) is equal to $\hat{Z} = 5.658 \times 10^9$ with a coefficient of variation of 0.335. Finally, the evolution of the prior to posterior update is illustrated in Figure 4.2 (without the auxiliary parameter v). We plot the prior and posterior marginals of the uncertain parameter. Note that in the posterior case, the marginals exhibit large correlations.

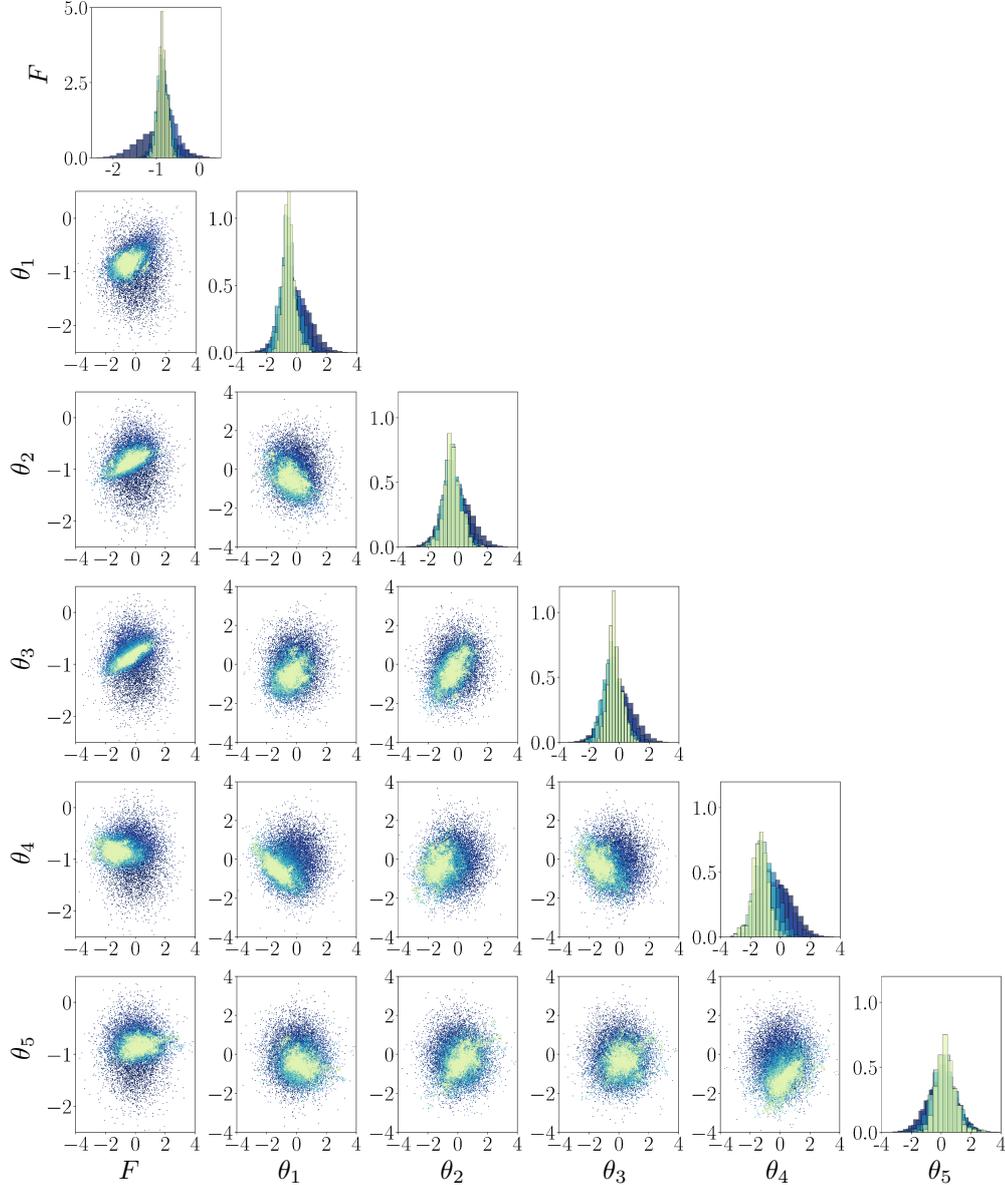


Figure 4.2: BUS-SuS algorithm: prior (blue) and posterior (light green) samples characterizing the 1D and 2D marginal distributions.

4.4. Updating of rare event probabilities

In the context of rare event simulation, it is also possible to evaluate failure probabilities conditional on the observational data. The updated probability of failure arises in many engineering applications, where data are collected to enhance the prediction of the probability of high-consequence events, e.g., the exceedance of a critical stress or displacement thresholds. This setting, which involves a combined Bayesian inversion and rare event estimation task, is referred to as *reliability updating* [176].

The information provided by measured or observed data can be incorporated into the analysis to improve the probability of failure estimate (3.1). This implies the computation of failure probabilities conditional on the observations $\tilde{\mathbf{y}}$. The *updated probability of failure* $\mathbb{P}[\mathcal{F} \mid \tilde{\mathbf{y}}]$ can be estimated using the posterior of the model parameters as

$$\mathbb{P}[\mathcal{F} \mid \tilde{\mathbf{y}}] := \int_{\Theta} \mathbb{1}_{\mathcal{F}}(\boldsymbol{\theta}) \pi_{\text{pos}}(\boldsymbol{\theta} \mid \tilde{\mathbf{y}}) d\boldsymbol{\theta} = \frac{1}{Z(\tilde{\mathbf{y}})} \int_{\Theta} \mathbb{1}_{\mathcal{F}}(\boldsymbol{\theta}) \pi_{\text{pr}}(\boldsymbol{\theta}) L(\boldsymbol{\theta}; \tilde{\mathbf{y}}) d\boldsymbol{\theta}. \quad (4.30)$$

The evaluation of the integral (4.30) is a more challenging task than (3.1), since in principle it requires sampling from the tails of the posterior distribution. Several strategies are proposed in the literature to estimate this posterior failure probability, see for example [176, 127, 88].

In BUS, the reliability updating problem can be re-expressed in terms of two rare event estimation tasks [224]. The resulting posterior failure probability is obtained by replacing the posterior density of (4.26a) into (4.30)

$$\mathbb{P}[\mathcal{F} \mid \tilde{\mathbf{y}}] = \frac{\int_{\Theta} \mathbb{1}_{\mathcal{F}}(\boldsymbol{\theta}) \int_0^1 \mathbb{1}_{\mathcal{H}}(\boldsymbol{\theta}, v) \pi_{\text{pr}}(\boldsymbol{\theta}) dv d\boldsymbol{\theta}}{\int_{\Theta} \int_0^1 \mathbb{1}_{\mathcal{H}}(\boldsymbol{\theta}, v) \pi_{\text{pr}}(\boldsymbol{\theta}) dv d\boldsymbol{\theta}} \quad (4.31a)$$

$$= \frac{\int_{g(\boldsymbol{\theta}) \leq 0 \cap h(\boldsymbol{\theta}, v) \leq 0} \pi_{\text{pr}}(\boldsymbol{\theta}) dv d\boldsymbol{\theta}}{\int_{h(\boldsymbol{\theta}, v) \leq 0} \pi_{\text{pr}}(\boldsymbol{\theta}) dv d\boldsymbol{\theta}} = \frac{\mathbb{P}[g(\boldsymbol{\theta}) \leq 0 \cap h(\boldsymbol{\theta}, v) \leq 0]}{\mathbb{P}[h(\boldsymbol{\theta}, v) \leq 0]}, \quad (4.31b)$$

which implies the computation of a so-called system reliability problem for the numerator and a component reliability problem for the denominator.

If the Bayesian inverse problem is solved with BUS-SuS (or any other sampling method), samples from the posterior distribution are available and they can be used to assist the estimation of the posterior probability of failure $\mathbb{P}[\mathcal{F} \mid \tilde{\mathbf{y}}]$. This is because the posterior samples belong to the domain \mathcal{H} associated to the LSF $h(\boldsymbol{\theta}, v)$, and they can be used to estimate the probability in the denominator of (4.31b). Therefore, the reliability updating task reduces to the estimation of the failure probability in the numerator, $\mathbb{P}[g(\boldsymbol{\theta}) \leq 0 \cap h(\boldsymbol{\theta}, v) \leq 0]$.

Particularly, the BUS-SuS algorithm can be modified to compute the updated failure probability in (4.31b). The idea is to employ the LSF describing the posterior $h(\boldsymbol{\theta}, v)$ (with fixed covering \bar{c}), the LSF describing the failure $g(\boldsymbol{\theta})$, and the estimated posterior samples $\{\boldsymbol{\theta}_{\text{pos}}\}$ to perform a last SuS run. This final SuS method requires some modifications of Algorithm 2, which read as follows:

- (i) *Initial Monte Carlo samples in SuS*: at the MC simulation level, we employ the estimated posterior samples $\boldsymbol{\theta}_{\text{pos}}$ to evaluate the LSF (cf. Line 1, Algorithm 2).
- (ii) *Acceptance/rejection MCMC criterion*: in the MCMC algorithm used within SuS at a given intermediate threshold level τ , the candidate sample needs to satisfy both, $g(\boldsymbol{\theta}) \leq \tau$ and $h(\boldsymbol{\theta}, v) \leq 0$; in this way the generated MCMC samples always belong to the intermediate failure domain, and at the same time, they remain in the space \mathcal{H} defining the posterior (cf. Line 18, Algorithm 2).

After this modified steps, the standard SuS algorithm continues until reaching the target failure domain intersected with the space of posterior samples $\mathcal{F} \cap \mathcal{H}$, for which $g(\boldsymbol{\theta}) \leq 0$ and $h(\boldsymbol{\theta}, v) \leq 0$. Details of this approach are given in [224], and we show an schematic representation of the problem in Figure 4.3.

One difficulty of this approach is that the coefficients of variation of the failure probability estimates become very large when the underlying probabilities of failure are very small. As an illustration, we show in Figure 4.4 the behavior of the simulation process for increasing dimension of the parameter space, in this case truncation orders of the KL expansion. This is a modified version of the example discussed later in section 4.6. It considers the updating of the

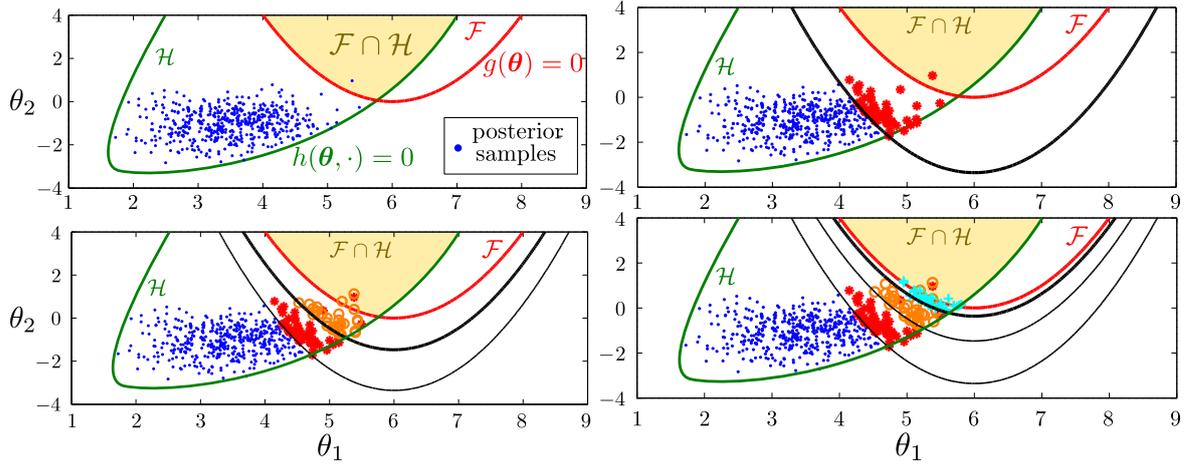


Figure 4.3: Schematic representation of SuS adapted for reliability updating: after obtaining posterior samples that belong to \mathcal{H} , SuS is used to “move” them to the failure set \mathcal{F} . During the entire simulation, the samples must lie in \mathcal{H} such that they always account for the observed data. Hence, the objective is to reach the set $\mathcal{F} \cap \mathcal{H}$.

Young’s modulus of the plate example in subsection 2.7.2 using measurements of the horizontal deformation only. The prior failure probability $p_{\mathcal{F}}$ is shown in solid black line estimated as the mean of several SuS runs. The application of BUS-SuS for the inference of the posterior parameters yields a probability $p_{\mathcal{H}}$ (shown in blue for the different simulation runs and their mean in dashed line); up to that point BUS-SuS requires on average 4 (inference) intermediate levels to obtain the posterior parameter samples. Thereafter, these samples are used to perform a final SuS run that computes the posterior probability of failure $\mathbb{P}[\mathcal{F} | \tilde{\mathbf{y}}]$ (shown in red for the different simulation runs and their mean in dashed line); for this SuS additionally requires on average 7 (failure) intermediate levels. Hence, at the end of the reliability updating computation, we effectively obtained 11 intermediate levels. Since the inherent correlation of the MCMC samples is being propagated from one level to the other, the updated probability estimates have large variability, as seen by the spread of the samples in Figure 4.4. This behavior will be discussed with further detail in section 4.6.

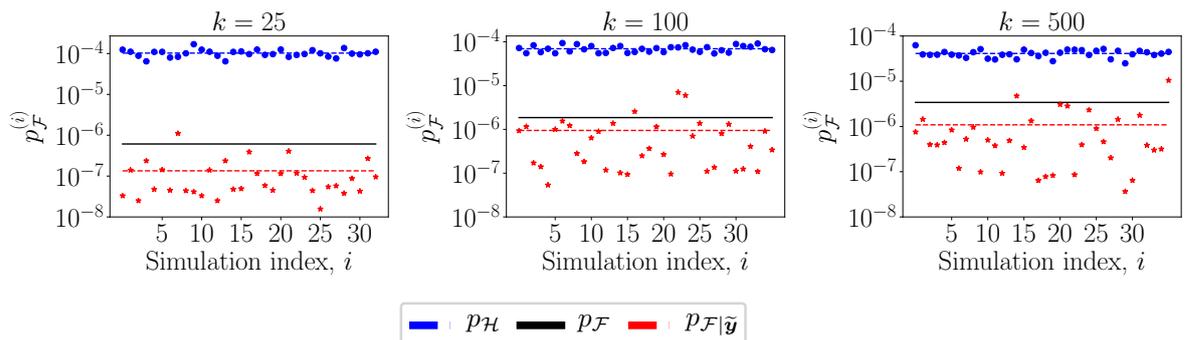


Figure 4.4: Probability of the domain \mathcal{H} together with the prior and posterior failure probabilities for 35 simulation runs of the SuS algorithm, and increasing terms in the KL expansion k (columns).

In the following, we discuss two examples that illustrate the Bayesian approach to inverse problems in the context of random fields: one for which it is possible to compute all the posterior quantities analytically, and a second one that requires the use of sampling-based approaches to estimate the posterior quantities. The idea is to explore the implications of different parameter choices for the prior random field modeling on the solution of the Bayesian inverse problem. We aim at showing these effects by carrying out several parameter studies.

4.5. Analytical example: one-dimensional cantilever beam

Consider the one-dimensional cantilever beam in subsection 2.7.1. We are now interested in inferring the flexibility random field using observations of the deflection response.

The observation noise is modeled as additive and mutually independent from the uncertain flexibility. The noise is described by a joint Gaussian distribution with mean zero and covariance matrix Σ_{obs} . The noise covariance is computed by assuming that the measurements are correlated with an exponential kernel, with standard deviation $\sigma_{\text{obs}} = 1 \times 10^{-3}$ and correlation length $\ell_{\text{obs}} = 1$ m. This results in the following likelihood function

$$L(\mathbf{F}; \tilde{\mathbf{y}}) = \frac{1}{\sqrt{(2\pi)^m \det(\Sigma_{\text{obs}})}} \exp\left(-\frac{1}{2}[\tilde{\mathbf{y}} - w(\tilde{\mathbf{x}}, \mathbf{F})]^\top \Sigma_{\text{obs}}^{-1}[\tilde{\mathbf{y}} - w(\tilde{\mathbf{x}}, \mathbf{F})]\right), \quad (4.32)$$

here the parameter \mathbf{F} denotes a realization of the flexibility random field, $w(\tilde{\mathbf{x}}, \mathbf{F})$ is the forward response operator defined via (2.38) and $\tilde{\mathbf{y}}$ is a set of m deflection observations measured at equally spaced points $\tilde{\mathbf{x}}$ of the domain (Figure 4.5).

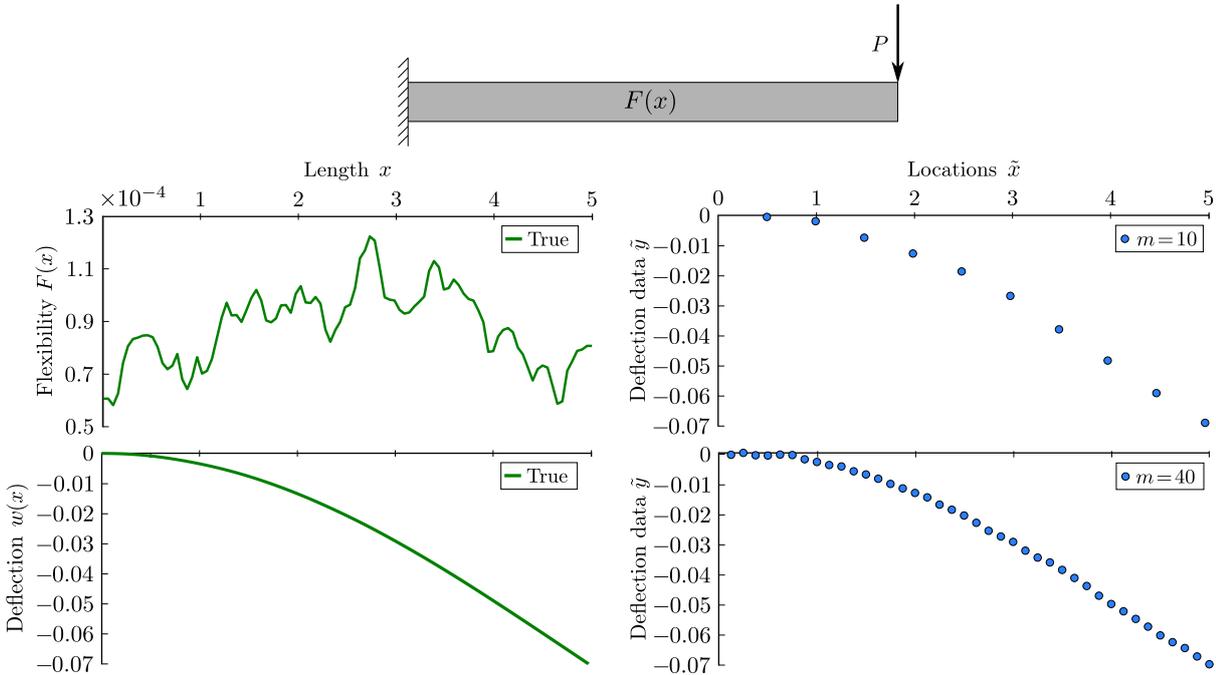


Figure 4.5: Cantilever beam: true values and two sets of deflection observations.

The observations are generated by simulation, assuming a true (but in real applications unknown) deflection of the beam. The underlying true flexibility is a realization of a random field modeled with an exponential kernel and $\ell_{\text{true}} = 2$ m. To avoid a so-called *inverse crime* [131, p.5], the numerical integration for obtaining the true displacements is performed on a finer grid (compared to the one used for the Bayesian inversion). The process involves evaluating the prior mean and covariance on the finer grid to compute a realization¹.

In the parameter studies, we consider the following random field settings:

- The number of terms in the KL expansion is $k \in \{5, 10, 20\}$;
- the correlation length of the prior flexibility varied among $\ell \in \{0.5, 2.5, 4.5\}$ m;
- two different sets of measurements are assumed with $m \in \{10, 40\}$ points (Figure 4.5); and

¹We remark that in [235] there is a typo in the value of the prior standard deviation of the flexibility. The correct value used in the studies is $\sigma_F = 0.35\mu_F = 3.5 \times 10^{-5}$, as we specified in subsection 2.7.1

- for each of these settings, the two covariance functions in (2.25) are used to represent the prior flexibility.

4.5.1. Analytical posterior

Since the prior and likelihood are Gaussian, the posterior distribution is also Gaussian [193]. We introduce the (joint) random vector $\mathbf{F}' = [\mathbf{F}, \tilde{\mathbf{w}}]$, comprised of the random vectors $\mathbf{F} = F(\mathbf{x}, \cdot) \in \mathbb{R}^n$, representing the flexibility field discretized at spatial locations $\mathbf{x} = [x_1, \dots, x_n]$ (the dimension becomes $d = n$), and $\tilde{\mathbf{w}} \in \mathbb{R}^m$ a random vector in the data space of deflections. The mean vector and covariance matrix of \mathbf{F}' can be partitioned accordingly;

$$\boldsymbol{\mu}_{F'} = \begin{bmatrix} \boldsymbol{\mu}_F \\ \boldsymbol{\mu}_{\tilde{\mathbf{w}}} \end{bmatrix} \quad \boldsymbol{\Sigma}_{F'F'} = \begin{bmatrix} \boldsymbol{\Sigma}_{FF} & \boldsymbol{\Sigma}_{F\tilde{\mathbf{w}}} \\ \boldsymbol{\Sigma}_{F\tilde{\mathbf{w}}}^\top & \boldsymbol{\Sigma}_{\tilde{\mathbf{w}}\tilde{\mathbf{w}}} \end{bmatrix}. \quad (4.33)$$

For a given deflection data set $\tilde{\mathbf{y}}$, the n -th order fi-di posterior distribution of the flexibility random field $\pi_{\text{pos}}(\mathbf{F} | \tilde{\mathbf{y}})$ can be obtained analytically from direct application of Bayes' theorem (see, e.g., [131, Sec. 3.4]). The resulting conditional density is given by

$$\pi_{\text{pos}}(\mathbf{F} | \tilde{\mathbf{y}}) = \frac{1}{\sqrt{(2\pi)^n \det(\boldsymbol{\Sigma}_{FF|\tilde{\mathbf{y}}})}} \exp\left(-\frac{1}{2} [\mathbf{F} - \boldsymbol{\mu}_{F|\tilde{\mathbf{y}}}]^\top \boldsymbol{\Sigma}_{FF|\tilde{\mathbf{y}}}^{-1} [\mathbf{F} - \boldsymbol{\mu}_{F|\tilde{\mathbf{y}}}] \right). \quad (4.34)$$

An analogous expression can be obtained for the posterior distribution of the deflection $\pi_{\text{pos}}(\mathbf{w} | \tilde{\mathbf{y}})$. Those multivariate distributions are characterized by the conditional mean vectors $\boldsymbol{\mu}_{F|\tilde{\mathbf{y}}}$, $\boldsymbol{\mu}_{w|\tilde{\mathbf{y}}}$ and the conditional covariance matrices $\boldsymbol{\Sigma}_{FF|\tilde{\mathbf{y}}}$, $\boldsymbol{\Sigma}_{ww|\tilde{\mathbf{y}}}$, which are respectively given by [239]

$$\boldsymbol{\mu}_{F|\tilde{\mathbf{y}}} = \boldsymbol{\mu}_F + \boldsymbol{\Sigma}_{F\tilde{\mathbf{w}}} \boldsymbol{\Sigma}_{\tilde{\mathbf{w}}\tilde{\mathbf{w}}}^{-1} (\tilde{\mathbf{y}} - \boldsymbol{\mu}_{\tilde{\mathbf{w}}}), \quad \boldsymbol{\Sigma}_{FF|\tilde{\mathbf{y}}} = \boldsymbol{\Sigma}_{FF} - \boldsymbol{\Sigma}_{F\tilde{\mathbf{w}}} \boldsymbol{\Sigma}_{\tilde{\mathbf{w}}\tilde{\mathbf{w}}}^{-1} \boldsymbol{\Sigma}_{F\tilde{\mathbf{w}}}^\top, \quad (4.35a)$$

$$\boldsymbol{\mu}_{w|\tilde{\mathbf{y}}} = \boldsymbol{\mu}_w + \boldsymbol{\Sigma}_{w\tilde{\mathbf{w}}} \boldsymbol{\Sigma}_{\tilde{\mathbf{w}}\tilde{\mathbf{w}}}^{-1} (\tilde{\mathbf{y}} - \boldsymbol{\mu}_{\tilde{\mathbf{w}}}), \quad \boldsymbol{\Sigma}_{ww|\tilde{\mathbf{y}}} = \boldsymbol{\Sigma}_{ww} - \boldsymbol{\Sigma}_{w\tilde{\mathbf{w}}} \boldsymbol{\Sigma}_{\tilde{\mathbf{w}}\tilde{\mathbf{w}}}^{-1} \boldsymbol{\Sigma}_{w\tilde{\mathbf{w}}}^\top. \quad (4.35b)$$

All quantities in (4.35) are known from the prior random fields or can be computed analytically since the inverse problem follows the additive relation $\tilde{\mathbf{w}} = \mathbf{w} + \boldsymbol{\eta}$. For instance, $\boldsymbol{\mu}_{\tilde{\mathbf{w}}}$ is given by the evaluation of (2.41) at the measurement locations $\tilde{\mathbf{x}}$, since $\mathbb{E}[\tilde{\mathbf{w}}] = \mathbb{E}[\mathbf{w} + \boldsymbol{\eta}] = \mathbb{E}[\mathbf{w}]$. Moreover, $\boldsymbol{\Sigma}_{\tilde{\mathbf{w}}\tilde{\mathbf{w}}} = \boldsymbol{\Sigma}_{ww} + \boldsymbol{\Sigma}_{\text{obs}}$, where $\boldsymbol{\Sigma}_{ww}$ is computed by evaluating (2.42) at the measurement locations.

The mean, standard deviation and correlation functions of the posterior flexibility and deflection random fields are plotted in Figure 4.6, for the Matérn kernels with $\nu = \{0.5, \infty\}$ and $\ell = (0.5L) = 2.5$ m. The structure of such parameters changes completely by the inclusion of data and the resulting posterior fields become non-homogeneous. Compare to the prior flexibility and deflection cases in Figures 2.5 to 2.7. Note that the standard deviation of the flexibility field changes considerably depending on whether one uses the exponential or squared-exponential kernel. When using smooth kernels there is higher correlation between the spatial points, and thus the information influences the updating in a more global manner. Notice also that this behavior does not appear for the deflection, where the smoothing action of the forward operator dominates.

4.5.2. Approximation of the posterior

When dealing with random fields, the Bayesian inference process involves analysis in infinite-dimensional spaces. In such cases, the uncertain function is typically represented by a suitable parametrization. We use the KL expansion to discretize the prior flexibility random field.

The approximated posterior random field can be subsequently computed following the ideas in subsection 4.5.1. In this case, however, it is applied to the random vector of KL coefficients

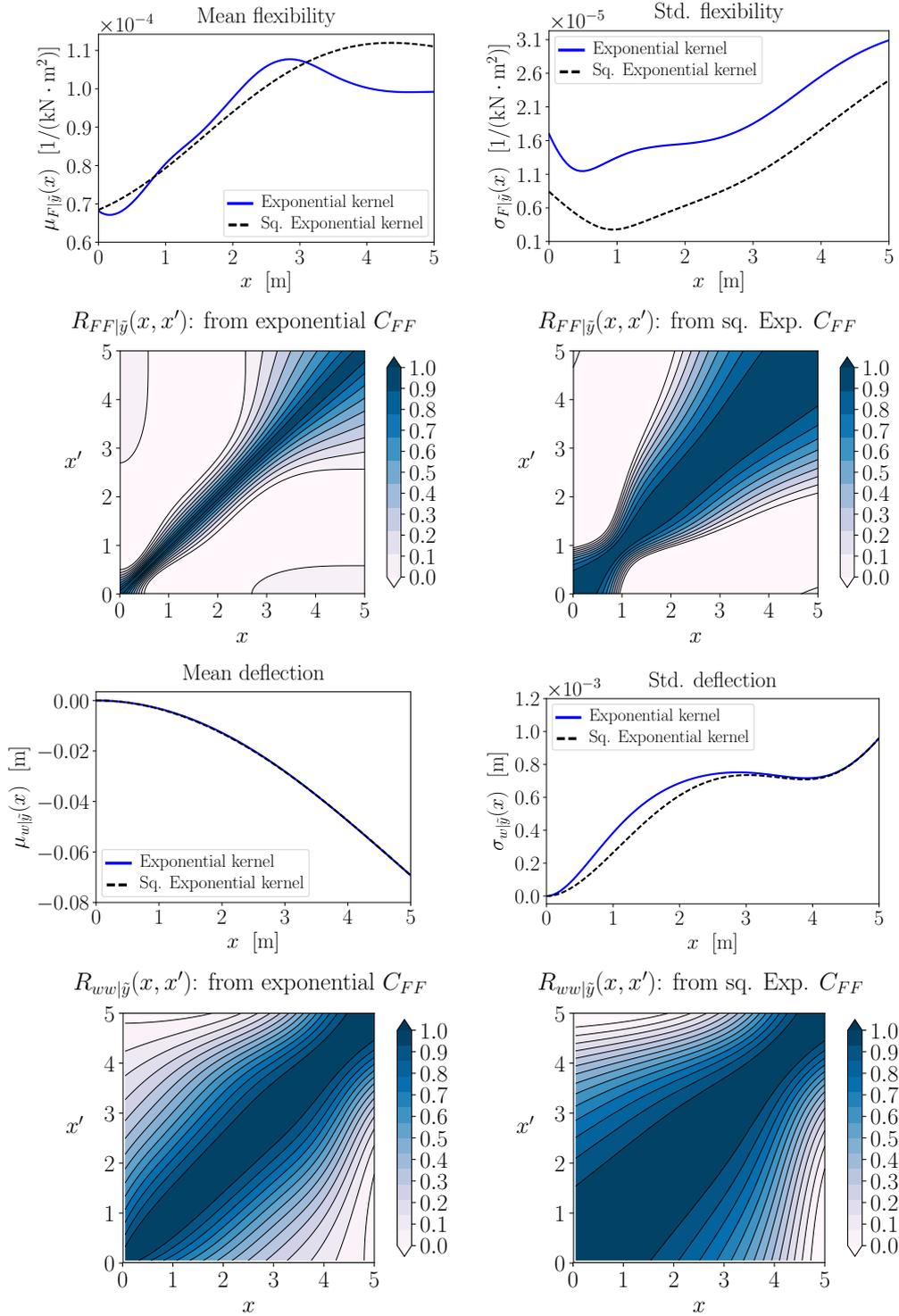


Figure 4.6: Mean, standard deviation and correlation function of the **posterior** flexibility (rows 1-2) and deflection (rows 3-4) random fields (from (4.35)). Using the exponential and squared exponential kernels in (2.25) to model C_{FF} (with $\ell = 2.5$ and $m = 10$).

$\theta \in \mathbb{R}^k$ (the dimension becomes $d = k \ll n$). We assume a Gaussian random vector composed by θ and \tilde{w} . The posterior distribution can be calculated as the conditional PDF of θ given a data set \tilde{y} . This posterior random field can be represented as a multivariate Gaussian distribution

with conditional mean vector $\boldsymbol{\mu}_{\theta|\tilde{y}}$ and conditional covariance matrix $\boldsymbol{\Sigma}_{\theta\theta|\tilde{y}}$ given by

$$\boldsymbol{\mu}_{\theta|\tilde{y}} = \boldsymbol{\mu}_{\theta} + \boldsymbol{\Sigma}_{\theta\tilde{w}}\boldsymbol{\Sigma}_{\tilde{w}\tilde{w}}^{-1}(\tilde{\mathbf{y}} - \boldsymbol{\mu}_{\tilde{w}}) \quad \text{and} \quad \boldsymbol{\Sigma}_{\theta\theta|\tilde{y}} = \boldsymbol{\Sigma}_{\theta\theta} - \boldsymbol{\Sigma}_{\theta\tilde{w}}\boldsymbol{\Sigma}_{\tilde{w}\tilde{w}}^{-1}\boldsymbol{\Sigma}_{\tilde{w}\theta}^{\top}, \quad (4.36)$$

where $\boldsymbol{\mu}_{\theta} = \mathbb{E}[\boldsymbol{\theta}] = \mathbf{0}$, $\boldsymbol{\Sigma}_{\theta\theta} = \mathbf{I}_k$ ($\mathbf{I}_k \in \mathbb{R}^{k \times k}$ is the identity matrix). From (4.36), one sees that the posterior mean and covariance are respectively positive and negative updates of the prior counterparts. The remaining covariance terms in (4.36) can be derived analytically from the approximated model (2.40), i.e., $\tilde{\mathbf{w}} = (\boldsymbol{\mu}_w - \mathbf{A}\boldsymbol{\theta}) + \boldsymbol{\eta}$. For example,

$$\text{Cov}[\boldsymbol{\theta}, \tilde{\mathbf{w}}] = \mathbb{E}\left[(\boldsymbol{\theta} - \boldsymbol{\mu}_{\theta})(\tilde{\mathbf{w}} - \boldsymbol{\mu}_{\tilde{w}})^{\top}\right] \quad (4.37a)$$

$$= \mathbb{E}\left[\boldsymbol{\theta}(-\mathbf{A}\boldsymbol{\theta} + \boldsymbol{\eta})^{\top}\right] = -\mathbf{A}^{\top} + \underbrace{\mathbb{E}\left[\boldsymbol{\eta}^{\top}\right]}_{=0} \implies \boldsymbol{\Sigma}_{\theta\tilde{w}} = -\mathbf{A}^{\top} \quad (4.37b)$$

$$\text{Cov}[\tilde{\mathbf{w}}, \tilde{\mathbf{w}}] = \mathbb{E}\left[(\tilde{\mathbf{w}} - \boldsymbol{\mu}_{\tilde{w}})(\tilde{\mathbf{w}} - \boldsymbol{\mu}_{\tilde{w}})^{\top}\right] \quad (4.37c)$$

$$= \mathbb{E}\left[(-\mathbf{A}\boldsymbol{\theta} + \boldsymbol{\eta})(-\mathbf{A}\boldsymbol{\theta} + \boldsymbol{\eta})^{\top}\right] \implies \boldsymbol{\Sigma}_{\tilde{w}\tilde{w}} = \mathbf{A}\mathbf{A}^{\top} + \boldsymbol{\Sigma}_{\text{obs}}. \quad (4.37d)$$

After deriving each term in (4.36), the mean vector and covariance matrix of the posterior distribution of $\boldsymbol{\theta}$ can be written as

$$\boldsymbol{\mu}_{\theta|\tilde{y}} = -\mathbf{A}^{\top} \left(\mathbf{A}\mathbf{A}^{\top} + \boldsymbol{\Sigma}_{\text{obs}}\right)^{-1} (\tilde{\mathbf{y}} - \boldsymbol{\mu}_{\tilde{w}}) \quad \text{and} \quad \boldsymbol{\Sigma}_{\theta\theta|\tilde{y}} = \mathbf{I} - \mathbf{A}^{\top} \left(\mathbf{A}\mathbf{A}^{\top} + \boldsymbol{\Sigma}_{\text{obs}}\right)^{-1} \mathbf{A}. \quad (4.38)$$

The posterior random fields of the flexibility and deflection after using the KL approximation can be obtained from the posterior of $\boldsymbol{\theta}$. In this case, both random fields are also represented by multivariate Gaussian distributions described by the following approximated mean and covariance functions:

$$\hat{\boldsymbol{\mu}}_{F|\tilde{y}}(x) = \boldsymbol{\mu}_F(x) + \sum_{i=1}^k \sqrt{\lambda_i} \phi_i(x) \boldsymbol{\mu}_{\theta|\tilde{y}}^{(i)}, \quad \hat{C}_{FF|\tilde{y}}(x, x') = \sum_{i=1}^k \sum_{j=1}^k \sqrt{\lambda_i \lambda_j} \phi_i(x) \phi_j(x') \boldsymbol{\Sigma}_{\theta\theta|\tilde{y}}^{(i,j)}, \quad (4.39a)$$

$$\hat{\boldsymbol{\mu}}_{w|\tilde{y}}(x) = \boldsymbol{\mu}_w(x) + P \sum_{i=1}^k \sqrt{\lambda_i} \Phi_i(x) \boldsymbol{\mu}_{\theta|\tilde{y}}^{(i)}, \quad \hat{C}_{ww|\tilde{y}}(x, x') = P^2 \sum_{i=1}^k \sum_{j=1}^k \sqrt{\lambda_i \lambda_j} \Phi_i(x) \Phi_j(x') \boldsymbol{\Sigma}_{\theta\theta|\tilde{y}}^{(i,j)}, \quad (4.39b)$$

where the superscripts in the vector $\boldsymbol{\mu}_{\theta|\tilde{y}}^{(i)}$ and matrix $\boldsymbol{\Sigma}_{\theta\theta|\tilde{y}}^{(i,j)}$ refer to element indexing. Recall from subsection 2.7.1 that $\{\lambda_i, \phi_i\}_{i=1}^k$ are the eigenpairs of the covariance operator and $\Phi_i(x)$ denotes the components of the integrated eigenfunction defined in (2.40) and plotted in Figure 2.4.

Evaluating the posterior approximation

The analytical posterior random field expressions (4.35) and the associated KL approximations (4.39) allow an assessment of the influence of different prior random field assumptions on the posterior solution. Note that the reference posterior random field is constructed with a covariance matrix with rank n (number of points in the domain discretization), since we employ closed-form expressions. For the KL representations, each of the posterior random field approximations is built with covariance matrices whose rank is defined by the truncation order of the expansion. Therefore, we are comparing low-dimensional k -rank approximations against the full n -rank reference.

The approximation of the posterior flexibility random field using an exponential kernel as the underlying prior flexibility covariance is illustrated in Figure 4.7. We show the 95% CI of

the analytical solution (shaded area) and the KL approximations as a function of the number of terms in the expansion for increasing correlation length. Those intervals are represented as the region between the 0.025 and 0.975 quantiles of the posterior. The complete set of KL representations are located inside the analytical CI, and they converge to this solution as the number of terms increases. For small correlation lengths, the posterior random field is more difficult to capture since one is learning a random field that has larger variability. Nevertheless, already $k = 10$ terms in the expansion are enough to have a good approximation of the posterior flexibility random field. Comparing the results from both sets of measurements, the number of data points controls the width of the CI bounds and it narrows as more information is available. Furthermore, the flexibility random field is no longer weakly homogeneous since the posterior mean vary through the domain (cf. Figure 4.6).

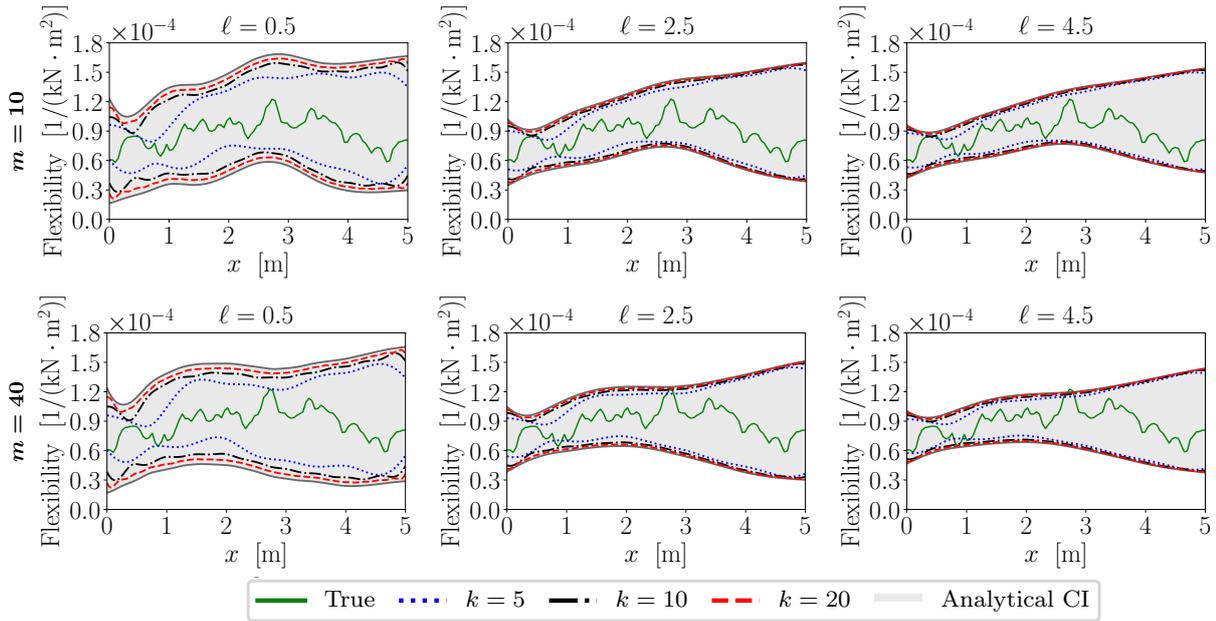


Figure 4.7: Posterior flexibility using an **exponential** kernel for the prior: 95% CI for different terms in the KL expansion, number of measurements (rows), and correlation lengths of the prior flexibility (columns). The shaded area corresponds to the analytical CI given by (4.35a).

Figure 4.8 presents the approximation of the posterior deflection random field with underlying exponential covariance for the prior flexibility. In order to better illustrate the solutions, the 95% CIs of a differential deflection are shown. They are computed as the difference between the prior mean of the random field and the 95% posterior CIs. In contrast to the posterior flexibility, the influence of the correlation length in the KL approximation of the posterior deflection is small. For all cases, the KL expansion represents the posterior deflection almost as exact as the analytical case, even when using a small number of terms in the expansion. The reason for this is that the posterior deflection is computed by averaging the KL expansion of the flexibility random field over the domain (see (2.39a)). As a result, the influence of the higher KL eigenfunctions becomes negligible, and mainly the first modes have a contribution to the random field representation.

Finally, the approximation of the posterior flexibility and deflection random fields assuming a squared exponential covariance function for the prior flexibility is shown in Figure 4.9. Here, only the results for the set of measurements with $m = 10$ points are shown. Even for small correlation lengths and KL expansions with at least $k = 10$ terms, the difference between the posterior flexibility random field and the analytical solution is negligible. As the correlation length increases, the inverse problem solution can be computed accurately with even a smaller

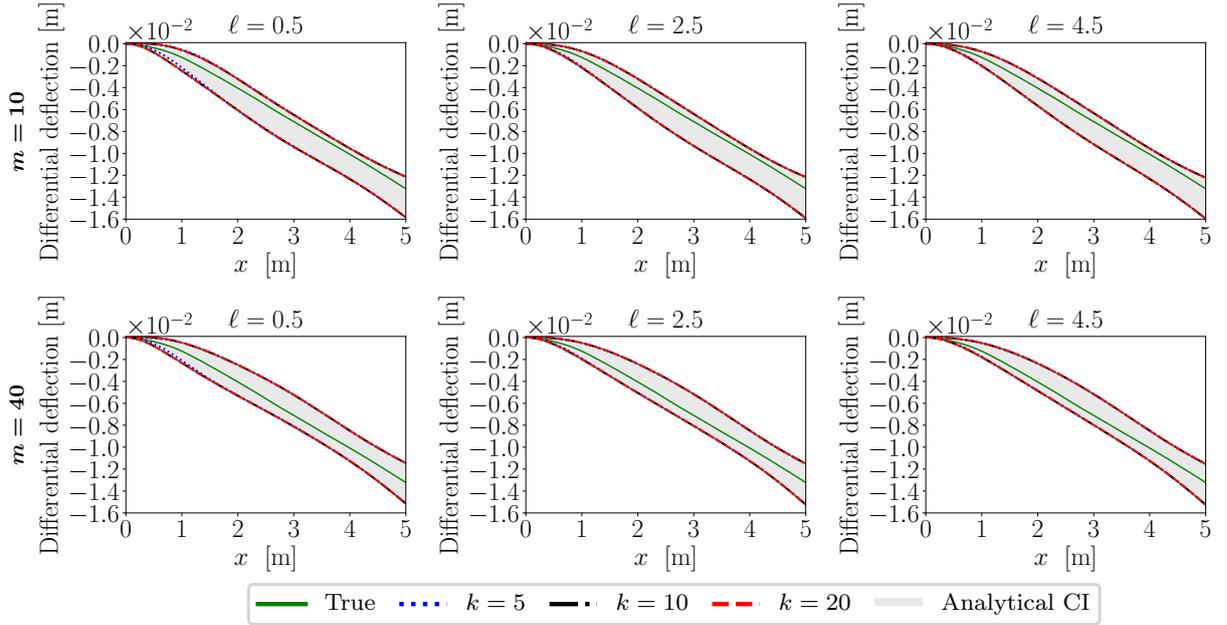


Figure 4.8: Differential posterior deflection using an **exponential** kernel for the prior: 95% CI for different terms in the KL expansion, number of measurements (rows), and correlation lengths of the prior flexibility (columns). The shaded area corresponds to the analytical CI given by (4.35b).

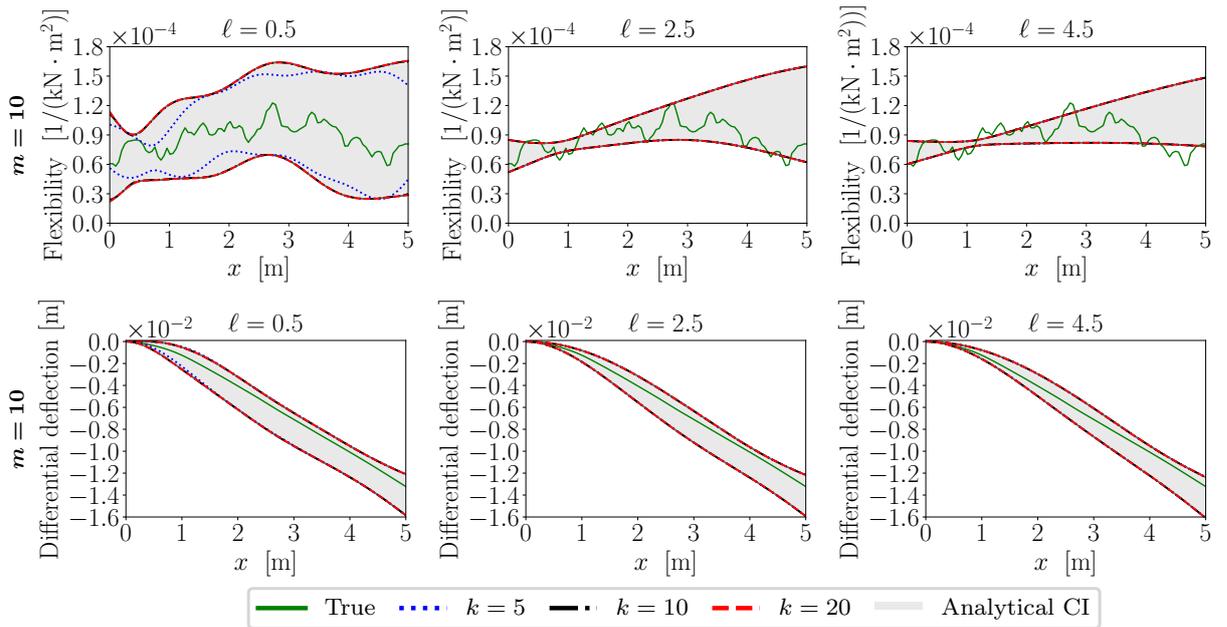


Figure 4.9: Posterior flexibility and differential posterior deflection using a **squared exponential** kernel for the prior: 95% CI for different terms in the KL expansion and correlation lengths of the prior flexibility. The shaded area corresponds to the analytical CI.

number of terms in the expansion ($k = 5$). We point out that the eigenvalue decay is stronger for the squared exponential kernel as compared to the exponential, which yields to a lower number of terms in the KL representation. Moreover, the true underlying flexibility is generated assuming an exponential kernel. This is reflected in the inverse problem solution, since sample paths generated from a random field with a squared exponential covariance smooth out faster. The resulting posterior approximation is not able to capture the true flexibility field with high confidence at all spatial points of the domain when the correlation length is large. Note that this

is not the case for the deflection field, which is always well-captured independent of the choice of the covariance kernel and correlation length. This is again due to the smoothing action of the forward operator. The combination of the squared exponential model with large correlation lengths leads to random fields that are too smooth and they are good at approximating the deflection field. However, this is not the case for the flexibility random field since smoothing oftentimes leads to loss of information that cannot be retrieved in an inverse operation; as a result the solution of the inverse problem is typically dominated by the prior.

Remark 4.6. Since the true solution is unknown in practice, there is no exact way to verify how well the approximations capture the truth. A possible way to proceed is via hypothesis testing [93, Ch. 6], which aims at retrieving the truth based on the data. Otherwise, the assumed models can be validated using Bayesian model selection or averaging [21].

Evaluating the error in the approximation

The error measures (2.34) discussed in section 2.6 can be evaluated for the prior and posterior flexibility and deflection random fields. The averaged relative mean error $\bar{\epsilon}_\mu(k)$ associated to the posterior flexibility and deflection random field representation is shown in Figure 4.10. For the prior random field, the error in the mean approximation is zero since the KL expansion represents the mean exactly. However, the KL discretization induces an error in the mean of the posterior random field. In all cases, this error is very small even when the truncation order in the expansion is small. The basis coefficients of a Gaussian posterior mean decay exponentially fast due to its smoothness (see [237] for a discussion). This indicates that most of the inference effort is necessary for representing the variability of the posterior field, rather than its mean. Moreover, the results show that the posterior mean error is about one order of magnitude higher for the flexibility than for the deflection, at the largest value of the KL truncation.

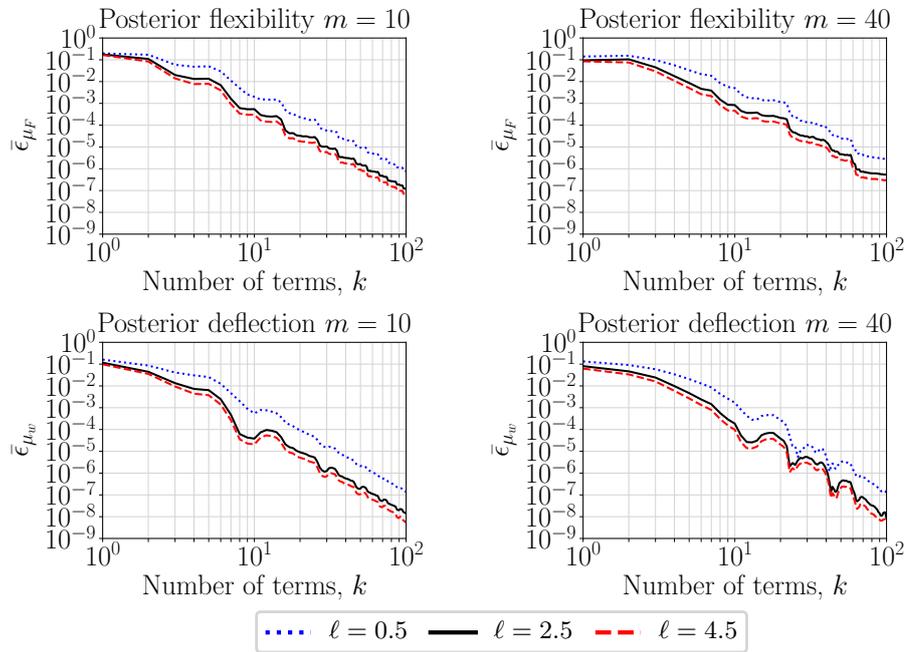


Figure 4.10: Averaged relative error mean in the posterior random fields using an **exponential** kernel for the prior: for the flexibility and deflection, number of measurements, different terms in the KL expansion and correlation lengths.

Figure 4.11 presents the averaged relative variance error $\bar{\epsilon}_{\sigma^2}(k)$ associated to the prior and posterior flexibility and deflection random field representation. The influence of the correlation length in the prior random field is larger compared to the posterior random field solution (the

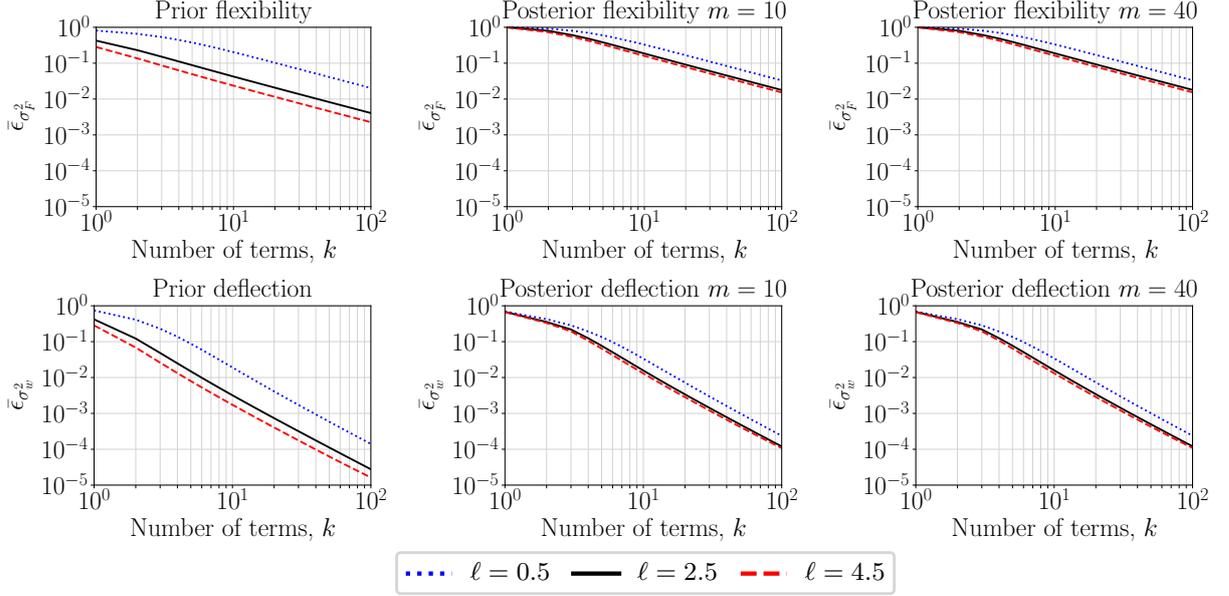


Figure 4.11: Averaged relative error variance in the prior and posterior random fields using an **exponential** kernel for the prior: for the flexibility and deflection (rows), number of measurements, different terms in the KL expansion, and correlation lengths.

error gap between correlation lengths is tighter in the posterior case). Moreover, the number of terms in the KL expansion required to attain a given variability of the random field is larger for the posterior case. This means that the use of a higher-order KL discretization is necessary when dealing with Bayesian inverse problems, compared to forward uncertainty propagation problems.

Furthermore, it is seen that more terms in the KL expansion are required in the posterior flexibility to obtain the same variance error magnitudes of the posterior deflection. This behavior is again a consequence of the deflection being computed as averaging of the KL expansion of the flexibility over the domain.

4.5.3. Model evidence

Consider a finite collection of possible models $\{\mathcal{M}_1, \mathcal{M}_2, \dots, \mathcal{M}_k, \dots, \mathcal{M}_{k_{\max}}\}$, where $k \in [1, k_{\max}]$ is a model indicator index. Each particular model \mathcal{M}_k has an associated vector of uncertain parameters $\boldsymbol{\theta} \in \mathbb{R}^k$, where the dimension k vary between different models. In the context of the KL discretization, these models correspond to the dimension of the stochastic space discretized by the truncated series, i.e., the number of terms in the KL expansion.

A closed-form expression for the model evidence can be derived for this example. The process involves a marginalization of the likelihood function over the parameters (integration); alternatively it can be computed as the product of prior with likelihood divided by the posterior. Following the latter approach, the natural logarithm of the model evidence in (4.2) is given by

$$\ln Z(\tilde{\mathbf{y}} | k) = \ln \pi_{\text{pr}}(\boldsymbol{\theta} | k) + \ln L(\boldsymbol{\theta}; \tilde{\mathbf{y}}, k) - \ln \pi_{\text{pos}}(\boldsymbol{\theta} | \tilde{\mathbf{y}}, k), \quad (4.40)$$

where the log-prior, log-likelihood and log-posterior conditional on the truncation are

$$\ln \pi_{\text{pr}}(\boldsymbol{\theta} | k) = -\frac{k}{2} \ln(2\pi) - \frac{1}{2} \boldsymbol{\theta}^T \boldsymbol{\theta} \quad (4.41a)$$

$$\ln L(\boldsymbol{\theta}; \tilde{\mathbf{y}}, k) = -\frac{m}{2} \ln(2\pi) - \frac{1}{2} \ln(\det(\boldsymbol{\Sigma}_{\text{obs}})) - \frac{1}{2} \left([\tilde{\mathbf{y}} - (\boldsymbol{\mu}_w - \mathbf{A}\boldsymbol{\theta})]^T \boldsymbol{\Sigma}_{\text{obs}}^{-1} [\tilde{\mathbf{y}} - (\boldsymbol{\mu}_w - \mathbf{A}\boldsymbol{\theta})] \right) \quad (4.41b)$$

$$\ln \pi_{\text{pos}}(\boldsymbol{\theta} \mid \tilde{\mathbf{y}}, k) = -\frac{k}{2} \ln(2\pi) - \frac{1}{2} \ln \left(\det(\boldsymbol{\Sigma}_{\theta\theta|\tilde{\mathbf{y}}}) \right) - \frac{1}{2} \left([\boldsymbol{\theta} - \boldsymbol{\mu}_{\theta|\tilde{\mathbf{y}}}]^{\top} \boldsymbol{\Sigma}_{\theta\theta|\tilde{\mathbf{y}}}^{-1} [\boldsymbol{\theta} - \boldsymbol{\mu}_{\theta|\tilde{\mathbf{y}}}] \right). \quad (4.41c)$$

The model evidence is calculated by inserting (4.41a)–(4.41c) into (4.40),

$$\begin{aligned} \ln Z(\tilde{\mathbf{y}} \mid k) &= -\frac{m}{2} \ln(2\pi) - \frac{1}{2} \ln \left(\frac{\det(\boldsymbol{\Sigma}_{\text{obs}})}{\det(\boldsymbol{\Sigma}_{\theta\theta|\tilde{\mathbf{y}}})} \right) \\ &\quad - \frac{1}{2} \left(\boldsymbol{\theta}^{\top} \boldsymbol{\theta} + [\tilde{\mathbf{y}} - (\boldsymbol{\mu}_w - \mathbf{A}\boldsymbol{\theta})]^{\top} \boldsymbol{\Sigma}_{\text{obs}}^{-1} [\tilde{\mathbf{y}} - (\boldsymbol{\mu}_w - \mathbf{A}\boldsymbol{\theta})] \right. \\ &\quad \left. - [\boldsymbol{\theta} - \boldsymbol{\mu}_{\theta|\tilde{\mathbf{y}}}]^{\top} \boldsymbol{\Sigma}_{\theta\theta|\tilde{\mathbf{y}}}^{-1} [\boldsymbol{\theta} - \boldsymbol{\mu}_{\theta|\tilde{\mathbf{y}}}] \right) \\ &= -\frac{m}{2} \ln(2\pi) - \frac{1}{2} \ln \left(\frac{\det(\boldsymbol{\Sigma}_{\text{obs}})}{\det(\boldsymbol{\Sigma}_{\theta\theta|\tilde{\mathbf{y}}})} \right) - \frac{1}{2} (\tilde{\mathbf{y}} - \boldsymbol{\mu}_w)^{\top} \boldsymbol{\Sigma}_{\text{obs}}^{-1} (\tilde{\mathbf{y}} - \boldsymbol{\mu}_w) \\ &\quad - \frac{1}{2} \left(\boldsymbol{\theta}^{\top} \boldsymbol{\theta} + \tilde{\mathbf{y}}^{\top} \boldsymbol{\Sigma}_{\text{obs}}^{-1} \mathbf{A}\boldsymbol{\theta} - \boldsymbol{\mu}_w^{\top} \boldsymbol{\Sigma}_{\text{obs}}^{-1} \mathbf{A}\boldsymbol{\theta} + \boldsymbol{\theta}^{\top} \mathbf{A}^{\top} \boldsymbol{\Sigma}_{\text{obs}}^{-1} \tilde{\mathbf{y}} - \boldsymbol{\theta}^{\top} \mathbf{A}^{\top} \boldsymbol{\Sigma}_{\text{obs}}^{-1} \boldsymbol{\mu}_w \right. \\ &\quad \left. + \boldsymbol{\theta}^{\top} \mathbf{A}^{\top} \boldsymbol{\Sigma}_{\text{obs}}^{-1} \mathbf{A}\boldsymbol{\theta} - \boldsymbol{\theta}^{\top} \boldsymbol{\Sigma}_{\theta\theta|\tilde{\mathbf{y}}}^{-1} \boldsymbol{\theta} + \boldsymbol{\theta}^{\top} \boldsymbol{\Sigma}_{\theta\theta|\tilde{\mathbf{y}}}^{-1} \boldsymbol{\mu}_{\theta|\tilde{\mathbf{y}}} + \boldsymbol{\mu}_{\theta|\tilde{\mathbf{y}}}^{\top} \boldsymbol{\Sigma}_{\theta\theta|\tilde{\mathbf{y}}}^{-1} \boldsymbol{\theta} - \boldsymbol{\mu}_{\theta|\tilde{\mathbf{y}}}^{\top} \boldsymbol{\Sigma}_{\theta\theta|\tilde{\mathbf{y}}}^{-1} \boldsymbol{\mu}_{\theta|\tilde{\mathbf{y}}} \right). \end{aligned}$$

We use two of the Searle identities in [204, p.151] to re-write the posterior mean and covariance of the KL coefficients in (4.38). For the posterior covariance, we employ the identity $(\mathbf{B} + \mathbf{C}\mathbf{C}^{\top})^{-1} \mathbf{C} = \mathbf{B}^{-1} \mathbf{C} (\mathbf{I} + \mathbf{C}^{\top} \mathbf{B}^{-1} \mathbf{C})^{-1}$, with $\mathbf{B} = \boldsymbol{\Sigma}_{\text{obs}}$ and $\mathbf{C} = \mathbf{A}$. For the posterior mean, we use the identity $(\mathbf{D} + \mathbf{B})^{-1} = \mathbf{D}^{-1} - \mathbf{D}^{-1} \mathbf{B} (\mathbf{I} + \mathbf{D}^{-1} \mathbf{B})^{-1} \mathbf{D}^{-1}$, with $\mathbf{B} = \boldsymbol{\Sigma}_{\text{obs}}$ and $\mathbf{D} = \mathbf{A}\mathbf{A}^{\top}$. The resulting expressions are

$$\begin{aligned} \boldsymbol{\Sigma}_{\theta\theta|\tilde{\mathbf{y}}} &= \mathbf{I}_k - \mathbf{A}^{\top} (\boldsymbol{\Sigma}_{\text{obs}} + \mathbf{A}\mathbf{A}^{\top})^{-1} \mathbf{A} \\ &= \mathbf{I}_k - \mathbf{A}^{\top} \boldsymbol{\Sigma}_{\text{obs}}^{-1} \mathbf{A} (\mathbf{I}_k + \mathbf{A}^{\top} \boldsymbol{\Sigma}_{\text{obs}}^{-1} \mathbf{A})^{-1} = (\mathbf{I}_k + \mathbf{A}^{\top} \boldsymbol{\Sigma}_{\text{obs}}^{-1} \mathbf{A})^{-1} \end{aligned} \quad (4.43a)$$

$$\begin{aligned} \boldsymbol{\mu}_{\theta|\tilde{\mathbf{y}}} &= -\mathbf{A}^{\top} (\mathbf{A}\mathbf{A}^{\top} + \boldsymbol{\Sigma}_{\text{obs}})^{-1} (\tilde{\mathbf{y}} - \boldsymbol{\mu}_w) \\ &= -(\mathbf{I}_k + \mathbf{A}^{\top} \boldsymbol{\Sigma}_{\text{obs}}^{-1} \mathbf{A})^{-1} \mathbf{A}^{\top} \boldsymbol{\Sigma}_{\text{obs}}^{-1} (\tilde{\mathbf{y}} - \boldsymbol{\mu}_w) = -\boldsymbol{\Sigma}_{\theta\theta|\tilde{\mathbf{y}}} \mathbf{A}^{\top} \boldsymbol{\Sigma}_{\text{obs}}^{-1} (\tilde{\mathbf{y}} - \boldsymbol{\mu}_w); \end{aligned} \quad (4.43b)$$

we remark that $\mathbf{A}^{\top} \boldsymbol{\Sigma}_{\text{obs}}^{-1} \mathbf{A}$ in (4.43a) is the Hessian of the negative log-likelihood parameterized by the KL expansion (compare to [218]). The natural logarithm of the model evidence is then computed as

$$\begin{aligned} \ln Z(\tilde{\mathbf{y}} \mid k) &= -\frac{m}{2} \ln(2\pi) - \frac{1}{2} \ln \left(\frac{\det(\boldsymbol{\Sigma}_{\text{obs}})}{\det(\boldsymbol{\Sigma}_{\theta\theta|\tilde{\mathbf{y}}})} \right) - \frac{1}{2} (\tilde{\mathbf{y}} - \boldsymbol{\mu}_w)^{\top} \boldsymbol{\Sigma}_{\text{obs}}^{-1} (\tilde{\mathbf{y}} - \boldsymbol{\mu}_w) + \frac{1}{2} \boldsymbol{\mu}_{\theta|\tilde{\mathbf{y}}}^{\top} \boldsymbol{\Sigma}_{\theta\theta|\tilde{\mathbf{y}}}^{-1} \boldsymbol{\mu}_{\theta|\tilde{\mathbf{y}}} \\ &\quad - \frac{1}{2} \left(\boldsymbol{\theta}^{\top} \boldsymbol{\theta} + \tilde{\mathbf{y}}^{\top} \boldsymbol{\Sigma}_{\text{obs}}^{-1} \mathbf{A}\boldsymbol{\theta} - \boldsymbol{\mu}_w^{\top} \boldsymbol{\Sigma}_{\text{obs}}^{-1} \mathbf{A}\boldsymbol{\theta} + \boldsymbol{\theta}^{\top} \mathbf{A}^{\top} \boldsymbol{\Sigma}_{\text{obs}}^{-1} \tilde{\mathbf{y}} - \boldsymbol{\theta}^{\top} \mathbf{A}^{\top} \boldsymbol{\Sigma}_{\text{obs}}^{-1} \boldsymbol{\mu}_w \right. \\ &\quad \left. + \boldsymbol{\theta}^{\top} \mathbf{A}^{\top} \boldsymbol{\Sigma}_{\text{obs}}^{-1} \mathbf{A}\boldsymbol{\theta} - \boldsymbol{\theta}^{\top} (\mathbf{I}_k + \mathbf{A}^{\top} \boldsymbol{\Sigma}_{\text{obs}}^{-1} \mathbf{A}) \boldsymbol{\theta} - \boldsymbol{\theta}^{\top} \mathbf{A}^{\top} \boldsymbol{\Sigma}_{\text{obs}}^{-1} (\tilde{\mathbf{y}} - \boldsymbol{\mu}_w) \right. \\ &\quad \left. - (\tilde{\mathbf{y}} - \boldsymbol{\mu}_w)^{\top} \boldsymbol{\Sigma}_{\text{obs}}^{-1} \mathbf{A}\boldsymbol{\theta} \right) \\ &= -\frac{m}{2} \ln(2\pi) - \frac{1}{2} \ln \left(\frac{\det(\boldsymbol{\Sigma}_{\text{obs}})}{\det(\boldsymbol{\Sigma}_{\theta\theta|\tilde{\mathbf{y}}})} \right) - \frac{1}{2} (\tilde{\mathbf{y}} - \boldsymbol{\mu}_w)^{\top} \boldsymbol{\Sigma}_{\text{obs}}^{-1} (\tilde{\mathbf{y}} - \boldsymbol{\mu}_w) + \frac{1}{2} \boldsymbol{\mu}_{\theta|\tilde{\mathbf{y}}}^{\top} \boldsymbol{\Sigma}_{\theta\theta|\tilde{\mathbf{y}}}^{-1} \boldsymbol{\mu}_{\theta|\tilde{\mathbf{y}}} \\ &\quad - \frac{1}{2} \underbrace{\left(\boldsymbol{\theta}^{\top} \boldsymbol{\theta} + \boldsymbol{\theta}^{\top} \mathbf{A}^{\top} \boldsymbol{\Sigma}_{\text{obs}}^{-1} \mathbf{A}\boldsymbol{\theta} - \boldsymbol{\theta}^{\top} (\mathbf{I}_k + \mathbf{A}^{\top} \boldsymbol{\Sigma}_{\text{obs}}^{-1} \mathbf{A}) \boldsymbol{\theta} \right)}_{=0} \\ &\quad \underbrace{+ \tilde{\mathbf{y}}^{\top} \boldsymbol{\Sigma}_{\text{obs}}^{-1} \mathbf{A}\boldsymbol{\theta} - \boldsymbol{\mu}_w^{\top} \boldsymbol{\Sigma}_{\text{obs}}^{-1} \mathbf{A}\boldsymbol{\theta} - (\tilde{\mathbf{y}} - \boldsymbol{\mu}_w)^{\top} \boldsymbol{\Sigma}_{\text{obs}}^{-1} \mathbf{A}\boldsymbol{\theta}}_{=0} \end{aligned}$$

$$\underbrace{+\boldsymbol{\theta}^\top \mathbf{A}^\top \boldsymbol{\Sigma}_{\text{obs}}^{-1} \tilde{\mathbf{y}} - \boldsymbol{\theta}^\top \mathbf{A}^\top \boldsymbol{\Sigma}_{\text{obs}}^{-1} \boldsymbol{\mu}_w - \boldsymbol{\theta}^\top \mathbf{A}^\top \boldsymbol{\Sigma}_{\text{obs}}^{-1} (\tilde{\mathbf{y}} - \boldsymbol{\mu}_w)}_{=0}, \quad (4.44a)$$

where the terms depending on $\boldsymbol{\theta}$ cancel out, leaving only the expression:

$$\ln Z(\tilde{\mathbf{y}} | k) = -\frac{1}{2} \left(m \ln(2\pi) + \ln \left(\frac{\det(\boldsymbol{\Sigma}_{\text{obs}})}{\det(\boldsymbol{\Sigma}_{\theta|\tilde{\mathbf{y}}})} \right) + (\tilde{\mathbf{y}} - \boldsymbol{\mu}_w)^\top \boldsymbol{\Sigma}_{\text{obs}}^{-1} (\tilde{\mathbf{y}} - \boldsymbol{\mu}_w) - \boldsymbol{\mu}_{\theta|\tilde{\mathbf{y}}}^\top \boldsymbol{\Sigma}_{\theta|\tilde{\mathbf{y}}}^{-1} \boldsymbol{\mu}_{\theta|\tilde{\mathbf{y}}} \right). \quad (4.45)$$

We employ the model evidence of (4.45) to perform model comparison between the investigated prior correlation lengths and truncation orders of the KL expansion. The dimension with the highest value of the model evidence is regarded as the best model, meaning that it gives an optimum balance between predictability and quality of the data fit [21].

Figure 4.12 shows the model evidence for different KL expansion terms, where the exponential kernel is used as the covariance of the prior flexibility. The “best” models are highlighted by a solid red line. Notice that different choices in the parameters of the prior random field lead to different optimal truncation orders in the KL expansion. As it is also evident in the posterior approximation results, random fields described by small correlation lengths require a larger number of terms in the expansion for their discretization. In particular, the information gained by the inclusion of additional terms is negligible once the best dimension is achieved, and it is lower than the penalty for the increased model complexity. Furthermore, the inclusion of more data points leads to a larger model evidence, which requires more KL parameters for an optimal random field representation. For both sets of measurements, the smaller correlation length yields the higher model evidence. We also evaluate the model evidence using a squared exponential covariance kernel for the prior flexibility (the plots are omitted). The results based on this assumption yield smaller model evidence values as compared to the exponential case. This agrees with the fact that the underlying true covariance is of the exponential type.

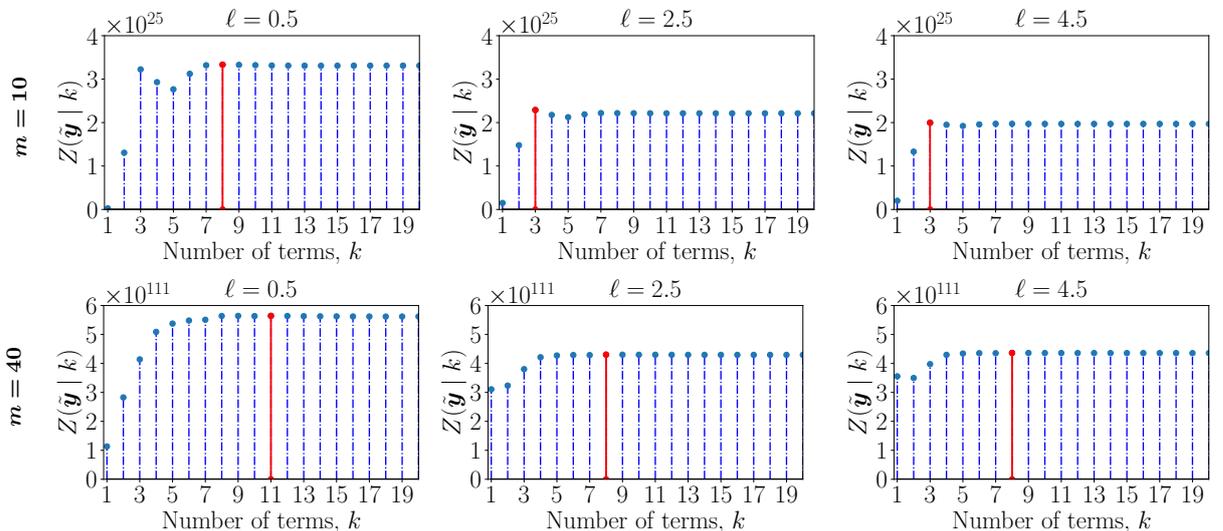


Figure 4.12: Model evidence using an **exponential** kernel for the prior for different number of measurements (rows) and correlation lengths of the prior flexibility (columns). The solid line corresponds to the best model.

Since the solution of the inverse problem is typically affected by changes in data, different measurements will yield different model evidence factors. In order to assess the overall contribution of the number of terms in the KL discretization, it is relevant to compute the model evidence without considering the measurement data. Therefore, the model evidence can be

marginalized with respect to the observational data as

$$\mathbb{E}_{\tilde{\mathbf{y}}} [Z(\tilde{\mathbf{y}} | k)] = \int_{\tilde{\mathbf{y}}} Z(\tilde{\mathbf{y}} | k) \pi_{\text{data}}(\tilde{\mathbf{y}}) d\tilde{\mathbf{y}}, \quad (4.46)$$

where $\pi_{\text{data}}(\tilde{\mathbf{y}})$ denotes the density used to generate the synthetic data. The value of (4.46) is shown in Figure 4.13 for different number of measurements and correlation lengths of the prior flexibility. The expected model evidence is also computed for the assumed true correlation length ($\ell = 2$ m), since such value yields the highest model evidence.

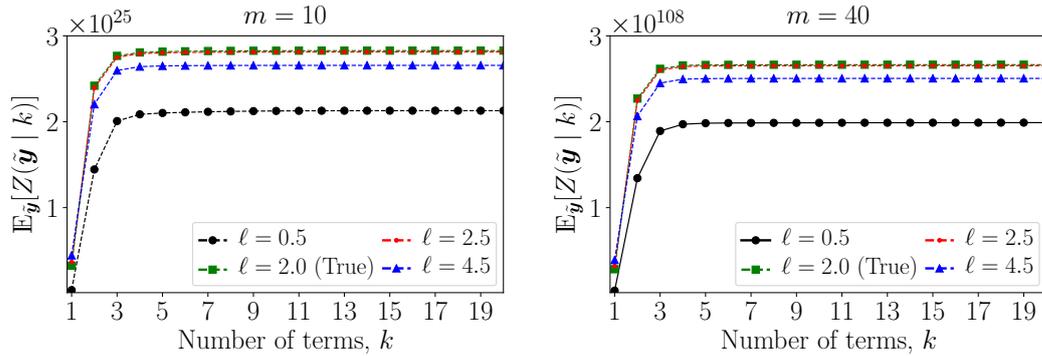


Figure 4.13: Expected model evidence using an **exponential** kernel as prior covariance with different number of data points (columns) and correlation lengths of the prior flexibility.

For the particular sets of measurements used during the inference process, the results in Figure 4.12 show that the random field modeled with the smaller correlation length has larger model evidence. However, if the observational data is averaged out, the smaller ℓ actually yields lower model evidence. In this case, the resulting random field generates sample paths with larger variability. This can be regarded as an increase in the model complexity, which in turn augments the penalization within the Bayesian framework. Moreover, note that the increasing number of data points has a positive scaling effect on the model evidence.

Remark 4.7. The natural logarithm of the model evidence can also be written as

$$\begin{aligned} \ln Z(\tilde{\mathbf{y}} | k) &= \int_{\Theta} \ln \left[\frac{\pi_{\text{pr}}(\boldsymbol{\theta} | k) L(\boldsymbol{\theta}; \tilde{\mathbf{y}}, k)}{\pi_{\text{pos}}(\boldsymbol{\theta} | \tilde{\mathbf{y}}, k)} \right] \pi_{\text{pos}}(\boldsymbol{\theta} | \tilde{\mathbf{y}}, k) d\boldsymbol{\theta} \\ &= \int_{\Theta} \ln L(\boldsymbol{\theta}; \tilde{\mathbf{y}}, k) \pi_{\text{pos}}(\boldsymbol{\theta} | \tilde{\mathbf{y}}, k) d\boldsymbol{\theta} - \int_{\Theta} \ln \left[\frac{\pi_{\text{pos}}(\boldsymbol{\theta} | \tilde{\mathbf{y}}, k)}{\pi_{\text{pr}}(\boldsymbol{\theta} | k)} \right] \pi_{\text{pos}}(\boldsymbol{\theta} | \tilde{\mathbf{y}}, k) d\boldsymbol{\theta} \\ &= \underbrace{\mathbb{E}_{\pi_{\text{pos}}} [\ln L(\boldsymbol{\theta}; \tilde{\mathbf{y}}, k)]}_{\text{average data fit}} - \underbrace{\mathbb{E}_{\pi_{\text{pos}}} \left[\ln \left(\frac{\pi_{\text{pos}}(\boldsymbol{\theta} | \tilde{\mathbf{y}}, k)}{\pi_{\text{pr}}(\boldsymbol{\theta} | k)} \right) \right]}_{\text{KLD}}, \end{aligned} \quad (4.47)$$

where the KLD can also be computed analytically as [32]

$$D_{\text{KL}}(\pi_{\text{pos}}(\boldsymbol{\theta} | \tilde{\mathbf{y}}, k) || \pi_{\text{pr}}(\boldsymbol{\theta})) = \frac{1}{2} \left(\ln \left(\det(\boldsymbol{\Sigma}_{\theta\theta|\tilde{\mathbf{y}}})^{-1} \right) - k + \text{tr}(\boldsymbol{\Sigma}_{\theta\theta|\tilde{\mathbf{y}}}) + \boldsymbol{\mu}_{\theta|\tilde{\mathbf{y}}}^{\text{T}} \boldsymbol{\mu}_{\theta|\tilde{\mathbf{y}}} \right). \quad (4.48)$$

As pointed out in [21], the data-fit term increases with the complexity of the models (in our case increasing KL terms), which often leads to over-fitting of the data. However, we see in (4.47) that the KLD term acts as a penalization to the data-fit. As a result, Bayesian inference at the model level has an incorporated penalty against more complex models. We will discuss this further in Chapter 5.

Remark 4.8. As shown in (4.47), Bayesian inference has an integrated penalization against complex models, which is expressed in terms of the KLD. This can be seen for instance in model selection applied to regression, where large polynomial orders are penalized since they cause over-fitting. Therefore, it seems counter-intuitive that the model evidence keeps increasing with the number of terms in the KL expansion. This particular behavior is related to the fact that the KL terms contribute to the representation of the field variability, and the more we add the better KL will represent the posterior covariance. However, from *some* order in the discretization, the inclusion of extra terms does not significantly improves the variance approximation. As a result, the “penalization” in the KL case is expressed by a *plateau* where only small increments in the model evidence occur.

Consider for example the results in Figure 4.14, where we compare the model evidence, data misfit and KLD for increasing correlation lengths in the prior random field, $\ell = \{0.1, 0.15, 0.2, 0.25, 0.5\}$. The prior covariance kernel is exponential and the number of data points is $m = 10$. For correlation length values $\ell < 0.2$, the KLD penalizes small number of KL terms since they are not sufficient for a proper representation of the random field variability. After roughly 10 terms, the KLD stabilizes giving same penalization to all KL terms. The same stable behavior is observed for the remainder correlation length values.

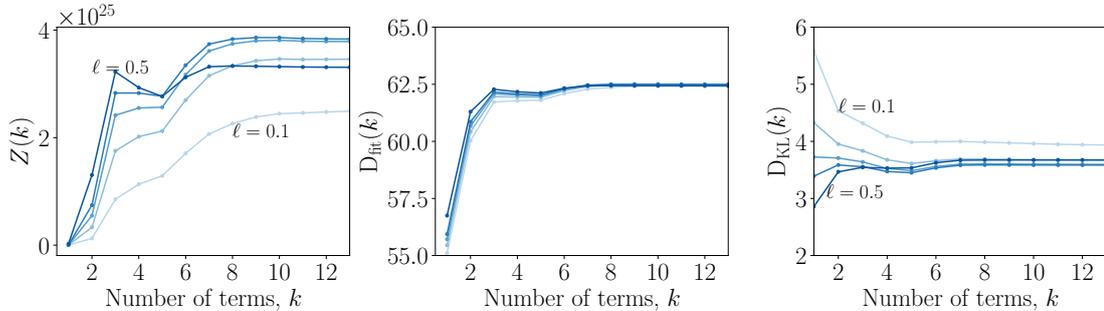


Figure 4.14: Model evidence, average data fit and KLD for increasing correlation lengths $\ell = \{0.1, 0.15, 0.2, 0.25, 0.5\}$.

4.6. Numerical example: two-dimensional plate with a hole

Consider the two-dimensional plate in subsection 2.7.2. We are now interested in inferring the Young’s modulus using deformation measurements. The idea is also to update the failure probability of the plate structure after identifying the posterior Young’s modulus field.

As in subsection 2.7.2, a lognormal random field with mean value $\mu_E = 2 \times 10^5$ MPa and standard deviation $\sigma_E = 3 \times 10^4$ MPa is used to represent the Young’s modulus $E(\mathbf{x})$. We select the isotropic exponential kernel (2.25) to model the covariance function of the underlying Gaussian field. The prior for the Young’s modulus is described in terms of the KL expansion of the form (2.44). The solution of the eigenvalue problem (2.31) is computed with the Nyström method using 160 GL points in each direction.

Strain sensing gauges are used to measure the deformation of the plate. These measured signals are affected by noise due to the misalignment of the strain gauges, temperature conditions, and other factors. The measurement error is modeled as additive and mutually independent from the uncertain quantities by a joint Gaussian density with mean zero and exponential covariance function. The standard deviation is $\sigma_{\text{obs}} = 1 \times 10^{-4}$, and the correlation length is $\ell_{\text{obs}} = 0.08$ m. The resulting likelihood function is a $2m$ -dimensional Gaussian since there are two strain components (horizontal x_1 and vertical x_2 directions) per sensor location. The measurement noises at the same location in the two directions have correlation coefficient 0.25,

and measurement noises at different locations are independent.

The set of $m = 20$ axial strain observations $\tilde{\mathbf{y}} = [\varepsilon_{x_1}(\tilde{\mathbf{x}}), \varepsilon_{x_2}(\tilde{\mathbf{x}})]^\top$ is generated by simulation at hot-spots locations of the plate $\tilde{\mathbf{x}}$ (Figure 4.15). The observations are computed by solving the forward problem using a finer mesh (779 elements). This mesh is also generated in **Gmsh** [86]. The underlying true Young’s modulus is generated using the full covariance information, assuming an exponential kernel with $\ell_{\text{true}} = 0.10$ m and applying the same noise used for the likelihood.

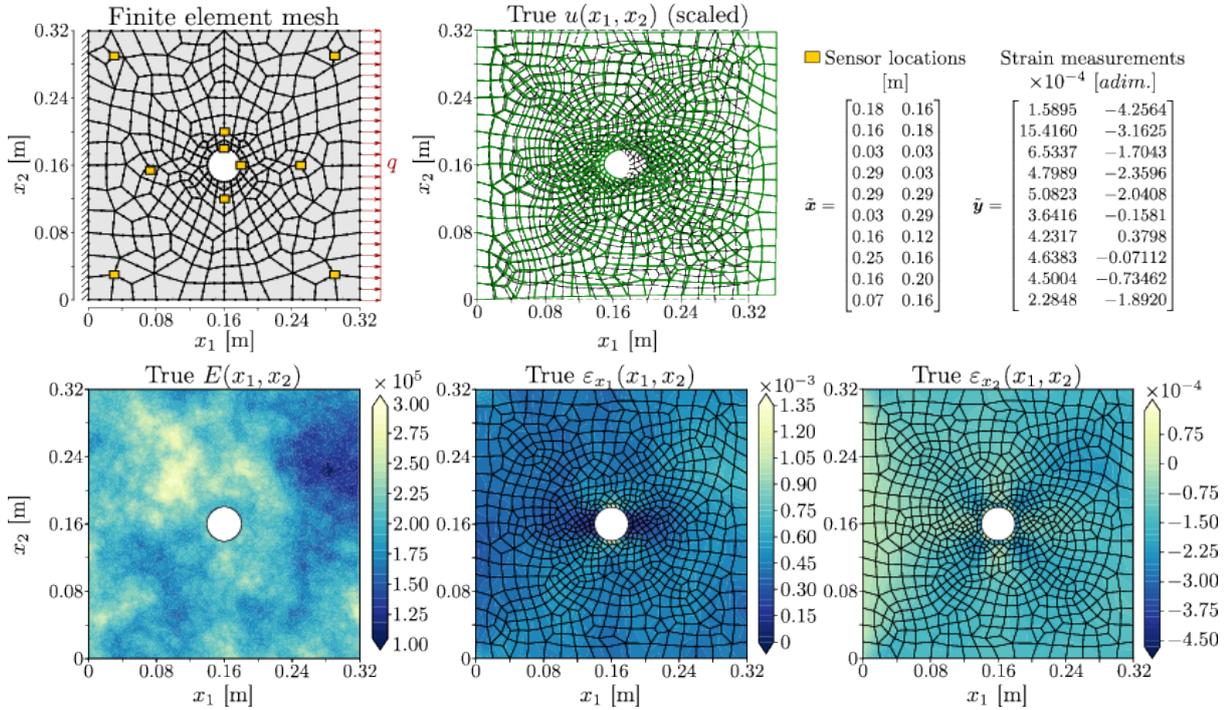


Figure 4.15: Thin plate: structural configuration, finite element mesh, sensor locations, observations, and true fields. Note that a finer mesh is used to generate the “truth” (and hence the data), and a coarser mesh is used for the Bayesian inversion.

4.6.1. Approximated posterior

The associated forward quantities (deformation components) are estimated after including the KL expansion (2.44) into the model PDE in (2.43). This represents a standard stochastic forward problem that is solved using a Monte Carlo-based stochastic finite element method [90]. The posterior random field is estimated by fixing the distributed load as $q = 1.6\mu_q$. This value guarantees that the model response lies in the linear range, i.e., the maximum principal stress σ_1 (defined in (B.19a)) is lower than the yield limit of the material.

Since closed-form expressions are cumbersome to derive in this case, the target posterior random field is explored through samples. This task is carried out using the BUS-SuS method as described in subsection 4.3.1. In the following, 25 independent simulations of the BUS-SuS method with $N = 10^4$ samples per level, and intermediate level probability of 0.1 are employed. The results are shown as the average of those simulations. The value of the constant c is estimated beforehand by optimization for each particular setting. This is to ensure that we are solving the same inference and reliability updating problems at each simulation.

The following parameter cases are considered:

- The number of terms in the KL expansion is chosen as $k \in \{5, 10, 50, 100, 500\}$; and
- the correlation length of the prior flexibility from $\ell \in \{0.02, 0.16, 0.30\}$ m.

The reference solution is computed using a KL expansion with $k_{\text{ref}} \in \{5311, 2609, 833\}$ terms, for the correlation lengths $\ell \in \{0.02, 0.16, 0.30\}$, respectively. The values are selected based on the averaged relative variance error in the prior random field, such that they represent the 95% of the variability for $\ell = 0.02$, and the 99% of the variability for $\ell \in \{0.16, 0.30\}$. We choose a lower precision for the smaller correlation length to reduce the computational cost of the simulation. Note that the reference values shown in Table 2.2 are different, since in this study we employ a finer GL grid for the solution of the KL eigenvalue problem.

Evaluating the posterior approximation

For different number of terms in the KL expansion and correlation lengths, we plot in Figure 4.16 the posterior mean random field. The posterior mean more closely resembles the reference solution with increasing number of terms in the KL expansion (Figure 4.15). We also evaluate the standard deviation fields (the plots are omitted); the magnitude of the standard deviation is larger in the vicinity of the boundaries of the domain (in the order of 0.7×10^5 MPa), and smaller at the measurement locations (in the order of 0.1×10^5 MPa), especially at the upper part of the hole where there is high stress concentration and two sensors are used to collect the strain data. We observe that for correlation lengths larger than the underlying true ℓ , the posterior mean estimates are smoother across the domain even at the measurement locations. These estimates should be used carefully when making predictions since they might lead to overconfident results.

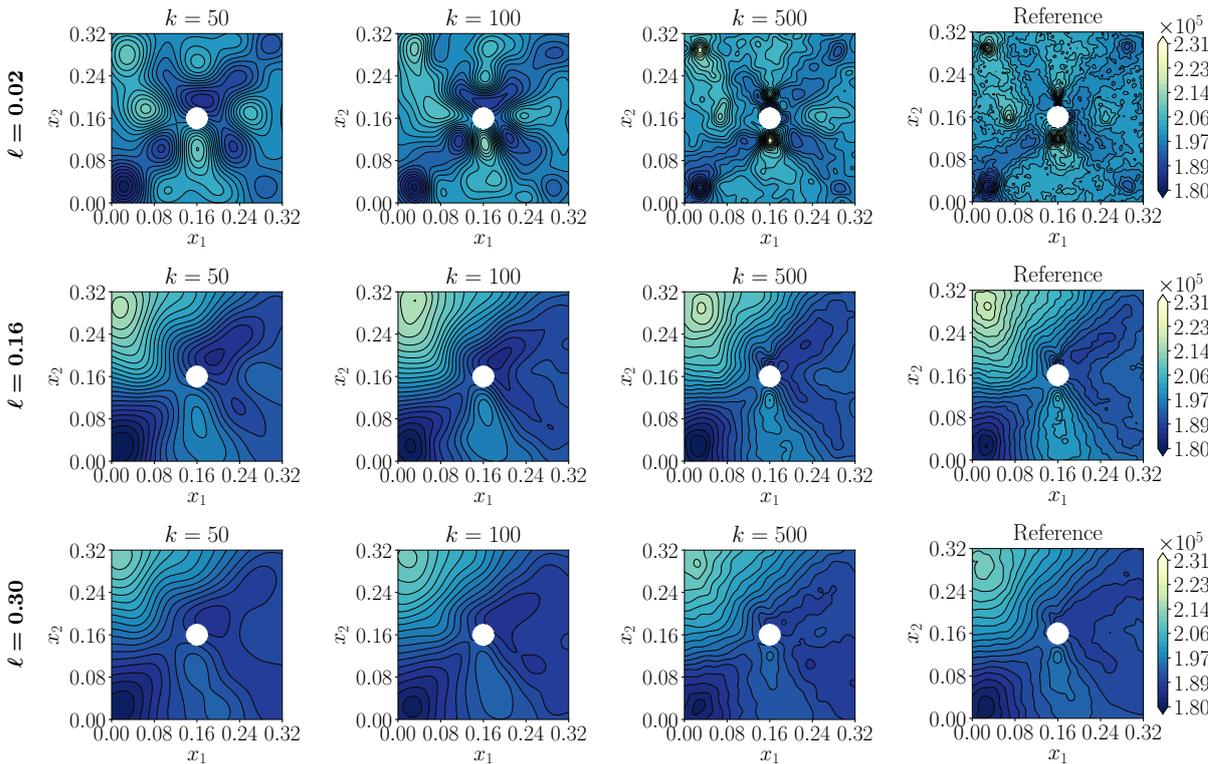


Figure 4.16: Posterior mean of the Young's modulus for different terms in the KL expansion (columns), and prior correlation lengths (rows).

In Figure 4.17, we plot the 95% CI of the posterior Young's modulus random field approximation. The reference solution is illustrated by the shaded area together with various KL approximations. The results are plotted at two different half cuts of the domain (i.e., fixing $x_1 = 0.16$ and $x_2 = 0.16$). As shown for the example in section 4.5, the KL representations are contained in the reference CI, and they converge to this solution as the truncation order increases.

For the correlation length $\ell = 0.02$, an accurate representation of the posterior random field is considerably more difficult to achieve. In this case, more than $k = 500$ terms are necessary for an accurate approximation of the reference solution.

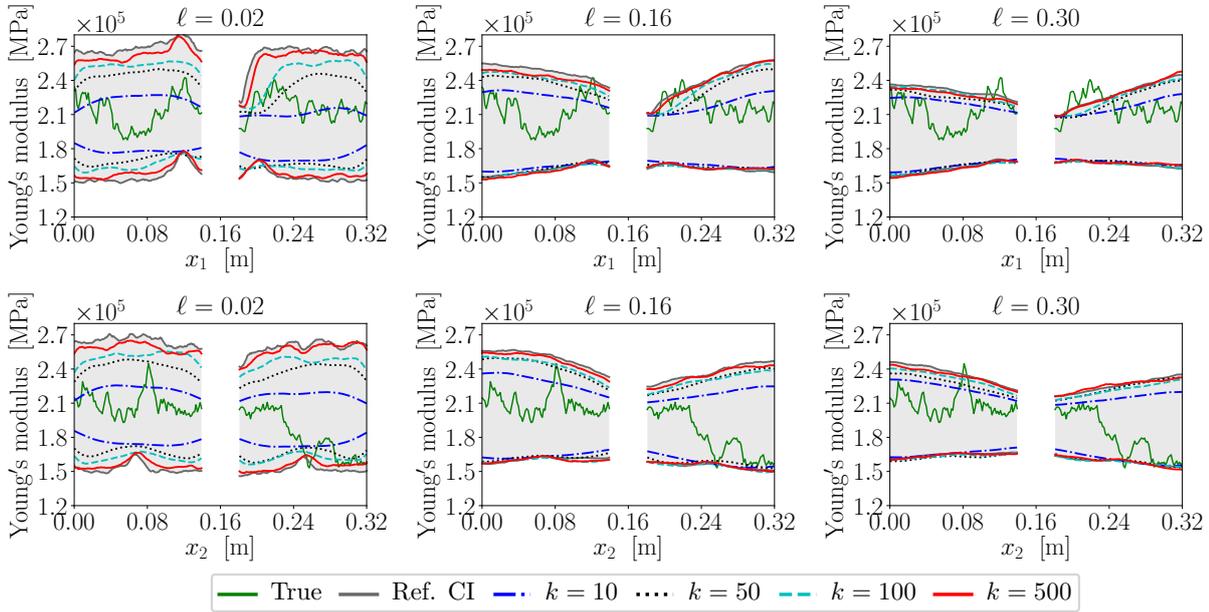


Figure 4.17: Posterior Young's modulus $E(x_1, x_2)$: cuts at $x_2 = 0.16$ m (1st row) and $x_1 = 0.16$ m (2nd row); 95% CI for different terms in the KL expansion and correlation lengths (columns) of the prior Young's modulus. The shaded area corresponds to the reference CI.

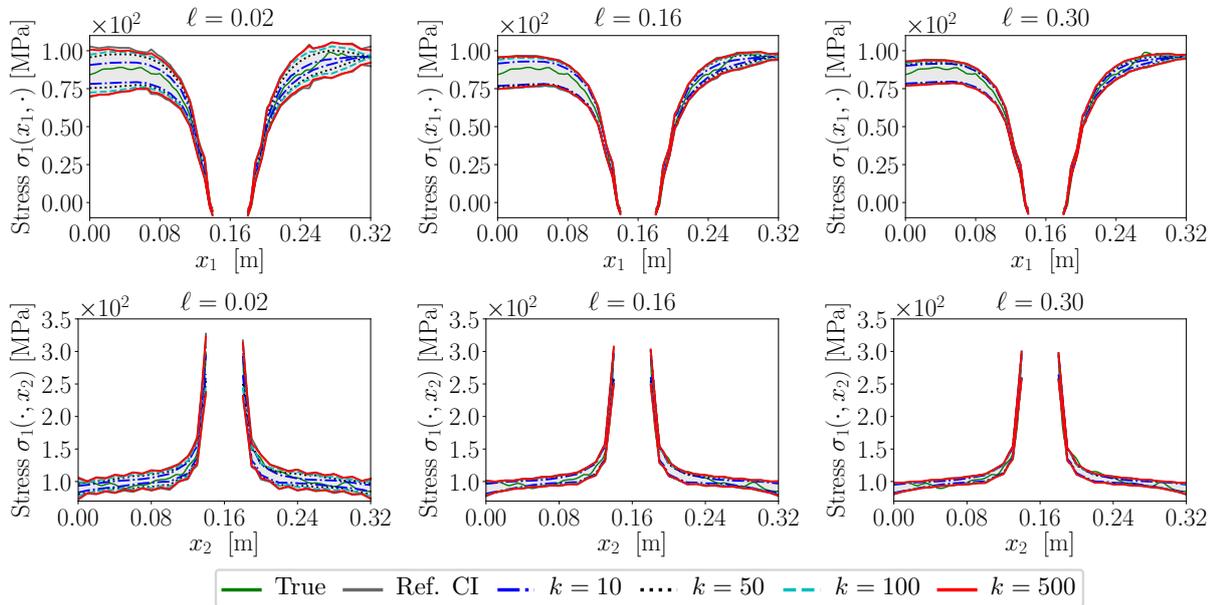


Figure 4.18: Posterior stress $\sigma_1(x_1, x_2)$: cuts at $x_2 = 0.16$ m (1st row) and $x_1 = 0.16$ m (2nd row); 95% CI for different terms in the KL expansion and correlation lengths (columns) of the prior Young's modulus. The shaded area corresponds to the reference CI.

The plots in Figure 4.18 present the approximation of the posterior σ_1 stress random field. Again, the results are shown at the vertical and horizontal half cuts of the domain. We select the principal stress (defined in (B.19a)), as a QoI since the stress field of a plate is highly dependent on the local values of material properties. Therefore, it is expected that the solution of the

Bayesian inverse problem (Young's modulus) will affect directly the values of the σ_1 stress. For larger correlation lengths, the KL expansion represents the posterior σ_1 stress almost as exact as the reference solution, even when using an small number of terms in the expansion. However, for the smaller correlation length, a large truncation order in the KL expansion is required. The situation is different for the displacements CIs (the plots are omitted) since the smoothing influence of the forward operator is larger on this QoI, generating accurate representations with very low number of terms in the KL expansion (similar to the example in section 4.5). A higher-order KL discretization is still necessary for the σ_1 stress given the strong dependence of the stress field on the local fluctuations of the Young's modulus.

Furthermore, the natural logarithm of the model evidence associated to each parameter setting is reported in Table 4.1. We represent the variability of the BUS approach in terms of the standard deviation of the natural logarithm of model evidence estimate. It is seen that the case $\ell = 0.02$ and $k = 500$ terms in the KL expansion yields the largest model evidence.

Evaluating the error in the approximation

The error measures (2.34) are computed for the prior and posterior random fields of the Young's modulus, principal stress σ_1 , and displacement in the horizontal direction u_{x_1} . Figure 4.19 presents the averaged relative variance error $\bar{\epsilon}_{\sigma^2}(k)$ associated to the prior and posterior random field representations of those quantities. The averaged relative mean error $\bar{\epsilon}_{\mu}(k)$ in the posterior is also shown in the third column.

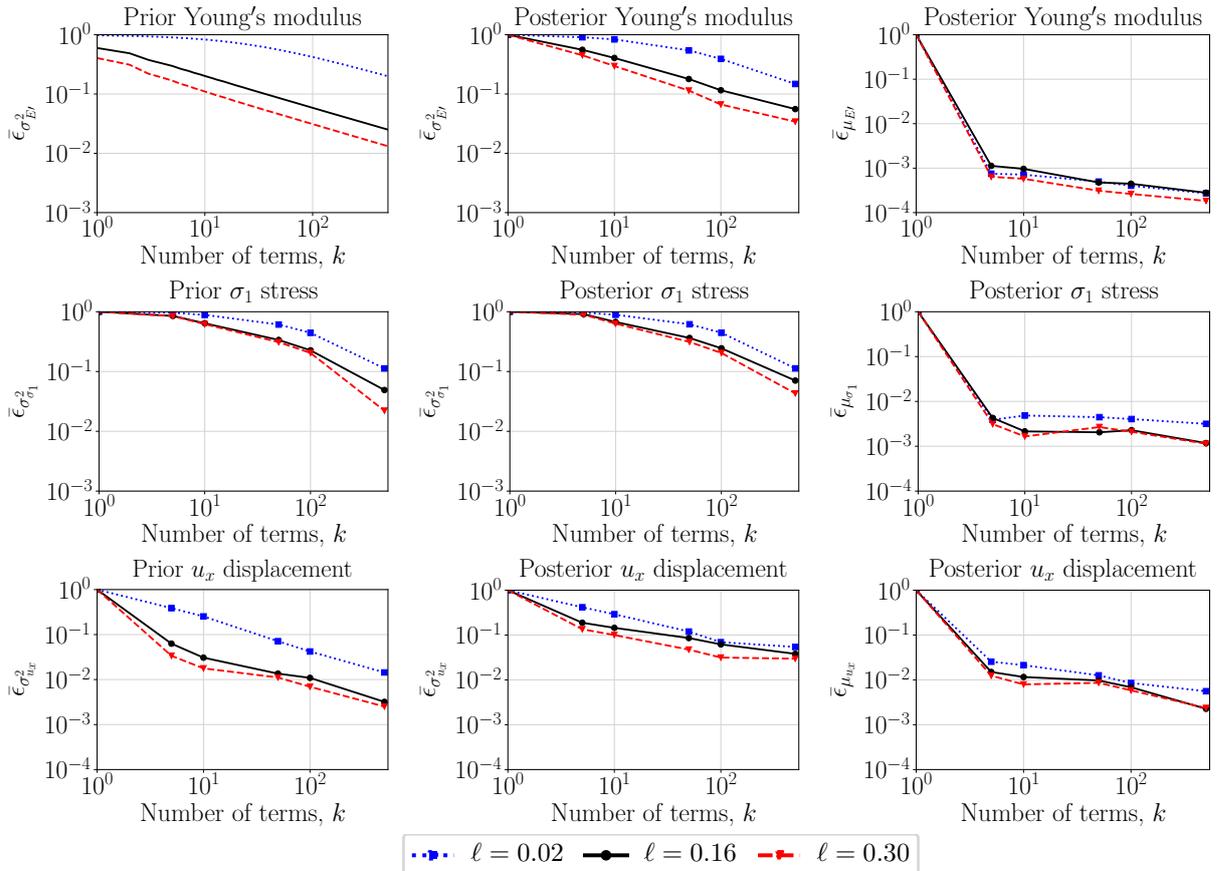


Figure 4.19: Averaged relative error variance in the prior and posterior random fields for the underlying Gaussian Young's modulus E' (1st row), σ_1 stress field (2nd row), and horizontal displacement u_{x_1} (3rd row), using different terms in the KL expansion and correlation lengths. The averaged relative error mean for the posterior random fields is also shown in the 3rd column.

Similar to the example in section 4.5, the errors in the posterior mean approximation are smaller than those associated to the posterior variance. Again, we observe that the variance errors in the prior are smaller than in the posterior. For $\ell = 0.02$ there is no substantial difference, which is attributed to the precision of the reference solution used in the computation of the errors. Since there is still a percentage of the variability that is not represented by the reference, it is expected that the error gap between prior and posterior will increase as the accuracy of the reference solution improves. Compare with the results of example in subsection 2.7.2; in Figure 2.14 the prior variance errors are shown for Matérn kernels with $\nu \in \{2, \infty\}$ and correlation length $\ell = 0.02$. The prior errors shown in the first column of Figure 4.19 are different since we use in this study $\nu = 0.5$ and the number of GL points for the solution of the KL eigenvalue problem is considerably larger. Nevertheless, we can see a similar trend between both solutions and also the presence of the smoothing action of the forward operator.

Different error values are obtained depending on the smoothing influence of the PDE on a given QoI. For instance, due to the Hooke's law in elasticity theory, stress values are linearly dependent on the Young's modulus. Since this parameter represents the inverse quantity, its local random variations will affect the values of the σ_1 stress. However, this does not hold when the forward operator has a "filtering/smoothing" effect on the QoI. This is the case of the displacement solution, for which integration of the PDE operator is required for its computation. Such an averaging process reduces the effect of the local fluctuations of the spatially varying material parameter, and thus the error decay for the σ_1 stress becomes slower as compared to the error decay for the u_{x_1} displacement. Consequently, if we are interested in posterior predictions that are based on displacements, the truncation order of the expansion is not required to be as high as the one required for the representation of the Young's modulus and stresses. We further note that there is a decrease in the accuracy of the computation of the integrals associated to the error measures with increasing dimension (also due to the quality of the reference solution for $\ell = 0.02$). This is seen for the values of the averaged relative mean error, which should decrease linearly (in the log-log scale) with the number of terms in the KL expansion (as shown for the example in section 4.5).

4.6.2. Reliability updating

This chapter concludes with a study on the influence of the random field discretization on the estimation of prior and posterior (updated) rare event probabilities.

We consider that the failure of the plate occurs when the maximum principal stress $\sigma_1(x_1, x_2)$ exceeds the yield tensile strength of the steel (320 MPa); hence the LSF is defined as

$$g(\boldsymbol{\vartheta}) = 320 - \max(\sigma_1(\boldsymbol{\vartheta}; x_1, x_2)), \quad (4.49)$$

where $\boldsymbol{\vartheta} = [\boldsymbol{\theta}, q] \in \mathbb{R}^{k+1}$ is an augmented random vector that includes the random surface load acting on the plate (Figure 4.15), and $\boldsymbol{\theta}$ is the random vector of KL expansion coefficients. Here, $\sigma_1(\boldsymbol{\vartheta}; x_1, x_2)$ stands for the principal stress σ_1 in (B.19a), obtained after solving the stochastic PDE in (2.43) for a given realization of the augmented uncertain parameter $\boldsymbol{\vartheta}$.

The prior random field is used for the reliability analysis (a similar study is presented in [5], with the field assumed to be Gaussian). The estimation of the failure probability in (3.1) is performed using the SuS method. In this case, $N = 10^4$ samples per level are employed with 25 independent simulations. The results are reported in Table 4.1 for increasing number of terms in the KL expansion and correlation lengths. The values of the coefficient of variation (\widehat{cv}) are also included and they are computed from the averaged simulation results. For the failure mode defined in (4.49), the probability of failure increases with the number of terms in the KL expansion. This is because a higher-order random field discretization leads to a larger variability in the spatially varying Young's modulus, which in turn leads to a higher uncertainty

Table 4.1.: Prior and posterior probability of failure using $\ell = \{0.02, 0.16, 0.30\}$ as correlation lengths of the prior Young's modulus (mean and coefficient of variation estimates). The natural logarithm of the model evidence computed by BUS-SuS is also reported (mean and standard deviation estimates).

$\ell = 0.02$						
k	Prior		BUS-SuS inference		Posterior	
	$\widehat{p}_{\mathcal{F}}$	$\widehat{c\bar{v}}$	$\ln Z_{\hat{y}}$	Std.	$\mathbb{P}[\mathcal{F} \hat{y}]$	$\widehat{c\bar{v}}$
5	4.192×10^{-5}	0.056	151.86	0.0298	3.832×10^{-5}	0.312
10	5.575×10^{-5}	0.105	152.15	0.0611	3.209×10^{-5}	0.319
50	1.362×10^{-4}	0.087	152.75	0.0781	3.473×10^{-5}	0.443
100	2.051×10^{-4}	0.081	153.13	0.1329	6.313×10^{-5}	0.462
500	3.993×10^{-4}	0.084	153.74	0.1369	5.237×10^{-5}	0.236
$\ell = 0.16$						
k	Prior		BUS-SuS inference		Posterior	
	$\widehat{p}_{\mathcal{F}}$	$\widehat{c\bar{v}}$	$\ln Z_{\hat{y}}$	Std.	$\mathbb{P}[\mathcal{F} \hat{y}]$	$\widehat{c\bar{v}}$
5	5.208×10^{-5}	0.087	151.45	0.0546	2.592×10^{-5}	0.318
10	6.601×10^{-5}	0.072	151.59	0.1079	1.941×10^{-5}	0.437
50	1.033×10^{-4}	0.105	151.78	0.1418	2.211×10^{-5}	0.541
100	1.196×10^{-4}	0.118	151.93	0.2184	1.441×10^{-5}	0.651
500	1.321×10^{-4}	0.047	152.13	0.1132	1.150×10^{-4}	0.590
$\ell = 0.30$						
k	Prior		BUS-SuS inference		Posterior	
	$\widehat{p}_{\mathcal{F}}$	$\widehat{c\bar{v}}$	$\ln Z_{\hat{y}}$	Std.	$\mathbb{P}[\mathcal{F} \hat{y}]$	$\widehat{c\bar{v}}$
5	4.554×10^{-5}	0.088	151.36	0.0479	1.935×10^{-5}	0.305
10	5.067×10^{-5}	0.119	151.46	0.1558	2.597×10^{-5}	0.518
50	7.600×10^{-5}	0.093	151.53	0.0778	2.398×10^{-5}	0.413
100	7.643×10^{-5}	0.073	151.63	0.2340	1.095×10^{-5}	0.766
500	8.656×10^{-5}	0.085	151.65	0.1401	1.059×10^{-5}	0.608

in the principal stress σ_1 (similar effects are observed in [254]). The magnitude of this effect can change if other types of failure are considered. However, as the truncation order tends to infinity, the contribution of additional terms in the KL expansion becomes insignificant (as also shown in the Bayesian inversion results). In this case, the failure probability converges to a target value that no longer depends on the KL order.

We also address the reliability updating problem in (4.30) using the posterior random field results from subsection 4.6.1. The rare event estimation associated to the LSF $g(\boldsymbol{\vartheta})$ is estimated with the SuS method after applying the modifications described in section 4.4. During the simulation, the $N = 10^4$ posterior samples computed from the 25 independent BUS-SuS simulations are used. The results are shown in Table 4.1 together with the coefficient of variation from the averaged simulations results. Recall that the updated failure probabilities are obtained as a result of two steps, one at which the BUS-SuS approach is used to estimate the Bayesian inverse problem solution, and other one that uses the resulting (correlated) posterior samples from the SuS algorithm to compute the updated probabilities of failure. Consequently, the variability associated to the rare event estimation process is propagated from one step to the other, increasing the coefficient of variation in the updated probabilities (cf. Figure 4.4). The lack of accuracy in the estimates complicates a definite assessment of the posterior failure probabilities. Nevertheless, in this particular example, the posterior failure probabilities tend to be smaller than those computed at the prior level, and thus the updating of the Young's

modulus random field leads to an increase in the reliability estimate of the plate.

In this chapter, we reviewed the theory of Bayesian inverse problems in the context of engineering models. The title of this chapter uses the term *Bayesian analysis* to highlight a more general presentation—the Bayesian approach is not only applied to classical inverse problems, but also to the updating of rare event probabilities. Our main focus was to show the effects of the prior random field choices on the posterior solution. Despite more optimal strategies can be utilized, we opted for a parametric approach in which the studies are performed step by step. This is common practice in engineering applications, where the analyst might be interested in the definition of modeling guidelines, rather than the application of optimal algorithms.

Bayesian inference of random fields: dimension selection and averaging

“... a scientist must also be concerned with a bigger problem: how to decide between different models when both seem able to account for the facts. Indeed, the progress of science requires comparison of different conceivable models; a false premise built into a model that is never questioned cannot be removed by any amount of new data”.

Edwin Thompson Jaynes, 1922-1998 [124, p.601].

Chapter 4 discussed computational approaches for the reduction of parameter uncertainties in numerical models using measurements of the physical system. This represents an inverse problem that is solved by the Bayesian inference framework.

This chapter presents a more general type of inference arising in model choice situations, whereby the model itself needs to be selected from a predefined collection of models. Each model in the set can have different parameters or can be based on particular mathematical assumptions. The Bayesian inference process consists of estimating the conditional posterior distribution induced by the model class, and determining the plausibility of each individual model class. Model selection also arises in Bayesian non-parametric settings where the models are associated to different dimensions of the parameter space. The dimension becomes part of the inference process, and as a result, we are dealing with problems where “the number of things you don’t know is one of the things you don’t know” [111]. This imposes an additional difficulty since in order to estimate the posterior, standard sampling algorithms used to solve the Bayesian inverse problem need to be able to jump between parameter spaces of different dimension.

We start by defining the Bayesian framework for inverse problems when the dimension of the parameter space is variable-dimensional. This is particularized to random field applications where the KL expansion is employed for the discretization. We also suggest a prior for the dimension parameter that penalizes increasing number of terms in the expansion. Thereafter, an overview of the methods that are commonly used to perform this type of inferences is presented. Finally, we propose a method that extends the applicability of the BUS formulation to variable-dimensional settings, and illustrate its applicability and limitations with two engineering examples involving random fields.

5.1. Mathematical framework

Consider an inference setting in which the set of observed data $\tilde{\mathbf{y}}$ is associated not only with one, but with a finite collection of plausible models $\{\mathcal{M}_1, \dots, \mathcal{M}_k, \dots, \mathcal{M}_{k_{\max}}\}$, where $k \in \mathcal{K}$ is a model indicator index, \mathcal{K} is the set of model indices under consideration, and $k_{\max} < \infty$ is a prescribed limit on the collection. The associated discrete-continuous parameter space can be

written as

$$\mathcal{X} = \bigcup_{k \in \mathcal{K}} \mathcal{X}_k = \bigcup_{k \in \mathcal{K}} (\{k\} \times \Theta_k), \quad (5.1)$$

where each \mathcal{X}_k is a parameter subspace related to the model \mathcal{M}_k . Observe that there exist different uncertain parameter vectors $\boldsymbol{\theta}_k \in \Theta_k := \mathbb{R}^k$ for each particular model k , and thus, the goal is to extract information from the data to infer jointly the pairs $(k, \boldsymbol{\theta}_k) \in \mathcal{X}$. For the sake of simplicity in notation, we shall henceforth use the model indicator index k to denote the model \mathcal{M}_k .

Let $\pi_{\text{pr}}(\boldsymbol{\theta}_k | k)$ be a first-level prior density imposed on the parameter $\boldsymbol{\theta}_k$ given the model k , and $\bar{\pi}_{\text{pr}}(k)$ a second-level discrete prior mass specified over the models k (we use the notation $\bar{\pi}$ to indicate probability mass functions). Here, level is meant in a hierarchical sense. The joint posterior density over both, model and parameters, is computed using Bayes' theorem as [85, 126]

$$\pi_{\text{pos}}(k, \boldsymbol{\theta}_k | \tilde{\mathbf{y}}) = \frac{1}{\bar{Z}_{\tilde{\mathbf{y}}}} \bar{\pi}_{\text{pr}}(k) \pi_{\text{pr}}(\boldsymbol{\theta}_k | k) \mathbb{L}(k, \boldsymbol{\theta}_k; \tilde{\mathbf{y}}), \quad (5.2)$$

wherein the evidence is given by the theorem of total probability [193]:

$$\bar{Z}_{\tilde{\mathbf{y}}} = \sum_{k' \in \mathcal{K}} \bar{\pi}_{\text{pr}}(k') Z_{\tilde{\mathbf{y}}}(k') = \sum_{k' \in \mathcal{K}} \bar{\pi}_{\text{pr}}(k') \int_{\Theta_{k'}} \pi_{\text{pr}}(\boldsymbol{\theta}_{k'} | k') \mathbb{L}(k', \boldsymbol{\theta}_{k'}; \tilde{\mathbf{y}}) d\boldsymbol{\theta}_{k'} \quad (5.3)$$

and $Z_{\tilde{\mathbf{y}}}(k)$ is the evidence of the individual model class k (i.e., the expression in (4.45)). The posterior density of the models is obtained by integrating out the parameters in (5.2) as

$$\bar{\pi}_{\text{pos}}(k | \tilde{\mathbf{y}}) = \frac{\bar{\pi}_{\text{pr}}(k) \int_{\Theta_k} \pi_{\text{pr}}(\boldsymbol{\theta}_k | k) \mathbb{L}(k, \boldsymbol{\theta}_k; \tilde{\mathbf{y}}) d\boldsymbol{\theta}_k}{\sum_{k' \in \mathcal{K}} \bar{\pi}_{\text{pr}}(k') \int_{\Theta_{k'}} \pi_{\text{pr}}(\boldsymbol{\theta}_{k'} | k') \mathbb{L}(k', \boldsymbol{\theta}_{k'}; \tilde{\mathbf{y}}) d\boldsymbol{\theta}_{k'}} = \bar{\pi}_{\text{pr}}(k) \frac{Z_{\tilde{\mathbf{y}}}(k)}{\bar{Z}_{\tilde{\mathbf{y}}}}. \quad (5.4)$$

In model selection, the posterior (5.4) can be used to perform the following tasks [245]:

- *Model choice or selection*, which requires the computation of the MAP estimator, $k_{\text{MAP}} = \arg \max_{k \in \mathcal{K}} \bar{\pi}_{\text{pos}}(k | \tilde{\mathbf{y}})$. Model choice is used as indicator of model complexity, i.e., the model that provides the best alignment with the observed data should be preferred over unnecessarily complicated ones. Furthermore, two different models \mathcal{M}_k and \mathcal{M}_l can be compared using Bayes' factors [193]:

$$\frac{\bar{\pi}_{\text{pos}}(k | \tilde{\mathbf{y}})}{\bar{\pi}_{\text{pos}}(l | \tilde{\mathbf{y}})} = \underbrace{\frac{\bar{\pi}_{\text{pr}}(k)}{\bar{\pi}_{\text{pr}}(l)}}_{\text{model prior}} \underbrace{\frac{\int_{\Theta_k} \pi_{\text{pr}}(\boldsymbol{\theta}_k | k) \mathbb{L}(k, \boldsymbol{\theta}_k; \tilde{\mathbf{y}}) d\boldsymbol{\theta}_k}{\int_{\Theta_l} \pi_{\text{pr}}(\boldsymbol{\theta}_l | l) \mathbb{L}(l, \boldsymbol{\theta}_l; \tilde{\mathbf{y}}) d\boldsymbol{\theta}_l}}_{\text{Bayes factor: } B_{k,l}}. \quad (5.5)$$

- *Model mixing or averaging*, which requires the consideration of the whole collection of parameters weighted by $\bar{\pi}_{\text{pos}}(k | \tilde{\mathbf{y}})$. The model mixing solution can be seen as a model posterior predictive distribution at the model level (recall the standard definition (4.7)). In this case, all the collection of models is used for future decisions and the underestimation of uncertainty resulting from choosing only a single model is avoided. Since this process leads to a higher computational cost, only models that are sufficiently likely compared to the MAP estimator may be considered in the analysis. Occam's window and Bayes' factors are used to perform such a model reduction ([193, p.368]).

A common approach in model choice is to use methods that are based on the likelihood function of individual models, including Akaike information, deviance information, Bayesian information, Hannan–Quinn information (fundamental methods are discussed in [203, 245, 22, 193, 148]). Some of these criteria approximate the logarithm of the model evidence such that the penalty term for model complexity (i.e., the KLD (4.48)) depends only on the number of

uncertain parameters [21]. In several cases, this simplification deviates considerably from the actual KLD (see for instance [169]).

Another flexible way to approach the model choice problem is via sampling methods. In this case, the characterization of the posterior requires the exploration of a discrete-continuous space. This task can be performed by performing several independent runs for each particular model; an approach referred to as *within-model simulation*. Several methods are described in the seminal work [85]; we also point out the work [51], where a post-processing step of the MH algorithm to estimate the model evidence is discussed. Alternatively, simultaneous inference on both, model and parameters, is performed by MCMC algorithms that explore the disjoint union space; this approach is called *across-model simulation*. The standard technique is the reversible jump MCMC algorithm [96, 97].

The formulations in (5.2) and (5.4) are also applicable to Bayesian non-parametric settings where in fact there exists only a single mathematical model, but one with variable-dimension parameter [111]. Some examples include, mixture models with an unknown number of components [192], polynomial regression where the degree of the polynomial is variable [198], object recognition where the image being analyzed is composed by an unknown number of shapes [184], or general functional representations that use series expansions for which the number of terms is unknown. In this context, the MCMC algorithms are also referred to as *trans-dimensional MCMC* [97].

The variable-dimensional setting is of particular relevance in the context of learning random fields. Series representations are typically applied in order to project the random field to a finite-dimensional space, e.g., the KL representation. It is common practice to truncate the KL expansion after a finite number of terms based on some variance-representation criterion. This heuristic is generally valid in prior situations when no information or observations about the field are available. In the inversion case, the optimal number of terms in the series expansion is unknown and it is controlled by the data [235].

In the KL expansion, the set of models can be defined by the model indicator indices $\mathcal{K} = \{1, 2, \dots, k, \dots, k_{\max}\}$, where each element defines a truncation order. This truncation order specifies the dimensionality of the standard Gaussian random coefficients of the expansion. Thus, each particular model/dimension k involves a vector of uncertain parameters $\boldsymbol{\theta}_k \in \boldsymbol{\Theta}_k := \mathbb{R}^k$. The aim is to perform simultaneous inference on the discrete random variable k (dimension), and the associated random vector $\boldsymbol{\theta}_k$ (coefficients) of the KL expansion. In the following, we often use the terms model and dimension interchangeably.

5.1.1. Selection of the prior distribution for the dimension

We address the specification of the model/dimension prior by defining a discrete distribution that penalizes increasing dimensionality of the parameter space. In model choice problems, imposing a penalty to complicated models is necessary (see [126, 22, 208] for a discussion). A model with more parameters usually fits the data better than a model with less parameters (recall the definition of the model evidence in (4.47)). However, the actual improvement might be negligible or possible over-fitting of the data can arise [22].

In the KL expansion, we have shown numerically that the information gained by the inclusion of additional terms becomes imperceptible once an optimal truncation is achieved [235] (cf. Figures 4.12 and 4.14, and remark 4.8). See also the work [74]. This behavior is described by small increments in the model evidence as the number of terms in the expansion increases. We conjecture that this result is related to the representation of the posterior covariance since its approximation improves as $k \rightarrow \infty$; however, a rigorous proof is still to be derived.

Based on this observation, we define a prior for the dimension parameter k that penalizes increasing dimensionality. The dimension prior is defined based on the geometry of the parameter space and prior information about the random fields. Our choice is the truncated geometric

distribution:

$$\bar{\pi}_{\text{pr}}(k) = \frac{(1-p)^{k-1}p}{1-(1-p)^{k_{\text{max}}}}, \quad k = 1, \dots, k_{\text{max}}, \quad (5.6)$$

where k_{max} is the upper truncation level, and the success probability $p \in (0, 1)$ marks the decay rate of the probability mass. This parameter allows us to control the shape of the distribution. In practice the parameter space is bounded, and typically some prior knowledge about these bounds is available. We select the parameter p by regulating the behavior of the distribution at the tails, such that $\mathbb{P}[k \leq k_{\text{min}}] = \bar{F}_{\text{pr}}(k_{\text{min}}; p) = \alpha$, where k_{min} is a prescribed bound, α is the probability associated to that event, and $\bar{F}_{\text{pr}}(k_{\text{min}}; p)$ is the cumulative distribution function associated to (5.6) (as function of p). This defines a nonlinear equation that is solved for p , given values of k_{min} , k_{max} and α .

Based on our experiments, k_{min} is chosen as the number of terms in the KL expansion that retains 50% of the variability in the prior random field, and we assign to the event $\{k \leq k_{\text{min}}\}$ a probability of $\alpha = 0.10$. Furthermore, the truncation value k_{max} is selected as the number of terms in the expansion that retains 99% of the variability. By doing so, approximately 90% of the probability mass is concentrated on truncation orders higher than those yielding the 50% variability and smaller than those producing the 99% variability. We found that this heuristic produces a decay p that does not excessively penalize high-order KL terms.

The prior defined in (5.6) fits well into the concept of *penalized complexity* (PC) priors proposed in [208]. Let us consider a *base* prior model $\pi_{\text{pr}}(\boldsymbol{\theta} \mid k = 1)$ and a *flexible* prior model $\pi_{\text{pr}}(\boldsymbol{\theta} \mid k)$, where $k > 1$ is a flexibility parameter that accounts here for the dimension. A PC prior is defined by applying the following propositions [208]:

- (i) Based on Occam's razor principle, a simpler model is preferred until there is enough support for a more complex model. The simplest model is defined in terms of the base prior $\pi_{\text{pr}}(\boldsymbol{\theta} \mid k = 1)$.
- (ii) A measure of increasing complexity is defined as the distance between the flexible and base models, $d(k) = K(\text{D}_{\text{KL}}(\pi_{\text{pr}}(\boldsymbol{\theta} \mid k) \parallel \pi_{\text{pr}}(\boldsymbol{\theta} \mid k = 1)))$, where $\text{D}_{\text{KL}}(\cdot \parallel \cdot)$ is the KLD [91] from the base to the flexible model and K is an increasing function, with $K(0) = 0$.
- (iii) A constant rate penalization is applied to deviations from the base model. This implies an exponential prior distribution on the distance scale, $\bar{\pi}_{\text{pr}}(d(k)) = a \cdot \exp(-a \cdot d(k))$. The prior of the flexibility parameter k is derived from $\bar{\pi}_{\text{pr}}(d(k))$ as, $\bar{\pi}_{\text{pr}}(k) \propto a \cdot \exp(-a \cdot d(k)) |\partial d(k) / \partial k|$, where the last term corresponds to the determinant of the Jacobian of the transformation from d to k .
- (iv) The user is able to define the scaling of the prior by controlling its mass at the tail, such that a can be selected by considering $\mathbb{P}[k \leq k_{\text{min}}] = \alpha$.

Through this construction a PC prior combines prior beliefs with the geometry of the parameter space [208]. The prior for the dimension in (5.6) can be seen as a PC prior in the discrete variable setting of the KL expansion. We derive a proof of this connection in [233]. Finally, we point out that different PC priors have been proposed in the context of random fields, e.g., for different values of a re-parameterized Matérn kernel [83], and an exponential prior with fixed decay rate for the truncation order in the KL expansion in [56].

In the remainder of this chapter, we discuss different MCMC-based methods for the solution of the Bayesian model choice problem (5.4). Recall that in section 4.2 we introduced two MCMC algorithms to sample from arbitrary distributions. MCMC can also be extended to the variable-dimensional context, as it is described next.

5.2. MCMC in varying dimensions

MCMC algorithms are also utilized to sample parameter spaces of varying dimensions associated to the joint posterior in (5.2). Since it is ideal to simultaneously infer the models/dimensions and parameters, we now discuss some methods that operate across the set of candidate models. These algorithms infer jointly models and parameters in a single simulation. In particular, we describe the reversible jump MCMC and saturated space algorithms.

5.2.1. Reversible jump MCMC

The idea of the *reversible jump* Markov chain Monte Carlo (RJMCMC) method [96] is to generate a Markov chain that is able to jump between parameter spaces of different dimension. If the current and proposed states have the same dimension, the proposal move explores different locations within the same parameter subspace, which yields a standard MCMC. If the current and proposed dimensions are different, the detailed balance holds by defining a proposal move that satisfies a dimension matching condition. This is achieved by constructing a one-to-one deterministic transformation (jumping function), ensuring that the image and the domain of the transformation have the same dimension.

Assume that for each dimension k , there exists a target density over \mathbb{R}^k . Excluding the move that implies staying in the same dimension, a particular move type j consists of both the forward and reverse moves [87]:

- *forward move* ($k' > k$), from $\mathbf{x} = (k, \boldsymbol{\theta}_k)$ to $\mathbf{x}' = (k', \boldsymbol{\theta}'_{k'})$: using $r_j = k' - k$, generate a random vector $\mathbf{u} \in \mathbb{R}^{r_j}$ from a known density q_j . The proposed state of the chain $\boldsymbol{\theta}'_{k'} \in \mathbb{R}^{k'}$ is constructed by a deterministic transformation $\mathcal{T}_j : \mathbb{R}^k \times \mathbb{R}^{r_j} \rightarrow \mathbb{R}^{k'} \times \mathbb{R}^{r'_j}$, such that $(\boldsymbol{\theta}'_{k'}, \mathbf{u}') = \mathcal{T}_j(\boldsymbol{\theta}_k, \mathbf{u})$. Here, $\mathbf{u}' \in \mathbb{R}^{r'_j}$ is a random vector generated from a known density q'_j , needed in case the next move is backward;
- *backward/reverse move* ($k' < k$), from $\mathbf{x}' = (k', \boldsymbol{\theta}'_{k'})$ to $\mathbf{x} = (k, \boldsymbol{\theta}_k)$: using a random vector $\mathbf{u}' \in \mathbb{R}^{r'_j}$ generated from q'_j , the proposed state of the chain $\boldsymbol{\theta}_k \in \mathbb{R}^k$ is constructed by the inverse of the transformation $\mathcal{T}_j^{-1} : \mathbb{R}^{k'} \times \mathbb{R}^{r'_j} \rightarrow \mathbb{R}^k \times \mathbb{R}^{r_j}$, such that $(\boldsymbol{\theta}_k, \mathbf{u}) = \mathcal{T}_j^{-1}(\boldsymbol{\theta}'_{k'}, \mathbf{u}')$. Here, $\mathbf{u} \in \mathbb{R}^{r_j}$ is needed in case the next move is forward.

The jumping function or transformation \mathcal{T} needs to be a diffeomorphism¹. The absolute value of the Jacobian determinant (henceforth called Jacobian) at a given point, provides the factor by which the transformation expands or shrinks areas near that point:

$$d\boldsymbol{\theta}' d\mathbf{u}' = \left| J_{(\boldsymbol{\theta}, \mathbf{u})} \right| d\boldsymbol{\theta} d\mathbf{u} \quad \text{where} \quad J_{(\boldsymbol{\theta}, \mathbf{u})} = \frac{\partial(k', \boldsymbol{\theta}'_{k'})}{\partial(k, \boldsymbol{\theta}_k)}. \quad (5.7)$$

One needs to guarantee that the image and the domain of the transformation have the same dimension. Such condition can be achieved by the relation

$$k + r_j = k' + r'_j, \quad (5.8)$$

which is referred to as *dimension matching condition*. If this equality fails, the transformation and its inverse may not be both differentiable [111]. Following the detailed balance condition (4.12) and considering the probabilities of each move type [111],

$$\int_{A \times B} \pi(\boldsymbol{\theta}) \rho_j q_j(\mathbf{u}) \alpha_j(\boldsymbol{\theta} | \boldsymbol{\theta}') d\boldsymbol{\theta} d\mathbf{u} = \int_{A \times B} \pi(\boldsymbol{\theta}') \rho'_j q'_j(\mathbf{u}') \alpha_j(\boldsymbol{\theta}' | \boldsymbol{\theta}) d\boldsymbol{\theta}' d\mathbf{u}', \quad (5.9)$$

¹an invertible function that maps one space to another, such that both, the function and its inverse are continuously differentiable.

where ρ_j denotes the probability of the move j to a given model. The complete transition kernel is obtained by summing over all the possible moves j in (5.9) [111]. As a result, it is sufficient to ensure that for each type of move j , the detailed balance is satisfied. The acceptance probability α_j associated with move type j is given by [96],

$$\alpha_j(k, \boldsymbol{\theta}_k; k', \boldsymbol{\theta}'_{k'}) = \min \left(1, \frac{\pi(k', \boldsymbol{\theta}'_{k'}) \rho_j(k') q'_j(\mathbf{u}')}{\pi(k, \boldsymbol{\theta}_k) \rho_j(k) q_j(\mathbf{u})} \Big|_{J(\boldsymbol{\theta}, \mathbf{u})} \right), \quad (5.10)$$

where the target density is $\pi(k, \boldsymbol{\theta}_k) = \bar{\pi}_{\text{pr}}(k) \pi_{\text{pr}}(\boldsymbol{\theta}_k | k) \text{L}(k, \boldsymbol{\theta}_k; \tilde{\mathbf{y}})$, $\rho_j(k')$ is the probability of the move to model k' , and the final term is the Jacobian of the jumping function. Hastie [110] points out that the Jacobian associated to the move type j , arises from the change of variables and not from the actual dimension jump.

The RJMCMC can suffer from poor sampling performance associated to the definition of the jumping function and the proposal distribution. The potential inefficiency of the method has motivated a considerable amount of tuning step procedures, such as controlling the acceptance probabilities of the jumps and saturated space approaches [43], using additional Markov chains to improve the acceptance rate of RJMCMC [4], marginal density estimators to construct jumping functions [77], hierarchical centering [173], among others. To conclude this subsection, we present a simple illustration of RJMCMC.

Illustration 2 (from [111]). The objective is to find the likelihood model that yields a better representation of a data set $\tilde{\mathbf{y}}$ consisting of $m = 1140$ values (Figure 5.1). Two models are defined:

$$\text{(Model 1: Poisson)} \quad \text{L}(\tilde{\mathbf{y}} | \lambda) = \prod_{i=1}^m \frac{\lambda^{\tilde{y}_i}}{\tilde{y}_i!} \exp(-\lambda). \quad (5.11a)$$

$$\text{(Model 2: Neg. Bin.)} \quad \text{L}(\tilde{\mathbf{y}} | \lambda, \kappa) = \prod_{i=1}^m \frac{\lambda^{\tilde{y}_i}}{\tilde{y}_i!} \frac{\Gamma(1/\kappa + \tilde{y}_i)}{\Gamma(1/\kappa)(1/\kappa + \lambda)^{\tilde{y}_i}} (1 + \kappa\lambda)^{-1/\kappa}. \quad (5.11b)$$

Consider the parameter $\boldsymbol{\theta} = [\theta_1, \theta_2] = [\lambda, \kappa]$. For $k = 1$ we have that $\tilde{y}_i \sim \text{Poisson}(\theta_1)$, and for $k = 2$ we have $\tilde{y}_i \sim \text{NegBin}(\theta_1, \theta_2)$. We define the priors as follows:

$$\text{(Gamma)} \quad \pi_{\text{pr}}(\theta_1 | k = 1) = \text{gamma}(\theta_1; \alpha_\lambda, \beta_\lambda) \quad (5.12a)$$

$$\text{(Gamma)} \quad \pi_{\text{pr}}(\theta_1, \theta_2 | k = 2) = \text{gamma}(\theta_1; \alpha_\lambda, \beta_\lambda) \times \text{gamma}(\theta_2; \alpha_\kappa, \beta_\kappa) \quad (5.12b)$$

$$\text{(Discrete uniform)} \quad \bar{\pi}_{\text{pr}}(k) = \text{Unif}(a, b) \quad (5.12c)$$

with fixed hyperparameters $\alpha_\lambda = 25$, $\beta_\lambda = 10$; $\beta_\kappa = 1$, $\beta_\kappa = 10$; and for the model prior $a = 1, b = 2$. The MCMC sampler needs to jump between models 1 and 2. Since these are related to parameter spaces of different dimensions, we employ the RJMCMC. In this case, the algorithm is tuned as follows:

- *Definition of the jumping function:* let us consider the forward move (from model 1 to 2). In this case, $r = 1$ and we consider the following mapping:

$$(\boldsymbol{\theta}', \mathbf{u}') = \mathcal{T}(\boldsymbol{\theta}, \mathbf{u}) = \mathcal{T}([\theta_1, \emptyset], \mathbf{u}) = \left(\underbrace{[\theta'_1 = \theta_1, \theta'_2 = c \cdot \exp(\mathbf{u})]}_{\boldsymbol{\theta}' \in \mathbb{R}^2}, \underbrace{\emptyset}_{\mathbf{u}' \in \mathbb{R}^0} \right), \quad (5.13)$$

the Jacobian of this forward move is given by:

$$|J_{(\boldsymbol{\theta}, u)}| = \begin{vmatrix} \frac{\partial \theta'_1}{\partial \theta_1} & \frac{\partial \theta'_1}{\partial u} \\ \frac{\partial \theta'_2}{\partial \theta_1} & \frac{\partial \theta'_2}{\partial u} \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ 0 & c \cdot \exp(u) \end{vmatrix} = c \cdot \exp(u); \quad (5.14)$$

for the reverse move (from model 2 to 1), $r' = 0$ and thus:

$$(\boldsymbol{\theta}, u) = \mathcal{T}^{-1}(\boldsymbol{\theta}', u') = \mathcal{T}^{-1}([\theta'_1, \theta'_2], u') = \left(\underbrace{\theta_1 = \theta'_1}_{\boldsymbol{\theta} \in \mathbb{R}^1}, \underbrace{\ln(\theta'_2/c)}_{u \in \mathbb{R}^1} \right), \quad (5.15)$$

and the associated Jacobian is $|J_{(\boldsymbol{\theta}', u')}| = 1/\theta'_2$. Note that this choice satisfies the dimension matching (5.8).

- *Definition of the proposal densities:* we generate the auxiliary random variable u from a Gaussian $q(u) = \mathcal{N}(u; 0, \sigma_q^2)$. In this example, u' is not required and hence, the proposal q' is not necessary.
- *Definition of the jump/move probabilities:* we restrict the dimension jumps only between consecutive models and give the same probability to them. Thus, each $\rho_{j'} = \rho_{1 \rightarrow 2} = \rho_{2 \rightarrow 1} = \rho_j$ is obtained from an discrete uniform law induced by a stochastic matrix $\mathbf{Q} \in \mathbb{R}^{k_{\max} \times k_{\max}}$ ($k_{\max} = 2$), defined as

$$\mathbf{Q} = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}. \quad (5.16)$$

We run the RJMCMC with $N = 5 \times 10^4$ samples, applying a burn-in period of $N_b = 0.1N$. The standard deviation of the proposal for the auxiliary parameters is chosen as $\sigma_q = 1.5$. Moreover, the free parameter included in the jumping function is selected as $c = 0.015$. The initial state of the Markov chain is defined at the model 1, such that $\mathbf{x}_0 = (k, \theta_k) = (1, 2.5)$.

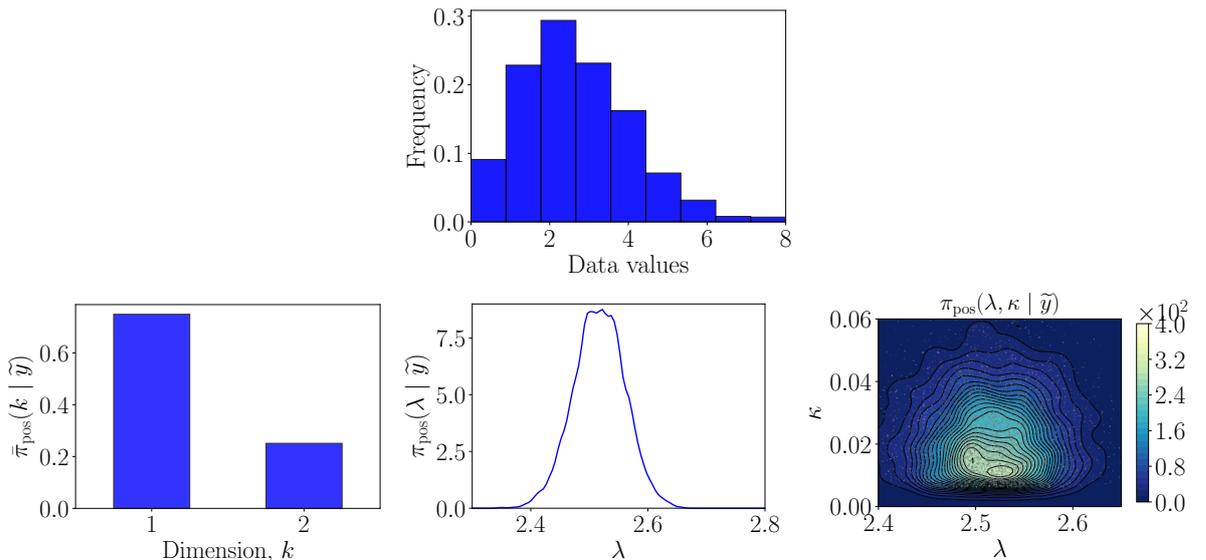


Figure 5.1: RJMCMC results: the histogram of the data set is shown in the 1st row. The posterior of the models and parameters in dimension 1 and 2 are shown in the 2nd row.

The results of a single simulation of the RJMCMC are shown in Figure 5.1. The individual

posterior masses are given by $\bar{\pi}_{\text{pos}}(k = 1 \mid \tilde{\mathbf{y}}) = 0.749$, and $\bar{\pi}_{\text{pos}}(k = 1 \mid \tilde{\mathbf{y}}) = 0.251$; this indicates via model choice that the most appropriate model to represent the data set is given by the Poisson likelihood function. From the RJMCMC simulation, one can also obtain samples of the parameter posterior for each dimension; we also plot those densities which are constructed from the posterior samples via kernel density estimation.

5.2.2. Saturated space methods

Another class of algorithms for model choice problems are the *saturated space* (also referred to as product or composite space) approaches ([47, 24, 94, 43]). The main characteristic is that the parameter space is not particularized to a given dimension k , instead the parameters lie in a space whose dimension contains all dimensions of interest, say k_{max} . The joint posterior in the saturated space is expressed as [94]

$$\pi_{\text{pos}}(k, \boldsymbol{\theta} \mid \tilde{\mathbf{y}}) = \frac{1}{Z_{\tilde{\mathbf{y}}}} \bar{\pi}_{\text{pr}}(k) \pi_{\text{pr}}(\boldsymbol{\theta}_k \mid k) \pi_{\text{pr}}(\boldsymbol{\theta}_{\sim k} \mid k, \boldsymbol{\theta}_k) \mathbf{L}(k, \boldsymbol{\theta}_k; \tilde{\mathbf{y}}), \quad (5.17)$$

where $\boldsymbol{\theta} = [\boldsymbol{\theta}_k, \boldsymbol{\theta}_{\sim k}]$, and the additional component $\pi_{\text{pr}}(\boldsymbol{\theta}_{\sim k} \mid k, \boldsymbol{\theta}_k)$ is the so-called linking density or *pseudo-prior*, where the $\boldsymbol{\theta}_{\sim k}$ denotes the parameters that are not used by the model k . This formulation allows one to employ standard MCMC procedures to variable-dimensional problems. We now motivate these techniques from the viewpoint of nested models.

Nested and variable selection models

For problems involving *nested models*, the parameters of the model k have the same interpretation as the first k parameters of the model $k + 1$. These type of models can also be related to *variable selection models*, where the dimension change amounts to switch on and off components in the parameter vector, i.e., exclusion of a component is equivalent to setting a parameter to zero. In these cases, the parameters associated to different models have a similar meaning, since they are not representing completely different mathematical models. Some examples of these classes include the KL expansion and polynomial regression.

Note that under the KL expansion, the likelihood in (5.2) is independent of $\boldsymbol{\theta}_i$ when $i > k$, and the prior of parameter vector $\boldsymbol{\theta}$ in the saturated space becomes independent of k . Therefore, the (saturated) joint posterior in (5.17) can be written as

$$\pi_{\text{pos}}(k, \boldsymbol{\theta} \mid \tilde{\mathbf{y}}) \propto \bar{\pi}_{\text{pr}}(k) \pi_{\text{pr}}(\boldsymbol{\theta}) \mathbf{L}(k, \boldsymbol{\theta}; \tilde{\mathbf{y}}), \quad (5.18)$$

with $\pi_{\text{pr}}(\boldsymbol{\theta})$ denoting the prior of the parameters in the saturated space $\Theta \subseteq \mathbb{R}^{k_{\text{max}}}$.

The saturated space is oftentimes high-dimensional when dealing with random field applications. In order to avoid convergence deterioration, dimension-independent MCMC algorithms are applied. As discussed in subsection 4.2.2, the standard algorithm is pCN [56, 199]. Given the high-dimensional nature of random fields and the structure of the KL expansion, we focus on saturated space approaches that are based on the pCN algorithm, namely: the Metropolis-within-Gibbs and step-wise samplers. We particularize these methods to model inference problems where the variables of the different models have a nested structure.

Metropolis-within-Gibbs sampler

The Metropolis-within-Gibbs (MwG) algorithm [92, 3] updates the KL coefficients $\boldsymbol{\theta}$ and the dimension k in an alternating manner. In the saturated space context, MwG explores the joint posterior (5.17) with a Gibbs sampling version of the Carlin and Chib [47] algorithm, after including MH steps (details are provided in [94]). The algorithm can also be derived by

expressing the joint posterior (5.18) as the product of the model posterior and the model specific parameter posterior [96]

$$\pi_{\text{pos}}(k, \boldsymbol{\theta} \mid \tilde{\mathbf{y}}) = \bar{\pi}(k \mid \tilde{\mathbf{y}})\pi(\boldsymbol{\theta} \mid k, \tilde{\mathbf{y}}). \quad (5.19)$$

In general Bayesian inference cases, the densities $\pi(\boldsymbol{\theta} \mid k, \tilde{\mathbf{y}})$ may differ abruptly for small changes in the variable k , and thus, the chain can always remain at a given state. However, under the KL formulation (2.29), the coefficients and the dimension are independent *a priori*. This property alleviates potential poor mixing properties in MwG [50].

The idea of MwG is to sample each conditional density in (5.19) by applying two different steps. In the first step, we fix the parameter $\boldsymbol{\theta}$ and sample the conditional distribution $\pi(k \mid \cdot)$ using a standard Metropolis–Hastings sampler to propose a candidate dimension k^* . In the second step, we fix the variable k (accepted in the first step), and sample the conditional distribution $\pi(\boldsymbol{\theta} \mid \cdot)$ to obtain a candidate parameter $\boldsymbol{\theta}^*$ using the pCN proposal in (4.21). The state update in MwG is formally described in Algorithm 6. Observe that this approach requires two likelihood model evaluations for the generation of one state of the chain.

Since k is a discrete variable, the proposal distribution in the standard MH algorithm is replaced by a proposal matrix $\mathbf{Q} \in \mathbb{R}^{k_{\max} \times k_{\max}}$. This proposal is a right stochastic matrix containing the probabilities of the moves [149]. Such probabilities are assigned using a discrete probability law that is controlled by a parameter $1 \leq \rho \leq k_{\max}$ defining the spread or jump lengths of the proposal.

Algorithm 6 State update in the MwG sampler.

Input : Let $(k, \boldsymbol{\theta})$ be the current state of the Markov chain

/ Step 1: for fixed $\boldsymbol{\theta}$, sample the conditional distribution $\bar{\pi}(k \mid \tilde{\mathbf{y}})$ */*

1 Sample the dimension, $k^* \sim \mathbf{Q}(k, :)$

2 Compute the acceptance probability

$$\alpha_k \leftarrow \min \left\{ 1, \frac{L(k^*, \boldsymbol{\theta}; \tilde{\mathbf{y}})}{L(k, \boldsymbol{\theta}; \tilde{\mathbf{y}})} \cdot \frac{\bar{\pi}_{\text{pr}}(k^*) \mathbf{Q}(k^*, k)}{\bar{\pi}_{\text{pr}}(k) \mathbf{Q}(k, k^*)} \right\}$$

3 Sample, $U_k \sim \text{Unif}(0, 1)$

4 **if** $U_k < \alpha_k$ **then**

5 | $k_{\text{next}} \leftarrow k^*$

6 **else**

7 | $k_{\text{next}} \leftarrow k$

8 **end**

/ Step 2: for fixed k_{next} , sample the conditional distribution $\pi(\boldsymbol{\theta} \mid k_{\text{next}}, \tilde{\mathbf{y}})$ */*

9 Sample the coefficients $\boldsymbol{\theta}^* \leftarrow \sqrt{1 - \beta^2} \boldsymbol{\theta} + \beta \boldsymbol{\xi}$, where $\boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\text{pr}})$

10 Compute the acceptance probability $\alpha_{\boldsymbol{\theta}} \leftarrow \min \{1, L(k_{\text{next}}, \boldsymbol{\theta}^*; \tilde{\mathbf{y}}) / L(k_{\text{next}}, \boldsymbol{\theta}; \tilde{\mathbf{y}})\}$

11 Sample, $U_{\boldsymbol{\theta}} \sim \text{Unif}(0, 1)$

12 **if** $U_{\boldsymbol{\theta}} < \alpha_{\boldsymbol{\theta}}$ **then**

13 | $\boldsymbol{\theta}_{\text{next}} \leftarrow \boldsymbol{\theta}^*$

14 **else**

15 | $\boldsymbol{\theta}_{\text{next}} \leftarrow \boldsymbol{\theta}$

16 **end**

Output : $(k_{\text{next}}, \boldsymbol{\theta}_{\text{next}})$

Step-wise sampler

We now construct a step-wise sampler based on the pCN proposal that only requires one acceptance probability step for both, dimension and parameters. The foundations and converge properties of this sampler follow from [24, 94]. It can also be connected to auxiliary variable MCMC algorithms, such as the Møller sampler [149, Ch.4].

Following the posterior (5.18), we can define a proposal distribution $q(k, \boldsymbol{\theta}) = q_1(k) q_2(\boldsymbol{\theta})$. This constitutes a joint proposal across the full state space that accounts for the proposal of the

dimension move $q_1(k) = \mathbf{Q}(k, :)$ and the proposal of the parameters in the saturated space $q_2(\boldsymbol{\theta})$. Under these assumptions, the acceptance probability of the standard MH algorithm becomes [94]

$$\alpha(k, \boldsymbol{\theta}; k^*, \boldsymbol{\theta}^*) = \min \left\{ 1, \frac{q_1(k^*) q_2(\boldsymbol{\theta}^*) \pi(k^*, \boldsymbol{\theta}^* | \tilde{\mathbf{y}})}{q_1(k) q_2(\boldsymbol{\theta}) \pi(k, \boldsymbol{\theta} | \tilde{\mathbf{y}})} \right\}. \quad (5.20)$$

We utilize the pCN proposal for the parameter vector $\boldsymbol{\theta}$ in the saturated space. In this case, the proposal $q_2(\boldsymbol{\theta})$ cancels out with the saturated parameter prior in the target posterior. The resulting acceptance probability simplifies to

$$\alpha(k, \boldsymbol{\theta}; k^*, \boldsymbol{\theta}^*) = \min \left\{ 1, \frac{L(k^*, \boldsymbol{\theta}^*; \tilde{\mathbf{y}})}{L(k, \boldsymbol{\theta}; \tilde{\mathbf{y}})} \cdot \frac{\bar{\pi}_{\text{pr}}(k^*) \mathbf{Q}(k^*, k)}{\bar{\pi}_{\text{pr}}(k) \mathbf{Q}(k, k^*)} \right\}. \quad (5.21)$$

This MH implementation in the saturated space proceeds in a “step-wise” fashion as follows: in the first step of the simulation, a candidate dimension k^* is proposed according to the proposal matrix \mathbf{Q} . In the second step, a candidate parameter $\boldsymbol{\theta}^*$ is proposed using the pCN proposal in (4.21). Afterwards, the candidate pair $(k^*, \boldsymbol{\theta}^*)$ is rejected or accepted jointly according to the probability (5.21). Algorithm 7 describes the procedure in detail.

Algorithm 7 State update in the step-wise sampler.

Input : Let $(k, \boldsymbol{\theta})$ be the current state of the Markov chain

/ Step 1: sample the dimension */*

1 Draw a candidate dimension $k^* \sim \mathbf{Q}(k, :)$

/ Step 2: sample the coefficients using pCN */*

2 Draw candidate coefficients $\boldsymbol{\theta}^* \leftarrow \sqrt{1 - \beta^2} \boldsymbol{\theta} + \beta \boldsymbol{\xi}$, where $\boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\text{pr}})$

3 Compute the acceptance probability

$$\alpha_{k, \boldsymbol{\theta}} \leftarrow \min \left\{ 1, \frac{L(k^*, \boldsymbol{\theta}^*; \tilde{\mathbf{y}})}{L(k, \boldsymbol{\theta}; \tilde{\mathbf{y}})} \cdot \frac{\bar{\pi}_{\text{pr}}(k^*) \mathbf{Q}(k^*, k)}{\bar{\pi}_{\text{pr}}(k) \mathbf{Q}(k, k^*)} \right\}$$

4 Sample, $U_{k, \boldsymbol{\theta}} \sim \text{Unif}(0, 1)$

5 **if** $U_{k, \boldsymbol{\theta}} < \alpha_{k, \boldsymbol{\theta}}$ **then**

6 | $k_{\text{next}} \leftarrow k^*$ and $\boldsymbol{\theta}_{\text{next}} \leftarrow \boldsymbol{\theta}^*$

7 **else**

8 | $k_{\text{next}} \leftarrow k$ and $\boldsymbol{\theta}_{\text{next}} \leftarrow \boldsymbol{\theta}$

9 **end**

Output : $(k_{\text{next}}, \boldsymbol{\theta}_{\text{next}})$

We remark that the jumps in the step-wise sampler take place between nested models. From the RJMCMC viewpoint, a natural jumping function linking the parameter spaces is the identity; this makes the determinant of the Jacobian of the jumping function (5.10) equal to one (see, e.g., [94, 43]).

Illustration 3 (*Polynomial regression*). We conclude this section by illustrating the performance of the previous MCMC samplers with a simple example (a modified version of Example 6.10 in [198]). The problem setup corresponds to inferring the polynomial degree and coefficients of a regression model using observations of a model response. The relationship between the input $\tilde{\mathbf{x}}$ and output $\tilde{\mathbf{y}}$ is modeled as

$$\tilde{y}_i = \sum_{j=0}^k \theta_j \tilde{x}_i^j + \eta_i, \quad i = 1, \dots, m, \quad (5.22)$$

where m is the number of data points, and each η_i is a Gaussian distributed error with mean zero and variance $\sigma_{\text{obs}}^2 = 1$. The model (5.22) can be expressed in matrix form as, $\tilde{\mathbf{y}} = \tilde{\mathbf{X}}\boldsymbol{\theta} + \boldsymbol{\eta}$, where $\tilde{\mathbf{y}}$ is the data vector, $\tilde{\mathbf{X}}$ is the design (Vandermonde) matrix, $\boldsymbol{\theta}$ the parameter vector, and

$\boldsymbol{\eta}$ the vector of random errors. A set of $m = 101$ data points $\tilde{\boldsymbol{y}}$ is generated from the relation (5.22) assuming a quadratic polynomial model with coefficients $\boldsymbol{\theta} = [1.0, 0.3, -0.2]^\top$. The design matrix is constructed at points $\tilde{x}_i = (i - 1)/20$, $i = 1, 2, \dots, m$.

The dimension of the problem is controlled by the polynomial degree in the regression. In order to evaluate which degree better fits the data, we employ the MCMC algorithms discussed in this section. The following settings are considered: (i) the dimension k is supported on the set $\{1, 2, 3\}$, thus we impose on k a discrete uniform distribution between 1 and 3; (ii) the parameters $\boldsymbol{\theta}$ are assumed to be Gaussian distributed with mean zero and covariance $\boldsymbol{\Sigma}_{\text{pr}} = \sigma_{\text{pr}}^2 \mathbf{I}_3$; the prior variance is $\sigma_{\text{pr}}^2 = 4$; and (iii) the jumping proposal matrix \mathbf{Q} is constructed such that all the dimensions are equally possible (i.e., with probability $1/3$).

We run $N_{\text{sim}} = 50$ independent simulations of the MwG and step-wise algorithms. At each simulation, the length of the Markov chains is $N = 1 \times 10^5$ and a burn-in period of $N_b = 0.2N$ is applied. For this illustration, the scaling of the pCN proposal is fixed as $\beta = 0.25$ for both algorithms. In MwG, this choice of β yields an average acceptance probability of 0.334 for the dimension step and 9.949×10^{-3} for the parameter step. In the step-wise sampler, the average joint acceptance probability is 3.019×10^{-3} .

The results of the simulation are shown in Figure 5.2 for MwG in the first row, and the step-wise algorithm in the second row. The values of the data-misfit function are plotted for the N samples of a single simulation; the sharp jumps in the values are due to the change in dimension. The regression solution is shown in terms of the mean and 95% credible intervals (shaded areas) for the two dimensions that give the highest model posterior (linear and quadratic models). The mean value of the parameters computed by MwG is $[1.652, -0.650]$ for the linear, and $[0.916, 0.247, -0.180]$ for the quadratic model; in the step-wise approach these values are respectively, $[1.651, -0.650]$ and $[0.924, 0.243, -0.179]$. The model posterior computed as the average from the N_{sim} independent simulations indicates that the quadratic model gives the best fit. We also show the mean and 95% CIs of the cumulative mean $\hat{\mu}(k_t) = \mathbb{E}[k_{1:t}]$, computed from N_{sim} independent simulations. Here, $k_{1:t}$ indicates the first t states of the dimension chain.

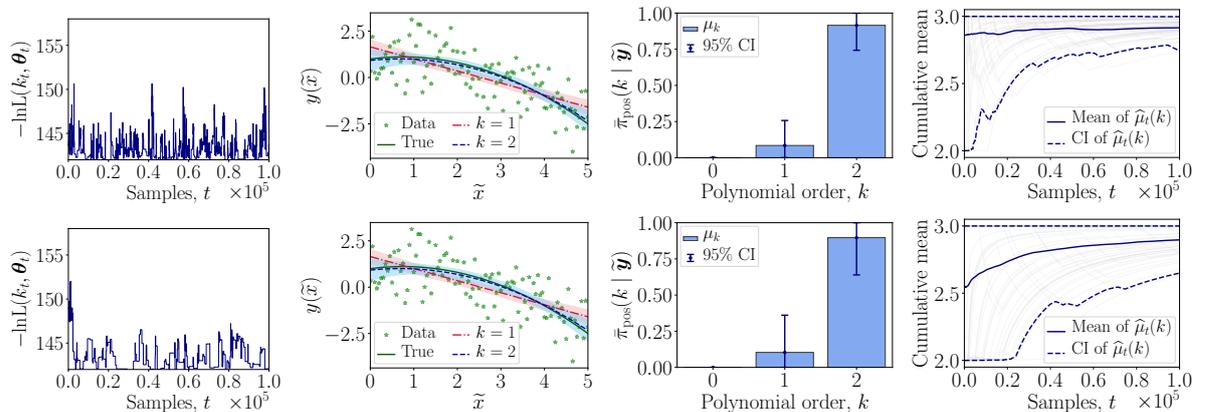


Figure 5.2: Comparison between MwG (1st row) and step-wise sampler (2nd row) algorithms. The evolution of the potential function is shown for a single simulation. The mean and quantile estimators in the remaining plots are computed from 50 independent simulations.

Most of the disadvantages of standard MCMC samplers, such as convergence rate deterioration with increasing dimensions and burn-in/thinning periods requirements, are also present in the MCMC samplers used for Bayesian inference in varying dimensions [78]. In fixed-dimensional cases, specialized sequential algorithms that gradually approach the posterior distribution alleviate some of the issues [174, 198]. These algorithms have been adapted to variable-dimensional settings, including the sequential MCMC approach proposed in [9] and the population-based

RJMCMC presented in [123]. Following this direction, we propose an extension of the classical BUS methodology that is able to perform inference in parameter spaces of different dimension. The idea is to exploit the sequential structure of the BUS-SuS approach to develop an efficient algorithm for Bayesian model choice problems.

5.3. Trans-dimensional BUS formulation

The BUS approach expresses a Bayesian inverse problem as a classical rare event estimation setting (cf., section 4.3). This construction allows us to employ efficient reliability estimation algorithms to sample from the posterior. In this section, we propose an extension of BUS to variable-dimensional inference problems [233]. The basic idea is to re-augment the parameter space by including the discrete dimension variable, which requires minor modifications of the target LSF. We denote this trans-dimensional extension of BUS as tBUS.

Consider the general Bayesian inverse problem (5.2). The rejection sampling algorithm in Section 4.3 can be reformulated accordingly. The joint posterior distribution is characterized by a target function in a discrete-continuous space, $\pi(k, \boldsymbol{\theta}_k) = \bar{\pi}_{\text{pr}}(k)\pi_{\text{pr}}(\boldsymbol{\theta}_k | k)L(k, \boldsymbol{\theta}_k; \tilde{\mathbf{y}}) \propto \bar{\pi}_{\text{pos}}(k, \boldsymbol{\theta}_k | \tilde{\mathbf{y}})$. We choose the proposal distribution to be equal to the joint prior $q(k, \boldsymbol{\theta}_k) = \bar{\pi}_{\text{pr}}(k)\pi_{\text{pr}}(\boldsymbol{\theta}_k | k)$. The acceptance probability in rejection sampling becomes

$$\alpha = \frac{\pi(k, \boldsymbol{\theta}_k)}{\bar{r} \cdot q(k, \boldsymbol{\theta}_k)} = \frac{\bar{\pi}_{\text{pr}}(k)\pi_{\text{pr}}(\boldsymbol{\theta}_k | k)L(k, \boldsymbol{\theta}_k; \tilde{\mathbf{y}})}{\bar{r} \cdot \bar{\pi}_{\text{pr}}(k)\pi_{\text{pr}}(\boldsymbol{\theta}_k | k)} = r \cdot L(k, \boldsymbol{\theta}_k; \tilde{\mathbf{y}}), \quad (5.23)$$

where $r = 1/\bar{r} \in \mathbb{R}_{>0}$. By analogy with the fixed-dimensional setting, the covering constant \bar{r} can be optimally chosen as $L_{\text{max,all}} = \max(L(k, \boldsymbol{\theta}_k; \tilde{\mathbf{y}}))$, i.e., as the maximum of the likelihood function across different dimensions. Note that in terms of the individual BUS covering constants, the tBUS covering factor is $\bar{r} = L_{\text{max,all}} = \max(\bar{c}_1, \dots, c_k, \dots, \bar{c}_{k_{\text{max}}})$ (see Figure 5.3).

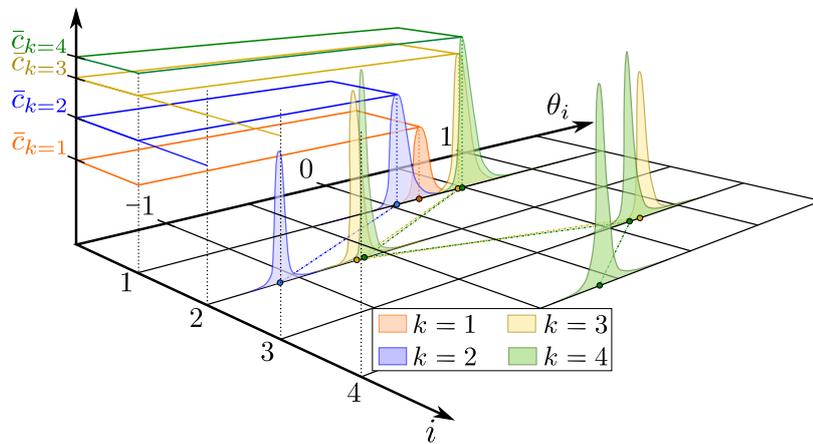


Figure 5.3: Schematic representation of tBUS for nested models. We show the marginal densities of the likelihood at each dimension. The maximum likelihood increases with the dimension of the parameter space.

Thereafter, samples drawn from the priors $k \sim \bar{\pi}_{\text{pr}}(k)$ and $\boldsymbol{\theta}_k \sim \pi_{\text{pr}}(\boldsymbol{\theta}_k | k)$ are accepted if $v \leq \alpha = r \cdot L(k, \boldsymbol{\theta}_k; \tilde{\mathbf{y}})$, otherwise they are rejected. In this case, the LSF in (4.25) and associated \mathcal{H} -space are re-defined as

$$h(k, \boldsymbol{\theta}_k, v) = v - r \cdot L(k, \boldsymbol{\theta}_k; \tilde{\mathbf{y}}), \quad \mathcal{H} = \{(k, \boldsymbol{\theta}_k, v) \in \bar{\mathcal{X}} : h(k, \boldsymbol{\theta}_k, v) \leq 0\}, \quad (5.24)$$

where $\bar{\mathcal{X}} = [K, \boldsymbol{\Theta}_k, \Upsilon]$ is the re-augmented discrete-continuous parameter space ($k \in K \subseteq \mathbb{Z}_{>0}$, $\boldsymbol{\theta}_k \in \boldsymbol{\Theta}_k \subseteq \mathbb{R}^k$ and $v \in \Upsilon \subseteq [0, 1]$).

Since we work in a saturated space, the parameter vector is such that $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^{k_{\max}}$, and thus, we re-define $\bar{\mathcal{X}} = [K, \Theta, \Upsilon]$. Moreover, the maximum of the likelihood function across different dimensions $L_{\max, \text{all}}$ will correspond to the maximum likelihood associated to the dimension k_{\max} (see Figure 5.3).

The posterior distribution in (5.2) can be equivalently computed in tBUS by conditioning the joint discrete-continuous distribution of $(k, \boldsymbol{\theta}, v) \in \bar{\mathcal{X}}$ on the discrete-continuous space \mathcal{H} and marginalizing over the auxiliary uniform random variable v , that is

$$\pi_{\text{pos}}(k, \boldsymbol{\theta} \mid \tilde{\mathbf{y}}) = \frac{\bar{\pi}_{\text{pr}}(k)\pi_{\text{pr}}(\boldsymbol{\theta} \mid k)}{Z_{k, \boldsymbol{\theta}}} \int_0^1 \mathbb{1}_{\mathcal{H}}(k, \boldsymbol{\theta}, v) dv \quad (5.25a)$$

$$\text{with } Z_{k, \boldsymbol{\theta}} = \sum_{k \in \mathcal{K}} \bar{\pi}_{\text{pr}}(k) \int_{\Theta} \int_0^1 \mathbb{1}_{\mathcal{H}}(k, \boldsymbol{\theta}, v) \pi_{\text{pr}}(\boldsymbol{\theta} \mid k) dv d\boldsymbol{\theta}, \quad (5.25b)$$

where the indicator function is defined as $\mathbb{1}_{\mathcal{H}}(k, \boldsymbol{\theta}, v) = 1$ when $(k, \boldsymbol{\theta}, v) \in \mathcal{H}$ and $\mathbb{1}_{\mathcal{H}}(k, \boldsymbol{\theta}, v) = 0$ otherwise.

As commented in remarks 4.3 and 4.4, we also solve the inverse problem in a standard Gaussian space, and express LSF (5.24) in terms of the log-likelihood to achieve numerical stability. This results in the variable-dimensional extension of the LSF in (4.29), given by

$$h_{\ln}(k, \boldsymbol{\vartheta}) = \ln(\Phi(\theta_v)) + \bar{z} - \ln L(k, \boldsymbol{\theta}; \tilde{\mathbf{y}}), \quad (5.26)$$

where $\boldsymbol{\vartheta} = [\boldsymbol{\theta}, \theta_v]^\top \in \mathbb{R}^{k_{\max}+1}$, and $\bar{z} = \ln(\bar{r})$ is optimally the maximum of the log-likelihood function across the different dimensions. In order to efficiently solve the Bayesian model choice problem associated to (5.26), the BUS-SuS algorithm in subsection 4.3.1 can also be extended to the variable-dimensional setting. This is described next.

5.3.1. Trans-dimensional BUS with subset simulation

The extension of BUS-SuS to varying dimensions requires the application of trans-dimensional MCMC algorithms to sample the intermediate conditional densities. We denote the trans-dimensional BUS-SuS methodology as *tBUS-SuS*.

In tBUS-SuS, each intermediate domain is defined as the set $\mathcal{H}_j = \{(k, \boldsymbol{\theta}, v) \in \bar{\mathcal{X}} : h(k, \boldsymbol{\theta}, v) \leq \tau_j\}$, with the threshold level sequence $\{\tau_j\}_{j=0}^{n_{\text{iv}}}$ adaptively selected as in the fixed-dimensional case (cf., subsection 3.2.3). Furthermore, under the LSF (5.26), the standard MCMC algorithms used within BUS-SuS are no longer suitable. In order to sample the intermediate densities $\pi(k, \boldsymbol{\theta}, v \mid \mathcal{H}_j)$, trans-dimensional MCMC methods, such as those discussed in subsection 5.2.2 are required.

When the MwG and step-wise samplers are applied within tBUS-SuS, the acceptance probabilities need to be modified to sample intermediate measures defined by the sequence of threshold levels $\{\tau_j\}_{j=0}^{n_{\text{iv}}}$. In this case, for either Algorithm 6 or Algorithm 7, the target function is equal to an intermediate density conditional on \mathcal{H}_j , i.e., $\pi(k, \boldsymbol{\theta}, v \mid \mathcal{H}_j) = \mathbb{1}_{\mathcal{H}_j}(k, \boldsymbol{\theta}, v) \bar{\pi}_{\text{pr}}(k) \pi_{\text{pr}}(\boldsymbol{\theta} \mid k)$. This implies substituting the likelihood function, for the indicator function in the acceptance probabilities of the two algorithms. For instance, in the step-wise sampler the modification reads as follows:

$$\alpha_{k, \boldsymbol{\theta}} \leftarrow \min \left\{ 1, \mathbb{1}_{\mathcal{H}_j}(k^*, \boldsymbol{\theta}^*) \frac{\bar{\pi}_{\text{pr}}(k^*) \mathbf{Q}(k^*, k)}{\bar{\pi}_{\text{pr}}(k) \mathbf{Q}(k, k^*)} \right\} = \mathbb{1}_{\mathcal{H}_j}(k^*, \boldsymbol{\theta}^*) \underbrace{\min \left\{ 1, \frac{\bar{\pi}_{\text{pr}}(k^*) \mathbf{Q}(k^*, k)}{\bar{\pi}_{\text{pr}}(k) \mathbf{Q}(k, k^*)} \right\}}_{(*)}; \quad (5.27)$$

for the implementation purposes, we use the right term of (5.27): first, a model k^* is proposed

which is accepted with probability (*); then $\boldsymbol{\theta}^*$ is drawn from a pCN proposal; finally the pair $(k^*, \boldsymbol{\theta}^*)$ is accepted, if it lies in the domain \mathcal{H}_j . The modification is analogous for each acceptance rate in MwG. The resulting tBUS-SuS method is summarized in Algorithm 8.

Algorithm 8 tBUS-SuS.

Input : number of samples per level N , conditional probability p_0 , covering constant $\bar{\varepsilon}$, maximum dimension k_{\max} , log-likelihood function $\ln L(\cdot, \cdot; \tilde{\mathbf{y}})$, dimension prior $\bar{\pi}_{\text{pr}}(k)$

```

1 Draw  $N$  samples from the dimension prior,  $k_0 \sim \bar{\pi}_{\text{pr}}(k)$ 
2 Draw  $N$  samples from the standard Gaussian,  $\boldsymbol{\vartheta}_0 = [\boldsymbol{\theta}_0, \theta_{v,0}] \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{k_{\max}+1})$ 
3 Compute initial log-likelihood function values,  $L_{\text{eval}} \leftarrow \ln L(k_0, \boldsymbol{\theta}_0; \tilde{\mathbf{y}})$ 
4 Set  $j \leftarrow 0$  and  $\tau_0 \leftarrow \infty$ 
5 while  $\tau_j > 0$  do
6   Increase intermediate level counter,  $j \leftarrow j + 1$ 
7   Compute LSF values,  $h_{\text{eval}} \leftarrow \ln(\Phi(\theta_{v,j-1})) + \bar{\varepsilon} - L_{\text{eval}}$ 
8   Sort  $h_{\text{eval}}$  in ascending order and create a vector  $\text{idx}$  to store the indices of this sorting
9   Create  $k_{\text{sort}}, \boldsymbol{\vartheta}_{\text{sort}}$  as the dimension and parameter samples  $k_{j-1}, \boldsymbol{\vartheta}_{j-1}$  sorted according to  $\text{idx}$ 
10  Set the intermediate threshold level  $\tau_j$  as the  $p_0$ -percentile of the values in  $h_{\text{eval}}$ 
11  Compute the number of samples in the  $j$ -th intermediate level,  $N_{\mathcal{H}_j} \leftarrow \sum_{i=1}^N (h_{\text{eval}}^{(i)} \leq \max(0, \tau_j))$ 
12  if  $\tau_j > 0$  then
13     $p_{\mathcal{H}_{j-1}} \leftarrow p_0$ 
14  else
15     $\tau_j \leftarrow 0$  and  $p_{\mathcal{H}_{j-1}} \leftarrow N_{\mathcal{H}_j}/N$ 
16  end
17  Select seeds for the MCMC step,  $(k_{\text{seed}}, \boldsymbol{\vartheta}_{\text{seed}}) \leftarrow \{k_{\text{sort}}^{(i)}, \boldsymbol{\vartheta}_{\text{sort}}^{(i)}\}_{i=1}^{N_{\mathcal{H}_j}}$ 
18  Randomize the ordering of the seeds
19  Generate next level values  $\{k_j^{(i)}, \boldsymbol{\vartheta}_j^{(i)}, L_{\text{eval}}^{(i)}\}_{i=1}^{N_{\mathcal{H}_j}}$  from the randomized seeds  $(k_{\text{seed}}, \boldsymbol{\vartheta}_{\text{seed}})$  and intermediate level  $\tau_j$  using a trans-dimensional MCMC algorithm (e.g., Algorithms 6 and 7 with likelihood set equal to the indicator function as in (5.27)). Here, each seed is used to construct a chain with  $N_c = \text{floor}(N/N_s)$  states, where  $N_s = N_{\mathcal{H}_j}$  is the number of seeds
20 end
21 Set the posterior samples,  $k_{\text{pos}} \leftarrow k_j$  and  $\boldsymbol{\vartheta}_{\text{pos}} \leftarrow \boldsymbol{\vartheta}_j$ 
22 for  $k \leftarrow 1$  to  $k_{\max}$  do
23   Find the number of posterior samples that lie in dimension  $k$ ,  $N_k \leftarrow \sum_{i=1}^N (k_{\text{pos}}^{(i)} = k)$  Estimate the model posterior,  $\hat{\pi}_{\text{pos}}^{(k)} \leftarrow N_k/N$ 
24 end
Output: sequence of posterior dimension samples  $\{k_{\text{pos}}^{(i)}\}_{i=1}^N$  and parameter samples  $\{\boldsymbol{\vartheta}_{\text{pos}}^{(i)}\}_{i=1}^N$ , and model posterior  $\hat{\pi}_{\text{pos}}$ .

```

Since finding the constant $\bar{\varepsilon} = \ln(\bar{r})$ poses an additional computational cost, it is convenient to introduce a tBUS-SuS algorithm for which the covering constant \bar{r} is not required as an input. We employ the adaptive BUS-SuS methodology proposed in [27]. The covering constant is updated at each level, leading to a set of values $\{\bar{r}_j\}_{j=0}^{n_{\text{lv}}}$. In order to guarantee the nestedness of the intermediate domains, the threshold levels τ_j are corrected after updating the value \bar{r}_j . First note that from (5.26), a j -th intermediate domain is defined as the set

$$\mathcal{H}_j = \{(k, \boldsymbol{\vartheta}) : \ln(\Phi(\theta_v)) + \bar{\varepsilon}_j - \ln L(k, \boldsymbol{\theta}; \tilde{\mathbf{y}}) \leq \tau_j\}, \quad (5.28)$$

where $\bar{\varepsilon}_j = \ln(\bar{r}_j)$ is the maximum of the log-likelihood function observed up to the j -th intermediate level. The event \mathcal{H}_j associated with the log-scaling constant $\bar{\varepsilon}_j$ and the threshold τ_j can be equivalently expressed by a scaling $\bar{\varepsilon}'_j$ and a modified threshold τ'_j selected as $\tau'_j = \tau_j - \bar{\varepsilon}_j + \bar{\varepsilon}'_j$. We can sequentially update the covering constant $\bar{\varepsilon}_j$ to a new value $\bar{\varepsilon}'_j$, without compromising the distribution of the samples. This is because adjusting the threshold value from τ_j to τ'_j (after updating the scaling from $\bar{\varepsilon}_j$ for $\bar{\varepsilon}'_j$) still defines the same intermediate

domain, as $\mathcal{H}_j = \{(k, \boldsymbol{\theta}) : \ln(\Phi(\theta_v)) \leq \tau'_j - \bar{z}'_j + \ln L(k, \boldsymbol{\theta}; \tilde{\mathbf{y}})\}$ which is equivalent to (5.28). In principle, τ'_j corrects the level τ_j using the residual of maximum log-likelihood values observed at different levels. At the last simulation level when $\bar{z}'_j = \bar{z}_j$, the covering reaches a value that is close or equal to the actual maximum log-likelihood, i.e., $\bar{z}_{n_{lv}} \leq \ln(L_{\max, \text{all}})$. In the limit $N \rightarrow \infty$, the value $\bar{z}_{n_{lv}}$ converges to $\ln(L_{\max, \text{all}})$. Despite the fact that $\bar{z}_{n_{lv}}$ is likely to be smaller than $\ln(L_{\max, \text{all}})$, the samples generated by the algorithm follow the posterior distribution; this is shown experimentally in [27, Fig. 4]. The adaptive tBUS-SuS method is summarized in Algorithm 9.

Algorithm 9 Adaptive tBUS-SuS.

Input : number of samples per level N , conditional probability p_0 , maximum dimension k_{\max}
log-likelihood function $\ln L(\cdot, \cdot; \tilde{\mathbf{y}})$, dimension prior $\bar{\pi}_{\text{pr}}(k)$

- 1 Repeat Lines 1-3 of Algorithm 8
 - 2 Compute the initial maximum log-likelihood, $\bar{z}_0 \leftarrow \max(L_{\text{eval}})$
 - 3 Set $j \leftarrow 0$ and $\tau_0 \leftarrow \infty$
 - 4 **while** $\tau_j > 0$ **do**
 - 5 Increase intermediate level counter, $j \leftarrow j + 1$
 - 6 Compute LSF values, $h_{\text{eval}} \leftarrow \ln(\Phi(\theta_{v, j-1})) + \bar{z}_j - L_{\text{eval}}$
 - 7 Repeat Lines 8-19 of Algorithm 8
 - 8 Compute a new value of the maximum log-likelihood, $\bar{z}'_j \leftarrow \max(\bar{z}_j, \{L_{\text{eval}}^{(i)}\}_{i=1}^N)$
 - 9 Compute the modified intermediate threshold level, $\tau_j \leftarrow \tau_j - \bar{z}_j + \bar{z}'_j$, and update $\bar{z}_j \leftarrow \bar{z}'_j$
 - 10 **end**
 - 11 Repeat Lines 21-25 and Output of Algorithm 8
-

5.3.2. Efficiency metrics for BUS-SuS and tBUS-SuS

In practice, a comparison between within-model BUS-SuS and across-model tBUS-SuS runs is useful to see the advantages of the trans-dimensional formulation. To match both alternatives, we derive two equivalent efficiency metrics.

Consider a collection of samples of a random variable Q representing a monitored QoI. A set of realizations of Q , $\{Q_i^{(j)}, j = 1, \dots, N_{\text{sim}}, i = 1, \dots, N\}$ can be obtained for N_{sim} independent simulation runs. Note that each element $Q^{(j)}$ contains N samples of the QoI. The MCMC steps within tBUS-SuS induce some correlation in the resulting posterior samples. We estimate the effective number of independent samples of Q in the set of all $N \cdot N_{\text{sim}}$ samples as [174]

$$N_{\text{eff}}(Q) = N_{\text{sim}} \cdot \mathbb{V}[Q] / \widehat{\mathbb{V}}[\widehat{\mu}_{Q^{(j)}}] \quad (5.29)$$

where $\mathbb{V}[Q]$ is the true variance of the QoI and $\widehat{\mu}_{Q^{(j)}}$ denotes the sample mean of the QoI obtained at the j -th simulation.

In addition to the value N_{eff} , we also take into account the computational cost. The number of model calls in BUS-SuS, is equal to $N_{\text{call}}^{(j)} = N + (N - N_s)(n_{lv}^{(j)} - 1)$, where $N_s = p_0 \cdot N$ is the number of seeds used in the MCMC steps, and $n_{lv}^{(j)}$ is the number of levels obtained at the j -th simulation. In tBUS-SuS, the number of model calls depends on the utilized trans-dimensional MCMC algorithm. If we use the step-wise sampler, the number of model calls per simulation run is, $N_{\text{call, a}}^{(j)} = N + (N - N_s)(n_{lv}^{(j)} - 1)$. If we employ the MwG algorithm, the number of model calls is $N_{\text{call, b}}^{(j)} = N + 2(N - N_s)(n_{lv}^{(j)} - 1)$ since the likelihood function is evaluated twice at the MCMC steps. We define the efficiency metrics as follows:

$$\text{eff}_{\text{BUS}}(Q) = \frac{N_{\text{eff, BUS}}^{(k)}(Q)}{\bar{N}_{\text{call, BUS}}^{(k)}} \quad \text{and} \quad \text{eff}_{\text{tBUS}}(Q) = \frac{N_{\text{eff, tBUS}}^{(k)}(Q)}{\bar{N}_{\text{call, tBUS}} \cdot \widehat{\bar{\pi}}_{\text{pos}}(k | \tilde{\mathbf{y}})}, \quad (5.30)$$

where $N_{\text{eff,BUS}}^{(k)}(Q)$ and $N_{\text{eff,tBUS}}^{(k)}(Q)$ are the effective number of independent samples of the QoI obtained at each dimension k for each method. Furthermore, in BUS-SuS the total number of calls is given by $\bar{N}_{\text{call,BUS}}^{(k)} = \sum_{j=1}^{N_{\text{sim}}} N_{\text{call}}^{(j)}$. These values are obtained per dimension since N_{sim} runs of BUS-SuS are performed at each specific $k = 1, \dots, k_{\text{max}}$. In tBUS-SuS, the total number of calls using the step-wise sampler is $\bar{N}_{\text{call,tBUS}} = \sum_{j=1}^{N_{\text{sim}}} N_{\text{call,a}}^{(j)}$, or with MwG is $\bar{N}_{\text{call,tBUS}} = \sum_{j=1}^{N_{\text{sim}}} N_{\text{call,b}}^{(j)}$. Since this cost is associated to the whole dimension spectrum and not to each individual k , the value is approximately distributed among dimensions by application of the estimated dimension posterior $\hat{\pi}_{\text{pos}}(k | \tilde{\mathbf{y}})$.

5.4. Numerical applications

We test the proposed method on two examples. The first problem allows analytical verification of the approximations performed by tBUS-SuS. In the second example, we compute several posterior dimension snapshots using a within-model BUS-SuS approach to verify the solution estimated by tBUS-SuS. In all cases, the conditional probabilities are fixed at $p_0 = 0.1$.

5.4.1. Example 1: one-dimensional cantilever beam

The first example involves the cantilever beam problem discussed in section 4.5.

From section 4.5, we know that the solution of the Bayesian inverse problem can be derived analytically. Recall also that the flexibility is modeled by a Gaussian random field prior $\mathcal{N}(\boldsymbol{\mu}_{\text{pr}}, \boldsymbol{\Sigma}_{\text{pr}})$, with constant mean vector $\boldsymbol{\mu}_{\text{pr}} = 1 \times 10^{-4}$ ($\text{kN}^{-1}\text{m}^{-2}$) and covariance matrix $\boldsymbol{\Sigma}_{\text{pr}}$ defined through an exponential kernel (2.25). We set the prior standard deviation to $\sigma_{\text{pr}} = 3.5 \times 10^{-5}$ ($\text{kN}^{-1}\text{m}^{-2}$) and perform a parameter study on the correlation length ℓ .

We use the data set with $m = 10$ point (cf. Figure 4.5). This generates a measurement vector $\tilde{\mathbf{y}} \in \mathbb{R}^m$ with additive and spatially correlated error described by a Gaussian PDF, $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\text{obs}})$, where the covariance structure of the error is constructed from an exponential kernel with standard deviation $\sigma_{\text{obs}} = 1 \times 10^{-3}$ and correlation length $\ell_{\text{obs}} = 1$ m.

We consider different correlation lengths to evaluate its influence on the model posterior estimation. Each correlation length also defines different dimension priors (5.6) as follows:

- for $\ell = 0.1$, $k_{\text{min}} = 17$ and $k_{\text{max}} = 1014$; this yields $p = 6.166 \times 10^{-3}$.
- for $\ell = 0.5$, $k_{\text{min}} = 4$ and $k_{\text{max}} = 204$; this yields $p = 2.586 \times 10^{-2}$.
- for $\ell = 0.9$, $k_{\text{min}} = 2$ and $k_{\text{max}} = 114$; this yields $p = 5.118 \times 10^{-2}$;

recall that the success probability p is computed by solving a nonlinear equation, given the truncation parameters $k_{\text{min}}, k_{\text{max}}$ (see subsection 5.1.1).

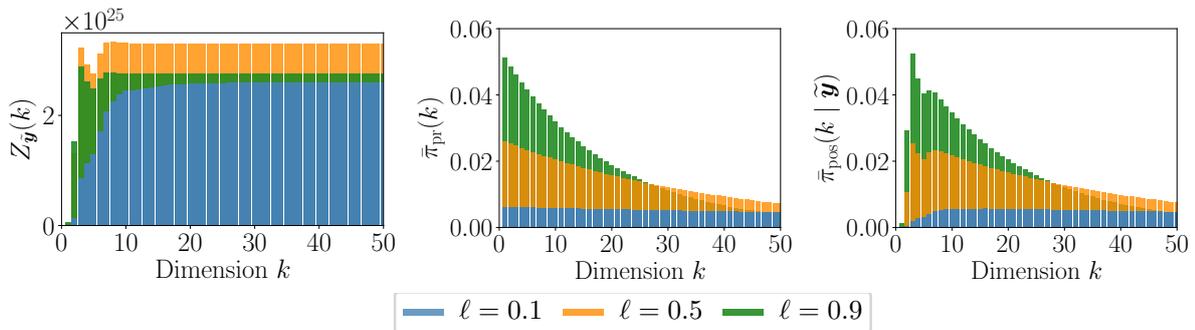


Figure 5.4: Closed-form solutions for different correlation lengths in the prior flexibility random field. Left: model evidence. Center: Model/dimension prior. Right: model/dimension posterior.

For this example, closed-form expressions of the model evidence for each dimension k are available (see subsection 4.5.3), this allows derivation of the model posterior analytically. The priors together with the analytical model posterior and evidence are shown in Figure 5.4. We employ these closed-form solutions as reference to test the performance of tBUS-SuS.

We initially perform parameter studies on the proposal scalings of the trans-dimensional MCMC methods. This allows finding near-optimal parameter settings that are subsequently fixed in the solution of Bayesian model choice problem. Later, we perform a comparison between within-model BUS-SuS runs and across-model tBUS-SuS. This is carried out by the efficiency metric derived in subsection 5.3.2, which is approximately equivalent for both approaches. At the end of this section, we use tBUS-SuS to estimate the posteriors of the dimension and random fields.

Influence of the proposal scaling

We first evaluate the performance of the tBUS-SuS algorithm for different proposal scalings, this includes the parameter ρ of the jump proposal \mathbf{Q} and the parameter β of the pCN proposal. The numerical studies are performed for a fixed prior correlation length $\ell = 0.5$ and for each MCMC sampler presented in subsection 5.2.2. The results are shown for an average of $N_{\text{sim}} = 100$ independent simulations using $N = 10^4$ samples per level.

In this example, the constant \bar{z} can be computed by optimization or from initial log-likelihood function evaluations at the largest dimension k_{max} . Therefore, the parameter studies are performed using standard tBUS-SuS without adaptation of \bar{z} . The performance of the algorithm is measured in terms of the RMSE in the estimation of a target QoI

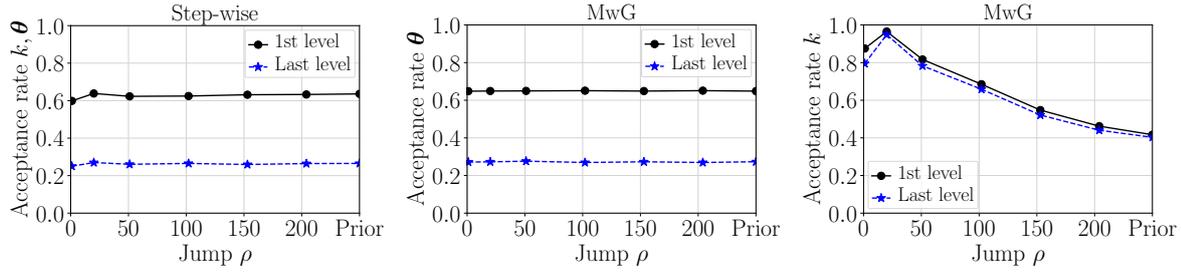
$$\text{RMSE}(k) = \sqrt{\mathbb{E} \left[\left(\widehat{Q}(k) - Q \right)^2 \right]}, \quad (5.31)$$

where $\widehat{Q}(k)$ is the approximated QoI using a KL expansion with k terms, and Q denotes the reference QoI value.

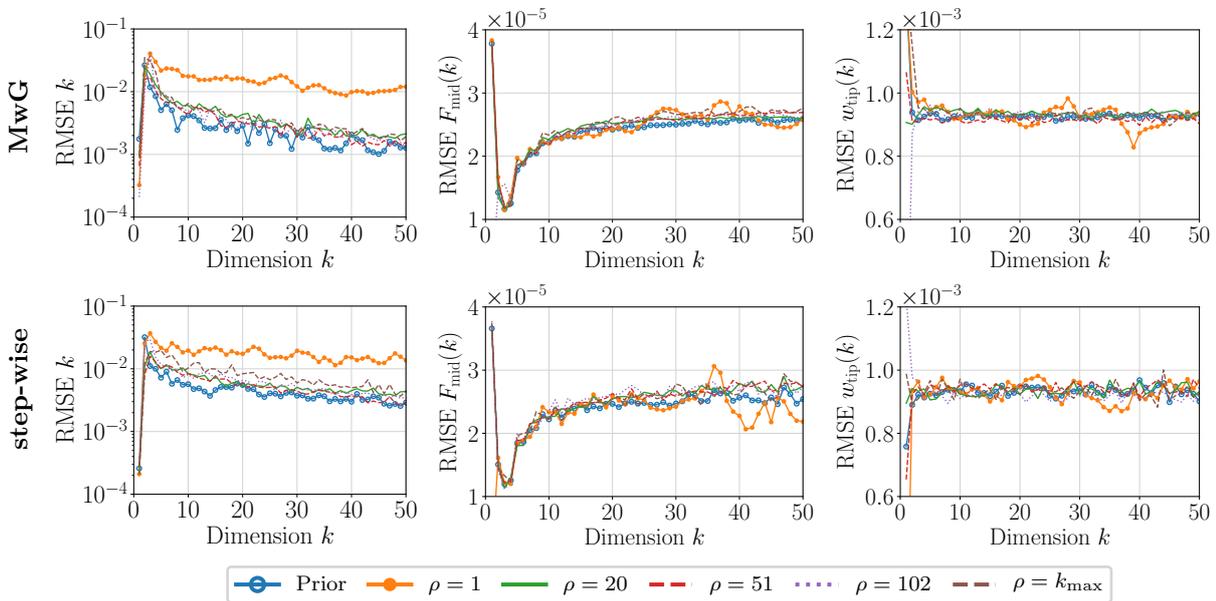
We monitor three posterior QoIs, namely, the dimension parameter k , the flexibility random field at the middle of the beam F_{mid} , and the deflection random field at the tip of the beam w_{tip} . For the evaluation of (5.31), we employ the closed-form model posterior as reference value for dimension QoI, and for the remaining random fields QoIs we use closed-form posterior means, $\mu_{F_{\text{mid}}} = 1.1114 \times 10^{-4}$ and $\mu_{w_{\text{tip}}} = -6.9346 \times 10^{-2}$, as reference values.

Study 1 (Jump length ρ). The pCN proposal scaling is fixed to $\beta = 0.2$ and we estimate the RMSE of the QoIs for different jump lengths $\rho \in \{1, [0.1, 0.25, 0.5, 1.0] \cdot k_{\text{max}}\} = \{1, 20, 51, 102, 204\}$. The jump proposal matrix \mathbf{Q} is constructed from a discrete triangular distribution with mode equal to a given sample k , and lower and upper limits defined by $k \pm \rho$. Alternatively, we also evaluate an independent-sampler-like setting in the dimension sampling steps, that is, instead of drawing samples from \mathbf{Q} , the variable k is sampled directly from the dimension prior $\bar{\pi}_{\text{pr}}(k)$.

Figure 5.5 shows the acceptance rates obtained in the initial and last levels of tBUS-SuS with step-wise (jointly for k and θ) and MwG (separately for θ and k). The acceptance rate in the step-wise sampler is almost insensitive to the jump length ρ , since the parameters are jointly accepted or rejected with the dimension, and the pCN proposal is dominating the acceptance rate. Similar values are obtained for the acceptance rate of the parameter θ in MwG. From the first to the last level of tBUS-SuS, there is a decrease of the acceptance rate since the samples in the posterior region are more concentrated and the scaling β remains fixed. Moreover, it is seen that for increasing values of the jump length ρ , the acceptance rate of k decreases since the spread of the jump proposal becomes larger.


 Figure 5.5: Parameter study on the **jump length** ρ : acceptance rates in the MCMC algorithms.

In Figure 5.6, we plot the evolution of the RMSE with the number of terms in the KL expansion, for each investigated QoI. The MwG sampler yields slightly better approximation of the model posterior as compared to the step-wise sampler. Note that the errors are larger around the model index that yields the largest value of the model posterior; the tBUS-SuS estimates become more accurate after exceeding this truncation level. There are no considerable differences between model posterior approximations obtained by sampling from the dimension prior or using the jumping proposal with an appropriate jump length ρ . Still, we observe that sampling from the prior produces smaller errors in the estimation of the model posterior. If we employ the jumping proposal \mathbf{Q} , it is seen that $\rho = (0.25k_{\max}) = 51$ is preferable for the MwG sampler, whereas intermediate to smaller jumps are more suitable for the step-wise sampler ($\rho \in [0.1k_{\max}, 0.25k_{\max}] = [20, 51]$). For this example, taking dimension samples from the prior appears to be the most appropriate choice, which is related to the model prior and posterior being close to each other (Figure 5.4). Based on these results, we henceforth choose to sample directly from the prior $\bar{\pi}_{\text{pr}}(k)$ at the dimension proposal step of both MCMC samplers.


 Figure 5.6: Parameter study on the **jump length** ρ : RMSE of the investigated QoIs using the MwG sampler (1st row) and the step-wise sampler (2nd row).

Study 2 (Scaling of pCN β). The dimension prior is used to sample k and we estimate the RMSE for different pCN proposal scalings $\beta \in \{0.01, 0.1, 0.2, 0.3, 0.4\}$. As in the previous study, we initially show in Figure 5.7 the acceptance rates obtained at the initial and last levels of tBUS-SuS using step-wise (jointly for k and θ) and MwG (separately for θ and k) methods. The acceptance rates obtained for the parameter θ are comparable in both MCMC samplers. In

this case, the scaling β has no effect on the acceptance of the dimension samples k of the MwG method.

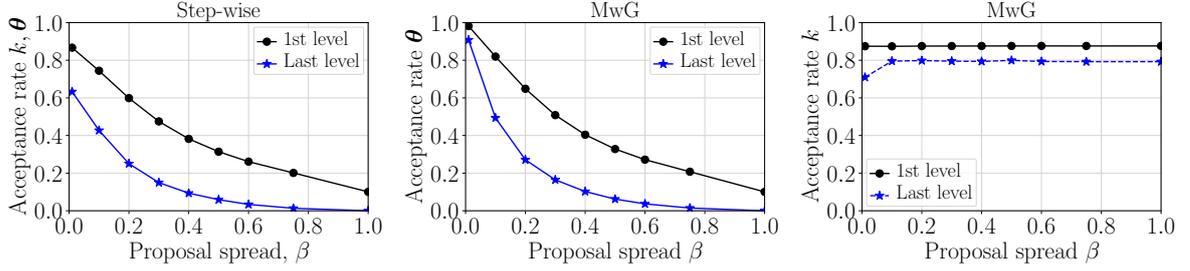


Figure 5.7: Parameter study on the **pCN scaling** β : acceptance rates in the MCMC algorithms.

The evolution of the RMSE with the truncation of the KL expansion for the investigated QoIs and for different β is shown in Figure 5.8. The MwG sampler yields a better estimation of the model posterior than the step-wise sampler. In this case, the scaling β has less effect on the approximation as compared to the jump length ρ . There are only minor differences in the RMSE values generated from the investigated proposal scaling. For the step-wise sampler, the differences in the error magnitudes for different β values are more noticeable compared to MwG. In both methods, the errors in the QoIs corresponding to the random fields are almost the same for all cases.

Note however, that for $\beta = 0.01$ the RMSE associated to the model posterior estimate is higher, and for the same value, the RMSE in the deflection w_{tip} is lower. To keep a balance between both approximations, scaling values $\beta \in [0.1, 0.2]$ are preferable for which the acceptance rate for the step-wise (jointly for k and θ) and for the MwG (for θ) ranges in $[0.2, 0.4]$, at the final simulation level (cf. Figure 5.7). Based on these results, we suggest to adapt the pCN scaling β , such that a target acceptance rate value of $\alpha \in [0.2, 0.4]$ is maintained through the simulation. We employ $\alpha = 0.4$ to keep a good approximation error in both, the monitored random field QoIs and the dimension. One could also increase the target α to obtain a smaller pCN scaling, which is beneficial to reduce the RMSE of QoIs, especially for the deflection.

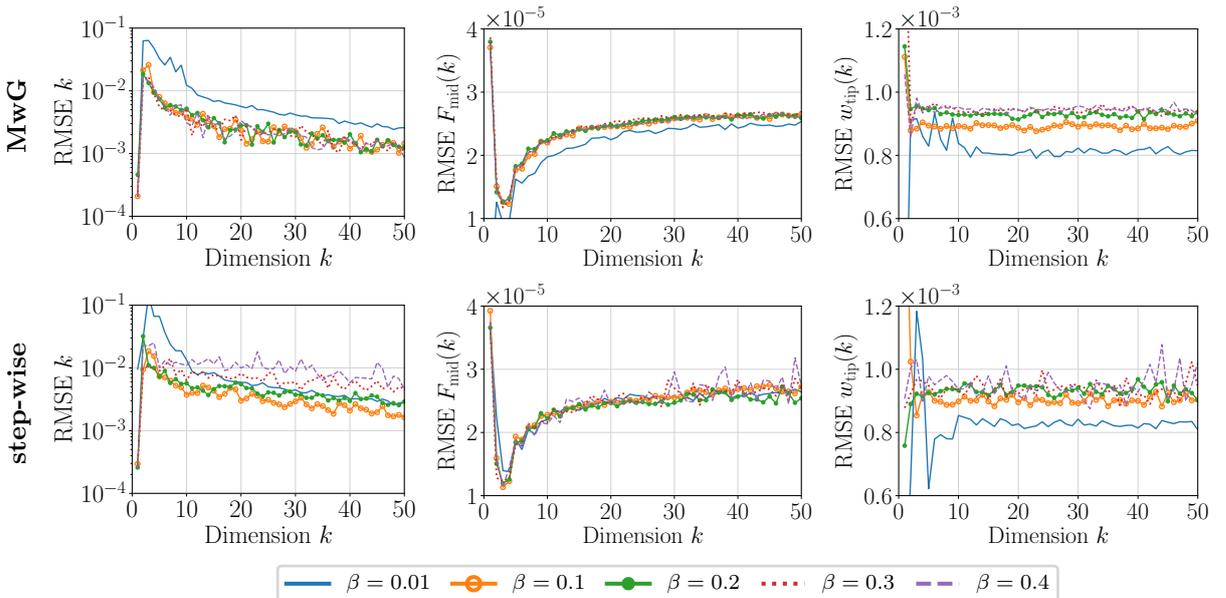


Figure 5.8: Parameter study on the **pCN scaling** β : RMSE of the investigated QoIs using the MwG sampler (1st row) and the step-wise sampler (2nd row).

Computational efficiency

We investigate the computational gain of using the across-model tBUS-SuS, compared to individual within-model runs of BUS-SuS. The objective is twofold: (i) identify the number of effective independent samples in the resulting set of posterior samples, and (ii) compare within-model and across-model simulation using the metric (5.30). In the following studies, the results are computed for an average of $N_{\text{sim}} = 100$ independent runs of the algorithms.

For each correlation length $\{0.1, 0.5, 0.9\}$, we use the reference variances of the QoIs for the estimation of the effective number of samples in (5.29):

- $\sigma_k^2 \in \{24161.22, 1250.80, 328.56\}$ for the dimension,
- $\sigma_{F_{\text{mid}}}^2 \in \{1.143 \times 10^{-9}, 7.484 \times 10^{-10}, 5.364 \times 10^{-10}\}$ for the flexibility, and
- $\sigma_{w_{\text{tip}}}^2 \in \{8.446 \times 10^{-7}, 9.042 \times 10^{-7}, 9.159 \times 10^{-7}\}$ for the deflection.

Study 3 (Standard tBUS-SuS vs. Adaptive tBUS-SuS). Before computing the efficiencies in the estimation of the QoIs related to the random fields, we compare the approximation of the model posterior between standard tBUS-SuS (with pre-defined constant \bar{z} , Algorithm 8) and adaptive tBUS-SuS (Algorithm 9). We employ $N = 10^4$ samples per level.

The adaptive version is a more general method since it is not always possible to define the constant \bar{z} a priori. Moreover, it can produce similar or better results than the standard tBUS-SuS depending on the accuracy of the constant \bar{z} in standard tBUS-SuS, as shown in Table 5.1.

Table 5.1.: Efficiency metrics (5.30) of the dimension parameter k for adaptive and standard tBUS-SuS.

ℓ	adaptive: $\text{eff}_{\text{tBUS}}(k)$		standard: $\text{eff}_{\text{tBUS}}(k)$	
	MwG	step-wise	MwG	step-wise
0.1	8.01×10^{-2}	4.24×10^{-2}	4.00×10^{-2}	1.79×10^{-2}
0.5	5.49×10^{-2}	2.38×10^{-2}	6.13×10^{-3}	1.85×10^{-2}
0.9	3.59×10^{-2}	2.41×10^{-2}	4.03×10^{-3}	2.08×10^{-3}

Recall that the efficiencies are defined as the number of effective samples normalized by the average number of forward model calls. Small efficiency values either indicate that the samples are potentially highly correlated, or that the forward model is evaluated many times. Thus, large efficiencies hint a better sampling method. The results show overall that the efficiency of adaptive tBUS-SuS is larger or comparable to the efficiency provided by standard tBUS-SuS.

We remark that the efficiency of adaptive tBUS-SuS is even larger since we did not account for the extra model evaluations required to find the constant $\bar{z} = \ln(\bar{r})$ in the standard algorithm. This constant is chosen *a priori* as the maximum of 10^5 independent log-likelihood evaluations; this value is also increased by 25% in an effort to make sure that $\bar{r} \geq L_{\text{max,all}}$. It is possible this way of selecting \bar{r} is too conservative ($\bar{r} \gg L_{\text{max,all}}$) and more levels in tBUS-SuS are required to estimate the solution. This is not an issue of the method, but of course leads to a reduced efficiency since more model evaluations are required. In the remainder of the studies, we employ the adaptive tBUS-SuS algorithm for the solution of the Bayesian model choice problem.

Study 4 (Within-model BUS-SuS vs. Across model tBUS-SuS). We compare the efficiencies in the estimation of the QoIs related to the random fields produced by within-model runs of adaptive BUS-SuS and adaptive tBUS-SuS. The number of samples per level used in BUS-SuS and tBUS-SuS are $N = 5 \times 10^3$ and $N = 10^4$, respectively.

Figure 5.9 shows the efficiencies computed with individual adaptive BUS-SuS runs (1st column) and adaptive tBUS-SuS using the two trans-dimensional MCMC algorithms (2nd

and 3rd columns). In the first row, we plot the total number of calls per dimension. In the fixed-dimensional BUS-SuS, the cost increases with the dimension since more intermediate levels are required to reach the posterior. Conversely, the cost is a single value for all the dimensions in across-model tBUS-SuS, and thus we need to distribute it according to the model posterior (cf., subsection 5.3.2). Note in the first row of Figure 5.9 that the total number of calls in within-model BUS-SuS is larger than tBUS-SuS, even when using a larger number of samples per level in tBUS-SuS; also tBUS-SuS with MwG almost doubles the cost compared to tBUS-SuS with the step-wise sampler. In the second and third rows of Figure 5.9, the efficiencies of the random field QoIs are shown per dimension, since different KL truncation orders yield different random field approximations. We remark that the number of effective samples obtained with MwG is larger than using the step-wise method. Nevertheless, the efficiencies in both approaches are similar. This is because the computational cost normalizes the number of effective number of samples in the efficiency metric (5.30), and MwG has almost twice the cost of the step-wise sampler. Finally, we clearly see the advantage of employing across-model simulation algorithms for the solution of Bayesian model choice problems, as compared to single model runs.

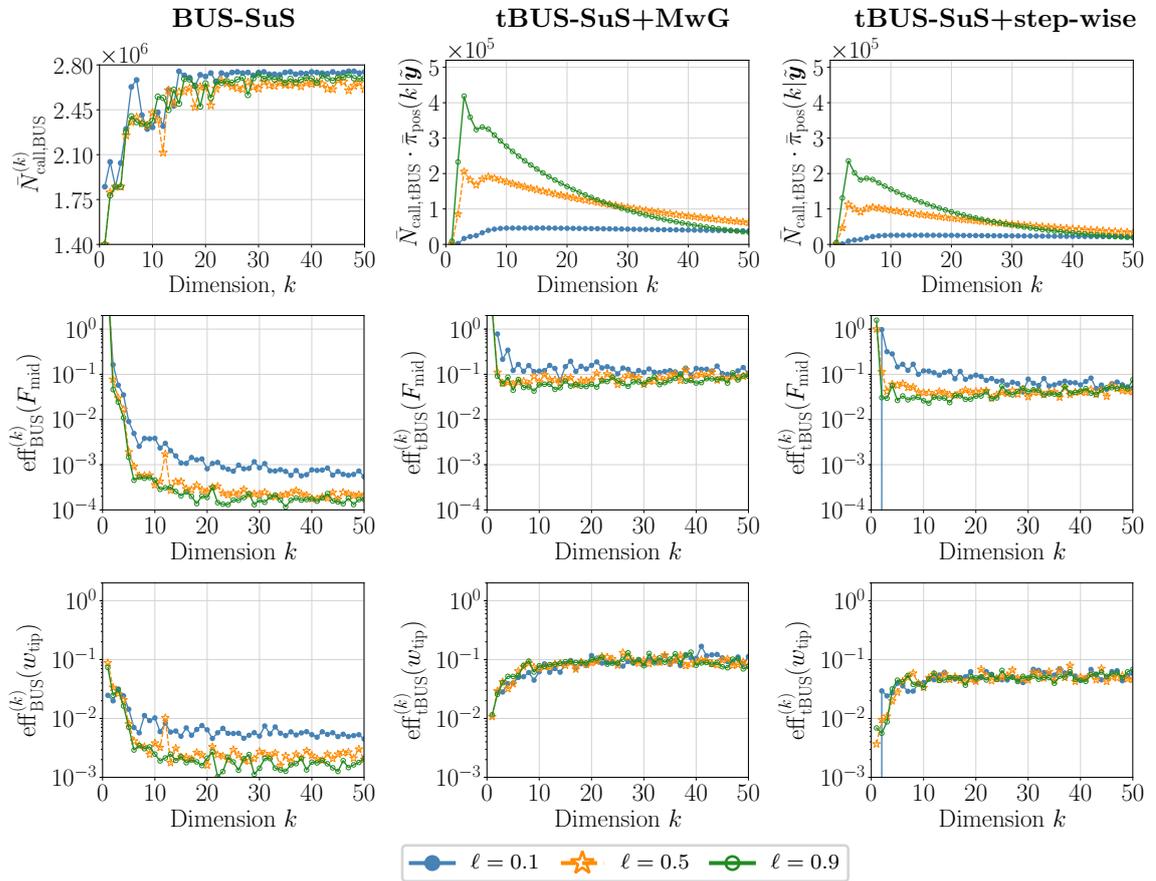


Figure 5.9: Comparison between within-model BUS-SuS and across-model tBUS-SuS: number of model calls and efficiency metrics (5.30) for different correlation lengths and random field QoIs $\mu_{F_{\text{mid}}}$ and $\mu_{w_{\text{tip}}}$. Adaptive BUS-SuS (1st col), adaptive tBUS with MwG sampler (2nd col) and adaptive tBUS with step-wise sampler (3rd col).

Study 5 (Number of samples per level). The influence of the number of sampler per level has been assessed for the adaptive BUS-SuS method in [27]. We perform an analogous study for adaptive tBUS-SuS, but focusing on the estimation of the dimension posterior statistics. For simplicity, we select MwG as the MCMC sampler, and fix the correlation length to $\ell = 0.5$.

Since analytical solutions are available, we compute a normalized RMSE of the approximation

of the model posterior, as N increases. This metric is computed pointwise for each dimension as, $\text{RMSE}^*(k) = \sqrt{\mathbb{E}[(\hat{\pi}_{\text{pos}}(k | \tilde{\mathbf{y}}) - \bar{\pi}_{\text{pos}}(k | \tilde{\mathbf{y}}))^2] / \bar{\pi}_{\text{pos}}(k | \tilde{\mathbf{y}})}$, where $\hat{\pi}_{\text{pos}}(k)$ denotes the estimated posterior. To obtain a global metric for each sample size, these normalized pointwise errors are averaged through all dimensions. The results are shown in Figure 5.10. We also plot the absolute bias and standard deviation components of the RMSE. Note that the standard deviation tends to dominate the RMSE. As expected, the approximation of the dimension posterior improves for increasing N .

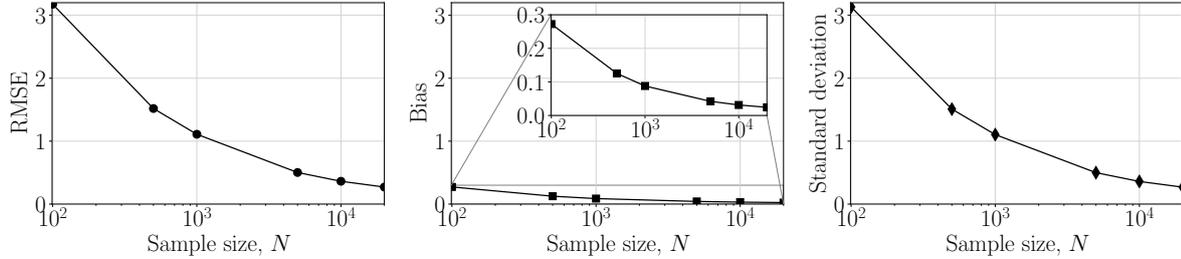


Figure 5.10: Influence of the number of samples per level on tBUS-SuS. Left: RMSE. Center: absolute bias. Right: standard deviation. These quantities are normalized by the exact solution and then averaged through dimensions to obtain a point estimate for each sample size.

We remark that the accuracy in the approximation of the quantities of interest in tBUS-SuS depends mostly on the correlation of the resulting MCMC chains. This correlation tends to increase with the number of levels in tBUS-SuS (a property inherited from SuS). Since this value is always problem specific, a recommendation is to increase N when the resulting number of intermediate levels becoming large.

Approximation of the dimension and random field posteriors

Based on the settings discussed in the previous parameter studies, we utilize the adaptive version of tBUS-SuS for the estimation of the model and random field posteriors. Again, the solution is computed as an average of $N_{\text{sim}} = 100$ simulations, using $N = 10^4$ samples per level.

The approximated model posteriors are shown in Figure 5.11 using the MwG and step-wise samplers. In this case, we plot the mean and standard deviation bounds of the approximation. The shape of the reference model posterior is well-captured for all investigated correlation length cases. The variability of the approximation using the MwG sampler is smaller than the one computed by the step-wise algorithm. Note that the differences in both samplers are larger for smaller correlation lengths. The tBUS-SuS simulations require in average $n_{\text{lv}} = 5$ intermediate levels to reach the posterior for all the investigated correlation lengths.

We also estimate the posterior flexibility and deflection random fields for different correlation lengths. We use as reference the closed-form expressions of the posterior random fields [235]. Figure 5.12 shows the model choice and model mixing solutions in terms of the posterior mean and posterior 95% CI. For the deflection response field, we compute the difference between the prior mean and the 95% posterior CIs (termed *differential* deflection), in order to differentiate the approximations.

The model choice estimate is given by the truncation order that yields the maximum model posterior (Figure 5.11), in this case $k_{\text{best}} = \{10, 3, 3\}$ for the correlation lengths $\{0.1, 0.5, 0.9\}$, respectively. The model mixing estimate takes into account the whole dimension spectrum (up to k_{max}). Note that the reference CIs agree closely with the model mixing estimates since we use all the KL expansions associated to the model posterior for the random field representation. For the larger correlation length, the model choice solution fails to capture the assumed true flexibility in different intervals of the domain.

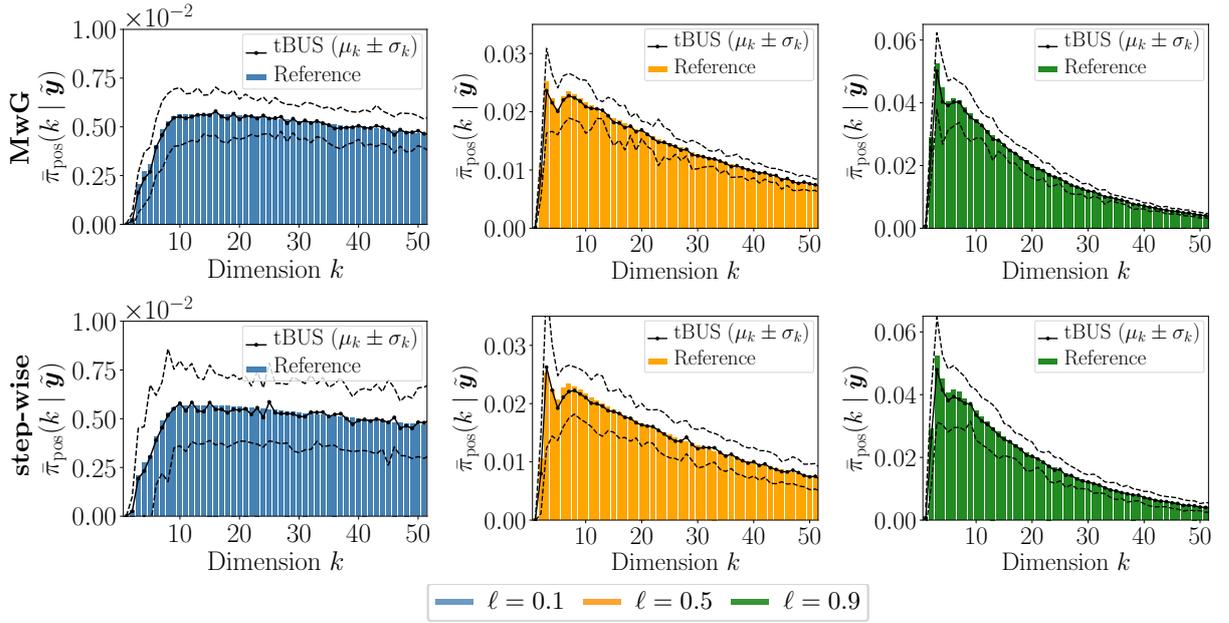


Figure 5.11: Estimation of the model posterior using adaptive tBUS-SuS for different correlation lengths in the prior flexibility random field: MwG (1st row); step-wise (2nd row).

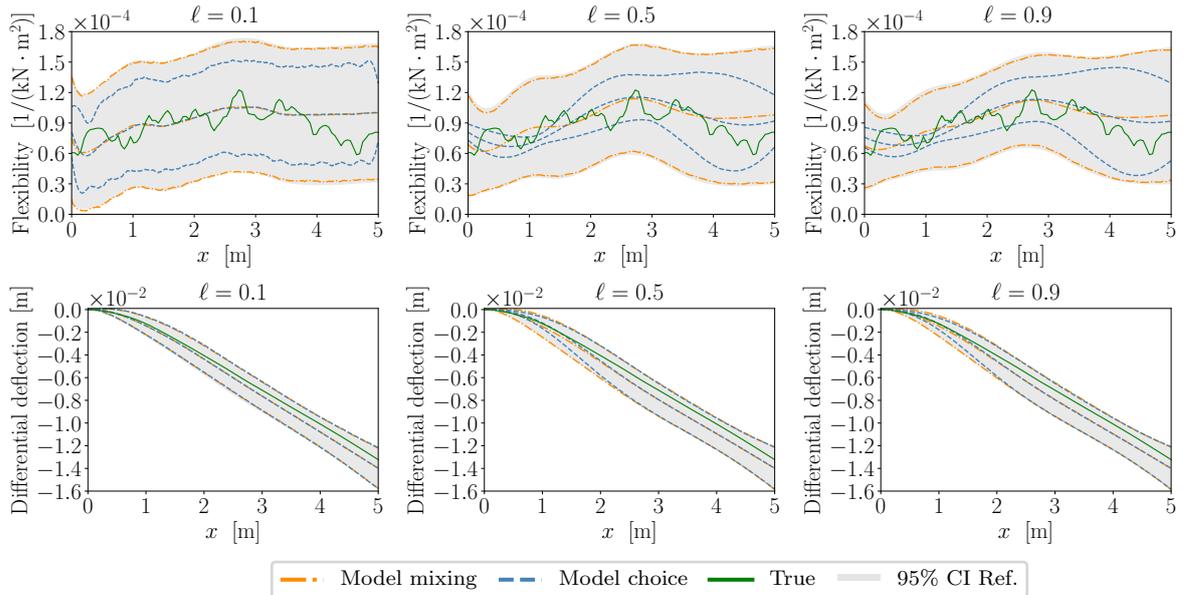


Figure 5.12: Posterior flexibility and deflection random fields for different correlation lengths in the prior flexibility field: estimated mean and 95% CI of the best model (model choice) and the averaging of models (model mixing) using adaptive tBUS-SuS with MwG. The reference 95% CI is highlighted in gray.

Moreover, Figure 5.13 plots the distribution of the KL coefficient samples for $k = k_{\text{best}} = 3$. These plot shows the one- and two-dimensional prior and posterior marginals of the KL coefficients. Significant correlations are apparent among the inferred posterior coefficients.

We conclude this subsection by illustrating the evolution of the samples in tBUS-SuS and its relation to the model posterior. The results are shown for the prior correlation length $\ell = 0.5$. Figure 5.14 shows the prior, second intermediate level, and posterior samples obtained from a single simulation of adaptive tBUS-SuS with MwG.

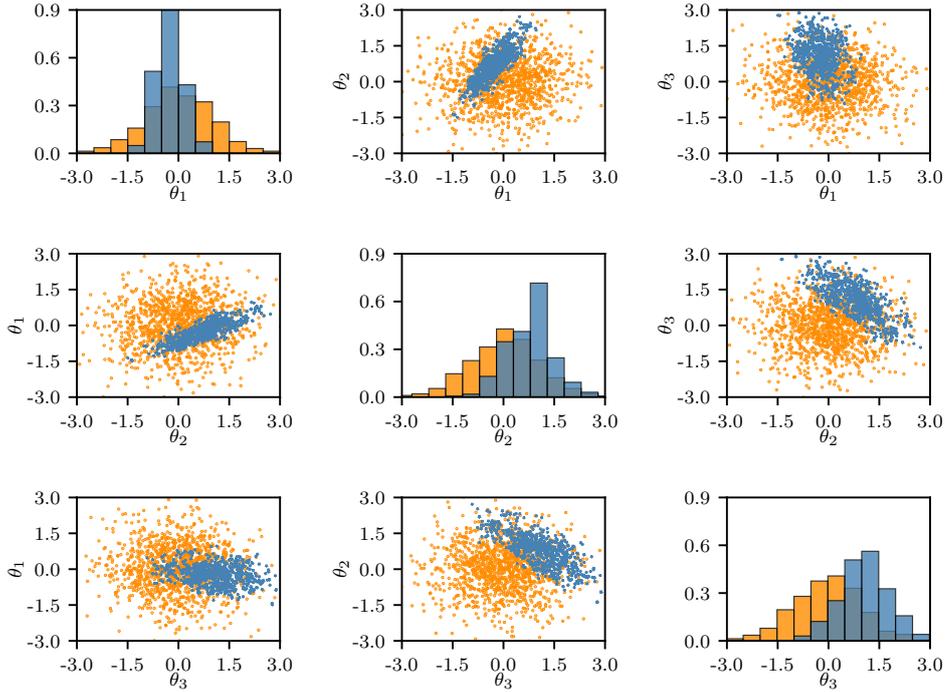


Figure 5.13: KL coefficient samples computed by tBUS-SuS at dimension $k_{\text{best}} = 3$ ($\ell = 0.5$): prior (orange) and posterior (blue).

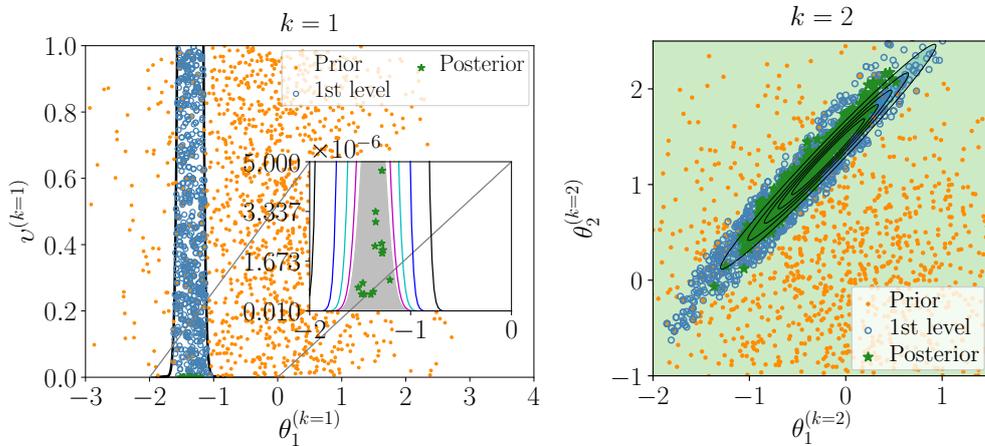


Figure 5.14: tBUS-SuS samples of the KL coefficients at dimensions $k = 1$ (left) and $k = 2$ (right). For $k = 1$, the posterior region is highlighted in gray. For $k = 2$, the contours of the log-likelihood function are also plotted.

The process of sequentially approximating the posterior is shown by the distribution of the samples, starting from the prior and narrowing down to the target posterior. For $k = 1$, we plot the samples that contribute to the model posterior at the first dimension, i.e., the one-dimensional KL coefficient against the auxiliary standard uniform random variable. The tBUS-SuS simulation required $n_{\text{IV}} = 5$ levels to reach the posterior region (highlighted in gray), $\tau = [14.74, 5.04, 2.67, 0.31, 0]$. Note that the value of the model posterior at $k = 1$ is almost zero (Figure 5.4), and hence the amount of samples is considerably reduced as the algorithm evolves from the prior to the posterior measure. Moreover, the maximum log-likelihood at dimension $k_{\text{max}} = 204$ is $\bar{c}_{204} = 68.170$, and at dimension $k = 1$ it is $\bar{c}_1 = 56.529$. Due to the nested structure of the KL expansion, the constant $\bar{\varepsilon} = \ln(\bar{r})$ in the LSF (5.26) is equal to the

maximum log-likelihood at the largest dimension. In this case, the scaling \bar{r} is significantly larger than the value of the covering constant at dimension 1. Thus, we observe that the posterior samples at $k = 1$ are located in a small region of the two-dimensional parameter space. This occurs mainly at lower dimensions, since there exist significant differences between lower- and higher-dimensional likelihood values. We remark that there is an associated reduction of the efficiency, but this does not prevent the algorithm from computing accurate posterior samples. Moreover, with increasing k the values of \bar{c}_k are closer to \bar{r} and the efficiency loss becomes negligible. For $k = 2$, Figure 5.14 plots the components of the two-dimensional KL coefficients; we also show the contours of the log-likelihood function with fixed dimension $k = 2$. In this case, the reduction in the number of samples when updating from prior to posterior is smaller than at dimension $k = 1$. Note that the value of the model posterior at $k = 2$ is larger than zero, and the difference in the probability mass between prior and posterior at $k = 2$ is less substantial (Figure 5.4).

5.4.2. Example 2: two-dimensional diffusion (groundwater flow)

We consider inference of the hydraulic conductivity field of an aquifer using observations of the hydraulic head measured at specific boreholes (see [161, 58]). We define the aquifer on the square domain $D = [0, 1] \times [0, 1]$ km² with boundary ∂D . Spatial coordinates are denoted by $\mathbf{x} = [x_1, x_2] \in D$. The steady-state Fick's second law of diffusion is used to describe the spatial variation of the hydraulic head inside the aquifer. Hence, for a given hydraulic conductivity of the soil $\kappa(\mathbf{x}, \omega)$ and sink or source terms $J(\mathbf{x})$, the hydraulic head $u(\mathbf{x})$ follows the elliptic PDE

$$-\nabla \cdot [\kappa(\mathbf{x}, \omega) \nabla u(\mathbf{x})] = J(\mathbf{x}), \quad (5.32)$$

with Dirichlet boundary condition, $u(\mathbf{x}) = 0$ for $\mathbf{x} \in \partial D$.

The source terms are defined as the superposition of nine weighted Gaussian plumes with standard width $\sigma_J = 1 \times 10^{-3}$ km. The plumes have equal and unitary strengths, and are centered at locations $\boldsymbol{\mu}_J = [0.25 \cdot i, 0.25 \cdot j]$ with $i, j = 1, \dots, 3$, that is

$$J(\mathbf{x}) = \sum_{i=1}^9 \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}_J^{(i)}, \sigma_J^2 \mathbf{I}_2). \quad (5.33)$$

We employ the finite element method to solve the PDE (5.32) using piecewise linear triangular elements on a regular mesh of 80×80 square blocks (each divided into two triangles). Hence, the total number of elements is $2 \times 80^2 = 1.28 \times 10^4$. This selection is based on parameter studies showing that this mesh is sufficiently fine to account for the random field variability associated to the smallest correlation length used in our experiments.

The prior hydraulic conductivity field is modeled as a lognormal random field, $\kappa(\mathbf{x}) := \exp(\kappa'(\mathbf{x}))$. The underlying Gaussian field $\kappa'(\mathbf{x})$ has mean zero and standard deviation $\sigma_{\kappa'} = 1$ km/day. The covariance operator of the Gaussian field is constructed from a Matérn kernel (2.24). We set the smoothing parameter to $\nu = 1.5$. The solution of the eigenvalue problem is computed with the Nyström method using 100 GL points in each direction. Bear in mind that the KL eigenvalue problem is solved only once for k_{\max} terms. This means that the problem is not re-computed for each k since we work on a saturated space where the KL coefficients have a nested structure.

The true conductivity $\kappa(\mathbf{x})$ is a realization of a random field with characteristics similar to the prior. In this case, we set the Matérn kernel parameters as $\nu_{\text{true}} = 2.0$ and $\ell_{\text{true}} = 0.1$. The truncation of the KL expansion used to generate this realization is 312, which captures 99% of the prior variance. The hydraulic head observations $\tilde{\mathbf{y}}$ are obtained at $m = 12$ sensor locations. They are computed from a PDE evaluation of the true conductivity field using a finer finite

element mesh. The measurement error is modeled as additive and mutually independent from the random field. It is defined by a joint Gaussian PDF with mean zero and noise covariance matrix $\Sigma_{\text{obs}} = \sigma_{\text{obs}}^2 \mathbf{I}_m$. The variance of the measurement noise is prescribed such that the observations have a signal-to-noise ratio $V[\tilde{\mathbf{y}}]/\sigma_{\text{obs}}^2 = 120$. The true hydraulic conductivity and hydraulic head fields, together with the source terms and the synthetic data are shown in Figure 5.15.

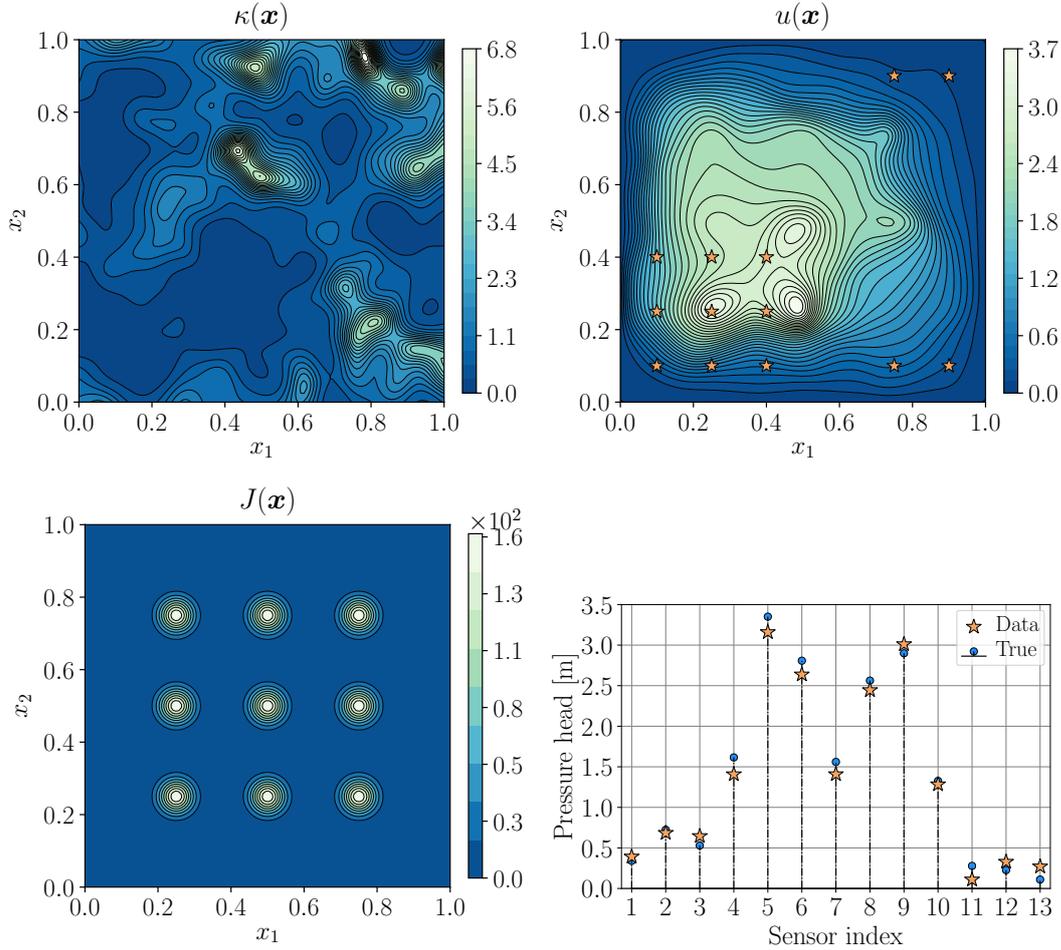


Figure 5.15: Groundwater flow problem: true hydraulic conductivity, true hydrostatic pressure with measurement locations, source term, measured and true (noise-free) hydraulic head.

In this example, we evaluate the posterior for different correlation lengths $\ell \in \{0.1, 0.2, 0.3\}$. Each correlation length defines a dimension prior as follows:

- for $\ell = 0.1$, we obtain $k_{\min} = 16$ and $k_{\max} = 520$; this yields $p = 6.307 \times 10^{-3}$.
- for $\ell = 0.2$, we obtain $k_{\min} = 5$ and $k_{\max} = 138$; this yields $p = 1.939 \times 10^{-2}$.
- for $\ell = 0.3$, we obtain $k_{\min} = 3$ and $k_{\max} = 65$; this yields $p = 2.939 \times 10^{-2}$.

We employ the same proposal scaling settings investigated in the previous example. However, in this case the dimension posterior differs significantly from the dimension prior. Therefore, additional to sampling k directly from the prior, the proposal \mathbf{Q} is used. The jump proposal matrix is constructed from a discrete triangular distribution with jump length $\rho = 0.25 \cdot k_{\max}$, which appears to be a good choice for both MCMC algorithms (cf., subsection 5.4.1).

Approximation of the posterior for the dimension and the random field

The adaptive tBUS-SuS is used to estimate the posterior of the dimension and the random field. The results are shown for an average of $N_{\text{sim}} = 60$ independent simulation runs using $N = 1.5 \times 10^4$ samples per level. To compare the tBUS-SuS approximations, we compute the reference solution from model evidences estimated by within-model runs of adaptive BUS-SuS using $N = 5 \times 10^3$ samples per level and averaged over $N_{\text{sim}} = 90$ simulations. For this example, it is not feasible to compute the full reference solution for the posterior of the dimension by means of within-model simulation algorithms. Thus, we estimate the reference at 6 dimension snapshots for each correlation length: $k_{\text{snap}} \in \{40, 45, 50, 55, 60, 70\}$ for $\ell = 0.1$, $k_{\text{snap}} \in \{20, 30, 35, 40, 50, 60\}$ for $\ell = 0.2$, and $k_{\text{snap}} \in \{20, 25, 30, 32, 35, 40\}$ for $\ell = 0.3$. Since the reference solutions are given in terms of the model evidences $Z_{\tilde{\mathbf{y}}}(k_{\text{snap}})$, we transform them to model posteriors using (5.4). This requires the knowledge of the evidence of all model classes, which is cumbersome to compute in this case. Instead, we apply the normalization

$$\bar{\pi}_{\text{pos}}(k_{\text{snap}} | \tilde{\mathbf{y}}) = \frac{\bar{\pi}_{\text{pr}}(k_{\text{snap}}) Z_{\tilde{\mathbf{y}}}(k_{\text{snap}})}{\sum_{k \in k_{\text{snap}}} \bar{\pi}_{\text{pr}}(k) Z_{\tilde{\mathbf{y}}}(k)} \sum_{k \in k_{\text{snap}}} \hat{\pi}_{\text{pos}}(k | \tilde{\mathbf{y}}), \quad (5.34)$$

such that the sum of the reference $\bar{\pi}_{\text{pos}}(k_{\text{snap}} | \tilde{\mathbf{y}})$ match the sum of the tBUS-SuS-estimated model posteriors $\hat{\pi}_{\text{pos}}(k_{\text{snap}} | \tilde{\mathbf{y}})$ at the given snapshots. Using this approach, the reference solution is limited to the tBUS-SuS solution. Nevertheless, it allows us to check that the algorithm is following the correct shape of the dimension posterior.

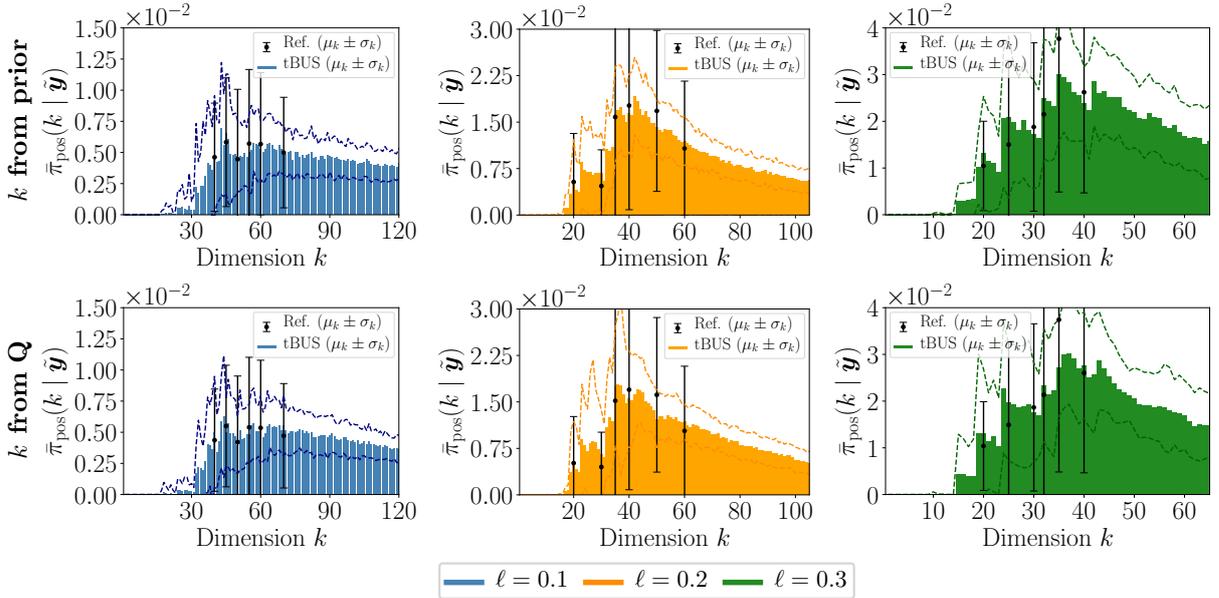


Figure 5.16: Diffusion example: model posterior using adaptive tBUS-SuS with **MwG sampler**, sampling k from the prior (1st row) and from proposal **Q** (2nd row).

The dimension posteriors estimated by adaptive tBUS-SuS with MwG are shown in Figure 5.16. We plot the mean and standard deviation bounds of the approximations. The solutions are computed when the dimension is sampled from the prior (1st row) and when it is sampled from the proposal **Q** (2nd row). Both alternatives yield comparable results. In general, we observe an increase in the variability around the MAP estimate. Note also that in this example, the dimension posteriors have several modes which can be related to the nonuniform distribution of the measurement locations together with the symmetry of the KL eigenfunctions. For instance, when using $\ell = 0.3$ there is a jump in the values of the probability mass after the 15-th dimension, every 5 dimensions until the MAP estimate. Furthermore, as an indicative of the algorithm

performance, tBUS-SuS requires on average $n_{lv} = 12$ intermediate levels and the proposal scaling β changes from 0.75 in the first level, to 0.09 in the last level (for the correlation length $\ell = 0.1$ and sampling from the prior).

The approximated dimension posteriors using adaptive tBUS-SuS with the step-wise sampler are shown in Figure 5.17. In this case, the step-wise sampler is more sensitive to the selection of the dimension proposal than MwG. We observed that the dimension prior is not a good proposal choice to sample the dimensions (the results are omitted). The main issue is that the resulting posterior samples are highly correlated since the values of the scaling β at the last level of the simulation are in the order of 10^{-4} . Therefore, instead of showing a comparison between the dimension proposal schemes, we employ the proposal matrix \mathbf{Q} with two different settings: using N samples per level, and using $2N$ samples per level. In both MCMC algorithms, increasing the number of samples per level considerably improves the variability of the estimates. Particularly, using $2N$ samples per level in the step-wise sampler yields comparable results to those of MwG, since we are evaluating the PDE model approximately the same number of times. The resulting dimension posteriors are able to capture the trend of the reference solutions.

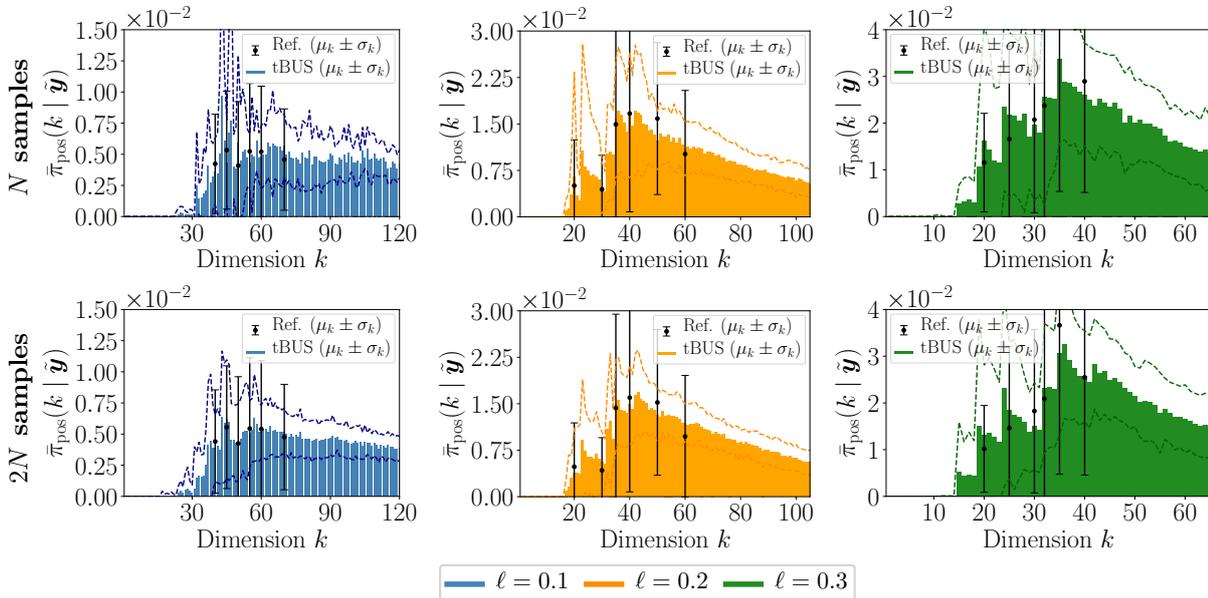


Figure 5.17: Diffusion example: model posterior using adaptive tBUS-SuS with **step-wise sampler** and sampling k from proposal \mathbf{Q} . Using N and $2N$ samples per level.

Finally, we estimate the posterior hydraulic conductivity and head random fields. Figure 5.18 shows the model choice and model mixing solutions in terms of the posterior mean and standard deviation of the hydraulic conductivity. The model choice estimate is given by the truncation order that yields the maximum model posterior, in this case $k_{\text{best}} \in \{43, 43, 33\}$ for the correlation lengths $\{0.1, 0.2, 0.3\}$ (Figure 5.17, 2nd row). The model mixing estimate takes into account the whole dimension spectrum. Note that the values in the posterior mean are smaller than those of the assumed truth. In this case, most of the measurements are concentrated in the lower left corner of the aquifer. In this area, the values of the true hydraulic conductivity are very small, and they are influencing the posterior solution. Nevertheless, the statistics are revealing the locations of lower and higher permeability values. The field modeled with the smaller correlation length is able to represent the small fluctuations better than those with larger correlation lengths, for which the resulting random field realizations are smoother. In contrast to the hydraulic conductivity field, the differences in the model choice and mixing solutions for the hydraulic head random field are minimal (Figure 5.19). This quantity is computed by integrating the PDE model, an averaging operation that reduces the effect of the spatial variability.

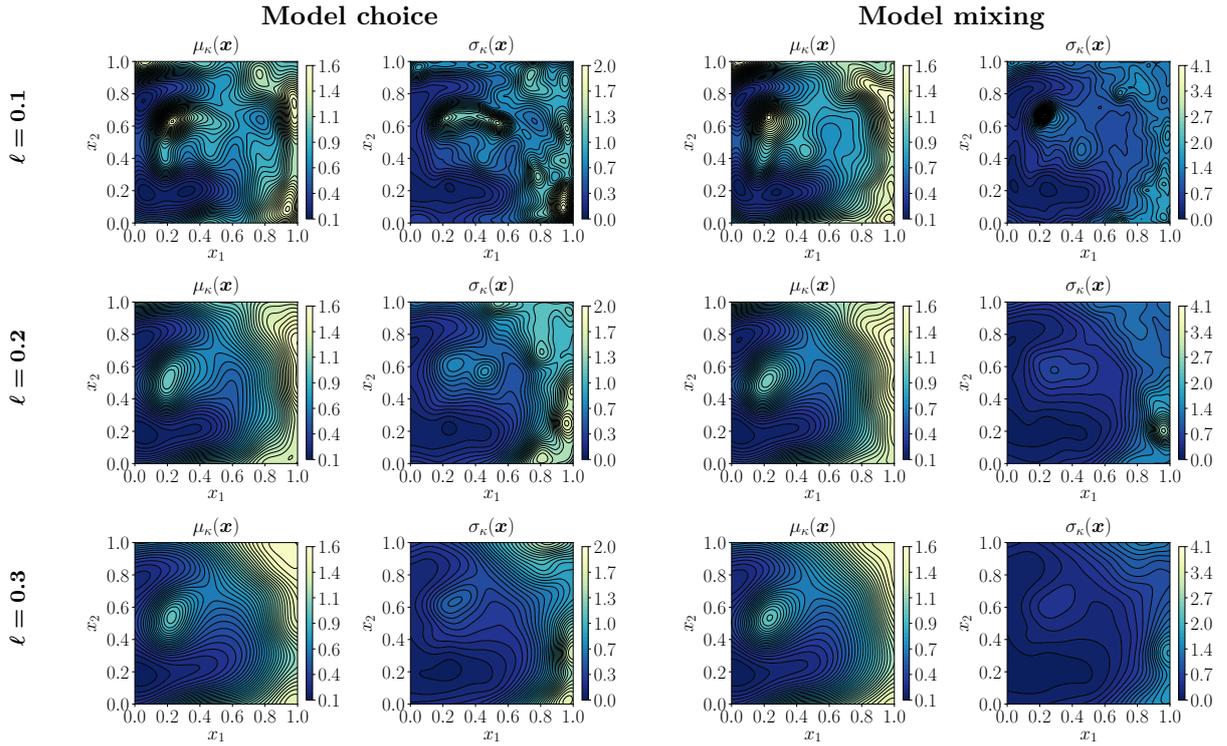


Figure 5.18: Diffusion example: posterior mean (1st and 3rd cols) and standard deviation (2nd and 4th cols) of the **hydraulic conductivity** random field using adaptive tBUS-SuS with the step-wise sampler for different prior correlation lengths (rows).

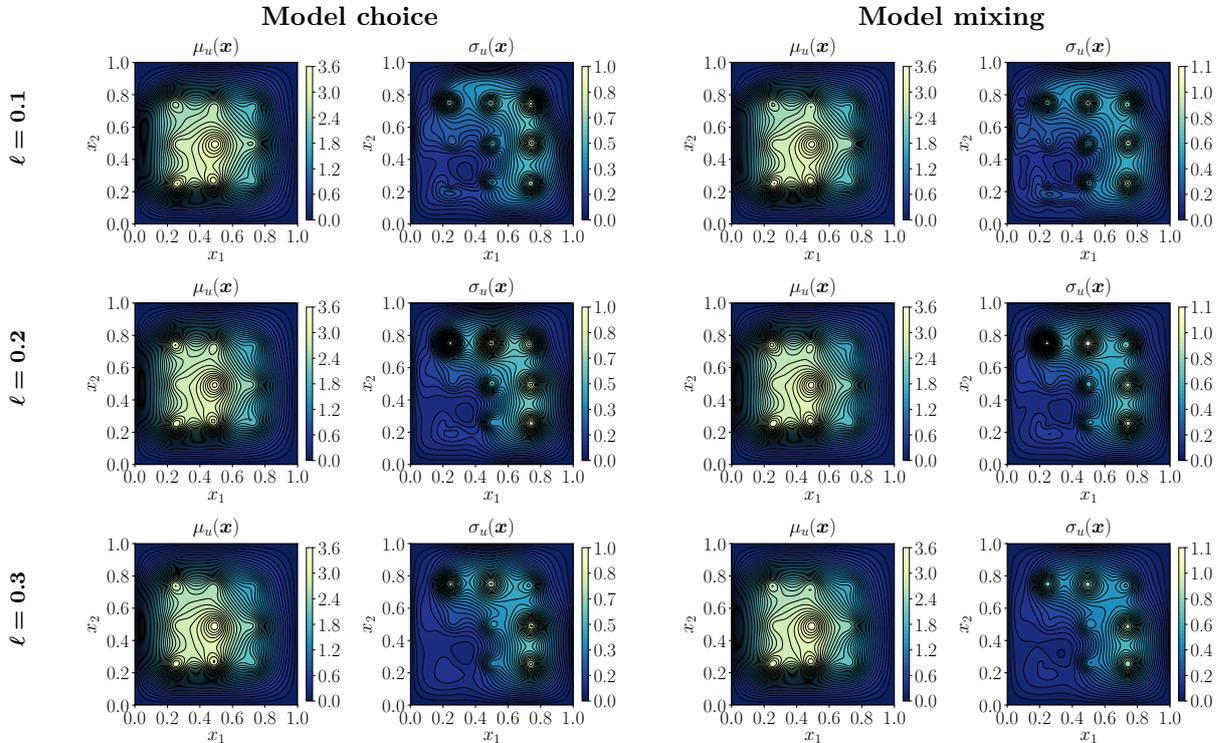


Figure 5.19: Diffusion example: posterior mean (1st and 3rd cols) and standard deviation (2nd and 4th cols) of the **hydrostatic pressure** random field using adaptive tBUS-SuS with the step-wise sampler for different prior correlation lengths (rows).

In this chapter, we have seen that the model evidence associated to the dimensions of the KL discretization does not reveal the Bayesian penalization that one sees for example in regression models, where increasing the polynomial order might cause over-fitting. As we comment in remark 4.8, increasing the KL terms contributes to a better representation of the posterior covariance, and the model evidence keeps increasing as one augments the KL truncation. This is the main reason for including a penalizing prior for the number of KL terms.

Due to this behavior, one could argue that performing model choice or mixing is not fundamentally necessary. Instead, one can fix the number of KL terms as k_{\max} to perform a standard within-model Bayesian inversion. However, one of the advantages of the methods developed in this chapter lies in the reduction of the computational cost. The value of k_{\max} can be in principle very large, whereas a smaller value (or a set of smaller values) could be sufficient for an appropriate posterior field representation. This is of course relevant in modeling scenarios where the PDE model is expensive to evaluate or that require gradient computations. In those situations, finding a solution in a parameter space with dimension k_{\max} could be potentially unfeasible.

Rare event simulation of random fields: dimension reduction

“... the determination of this maximum is quite definitely not routine when the number of variables is large. All this may be subsumed under the heading the curse of dimensionality. Since this is a curse which has hung over the head of the physicist and astronomer for many years, there is no need to feel discouraged about the possibility of obtaining significant results despite it”.

Richard Ernest Bellman, 1920-1984 [23, p.ix].

When the rare event of interest is expressed in terms of a computationally costly numerical model, finding the probability of failure *accurately* becomes a cumbersome task for the standard simulation methods discussed in Chapter 3. IS with the CE method is an efficient way to address this type of problems provided that a suitable parametric family of biasing densities is employed. Despite some existing distribution families are designed to sample efficiently in high dimensions (see, e.g., [244, 179]), their applicability within the CE method is limited to problems with dimension of $\mathcal{O}(10^2)$. Some extended works involve the construction of surrogate models [146], the use of kernel mixture distributions [37], or application of multi-fidelity methods [183] to increase the range of applications of the CE method. Although these algorithms can estimate small probabilities of failure, the studied applications still involve parameter spaces with low to moderate dimensions.

In this chapter, we develop a method that uses a different strategy. Rather than identifying sampling densities that are effective in high dimensions, we focus on finding the intrinsic low-dimensional structure of the rare event simulation problem. To this end, we exploit the connection between rare event simulation and Bayesian inversion shown in (3.2). The resulting posterior failure distribution coincides in particular with the optimal zero-variance distribution of IS (3.15). This allows one to adapt dimension reduction techniques from Bayesian inference to construct effective low-dimensional biasing distribution models within IS with the CE method.

We start by summarizing well-known dimension reduction techniques in Bayesian inverse problems. Then, we use the link between Bayesian inversion and rare event estimation to build the so-called failure-informed subspace, where efficient biasing densities for the CE method are defined. We illustrate the proposed method by means of high-dimensional numerical examples; two of them are related to random fields.

6.1. General comments on dimension reduction

Dimension reduction can be defined as the action of mapping the input parameters from the original space to a low-dimensional subspace. Such a subspace is usually defined by directions in the original space in which the model response changes the most. One can roughly group

dimension reduction strategies into three classes: (i) manifold embedding techniques, (ii) sensitivity analysis, and (iii) projection methods.

A classical method in manifold embedding is multidimensional scaling [46]. The idea is to use a measure of dissimilarity between each pair of elements in the parameter space. Thereafter, one seeks a transformation of the dissimilarities onto a low-dimensional Euclidean space, such that the transformed pairwise dissimilarities become squared distances. The resulting low-dimensional parameters can be used to build response surfaces. Other manifold modeling methods include, locally linear embedding, Laplacian eigenmaps, spectral clustering, along with others (see, [46] for a review). Nonlinear transformations of the input parameter space can also be applied to perform dimension reduction. For example, in the context of rare event simulation, a polar coordinate transformation to express the reliability problem in a two-dimensional space is proposed in [119].

In global sensitivity analysis, the objective is to determine the input parameters that have the most influence on the model response. Classical methods employ high-dimensional model representation (i.e., a finite expansion for the multivariate function of interest) to perform variance decomposition of the model output [200, 174]. For instance, consider the model $y = f(\boldsymbol{\theta})$ (e.g., a likelihood function in Bayesian inference, or a LSF in rare event simulation), where the parameter vector has independent components $\boldsymbol{\theta} = [\theta_1, \dots, \theta_i, \dots, \theta_d]$. In variance-based sensitivity analysis, one wants to determine how much uncertainty in the output y can be attributed to each input θ_i . This provides a ranking of the inputs that can be expressed in terms of Sobol indices [213]. One can use this information to build a response surface that concentrates the approximation on the most important input parameters. Note that this type of dimension reduction acts along the coordinate system imposed by the input parameters.

Ideas on dimension reduction using projection methods probably trace back to principal component analysis [116], which utilizes an orthogonal operator to map a set of data points into a set of linearly uncorrelated variables (principal components). The construction of the projection requires the assembly of a covariance matrix of the data set and its eigenvalue decomposition. In the context of random fields, principal component analysis is equivalent to the KL expansion (2.29). A strong decay of the spectrum of the covariance operator of the field guarantees that the mean-squared error of the KL approximation is small for a suitable truncation k (see (2.32)). In Bayesian inference, the dimension reduction associated to the KL expansion typically relies only on the prior. This means that the low-dimensional structure induced by the computational model is not exploited. However, note that the modeling strategy used for the KL expansion in Chapter 5, considers both prior and model to identify the dimensions that yield the best representation of the posterior. In this sense, Bayesian model inference can also be considered a dimension reduction technique.

In this chapter, we focus on projection techniques that have the following properties: (i) they consider both, the prior and the potential low-dimensional structure induced by the mathematical model, to perform the dimension reduction (in contrast to the standard KL expansion); and (ii) they exploit the fact that the underlying mathematical model oftentimes varies most, along directions of the parameter space that are not aligned with the coordinate system (in contrast to sensitivity analysis). In the following, we present a short summary of methods used for dimension reduction in nonlinear Bayesian inverse problems. This is because they can be adapted to the rare event context via the connection (3.2), which (again) expresses a rare event simulation problem as a Bayesian inversion task.

6.1.1. Likelihood-informed subspace

A nonlinear extension of the optimal low-rank approximations derived in [218] for linear Bayesian inverse problems is proposed in [57]. The idea is to identify a subspace of the parameter space that is likelihood-informed, by describing the relative influences of the prior and likelihood over

the support of the posterior distribution.

The posterior distribution is approximated as the product of a low-dimensional posterior (on the likelihood-informed directions) and the marginalization of the prior onto the complementary directions. The discovery of the new coordinate system requires gradient information.

In contrast to [218], the Hessian of the negative log-likelihood is no longer constant since the forward model is nonlinear. A linearization of the forward operator $\mathcal{G}(\boldsymbol{\theta})$ is employed to build a local Gauss-Newton approximation of the Hessian, which is averaged over the posterior support [49, 57]:

$$\mathbf{H}(\boldsymbol{\theta}) = \mathbb{E}_{\pi_{\text{pos}}} \left[(\nabla \mathcal{G}(\boldsymbol{\theta}))^\top \boldsymbol{\Sigma}_{\text{obs}}^{-1} (\nabla \mathcal{G}(\boldsymbol{\theta})) \right]. \quad (6.1)$$

A MC estimator is used in practice to compute this expectation. The likelihood-informed directions are identified from the generalized eigenvalue problem associated to the matrix pair $(\mathbf{H}(\boldsymbol{\theta}), \boldsymbol{\Sigma}_{\text{pr}}^{-1})$, that is:

$$\mathbf{L}^\top \mathbf{H}(\boldsymbol{\theta}) \mathbf{L} \boldsymbol{\phi}_i = \lambda_i \boldsymbol{\phi}_i, \quad (6.2)$$

where $\mathbf{L}\mathbf{L}^\top = \boldsymbol{\Sigma}_{\text{pr}}$ is a symmetric decomposition of the prior covariance matrix and $\{\lambda_i, \boldsymbol{\phi}_i\}_{i=1}^d$ are the generalized eigenpairs. The first r eigenvectors (grouped into a matrix $\boldsymbol{\Phi}_r \in \mathbb{R}^{d \times r}$) associated to the first r leading eigenvalues, are used to build a projector onto the *likelihood-informed subspace* (LIS). This operator can be written as $\mathbf{P}_r = \mathbf{L}\boldsymbol{\Phi}_r\boldsymbol{\Phi}_r^\top\mathbf{L}^{-1}$. In this setting, the subspace $\boldsymbol{\Theta}_r = \text{Im}(\mathbf{P}_r)$ defines the LIS and $\boldsymbol{\Theta}_\perp = \text{Im}(\mathbf{I}_d - \mathbf{P}_r)$ defines the complementary subspace.

Cui et al. [57] also discuss a local LIS approach in which the LIS basis $\boldsymbol{\Phi}_r$ is used instead of the projector. The local LIS has similarities to the coordinate transformation applied by the active subspace method, which is described next.

6.1.2. Active subspace

In the active subspace (AS) method applied to Bayesian inverse problems [55], one seeks a low-rank approximation of the potential or data-misfit function $f(\boldsymbol{\theta}) = \Phi(\boldsymbol{\theta}; \tilde{\mathbf{y}}) = -\ln L(\boldsymbol{\theta}; \tilde{\mathbf{y}})$ (see, e.g., (4.3)). If f is such that the products of its partial derivatives are square integrable, then one can define the matrix [53]:

$$\mathbf{H} = \mathbb{E}_{\pi_{\text{pr}}} \left[\nabla f(\boldsymbol{\theta}) \nabla f(\boldsymbol{\theta})^\top \right], \quad (6.3)$$

which can be seen as the covariance of the gradient vector. Notice that in LIS the expectation is with respect to the posterior. In practice, \mathbf{H} is approximated using MC as

$$\mathbf{H} \approx \hat{\mathbf{H}} = \sum_{i=1}^N \nabla f(\boldsymbol{\theta}_i) \nabla f(\boldsymbol{\theta}_i)^\top, \quad (6.4)$$

where $\{\boldsymbol{\theta}_i\}_{i=1}^N \sim \pi_{\text{pr}}$. A heuristic relation used to define the number of samples required to estimate $\hat{\mathbf{H}}$ is $N = \alpha \cdot r \ln(d)$, where $r \leq d$ is the rank and $\alpha \in [2, 10]$ is the so-called oversampling factor [54].

Note that the matrix \mathbf{H} admits the spectral decomposition, $\mathbf{H} = \boldsymbol{\Phi}\boldsymbol{\Lambda}\boldsymbol{\Phi}^\top$, where $\boldsymbol{\Phi} \in \mathbb{R}^{d \times d}$ contains the eigenvectors and $\boldsymbol{\Lambda} \in \mathbb{R}^{d \times d}$ the eigenvalues of \mathbf{H} . The eigenvectors define a basis for the input parameter space $\boldsymbol{\Theta}$. The idea is to separate the components of the coordinate system induced by $\boldsymbol{\Phi}$, such that one groups the directions carrying most of the variation (larger eigenvalues) and those corresponding to almost no variation (smaller eigenvalues):

$$\boldsymbol{\Lambda} = \begin{bmatrix} \boldsymbol{\Lambda}_r & \\ & \boldsymbol{\Lambda}_\perp \end{bmatrix} \quad \text{and} \quad \boldsymbol{\Phi} = [\boldsymbol{\Phi}_r, \boldsymbol{\Phi}_\perp], \quad (6.5)$$

where $\mathbf{\Lambda}_r = \text{diag}(\lambda_1, \dots, \lambda_r)$ and $\mathbf{\Phi}_r \in \mathbb{R}^{d \times r}$. One can employ the resulting bases, $\mathbf{\Phi}_r$ and $\mathbf{\Phi}_\perp$, to project the parameter vector into two subspaces; the transformation reads as follows:

$$\tilde{\boldsymbol{\theta}}_r = \mathbf{\Phi}_r^\top \boldsymbol{\theta} \quad \text{and} \quad \tilde{\boldsymbol{\theta}}_\perp = \mathbf{\Phi}_\perp^\top \boldsymbol{\theta}, \quad (6.6)$$

where $\tilde{\boldsymbol{\theta}}_r \in \mathbb{R}^r$ and $\tilde{\boldsymbol{\theta}}_\perp \in \mathbb{R}^{d-r}$. Based on the previous decomposition, AS approximates the d -dimensional function f by a function of fewer variables r (on the ‘‘active’’ coordinates). For a fixed $\tilde{\boldsymbol{\theta}}_r$, the optimal (in the mean-squared sense) approximation of f given $\tilde{\boldsymbol{\theta}}_r$ is the conditional expectation [54]

$$f(\boldsymbol{\theta}) \approx h(\mathbf{\Phi}_r^\top \boldsymbol{\theta}) = h(\tilde{\boldsymbol{\theta}}_r) = \int_{\mathbb{R}^{d-r}} f(\mathbf{\Phi}_r \tilde{\boldsymbol{\theta}}_r + \mathbf{\Phi}_\perp \tilde{\boldsymbol{\theta}}_\perp) \pi(\tilde{\boldsymbol{\theta}}_\perp | \tilde{\boldsymbol{\theta}}_r) d\tilde{\boldsymbol{\theta}}_\perp = \mathbb{E}_{\pi_{\text{pr}}} [f(\boldsymbol{\theta}) | \mathbf{\Phi}_r]; \quad (6.7)$$

similar to the computation of the matrix \mathbf{H} , a MC estimator is also used to compute this expectation.

We remark that the authors in [229] found a theoretical inconsistency in the AS method. They argue that there exist unbounded constants in the probabilistic Poincaré inequalities that are used to derive the methodology. This affects the application of AS to log-concave distributions, which includes the exponential family. They propose a framework that enables the derivation of generalized estimates in an attempt to address the issue.

6.1.3. Certified dimension reduction

Certified dimension reduction (CDR) [253] is a generalization of the previous approaches to nonlinear and non-Gaussian Bayesian inverse problems. In principle, the framework can be applied to any type of measurable likelihood function, as long as the gradient of its logarithm is square integrable. In contrast to the previous methods, the posterior approximation in CDR is derived with certified error bounds in the sense of the KLD. As a result, the approximation can be controlled by some user-defined threshold bounding the KLD from the exact to the approximated posterior. In our framework, we employ CDR as the dimension reduction technique. Thus, we skip the details on CDR here and leave them for section 6.2.

We conclude the short presentation in this section by summarizing in Table 6.1 the main ingredients and differences between the discussed dimension reduction methods.

Table 6.1.: Summary of some approaches for dimension reduction in Bayesian inverse problems.

	Posterior approximation	Matrix \mathbf{H}
LIS	$\pi_{\text{pos}}(\boldsymbol{\theta}) \propto \text{L}(\mathbf{P}_r \boldsymbol{\theta} + \mathbf{P}_\perp \boldsymbol{\mu}_{\text{pr}}) \pi_{\text{pr}}(\boldsymbol{\theta})$	$\int_{\boldsymbol{\theta}} \left(\nabla \mathcal{G}(\boldsymbol{\theta})^\top \boldsymbol{\Sigma}_{\text{obs}}^{-1} \nabla \mathcal{G}(\boldsymbol{\theta}) \right) \pi_{\text{pos}}(\boldsymbol{\theta}) d\boldsymbol{\theta}$
AS	$\pi_{\text{pos}}(\boldsymbol{\theta}) \propto \exp(\mathbb{E}_{\pi_{\text{pr}}}[\ln \text{L}(\boldsymbol{\theta}) \mathbf{P}_r]) \pi_{\text{pr}}(\boldsymbol{\theta})$	$\int_{\boldsymbol{\theta}} \left(\nabla \ln \text{L}(\boldsymbol{\theta}) \nabla \ln \text{L}(\boldsymbol{\theta})^\top \right) \pi_{\text{pr}}(\boldsymbol{\theta}) d\boldsymbol{\theta}$
CDR	$\pi_{\text{pos}}(\boldsymbol{\theta}) \propto \mathbb{E}_{\pi_{\text{pr}}}[\text{L}(\boldsymbol{\theta}) \mathbf{P}_r] \pi_{\text{pr}}(\boldsymbol{\theta})$	$\int_{\boldsymbol{\theta}} \left(\nabla \ln \text{L}(\boldsymbol{\theta}) \nabla \ln \text{L}(\boldsymbol{\theta})^\top \right) \pi_{\text{pos}}(\boldsymbol{\theta}) d\boldsymbol{\theta}$

6.2. Construction of the failure-informed subspace

We build on the ideas of [253] to construct low-dimensional approximations of the posterior-failure distribution (3.2). By analogy to the Bayesian inversion setting, we refer to the resulting low-dimensional failure subspace as the *failure-informed subspace*.

6.2.1. Rare event simulation as a Bayesian inversion

Recall from Chapter 3 that the probability of failure $p_{\mathcal{F}}$ and indicator function $\mathbb{1}_{\mathcal{F}}(\boldsymbol{\theta})$ are analogous to the model evidence and likelihood function in Bayesian inference (compare (3.2)

with (4.2)). The resulting posterior density is equivalent to the prior (nominal) density censored on the failure set \mathcal{F} . As a result, the rare event simulation setting can be interpreted as a nonlinear Bayesian inverse problem.

The indicator function $\mathbf{1}_{\mathcal{F}}(\boldsymbol{\theta})$ is upper semi-continuous and does not provide sufficient regularity to construct the certified approximation ($\nabla \ln \mathbf{1}_{\mathcal{F}}(\boldsymbol{\theta})$ is not square-integrable). Hence, we employ one of the smooth representations $f(\boldsymbol{\theta}; s) \approx \mathbf{1}_{\mathcal{F}}(\boldsymbol{\theta})$ in (3.3). The posterior-failure density $\pi_{\mathcal{F}} \propto \mathbf{1}_{\mathcal{F}}(\boldsymbol{\theta})\pi_{\text{pr}}(\boldsymbol{\theta})$ is retrieved as the smoothing parameter goes to zero; thus, we write $\pi_{\mathcal{F}}$ in terms of the smooth approximation:

$$\pi_{\mathcal{F}}(\boldsymbol{\theta}) = \frac{1}{p_{\mathcal{F}}} f(\boldsymbol{\theta}; s)\pi_{\text{pr}}(\boldsymbol{\theta}) \quad \text{as } s \rightarrow 0. \quad (6.8)$$

Our objective is to find a low-dimensional approximation of (6.8), denoted $\pi_{\mathcal{F}}^{(r)}$ such that

$$\pi_{\mathcal{F}}^{(r)}(\boldsymbol{\theta}) \propto (h \circ \mathbf{P}_r)(\boldsymbol{\theta})\pi_{\text{pr}}(\boldsymbol{\theta}) \approx \pi_{\mathcal{F}}(\boldsymbol{\theta}), \quad (6.9)$$

where $h : \mathbb{R}^d \rightarrow \mathbb{R}_{>0}$ is a *profile function* and the *projector* $\mathbf{P}_r \in \mathbb{R}^{d \times d}$ is a linear transformation such that $\mathbf{P}_r^2 = \mathbf{P}_r$, but not necessarily $\mathbf{P}_r^T = \mathbf{P}_r$; we can also define the *complementary projector* to \mathbf{P}_r as the matrix $\mathbf{P}_{\perp} = \mathbf{I}_d - \mathbf{P}_r$, which satisfies $\text{Im}(\mathbf{P}_{\perp}) = \text{Ker}(\mathbf{P}_r)$.

The profile function in (6.9) depends only on $\mathbf{P}_r \boldsymbol{\theta} = \boldsymbol{\theta}_r$, which is defined on the *failure-informed subspace* (FIS) $\boldsymbol{\Theta}_r = \text{Im}(\mathbf{P}_r)$, and it is essentially constant along the *complementary subspace* (CS) $\boldsymbol{\Theta}_{\perp} = \text{Ker}(\mathbf{P}_r)$. As a result, if $r \ll d$, the goal of the approximation (6.9) is to replace the high-dimensional smooth indicator by a function of fewer variables.

Following the CDR framework [253], we now describe how the optimal profile function and projector are derived.

6.2.2. Optimal profile function

For a fixed projector \mathbf{P}_r , $f(\boldsymbol{\theta}; s)$ can be approximated by its average over all values of $\boldsymbol{\theta}$ that map to $\boldsymbol{\theta}_r = \mathbf{P}_r \boldsymbol{\theta}$ [195]. This is the *conditional expectation* of the smooth indicator given the projector under the prior distribution [253]

$$\mathbb{E}_{\pi_{\text{pr}}}[f(\boldsymbol{\theta}; s) \mid \mathbf{P}_r \boldsymbol{\theta}] = \int_{\tilde{\boldsymbol{\Theta}}_{\perp}} f(\mathbf{P}_r \boldsymbol{\theta} + \boldsymbol{\Phi}_{\perp} \boldsymbol{\xi}_{\perp}; s) \underbrace{\frac{\pi_{\text{pr}}(\mathbf{P}_r \boldsymbol{\theta} + \boldsymbol{\Phi}_{\perp} \boldsymbol{\xi}_{\perp})}{\int_{\tilde{\boldsymbol{\Theta}}_{\perp}} \pi_{\text{pr}}(\mathbf{P}_r \boldsymbol{\theta} + \boldsymbol{\Phi}_{\perp} \boldsymbol{\xi}'_{\perp}) d\boldsymbol{\xi}'_{\perp}}}_{\pi(\boldsymbol{\xi}_{\perp} \mid \mathbf{P}_r \boldsymbol{\theta})} d\boldsymbol{\xi}_{\perp}, \quad (6.10)$$

where $\tilde{\boldsymbol{\Theta}}_{\perp} \subseteq \mathbb{R}^{d-r}$, $\boldsymbol{\xi}_{\perp} \in \tilde{\boldsymbol{\Theta}}_{\perp}$, and the columns of $\boldsymbol{\Phi}_{\perp} \in \mathbb{R}^{d \times d-r}$ form a basis for $\text{Ker}(\mathbf{P}_r)$.

The optimal profile function h^* is obtained by minimizing the KLD between the exact $\pi_{\mathcal{F}}$ and approximated $\pi_{\mathcal{F}}^{(r)}$ densities. We can use the conditional expectation (6.10) to define the density

$$\pi_{\mathcal{F}}^{(r)*} \propto \underbrace{\mathbb{E}_{\pi_{\text{pr}}}[f(\boldsymbol{\theta}; s) \mid \mathbf{P}_r \boldsymbol{\theta}]}_{(h^* \circ \mathbf{P}_r)(\boldsymbol{\theta})} \pi_{\text{pr}}(\boldsymbol{\theta}), \quad (6.11)$$

for which the relation, $D_{\text{KL}}(\pi_{\mathcal{F}} \parallel \pi_{\mathcal{F}}^{(r)}) - D_{\text{KL}}(\pi_{\mathcal{F}} \parallel \pi_{\mathcal{F}}^{(r)*}) = D_{\text{KL}}(\pi_{\mathcal{F}}^{(r)*} \parallel \pi_{\mathcal{F}}^{(r)}) \geq 0$ holds (see [253]). In particular, $D_{\text{KL}}(\pi_{\mathcal{F}} \parallel \pi_{\mathcal{F}}^{(r)}) \geq D_{\text{KL}}(\pi_{\mathcal{F}} \parallel \pi_{\mathcal{F}}^{(r)*})$, and hence a profile function of the form

$$(h^* \circ \mathbf{P}_r)(\boldsymbol{\theta}) = \mathbb{E}_{\pi_{\text{pr}}}[f(\boldsymbol{\theta}; s) \mid \mathbf{P}_r \boldsymbol{\theta}] \quad (6.12)$$

is a minimizer of $D_{\text{KL}}(\pi_{\mathcal{F}} \parallel \pi_{\mathcal{F}}^{(r)})$.

6.2.3. Optimal projector

Consider a Gaussian prior density $\pi_{\text{pr}}(\boldsymbol{\theta}) \propto \exp(-V(\boldsymbol{\theta}))$, where the function $V(\boldsymbol{\theta}) = \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\mu}_{\text{pr}})^\top \boldsymbol{\Sigma}_{\text{pr}}^{-1}(\boldsymbol{\theta} - \boldsymbol{\mu}_{\text{pr}})$ is twice differentiable and satisfies $\nabla^2 V(\boldsymbol{\theta}) = \boldsymbol{\Sigma}_{\text{pr}}^{-1} \succeq c \cdot \mathbf{I}_d$, with $c > 0$. Theorem 1 in [253] shows that, for any continuously differentiable function $h : \mathbb{R}^d \rightarrow \mathbb{R}$, such that $\mathbb{E}_{\pi_{\text{pr}}}[\|\nabla h(\boldsymbol{\theta})\|_{\boldsymbol{\Sigma}_{\text{pr}}}^2] < \infty$, the following *subspace logarithmic Sobolev inequality* holds

$$\mathbb{E}_{\pi_{\text{pr}}} \left[h^2(\boldsymbol{\theta}) \ln \left(\frac{h^2(\boldsymbol{\theta})}{\mathbb{E}_{\pi_{\text{pr}}}[h^2(\boldsymbol{\theta}) \mid \mathbf{P}_r \boldsymbol{\theta}]} \right) \right] \leq 2 \mathbb{E}_{\pi_{\text{pr}}} \left[\left\| (\mathbf{I}_d - \mathbf{P}_r^\top) \nabla h(\boldsymbol{\theta}) \right\|_{\boldsymbol{\Sigma}_{\text{pr}}}^2 \right], \quad (6.13)$$

with the notation $\|\nabla h(\boldsymbol{\theta})\|_{\boldsymbol{\Sigma}_{\text{pr}}}^2 = \nabla h(\boldsymbol{\theta})^\top \boldsymbol{\Sigma}_{\text{pr}} \nabla h(\boldsymbol{\theta})$. The inequality (6.13) allows one to bound $\text{D}_{\text{KL}}(\pi_{\mathcal{F}} \parallel \pi_{\mathcal{F}}^{(r)\star})$, which in turn provides a way to characterize the optimal projector in the approximation $\pi_{\mathcal{F}}^{(r)\star}$ written in (6.11). Roughly speaking, the original logarithmic Sobolev inequality states that the entropy of h^2 is bounded, up to scaling, by its Fisher information [34].

Let $h^2(\boldsymbol{\theta})$ be equal to the normalized smooth indicator function $f(\boldsymbol{\theta}; s)/p$, such that $\nabla h(\boldsymbol{\theta}) = \frac{1}{2}(f(\boldsymbol{\theta}; s)/p)^{1/2} \nabla \ln f(\boldsymbol{\theta}; s)$, then (6.13) can be written as

$$\mathbb{E}_{\pi_{\text{pr}}} \left[\frac{f(\boldsymbol{\theta}; s)}{p} \ln \left(\frac{f(\boldsymbol{\theta}; s)/p}{\mathbb{E}_{\pi_{\text{pr}}}[f(\boldsymbol{\theta}; s) \mid \mathbf{P}_r \boldsymbol{\theta}] / p_{\mathcal{F}}} \right) \right] \leq \frac{1}{2} \mathbb{E}_{\pi_{\text{pr}}} \left[\frac{f(\boldsymbol{\theta}; s)}{p} \left\| (\mathbf{I}_d - \mathbf{P}_r^\top) \nabla \ln f(\boldsymbol{\theta}; s) \right\|_{\boldsymbol{\Sigma}_{\text{pr}}}^2 \right]. \quad (6.14)$$

The left-hand side in (6.14) is equal to $\text{D}_{\text{KL}}(\pi_{\mathcal{F}} \parallel \pi_{\mathcal{F}}^{(r)\star})$; hence we can obtain the bound

$$\begin{aligned} \text{D}_{\text{KL}}(\pi_{\mathcal{F}} \parallel \pi_{\mathcal{F}}^{(r)\star}) &\leq \frac{1}{2} \mathbb{E}_{\pi_{\mathcal{F}}} \left[\left\| (\mathbf{I}_d - \mathbf{P}_r^\top) \nabla \ln f(\boldsymbol{\theta}; s) \right\|_{\boldsymbol{\Sigma}_{\text{pr}}}^2 \right] \\ &\leq \frac{1}{2} \mathbb{E}_{\pi_{\mathcal{F}}} \left[\text{tr} \left(\boldsymbol{\Sigma}_{\text{pr}} (\mathbf{I}_d - \mathbf{P}_r^\top) \nabla \ln f(\boldsymbol{\theta}; s) \nabla \ln f(\boldsymbol{\theta}; s)^\top (\mathbf{I}_d - \mathbf{P}_r) \right) \right] \\ &\leq \frac{1}{2} \text{tr} \left(\boldsymbol{\Sigma}_{\text{pr}} (\mathbf{I}_d - \mathbf{P}_r^\top) \mathbf{H} (\mathbf{I}_d - \mathbf{P}_r) \right) =: \frac{1}{2} \mathcal{R}(\mathbf{P}_r, \mathbf{H}), \end{aligned} \quad (6.15)$$

where $\mathcal{R}(\mathbf{P}_r, \mathbf{H})$ is the mean-squared error of the approximation of $\nabla \ln f(\boldsymbol{\theta}; s)$ by $\mathbf{P}_r^\top \nabla \ln f(\boldsymbol{\theta}; s)$ (with the parameter $\boldsymbol{\theta} \sim \pi_{\mathcal{F}}$), and $\mathbf{H} \in \mathbb{R}^{d \times d}$ is the symmetric and positive-definite second moment matrix of the gradient of the log-smooth indicator function [253]

$$\mathbf{H} = \int_{\Theta} \nabla \ln f(\boldsymbol{\theta}; s) \nabla \ln f(\boldsymbol{\theta}; s)^\top \pi_{\mathcal{F}}(\boldsymbol{\theta}) d\boldsymbol{\theta} = \mathbb{E}_{\pi_{\mathcal{F}}} \left[\nabla \ln f(\boldsymbol{\theta}; s) \nabla \ln f(\boldsymbol{\theta}; s)^\top \right]. \quad (6.16)$$

As in the LIS and AS methods, analytical evaluation of this equation is typically unfeasible and MC estimators are used in practice. Note from (6.15) that the mean-squared error is quadratic in \mathbf{P}_r . Therefore, we can minimize the upper bound in (6.15) over the collection of r -rank projectors to find an optimal projector:

$$\mathbf{P}_r = \arg \min_{\mathbf{P}'_r \in \mathbb{R}^{d \times d}} \mathcal{R}(\mathbf{P}'_r, \mathbf{H}). \quad (6.17)$$

As shown by Proposition 2 in [253] (see also Proposition 2.6 in [252]), a solution to (6.17) is

$$\mathbf{P}_r = \left(\sum_{i=1}^r \phi_i \phi_i^\top \right) \boldsymbol{\Sigma}_{\text{pr}}^{-1} \quad \text{and} \quad \min(\mathcal{R}(\mathbf{P}_r, \mathbf{H})) = \sum_{i=r+1}^d \lambda_i, \quad (6.18)$$

where the eigenpairs (λ_i, ϕ_i) correspond to the solution of the generalized eigenvalue problem $\mathbf{H} \phi_i = \lambda_i \boldsymbol{\Sigma}_{\text{pr}}^{-1} \phi_i$, $i = 1, \dots, d$. Hence, if π_{pr} satisfies the inequality (6.13) and \mathbf{P}_r is defined as

in (6.18), the approximation $\pi_{\mathcal{F}}^{(r)\star}$ in (6.11) can be controlled by a desired tolerance $\varepsilon \geq 0$ as

$$D_{\text{KL}}\left(\pi_{\mathcal{F}} \parallel \pi_{\mathcal{F}}^{(r)\star}\right) \leq \frac{1}{2} \sum_{i=r+1}^d \lambda_i \leq \varepsilon. \quad (6.19)$$

In this case, bounding $D_{\text{KL}}(\pi_{\mathcal{F}} \parallel \pi_{\mathcal{F}}^{(r)\star})$ amounts to select the rank r as the smallest integer, such that the left-hand side of (6.19) is below a prescribed ε , i.e.,

$$r = \min \left(r' : \frac{1}{2} \sum_{i=r'+1}^d \lambda_i \leq \varepsilon \right). \quad (6.20)$$

Note that along with the prior, the second-moment matrix of the log-smooth indicator function \mathbf{H} reveals the effective low dimension of the posterior-failure distribution. A sharp decay in the generalized spectrum of $(\mathbf{H}, \Sigma_{\text{pr}}^{-1})$ guarantees the existence of a low-rank approximation $r \ll d$.

Remark 6.1. The computational evaluation of the “optimal” conditional expectation (6.10) is demanding since the $\pi(\boldsymbol{\xi}_{\perp} \mid \mathbf{P}_r \boldsymbol{\theta})$ depends also on the parameter $\boldsymbol{\theta}$. Therefore, following [253], we employ a “practical” conditional expectation

$$\mathbb{E}_{\pi_{\text{pr}}}[f(\boldsymbol{\theta}; s) \mid \mathbf{P}_r \boldsymbol{\theta}] \approx \int_{\Theta} f(\mathbf{P}_r \boldsymbol{\theta} + \mathbf{P}_{\perp} \boldsymbol{\xi}; s) \pi_{\text{pr}}(\boldsymbol{\xi}) \, d\boldsymbol{\xi}, \quad (6.21)$$

where $\boldsymbol{\xi}$ is a random variable distributed according to π_{pr} . Proposition 6 in [253] shows that if π_{pr} is Gaussian, the KL divergence from the exact to the approximation constructed with (6.21) instead of (6.10) can still be controlled by the mean-squared error as in (6.19).

6.3. Failure-informed cross-entropy-based importance sampling

The optimal IS biasing density $\pi_{\text{bias}}^{\star}$ is equivalent to the posterior-failure density $\pi_{\mathcal{F}}$. Hence, we can employ the theory of section 6.2 to derive low-dimensional biasing distributions that extend the application of the iCE method to high-dimensional parameter spaces that have an intrinsic low-dimensional structure.

Recall that the goal is to represent (6.8), written now as the optimal biasing density:

$$\pi_{\mathcal{F}}(\boldsymbol{\theta}) = \pi_{\text{bias}}^{\star}(\boldsymbol{\theta}) = \frac{1}{p_{\mathcal{F}}} f(\boldsymbol{\theta}; s) \pi_{\text{pr}}(\boldsymbol{\theta}) \quad \text{as } s \rightarrow 0, \quad (6.22)$$

on the FIS using the certified approximation (6.11).

6.3.1. Formulation

We assume that the prior $\pi_{\text{pr}}(\boldsymbol{\theta})$ is standard Gaussian, or that at least an isoprobabilistic transformation can be applied to define such prior. The solution of the generalized eigenvalue problem $(\mathbf{H}, \Sigma_{\text{pr}}^{-1})$ required for the construction of \mathbf{P}_r in (6.18), reduces to a standard eigenvalue problem for \mathbf{H} . As a result, the projector can be written as $\mathbf{P}_r = \boldsymbol{\Phi}_r \boldsymbol{\Phi}_r^{\top}$, where $\boldsymbol{\Phi}_r \in \mathbb{R}^{d \times r}$ contains the first r eigenvectors that define a basis for the FIS. Similarly, the complementary projector can be obtained as $\mathbf{P}_{\perp} = \boldsymbol{\Phi}_{\perp} \boldsymbol{\Phi}_{\perp}^{\top}$, where $\boldsymbol{\Phi}_{\perp} \in \mathbb{R}^{d \times d-r}$ is defined by the remaining eigenvectors, generating a basis for the CS.

The projector \mathbf{P}_r is orthogonal with respect to the prior precision matrix $\Sigma_{\text{pr}}^{-1} = \mathbf{I}_d$, and thus, it satisfies $\boldsymbol{\theta}^{\top} \Sigma_{\text{pr}}^{-1} \boldsymbol{\theta} = \|\boldsymbol{\theta}\|_{\Sigma_{\text{pr}}^{-1}}^2 = \|\mathbf{P}_r \boldsymbol{\theta}\|_{\Sigma_{\text{pr}}^{-1}}^2 + \|\mathbf{P}_{\perp} \boldsymbol{\theta}\|_{\Sigma_{\text{pr}}^{-1}}^2$, for all $\boldsymbol{\theta} \in \Theta$. This induces a decomposition of the parameter space as the direct sum $\Theta = \Theta_r \oplus \Theta_{\perp}$, where any element of Θ

can be uniquely represented as

$$\boldsymbol{\theta} = \boldsymbol{\theta}_r + \boldsymbol{\theta}_\perp, \quad (6.23)$$

where $\boldsymbol{\theta}_r = \mathbf{P}_r \boldsymbol{\theta} \in \Theta_r$ and $\boldsymbol{\theta}_\perp = \mathbf{P}_\perp \boldsymbol{\theta} \in \Theta_\perp$. Moreover, we can use the operators Φ_r and Φ_\perp to map the parameters onto a local FIS $\tilde{\Theta}_r \subseteq \mathbb{R}^r$ and local CS $\tilde{\Theta}_\perp \subseteq \mathbb{R}^{d-r}$, respectively. In this case, the parameter vector can be written as

$$\boldsymbol{\theta} = \Phi_r \tilde{\boldsymbol{\theta}}_r + \Phi_\perp \tilde{\boldsymbol{\theta}}_\perp, \quad (6.24)$$

where $\tilde{\boldsymbol{\theta}} = [\tilde{\boldsymbol{\theta}}_r, \tilde{\boldsymbol{\theta}}_\perp]$ has components $\tilde{\boldsymbol{\theta}}_r = \Phi_r^\top \boldsymbol{\theta} \in \tilde{\Theta}_r$ and $\tilde{\boldsymbol{\theta}}_\perp = \Phi_\perp^\top \boldsymbol{\theta} \in \tilde{\Theta}_\perp$. Figure 6.1 illustrates the relation between the global and local FIS together with the action of the projectors and basis.

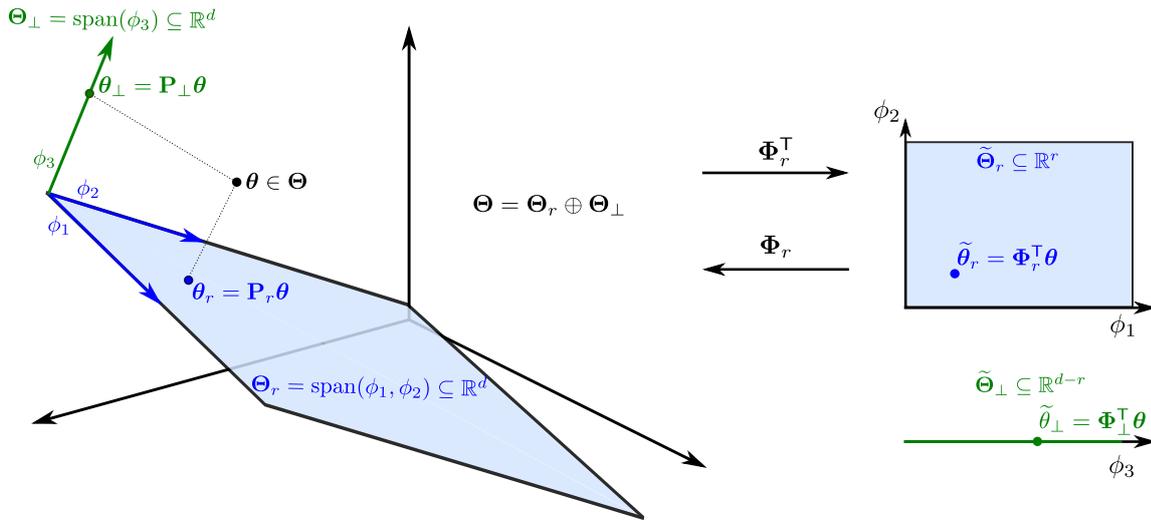


Figure 6.1: Schematic representation of a parameter $\boldsymbol{\theta} \in \mathbb{R}^3$ projected onto the FIS with $r = 2$ (in blue): global $\boldsymbol{\theta}_r \in \mathbb{R}^d$, local $\tilde{\boldsymbol{\theta}}_r \in \mathbb{R}^r$ (adapted from [57]).

From the discussion above, we see that the prior distribution can be factorized as $\pi_{\text{pr}}(\tilde{\boldsymbol{\theta}}) = \pi_{\text{pr}}^{(r)}(\tilde{\boldsymbol{\theta}}_r) \pi_{\text{pr}}^{(\perp)}(\tilde{\boldsymbol{\theta}}_\perp)$, where $\pi_{\text{pr}}^{(r)}(\tilde{\boldsymbol{\theta}}_r)$ and $\pi_{\text{pr}}^{(\perp)}(\tilde{\boldsymbol{\theta}}_\perp)$ are densities on the local FIS and CS, respectively [57]. This choice, together with the certified approximation in (6.11), allows us to define the optimal low-dimensional biasing density:

$$\pi_{\mathcal{F}}^{(r)*}(\tilde{\boldsymbol{\theta}}) = \pi_{\text{bias}}^*(\tilde{\boldsymbol{\theta}}) \propto \underbrace{\mathbb{E}_{\pi_{\text{pr}}} [f(\boldsymbol{\theta}; s) \mid \Phi_r \tilde{\boldsymbol{\theta}}_r]}_{\text{reduced optimal biasing}} \pi_{\text{pr}}^{(r)}(\tilde{\boldsymbol{\theta}}_r) \underbrace{\pi_{\text{pr}}^{(\perp)}(\tilde{\boldsymbol{\theta}}_\perp)}_{\text{complementary prior}}. \quad (6.25)$$

We select the parametric biasing distribution from a Gaussian family that decomposes as:

$$\pi_{\text{bias}}(\tilde{\boldsymbol{\theta}}; \mathbf{v}_r) = \pi_{\text{bias}}^{(r)}(\tilde{\boldsymbol{\theta}}_r; \mathbf{v}_r) \pi_{\text{pr}}^{(\perp)}(\tilde{\boldsymbol{\theta}}_\perp); \quad (6.26)$$

this family is allowed to have a general Gaussian form on the local FIS, where $\mathbf{v}_r \in \Upsilon_r$ represents the mean and covariance matrix of this r -dimensional Gaussian, and is equal to the prior in the complementary directions.

Aiming at the optimal biasing density (6.25) using biasing densities in the form (6.26), the CE optimization problem can be re-defined on a low-dimensional space. The KLD between (6.25) and (6.26) reads

$$\text{D}_{\text{KL}}(\pi_{\text{bias}}^* \parallel \pi_{\text{bias}}) = \mathbb{E}_{\pi_{\text{bias}}^*} \left[\ln \left(\frac{p_{\mathcal{F}}^{-1} \mathbb{E}_{\pi_{\text{pr}}} [f(\boldsymbol{\theta}; s) \mid \Phi_r \tilde{\boldsymbol{\theta}}_r] \pi_{\text{pr}}^{(r)}(\tilde{\boldsymbol{\theta}}_r)}{\pi_{\text{bias}}^{(r)}(\tilde{\boldsymbol{\theta}}_r; \mathbf{v}_r)} \right) \right]. \quad (6.27)$$

Similarly to subsection 3.2.4, minimizing $D_{\text{KL}}(\pi_{\text{bias}}^* || \pi_{\text{bias}})$ is equivalent to the maximization:

$$\begin{aligned}
 \mathbf{v}_r^* &= \arg \max_{\mathbf{v}_r \in \Upsilon_r} \mathbb{E}_{\pi_{\text{bias}}^*} \left[\ln \pi_{\text{bias}}^{(r)}(\tilde{\boldsymbol{\theta}}_r; \mathbf{v}_r) \right] \\
 &= \arg \max_{\mathbf{v}_r \in \Upsilon_r} \int_{\tilde{\boldsymbol{\theta}}_r} \int_{\tilde{\boldsymbol{\theta}}_{\perp}} \ln \pi_{\text{bias}}^{(r)}(\tilde{\boldsymbol{\theta}}_r; \mathbf{v}_r) \mathbb{E}_{\pi_{\text{pr}}} \left[f(\boldsymbol{\theta}; s) \mid \boldsymbol{\Phi}_r \tilde{\boldsymbol{\theta}}_r \right] \pi_{\text{pr}}^{(r)}(\tilde{\boldsymbol{\theta}}_r) \pi_{\text{pr}}^{(\perp)}(\tilde{\boldsymbol{\theta}}_{\perp}) d\tilde{\boldsymbol{\theta}}_{\perp} d\tilde{\boldsymbol{\theta}}_r \\
 &= \arg \max_{\mathbf{v}_r \in \Upsilon_r} \int_{\tilde{\boldsymbol{\theta}}_r} \int_{\tilde{\boldsymbol{\theta}}_{\perp}} \ln \pi_{\text{bias}}^{(r)}(\tilde{\boldsymbol{\theta}}_r; \mathbf{v}_r) \left[\int_{\tilde{\boldsymbol{\theta}}_{\perp}} f(\boldsymbol{\Phi}_r \tilde{\boldsymbol{\theta}}_r + \boldsymbol{\Phi}_{\perp} \tilde{\boldsymbol{\xi}}_{\perp}; s) \pi_{\text{pr}}^{(\perp)}(\tilde{\boldsymbol{\xi}}_{\perp}) d\tilde{\boldsymbol{\xi}}_{\perp} \right] \pi_{\text{pr}}^{(r)}(\tilde{\boldsymbol{\theta}}_r) \pi_{\text{pr}}^{(\perp)}(\tilde{\boldsymbol{\theta}}_{\perp}) d\tilde{\boldsymbol{\theta}}_{\perp} d\tilde{\boldsymbol{\theta}}_r \\
 &= \arg \max_{\mathbf{v}_r \in \Upsilon_r} \int_{\tilde{\boldsymbol{\theta}}_r} \left[\int_{\tilde{\boldsymbol{\theta}}_{\perp}} f(\boldsymbol{\Phi}_r \tilde{\boldsymbol{\theta}}_r + \boldsymbol{\Phi}_{\perp} \tilde{\boldsymbol{\xi}}_{\perp}; s) \pi_{\text{pr}}^{(\perp)}(\tilde{\boldsymbol{\xi}}_{\perp}) d\tilde{\boldsymbol{\xi}}_{\perp} \right] \ln \pi_{\text{bias}}^{(r)}(\tilde{\boldsymbol{\theta}}_r; \mathbf{v}_r) \pi_{\text{pr}}^{(r)}(\tilde{\boldsymbol{\theta}}_r) d\tilde{\boldsymbol{\theta}}_r \\
 &= \arg \max_{\mathbf{v}_r \in \Upsilon_r} \mathbb{E}_{\pi_{\text{pr}}} \left[f(\boldsymbol{\Phi}_r \tilde{\boldsymbol{\theta}}_r + \boldsymbol{\Phi}_{\perp} \tilde{\boldsymbol{\theta}}_{\perp}; s) \ln \pi_{\text{bias}}^{(r)}(\tilde{\boldsymbol{\theta}}_r; \mathbf{v}_r) \right]. \tag{6.28}
 \end{aligned}$$

Note that in this CE formulation there is no need for evaluating the conditional expectation in (6.25), as it is normally the case when applying sequential IS in the Bayesian inference setting where MCMC techniques are required to draw samples from (6.25) (see, e.g., [57, 253]). Indeed, the expectation is with respect to the full prior density, thus alleviating computational demands associated with double-loop procedures.

Instead of taking the expectation with respect to the prior in (6.28), we apply IS with biasing distribution $\pi_{\text{bias}}(\boldsymbol{\theta}; \mathbf{v}'_r)$ in (6.26):

$$\mathbf{v}_r^* = \arg \max_{\mathbf{v}_r \in \Upsilon_r} \mathbb{E}_{\pi_{\text{bias}}(\cdot; \mathbf{v}'_r)} \left[f(\boldsymbol{\Phi}_r \tilde{\boldsymbol{\theta}}_r + \boldsymbol{\Phi}_{\perp} \tilde{\boldsymbol{\theta}}_{\perp}; s) \ln \pi_{\text{bias}}^{(r)}(\tilde{\boldsymbol{\theta}}_r; \mathbf{v}_r) \frac{\pi_{\text{pr}}(\tilde{\boldsymbol{\theta}})}{\pi_{\text{bias}}(\tilde{\boldsymbol{\theta}}; \mathbf{v}'_r)} \right], \tag{6.29}$$

for reference parameters \mathbf{v}'_r . We apply the IS estimator of the expectation (6.29) to define the stochastic optimization problem:

$$\mathbf{v}_r^* \approx \hat{\mathbf{v}}_r^* = \arg \max_{\mathbf{v}_r \in \Upsilon_r} \frac{1}{N} \sum_{i=1}^N f(\boldsymbol{\Phi}_r \tilde{\boldsymbol{\theta}}_{r,i} + \boldsymbol{\Phi}_{\perp} \tilde{\boldsymbol{\theta}}_{\perp,i}; s) \ln \pi_{\text{bias}}^{(r)}(\tilde{\boldsymbol{\theta}}_{r,i}; \mathbf{v}_r) w(\tilde{\boldsymbol{\theta}}_i; \mathbf{v}'_r), \tag{6.30}$$

where $\{\tilde{\boldsymbol{\theta}}_i = [\tilde{\boldsymbol{\theta}}_{r,i}, \tilde{\boldsymbol{\theta}}_{\perp,i}]\}_{i=1}^N \stackrel{\text{i.i.d.}}{\sim} \pi_{\text{bias}}(\cdot; \mathbf{v}'_r)$ and $w(\tilde{\boldsymbol{\theta}}_i; \mathbf{v}'_r) = \pi_{\text{pr}}(\tilde{\boldsymbol{\theta}}) / \pi_{\text{bias}}(\tilde{\boldsymbol{\theta}}; \mathbf{v}'_r)$. As in the iCE method, the optimization (6.30) is performed sequentially by introducing a set of smoothing parameters $\{s_j > 0\}_{j=0}^{n_{\text{iv}}}$.

Starting from the full-dimensional space at level $j = 0$, the iCE method applied to the FIS requires two main steps: (i) finding the projector and associated basis operators, and (ii) updating the reduced reference parameters $\hat{\mathbf{v}}_r^{(j+1)}$ of the next intermediate biasing density. To achieve both steps, we use samples from the current biasing density $\pi_{\text{bias}}(\cdot; \hat{\mathbf{v}}_r^{(j)})$. In the following, we slightly abuse the notation to define $\boldsymbol{\theta}^{(j)} \in \mathbb{R}^{d \times N}$ as the matrix containing the full set of samples, such that $\boldsymbol{\theta}^{(j)} = \{\boldsymbol{\theta}_i\}_{i=1}^N \sim \pi_{\text{bias}}(\cdot; \hat{\mathbf{v}}_r^{(j)})$ for each sample $\boldsymbol{\theta}_i \in \mathbb{R}^d$.

- For the first step, we approximate the matrix \mathbf{H} in (6.16) at each level by the self-normalized IS estimator:

$$\mathbf{H}^{(j+1)} \approx \hat{\mathbf{H}}^{(j+1)} = \frac{1}{\widetilde{W}^{(j+1)}} \sum_{i=1}^N \tilde{w}_i^{(j+1)} \left[\nabla \ln f(\boldsymbol{\theta}_i^{(j)}; s_{j+1}) \right] \left[\nabla \ln f(\boldsymbol{\theta}_i^{(j)}; s_{j+1}) \right]^{\text{T}}, \tag{6.31}$$

where $\tilde{w}^{(j+1)} = f(\boldsymbol{\theta}^{(j)}; s_{j+1}) \mathbf{w}^{(j)}$ represents the weight vector computed with respect to the smooth indicator and $\widetilde{W}^{(j+1)} = \sum_{i=1}^N \tilde{w}_i^{(j)}$ is the sum of the weights.

Thereafter, we compute the eigenpairs of the estimator (6.31) to construct $\boldsymbol{\Phi}_r^{(j+1)}, \boldsymbol{\Phi}_{\perp}^{(j+1)}$, and each basis is used to project the full set of samples $\boldsymbol{\theta}^{(j)}$ onto the local FIS and CS as

$$\tilde{\boldsymbol{\theta}}^{(j+1)} = [\tilde{\boldsymbol{\theta}}_r^{(j+1)}, \tilde{\boldsymbol{\theta}}_{\perp}^{(j+1)}] = [\boldsymbol{\Phi}_r^{(j+1),\top}, \boldsymbol{\Phi}_{\perp}^{(j+1),\top}] \cdot \boldsymbol{\theta}^{(j)}, \quad (6.32)$$

with the notation $\tilde{\boldsymbol{\theta}}_r^{(j+1)} \in \mathbb{R}^{r \times N}$ and $\tilde{\boldsymbol{\theta}}_{\perp}^{(j+1)} \in \mathbb{R}^{d-r \times N}$.

- For the second step, we update the reference parameters at each level solving the iCE stochastic optimization problem (6.30) in the reduced coordinate system

$$\hat{\boldsymbol{v}}_r^{(j+1)} = \arg \max_{\boldsymbol{v}_r \in \Upsilon_r} \frac{1}{N} \sum_{i=1}^N \ln \pi_{\text{bias}}^{(r)}(\tilde{\boldsymbol{\theta}}_{r,i}^{(j+1)}; \boldsymbol{v}_r) \bar{w}_i^{(j+1)}, \quad (6.33a)$$

$$\text{with } \bar{w}_i^{(j+1)} = f(\boldsymbol{\theta}_i^{(j)}; \boldsymbol{s}_{j+1}) \frac{\pi_{\text{pr}}(\tilde{\boldsymbol{\theta}}_i^{(j+1)})}{\pi_{\text{bias}}(\tilde{\boldsymbol{\theta}}_i^{(j+1)}; \bar{\boldsymbol{v}}^{(j+1)})}. \quad (6.33b)$$

where $\bar{\boldsymbol{w}}^{(j+1)} = \{\bar{w}_i^{(j+1)}\}_{i=1}^N$ denotes adjusted weights. This correction is required because the parametric biasing density is evaluated on samples that belong to the new basis, but it is defined with reference parameters estimated at the previous level. Therefore, $\hat{\boldsymbol{v}}_r^{(j)}$ (computed using the basis $\boldsymbol{\Phi}_r^{(j)}$) needs to be expressed in the coordinate system induced by $\boldsymbol{\Phi}_r^{(j+1)}, \boldsymbol{\Phi}_{\perp}^{(j+1)}$. Since we employ the Gaussian parametric family and the coordinate transformation is linear, the corrected mean and covariance is computed as

$$\mathbb{E}[\tilde{\boldsymbol{\theta}}^{(j+1)}] = [\boldsymbol{\Phi}_r^{(j+1),\top}, \boldsymbol{\Phi}_{\perp}^{(j+1),\top}] \cdot \boldsymbol{\mu} \quad (6.34a)$$

$$\text{Cov}[\tilde{\boldsymbol{\theta}}^{(j+1)}, \tilde{\boldsymbol{\theta}}^{(j+1)}] = \begin{bmatrix} \boldsymbol{\Phi}_r^{(j+1),\top} \boldsymbol{\Sigma}_1 \boldsymbol{\Phi}_r^{(j+1)} + \boldsymbol{\Phi}_r^{(j+1),\top} \boldsymbol{\Sigma}_2 \boldsymbol{\Phi}_r^{(j+1)} & \boldsymbol{\Phi}_r^{(j+1),\top} \boldsymbol{\Sigma}_1 \boldsymbol{\Phi}_{\perp}^{(j+1)} + \boldsymbol{\Phi}_r^{(j+1),\top} \boldsymbol{\Sigma}_2 \boldsymbol{\Phi}_{\perp}^{(j+1)} \\ \boldsymbol{\Phi}_{\perp}^{(j+1),\top} \boldsymbol{\Sigma}_1 \boldsymbol{\Phi}_r^{(j+1)} + \boldsymbol{\Phi}_{\perp}^{(j+1),\top} \boldsymbol{\Sigma}_2 \boldsymbol{\Phi}_r^{(j+1)} & \boldsymbol{\Phi}_{\perp}^{(j+1),\top} \boldsymbol{\Sigma}_1 \boldsymbol{\Phi}_{\perp}^{(j+1)} + \boldsymbol{\Phi}_{\perp}^{(j+1),\top} \boldsymbol{\Sigma}_2 \boldsymbol{\Phi}_{\perp}^{(j+1)} \end{bmatrix}, \quad (6.34b)$$

where $\boldsymbol{\mu} = \boldsymbol{\Phi}_r^{(j)} \boldsymbol{\mu}_r^{(j)}$, $\boldsymbol{\Sigma}_1 = \boldsymbol{\Phi}_r^{(j)} \boldsymbol{\Sigma}_r^{(j)} \boldsymbol{\Phi}_r^{(j),\top}$, $\boldsymbol{\Sigma}_2 = \boldsymbol{\Phi}_{\perp}^{(j)} \boldsymbol{\Sigma}_{\perp}^{(j)} \boldsymbol{\Phi}_{\perp}^{(j),\top}$, and the parameters $\boldsymbol{\mu}_r^{(j)}$ and $\boldsymbol{\Sigma}_r^{(j)}$ constitute the reference parameter $\hat{\boldsymbol{v}}_r^{(j)}$ estimated at level j . The components of the adjusted reference parameter $\bar{\boldsymbol{v}}^{(j+1)}$ used for the computation of the adjusted weights are then given by (6.34).

Note that the full covariance (6.34b) has correlated components. Hence, the biasing density used for the computation of the adjusted weights $\bar{w}_i^{(j+1)}$ does not necessarily admits the decomposition (6.26). In this case, π_{bias} becomes a d -dimensional Gaussian with parameters given by (6.34). This is only required to compute the corrected weights, the optimization problem in (6.33b) is still performed in the reduced space.

The complete procedure of the improved cross-entropy method with failure-informed dimension reduction (called iCEred), is summarized in Algorithm 10.

Remark 6.2. A key question for iCEred is how large N must be for an accurate MC approximation of $\hat{\mathbf{H}}$. At least d evaluations are necessary to approximate a full rank \mathbf{H} . However, the number of samples needed for accurate approximation is related to the effective rank of \mathbf{H} [54]. The heuristic rule $N = \alpha \cdot r \ln(d)$ provides some intuition about the required sample size; here $\alpha \in [2, 10]$ is an oversampling factor (see, e.g., [54, p.35]). As a result, we can modify Algorithm 10 such that we employ two types of sample sizes, one N_{LSF} associated to LSF evaluations and another N_{grad} related to the gradient evaluations. The value of N_{grad} can be adapted using the heuristic rule at each intermediate level of iCEred. This suggested adaption strategy is left for future study.

Algorithm 10 iCERed: iCE method with failure-informed dimension reduction.

Input : LSF $g(\boldsymbol{\theta})$, smooth indicator function $f(\boldsymbol{\theta}; s)$, gradient $\nabla \ln f(\boldsymbol{\theta}; s)$, number of samples per level N , tolerance in the approximation ε , target coefficient of variation δ^* , maximum iterations t_{\max}

- 1 Set the standard Gaussian prior density $\pi_{\text{pr}}(\boldsymbol{\theta})$
- 2 Set the parametric family of standard Gaussian biasing densities $\pi_{\text{bias}}(\boldsymbol{\theta}; \mathbf{v})$
- 3 Set initial reference parameters $\hat{\mathbf{v}}^{(0)}$ from the prior parameters, and the smoothing parameter $s_0 \leftarrow \infty$
- 4 Initial level $j \leftarrow 0$
- 5 **while** *True* **do**
- 6 **if** $j = 0$ **then**
- 7 Generate N samples from the biasing distribution $\boldsymbol{\theta}^{(j)} \sim \pi_{\text{bias}}(\cdot; \hat{\mathbf{v}}^{(j)})$
- 8 Since $\hat{\mathbf{v}}^{(0)}$ are selected from the prior, the weights $\mathbf{w}^{(j)}$ are equal to one
- 9 **else**
- 10 Generate N samples from the reduced biasing distribution $\tilde{\boldsymbol{\theta}}_r^{(j)} \sim \pi_{\text{bias}}(\cdot; \hat{\mathbf{v}}_r^{(j)})$
- 11 Compute the weights $\mathbf{w}^{(j)} \leftarrow \exp\left(\ln \pi_{\text{pr}}^{(r)}(\tilde{\boldsymbol{\theta}}_r^{(j)}) - \ln \pi_{\text{bias}}^{(r)}(\tilde{\boldsymbol{\theta}}_r^{(j)}; \hat{\mathbf{v}}_r^{(j)})\right)$
- 12 Construct the full set of parameters $\boldsymbol{\theta}^{(j)} \leftarrow \Phi_r^{(j)} \tilde{\boldsymbol{\theta}}_r^{(j)} + \Phi_{\perp}^{(j)} \tilde{\boldsymbol{\theta}}_{\perp}^{(j)}$
- 13 **end**
- 14 Evaluate LSF $\mathbf{g}_{\text{eval}} \leftarrow g(\boldsymbol{\theta}^{(j)})$ and indicator function $\mathbf{d}_{\text{eval}} \leftarrow \mathbf{g}_{\text{eval}} \leq 0 = \mathbf{1}_{\mathcal{F}}(\boldsymbol{\theta}^{(j)})$
- 15 Evaluate smooth indicator function $\mathbf{f}_{\text{eval}} \leftarrow f(\boldsymbol{\theta}^{(j)}; s_j)$
- 16 Compute the coefficient of variation of the ratio between the indicator function and its smooth approximation

$$\hat{\text{cV}}^{(j)} = \frac{\sqrt{\hat{\mathbf{V}}[\mathbf{d}_{\text{eval}}/\mathbf{f}_{\text{eval}}]}}{\hat{\mathbf{E}}[\mathbf{d}_{\text{eval}}/\mathbf{f}_{\text{eval}}]}$$
- 17 **if** $(\hat{\text{cV}}^{(j)} \leq \delta^*)$ or $(j \geq t_{\max})$ **then**
- 18 **Break**
- 19 **end**
- 20 Use the samples $\boldsymbol{\theta}^{(j)}$ and the smooth indicator function to compute s_{j+1} as per (3.36)
- 21 Update the smooth indicator values $\mathbf{f}_{\text{eval}} \leftarrow f(\boldsymbol{\theta}^{(j)}; s_{j+1})$
- 22 Compute the weights associated to the smooth indicator $\tilde{\mathbf{w}}^{(j+1)} \leftarrow \mathbf{w}^{(j)} \cdot \mathbf{f}_{\text{eval}}$ and its sum \tilde{W}
- 23 Compute the gradient of the log-smooth indicator function, $\nabla \ln f(\boldsymbol{\theta}^{(j)}; s_{j+1})$
- 24 Use $\nabla \ln f(\boldsymbol{\theta}^{(j)}; s_{j+1})$ to compute the estimator $\hat{\mathbf{H}}$ as per (6.31) and solve the eigenvalue problem $\hat{\mathbf{H}}\Phi = \Lambda\Phi$
- 25 Use the tolerance ε to find the rank r as per (6.20)
- 26 Construct the FIS basis $\Phi_r^{(j+1)} \leftarrow \Phi_{1:r}$, and the CS basis $\Phi_{\perp}^{(j+1)} \leftarrow \Phi_{r+1:d}$
- 27 Project the samples to the local FIS and CS, $\tilde{\boldsymbol{\theta}}_r \leftarrow \Phi_r^{(j+1),\text{T}} \boldsymbol{\theta}^{(j)}$, $\tilde{\boldsymbol{\theta}}_{\perp} \leftarrow \Phi_{\perp}^{(j+1),\text{T}} \boldsymbol{\theta}^{(j)}$
- 28 **if** $j > 0$ **then**
- 29 Samples in the new coordinates $\tilde{\boldsymbol{\theta}} = [\tilde{\boldsymbol{\theta}}_r, \tilde{\boldsymbol{\theta}}_{\perp}]$
- 30 Compute the corrected reference parameters $\bar{\mathbf{v}}$ using the mean and covariance in (6.34)
- 31 Compute the adjusted weights $\bar{\mathbf{w}} \leftarrow \mathbf{f}_{\text{eval}} \cdot \exp\left(\ln \pi_{\text{pr}}(\tilde{\boldsymbol{\theta}}) - \ln \pi_{\text{bias}}(\tilde{\boldsymbol{\theta}}; \bar{\mathbf{v}})\right)$
- 32 Find the next reference parameters $\hat{\mathbf{v}}_r^{(j+1)}$ by solving the stochastic optimization problem (6.33b) using samples $\tilde{\boldsymbol{\theta}}_r$ and weights $\bar{\mathbf{w}}$.
- 33 $j \leftarrow j + 1$
- 34 **end**
- 35 Compute the IS estimator $\hat{p}_{\mathcal{F}} \leftarrow \frac{1}{N} \sum_{i=1}^N d_{\text{eval}}(i) \cdot w_i^{(j)}$

Output : $\hat{p}_{\mathcal{F}}$

6.3.2. Refinement step

Computational approaches such as adjoint methods [11, 53] and automatic differentiation [157] can be used to calculate efficiently the gradient of the log-smooth indicator function in (6.31). However, when direct application of any of these tools is unfeasible, gradient evaluations are a computationally demanding task.

This motivates the definition of a refinement step for the iCEred method that allows improving the failure probability estimate, keeping the number of gradient evaluation fixed. The idea is to employ the reference parameters of the low-dimensional biasing density at the last level, to draw additional samples from the fitted parametric density. Such strategy has been applied to the standard CE method in [244]. This complementary step to iCEred adaptively increases the LSF evaluations but it does not involve additional gradient computations. The sample size is augmented at each iteration to obtain new LSF evaluations that will improve the IS estimate. The coefficient of variation of the updated failure probability estimate $\widehat{cv}(\widehat{p}_{\mathcal{F}})$ is then monitored until it reaches a predefined threshold $\bar{\delta}$. Since the value of $\widehat{cv}(\widehat{p}_{\mathcal{F}})$ at each iteration is typically noisy, we control its mean from a set of previous iteration values. Algorithm 11 describes in detail this optional “post-processing” procedure.

Algorithm 11 Refinement step for iCEred.

Input : target coefficient of variation $\bar{\delta}$, number of coefficient of variation checks m , sample size increment M . Last iCEred iteration values of: reference parameters $\widehat{\mathbf{v}}_r$, indicator function \mathbf{d}_{eval} , IS weights \mathbf{w} , and bases Φ_r, Φ_{\perp}

```

1  $k = 1$ 
2 while True do
3   Compute IS estimator  $\widehat{p}_{\mathcal{F}} \leftarrow \frac{1}{N} \sum_{i=1}^N d_{\text{eval}}(i) \cdot w_i$ 
4   Estimate the variance of the IS estimator  $\widehat{V}[\widehat{p}_{\mathcal{F}}] \leftarrow \frac{1}{N-1} \left[ \frac{1}{N} \sum_{i=1}^N d_{\text{eval}}(i) \cdot w_i^2 - \widehat{p}_{\mathcal{F}}^2 \right]$ 
5   Estimate the coefficient of variation  $\widehat{cv}(k) \leftarrow \sqrt{\widehat{V}[\widehat{p}_{\mathcal{F}}]} / \widehat{p}_{\mathcal{F}}$ 
6   if  $k \bmod m = 0$  then
7     Compute the mean of the previous  $m$  values stored in  $\widehat{cv}, \mu_{cv}$ 
8     if  $\mu_{cv} \leq \bar{\delta}$  then
9       | Break
10    end
11  else
12    Generate  $M$  samples from the biasing density in the local FIS  $\widetilde{\boldsymbol{\theta}}_r \sim \pi_{\text{bias}}^{(r)}(\cdot; \widehat{\mathbf{v}}_r)$ 
13    Generate  $M$  samples from the prior density in the local CS  $\widetilde{\boldsymbol{\theta}}_{\perp} \sim \pi_{\text{pr}}^{(1)}(\cdot)$ 
14    Construct the full set of parameter samples  $\boldsymbol{\theta} \leftarrow \Phi_r \widetilde{\boldsymbol{\theta}}_r + \Phi_{\perp} \widetilde{\boldsymbol{\theta}}_{\perp}$ 
15    Evaluate LSF  $\mathbf{g}_{\text{eval}} \leftarrow g(\boldsymbol{\theta})$  and indicator function  $\mathbf{d}_{\text{extra}} \leftarrow \mathbf{g}_{\text{eval}} \leq 0 = \mathbf{1}_{\mathcal{F}}(\boldsymbol{\theta})$ 
16    Compute the weights  $\mathbf{w}_{\text{extra}} \leftarrow \exp \left( \ln \pi_{\text{pr}}^{(r)}(\widetilde{\boldsymbol{\theta}}_r) - \ln \pi_{\text{bias}}^{(r)}(\widetilde{\boldsymbol{\theta}}_r; \widehat{\mathbf{v}}_r) \right)$ 
17    Append the extra indicator function and weight values  $\mathbf{d}_{\text{eval}} \leftarrow [\mathbf{d}_{\text{eval}}, \mathbf{d}_{\text{extra}}]$  and  $\mathbf{w} \leftarrow [\mathbf{w}, \mathbf{w}_{\text{extra}}]$ 
18     $k \leftarrow k + 1$  and  $N \leftarrow N + M$ 
19  end
20 end
Output :  $\widehat{p}_{\mathcal{F}}$ 

```

We conclude this section by illustrating the evolution of the iCEred method for two different examples. For both problems, the dimension of the parameter space is $d = 2$ and we select the approximation tolerance as $\varepsilon = 0.01$. The first row of Figure 6.2 shows the results for a linear LSF problem that has a clear direction in which the parameter space varies the most; the rank of the projector is found to be $r = 1 < d$. The second row of Figure 6.2 plots the solution for a quadratic LSF in which the parameter space does not have a low-dimensional structure; the rank of the projector is $r = 2 = d$ and the iCEred algorithm reduces to the standard iCE method.

6.4. Numerical applications

We test the proposed method on four examples. For the first two experiments, the gradient of the LSF is available analytically and the failure probabilities are independent of the dimension of the input parameter space. This allows us to perform several parameter studies to validate

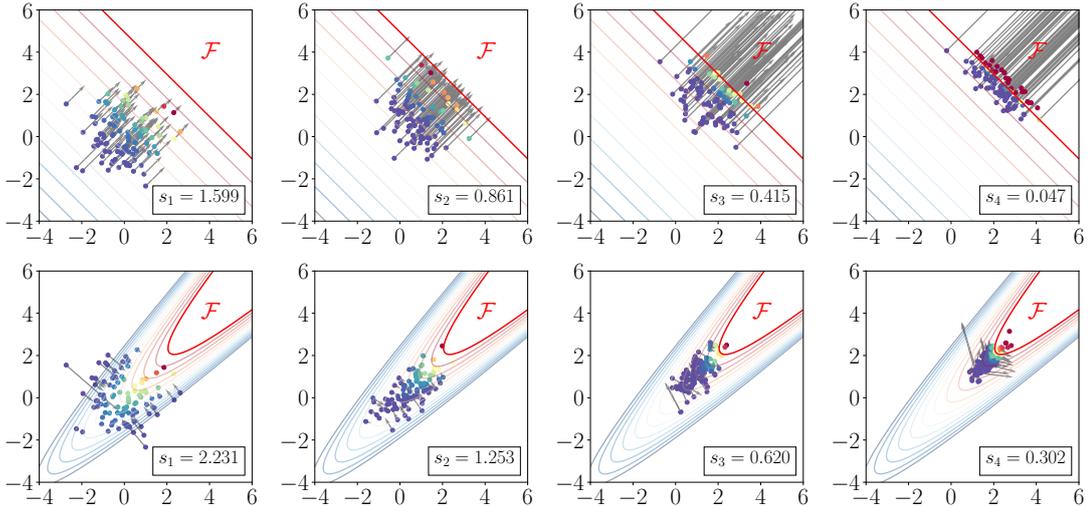


Figure 6.2: Evolution of the iCEred method for problems *with* (top $r = 1 < d$) and *without* (bottom $r = 2 = d$) intrinsic low-dimensional structure of the parameter space. The arrows show the values of $\nabla \ln f(\boldsymbol{\theta}; s_j)$ and the color in the samples accounts for the smooth indicator values.

the applicability and accuracy of the iCEred method. In the final examples, we consider models involving spatially variable parameters modeled as random fields. For the latter example, the LSF gradient is computed using the adjoint method, which is derived in Appendix B.3.1 for this particular problem.

Moreover, we consider the gradients of the natural logarithm of the smooth indicators (3.3)

$$\nabla \ln f^{\log}(\boldsymbol{\theta}; s) = -\frac{\nabla g(\boldsymbol{\theta})}{s} \left[1 + \tanh\left(\frac{g(\boldsymbol{\theta})}{s}\right) \right], \quad \nabla \ln f^{\text{erf}}(\boldsymbol{\theta}; s) = -\frac{\nabla g(\boldsymbol{\theta})}{s} \cdot \frac{\phi\left(-\frac{g(\boldsymbol{\theta})}{s}\right)}{\Phi\left(-\frac{g(\boldsymbol{\theta})}{s}\right)}, \quad (6.35)$$

for the computation of the estimator $\widehat{\mathbf{H}}$ in (6.31).

We note that the cost of the eigenvalue problem associated to $\widehat{\mathbf{H}}$ at each iteration is negligible compared to the LSF/gradient evaluations. The dimension d needs to be considerably large ($\mathcal{O}(10^4)$) to make a significant contribution to the computational cost of iCEred. Moreover, the parametric family of biasing densities is chosen as single Gaussian densities, i.e., $\Pi = \{\pi_{\text{bias}}(\boldsymbol{\theta}; \boldsymbol{v}) = \mathcal{N}(\boldsymbol{\theta}; \boldsymbol{v}) \mid \boldsymbol{v} = [\boldsymbol{\mu}, \boldsymbol{\Sigma}] \in \Upsilon\}$.

The inputs to iCEred are studied/selected as follows: (i). The choice of the smooth indicator function and its gradient is analyzed in the first example; (ii). The influence of the number of sampler per level N is studied in all examples; (iii). The target coefficient of variation for the iCEred method is fix to $\delta^* = 1.5$. This is based on the studies [180, 179], which directly translate to the iCEred framework. This value yields a probability estimate whose coefficient of variation is around $\delta^*/\sqrt{N} = 0.047$ (using $N = 10^3$ samples); and (iv). Based on [253], we fix the tolerance of the certified approximation as $\varepsilon = 0.01$.

Unless specified, the inputs to the refinement step are: (i). The target coefficient of variation of the failure probability estimate is $\bar{\delta} = 0.05$; (ii). We use $m = 10$ to explore the coefficient of variation throughout iterations. Since the value of $\widehat{c\hat{v}}(\widehat{p}_{\mathcal{F}})$ is typically noisy, we control its average every m iterations. This is not a crucial parameter in the refinement step because the variability of $\widehat{c\hat{v}}$ decreases as the iterations evolve; and (iii). The increment in the sample size is $M = 50$. Using larger M implies more LSF evaluations, which in turn helps to obtain a better estimate of $\widehat{c\hat{v}}$ and the refinement step potentially finishes faster. However, if M is too large, one can perform unnecessary LSF evaluations. As a result, it is recommended to define an $M \ll N$

and let the refinement step iterate until it matches the target $\bar{\delta}$.

Finally, we provide Python scripts for the first two examples in a [GitHub repository](#).

6.4.1. Example 1: linear LSF in varying dimensions

We consider a LSF expressed as a linear combination of d independent standard Gaussian random variables [76]:

$$g(\boldsymbol{\theta}) = \beta - \frac{1}{\sqrt{d}} \sum_{i=1}^d \theta_i \quad \text{with gradient} \quad \nabla g(\boldsymbol{\theta}) = \left[-\frac{1}{\sqrt{d}} \cdot \mathbf{1}_d \right]^\top, \quad (6.36)$$

where β is a given maximum threshold, and $\mathbf{1}_d$ denotes a d -dimensional vector containing ones. The probability of failure is independent of the dimension d and it can be computed analytically as $p_{\mathcal{F}} = \Phi(-\beta)$, with $\Phi(\cdot)$ denoting the standard Gaussian CDF.

We initially test the two smooth approximations of the log-indicator gradient (6.35) used within the iCEred method. In this case, the refinement step of the algorithm is not implemented. This is to evaluate the standalone performance of the method. Ten different values of decreasing target failure probabilities are fixed as $p_{\mathcal{F}} = [1 \times 10^{-1}, 1 \times 10^{-2}, \dots, 1 \times 10^{-10}]$, which have an associated list of thresholds $\beta = -\Phi^{-1}(p_{\mathcal{F}}) \approx [1.282, 2.326, \dots, 6.361]$. Moreover, for each threshold β , three different dimensions $d \in \{2, 358, 1000\}$ are employed to define the LSF (6.36). These allows us to test the method for low- and high-dimensional settings. A parameter study on the number of samples per level is also performed and we choose this value from $N \in \{100, 250, 500, 1000\}$. In this example, there exists a clear low-dimensional structure of the parameter space and the rank of the projector \mathbf{P}_r is $r = 1$ for all parameter cases.

Figure 6.3 shows the coefficient of variation of the failure probability estimates, computed as an average of 100 independent iCEred runs. One observes that under the LSF (6.36), the performance of the iCEred method is independent of the dimension and magnitude of the failure probability. As the number of samples increases, the smooth approximations yield similar results. However, when the number of samples per level is small, the smooth approximation f^{erf} which is based on the standard Gaussian CDF produces larger $\widehat{\text{cv}}(\widehat{p}_{\mathcal{F}})$ values. Note in (6.35) that the standard Gaussian CDF is evaluated in the denominator. As the number of iCEred levels increases, the smoothing parameter $s \rightarrow 0$. For very small values of s , $\Phi(-g(\boldsymbol{\theta})/s)$ takes values close to zero, and hence the value of $\nabla \ln f(\boldsymbol{\theta}; s)$ is not defined. When this occurs, the matrix $\widehat{\mathbf{H}}$ cannot be estimated and the basis of the previous level is used. Although this problem is inherent to any approximation of the indicator function, it appears that f^{log} based on the logistic function is more robust than f^{erf} . This is probably because f^{erf} involves numerical approximation of the error function. As a result, f^{log} yields more stable results for sample size $N = 100$. Nevertheless, under the LSF (6.36), we observe that the performance of the iCEred method is independent of the dimension and magnitude of the failure probability estimate.

We now fix the smooth approximation of the indicator function to f^{log} and the threshold to $\beta = 3.5$ ($p_{\mathcal{F}} \approx 2.33 \times 10^{-4}$). In Figure 6.4, we plot a comparison between the iCEred method (without refinement) with the standard CE and iCE methods for increasing dimension of the parameter space and number of samples per level $N \in \{250, 1000, 2500\}$. The CE method is implemented with $\rho = 0.1$ (cf., subsection 3.2.4) and the iCE method with $\delta^* = 1.5$ (cf., subsection 3.2.5); for both approaches the biasing distribution family consist of single Gaussian densities. We observe that the performance of standard CE and iCE methods deteriorates with increasing dimension of the parameter space, and that augmenting the sample size improves the accuracy in the estimation of the target failure probability. This is related to the solution of the stochastic optimization problem required in both approaches, which amounts to fitting high-dimensional Gaussian densities. The number of samples required to properly perform this step depends on the rank of the covariance matrix of the Gaussian biasing density, and it roughly

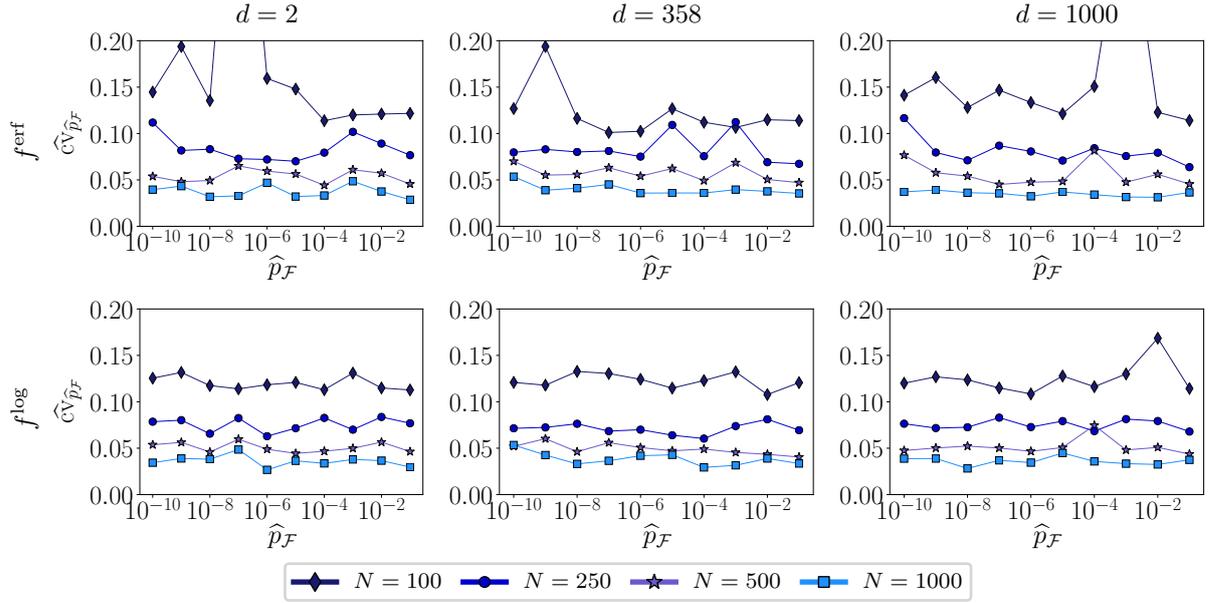


Figure 6.3: Coefficient of variation of the failure probability estimate for example 6.4.1 using different target probabilities of failure, number of samples, smooth indicators (rows), and increasing dimension (columns).

scales quadratically with the dimension. Although the results can improve if other parametric families of biasing densities are employed (as those proposed in [244, 179]), under the Gaussian parametric family the standard methods are inadequate to perform reliability analysis in high dimensions. The iCERed method estimates consistently the target failure probability for all the dimensions and even at small sample size. Since the rank of the projector is $r = 1$, the updating of reference parameters in iCERed amounts to fitting one-dimensional Gaussian densities for all the dimension cases.

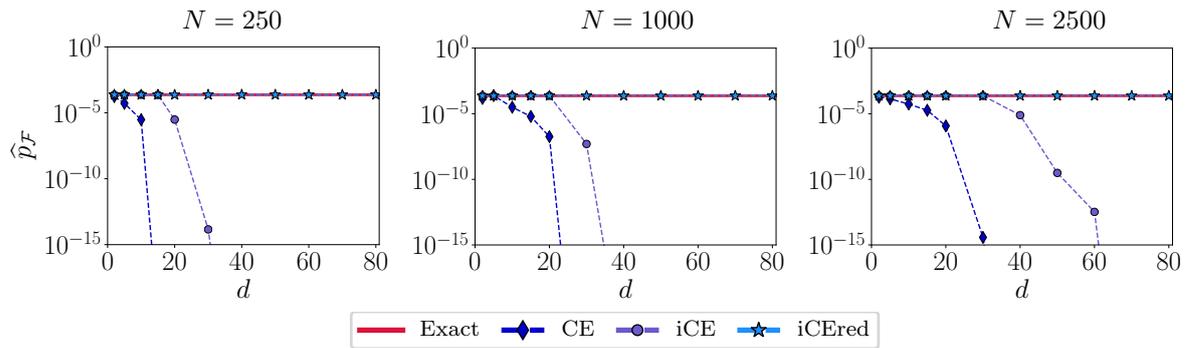


Figure 6.4: Failure probability estimate for example 6.4.1 using the CE, iCE, and iCERed methods (with single Gaussian parametric family) for increasing dimension of the parameter space and different number of samples per level (columns).

We conclude this example by showing the advantage of the refinement step for iCERed. In this case, the number of samples per level is selected from the set $N \in \{100, 250, 500, 1000, 2500, 5000\}$. Figure 6.5 shows the number of LSF and gradient calls, together with the coefficient of variation of the failure probability estimate (again, as an average of 100 independent simulations). Note that for sample sizes 100, 250 and 500, there is an increase in the number of LSF evaluations, due to the refinement. This is because small values of N are not sufficient to compute a probability of failure estimate that has $\widehat{\text{cv}}(\widehat{p}_{\mathcal{F}})$ smaller than the value of $\bar{\delta}$. Therefore, extra LSF computations are performed in the refinement algorithm to reduce the $\widehat{\text{cv}}(\widehat{p}_{\mathcal{F}})$ to $\bar{\delta}$. For large sample sizes,

this step has no effect since the $\widehat{c\mathbf{v}}(\widehat{p}_{\mathcal{F}})$ obtained from iCERed is already below the predefined $\bar{\delta}$.

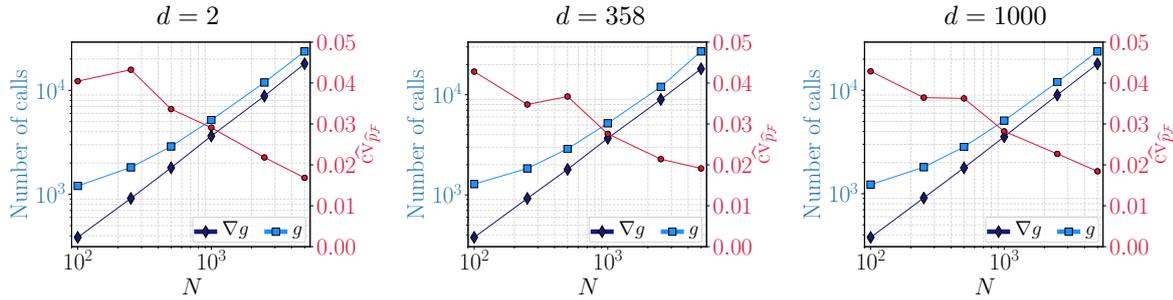


Figure 6.5: iCERed with refinement for example 6.4.1: number of LSF and gradient calls for different number of samples, and dimensions (columns). The $\widehat{c\mathbf{v}}$ of the failure probability estimate is also shown (in red).

6.4.2. Example 2: quadratic LSF in varying dimensions

We employ the LSF discussed in subsection 3.3.1, which has the structure of the LSF in example 6.4.1 but adding a quadratic term:

$$g(\boldsymbol{\theta}) = \beta + \frac{\kappa}{4} (\theta_1 - \theta_2)^2 - \frac{1}{\sqrt{d}} \sum_{i=1}^d \theta_i \quad \text{with gradient} \quad (6.37a)$$

$$\nabla g(\boldsymbol{\theta}) = \left[\frac{\kappa}{2} (\theta_1 - \theta_2) - \frac{1}{\sqrt{d}}, \frac{\kappa}{2} (\theta_2 - \theta_1) - \frac{1}{\sqrt{d}}, -\frac{1}{\sqrt{d}} \cdot \mathbf{1}_{d-2} \right]^T, \quad (6.37b)$$

where $\beta = 4$ and the parameter κ defines the curvature of the LSF at the point in the parameter space with the largest probability density (the larger the curvature, the smaller $p_{\mathcal{F}}$). The reference probability of failure is independent of the dimension and it can be computed analytically using (3.38) [68, 177]. We employ the iCERed method with adaptation to estimate the probability of failure associated to the LSF (6.37a). The target failure probabilities are $p_{\mathcal{F}} \approx \{6.62 \times 10^{-6}, 4.73 \times 10^{-6}\}$ for curvature values $\kappa = \{5, 10\}$. Three different dimensions $d = \{2, 334, 1000\}$ are also employed to define the LSF.

In Figure 6.6, we plot the eigenvalue decay and first two eigenvectors of the estimator $\widehat{\mathbf{H}}$, for different number of samples per level $N \in \{100, 500, 1000\}$ and a fixed curvature $\kappa = 5$. The results are illustrative since they correspond to a single iCERed simulation. We note that these eigenpairs correspond to the eigendecomposition of $\widehat{\mathbf{H}}$ at the last iCERed iteration. The rank of the projector is $r = 2$ and it is constant through all the iCERed levels. As in example 6.4.1, this is a problem with a clear low-dimensional structure since only the parameter components θ_1 and θ_2 define the geometry of the LSF; the remaining components represent a linear extrusion in the complementary directions. Thus, the rank remains unchanged even when using small number of samples per level. Note also that for $d = 2$ there is no associated dimension reduction and the iCERed method reduces to standard iCE (as seen in Figure 6.2).

Figure 6.7 shows the number of calls of the LSF and its gradient, together with the coefficient of variation of the failure probability estimates; the number of samples per level is now chosen as $N \in \{100, 250, 500, 1000, 2500, 5000\}$. The results are computed as an average of 100 independent simulation runs. Since we perform refinement, all $\widehat{c\mathbf{v}}(\widehat{p}_{\mathcal{F}})$ values are close to the predefined $\bar{\delta}$ for sample size smaller than $N = 1000$. This behavior is similar for all studied dimension cases. Although the curvature parameter increases the nonlinearity of the problem, the performance of iCERed remains insensitive to this value. For instance, when using $N = 1000$ samples per level, the failure probability estimates are $\widehat{p}_{\mathcal{F}} \approx [6.61 \times 10^{-6}, 4.74 \times 10^{-6}]$.

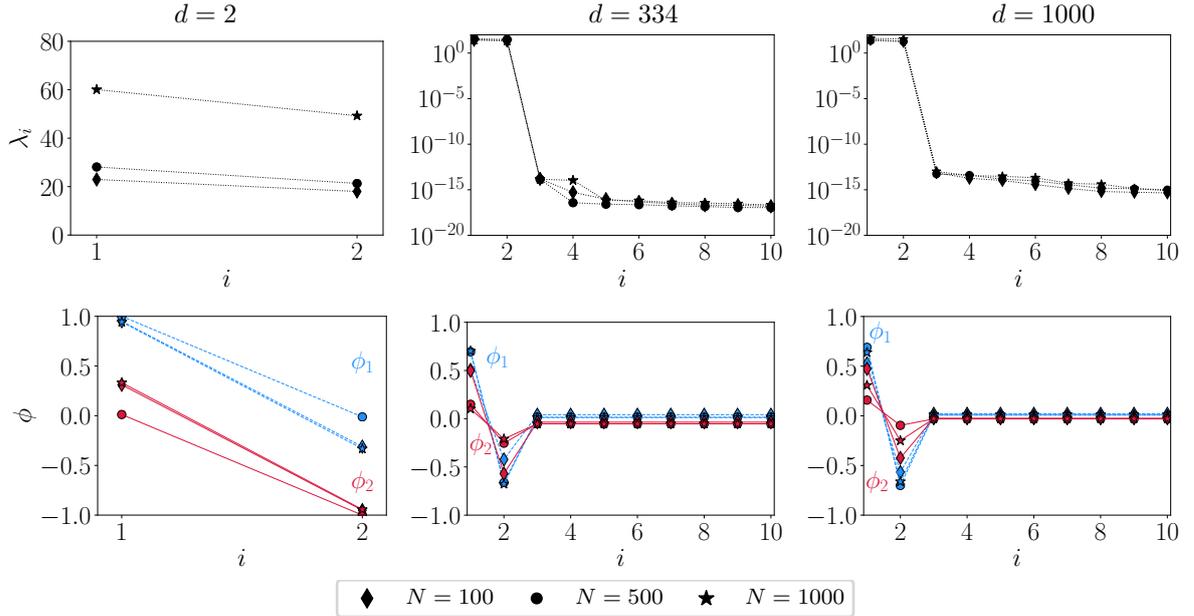


Figure 6.6: Eigenvalue decay (1st row) and first two eigenvectors (2nd row) of the matrix $\hat{\mathbf{H}}$ corresponding to example 6.4.2 with $\kappa = 5$, for sample size and increasing dimension (columns).

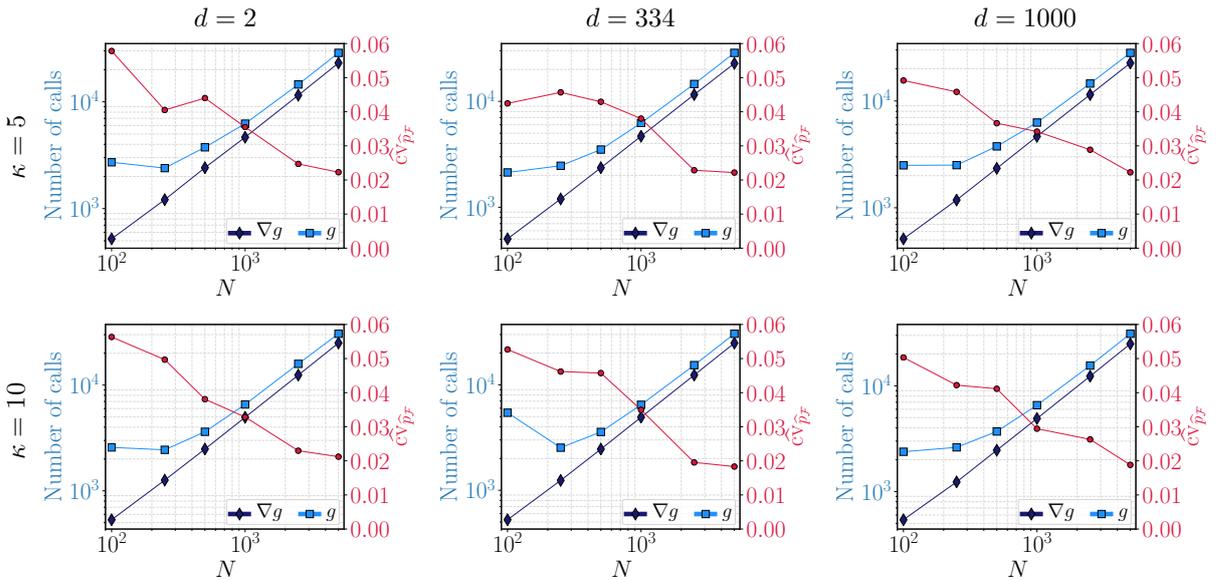


Figure 6.7: iCERed with refinement for example 6.4.2: number of LSF and gradient calls for different sample size, curvature values (rows) and dimensions (columns). The $\hat{c}_{\hat{p}_F}$ of the failure probability estimate is also shown (in red).

6.4.3. Example 3: one-dimensional diffusion

We consider the stochastic linear elliptic equation defined in subsection 3.3.2. In this case, the covariance kernel of the Gaussian field is exponential with correlation length $\ell = 0.1$. We employ the KL expansion to represent the conductivity field using truncation orders $k = 200$. The dimension of the parameter space is $d = 201$ after including the flux random variable.

Recall that the failure of the system occurs when the maximum pressure head exceeds a given admissible threshold. The LSF is defined in (3.44). The forward problem has an analytical solution (3.43), and the gradient of the LSF is estimated by finite differences. The reference value

of the probability of failure is estimated with $N = 5000$ samples per level using SIS, CE, iCE and SuS methods in subsection 3.3.2. The values are listed in Table 3.2, which are respectively for each method, $p_{\mathcal{F}} = \{3.36, 3.44, 3.41, 3.23\} \times 10^{-5}$. The value with the smallest coefficient of variation is given by standard iCE, which is $p_{\mathcal{F}} = 3.41 \times 10^{-5}$.

We assess the accuracy and performance of iCEred employing sample sizes $N = \{100, 250, 500, 1000\}$. For each N , we perform $N_{\text{sim}} = 50$ independent runs of the algorithm. The logistic function is chosen as the smooth approximation of the failure indicator function. In the refinement step, the target coefficient of variation is set to $\bar{\delta} = 0.07$, and the increment in the sample size is $M = 100$.

Table 6.2 shows for different sample sizes the failure probability estimates, together with the number of LSF and LSF gradient calls, denoted g_{call} and $(\nabla g)_{\text{call}}$, respectively. For each of the sample sizes, the average number of levels obtained in iCEred is $n_{\text{lv}} = \{6, 5, 5, 4\}$, and the average final ranks are $r = \{8, 10, 10, 10\}$. The mean value of the probability of failure computed by repeated runs of iCEred is close to the reference, whereas the coefficient of variation of the estimate by the methods in Chapter 3 are significantly larger than the one of iCEred (under same sample size). The iCEred method requires considerably less number of LSF evaluations. In this example, iCEred proves to be more efficient than the standard methods since the gradient computations are easy to compute.

However, from Table 6.2 it is seen that $N = 100$ is not sufficient to obtain an accurate estimator of the probability of failure. This has to do with the MC approximation of the second moment matrix of the log-smooth indicator function gradient. The final rank of the projector is approximately $r = 10$. If we follow the heuristic rule to determine the sample size provided by AS in subsection 6.1.2, the number of samples required to estimate $r = 10$ eigenvalues using the limiting oversampling factors $\alpha = \{2, 10\}$, is $N = \{103, 530\}$. In this example, $\alpha = 2$ is not sufficient to get a stable solution of the eigenvalues of $\hat{\mathbf{H}}$.

Table 6.2.: Failure probability estimates by iCEred of the example in subsection 6.4.3: the iCEred results are shown as an average of 50 simulations.

N	g_{call}	$(\nabla g)_{\text{call}}$	$\hat{p}_{\mathcal{F}}$	$\widehat{\text{cv}}(p_{\mathcal{F}})$
100	15178	582	2.94×10^{-5}	0.279
250	4098	1340	3.39×10^{-5}	0.077
500	3992	2260	3.42×10^{-5}	0.057
1000	6226	4160	3.40×10^{-5}	0.049

This can also be seen in Figure 6.8. In the first row, we plot the evolution of the rank with the smoothing parameters for the investigated sample sizes and the 50 independent simulations. Using $N = 100$, the relation between rank and smoothing parameter is more noisy compared to the results with increasing N . In the second row, we plot the evolution of the iCEred iterations and the rank for 50 independent simulations. One requires in average more levels to estimate the failure probability when using small sample sizes. As we increase the sample size, the relation between the intermediate levels in iCEred and the rank seems to be linear.

Note also the effect of the parametrization induced by the smoothing parameters: at initial levels when $s > 0$, the rank of the projector is small since the approximation is still close to the prior; as we approach the posterior with $s \rightarrow 0$, the rank of the projector increases to estimate the appropriate low-dimensional approximation.

6.4.4. Example 4: two-dimensional plate with a hole

We consider the steel plate model used subsection 2.7.2. The Young's modulus is a random field modeled with the isotropic exponential kernel and correlation length $\ell = 0.04$ m. In this case,

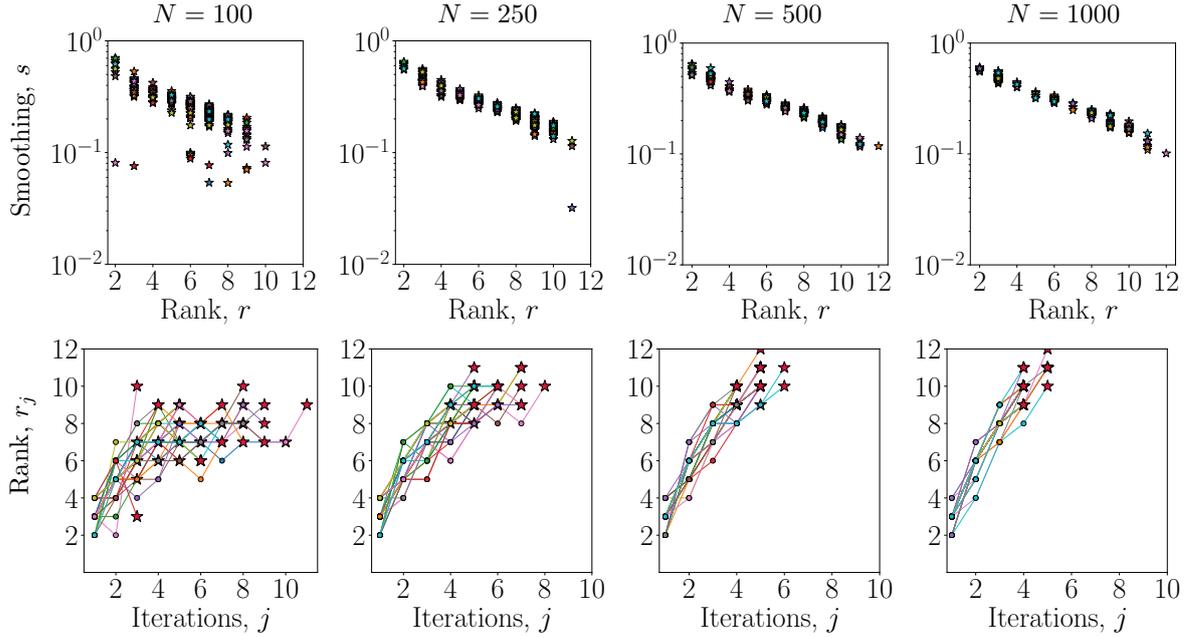


Figure 6.8: Evolution of the rank for different simulation runs and $N = \{100, 250, 500, 1000\}$. 1st row: rank and smoothing parameters. 2nd row: iCEred iterations and rank. The star marks the final value of the rank at the given run.

the body forces in (2.43) are assumed to be approximately zero. The number of terms in the expansion is fixed to $k = 868$, which accounts for 92.5% of the spatial average of the variance of the Gaussian random field $\ln E$. The eigenpairs are estimated with the Nyström method using 107 GL points in each direction.

As in subsection 4.6.2, we select the principal stress σ_1 as the target QoI. Recall that this quantity depends on the stress field of the plate $\boldsymbol{\sigma}(\mathbf{x}) := [\sigma_{x_1}(\mathbf{x}), \sigma_{x_2}(\mathbf{x}), \tau_{x_1x_2}(\mathbf{x})]^\top$, which is computed after obtaining the displacement field via (2.43) and applying Hooke's law for continuous media. The failure of the plate occurs when the value of σ_1 at a control point $\mathbf{x}_{\text{ctr}} = [0.16, 0.18]$ (see Figure 2.11), exceeds a yield tensile strength of 320 MPa; the LSF is defined as

$$g(\boldsymbol{\theta}) = 320 - \sigma_1(\boldsymbol{\sigma}(\mathbf{x}_{\text{ctr}}; \boldsymbol{\theta})); \quad (6.38)$$

we remark that for computational reasons the σ_1 stress defining the LSF is evaluated at the GL point of the element closest to the control node \mathbf{x}_{ctr} . The gradient of the LSF is computed by the adjoint method derived in Appendix B.3.1.

The uncertain parameter vector is given by $\boldsymbol{\theta} = [\theta^{(q)}, \boldsymbol{\theta}^{(\text{KL})}]$ to account for both, the load random variable and Young's modulus random field. Here, $\theta^{(q)}$ is also a standard Gaussian random variable since we apply the transformation $\theta^{(q)} = (q - \mu_q)/\sigma_q$. As a result, the stochastic dimension of the problem is $d = K + 1 = 869$. This example cannot be solved efficiently using standard CE or iCE, since a large number of effective samples per level are required for fitting high-dimensional parametric densities. Therefore, to compare the results of iCEred, we estimate the failure probability by an average of 100 independent runs of SuS (cf., subsection 3.2.3) using $N = 3000$ samples per level.

We apply the iCEred method with number of samples per level selected from the set $N \in \{100, 250, 500\}$. Table 6.3 shows the results as an average of 40 independent simulations. Note that the coefficient of variation of the failure probabilities are close to the target $\bar{\delta}$. This is also reflected in the averaged number of LSF and gradient calls, denoted respectively as g_{call} and $(\nabla g)_{\text{call}}$. Both number of evaluations are the same when refinement is not required, and

the values are closer to each other when number of required iterations in the refinement step is small. As seen also in the previous examples, the value of g_{call} increases considerably at a small value of N in order to match $\bar{\delta}$. However, there is a trade-off between the computational cost of the extra LSF and the LSF gradient calls. If the gradient is expensive to evaluate, a small number of samples is recommended since the value of the probability estimate is later improved by the refinement. On the contrary, if an efficient way to compute the LSF gradient is available, the extra LSF evaluations in the refinement might exceed the cost of the gradient computations; in this case using a large sample size is recommended. Note also in Table 6.3 that the mean value of the probability of failure computed by repeated runs of SuS is close to the values estimated by repeated runs of iCEred, whereas the coefficient of variation of the SuS estimate is significantly higher than the one of iCEred. Moreover, the iCEred method requires considerably fewer LSF evaluations, and thus can be much more efficient than SuS provided an effective gradient computation is feasible.

Table 6.3.: Failure probability estimates for the example 6.4.4: the iCEred results are shown as an average of 40 simulations; the SuS estimate is computed as an average of 100 simulations.

iCEred					SuS				
N	g_{call}	$(\nabla g)_{\text{call}}$	$\hat{p}_{\mathcal{F}}$	$\widehat{\text{cv}}(p_{\mathcal{F}})$	N	n_{lv}	g_{call}	$\hat{p}_{\mathcal{F}}$	$\widehat{\text{cv}}(p_{\mathcal{F}})$
100	2396	675	3.57×10^{-6}	0.050					
250	2416	1612	3.60×10^{-6}	0.046	3000	6	1.62×10^4	3.75×10^{-6}	0.215
500	3685	3037	3.62×10^{-6}	0.043					

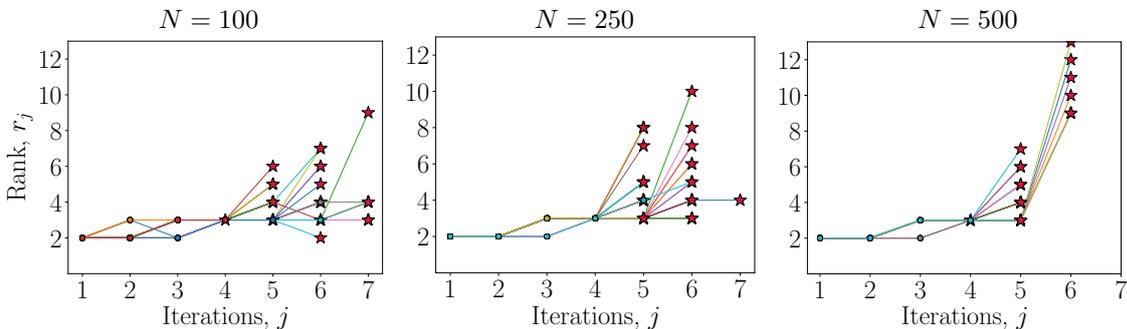


Figure 6.9: Evolution of the rank with the iCEred iterations for several simulation runs and $N \in \{100, 250, 500\}$ (columns). The star marks the final value of the rank at the given simulation run.

We plot in Figure 6.9 the evolution of the rank with the iCEred iterations (intermediate levels). The results are shown for different sample sizes and independent simulation runs. As N increases, the number of iterations becomes smaller; on average, for each of the employed sample sizes, the number of levels are $n_{\text{lv}} = [5.75, 5.45, 5.08]$. For the investigated number of samples per level, the maximum observed ranks across the 40 independent simulations are $r = [8, 9, 12]$; however, the final rank is on average $r = 3$ for $N = 100$, and $r = 4$ for $N = [250, 500]$. We also observe that the value of the smoothing parameters at the final iteration are on average $s_{n_{\text{lv}}} = [0.098, 0.084, 0.051]$, which shows that s reaches values closer to zero for larger sample sizes. Note in (6.35) that the gradient of the log-smooth indicator is essentially driven by the gradient of the LSF; the smoothing parameter indirectly determines the locations at which the gradient is evaluated. In this example, the increment in the rank with the number of intermediate levels seems to be related to the LSF gradient discovering a larger FIS as the samples move towards the failure domain where the nonlinearity of the LSF increases. Moreover, in contrast to the

previous example, $N = 100$ samples yields an appropriate estimate of the probability of failure. In this case, the AS rule gives a required number of samples of $N \approx [54, 271]$, for each of the oversampling factor $\alpha \in \{2, 10\}$ and $r = 4$.

Finally, Figure 6.10 shows the evolution of the probability of failure for increasing terms in the KL expansion (2.44). Different failure probabilities are computed by SuS for truncation orders $K \in \{25, 100, 500, 868\}$; the iCEred method is only evaluated for $K = 868$. In both approaches, the estimators are given as average of 40 independent simulations. Note the difference between the failure probabilities computed by reducing the number of input dimensions (KL terms) and those estimated by identifying the FIS. Solving the rare event simulation problem on the FIS is *not* equivalent to truncating the parameter space, as in the latter case, there is an associated loss of information (e.g., random field variability). The effective dimension of the local FIS in this example is on average $r = 4$ (when using $N = 500$ samples), and thus the iCEred method can compute the failure probability associated with an input dimension of $K = 868$ efficiently and with a very small coefficient of variation.

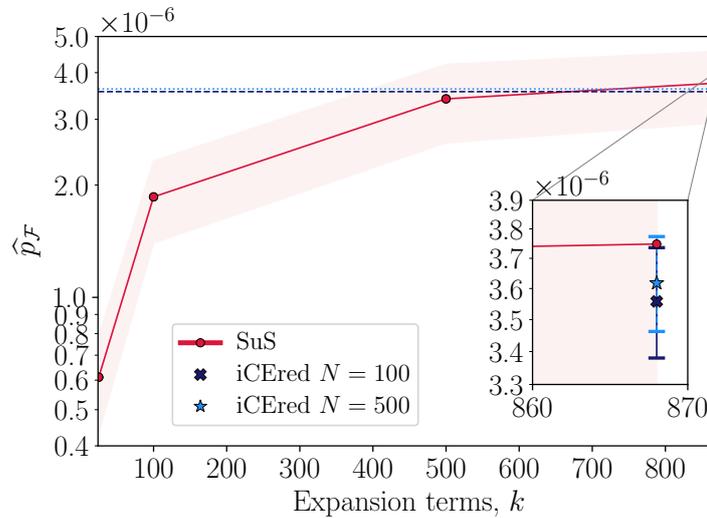


Figure 6.10: Probability of failure estimated by SuS for increasing number of terms in the KL expansion, and probability of failure estimated by iCEred at 868 KL terms.

In this chapter, we discussed a dimension reduction technique that enables the construction of a so-called failure-informed subspace. This subspace allows the definition of a CE-based IS algorithm that can be used to approach high-dimensional reliability problems; especially those involving random field parameters. The method, called iCEred, is able to effectively estimate rare event probabilities in high dimensions with considerably smaller variability and sample size compared to standard simulation approaches (cf. Chapter 3). This applies to cases where an efficient gradient computation is feasible. In other circumstances, there is a trade-off between the computational cost of LSF evaluations and their gradient counterparts. As a result, we also defined a refinement step that can be used to improve the failure probability estimate without further increasing the number of gradient evaluations. We remark that in applications where there is no significant dimension reduction associated to the reliability problem, iCEred can be combined with more general parametric distribution models (see, e.g., [179]).

Conclusions and outlook

“If a man will begin with certainties, he shall end in doubts; but if he will be content to begin with doubts he shall end in certainties”.

Francis Bacon, 1561-1626.

7.1. Conclusions

The uncertainty related to spatially varying parameters of engineering models is represented by random fields, which are discretized into a finite (but typically large) number of random variables. We study the influence of different parameter choices in the prior random field on the updated posterior solution, especially the effect of the number of terms in the random field discretization. The analysis is also performed for the evaluation of rare event probabilities. This initial study motivated the development of two algorithms, one for Bayesian inference and the other for rare event simulation. The methods are particularly useful and efficient to address engineering models involving random fields.

The main contributions of this thesis are listed in the following.

Bayesian analysis of engineering models with random fields

Bayesian inference is a popular framework for identifying spatially varying parameters modeled as random fields. The definition of an appropriate prior for the covariance kernel of the random field is often challenging. An arbitrary choice of the prior distribution and its parameters can have a significant impact on the posterior solution and predictions. To assess the influence of the prior random field model, we investigated two examples, one for which we derived analytical solutions (section 4.5) and another one that is approached numerically (section 4.6). For the latter, we also evaluated the influence of the random field discretization on the estimation of prior and posterior rare event probabilities. The main findings of this study are:

- Oftentimes in engineering applications, the same truncation order in the prior and posterior random field representations is employed (see, e.g., [202]). However, when dealing with Bayesian inverse problems, a larger number of terms in the KL expansion is required for an accurate approximation, compared to the prior random field. The common choice for the KL truncation order to retain a given percentage (typically the 90% or 95%) of the spatially averaged variability of the prior random field is not generally sufficient to obtain the same error levels in the posterior solution.
- We observe that after including the observations, the reduction in the uncertainty is more significant for random field models that use priors with smooth correlation kernels, such as the squared exponential (cf. Figure 4.6). The smoother the kernel, the more correlated the spatial points are, and thus the information impacts the inference in a more global manner. However, one should be careful when using those type of priors since the posterior

fields can be too smooth to explain the observed data (cf. Figure 4.9). In general, the application non-smooth kernels for prior modeling generate more conservative results. This is also seen in the values of the failure probabilities, which are larger compared to those resulting from utilizing smooth kernels (cf. Table 4.1).

- The solution of Bayesian inverse problems is highly influenced by the smoothing action of the forward operator. The smoothing effect appears oftentimes in inverse problems where the forward operator involves integration of the inverse quantity, and it can be explained by the Riemann-Lebesgue lemma (see, e.g., [107, p.8]). Specifically, QoIs that are computed by integration of the forward model operator yield smaller errors in the mean and variance of the random field, as compared to those obtained for QoIs depending locally on the fluctuations of the posterior random field. This information can support the selection of the QoIs that are relevant for the analysis (e.g., when performing posterior predictions, such as the estimation of rare events), in order to avoid unnecessary computations when solving the Bayesian inverse problem. Moreover, the smoothing effect of the forward operator typically induces a low-dimensional structure on the input parameter space, which can be exploited to improve the efficiency of standard Bayesian inversion algorithms (see, e.g., [218, 53, 57, 253]).
- An additional effect of the Bayesian inversion is the modification of the homogeneous prior random field structure. The resulting posterior fields are no longer homogeneous, since the posterior mean is variable in space and the covariance is anisotropic (cf. Figure 4.6). The KL modes associated to the posterior non-homogeneous random fields decay more slowly, and hence, more terms in the expansion are required than in the prior (homogeneous) fields.
- We derived a closed-form expression for the model evidence in the context of the KL expansion of random field (subsection 4.5.3). This quantity reaches a *plateau* at a relatively small number of KL terms (for our numerical experiments) and it does not increase substantially with further addition of terms. The model choice/selection solution is generally not sufficient to represent the whole variability of the posterior random field. For instance, Figure 4.11 shows that for the smaller correlation length one requires at least 30 terms in the KL expansion to explain 90% of the posterior variability. If one performs model selection (Figure 4.12), only 8 terms in the expansion are enough to reach the maximum value of the model evidence. However, this solution is only able to represent 60% of the variability. A better strategy when dealing with random fields is to perform model mixing/averaging, in which one employs all the collection of models for future decisions.
- At the prior level, the probability of failure increases with the number of terms in the KL expansion. This is due to the investigated type of failure mode in the example of section 4.6, which is local in the sense that the σ_1 stress is highly dependent on the local spatial fluctuations of the Young's modulus field. In this case, increasing the KL terms contributes to a better representation the stress field variability. However, if the failure mode is global, e.g., LSF based on displacements, only a few terms are required for representing the variability of the displacement field, and as a result the probability of failure will remain almost constant with increasing KL terms (see also the study in [89]). Nevertheless, as the truncation order tends to infinity, the probabilities converge to a target value corresponding to the probability of failure defined by the fully-represented random field. Since in practice it is not possible to account for the total uncertainty of the field, one is typically estimating a bound of the target failure probability when applying the KL discretization. This is because the approximate variance from the truncated KL

expansion is always smaller than the exact variance, underestimating the variability of the random field.

- At the posterior level, the example in section 4.6 shows that the probability of failure decreases due to the inclusion of information. This is not always the case, e.g., if the measurements indicate a much weaker material than assumed in the prior, the posterior probability of failure can increase. The values of the updated/posterior probabilities depend on the quality of the information, the accuracy of the random field representation, the LSF and the numerical method used for the estimation. Due to the two-step approach that is used in the method presented in section 4.4, the variability of the posterior probability estimates generated by the SuS algorithm is large; especially when the likelihood function is strongly peaked and the failure probabilities are very small. In those cases, the sequential exploration of small probability regions in the posterior requires several intermediate levels, which accumulates the variability of the updated probability estimate. Hence, either a different sampling strategy, or an extension of the SuS method is required when dealing with the updating of rare event probabilities involving random fields.

Dimension selection and averaging: trans-dimensional BUS

We introduced a sequential methodology for the solution of Bayesian model choice problems in section 5.3. The method (termed tBUS-SuS) is a variable-dimensional extension of the classical BUS formulation when used in combination with the SuS algorithm. Our applications involved problems where the KL expansion is used to represent random fields. Hence, the truncation/dimension of the KL discretization is also a part of the inference. The main ideas behind our method are: (i) the re-definition of the parameter space to account for the discrete dimension random variable, and (ii) the application of trans-dimensional MCMC algorithms to sample the resulting sequence of intermediate discrete-continuous distributions. In the first case, we proposed a dimension prior that penalizes increasing model complexity. In the second case, two MCMC algorithms that work in a saturated space where the dimensions are fixed were discussed. Particularly, we derived a step-wise sampler based on a simplification of the RJMCMC algorithm in the saturated space. The main conclusions from this study are:

- The inclusion of a prior that penalizes models with increased number of parameters is beneficial in the context of the KL expansion. Note that if we employ a uniform prior for the dimension, the individual model evidences define the model posterior. In this case, visits to models with high complexity have the same probability of occurrence despite the contribution of both models is almost the same (cf. Figure 5.4). We use the proposed dimension prior precisely to avoid such situations, since it allows us to emphasize dimensions with the largest model evidence, and at the same time, to alleviate the computational cost since the evaluation of unnecessarily complicated models is significantly reduced. This prior is tailored to random field models parameterized with the KL expansion.
- The choice of the dimension proposal of the MCMC samplers plays an important role in the estimation of the model posterior. In general, we observed that the step-wise sampler is more sensitive to this tuning step. Particularly, if the change from model prior to posterior is minimal, drawing samples from the prior itself gives the best results. However, using a proposal matrix to control the dimension moves is beneficial when the changes in the prior-posterior update are significant. Moreover, from the examples in section 5.4, we see that using a jump length between 10% and 50% of the maximum truncation order yields good results.

- We found that in the context of random fields, the model choice solution is not guaranteed to produce the best representation. It can be seen as a “point” estimate of the dimension posterior which in some cases can fail to approximate the variability of the underlying random field. As a result, we recommend the model mixing solution (or a subset of it) in this type of applications.
- The computational cost of tBUS-SuS seems elevated since a large number of samples per level N is required (cf. Figure 5.10). However, this cost is small compared to those of standard within-model simulation approaches, trans-dimensional MCMC algorithms that are not used in combination with a sequential approach, or other sequential algorithms used to solve Bayesian model choice problems. For instance, the work in [9] uses $N = 5 \times 10^4$ particles per level, and in [123] the algorithm is executed for $N = 10^6$ states, storing every 500-th sample.
- The proposed step-wise sampler is a simplification of the RJMCMC sampler in the saturated or composite space. This produces a tBUS-SuS algorithm that is more “automatic” in its definition, since the tuning of a jumping function is not required. In more general model choice situations (other than nested/variable-selection models), the inclusion of a jumping function can be necessary to satisfy the detailed balance condition, and to improve the efficiency of the algorithm.
- The variability in the dimension posterior estimate is larger when employing the step-wise sampler. However, recall that MwG evaluates the joint posterior in two levels, one at which we fix the parameters and sample the dimension, and another one for which we fix the dimension and sample the parameters. This mechanism requires twice the number of PDE model evaluations as compared to the step-wise algorithm, which uses a single joint acceptance step. We have seen that under the same computational cost, both algorithms have comparable performance.

Dimension reduction: failure-informed cross-entropy-based importance sampling

We developed a computational framework for solving high-dimensional and numerically intensive rare event simulation problems. This includes settings where the underlying system properties are spatially variable and random field models are required for their representation. The approach (termed iCEred) is an extension of importance sampling with the CE method that enhances its efficiency in high dimensions. The main idea is to adapt dimension reduction techniques developed for Bayesian inference through expressing the rare event simulation problem as a Bayesian inversion problem. This enables the identification of the potential low-dimensional structure of the rare event simulation problem. As a result, we construct a so-called failure-informed subspace (FIS), where efficient low-dimensional biasing distributions can be defined. This process ultimately requires the evaluation of the LSF gradient. From this study, we can outline the following:

- The iCEred method is able to effectively compute rare event probability estimators in high dimensions that have a considerably smaller variability compared to the standard approaches that typically rely on MCMC simulation. This is also true, even in cases where a considerably small number of samples per level is employed. Although the sample size will depend on the effective rank of the projector.
- As the rank increases, more samples per level are required to obtain a good fitting of the single Gaussian parametric densities. Hence, in applications where there is no significant dimension reduction, iCEred can be coupled with more general parametric distribution models, such as the von Mises–Fisher–Nakagami mixture [179].

- iCEred applies mostly to cases where an efficient gradient computation is feasible. In other circumstances, there is a trade-off between the computational cost of LSF evaluations and their gradient counterparts. The suggested refinement step helps to address these situations by improving the failure probability estimate without performing extra gradient calls.
- The rank of the projectors that map the parameters to the FIS increases as one approaches the final iterations of the algorithm when the LSF is highly nonlinear. Note in (6.35) that the gradient of the log-smooth indicator is essentially driven by the gradient of the LSF; the smoothing parameter indirectly determines the locations at which the gradient is evaluated. The increment in the rank with the number of intermediate levels seems to be related to the LSF gradient discovering a larger FIS, as the samples move towards the failure domain where the potential nonlinearity of the LSF increases.
- In iCEred, the tolerance value ε is used to control the KLD between the exact and approximated optimal biasing densities. However, if we instead choose to bound the MSE, the values of the rank in the plate example of subsection 6.4.4 are half in magnitude. It might be sufficient for the iCEred method to define the controlled approximation in terms of the MSE instead of the KLD. This is because the main objective in rare event simulation is to provide an estimate for the failure probability (normalizing constant) and not the reconstruction of the optimal biasing distribution itself. Further studies are required to formally analyze the approximation in this case. Specifically, an investigation on the user-defined tolerance ε is necessary to assess its influence on the performance of iCEred and the accuracy of the probability estimate.
- When dealing with problems that have multiple failure points (multi-modal failure hypersurface), mixture distributions can also be adapted within the iCEred framework.

7.2. Outlook

The following is a list of ideas that could inspire new research.

Bayesian analysis of random fields

The analysis of random fields requires the specification of several parameters, e.g., the smoothing parameter and correlation length of the Matérn covariance kernel, the dimension of the discretization, and the mean and variance of the field. Such quantities can also be modeled in a probabilistic manner. Our focus was on the dimension of the random field discretization in the context of the KL expansion. However, a fully hierarchical framework that imposes priors on each of random field parameters is still to be developed. Such a robust Bayesian setting can be addressed by the careful combination and adaptation of several existing methodologies (as those described in the Introduction, e.g., [161, 220, 142, 194]). The tBUS methodology can be combined with those approaches to further expand the hierarchies in the Bayesian inversion.

The computational machinery needs to be further developed for problems dealing with the updating of rare events in the context of random fields, i.e., exploring tails of high-dimensional posterior distributions. We have observed that the KL modes of the posterior covariance operator are almost the same as those of the prior covariance, for large components of the expansion. Dimension reduction techniques that exist in Bayesian inverse problems (section 6.1) can be used to find this potential low-dimensional structure. Thereafter, the MCMC algorithm within SuS can be replaced for instance, by the DILI (dimension-independent likelihood-informed) algorithm developed in [58]. Such application can potentially increase the efficiency of the SuS method and improve its accuracy.

In order to perform a particular UQ task (whether is forward propagation or inversion), further studies can address the following questions: (i) in which sense the optimal approximations should be defined, and (ii) how precise the resulting optimal representation needs to be. In this direction, exploring the relation between the error metrics (2.33), which are standard in engineering practice, and theoretical bounds on probability distances, such as those in [91, 225, 147, 115, 117], can be relevant to define optimal decision tools in the specification of random field approximations. This holds in particular when one requires a precise representation of the distribution tails, and not only of the mean and variance. Recent developments in goal-oriented UQ can be used in this case.

Improvements and extensions for tBUS

One possible improvement of tBUS is to use the covering constants $\{\bar{c}_1, \dots, \bar{c}_{k_{\max}}\}$ of each particular dimension in the LSF (5.24), instead of the single constant $\bar{r} = \max(\bar{c}_k)_{k=1}^{k_{\max}}$. The value \bar{r} can be considerably larger than the magnitude of the maximum log-likelihood at the lower dimensions, inducing a loss of sampling efficiency in those dimensions. One can design the algorithm such that one keeps track of the dimensions that are being evaluated, and define the LSF (5.26) as a function of the individual \bar{c}_k . This will also require the definition of multiple intermediate levels, since every dimension variable defines a particular posterior domain.

The exploration of more general model choice settings is still to be addressed by tBUS. This includes cases where the uncertain parameters associated to different dimensions have a complete different meaning. For example, in cases where several PDE models are considered to represent a given physical process. The saturated space methods discussed in subsection 5.2.2 are potentially no longer valid when there is no nested structure in the models.

Improvements and extensions for iCEred

The FIS concept provides a new perspective in high-dimensional rare event simulation. Different simulation algorithms, such as [180, 241], can be adapted to work on the FIS in order to improve their performance and accuracy in high dimensions.

Furthermore, several improvements and adaptations can be performed on iCEred. The following suggestions could potentially save significant computational demands related to gradient evaluations in iCEred:

- a clear direction is to employ gradient-free methods to construct the FIS — partial least squares [178] might be useful to accomplish such task;
- another idea is to define a measure to assess whether or not the second moment matrix of the gradient of the log-smooth indicator function needs to be computed at a given iCEred level. Since the main interest is the failure probability computation and not the construction of the optimal biasing distribution itself, it might be sufficient to build the FIS at the first one or two iterations, and continue with the standard methodology in the remainder levels. This requires a study on the influence of the user-defined tolerance ε ; and
- a potential extension is to employ two different sample sizes, one for the LSF evaluations and another one for the gradient computations. The number of samples per level required to evaluate the gradient can be adapted based on the value of the rank at the given intermediate level.

Finally, the iCEred method can also be adapted to perform reliability updating. One can for instance employ the standard CDR approach [253], which uses SIS to obtain samples

from the posterior parameters, and then employ iCERed to estimate the posterior failure probability. Alternatively, one can utilize the BUS formulation to solve the inverse problem, but in combination with the iCERed method. This requires evaluating the gradient of the LSF $h_{\ln}(\boldsymbol{\vartheta})$ in (4.29)

$$\frac{\partial h_{\ln}(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} = \left[\frac{\partial h_{\ln}(\boldsymbol{\vartheta})}{\partial \theta_v}, \frac{\partial h_{\ln}(\boldsymbol{\vartheta})}{\partial \boldsymbol{\theta}} \right]^T, \quad (7.1)$$

where each component is given by:

$$\frac{\partial h_{\ln}(\boldsymbol{\vartheta})}{\partial \theta_v} = \frac{\partial}{\partial \theta_v} \ln(\Phi(\theta_v)) = \sqrt{\frac{2}{\pi}} \frac{\exp\left(-\frac{\theta_v^2}{2}\right)}{\operatorname{erf}\left(\frac{\sqrt{2}\theta_v}{2}\right) + 1}, \quad \frac{\partial h_{\ln}(\boldsymbol{\vartheta})}{\partial \boldsymbol{\theta}} = -\frac{\partial}{\partial \boldsymbol{\theta}} \ln L(\boldsymbol{\theta}; \tilde{\mathbf{y}}). \quad (7.2)$$

Ultimately, one obtains a framework that fully works on the underlying low-dimensional subspaces and that requires the gradients of both, the likelihood function for the Bayesian inversion and the LSF for the reliability updating.

Fundamentals on probability theory

The properties and characterization of random fields follow from the theory of stochastic processes, which is rooted in the study of probability spaces and random variables. In this Appendix, we review some introductory concepts on probability theory.

A.1. Setting a probabilistic framework

We define a probability space [137] as the triple $(\Omega, \mathcal{F}, \mathbb{P})$, where

- *the sample space* Ω is a non-empty set containing the outcomes of a random experiment (elementary events),
- *the σ -algebra* \mathcal{F} is a collection of subsets of Ω , satisfying: (i) \mathcal{F} contains the empty set and the sample space, i.e., $\emptyset \in \mathcal{F}$ and $\Omega \in \mathcal{F}$, (ii) $A \in \mathcal{F}$ implies that $A^c \in \mathcal{F}$, and (iii) $A_1, A_2, \dots \in \mathcal{F}$ implies that $\cup_{i=1}^{\infty} A_i \in \mathcal{F}$ (this also holds for countable intersections). The elements of \mathcal{F} are called *events* (or measurable sets), and
- *the probability measure* \mathbb{P} is a mapping $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$, such that (i) \mathbb{P} is real and non-negative, (ii) \mathbb{P} is σ -additive, i.e., $\mathbb{P}[\cup_{i=1}^{\infty} A_i] = \sum_{i=1}^{\infty} \mathbb{P}[A_i]$, for mutually disjoint events A_i , and (iii) $\mathbb{P}[\emptyset] = 0$ and $\mathbb{P}[\Omega] = 1$.

Remark A.1. There are some practicalities underlying the previous definition. For instance, it is important for \mathcal{F} to be closed under set operations, since this allows us to take limits and define asymptotic expansions when applying probability theory [195]. Moreover, \mathcal{F} must be a σ -algebra since we are only interested in “nice” events, meaning those that can be measured (so-called *Borel sets*¹). Non-measurable sets typically lead to some paradoxes in mathematics. Therefore, any subset of the sample space not belonging to \mathcal{F} has non-defined probability measure.

Remark A.2. Another point about σ -algebras is that they are strictly necessary only when the sample space is not countable, say all possible subsets of the interval $[0, 1]$. In such cases, the Carathéodory’s extension theorem (see [12, p.19]) enables one to define valid probability spaces. As a result, one can specify for example the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where $\Omega = [0, 1]$, \mathcal{F} contains all intervals in $[0, 1]$, and for any interval $I \in \mathcal{F}$ the probability measure $\mathbb{P}[I]$ is equal to the length of the interval. This triple defines a *uniform distribution* on $[0, 1]$, or a so-called *Lebesgue measure* on $[0, 1]$ (see [195, p.15] for a formal construction).

Probability theory is mainly concerned with the study of random variables. Intuitively, a random variable is a quantity that is measured in a random experiment. Formally, given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a real-valued *random variable* is a function $X : \Omega \rightarrow \mathbb{R}$, such that $\{X(\omega) \leq x\} \in \mathcal{F}$, with $\omega \in \Omega$ and $x \in \mathbb{R}$ (i.e., random variables are \mathcal{F} -measurable or just measurable functions).

A class of random variables which have finite range and discrete support, are the so-called simple or *discrete*. A simple random variable can be expressed in terms of the series expansion

¹A Borel set is any set in a topological space that can be formed from open/closed sets through the operations of countable union/intersection, and relative complement.

$X = \sum_{i=1}^n x_i \mathbb{1}_{A_i}$, where x_1, x_2, \dots, x_n are elements of its range, the sets $A_i = \{\omega \in \Omega; X(\omega) = x_i\} = X^{-1}(\{x_i\})$ form a finite partition of the sample space Ω , and $\mathbb{1}_{A_i}$ is the indicator function for some event A_i (it is equal to one, if $x_i \in A_i$ and zero otherwise). Random variables are usually described in terms of statistics, the most fundamental being the expected value (or expectation or mean). For a simple random variable X , the *expected value*, denoted $\mathbb{E}[X]$ or μ_X , is defined as [195]

$$\mathbb{E}[X] = \mathbb{E} \left[\sum_{i=1}^n x_i \mathbb{1}_{A_i} \right] = \sum_{i=1}^n x_i \mathbb{P}[A_i], \quad \text{where each } A_i \text{ is a finite partition of } \Omega. \quad (\text{A.1})$$

Consider the random variables X, Y and scalars a, b , then the following statements about the expectation operator hold [195]: (i). The expectation is linear, $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$, (ii). The expectation is order-preserving, if $X \leq Y$, then $\mathbb{E}[X] \leq \mathbb{E}[Y]$, (iii). The expectation follows the triangle inequality, $|\mathbb{E}[X]| \leq \mathbb{E}[|X|]$, (iv). If $f : \mathbb{R} \rightarrow \mathbb{R}$ is any function and X is simple, then $f(X) = \sum_{i=1}^n f(x_i) \mathbb{1}_{A_i}$ is also simple with expected value $\mathbb{E}[f(X)] = \sum_{i=1}^n f(x_i) \mathbb{P}[A_i]$, (v). If X and Y are *independent*, then $\mathbb{E}[XY] = \mathbb{E}[X] \mathbb{E}[Y]$.

Besides the expected value, another useful concept in probability arises by considering $\mathbb{E}[f(X)]$ with $f(X) = (X - \mathbb{E}[X])^2$. This is the second central moment or *variance* of the random variable X , denoted $\mathbb{V}[X]$ or σ_X^2 . This function gives information about the variation of a random variable X and it is defined as $\mathbb{V}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2]$. Some immediate properties of the variance include [195]: (i). $0 \leq \mathbb{V}[X] \leq \mathbb{E}[X^2]$, (ii). $\mathbb{V}[aX + bY] = a^2\mathbb{V}[X] + b^2\mathbb{V}[Y] + 2ab \text{Cov}[X, Y]$, where $\text{Cov}[X, Y] = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)]$ is the *covariance* between X and Y , (iii). If X and Y are *independent*, then $\mathbb{V}[X + Y] = \mathbb{V}[X] + \mathbb{V}[Y]$, (iv). If $\mathbb{V}[X] > 0$ and $\mathbb{V}[Y] > 0$, then the *correlation* between X and Y is $\text{Corr}[X, Y] = \text{Cov}[X, Y] / \sqrt{\mathbb{V}[X] \mathbb{V}[Y]}$, (v). The positive square root $\sigma_X = \sqrt{\mathbb{V}[X]}$ is called the *standard deviation* of X .

In general continuous settings, the expectation is given by integration of the random variable with respect to the associated probability measure:

$$\mathbb{E}[X] := \int_{\Omega} X d\mathbb{P} = \int_{\Omega} X(\omega) \mathbb{P}(d\omega). \quad (\text{A.2})$$

In many cases, the expectation still exists for random variables that are defined in a probability space with a Lebesgue probability measure, despite they are not Riemann integrable (see [195, p.51] for an example). This is because the latter is only applicable to continuous functions, or at least to functions that do not have too many points of discontinuity [138, p.293]. Therefore, definition (A.2) is more flexible since it is applicable to any type of random variable. The expected value in (A.2) can be understood as the Lebesgue integral of the random variable X . If $\mathbb{E}[|X|] < \infty$, X is integrable, and if $\mathbb{E}[|X|^2] < \infty$, we call X square-integrable.

A.2. Distributions of random variables

In probability theory, one is particularly interested in computing probabilities of events associated to a random variable X , defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Such probabilities are specified by the *law* or distribution of X

$$\mu(B) := \mathbb{P}[\omega : X(\omega) \in B] = \mathbb{P}[X \in B] = \mathbb{P}[X^{-1}(B)], \quad (\text{A.3})$$

where $B \in \mathcal{B}$, and \mathcal{B} is the collection of Borel sets on \mathbb{R} . Furthermore, the *cumulative distribution function* (CDF) of X is defined as

$$F_X(x) := \mathbb{P}[\omega : X(\omega) \leq x] = \mathbb{P}[X \leq x], \quad (\text{A.4})$$

the map $F_X : \mathbb{R} \rightarrow [0, 1]$ is increasing ($a \leq b \implies F_X(a) \leq F_X(b)$), right-continuous ($\lim_{x \rightarrow x_0^+} F_X(x) = F_X(x_0)$), and satisfies $\lim_{x \rightarrow -\infty} F_X(x) = 0, \lim_{x \rightarrow \infty} F_X(x) = 1$. There is a pivotal connection between the CDF F_X and the law μ . By definition of the CDF, the probability that X lies in the semi-closed interval $(a, b]$, where $a \leq b$, is [12]

$$\mathbb{P}[\omega : a < X(\omega) \leq b] = \mathbb{P}[a < X \leq b] = F_X(b) - F_X(a) = \mu((a, b]), \quad (\text{A.5})$$

this is because the intervals $(a, b]$ are in particular Borel sets on \mathbb{R} . The measure of arbitrary intervals can also be expressed in terms of the CDF (see [12, p.25]). Due to (A.5), the law μ is completely determined by F_X , and in this case we say that μ is the *Lebesgue-Stieltjes measure* corresponding to F_X (see [12, p.22]).

Remark A.3. It is common in practice to define uncertain quantities via the following statement: “Let X be a random variable with distribution function F_X ”, that is, without any reference to the probability space characterizing the uncertainty. In principle this is not an issue since the main interest is on computing the probability of all events involving X , and of course F_X itself determines the probability law μ , which in turn determines probability values. Despite the nature of the underlying probability space seems to become irrelevant, one still needs to check whether there exist at least one $(\Omega, \mathcal{F}, \mathbb{P})$ on which $X \sim F_X$ can be specified.

Remark A.4. There exists a *canonical way* of constructing a probability space, which is by defining $\Omega = \mathbb{R}$, $\mathcal{F} = \mathcal{B}(\mathbb{R})$, $\mathbb{P} = \mu$ equal to the Lebesgue measure corresponding to F_X , and writing the random variable as $X(\omega) = \omega$, which is the so-called *identity map* [12, p.174]. In most engineering applications, for example when performing simulation, one usually forgets about Ω and \mathcal{F} , and use samples to build probability distributions. However, it is not always clear how to reconstruct the random variable itself, meaning the function that connects the sample space with values on the real numbers, which is ultimately the quantity we want to define. The identity map implicitly fills such void, since elements of the sample space are completely determined by values of the random variable.

One can further specify a property of distribution functions that yield another important function employed in probability theory:

$$F_X(x) = \mathbb{P}[X \leq x] = \mathbb{P}[-\infty \leq X \leq x] = \int_{-\infty}^x \pi_X(x) \lambda(dx), \quad (\text{A.6})$$

where π_X is a positive Lebesgue-integrable function called the *probability density function* (PDF) of X , and λ denotes the Lebesgue measure on \mathbb{R} . The PDF is such that $\int_{-\infty}^{\infty} \pi_X(x) \lambda(dx) = 1$. We remark that for discrete random variables the PDF is called the *probability mass function*, which is a density with respect to the *counting measure* over \mathbb{Z} .

In general, (A.6) also expresses the connection between the PDF and the law μ . Note that one can also write (A.6) in terms of the law as $\mu(B) = \int_B \pi_X(x) \lambda(dx)$. This can also be written as the differentials $d\mu = \pi_X d\lambda$. Therefore, $\pi_X = d\mu/d\lambda$ corresponds to the *Radon-Nikodym derivative* of μ with respect to λ . Since λ is the Lebesgue measure on \mathbb{R} , one usually says that π_X is the density of μ , or analogously of F_X [12]. A classical example is the *Gaussian distribution*, denoted $\mathcal{N}(\mu_X, \sigma_X^2)$ which has probability law

$$\mu(B) = \int_{-\infty}^{\infty} \frac{1}{\sigma_X} \phi\left(\frac{x - \mu_X}{\sigma_X}\right) \mathbb{1}_B(x) \lambda(dx) = \int_B \frac{1}{\sigma_X} \phi\left(\frac{x - \mu_X}{\sigma_X}\right) \lambda(dx), \quad (\text{A.7a})$$

where $\phi(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$ is the standard Gaussian PDF, denoted $\mathcal{N}(0, 1)$. It is common to drop the notation λ and one simply writes $\lambda(dx) = dx$.

The expected value (A.2) can also be defined in terms of the distribution of the random

variable. Consider a random variable $X \sim \mu$. For any Lebesgue-integrable function $g : \mathbb{R} \rightarrow \mathbb{R}$:

$$\mathbb{E}_{\mathbb{P}}[g(X)] = \int_{\Omega} g(X(\omega))\mathbb{P}(d\omega) = \int_{-\infty}^{\infty} g(x)\mu(dx) = \mathbb{E}_{\mu}[g]; \quad (\text{A.8})$$

this result is called the *change of variable theorem* (see [195, p.67]) and shows that distributions completely specify expectations. This means that the expectation of a random variable $g(X)$ with respect to the probability measure \mathbb{P} on Ω , is equal to the expectation of the function g with respect to the distribution μ on \mathbb{R} . Moreover, assuming that μ has a density with respect to the Lebesgue measure then (A.8) becomes

$$\mathbb{E}_{\mu}[g] = \int_{-\infty}^{\infty} g(x)\mu(dx) = \int_{-\infty}^{\infty} g(x)\pi_X(x) dx; \quad (\text{A.9})$$

hence, we can either compute the expectation by integrating the function with respect to the distribution μ as in (A.8), or by integrating the function times the density with respect to the Lebesgue measure λ as in (A.9).

A.3. Conditioning

The previous concepts generalize to settings where we consider collections of random variables. In this case, the canonical probability space takes the form $\Omega = \mathbb{R}^d$, $\mathcal{F} = \mathcal{B}(\mathbb{R}^d)$, and \mathbb{P} requires the extension of Lebesgue measures on \mathbb{R} to \mathbb{R}^d , where d is the number of random variables. We skip many of the details that imply such an extension (and refer to [12] for the details). Nevertheless, we discuss important concepts in probability that arise when dealing with multiple random variables.

Consider a d -dimensional *random vector* on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is the function $\mathbf{X} : \Omega \rightarrow \mathbb{R}^d$. Random vectors can be regarded as d -tuples of random variables, i.e., $\mathbf{X} = [X_1, \dots, X_i, \dots, X_d]$, where each X_i is the projection of \mathbf{X} onto the i -th coordinate space. The joint CDF of the random vector \mathbf{X} is defined analogously from (A.6) as

$$F_{\mathbf{X}}(\mathbf{x}) = \mathbb{P}[\omega : X_i(\omega) \leq x_i, i = 1, \dots, d] = \mathbb{P}[X_1 \leq x_1 \cap X_2 \leq x_2 \cap \dots \cap X_d \leq x_d], \quad (\text{A.10})$$

where $\mathbf{x} = [x_1, x_2, \dots, x_d] \in \mathbb{R}^d$. The associated joint PDF can be obtained as

$$\pi_{\mathbf{X}}(\mathbf{x}) = \frac{\partial^d F_{\mathbf{X}}(x_1, x_2, \dots, x_d)}{\partial x_1 \partial x_2 \dots \partial x_d}. \quad (\text{A.11})$$

Two important classes of functions are obtained by manipulating (A.11). The first ones are the *marginal probability distributions*, which give the probabilities for any of the individual variables without reference to the values of the other ones; and the second ones are the *conditional probability distributions*, which give the probabilities for any grouping of the random variables, conditional on particular values of the remaining ones. The *conditional* PDF between two random variables, say X_j given the occurrence of a value x_i of X_i , is given by

$$\pi_{X_j|X_i}(x_j | x_i) = \frac{\pi_{X_i X_j}(x_i, x_j)}{\pi_{X_i}(x_i)} \quad \text{for any } i, j = 1, \dots, d, \quad (\text{A.12})$$

where each $\pi_{X_i}(x_i)$ represents the *marginal* PDF of the random variable X_i , which is computed by integration (marginalization) of the joint $\pi_{X_i X_j}(x_i, x_j)$ with respect to X_j .

Note that the joint, conditional and marginal PDFs are related as $\pi_{X_i X_j}(x_i, x_j) = \pi_{X_j|X_i}(x_j | x_i) \pi_{X_i}(x_i) = \pi_{X_i|X_j}(x_i | x_j) \pi_{X_j}(x_j)$. The conditional PDF $\pi_{X_j|X_i}(x_j | x_i)$ can be interpreted as a normalized *profile function*. For a fixed $x_i = c$, the function $\pi_{X_i X_j}(c, x_j)$ is a *profile* of the

joint PDF, since it equals the intersection of the surface $\pi_{X_i X_j}(x_i, x_j)$ by the plane $x_i = c$. The conditional density is the distribution of this profile, normalized by $\pi_{X_i}(x_i)$. Figure A.1 illustrates the relation between joint, marginals and conditional distributions, for $d = 2$. Without loss of generality the functions are represented via histograms.

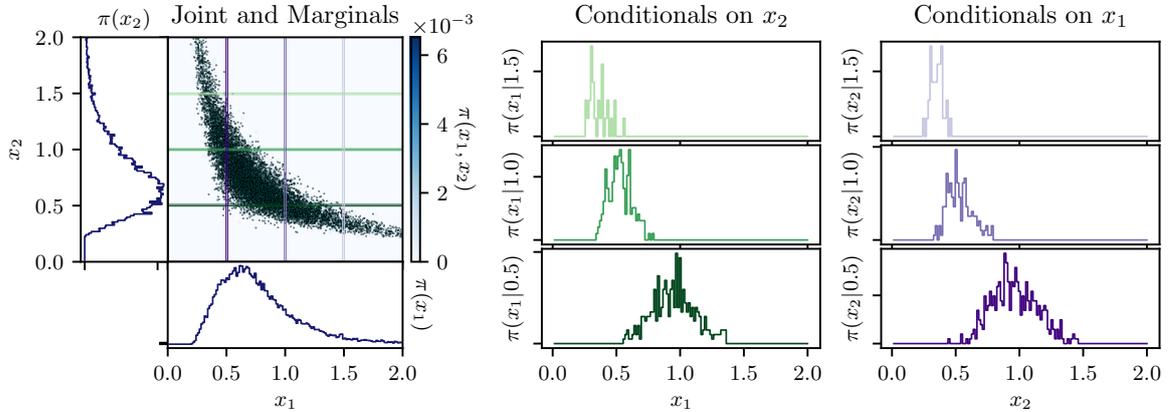


Figure A.1: Joint, marginal and conditional densities. Samples from a banana-shaped distribution are generated to represent the joint PDF. Lines along which the profiles are taken are also shown. The marginals are depicted by histograms at each side of the 2D plot. Three conditionals of x_1 given $x_2 = [0.5, 1.0, 1.5]$ are shown in green. The associated conditionals for x_2 given $x_1 = [0.5, 1.0, 1.5]$ are shown on purple.

Now, consider the Hilbert space of second-order random variables (with finite second moment) $L^2(\Omega, \mathcal{F}, \mathbb{P})$, with the inner product defined via the expectation operator (A.9). For two random variables $X, Y \in L^2(\mathbb{R}^d, \mathbb{P})$, the *conditional expectation* of X given Y is a random variable $Z = \mathbb{E}[X | Y]$, such that for a fixed value $Y(\omega) = y$, we obtain $Z(\omega) = \mathbb{E}[X | Y = y]$ [248].

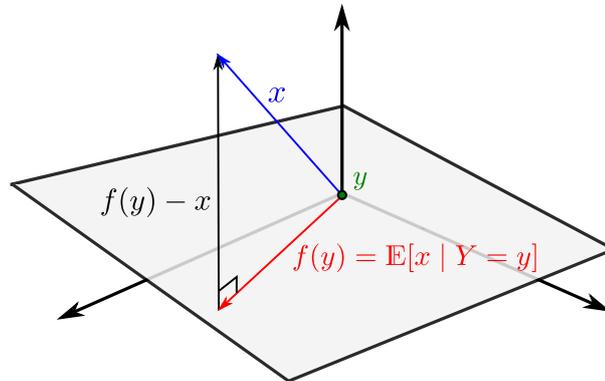


Figure A.2: Simple illustration of the conditional expectation as the best least squares projector in the Hilbert space of second-order random variables.

Conditional expectations are widely used in cases where one needs to estimate a random variable X using observations of another Y . For instance, given Y one can be interested in a function $f(Y)$ that acts as a predictor for X . Typically, the complexity of finding such predictor is reduced by choosing f from a class of functions that minimize the MSE of the predictor. This is the idea in least-squares, where one tries to find the $f(Y)$ that is closest to X in the MSE sense, i.e., such that $\mathbb{E}[(f(Y) - X)^2]$ is minimized. The solution to this problem is an optimal predictor that is equal to the conditional expectation $f(Y) = \mathbb{E}[X | Y]$ [248, p.85]. Therefore, $f(Y)$ is basically the orthogonal projection of X onto the linear subspace functions that can be computed from the available information encoded in Y (cf. Figure A.2; *note that this figure*

motivates the epigraph of Chapter 4). Moreover, for every measurable function g , we can write

$$f(y) = \mathbb{E}[g(X) | Y = y] = \int_{-\infty}^{\infty} g(x) \pi_{X|Y}(x | y) dx, \quad (\text{A.13})$$

one can see that for a fixed $Y = y$, the best approximation of g is its average over all values of X that map to $Y = y$. We remark that the so-called “modern” probability theory is based on the concept of conditional expectation (see fundamental theorem in [248, p.84]).

Finite element method for 2D elasticity problems

We present a brief introduction to some key points on the theory of linear continuum models and the finite element method applied to basic structural systems (this overview follows from [234]). We also discuss the case where the underlying system properties are random and spatially variable; the resulting random field is then incorporated into the finite element formulation. Moreover, the adjoint method used to efficiently compute the gradient of the LSF (6.38) in the example of subsection 6.4.4 is also derived.

B.1. Fundamental equations

In engineering mechanics, the theory of elasticity provides a way to describe the behavior of homogeneous isotropic elastic bodies. This theory formulates a set of PDEs representing the stress, deformation and displacement fields of a structure subjected to certain kind of forces and boundary and initial conditions [41].

Several problems in 3D spatial elasticity can be treated in a 2D domain. This kind of model reduction alleviates the computational cost generated during the simulation of complex structures, without affecting the accuracy of the results. Based on the dimensions of the structure and the arrangement of the applied loads, there exists two general model reduction cases: (i) the *plane stress* model can be used when the thickness of the structure is much smaller compared to the other two dimensions; in this case the structure is only subjected to load forces that act parallel to it (e.g., thin plates); (ii) the *plane strain* model can be used when the length of the structure is much larger compared to the other two dimensions; in this case only uniformly distributed forces act along the length of the structure (e.g., tunnels).

Consider an homogeneous isotropic elastic body in equilibrium under the assumption of plane stress. It is defined by a bounded domain $D \in \mathbb{R}^2$, with thickness t and boundaries Γ_1 and Γ_2 . The solid is fixed along boundary Γ_1 and it is subjected to *volume or body forces* \mathbf{b} , *surface forces* \mathbf{s} and *external point loads* \mathbf{p}_i , $i = 1, \dots, n$, along the boundary Γ_2 (Figure B.1).

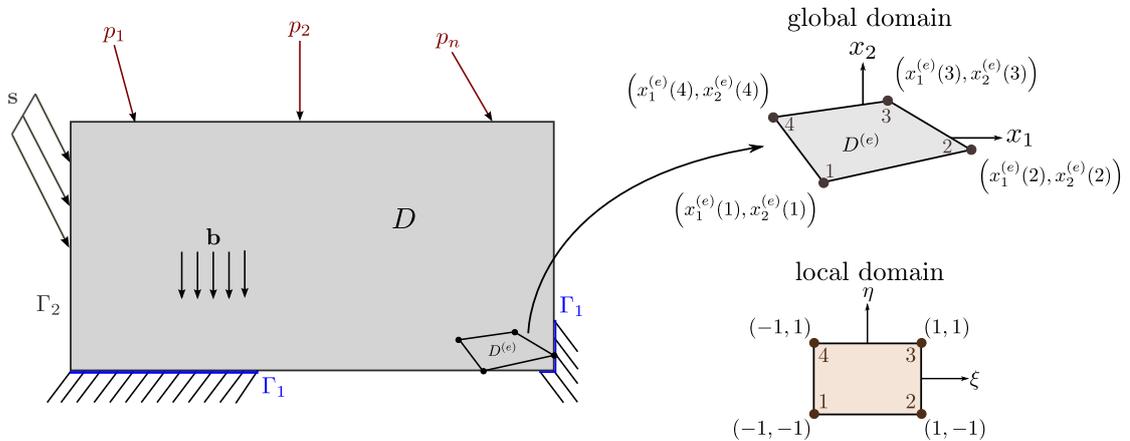


Figure B.1: Boundary and body forces acting on a body in equilibrium. Global and local domains of a four-noded finite element.

Under the forces \mathbf{b} , \mathbf{p} and \mathbf{s} , the objective is to determine the displacement, deformation and stress fields of the structure. For plane stress, such fields are defined as:

- The displacement field is $\mathbf{u}(x_1, x_2) = [u(x_1), v(x_2)]^\top$, with $w(x_3) := 0$,
- the symmetric deformation tensor $\boldsymbol{\varepsilon} = [\varepsilon_{ij}]$ given by:

$$\varepsilon_{11} = \varepsilon_{x_1} = \frac{\partial u}{\partial x_1} \qquad \varepsilon_{12} = \gamma_{x_1 x_2} = \frac{\partial u}{\partial x_2} + \frac{\partial v}{\partial x_1} \qquad (\text{B.1a})$$

$$\varepsilon_{22} = \varepsilon_{x_2} = \frac{\partial v}{\partial x_2} \qquad \varepsilon_{13} = \gamma_{x_1 x_3} = \frac{\partial u}{\partial x_3} + \frac{\partial w}{\partial x_1} := 0 \qquad (\text{B.1b})$$

$$\varepsilon_{33} = \varepsilon_{x_3} = \frac{\partial w}{\partial x_3} \qquad \varepsilon_{23} = \gamma_{x_2 x_3} = \frac{\partial v}{\partial x_3} + \frac{\partial w}{\partial x_2} := 0, \qquad (\text{B.1c})$$

- the symmetric stress tensor $\boldsymbol{\sigma} = [\sigma_{ij}]$, with σ_{ii} representing the *normal stresses* and $\sigma_{ij}, i \neq j$ the *shear stresses*. Assuming that the displacements of the solid are small, the stresses and deformations of an isotropic elastic material can be computed by the *generalized Hooke's law*:

$$\sigma_{11} = \sigma_{x_1} = \frac{E}{1 - \nu^2} (\varepsilon_{x_1} + \nu \varepsilon_{x_2}) \qquad \sigma_{13} = \tau_{x_1 x_3} := 0 \qquad (\text{B.2a})$$

$$\sigma_{12} = \tau_{x_1 x_2} = G \gamma_{x_1 x_2} \qquad \sigma_{33} = \sigma_{x_3} := 0 \qquad (\text{B.2b})$$

$$\sigma_{22} = \sigma_{x_2} = \frac{E}{1 - \nu^2} (\varepsilon_{x_2} + \nu \varepsilon_{x_1}) \qquad \sigma_{23} = \tau_{x_2 x_3} := 0, \qquad (\text{B.2c})$$

which can be expressed in matrix form as

$$\begin{pmatrix} \sigma_{x_1} \\ \sigma_{x_2} \\ \tau_{x_1 x_2} \end{pmatrix} = \underbrace{\frac{E}{1 - \nu^2} \begin{bmatrix} 1 & \nu & 0 \\ \nu & 1 & 0 \\ 0 & 0 & \frac{1-\nu}{2} \end{bmatrix}}_{\mathbf{D}} \begin{pmatrix} \varepsilon_{x_1} \\ \varepsilon_{x_2} \\ \gamma_{x_1 x_2} \end{pmatrix}, \qquad (\text{B.3})$$

where $G := E/2(1 + \nu)$ is the *shear modulus*, E is the *Young's modulus*, ν is the *Poisson ratio* of the elastic solid. The symmetric matrix \mathbf{D} is called the *constitutive matrix*.

The 2D elasticity problem can be formally stated from two different but equivalent viewpoints. In both cases, we are given body forces $\mathbf{b} = [b_{x_1}, b_{x_2}]^\top$, surface forces $\mathbf{s} = [s_{x_1}, s_{x_2}]^\top$ and external point loads $\mathbf{p}_i = [p_{x_1(i)}, p_{x_2(i)}]^\top$.

The first choice is formulated through a set of elliptic PDEs called the *Cauchy–Navier equations* (or elastostatic equations). The task is to find $\mathbf{u}(\mathbf{x}) = \mathbf{u}(x_1, x_2) = [u(x_1), v(x_2)]^\top$ such that,

$$G(\mathbf{x}) \nabla^2 \mathbf{u}(\mathbf{x}) + \frac{E(\mathbf{x})}{2(1 - \nu)} \nabla(\nabla \cdot \mathbf{u}(\mathbf{x})) + \mathbf{b} = 0. \qquad (\text{B.4})$$

Alternatively, one can formulate the problem through the *virtual work principle*. In this case, the task is to find $\mathbf{u} = [u, v]^\top$ such that,

$$t \iint_D \delta \boldsymbol{\varepsilon}^\top \boldsymbol{\sigma} \, dA = t \iint_D \delta \mathbf{u}^\top \mathbf{b} \, dA + t \oint_{\Gamma_2} \delta \mathbf{u}^\top \mathbf{s} \, ds + \sum_{i=1}^n \delta \mathbf{u}_i^\top \mathbf{p}_i, \qquad (\text{B.5})$$

where $\boldsymbol{\sigma} = [\sigma_{x_1}, \sigma_{x_2}, \tau_{x_1 x_2}]^\top$ are the stresses, $\delta \boldsymbol{\varepsilon} = [\delta \varepsilon_{x_1}, \delta \varepsilon_{x_2}, \delta \gamma_{x_1 x_2}]^\top$ are the associated virtual deformation components, and $\delta \mathbf{u} = [\delta u, \delta v]^\top$ are the virtual displacements. It can be shown that (B.4) and (B.5) are equivalent and also that the solution \mathbf{u} is unique [118, 129].

B.2. Finite element formulation

In order to define a suitable finite-dimensional vector space of solutions, the physical domain D of a structure is divided into a set of n_{fe} *elements*:

$$D = \bigcup_{e=1}^{n_{fe}} r_e = r_1 \cup r_2 \cup \dots \cup r_{n_{fe}}, \quad (\text{B.6})$$

which are defined such that they cannot overlap, the element nodal points cannot be located on the side of an adjacent element, and each element nodal point must coincide between the elements.

A number of finite elements are available in the literature. They can be identified by their number of nodes, their shape, the type of interpolation and continuity requirements. In general, finite elements can be characterized by their *shape or test functions*, which must satisfy certain convergence requirements: smoothness over the element interior, continuity across each element boundary, and completeness, so that the element interpolation function is able to represent an arbitrary polynomial at the degrees of freedom (dof) of the element nodes. In the following, we employ the bilinear (four-noded) quadrilateral element to make the explanation more clear.

The isoparametric formulation is applied in order to make use of the same shape interpolation functions of the element coordinates and the displacements. Hence, a point in the local domain (ξ, η) is related to a point (x_1, x_2) in the global domain by mappings of the form

$$x_1(\xi, \eta) = \sum_{i=1}^4 N_i(\xi, \eta) x_1^{(e)}(i) \quad x_2(\xi, \eta) = \sum_{i=1}^4 N_i(\xi, \eta) x_2^{(e)}(i), \quad (\text{B.7})$$

where (ξ, η) are called the *natural coordinates* and N_i are element shape functions. An important relationship between the derivatives of the shape functions with respect to the Cartesian and the natural coordinates is given by the chain rule of derivation

$$\begin{bmatrix} \frac{\partial N_i}{\partial \xi} \\ \frac{\partial N_i}{\partial \eta} \end{bmatrix} = \underbrace{\begin{bmatrix} \frac{\partial x_1}{\partial \xi} & \frac{\partial x_2}{\partial \xi} \\ \frac{\partial x_1}{\partial \eta} & \frac{\partial x_2}{\partial \eta} \end{bmatrix}}_{\mathbf{J}^{(e)}} \begin{bmatrix} \frac{\partial N_i}{\partial x_1} \\ \frac{\partial N_i}{\partial x_2} \end{bmatrix}, \quad (\text{B.8})$$

where $\mathbf{J}^{(e)}$ is the Jacobian matrix of the transformation of the derivatives of the shape functions in the natural and Cartesian axes. Moreover,

$$\begin{bmatrix} \frac{\partial N_i}{\partial x_1} \\ \frac{\partial N_i}{\partial x_2} \end{bmatrix} = [\mathbf{J}^{(e)}]^{-1} \begin{bmatrix} \frac{\partial N_i}{\partial \xi} \\ \frac{\partial N_i}{\partial \eta} \end{bmatrix} = \frac{1}{|\mathbf{J}^{(e)}|} \begin{bmatrix} \frac{\partial x_2}{\partial \eta} & -\frac{\partial x_2}{\partial \xi} \\ -\frac{\partial x_1}{\partial \eta} & \frac{\partial x_1}{\partial \xi} \end{bmatrix} \begin{bmatrix} \frac{\partial N_i}{\partial \xi} \\ \frac{\partial N_i}{\partial \eta} \end{bmatrix}, \quad (\text{B.9})$$

where the determinant of the Jacobian matrix relates the differential of area in the two coordinate systems, i.e., $dx_1 dx_2 = |\mathbf{J}^{(e)}| d\xi d\eta$. In case of the four-noded quadrilateral element, the shape functions are obtained by assuming a bilinear expansion of the form

$$x_1(\xi, \eta) = a_0 + a_1 \xi + a_2 \eta + a_3 \xi \eta \quad (\text{B.10a})$$

$$x_2(\xi, \eta) = b_0 + b_1 \xi + b_2 \eta + b_3 \xi \eta, \quad (\text{B.10b})$$

where the coefficients a 's and b 's need to be determined. The expressions (B.10) satisfy the conditions, $x_1(\xi_i, \eta_i) = x_1^{(e)}(i)$ and $x_2(\xi_i, \eta_i) = x_2^{(e)}(i)$. The values of the (ξ_i, η_i) coordinates in the

local domain are shown in Figure B.1. After imposing these restrictions on (B.7), two different system of equations are obtained that are solved for the coefficients a 's and b 's, respectively. These values are then back-substituted into (B.7) to obtain the corresponding shape functions of the bilinear quadrilateral element:

$$N_1(\xi, \eta) = ((\eta - 1)(\xi - 1))/4 \quad N_2(\xi, \eta) = -((\eta - 1)(\xi + 1))/4 \quad (\text{B.11a})$$

$$N_3(\xi, \eta) = ((\eta + 1)(\xi + 1))/4 \quad N_4(\xi, \eta) = -((\eta + 1)(\xi - 1))/4. \quad (\text{B.11b})$$

B.2.1. Discretization of the displacement field

The displacements can be expressed in the x_1, x_2 directions at the interior of an element e as a function of its nodal displacements as

$$u^{(e)} = u_1^{(e)} N_1^{(e)} + u_2^{(e)} N_2^{(e)} + u_3^{(e)} N_3^{(e)} + u_4^{(e)} N_4^{(e)} \quad (\text{B.12a})$$

$$v^{(e)} = v_1^{(e)} N_1^{(e)} + v_2^{(e)} N_2^{(e)} + v_3^{(e)} N_3^{(e)} + v_4^{(e)} N_4^{(e)}, \quad (\text{B.12b})$$

or in matrix form

$$\underbrace{\begin{bmatrix} u^{(e)} \\ v^{(e)} \end{bmatrix}}_{\mathbf{u}^{(e)}} = \underbrace{\begin{bmatrix} N_1^{(e)} & 0 & \vdots & N_2^{(e)} & 0 & \vdots & N_3^{(e)} & 0 & \vdots & N_4^{(e)} & 0 \\ 0 & N_1^{(e)} & \vdots & 0 & N_2^{(e)} & \vdots & 0 & N_3^{(e)} & \vdots & 0 & N_4^{(e)} \end{bmatrix}}_{\mathbf{N}^{(e)}} \underbrace{\begin{bmatrix} u_1^{(e)} \\ v_1^{(e)} \\ u_2^{(e)} \\ v_2^{(e)} \\ u_3^{(e)} \\ v_3^{(e)} \\ u_4^{(e)} \\ v_4^{(e)} \end{bmatrix}}_{\mathbf{a}^{(e)}}, \quad (\text{B.13})$$

where the matrix $\mathbf{N}^{(e)}$ can be represented in blocks as

$$\mathbf{N}^{(e)} = \begin{bmatrix} \mathbf{N}_1^{(e)} & \mathbf{N}_2^{(e)} & \mathbf{N}_3^{(e)} & \mathbf{N}_4^{(e)} \end{bmatrix}, \quad \text{with } \mathbf{N}_i^{(e)} = \begin{bmatrix} N_i^{(e)} & 0 \\ 0 & N_i^{(e)} \end{bmatrix}; \quad (\text{B.14})$$

here $\mathbf{N}^{(e)}$ is the *shape function matrix* of the element e , and $\mathbf{N}_i^{(e)}$ denotes the shape function matrix of the element e at node i . Analogously, the displacement vector $\mathbf{u}^{(e)}$ can be represented as

$$\mathbf{u}^{(e)} = \begin{bmatrix} \mathbf{u}_1^{(e)} & \mathbf{u}_2^{(e)} & \mathbf{u}_3^{(e)} & \mathbf{u}_4^{(e)} \end{bmatrix}^\top, \quad \text{where } \mathbf{u}_i^{(e)} = \begin{bmatrix} u_i^{(e)} & v_i^{(e)} \end{bmatrix}^\top \quad (\text{B.15})$$

and $\mathbf{u}^{(e)}$ is the *nodal displacement vector* of the element e and $\mathbf{u}_i^{(e)}$ denotes the nodal displacement vector of the element e at node i .

B.2.2. Discretization of the deformation field

From the definition of each component of the deformation field, they can be interpolated as

$$\varepsilon_{x_1}^{(e)} = \frac{\partial u^{(e)}}{\partial x_1} = \frac{\partial N_1^{(e)}}{\partial x_1} u_1^{(e)} + \frac{\partial N_2^{(e)}}{\partial x_1} u_2^{(e)} + \frac{\partial N_3^{(e)}}{\partial x_1} u_3^{(e)} + \frac{\partial N_4^{(e)}}{\partial x_1} u_4^{(e)} \quad (\text{B.16a})$$

$$\varepsilon_{x_2}^{(e)} = \frac{\partial v^{(e)}}{\partial x_2} = \frac{\partial N_1^{(e)}}{\partial x_2} v_1^{(e)} + \frac{\partial N_2^{(e)}}{\partial x_2} v_2^{(e)} + \frac{\partial N_3^{(e)}}{\partial x_2} v_3^{(e)} + \frac{\partial N_4^{(e)}}{\partial x_2} v_4^{(e)} \quad (\text{B.16b})$$

$$\gamma_{x_1 x_2}^{(e)} = \frac{\partial u^{(e)}}{\partial x_2} + \frac{\partial v^{(e)}}{\partial x_1} = \frac{\partial N_1^{(e)}}{\partial x_2} u_1^{(e)} + \frac{\partial N_1^{(e)}}{\partial x_1} v_1^{(e)} + \frac{\partial N_2^{(e)}}{\partial x_2} u_2^{(e)} + \frac{\partial N_2^{(e)}}{\partial x_1} v_2^{(e)} + \quad (\text{B.16c})$$

$$\frac{\partial N_3^{(e)}}{\partial x_2} u_3^{(e)} + \frac{\partial N_3^{(e)}}{\partial x_1} v_3^{(e)} + \frac{\partial N_4^{(e)}}{\partial x_2} u_4^{(e)} + \frac{\partial N_4^{(e)}}{\partial x_1} v_4^{(e)}, \quad (\text{B.16d})$$

which can be written in matrix form as

$$\underbrace{\begin{bmatrix} \frac{\partial u^{(e)}}{\partial x_1} \\ \frac{\partial u^{(e)}}{\partial x_2} \\ \frac{\partial u^{(e)}}{\partial x_2} + \frac{\partial v^{(e)}}{\partial x_1} \end{bmatrix}}_{\boldsymbol{\varepsilon}^{(e)}} = \underbrace{\begin{bmatrix} \frac{\partial N_1^{(e)}}{\partial x_1} & 0 & \vdots & \frac{\partial N_2^{(e)}}{\partial x_1} & 0 & \vdots & \frac{\partial N_3^{(e)}}{\partial x_1} & 0 & \vdots & \frac{\partial N_4^{(e)}}{\partial x_1} & 0 \\ 0 & \frac{\partial N_1^{(e)}}{\partial x_2} & \vdots & 0 & \frac{\partial N_2^{(e)}}{\partial x_2} & \vdots & 0 & \frac{\partial N_3^{(e)}}{\partial x_2} & \vdots & 0 & \frac{\partial N_4^{(e)}}{\partial x_2} \\ \frac{\partial N_1^{(e)}}{\partial x_2} & \frac{\partial N_1^{(e)}}{\partial x_1} & \vdots & \frac{\partial N_2^{(e)}}{\partial x_2} & \frac{\partial N_2^{(e)}}{\partial x_1} & \vdots & \frac{\partial N_3^{(e)}}{\partial x_2} & \frac{\partial N_3^{(e)}}{\partial x_1} & \vdots & \frac{\partial N_4^{(e)}}{\partial x_2} & \frac{\partial N_4^{(e)}}{\partial x_1} \end{bmatrix}}_{\mathbf{B}^{(e)}} \underbrace{\begin{bmatrix} u_1^{(e)} \\ v_1^{(e)} \\ u_2^{(e)} \\ v_2^{(e)} \\ u_3^{(e)} \\ v_3^{(e)} \\ u_4^{(e)} \\ v_4^{(e)} \end{bmatrix}}_{\mathbf{a}^{(e)}};$$

note that the matrix $\mathbf{B}^{(e)}$ can be represented in blocks as

$$\mathbf{B}^{(e)} = \begin{bmatrix} \mathbf{B}_1^{(e)} & \mathbf{B}_2^{(e)} & \mathbf{B}_3^{(e)} & \mathbf{B}_4^{(e)} \end{bmatrix} \quad \text{with} \quad \mathbf{B}_i^{(e)} = \begin{bmatrix} \frac{\partial N_i^{(e)}}{\partial x_1} & 0 \\ 0 & \frac{\partial N_i^{(e)}}{\partial x_2} \\ \frac{\partial N_i^{(e)}}{\partial x_2} & \frac{\partial N_i^{(e)}}{\partial x_1} \end{bmatrix}; \quad (\text{B.17})$$

here $\mathbf{B}^{(e)}$ is the *deformation matrix* of the element e and $\mathbf{B}_i^{(e)}$ denotes the deformation matrix of the element e at node i .

B.2.3. Discretization of the stress field

From Hooke's law, the stress field at the element level is given by [214]

$$\boldsymbol{\sigma}^{(e)} = \mathbf{D}\boldsymbol{\varepsilon}^{(e)} = \mathbf{D}\mathbf{B}^{(e)}\mathbf{u}^{(e)}. \quad (\text{B.18})$$

This relation represents the discretization of the stress field inside a given element. However, since the constitutive matrix \mathbf{D} and the deformation matrix $\mathbf{B}^{(e)}$ are constant, the deformations and stresses of an element are also constant in the whole element. This requires the use of a refined mesh in order to obtain a better approximation of the deformation and stresses of the solid, especially in regions of high stress gradients.

At each point in an elastic body there exist principal planes where the stress vector is perpendicular [214, 41]. The associated stresses normal to these principal planes in 2D are given by

$$\sigma_1 = \frac{1}{2}(\sigma_{x_1} + \sigma_{x_2}) + \sqrt{\left(\frac{1}{2}(\sigma_{x_1} - \sigma_{x_2})\right)^2 + \tau_{x_1 x_2}^2} \quad (\text{B.19a})$$

$$\sigma_2 = \frac{1}{2}(\sigma_{x_1} + \sigma_{x_2}) - \sqrt{\left(\frac{1}{2}(\sigma_{x_1} - \sigma_{x_2})\right)^2 + \tau_{x_1 x_2}^2}. \quad (\text{B.19b})$$

In practice, one is especially interested in the principal stresses and their directions, since these are directly related to the maximum allowed critical stress of the elastic body.

B.2.4. Stiffness matrix and force vector

The equilibrium of forces acting on an element is only enforced at the nodes. Therefore, *equilibrium nodal forces* $\mathbf{q}_i = [q_{x_i}, q_{y_i}]^T$, which balance the external and internal forces due to the element deformation, can be obtained by applying the virtual work principle (B.5) to an individual element. For a four-noded element with domain $D^{(e)}$:

$$t \iint_{D^{(e)}} \delta \boldsymbol{\varepsilon}^T \boldsymbol{\sigma} \, dA - t \iint_{D^{(e)}} \delta \mathbf{u}^T \mathbf{b} \, dA - t \oint_{\Gamma_2^{(e)}} \delta \mathbf{u}^T \mathbf{s} \, ds - [\delta \mathbf{u}^{(e)}]^T \mathbf{p}^{(e)} = [\delta \mathbf{u}^{(e)}]^T \mathbf{q}^{(e)},$$

where $\delta \mathbf{u}^{(e)} = [\delta u_1^{(e)}, \delta v_1^{(e)}, \delta u_2^{(e)}, \delta v_2^{(e)}, \delta u_3^{(e)}, \delta v_3^{(e)}, \delta u_4^{(e)}, \delta v_4^{(e)}]^T$ are the virtual nodal displacements of the element, and $\mathbf{q}^{(e)} = [q_{x_1(1)}, q_{x_2(1)}, q_{x_1(2)}, q_{x_2(2)}, q_{x_1(3)}, q_{x_2(3)}, q_{x_1(4)}, q_{x_2(4)}]^T$ are the equilibrium nodal forces of the element.

Using the discretization of the displacement and deformation fields, $\delta \mathbf{u}^T = [\delta \mathbf{u}^{(e)}]^T \mathbf{N}^T$ and $\delta \boldsymbol{\varepsilon}^T = [\delta \mathbf{u}^{(e)}]^T \mathbf{B}^T$. Then, (B.20) can be expressed as

$$[\delta \mathbf{u}^{(e)}]^T \left(t \iint_{D^{(e)}} \mathbf{B}^T \boldsymbol{\sigma} \, dA - t \iint_{D^{(e)}} \mathbf{N}^T \mathbf{b} \, dA - t \oint_{\Gamma_2^{(e)}} \mathbf{N}^T \mathbf{s} \, ds - \mathbf{p}^{(e)} \right) = \mathbf{q}^{(e)}. \quad (\text{B.20})$$

Since by definition the functional $[\delta \mathbf{u}]^T$ represents the field of all possible displacements, it cannot be equal to zero. Thus, the residual in (B.20) holds if

$$t \iint_{D^{(e)}} \mathbf{B}^T \boldsymbol{\sigma} \, dA - t \iint_{D^{(e)}} \mathbf{N}^T \mathbf{b} \, dA - t \oint_{\Gamma_2^{(e)}} \mathbf{N}^T \mathbf{s} \, ds - \mathbf{p}^{(e)} = \mathbf{q}^{(e)}. \quad (\text{B.21})$$

Equation (B.21) expresses the equilibrium between the element deformation forces, element volume forces, element distributed forces, element point loads and element nodal equilibrium forces. Substituting the expression for the stress field (B.18) into (B.21) yields

$$t \iint_{D^{(e)}} \mathbf{B}^T (\mathbf{D} \mathbf{B} \mathbf{u}^{(e)}) \, dA - t \iint_{D^{(e)}} \mathbf{N}^T \mathbf{b} \, dA - t \oint_{\Gamma_2^{(e)}} \mathbf{N}^T \mathbf{s} \, ds - \mathbf{p}^{(e)} = \mathbf{q}^{(e)}, \quad (\text{B.22})$$

which is equivalent to

$$\left(t \iint_{D^{(e)}} \mathbf{B}^T \mathbf{D} \mathbf{B} \, dA \right) \mathbf{u}^{(e)} - t \iint_{D^{(e)}} \mathbf{N}^T \mathbf{b} \, dA - t \oint_{\Gamma_2^{(e)}} \mathbf{N}^T \mathbf{s} \, ds - \mathbf{p}^{(e)} = \mathbf{q}^{(e)}. \quad (\text{B.23})$$

Equation (B.23) can be expressed in matrix form as,

$$\mathbf{K}^{(e)} \mathbf{u}^{(e)} - \mathbf{f}^{(e)} = \mathbf{q}^{(e)}, \quad (\text{B.24})$$

where the so-called element *stiffness matrix* is given by

$$\mathbf{K}^{(e)} = t \iint_{D^{(e)}} \mathbf{B}^T \mathbf{D} \mathbf{B} \, dA \quad (\text{B.25})$$

and the element equivalent nodal *force vector* by

$$\mathbf{f}^{(e)} = t \iint_{D^{(e)}} \mathbf{N}^T \mathbf{b} \, dA + t \oint_{\Gamma_2} \mathbf{N}^T \mathbf{s} \, ds + \mathbf{p}^{(e)}. \quad (\text{B.26})$$

Numerical integration quadratures are used for the computation of integrals (B.25) and (B.26). A coordinate transformation is performed such that the domain of integration is changed

from Cartesian coordinates to natural coordinates (as shown in Figure B.1). In this way, the integration can be performed over the region $[-1, 1] \times [-1, 1]$. Hence, (B.25) expressed as

$$\mathbf{K}^{(e)} = t \int_{-1}^1 \int_{-1}^1 \mathbf{B}^\top(\xi, \eta) \mathbf{D} \mathbf{B}(\xi, \eta) \left| \mathbf{J}^{(e)}(\xi, \eta) \right| d\xi d\eta, \quad (\text{B.27})$$

where $\left| \mathbf{J}^{(e)} \right|$ is the determinant of the Jacobian matrix, which arises from the coordinate transformation (B.7). Therefore, the element stiffness matrix in (B.27) can be evaluated using a GL quadrature with n_{GP} points

$$\mathbf{K}^{(e)} = \sum_{p=1}^{n_{\text{GP}}} \sum_{q=1}^{n_{\text{GP}}} \mathbf{B}^\top(\xi_p, \eta_q) \mathbf{D} \mathbf{B}(\xi_p, \eta_q) t \left| \mathbf{J}^{(e)}(\xi_p, \eta_q) \right| w_p w_q, \quad (\text{B.28})$$

where w_p and w_q are the weights of quadrature rule and the points ξ_p and η_q are the associated abscissas of the Legendre polynomial. The quadrature points n_{GP} are chosen based on the fact that a n_{GP} -order quadrature integrates exactly a polynomial of degree $2n_{\text{GP}} - 1$ or less [118]. An analogous procedure applies to the force vector components in (B.26). Finally, after each element stiffness and force components are computed, the contributions are assembled in order to obtain the global stiffness matrix and force vector [118]. This assembly procedure leads to the well-known system of equations

$$\mathbf{K} \mathbf{u} - \mathbf{f} = \mathbf{q}. \quad (\text{B.29})$$

B.2.5. The random field case: spatially variable Young's modulus

Oftentimes, material properties of structural system are subjected to spatial variation. If the Young's modulus is spatially variable and modeled by a random field, one can employ the KL expansion to represent the associated uncertainty. In this case, the constitutive matrix is no longer constant through the finite element formulation and it will depend on the values of the Young's modulus $E(\boldsymbol{\theta})$. Under the plane stress assumption, the resulting matrix is written as

$$\mathbf{D}(\boldsymbol{\theta}) = \frac{E(\boldsymbol{\theta})}{1 - \nu^2} \begin{bmatrix} 1 & \nu & 0 \\ \nu & 1 & 0 \\ 0 & 0 & \frac{(1-\nu)}{2} \end{bmatrix} = E(\boldsymbol{\theta}) \mathbf{D}_0, \quad (\text{B.30})$$

where $\mathbf{D}_0 \in \mathbb{R}^{3 \times 3}$ is a constant matrix.

The KL expansion is incorporated in the finite element solution, such that the eigenfunctions of the covariance operator are interpolated at the GL points of the finite element mesh using Nyström formula [186]. Therefore, for a n_{GP} -point spatial discretization, the KL representation of the Young's modulus truncated at the k -th term is expressed in matrix form as

$$\widehat{\mathbf{E}} = \exp \left[\mu_{E'} + \mathbf{A} \boldsymbol{\theta}^{(\text{KL})} \right], \quad (\text{B.31})$$

where $\mu_{E'}$ is the mean of the underlying Gaussian random field, $\Phi \mathbf{\Lambda} = \mathbf{A} \in \mathbb{R}^{n_{\text{GP}} \times k}$, $\mathbf{\Lambda} = \text{diag}(\sqrt{\boldsymbol{\lambda}}) \in \mathbb{R}^{k \times k}$ is a diagonal matrix with the square root of the eigenvalues of the covariance operator, $\Phi \in \mathbb{R}^{n_{\text{GP}} \times k}$ is a matrix containing the associated eigenfunctions evaluated at the GL points, and $\boldsymbol{\theta}^{(\text{KL})} \in \mathbb{R}^k$ is the standard Gaussian random vector of KL coefficients.

One can also consider cases where the applied load to the structure is also random. We can assume that a random variable $\theta^{(q)}$ represents this uncertainty. As a result, the stress field (B.32) at the element level is now given by

$$\boldsymbol{\sigma}^{(e)}(\boldsymbol{\theta}) = \mathbf{D}^{(e)}(\boldsymbol{\theta}) \boldsymbol{\varepsilon}^{(e)} = \mathbf{D}^{(e)}(\boldsymbol{\theta}) \mathbf{B}^{(e)} \mathbf{u}^{(e)}(\boldsymbol{\theta}). \quad (\text{B.32})$$

The uncertainty parameter vector is given by $\boldsymbol{\theta} = [\theta^{(q)}, \boldsymbol{\theta}^{(\text{KL})}] \in \mathbb{R}^{k+1}$. The finite element formulation (B.29) is then expressed as the global matrix equilibrium equation in function of a given parameter vector realization

$$\mathbf{K}(\boldsymbol{\theta})\mathbf{u}(\boldsymbol{\theta}) = \mathbf{f}(\boldsymbol{\theta}), \quad (\text{B.33})$$

where $\mathbf{K} \in \mathbb{R}^{n_{\text{dof}} \times n_{\text{dof}}}$ is the stiffness matrix, $\mathbf{f} \in \mathbb{R}^{n_{\text{dof}}}$ is the force vector, and $\mathbf{u} \in \mathbb{R}^{n_{\text{dof}}}$ is the vector of displacements; n_{dof} denotes the total number of degrees of freedom.

B.3. Adjoint differentiation for plane stress problems

The following derivative elements are computed at the GL point closest to the control node (cf. Figure 2.11).

B.3.1. Adjoint for the rare event simulation setting

We consider the case where the LSF is defined by the principal stress σ_1 . The stress field (Appendix B.2.3) evaluated at a specific GL point can be computed as

$$\boldsymbol{\sigma}^*(\boldsymbol{\theta}) = (\mathbf{D}^*(\boldsymbol{\theta}) \mathbf{B}^* \mathbf{M}) \mathbf{u}(\boldsymbol{\theta}), \quad (\text{B.34})$$

where the constitutive matrix $\mathbf{D}^* \in \mathbb{R}^{3 \times 3}$ and deformation matrix $\mathbf{B}^* \in \mathbb{R}^{3 \times n_{\text{eq}}}$ are those evaluated at the GL point closest to the control point. The value n_{eq} represents the number of finite element equations per element (e.g., number of element nodes times number of element dofs). Note that in (B.34), we employ a matrix $\mathbf{M} \in \mathbb{R}^{n_{\text{eq}} \times n_{\text{dof}}}$ to activate the dofs corresponding to the nodes of the element that has the control point as one of its nodes.

Our aim is to compute directly the gradient of the principal stress $\sigma_1(\boldsymbol{\sigma}^*(\mathbf{D}(\boldsymbol{\theta}), \mathbf{u}(\boldsymbol{\theta})))$ with respect to $\boldsymbol{\theta}$. We consider the total derivative

$$\frac{d\sigma_1}{d\boldsymbol{\theta}} = \frac{d\sigma_1}{d\boldsymbol{\sigma}^*} \left(\frac{d\boldsymbol{\sigma}^*}{d\mathbf{D}^*} \frac{d\mathbf{D}^*}{d\boldsymbol{\theta}} + \frac{d\boldsymbol{\sigma}^*}{d\mathbf{u}} \frac{d\mathbf{u}}{d\boldsymbol{\theta}} \right). \quad (\text{B.35})$$

The multiplicative term in (B.35) is obtained from the definition of the σ_1 stress

$$\frac{d\sigma_1}{d\boldsymbol{\sigma}^*} = \left[\frac{1}{2} \left(1 + \frac{\sigma_x^* - \sigma_y^*}{2e} \right), \frac{1}{2} \left(1 - \frac{\sigma_x^* - \sigma_y^*}{2e} \right), \frac{\tau_{xy}^*}{e} \right], \quad e = \sqrt{\left(\frac{\sigma_x^* - \sigma_y^*}{2} \right)^2 + \tau_{xy}^{*2}}. \quad (\text{B.36})$$

To compute the first term inside the parenthesis (B.35), we use the fact that the element constitutive matrix can be factored as $\hat{\mathbf{E}} \cdot \mathbf{D}_0 = \mathbf{D}$ (plane stress assumption), and that the derivative of the KL expansion of a lognormal field is $\hat{\mathbf{E}}\mathbf{A} = \hat{\mathbf{E}}' \in \mathbb{R}^{n_{\text{GP}} \times K}$. Hence,

$$\frac{d\boldsymbol{\sigma}^*}{d\mathbf{D}^*} = (\mathbf{B}^* \mathbf{M}) \mathbf{u} \quad \text{and} \quad \frac{d\mathbf{D}^*}{d\boldsymbol{\theta}} = \frac{d\mathbf{D}^*}{d\hat{\mathbf{E}}} \frac{d\hat{\mathbf{E}}}{d\boldsymbol{\theta}^{(\text{KL})}} = \mathbf{D}_0 \hat{\mathbf{E}}'^* \quad (\text{B.37})$$

where $\hat{\mathbf{E}}'^* \in \mathbb{R}^K$ is the KL expansion derivative at the GL point closest to the control node. Moreover, the second term inside the parenthesis in (B.35) is

$$\frac{d\boldsymbol{\sigma}^*}{d\mathbf{u}} = \mathbf{D}^* \mathbf{B}^* \mathbf{M} \quad \text{and} \quad \frac{d\mathbf{u}}{d\boldsymbol{\theta}} = \mathbf{K}^{-1} \left(\frac{d\mathbf{f}}{d\theta^{(q)}} - \frac{d\mathbf{K}}{d\boldsymbol{\theta}^{(\text{KL})}} \mathbf{u} \right), \quad (\text{B.38})$$

and the components of the derivative $\frac{d\mathbf{u}}{d\boldsymbol{\theta}}$ are obtained from the finite element formulation as

$$\frac{d\mathbf{f}}{d\theta^{(q)}} = \frac{d\mathbf{f}}{dq} \frac{dq}{d\theta^{(q)}} = \mathbf{c} \cdot \boldsymbol{\sigma}_q \quad \text{and} \quad \frac{d\mathbf{K}}{d\boldsymbol{\theta}^{(\text{KL})}} = \bigcup_e \int_{A^{(e)}} \mathbf{B}^{(e),\text{T}} \left(\mathbf{D}_0 \hat{\mathbf{E}}'^{(e)} \right) \mathbf{B}^{(e)} t \, dA^{(e)},$$

where $\mathbf{c} \in \mathbb{R}^{n_{\text{dof}}}$ is a constant vector that maps the surface load q to equivalent nodal forces, \bigcup_e denotes assembly procedure over all elements, t is the thickness of the plate, and the integration is performed over the area of each element $A^{(e)}$. The term $\frac{d\mathbf{K}}{d\boldsymbol{\theta}^{(\text{KL})}}$ is a rank-3 tensor with dimensions $K \times n_{\text{dof}} \times n_{\text{dof}}$, and its assembly is in general the most computationally intensive part of the adjoint method.

The idea of the adjoint method [11] is to expand and re-organize the terms in (B.35) to avoid the computation of expensive matrix operations. Specifically consider the term

$$\frac{d\sigma_1}{d\boldsymbol{\sigma}^*} \frac{d\boldsymbol{\sigma}^*}{d\mathbf{u}} \frac{d\mathbf{u}}{d\boldsymbol{\theta}} = \frac{d\sigma_1}{d\boldsymbol{\sigma}^*} \frac{d\boldsymbol{\sigma}^*}{d\mathbf{u}} \left(\mathbf{K}^{-1} \left(\frac{d\mathbf{f}}{d\theta^{(q)}} - \frac{d\mathbf{K}}{d\boldsymbol{\theta}^{(\text{KL})}} \mathbf{u} \right) \right) = \underbrace{\frac{d\sigma_1}{d\boldsymbol{\sigma}^*} \frac{d\boldsymbol{\sigma}^*}{d\mathbf{u}} \mathbf{K}^{-1}}_{\boldsymbol{\lambda}} \underbrace{\left(\frac{d\mathbf{f}}{d\theta^{(q)}} - \frac{d\mathbf{K}}{d\boldsymbol{\theta}^{(\text{KL})}} \mathbf{u} \right)}_{\frac{d\mathbf{p}}{d\boldsymbol{\theta}}},$$

where $\mathbf{p} = \mathbf{K}\mathbf{u} - \mathbf{f}$, and the adjoint multiplier $\boldsymbol{\lambda} \in \mathbb{R}^{n_{\text{dof}}}$ is computed by solving the adjoint equation

$$\mathbf{K}^\top \boldsymbol{\lambda} = \left[\frac{d\sigma_1}{d\boldsymbol{\sigma}^*} \frac{d\boldsymbol{\sigma}^*}{d\mathbf{u}} \right]^\top, \quad (\text{B.39})$$

we remark that the boundary conditions defining the finite element problem also need to be imposed on (B.39). Therefore, finding the gradient

$$\frac{d\sigma_1}{d\boldsymbol{\theta}} = \left[\frac{d\sigma_1}{d\theta^{(q)}}, \frac{d\sigma_1}{d\boldsymbol{\theta}^{(\text{KL})}} \right] \quad (\text{B.40})$$

amounts to

$$\frac{d\sigma_1}{d\theta^{(q)}} = \boldsymbol{\lambda}^\top \frac{d\mathbf{f}}{d\theta^{(q)}} \quad \text{and} \quad \frac{d\sigma_1}{d\boldsymbol{\theta}^{(\text{KL})}} = \frac{d\sigma_1}{d\boldsymbol{\sigma}^*} \frac{d\boldsymbol{\sigma}^*}{d\mathbf{D}^*} \frac{d\mathbf{D}^*}{d\boldsymbol{\theta}} + \boldsymbol{\lambda}^\top \left(-\frac{d\mathbf{K}}{d\boldsymbol{\theta}^{(\text{KL})}} \mathbf{u} \right). \quad (\text{B.41})$$

Figure B.2 shows a comparison between the gradient of the LSF (6.38) estimated by finite differences and the adjoint (B.41) at 6 different parameter realizations.

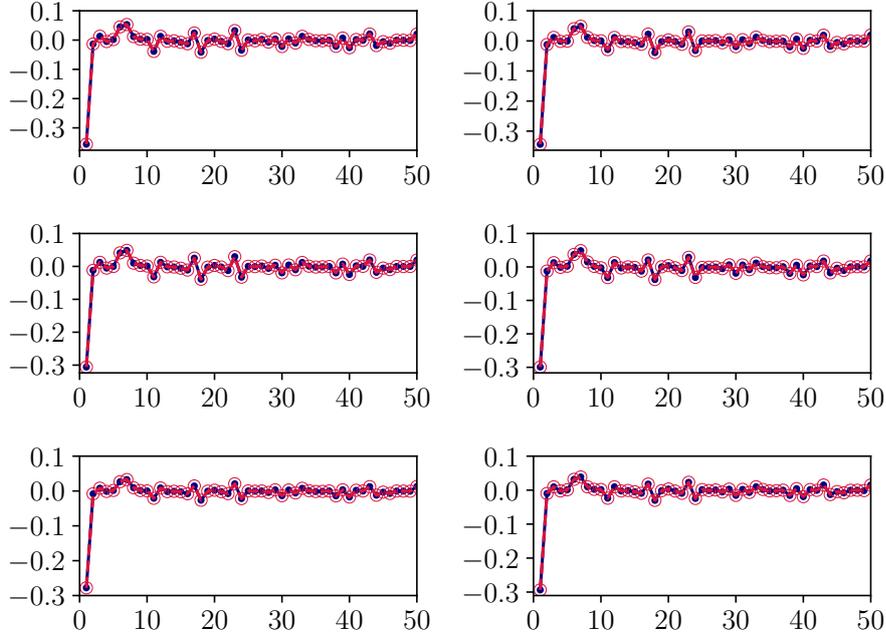


Figure B.2: Gradient of the LSF $\nabla g(\boldsymbol{\theta})$ (6.38): comparison between finite differences (blue) and the adjoint method (red) for 6 different samples $\boldsymbol{\theta}$. We plot the first 50 components of the gradient vector.

B.3.2. Adjoint for the Bayesian inversion setting

We consider the case where the likelihood function is defined by deformation measurements ε_{x_1} (as in section 4.6). In this case, the derivation of the adjoint for the PDE forward model defining the Bayesian inversion is easier to obtain compared to the one required for the rare event simulation. The deformation field (Appendix B.2.2) evaluated at a specific node can be computed at,

$$\boldsymbol{\varepsilon}^*(\boldsymbol{\theta}) = (\mathbf{B}^* \mathbf{M}) \mathbf{u}(\boldsymbol{\theta}). \quad (\text{B.42})$$

In this case, we need to compute the gradient:

$$\frac{d\varepsilon_{x_1}}{d\boldsymbol{\theta}^{(\text{KL})}} = \frac{d\varepsilon_{x_1}}{d\mathbf{u}} \frac{d\mathbf{u}}{d\boldsymbol{\theta}^{(\text{KL})}}. \quad (\text{B.43})$$

where each term is given by:

$$\frac{d\varepsilon_{x_1}}{d\mathbf{u}} = \mathbf{B}^* \mathbf{M} \quad \text{and} \quad \frac{d\mathbf{u}}{d\boldsymbol{\theta}^{(\text{KL})}} = \mathbf{K}^{-1} \left(-\frac{d\mathbf{K}}{d\boldsymbol{\theta}^{(\text{KL})}} \mathbf{u} \right), \quad (\text{B.44})$$

here deformation matrix $\mathbf{B}^* \in \mathbb{R}^{3 \times n_{\text{eq}}}$ are those evaluated at the GL point closest to the measurement locations. Therefore, similar to the previous section:

$$\frac{d\varepsilon_{x_1}}{d\mathbf{u}} \frac{d\mathbf{u}}{d\boldsymbol{\theta}^{(\text{KL})}} = \underbrace{\frac{d\varepsilon_{x_1}}{d\mathbf{u}} \mathbf{K}^{-1}}_{\boldsymbol{\lambda}} \underbrace{\left(-\frac{d\mathbf{K}}{d\boldsymbol{\theta}^{(\text{KL})}} \mathbf{u} \right)}_{\frac{d\mathbf{p}}{d\boldsymbol{\theta}}},$$

and the adjoint multiplier $\boldsymbol{\lambda} \in \mathbb{R}^{n_{\text{dof}}}$ is computed by solving the adjoint equation

$$\mathbf{K}^T \boldsymbol{\lambda} = \left[\frac{d\varepsilon_{x_1}}{d\mathbf{u}} \right]^T, \quad (\text{B.45})$$

from which the gradient can be computed as

$$\frac{d\varepsilon_{x_1}}{d\boldsymbol{\theta}^{(\text{KL})}} = \boldsymbol{\lambda}^T \left(-\frac{d\mathbf{K}}{d\boldsymbol{\theta}^{(\text{KL})}} \mathbf{u} \right). \quad (\text{B.46})$$

Bibliography

- [1] P. Abrahamsen. *A review of Gaussian random fields and correlation functions*. 2nd ed. Norwegian Computing Center, 1997. (Cited on pp. 12–17).
- [2] R. J. Adler. *The geometry of random fields*. Society for Industrial and Applied Mathematics (SIAM), 2010 (Original in 1981). (Cited on pp. 9, 10, 13–15).
- [3] S. Agapiou, J. M. Bardsley, O. Papaspiliopoulos, and A. M. Stuart. “Analysis of the Gibbs sampler for hierarchical inverse problems”. In: *SIAM/ASA Journal on Uncertainty Quantification* 2 (2014), pp. 511–544. (Cited on p. 92).
- [4] F. Al-Awadhi, M. Hurn, and C. Jennison. “Improving the acceptance rate of reversible jump MCMC proposals”. In: *Statistics & Probability Letters* 69 (2004), pp. 189–198. (Cited on p. 90).
- [5] D. L. Alaix and V. I. Carbone. “An efficient coupling of FORM and Karhunen-Loève series expansion”. In: *Engineering with Computers* 32.1 (2016), pp. 1–13. (Cited on pp. 6, 82).
- [6] A. Alexanderian. “A brief note on the Karhunen-Loève expansion”. In: [arXiv:1509.07526v2 eprint](https://arxiv.org/abs/1509.07526v2) (2015), pp. 1–14. (Cited on p. 16).
- [7] D. A. Alvarez. “Infinite random sets and applications in uncertainty analysis”. Ph.D. Thesis. Leopold-Franzens-Universität Innsbruck, 2007. (Cited on pp. 1, 2).
- [8] D. A. Alvarez, F. Uribe, and J. E. Hurtado. “Estimation of the lower and upper bounds on the probability of failure using subset simulation and random set theory”. In: *Mechanical Systems and Signal Processing* 100 (2018), pp. 782–801. (Cited on p. 2).
- [9] C. Andrieu, N. De Freitas, and A. Doucet. “Sequential MCMC for Bayesian model selection”. In: *Proceedings of the IEEE Signal Processing Workshop on Higher-Order Statistics. SPW-HOS '99*. 1999, pp. 130–134. (Cited on pp. 4, 95, 140).
- [10] C. Andrieu and J. Thoms. “A tutorial on adaptive MCMC”. In: *Statistics and Computing* 18 (2008), 343–373. (Cited on p. 62).
- [11] J. S. Arora and E. J. Haug. “Methods of design sensitivity analysis in structural optimization”. In: *AIAA Journal* 17.9 (1979), pp. 970–974. (Cited on pp. 125, 159).
- [12] R. B. Ash and C. Doléans-Dade. *Probability and measure theory*. 2nd ed. Harcourt/Academic Press, 2000. (Cited on pp. 33, 145, 147, 148).
- [13] S. Asmussen. “Large deviations in rare events simulation: Examples, counterexamples and alternatives”. In: *Monte Carlo and Quasi-Monte Carlo Methods 2000*. Ed. by K.-T. Fang, H. Niederreiter, and F. J. Hickernell. Springer, 2002, pp. 1–9. (Cited on p. 37).
- [14] K. E. Atkinson. *The numerical solution of integral equations of the second kind*. Cambridge University Press, 2010. (Cited on p. 19).
- [15] S.-K. Au and J. L. Beck. “Estimation of small failure probabilities in high dimensions by subset simulation”. In: *Probabilistic Engineering Mechanics* 16.4 (2001), pp. 263–277. (Cited on pp. 3, 4, 37, 40).

- [16] S. Bansal and S. H. Cheung. “A new stochastic simulation algorithm for updating robust reliability of linear structural dynamic systems subjected to future Gaussian excitations”. In: *Computer Methods in Applied Mechanics and Engineering* 326 (2017), pp. 481–504. (Cited on p. 4).
- [17] J. M. Bardsley. *Computational uncertainty quantification for inverse problems*. Society for Industrial and Applied Mathematics (SIAM), 2019. (Cited on p. 18).
- [18] A. Basudhar and S. Missoum. “A sampling-based approach for probabilistic design with random fields”. In: *49th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*. 2008, pp. 1–13. (Cited on p. 6).
- [19] T. Bayes and R. Price. “An essay towards solving a problem in the doctrine of chances (By the late Rev. Mr. Bayes, F. R. S. communicated by Mr. Price)”. In: *Philosophical Transactions of the Royal Society of London* 53 (1763), pp. 370–418. (Cited on p. 54).
- [20] J. L. Beck and L. S. Katafygiotis. “Updating models and their uncertainties. I: Bayesian statistical framework”. In: *Journal of Engineering Mechanics* 124.4 (1998), pp. 455–461. (Cited on p. 53).
- [21] J. L. Beck. “Bayesian system identification based on probability logic”. In: *Structural Control and Health Monitoring* 17 (2010), pp. 825–847. (Cited on pp. 72, 75, 76, 87).
- [22] J. L. Beck and K.-V. Yuen. “Model selection using response measurements: Bayesian probabilistic approach”. In: *Journal of Engineering Mechanics* 130.2 (2004), pp. 192–203. (Cited on pp. 86, 87).
- [23] R. Bellman. *Dynamic programming*. Princeton University Press, 1957. (Cited on p. 115).
- [24] J. Besag. *Markov chain Monte Carlo for statistical inference*. Center for Statistics and the Social Sciences, University of Washington, USA. 2001. (Cited on pp. 4, 92, 93).
- [25] A. Beskos, G. Roberts, A. Stuart, and J. Voss. “MCMC methods for diffusion bridges”. In: *Stochastics and Dynamics* 8.3 (2008), pp. 319–350. (Cited on p. 58).
- [26] W. Betz. “Bayesian inference of engineering models”. [Ph.D. Thesis](#). Technische Universität München, 2017. (Cited on p. 61).
- [27] W. Betz, I. Papaioannou, J. L. Beck, and D. Straub. “Bayesian inference with subset simulation: strategies and improvements”. In: *Computer Methods in Applied Mechanics and Engineering* 331 (2018), pp. 72–93. (Cited on pp. 60, 98, 99, 105).
- [28] W. Betz, I. Papaioannou, and D. Straub. “Numerical methods for the discretization of random fields by means of the Karhunen-Loève expansion”. In: *Computer Methods in Applied Mechanics and Engineering* 271 (2014), pp. 109–129. (Cited on pp. 19, 21).
- [29] J. Bierkens. “Non-reversible Metropolis-Hastings”. In: *Statistics and Computing* 26 (2016), pp. 1213–1228. (Cited on p. 57).
- [30] P. Billingsley. *Probability and measure*. 3rd ed. John Wiley & Sons, 1995. (Cited on p. 11).
- [31] J. R. Birge and F. Louveaux. *Introduction to stochastic programming*. 2nd ed. Springer, 2011. (Cited on p. 2).
- [32] C. M. Bishop. *Pattern recognition and machine learning*. Springer, 2006. (Cited on pp. 2, 76).
- [33] E. Bismut and D. Straub. “Optimal adaptive inspection and maintenance planning for deteriorating structural systems”. In: *Reliability Engineering & System Safety* Under review (2019). (Cited on p. 2).

-
- [34] S. G. Bobkov and M. Ledoux. “From Brunn-Minkowski to Brascamp-Lieb and to logarithmic Sobolev inequalities”. In: *Geometric & Functional Analysis GAFA* 10.5 (2000), pp. 1028–1052. (Cited on p. 120).
- [35] S. Bochner. “A theorem on Fourier-Stieltjes integrals”. In: *Bulletin of the American Mathematical Society* 40 (1934), pp. 271–276. (Cited on p. 16).
- [36] D. Bolin. “Spatial Matérn fields Driven by non-Gaussian noise”. In: *Scandinavian Journal of Statistics* 41.3 (2014), pp. 557–579. (Cited on p. 7).
- [37] Z. I. Botev, D. P. Kroese, and T. Taimre. “Generalized cross-entropy methods with applications to rare-event simulation and optimization”. In: *SIMULATION* 83.11 (2007), pp. 785–806. (Cited on p. 115).
- [38] Z. I. Botev and D. P. Kroese. “Efficient Monte Carlo simulation via the generalized splitting method”. In: *Statistics and Computing* 22.1 (2012), pp. 1–16. (Cited on pp. 4, 37, 41).
- [39] Z. I. Botev and D. P. Kroese. “The generalized cross entropy method, with applications to probability density estimation”. In: *Methodology and Computing in Applied Probability* 13.1 (2011), pp. 1–27. (Cited on p. 43).
- [40] A. Bouchard-Côté, S. J. Vollmer, and A. Doucet. “The Bouncy Particle Sampler: A non-reversible rejection-free Markov Chain Monte Carlo method”. In: *Journal of the American Statistical Association* 113.522 (2018), pp. 855–867. (Cited on p. 57).
- [41] A. F. Bower. *Applied mechanics of solids*. CRC press Taylor & Francis Group, 2010. (Cited on pp. 21, 151, 155).
- [42] G. E. P. Box and M. E. Muller. “A Note on the generation of random normal deviates”. In: *The Annals of Mathematical Statistics* 29.2 (1958), pp. 610–611. (Cited on p. 35).
- [43] S. P. Brooks, P. Giudici, and G. O. Roberts. “Efficient construction of reversible jump Markov chain Monte Carlo proposal distributions”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 65.1 (2003), pp. 3–55. (Cited on pp. 90, 92, 94).
- [44] C. Bucher. *Computational analysis of randomness in structural mechanics*. CRC Press. Structures and Infrastructures Series V.3, 2009. (Cited on p. 2).
- [45] J. A. Bucklew. *An introduction to rare event simulation*. Springer, 2004. (Cited on pp. 2, 33, 36, 37).
- [46] C. J. C. Burges. “Dimension reduction: A guided tour”. In: *Foundations and Trends in Machine Learning* 2.4 (2009), pp. 271–280. (Cited on pp. 2, 116).
- [47] B. P. Carlin and S. Chib. “Bayesian model choice via Markov chain Monte Carlo methods”. In: *Journal of the Royal Statistical Society: Series B* 57.3 (1995), pp. 473–484. (Cited on pp. 4, 92).
- [48] F. Cérou, P. Del Moral, T. Furon, and A. Guyader. “Sequential Monte Carlo for rare event estimation”. In: *Statistics and Computing* 22.3 (2012), pp. 795–808. (Cited on pp. 4, 37, 41).
- [49] P. Chen. “Hessian Matrix vs. Gauss-Newton Hessian Matrix”. In: *SIAM Journal on Numerical Analysis* 49.4 (2011), pp. 1417–1435. (Cited on p. 117).
- [50] V. Chen, M. M. Dunlop, O. Papaspiliopoulos, and A. M. Stuart. “Dimension-robust MCMC in Bayesian inverse problems”. In: [arXiv:1803.03344v2 eprint](https://arxiv.org/abs/1803.03344v2) (2019), pp. 1–29. (Cited on p. 93).

- [51] S. Chib and I. Jeliazkov. “Marginal likelihood from the Metropolis-Hastings output”. In: *Journal of the American Statistical Association* 96.453 (2001), pp. 270–281. (Cited on pp. 4, 87).
- [52] P. R. Conrad and Y. M. Marzouk. “Adaptive Smolyak pseudospectral approximations”. In: *SIAM Journal on Scientific Computing* 35.6 (2013), A2643–A2670. (Cited on p. 2).
- [53] P. G. Constantine, E. Dow, and Q. Wang. “Active subspace methods in theory and practice: applications to kriging surfaces”. In: *SIAM Journal on Scientific Computing* 36.4 (2014), A1500–A1524. (Cited on pp. 117, 125, 138).
- [54] P. G. Constantine. *Active subspaces: emerging ideas for dimension reduction in parameter studies*. Society for Industrial and Applied Mathematics (SIAM), 2015. (Cited on pp. 2, 117, 118, 124).
- [55] P. G. Constantine, C. Kent, and T. Bui-Thanh. “Accelerating Markov chain Monte Carlo with active subspaces”. In: *SIAM Journal on Scientific Computing* 38.5 (2016), A2779–A2805. (Cited on p. 117).
- [56] S. Cotter, G. Roberts, A. Stuart, and A. White. “MCMC methods for functions: modifying old algorithms to make them faster”. In: *Statistical Science* 28.3 (2013), pp. 424–446. (Cited on pp. 58, 59, 88, 92).
- [57] T. Cui, J. Martin, Y. M. Marzouk, A. Solonen, and A. Spantini. “Likelihood-informed dimension reduction for nonlinear inverse problems”. In: *Inverse Problems* 30 (2014), p. 0114015. (Cited on pp. 2, 6, 116, 117, 122, 123, 138).
- [58] T. Cui, K. J. H. Law, and Y. M. Marzouk. “Dimension-independent likelihood-informed MCMC”. In: *Journal of Computational Physics* 304.6 (2016), pp. 109–137. (Cited on pp. 109, 141).
- [59] F. Cérou, A. Guyader, and M. Rousset. “Adaptive multilevel splitting: Historical perspective and recent results”. In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 29 (4 2019). (Cited on p. 41).
- [60] M. Dashti and A. M. Stuart. “The Bayesian approach to inverse problems”. In: *Handbook of uncertainty quantification*. Ed. by R. Ghanem, D. Higdon, and H. Owhadi. Springer International Publishing, 2017. Chap. 10, pp. 311–428. (Cited on pp. 54, 58, 59).
- [61] V. De Oliveira, B. Kedem, and D. A. Short. “Bayesian prediction of transformed Gaussian random fields”. In: *Journal of the American Statistical Association* 92.440 (1997), pp. 1422–1433. (Cited on pp. 3, 19, 55).
- [62] P. Del Moral, A. Doucet, and A. Jasra. “Sequential Monte Carlo samplers”. In: *Journal of the Royal Statistical Society: Series B* 68.3 (2006), pp. 411–436. (Cited on pp. 3, 38, 39, 59).
- [63] A. P. Dempster. “Upper and lower probabilities induced by a multivalued mapping”. In: *Annals of Mathematical Statistics* 38 (1967), pp. 325–339. (Cited on p. 1).
- [64] G. Deodatis. “Weighted integral method. I: Stochastic stiffness matrix”. In: *Journal of Engineering Mechanics* 117.8 (1991), pp. 1851–1864. (Cited on p. 18).
- [65] A. Der Kiureghian and J.-B. Ke. “The stochastic finite element method in structural reliability”. In: *Probabilistic Engineering Mechanics* 3.2 (1988), pp. 83–91. (Cited on pp. 2, 18).
- [66] F. DiazDelaO, A. Garbuno-Inigo, S. Au, and I. Yoshida. “Bayesian updating and model class selection with subset simulation”. In: *Computer Methods in Applied Mechanics and Engineering* 317 (2017), pp. 1102–1121. (Cited on p. 61).

-
- [67] C. R. Dietrich and G. N. Newsam. “Fast and exact simulation of stationary Gaussian processes through circulant embedding of the covariance matrix”. In: *SIAM Journal on Scientific Computing* 18.4 (1997), pp. 1088–1107. (Cited on p. 18).
- [68] O. Ditlevsen and H. O. Madsen. *Structural reliability methods*. 1st ed. John Wiley & Sons, 1996. (Cited on pp. 47, 130).
- [69] O. Ditlevsen, R. E. Melchers, and H. Gluwer. “General multi-dimensional probability integration by directional simulation”. In: *Computers & Structures* 36.2 (1990), pp. 355–368. (Cited on pp. 4, 37).
- [70] J. L. Doob. *Stochastic processes*. John Wiley & Sons, 1953. (Cited on p. 13).
- [71] A. B. Duncan, N. Nüsken, and G. A. Pavliotis. “Using perturbed underdamped Langevin dynamics to efficiently sample from probability distributions”. In: *Journal of Statistical Physics* 169 (2017), pp. 1098–1131. (Cited on p. 58).
- [72] M. Ehre, I. Papaioannou, and D. Straub. “A framework for global reliability sensitivity analysis in the presence of multi-uncertainty”. In: *Reliability Engineering & System Safety* 195 (2020), p. 106726. (Cited on p. 1).
- [73] T. A. El Moselhy and Y. M. Marzouk. “Bayesian inference with optimal maps”. In: *Journal of Computational Physics* 231.23 (2012), pp. 7815–7850. (Cited on p. 3).
- [74] A. H. Elsheikh, I. Hoteit, and M. F. Wheeler. “Efficient Bayesian inference of subsurface flow models using nested sampling and sparse polynomial chaos surrogates”. In: *Computer Methods in Applied Mechanics and Engineering* 269 (2014), pp. 515–537. (Cited on p. 87).
- [75] A. F. Emery and A. V. Nenarokomov. “Optimal experiment design”. In: *Measurement Science and Technology* 9.6 (1998), pp. 864–876. (Cited on p. 2).
- [76] S. Englund and R. Rackwitz. “A benchmark study on importance sampling techniques in structural reliability”. In: *Structural Safety* 12.4 (1993), pp. 255–276. (Cited on pp. 4, 128).
- [77] Y. Fan, Q. Peters, and S. Sisson. “Automating and evaluating reversible jump MCMC proposal distributions”. In: *Statistics and Computing* 19 (2009), pp. 409–421. (Cited on p. 90).
- [78] Y. Fan and S. A. Sisson. “Reversible jump Markov chain Monte Carlo”. In: *Handbook of Markov chain Monte Carlo*. Ed. by S. Brooks, A. Gelman, G. L. Jones, and X.-L. Meng. Chapman & Hall/CRC, 2011. Chap. 3, pp. 67–91. (Cited on p. 95).
- [79] B. Fiessler, H.-J. Neumann, and R. Rackwitz. “Quadratic limit states in structural reliability”. In: *Journal of the Engineering Mechanics Division (ASCE)* 105.4 (1979), pp. 661–676. (Cited on p. 34).
- [80] I. M. Franck and P. S. Koutsourelakis. “Multimodal, high-dimensional, model-based, Bayesian inverse problems with applications in biomechanics”. In: *Journal of Computational Physics* 329 (2017), pp. 91–125. (Cited on p. 3).
- [81] P. S. Frederiksen. “Parameter uncertainty and design of optimal experiments for the estimation of elastic constants”. In: *International Journal of Solids and Structures* 35.12 (1998), pp. 1241–1260. (Cited on p. 2).
- [82] A. M. Freudenthal. “Introductory remarks”. In: *1st International Conference on Structural Safety & Reliability (ICOSSAR)*. Ed. by A. M. Freudenthal. Pergamon, 1972, pp. 5–9. (Cited on p. 33).
- [83] G.-A. Fuglstad, D. Simpson, F. Lindgren, and H. Rue. “Constructing priors that penalize the complexity of Gaussian random fields”. In: *Journal of the American Statistical Association* 114.525 (2019), pp. 445–452. (Cited on pp. 17, 88).

- [84] G.-A. Fuglstad, D. Simpson, F. Lindgren, and H. Rue. “Does non-stationary spatial data always require non-stationary random fields?” In: *Spatial Statistics* 14 (2015), pp. 505–531. (Cited on p. 13).
- [85] A. Gelman and X.-L. Meng. “Simulating normalizing constants: from importance sampling to bridge sampling to path sampling”. In: *Statistical Science* 13.2 (1998), pp. 163–185. (Cited on pp. 2, 4, 86, 87).
- [86] C. Geuzaine and J.-F. Remacle. “Gmsh: a three-dimensional finite element mesh generator with built-in pre- and post-processing facilities”. In: *International Journal for Numerical Methods in Engineering* 79.11 (2009), pp. 1309–1331. (Cited on pp. 28, 78).
- [87] C. J. Geyer. “Introduction to Markov chain Monte Carlo”. In: *Handbook of Markov chain Monte Carlo*. Ed. by S. Brooks, A. Gelman, G. L. Jones, and X.-L. Meng. Chapman & Hall/CRC, 2011. Chap. 1, pp. 3–48. (Cited on pp. 56, 57, 89).
- [88] S. Geyer, I. Papaioannou, C. Kunz, and D. Straub. “Reliability assessment of large hydraulic structures with spatially distributed measurements”. In: *Structure and Infrastructure Engineering* (2019). (Cited on pp. 4, 64).
- [89] S. Geyer, I. Papaioannou, and D. Straub. “Cross entropy-based importance sampling using Gaussian densities revisited”. In: *Structural Safety* 76 (2019), pp. 15–27. (Cited on pp. 42, 45, 138).
- [90] R. G. Ghanem and P. D. Spanos. *Stochastic finite elements: a spectral approach*. Revised edition. Dover Publications, 2012 (Original in 1991). (Cited on pp. 1, 2, 6, 16, 19, 20, 78).
- [91] A. L. Gibbs and F. E. Su. “On choosing and bounding probability metrics”. In: *International Statistical Review* 70.3 (2002), pp. 419–435. (Cited on pp. 42, 88, 142).
- [92] W. R. Gilks, N. G. Best, and K. K. C. Tan. “Adaptive rejection Metropolis sampling within Gibbs sampling”. In: *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 44.4 (1995), pp. 455–472. (Cited on p. 92).
- [93] E. Giné and R. Nickl. *Mathematical foundations of infinite-dimensional statistical models*. Cambridge University Press, 2016. (Cited on p. 72).
- [94] S. J. Godsill. “On the relationship between MCMC methods for model uncertainty”. In: *Journal of Computational and Graphical Statistics* 10.2 (2001), pp. 230–248. (Cited on pp. 4, 92–94).
- [95] L. Grasedyck, D. Kressner, and C. Tobler. “A literature survey of low-rank tensor approximation techniques”. In: *GAMM-Mitteilungen* 36.1 (2013), pp. 53–78. (Cited on p. 2).
- [96] P. J. Green. “Reversible jump Markov chain Monte Carlo computation and Bayesian model determination”. In: *Biometrika* 82.4 (1995), pp. 711–732. (Cited on pp. 4, 87, 89, 90, 93).
- [97] P. J. Green. “Trans-dimensional Markov chain Monte Carlo”. In: *Highly structured stochastic systems*. Ed. by P. J. Green, N. L. Hjort, and S. Richardson. Oxford University Press, 2003. Chap. 6, pp. 179–198. (Cited on pp. 4, 87).
- [98] D. A. Griffith and G. Arbia. “Detecting negative spatial autocorrelation in georeferenced random variables”. In: *International Journal of Geographical Information Science* 24.3 (2010), pp. 417–437. (Cited on p. 51).
- [99] D. V. Griffiths, J. Huang, and G. A. Fenton. “Influence of spatial variability on slope reliability using 2-D random fields”. In: *Journal of Geotechnical and Geoenvironmental Engineering* 135.10 (2009), pp. 1367–1378. (Cited on p. 5).

-
- [100] M. Grigoriu. “Evaluation of Karhunen-Loève, spectral, and sampling representations for stochastic processes”. In: *Journal of Engineering Mechanics* 132.2 (2006), pp. 179–189. (Cited on p. 15).
- [101] M. Grigoriu. “Existence and construction of translation models for stationary non-Gaussian processes”. In: *Probabilistic Engineering Mechanics* 24.4 (2009), pp. 545–551. (Cited on pp. 15, 19).
- [102] M. D. Grigoriu. *Stochastic systems: uncertainty quantification and propagation*. Springer, 2012. (Cited on pp. 15, 58).
- [103] J. Hadamard. “Sur les problèmes aux dérivées partielles et leur signification physique”. In: *Bulletin of the University of Princeton* 13 (1902). (Cited on p. 53).
- [104] M. Hairer, A. M. Stuart, J. Voss, and P. Wiberg. “Analysis of SPDEs arising in path sampling part I: the Gaussian case”. In: *Communications in Mathematical Sciences* 3.4 (2005), pp. 587–603. (Cited on p. 58).
- [105] M. S. Handcock and M. L. Stein. “A Bayesian analysis of Kriging”. In: *Technometrics* 35 (4 1993). (Cited on p. 17).
- [106] M. S. Handcock and J. R. Wallis. “An approach to statistical spatial-temporal modeling of meteorological fields”. In: *Journal of the American Statistical Association* 89 (426 1994). (Cited on p. 17).
- [107] P. C. Hansen. *Discrete inverse problems: Insight and algorithms*. Society for Industrial and Applied Mathematics (SIAM), 2010. (Cited on pp. 22, 53, 54, 138).
- [108] P. C. Hansen, J. G. Nagy, and D. P. O’Leary. *Deblurring images: Matrices, spectra, and filtering*. Society for Industrial and Applied Mathematics (SIAM), 2006. (Cited on p. 7).
- [109] K. A. Haskard. “An anisotropic Matern spatial covariance model: REML estimation and properties”. PhD thesis. BiometricsSA, School of Agriculture and Wine, University of Adelaide, 2007. (Cited on p. 17).
- [110] D. Hastie. “Towards automatic reversible jump Markov chain Monte Carlo”. PhD thesis. Department of Mathematics, Faculty of Science, University of Bristol, 2005. (Cited on p. 90).
- [111] D. I. Hastie and P. J. Green. “Model choice using reversible jump Markov chain Monte Carlo”. In: *Statistica Neerlandica* 66.3 (2012), pp. 309–338. (Cited on pp. 85, 87, 89, 90).
- [112] W. K. Hastings. “Monte Carlo sampling methods using Markov chains and their applications”. In: *Biometrika* 57.1 (1970), pp. 97–109. (Cited on p. 57).
- [113] J. C. Helton. “Uncertainty and sensitivity analysis in the presence of stochastic and subjective uncertainty”. In: *Journal of Statistical Computation and Simulation* 57 (1997), pp. 3–76. (Cited on p. 1).
- [114] M. Hohenbichler and R. Rackwitz. “Improvement of second-order reliability estimates by importance sampling”. In: *Journal of Engineering Mechanics* 114.12 (1988), pp. 2195–2199. (Cited on pp. 35, 37).
- [115] B. Hosseini and N. Nigam. “Well-posed Bayesian inverse problems: priors with exponential tails”. In: *SIAM/ASA Journal on Uncertainty Quantification* 5 (2017), pp. 436–465. (Cited on pp. 19, 54, 142).
- [116] H. Hotelling. “Analysis of a complex of statistical variables into principal components.” In: *Journal of Educational Psychology* 24.6 (1933). (Cited on p. 116).

- [117] J. H. Huggins, T. Campbell, M. Kasprzak, and T. Broderick. “Practical bounds on the error of Bayesian posterior approximations: a nonasymptotic approach”. In: [arXiv:1809.09505v2 eprint](https://arxiv.org/abs/1809.09505v2) (2018), pp. 1–22. (Cited on p. 142).
- [118] T. J. Hughes. *The finite element method: linear static and dynamic finite element analysis*. Dover Publications, 2000. (Cited on pp. 1, 152, 157).
- [119] J. E. Hurtado. “Dimensionality reduction and visualization of structural reliability problems using polar features”. In: *Probabilistic Engineering Mechanics* 29 (2012), pp. 16–31. (Cited on p. 116).
- [120] J. E. Hurtado. *Structural reliability: Statistical learning perspectives*. Springer, 2004. (Cited on p. 2).
- [121] J. E. Hurtado and D. A. Alvarez. “Neural-network-based reliability analysis: a comparative study”. In: *Computer methods in applied mechanics and engineering* 191.1-2 (2001), pp. 113–132. (Cited on p. 2).
- [122] L. I. Rudin, S. Osher, and E. Fatemi. “Nonlinear total variation based noise removal algorithms”. In: *Physica D: Nonlinear Phenomena* 60.1-4 (1992), pp. 259–268. (Cited on p. 53).
- [123] A. Jasra, D. A. Stephens, and C. C. Holmes. “Population-based reversible jump Markov chain Monte Carlo”. In: *Biometrika* 94.4 (2007), pp. 787–807. (Cited on pp. 4, 96, 140).
- [124] E. T. Jaynes. *Probability theory: The logic of science*. Cambridge University Press, 2003. (Cited on p. 85).
- [125] H. Jeffreys. *Scientific inference*. Muller Press, 2007. (Cited on p. 53).
- [126] H. Jeffreys. *Theory of probability*. 3rd ed. Oxford University Press, 2003. (Cited on pp. 86, 87).
- [127] H. Jensen, C. Vergara, C. Papadimitriou, and E. Millas. “The use of updated robust reliability measures in stochastic dynamical systems”. In: *Computer Methods in Applied Mechanics and Engineering* 267 (2013), pp. 293–317. (Cited on pp. 4, 64).
- [128] Z. Jiang and J. Li. “High dimensional structural reliability with dimension reduction”. In: *Structural Safety* 69 (2017), pp. 35–46. (Cited on p. 6).
- [129] C. Johnson. *Numerical solution of partial differential equations by the finite element method*. Dover Publications, 2009. (Cited on pp. 28, 152).
- [130] H. Kahn and A. W. Marshall. “Methods of reducing sample size in Monte Carlo computations”. In: *Journal of the Operations Research Society of America* 1.5 (1953), pp. 263–278. (Cited on p. 36).
- [131] J. Kaipio and E. Somersalo. *Statistical and computational inverse problems*. Springer, 2005. (Cited on pp. 3, 53, 54, 66, 67).
- [132] K. Karhunen. “Über lineare Methoden in der Wahrscheinlichkeitsrechnung”. In: *Annales Academiæ Scientiarum Fennicæ. Series A.1, Mathematica-Physica* 37 (1947), pp. 1–79. (Cited on p. 18).
- [133] A. Keese. *A review of recent developments in the numerical solution of stochastic partial differential equations (stochastic finite elements)*. Review. Institute of Scientific Computing, Technical University Braunschweig, 2003. (Cited on p. 20).
- [134] D. G. Kendall. “Foundations of a theory of random sets”. In: *Stochastic geometry*. Ed. by E. F. Harding and D. G. Kendall. Wiley, 1974, pp. 322–376. (Cited on p. 1).

-
- [135] M. C. Kennedy and A. O’Hagan. “Bayesian calibration of computer models”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63.3 (2001), pp. 425–464. (Cited on p. 1).
- [136] M. J. Kochenderfer. *Decision making under uncertainty: Theory and application*. MIT Press, 2015. (Cited on p. 2).
- [137] A. N. Kolmogorov. *Foundations of the theory of probability*. Chelsea Publishing Company, 1956. (Cited on pp. 1, 145).
- [138] A. N. Kolmogorov and S. V. Fomin. *Introductory real analysis*. Dover Publications, 1975. (Cited on pp. 13, 16, 146).
- [139] P. S. Koutsourelakis, H. J. Pradlwarter, and G. I. Schuëller. “Reliability of structures in high dimensions, part I: Algorithms and applications”. In: *Probabilistic Engineering Mechanics* 19 (2004), pp. 409–417. (Cited on pp. 4, 37).
- [140] S. Lacaze, L. Brevault, S. Missoum, and M. Balesdent. “Probability of failure sensitivity with respect to decision variables”. In: *Structural and Multidisciplinary Optimization* 52.2 (2015), pp. 375–381. (Cited on p. 34).
- [141] P. Laplace. “Mémoire sur la probabilité des causes par les événements”. In: *Mémoires de l’Académie royale des Sciences de MI (Savants étrangers)* 4 (1774), pp. 621–656. (Cited on p. 54).
- [142] J. Latz, M. Eisenberger, and E. Ullmann. “Fast sampling of parameterised Gaussian random fields”. In: *Computer Methods in Applied Mechanics and Engineering* 348 (2019), pp. 978–1012. (Cited on pp. 3, 141).
- [143] M. A. Lawrence. “Basis random variables in finite element analysis”. In: *International Journal for Numerical Methods in Engineering* 24.10 (1987), pp. 1849–1863. (Cited on p. 2).
- [144] M. Lemaire, A. Chateauneuf, and J. Mitteau. *Structural reliability*. Wiley-ISTE, 2009. (Cited on pp. 7, 34).
- [145] C.-C. Li and A. Der Kiureghian. “Optimal discretization of random fields”. In: *Journal of Engineering Mechanics* 119.6 (1993), pp. 1136–1154. (Cited on pp. 2, 18, 20).
- [146] J. Li, J. Li, and D. Xiu. “An efficient surrogate-based method for computing rare failure probability”. In: *Journal of Computational Physics* 230 (24 2011). (Cited on p. 115).
- [147] J. Li. “A note on the Karhunen-Loève expansions for infinite-dimensional Bayesian inverse problems”. In: *Statistics & Probability Letters* 106 (2015), pp. 1–4. (Cited on p. 142).
- [148] X. Li and F. T.-C. Tsai. “Bayesian model averaging for groundwater head prediction and uncertainty analysis using multimodel and multimethod”. In: *Water Resources Research* 45.9 (2009). (Cited on p. 86).
- [149] F. Liang, C. Liu, and R. Carroll. *Advanced Markov chain Monte Carlo methods: learning from past samples*. Springer, 2010. (Cited on pp. 61, 93).
- [150] F. Lindgren, H. Rue, and J. Lindström. “An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach”. In: *Journal of the Royal Statistical Society: Series B* 73.4 (2011), pp. 423–498. (Cited on pp. 17, 18).
- [151] J. Liu, J. Lu, and X. Zhou. “Efficient rare event simulation for failure problems in random media”. In: *SIAM Journal on Scientific Computing* 37.2 (2015), A609–A624. (Cited on p. 5).

- [152] P.-L. Liu and K.-G. Liu. “Selection of random field mesh in finite element reliability analysis”. In: *Journal of Engineering Mechanics* 119.4 (1993), pp. 667–680. (Cited on p. 5).
- [153] Q. Liu and X. Zhang. “A Chebyshev polynomial-based Galerkin method for the discretization of spatially varying random properties”. In: *Acta Mechanica* 228.6 (2017), pp. 2063–2081. (Cited on p. 2).
- [154] W. K. Liu, T. Belytschko, and A. Mani. “Random field finite elements”. In: *International Journal for Numerical Methods in Engineering* 23.10 (1986), pp. 1831–1845. (Cited on pp. 2, 18).
- [155] M. Loève. “Fonctions aléatoires du second ordre”. In: *La Revue Scientifique* 84 (1946), pp. 195–206. (Cited on p. 18).
- [156] A. Logg, K.-A. Mardal, and G. Wells. *Automated solution of differential equations by the finite element method: The FEniCS book*. Springer, 2012. (Cited on p. 1).
- [157] C. C. Margossian. “A review of Automatic Differentiation and its efficient implementation”. In: [arXiv:1811.05031v2 eprint](https://arxiv.org/abs/1811.05031v2) (2019), pp. 1–32. (Cited on p. 125).
- [158] G. Marsaglia and T. A. Bray. “A convenient method for generating normal variables”. In: *SIAM Review* 6.3 (1964), pp. 260–264. (Cited on p. 35).
- [159] G. Marsaglia and W. W. Tsang. “The Ziggurat method for generating random variables”. In: *Journal of Statistical Software* 5.8 (2000), pp. 1–7. (Cited on p. 35).
- [160] Y. M. Marzouk. *Lecture exercise: Numerical Methods for Stochastic Modeling and Inference*. Sept. Massachusetts Institute of Technology. Fall 2018. (Cited on pp. 48, 62).
- [161] Y. M. Marzouk and H. N. Najm. “Dimensionality reduction and polynomial chaos acceleration of Bayesian inference in inverse problems”. In: *Journal of Computational Physics* 228.6 (2009), pp. 1862–1902. (Cited on pp. 3, 109, 141).
- [162] G. Matheron. *Random sets and integral geometry*. Wiley, 1975. (Cited on p. 1).
- [163] H. G. Matthies, C. E. Brenner, C. G. Bucher, and C. Guedes Soares. “Uncertainties in probabilistic numerical analysis of structures and solids – Stochastic finite elements”. In: *Structural Safety* 19 (3 1997), pp. 283–336. (Cited on p. 18).
- [164] B. Matérn. *Spatial variation: stochastic models and their application to some problems in forest surveys and other sampling investigations*. Report No. 49/5. Forest Research Institute of Sweden, 1960. (Cited on pp. 9, 13, 16, 17).
- [165] M. McKay, R. Beckman, and W. Conover. “A Comparison of three methods for selecting values of input variables in the analysis of output from a computer code”. In: *Technometrics* 21.2 (1979), pp. 239–245. (Cited on p. 35).
- [166] J. Mercer. “Functions of positive and negative type, and their connection the theory of integral equations”. In: *Philosophical Transactions of the Royal Society of London. Series A: Containing Papers of a Mathematical or Physical Character* 209 (1909). (Cited on pp. 16, 18).
- [167] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. “Equation of state calculations by fast computing machines”. In: *Chemical Physics* 21.6 (1953), pp. 1087–1092. (Cited on p. 57).
- [168] C.-H. Michael Tso, Y. Zha, T.-C. Jim Yeh, and J.-C. Wen. “The relative importance of head, flux, and prior information in hydraulic tomography analysis”. In: *Water Resources Research* 52.1 (2016), pp. 3–20. (Cited on p. 2).

-
- [169] M. Muto and J. L. Beck. “Bayesian updating and model class selection for hysteretic structural models using stochastic simulation”. In: *Journal of Vibration and Control* 14.1-2 (2008), pp. 7–34. (Cited on p. 87).
- [170] P. Nath, Z. Hu, and S. Mahadevan. “Sensor placement for calibration of spatially varying model parameters”. In: *Journal of Computational Physics* 343 (2017), pp. 150–169. (Cited on p. 2).
- [171] R. M. Neal. “Slice sampling”. In: *Annals of Statistics* 31.3 (2003), pp. 705–767. (Cited on p. 61).
- [172] J. von Neumann. “Various techniques used in connection with random digits”. In: *Monte Carlo Method*. Ed. by A. S. Householder, G. E. Forsythe, and H. H. Germond. Vol. 12. National Bureau of Standards Applied Mathematics Series. US Government Printing Office, 1951. Chap. 13, pp. 36–38. (Cited on p. 35).
- [173] C. Oedekoven, R. King, S. Buckland, M. Mackenzie, K. Evans, and L. Burger. “Using hierarchical centering to facilitate a reversible jump MCMC algorithm for random effects models”. In: *Computational Statistics & Data Analysis* 98 (2016), pp. 79–90. (Cited on p. 90).
- [174] A. B. Owen. *Monte Carlo theory, methods and examples*. statweb.stanford.edu/~owen/mc/, 2013. (Cited on pp. 35–37, 59, 95, 99, 116).
- [175] E. Oñate. *Structural analysis with the finite element method: Linear statics. Volume 1 Basis and Solids*. Springer, 2013. (Cited on p. 28).
- [176] C. Papadimitriou, J. Beck, and L. Katafygiotis. “Updating robust reliability using structural test data”. In: *Probabilistic Engineering Mechanics* 16.2 (2001), pp. 103–113. (Cited on pp. 4, 63, 64).
- [177] I. Papaioannou, W. Betz, K. Zwirgmaier, and D. Straub. “MCMC algorithms for subset simulation”. In: *Probabilistic Engineering Mechanics* 41 (2015), pp. 89–103. (Cited on pp. 41, 47, 58, 59, 62, 130).
- [178] I. Papaioannou, M. Ehre, and D. Straub. “PLS-based adaptation for efficient PCE representation in high dimensions”. In: *Journal of Computational Physics* 387 (2019), pp. 186–204. (Cited on p. 142).
- [179] I. Papaioannou, S. Geyer, and D. Straub. “Improved cross entropy-based importance sampling with a flexible mixture model”. In: *Reliability Engineering & System Safety* 191 (2019), 1–11 (106564). (Cited on pp. 6, 44, 45, 115, 127, 129, 135, 140).
- [180] I. Papaioannou, C. Papadimitriou, and D. Straub. “Sequential importance sampling for structural reliability analysis”. In: *Structural Safety* 62 (2016), pp. 66–75. (Cited on pp. 4, 37–39, 127, 142).
- [181] I. Papaioannou and D. Straub. “Learning soil parameters and updating geotechnical reliability estimates under spatial variability – theory and application to shallow foundations”. In: *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards* 11.1 (2017), pp. 116–128. (Cited on p. 4).
- [182] A. Papoulis and S. U. Pillai. *Probability, random variables and stochastic processes*. 4th ed. McGraw-Hill, 2002. 678 pp.. (Cited on p. 1).
- [183] B. Peherstorfer, B. Kramer, and K. Willcox. “Multifidelity preconditioning of the cross-entropy method for rare event simulation and failure probability estimation”. In: *SIAM/ASA Journal on Uncertainty Quantification* 6.2 (2018), pp. 737–761. (Cited on p. 115).

- [184] D. B. Phillips and A. F. M. Smith. “Bayesian model comparison via jump diffusions”. In: *Markov chain Monte Carlo in practice*. Ed. by W. R. Gilks, S. Richardson, and D. J. Spiegelhalter. Springer-Science+Business media, 1996. Chap. 13, pp. 215–239. (Cited on pp. 4, 87).
- [185] K. Phoon, S. Huang, and S. Quek. “Implementation of Karhunen-Loève expansion for simulation using a wavelet-Galerkin scheme”. In: *Probabilistic Engineering Mechanics* 17 (2002), pp. 293–303. (Cited on p. 19).
- [186] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical recipes in C: the art of scientific computing*. 3rd ed. Cambridge University Press, 2007. 1256 pp.. (Cited on pp. 19, 20, 25, 58, 157).
- [187] R. Rackwitz and B. Fiessler. “Structural reliability under combined random load sequences”. In: *Computers & Structures* 9.5 (1978), pp. 489–494. (Cited on pp. 7, 34).
- [188] M. Raissi, P. Perdikaris, and G. Karniadakis. “Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations”. In: *Journal of Computational Physics* 378 (2019), pp. 686–707. (Cited on p. 2).
- [189] M. Raissi. “Deep hidden physics models: Deep learning of nonlinear partial differential equations”. In: *The Journal of Machine Learning Research* 19.1 (2018), pp. 932–955. (Cited on p. 2).
- [190] C. E. Rasmussen and C. K. I. Williams. *Gaussian processes for machine learning*. The MIT Press, 2006. (Cited on pp. 2, 3, 17).
- [191] S. Reich. *Probabilistic forecasting and Bayesian data assimilation*. Reprint edition. Cambridge University Press, 2015. (Cited on p. 7).
- [192] S. Richardson and P. J. Green. “On Bayesian analysis of mixtures with an unknown number of components (with discussion)”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 59.4 (1997), pp. 731–792. (Cited on p. 87).
- [193] C. P. Robert. *The Bayesian choice: from decision-theoretic foundations to computational implementation*. 2nd ed. Springer, 2007. (Cited on pp. 3, 55, 56, 67, 86).
- [194] L. Roininen, M. Girolami, S. Lasanen, and M. Markkanen. “Hyperpriors for Matérn fields with applications in Bayesian inversion”. In: *Inverse Problems & Imaging* 13.1 (2019), pp. 1–29. (Cited on pp. 3, 141).
- [195] J. S. Rosenthal. *A first look at rigorous probability theory*. 2nd ed. World Scientific Publishing Company, 2006. (Cited on pp. 119, 145, 146, 148).
- [196] R. Y. Rubinstein. “The stochastic minimum cross-entropy method for combinatorial optimization and rare-event estimation”. In: *Methodology and Computing in Applied Probability* 7.1 (2005), pp. 5–50. (Cited on p. 43).
- [197] R. Y. Rubinstein. “Optimization of computer simulation models with rare events”. In: *European Journal of Operational Research* 99.1 (1997), pp. 89–112. (Cited on pp. 6, 37, 41, 42).
- [198] R. Y. Rubinstein and D. P. Kroese. *Simulation and the Monte Carlo method*. 3rd ed. John Wiley & Sons, 2017. (Cited on pp. 3, 4, 34, 35, 43, 87, 94, 95).
- [199] D. Rudolf and B. Sprungk. “On a generalization of the preconditioned Crank-Nicolson Metropolis algorithm”. In: *Foundations of Computational Mathematics* 18.2 (2018), pp. 309–343. (Cited on pp. 58, 92).

-
- [200] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola. *Global Sensitivity Analysis: The Primer*. Wiley-Interscience, 2008. (Cited on pp. 2, 116).
- [201] R. L. Schilling. *Measures, integrals and martingales*. Cambridge University Press, 2005. (Cited on p. 34).
- [202] R. Schöbi and B. Sudret. “Application of conditional random fields and sparse polynomial chaos expansions in structural reliability analysis”. In: *12th International Conference on Structural Safety & Reliability (ICOSSAR)*. Ed. by C. Bucher, B. R. Ellingwood, and D. M. Frangopol. TU-Verlag Vienna, 2017, pp. 1356–1363. (Cited on p. 137).
- [203] G. Schwarz. “Estimating the Dimension of a Model”. In: *Annals of Statistics* 6.2 (1978), pp. 461–464. (Cited on p. 86).
- [204] S. R. Searle. *Matrix algebra useful for statistics*. 1st ed. John Wiley & Sons, 1982. (Cited on p. 74).
- [205] G. Shafer. *A mathematical theory of evidence*. Princeton University Press, 1976. (Cited on p. 1).
- [206] M. Shinozuka. “Basic analysis of structural safety”. In: *Journal of Structural Engineering* 109.3 (1983), pp. 721–740. (Cited on pp. 4, 36).
- [207] M. Shinozuka and G. Deodatis. “Simulation of multi-dimensional Gaussian stochastic fields by spectral representation”. In: *Applied Mechanics Reviews* 49.1 (1996), pp. 29–53. (Cited on pp. 2, 18).
- [208] D. Simpson, H. Rue, T. G. Martins, A. Riebler, and S. H. Sørbye. “Penalising model component complexity: a principled, practical approach to constructing priors”. In: *Statistical Science* 32.1 (2017), pp. 1–28. (Cited on pp. 87, 88).
- [209] S. A. Sisson and Y. Fan. “Likelihood-free MCMC”. In: *Handbook of Markov chain Monte Carlo*. Ed. by S. Brooks, A. Gelman, G. L. Jones, and X.-L. Meng. Chapman & Hall/CRC, 2011. Chap. 12, pp. 313–333. (Cited on p. 3).
- [210] D. S. Sivia and J. Skilling. *Data analysis: a Bayesian tutorial*. 2nd ed. Oxford University Press, 2006. (Cited on pp. 3, 55).
- [211] J. Skilling. “Nested sampling for general Bayesian computation”. In: *Bayesian Analysis* 1.4 (2006), pp. 833–859. (Cited on pp. 59, 61).
- [212] R. C. Smith. *Uncertainty quantification: theory, implementation, and applications*. Society for Industrial and Applied Mathematics (SIAM), 2014. (Cited on p. 1).
- [213] I. M. Sobol. “Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates”. In: *Mathematics and Computers in Simulation* 55.1 (2001), pp. 271–280. (Cited on pp. 2, 116).
- [214] R. Sulecki and R. J. Conant. *Advanced mechanics of materials*. Oxford University Press, 2003. (Cited on p. 155).
- [215] T. T. Soong and M. D. Grigoriu. *Random vibration of mechanical and structural systems*. Prentice-Hall, 1993. (Cited on p. 10).
- [216] P. Spanos and R. Ghanem. “Stochastic finite element expansion for random media”. In: *Journal of Engineering Mechanics* 115.5 (1989), pp. 1035–1053. (Cited on pp. 2, 18).
- [217] P. D. Spanos, M. Beer, and J. Red-Horse. “Karhunen-Loève expansion of stochastic processes with a modified exponential covariance kernel”. In: *Journal of Engineering Mechanics* 133.7 (2007), pp. 773–779. (Cited on p. 17).

- [218] A. Spantini, A. Solonen, T. Cui, J. Martin, L. Tenorio, and Y. Marzouk. “Optimal low-rank approximations of Bayesian linear inverse problems”. In: *SIAM Journal of Scientific Computing* 37.6 (2015), A2451–A2487. (Cited on pp. 74, 116, 117, 138).
- [219] E. Spodarev, E. Shmileva, and S. Roth. “Extrapolation of stationary random fields”. In: *Stochastic geometry, spatial statistics and random fields: Models and algorithms*. Springer, 2015. Chap. 11, pp. 321–368. (Cited on p. 16).
- [220] I. Sraj, O. P. Le Maître, O. M. Knio, and I. Hoteit. “Coordinate transformation and polynomial chaos for the Bayesian inference of a Gaussian process with parametrized prior covariance function”. In: *Computer Methods in Applied Mechanics and Engineering* 298 (2016), pp. 205–228. (Cited on pp. 3, 141).
- [221] M. L. Stein. *Interpolation of spatial data: some theory for kriging*. Springer, 1999. (Cited on p. 17).
- [222] M. Stephens. “Bayesian analysis of mixture models with an unknown number of components—an alternative to reversible jump methods”. In: *The Annals of Statistics* 28.1 (2000), pp. 40–74. (Cited on p. 4).
- [223] D. Straub and I. Papaioannou. “Bayesian Updating with Structural Reliability Methods”. In: *Journal of Engineering Mechanics* 141.3 (2015), p. 04014134. (Cited on pp. 3, 21, 59, 61).
- [224] D. Straub, I. Papaioannou, and W. Betz. “Bayesian analysis of rare events”. In: *Journal of Computational Physics* 314 (2016), pp. 538–556. (Cited on pp. 2–4, 64).
- [225] A. M. Stuart. “Inverse problems: a Bayesian perspective”. In: *Acta Numerica* 19 (2010), pp. 451–559. (Cited on pp. 54, 142).
- [226] B. Sudret and A. Der Kiureghian. *Stochastic finite element methods and reliability: a state-of-the-art report*. Report on research No. UCB/SEMM-2000/08. University of California, Berkeley, 2000. (Cited on p. 6).
- [227] T. J. Sullivan. *Introduction to uncertainty quantification*. Springer, 2015. (Cited on p. 2).
- [228] S. Särkkä. *Bayesian filtering and smoothing*. Cambridge University Press, 2014. (Cited on p. 7).
- [229] M. Teixeira Parente, J. Wallin, and B. Wohlmuth. “Generalized bounds for active subspaces”. In: *Electronic Journal of Statistics* 14.1 (2020), pp. 917–943. (Cited on p. 118).
- [230] A. N. Tikhonov and V. Y. Arsenin. *Solutions of ill-posed problems*. V. H. Winston & Sons, 1977. (Cited on p. 53).
- [231] F. G. Tricomi. *Integral Equations*. Dover Publications, 1985. (Cited on p. 19).
- [232] E. Ullmann and I. Papaioannou. “Multilevel estimation of rare events”. In: *SIAM/ASA Journal on Uncertainty Quantification* 3.1 (2015), pp. 922–953. (Cited on pp. 4, 37).
- [233] F. Uribe, I. Papaioannou, J. Latz, W. Betz, E. Ullmann, and D. Straub. “Bayesian inference with subset simulation in varying dimension applied to the Karhunen–Loève expansion”. In: [eprint](#) (2020). (Cited on pp. 4, 88, 96).
- [234] F. Uribe. “Probabilistic analysis of structures using stochastic finite elements”. Master thesis. Department of Electrical, Electronic Engineering and Computing. Universidad Nacional de Colombia, 2015. (Cited on p. 151).
- [235] F. Uribe, I. Papaioannou, W. Betz, and D. Straub. “Bayesian inference of random fields represented with the Karhunen–Loève expansion”. In: *Computer Methods in Applied Mechanics and Engineering* 358 (2020), p. 112632. (Cited on pp. 3–5, 66, 87, 106).

- [236] F. Uribe, I. Papaioannou, Y. M. Marzouk, and D. Straub. “Cross-entropy-based importance sampling with failure-informed dimension reduction for rare event simulation”. In: [arXiv:2006.05496 eprint](#) (2020), pp. 1–24. (Cited on p. 6).
- [237] A. van der Vaart and H. van Zanten. “Bayesian inference with rescaled Gaussian process priors”. In: *Electronic Journal of Statistics* 1 (2007), pp. 433–448. (Cited on p. 72).
- [238] M. A. Valdebenito, H. J. Pradlwarter, and G. I. Schuëller. “The role of the design point for calculating failure probabilities in view of dimensionality and structural nonlinearities”. In: *Structural Safety* 32.2 (2010), pp. 101–111. (Cited on p. 35).
- [239] E. Vanmarcke. *Random fields: analysis and synthesis*. Revised and Expanded New. World Scientific Publishing Company, 2010. (Cited on pp. 2, 9, 14, 18, 67).
- [240] E. Vanmarcke and M. Grigoriu. “Stochastic finite element analysis of simple beams”. In: *Journal of Engineering Mechanics* 109.5 (1983), pp. 1203–1214. (Cited on pp. 2, 18).
- [241] S. Wahal and G. Biros. “BIMC: the Bayesian inverse Monte Carlo method for goal-oriented uncertainty quantification. Part II.” In: [arXiv:1911.01268 eprint](#) (2019), pp. 1–35. (Cited on pp. 34, 142).
- [242] C. Walter. “Moving particles: A parallel optimal multilevel splitting method with application in quantiles estimation and meta-model based algorithms”. In: *Structural Safety* 55 (2015), pp. 10–25. (Cited on p. 37).
- [243] F. Wang and H. Li. “The role of copulas in random fields: Characterization and application”. In: *Structural Safety* 75 (2018), pp. 75–88. (Cited on p. 15).
- [244] Z. Wang and J. Song. “Cross-entropy-based adaptive importance sampling using von Mises-Fisher mixture for high dimensional reliability analysis”. In: *Structural Safety* 59 (2016), pp. 42–52. (Cited on pp. 115, 126, 129).
- [245] L. Wasserman. “Bayesian model selection and model averaging”. In: *Journal of Mathematical Psychology* 44.1 (2000), pp. 92–107. (Cited on pp. 2, 86).
- [246] P. Whittle. “On stationary processes in the plane”. In: *Biometrika* 41 (3-4 1954). (Cited on pp. 17, 18).
- [247] K. Willcox and J. Peraire. “Balanced model reduction via the proper orthogonal decomposition”. In: *AIAA Journal* 40.11 (2002), pp. 2323–2330. (Cited on p. 2).
- [248] D. Williams. *Probability with martingales*. Cambridge University Press, 1991. (Cited on pp. 149, 150).
- [249] D. Xiu. *Numerical methods for stochastic computations. A spectral method approach*. Princeton University Press, 2010. (Cited on pp. 2, 13).
- [250] A. M. Yaglom. *Correlation theory of stationary and related random functions. Volume I: basic results*. Springer, 1987. (Cited on p. 16).
- [251] L. A. Zadeh. “Fuzzy sets”. In: *Information and Control* 8 (1965), pp. 338–353. (Cited on p. 1).
- [252] O. Zahm, P. Constantine, C. Prieur, and Y. Marzouk. “Gradient-based dimension reduction of multivariate vector-valued functions”. In: [arXiv:1801.07922v2 eprint](#) (2018), pp. 1–25. (Cited on p. 120).
- [253] O. Zahm, T. Cui, K. Law, A. Spantini, and Y. Marzouk. “Certified dimension reduction in nonlinear Bayesian inverse problems”. In: [arXiv:1807.03712v2 eprint](#) (2018), pp. 1–41. (Cited on pp. 6, 20, 118–121, 123, 127, 138, 142).

- [254] J. Zhang and B. Ellingwood. “Orthogonal series expansion of random fields in reliability analysis”. In: *Journal of Engineering Mechanics* 120.12 (1994), pp. 2660–2677. (Cited on pp. 2, 5, 18, 83).
- [255] K. M. Zuev, J. L. Beck, S.-K. Au, and L. S. Katafygiotis. “Bayesian post-processor and other enhancements of Subset Simulation for estimating failure probabilities in high dimensions”. In: *Computers & Structures* 92-93 (2012), pp. 283–293. (Cited on p. 40).
- [256] B. Øksendal. *Stochastic differential equations: An introduction with applications*. 6th ed. Springer, 2003. (Cited on p. 10).

List of Figures

1.1. Schematic representation of inverse uncertainty quantification (red flow), forward rare event simulation (blue flow), and inverse rare event simulation (red and green flows) of an engineering model with random field parameters.	5
2.1. Evolution of marginal densities and realizations of a 1D random field	10
2.2. Some marginal densities and realizations of a 2D random field	11
2.3. Cantilever beam configuration.	21
2.4. Function $\Phi_i(x)$ in (2.39c) using an exponential kernel, for different correlation lengths and first 5 terms in the KL expansion.	22
2.5. Mean and standard deviation functions of the flexibility (1st row) and deflection (2nd row) random fields	23
2.6. Correlation matrix of the flexibility random field associated to the Matérn kernel (2.24) for different correlation lengths (rows) and smoothing parameters (cols).	24
2.7. Correlation matrix of the deflection random field associated to the Matérn kernel by application of (2.42) for different correlation lengths (rows) and smoothing parameters (cols).	25
2.8. Evolution of the global variance error for the flexibility random field	26
2.9. Flexibility random field realizations for different correlation lengths and smoothing parameters of the Matérn kernel modeling the flexibility covariance.	27
2.10. Deflection random field realizations for different correlation lengths and smoothing parameters of the Matérn kernel modeling the flexibility covariance.	27
2.11. Plate with a hole. The red star marks a control point.	28
2.12. Plate example: evolution of the eigenvalues of the Matérn covariance operator with the truncation order of the KL expansion, for different smoothing parameters and correlation lengths.	29
2.13. Plate example: selected eigenfunctions of the Matérn operator for $\nu = 0.5$ and $\ell = 2$. As a comparison, we also illustrate the eigenfunctions when there is no hole in the plate.	29
2.14. Plate example: evolution of the global variance error (2.36) with the number of terms in the KL expansion k	30
2.15. Plate example: mean and standard deviation fields estimated from 1000 realizations. Using a Matérn covariance with $\nu = 0.5$ and $\ell = 2$ (1st row), $\ell = 16$ (2nd row), $\ell = 32$ (3rd row).	31
3.1. Relationships between the sequential rare event simulation methods discussed in this section.	45
3.2. Samples and contours of the intermediate optimal biasing densities at three initial simulation levels for SIS, SuS, CE and iCE methods	46
3.3. Failure probability estimate obtained by each method: for increasing dimension of the parameter space and number of samples per level	47

3.4. Source term, integrated source term, and some realizations of the hydraulic conductivity and pressure head fields using an exponential kernel with $\ell = 0.1$ (1st row) and $\ell = 0.9$ (2nd row).	49
4.1. BUS as a rejection sampling algorithm.	60
4.2. BUS-SuS algorithm: prior (blue) and posterior (light green) samples characterizing the 1D and 2D marginal distributions.	63
4.3. Schematic representation of SuS adapted for reliability updating.	65
4.4. Probability of the domain \mathcal{H} together with the prior and posterior failure probabilities for 35 simulation runs of the SuS algorithm, and increasing terms in the KL expansion k (columns).	65
4.5. Cantilever beam: true values and two sets of deflection observations.	66
4.6. Mean, standard deviation and correlation function of the posterior flexibility (rows 1-2) and deflection (rows 3-4) random fields. Using the exponential and squared exponential kernels in to model C_{FF} (with $\ell = 2.5$ and $m = 10$).	68
4.7. Posterior flexibility using an exponential kernel for the prior: 95% CI for different terms in the KL expansion, number of measurements (rows), and correlation lengths of the prior flexibility (columns).	70
4.8. Differential posterior deflection using an exponential kernel for the prior: 95% CI for different terms in the KL expansion, number of measurements (rows), and correlation lengths of the prior flexibility (columns).	71
4.9. Posterior flexibility and differential posterior deflection using a squared exponential kernel for the prior: 95% CI for different terms in the KL expansion and correlation lengths of the prior flexibility.	71
4.10. Averaged relative error mean in the posterior random fields using an exponential kernel for the prior: for the flexibility and deflection, number of measurements, different terms in the KL expansion and correlation lengths.	72
4.11. Averaged relative error variance in the prior and posterior random fields using an exponential kernel for the prior: for the flexibility and deflection (rows), number of measurements, different terms in the KL expansion, and correlation lengths.	73
4.12. Model evidence using an exponential kernel for the prior for different number of measurements (rows) and correlation lengths of the prior flexibility (columns).	75
4.13. Expected model evidence using an exponential kernel as prior covariance with different number of data points (columns) and correlation lengths of the prior flexibility.	76
4.14. Model evidence, average data fit and KLD for different correlation lengths.	77
4.15. Thin plate: structural configuration, finite element mesh, sensor locations, observations, and true fields.	78
4.16. Posterior mean of the Young's modulus for different terms in the KL expansion (columns), and prior correlation lengths (rows).	79
4.17. Posterior Young's modulus $E(x_1, x_2)$: cuts at $x_2 = 0.16$ m (1st row) and $x_1 = 0.16$ m (2nd row); 95% CI for different terms in the KL expansion and correlation lengths (columns) of the prior Young's modulus.	80
4.18. Posterior stress $\sigma_1(x_1, x_2)$: cuts at $x_2 = 0.16$ m (1st row) and $x_1 = 0.16$ m (2nd row); 95% CI for different terms in the KL expansion and correlation lengths (columns) of the prior Young's modulus.	80
4.19. Averaged relative error variance in the prior and posterior random fields for the underlying Gaussian Young's modulus E' (1st row), σ_1 stress field (2nd row), and horizontal displacement u_{x_1} (3rd row), using different terms in the KL expansion and correlation lengths.	81

5.1. RJMCMC results: the histogram of the data set is shown in the 1st row. The posterior of the models and parameters in dimension 1 and 2 are shown in the 2nd row.	91
5.2. Comparison between MwG (1st row) and step-wise sampler (2nd row) algorithms. The evolution of the potential function is shown for a single simulation.	95
5.3. Schematic representation of tBUS for nested models. We show the marginal densities of the likelihood at each dimension The maximum likelihood increases with the dimension of the parameter space.	96
5.4. Closed-form solutions for different correlation lengths in the prior flexibility random field. Left: model evidence. Center: Model/dimension prior. Right: model/dimension posterior.	100
5.5. Parameter study on the jump length ρ : acceptance rates in the MCMC algorithms.	102
5.6. Parameter study on the jump length ρ : RMSE of the investigated QoIs using the MwG sampler (1st row) and the step-wise sampler (2nd row).	102
5.7. Parameter study on the pCN scaling β : acceptance rates in the MCMC algorithms.	103
5.8. Parameter study on the pCN scaling β : RMSE of the investigated QoIs using the MwG sampler (1st row) and the step-wise sampler (2nd row)	103
5.9. Comparison between within-model BUS-SuS and across-model tBUS-SuS.	105
5.10. Influence of the number of samples per level on tBUS-SuS.	106
5.11. Estimation of the model posterior using adaptive tBUS-SuS for different correlation lengths in the prior flexibility random field: MwG (1st row); step-wise (2nd row).	107
5.12. Posterior flexibility and deflection random fields for different correlation lengths in the prior flexibility field: estimated mean and 95% CI of the best model and the averaging of models using adaptive tBUS-SuS with MwG.	107
5.13. KL coefficient samples computed by tBUS-SuS at dimension $k_{\text{best}} = 3$ ($\ell = 0.5$): prior (orange) and posterior (blue).	108
5.14. tBUS-SuS samples of the KL coefficients at dimensions $k = 1$ (left) and $k = 2$ (right). For $k = 1$, the posterior region is highlighted in gray. For $k = 2$, the contours of the log-likelihood function are also plotted.	108
5.15. Groundwater flow problem: hydraulic conductivity, hydrostatic pressure with measurement locations, source term, measured and noise-free hydraulic head. . .	110
5.16. Diffusion example: model posterior using adaptive tBUS-SuS with MwG sampler	111
5.17. Diffusion example: model posterior using adaptive tBUS-SuS with step-wise sampler	112
5.18. Diffusion example: posterior mean and standard deviation of the hydraulic conductivity random field using adaptive tBUS-SuS with the step-wise sampler, employing model choice or model mixing	113
5.19. Diffusion example: posterior mean and standard deviation of the hydrostatic pressure random field using adaptive tBUS-SuS with the step-wise sampler, employing model choice or model mixing	113
6.1. Schematic representation of a parameter $\boldsymbol{\theta} \in \mathbb{R}^3$ projected onto the FIS with $r = 2$ (in blue): global $\boldsymbol{\theta}_r \in \mathbb{R}^d$, local $\tilde{\boldsymbol{\theta}}_r \in \mathbb{R}^r$ (adapted from [57]).	122
6.2. Evolution of the iCERed method for problems <i>with</i> (top $r = 1 < d$) and <i>without</i> (bottom $r = 2 = d$) intrinsic low-dimensional structure of the parameter space. . .	127
6.3. Coefficient of variation of the failure probability estimate for example 6.4.1 using different target probabilities of failure, number of samples, smooth indicators (rows), and increasing dimension (columns).	129
6.4. Failure probability estimate for example 6.4.1 using the CE, iCE, and iCERed methods (with single Gaussian parametric family) for increasing dimension of the parameter space and different number of samples per level (columns).	129

6.5.	iCEred with refinement for example 6.4.1: number of LSF and gradient calls for different number of samples, and dimensions (columns). The $\widehat{c\mathbf{v}}$ of the failure probability estimate is also shown (in red).	130
6.6.	Eigenvalue decay (1st row) and first two eigenvectors (2nd row) of the matrix $\widehat{\mathbf{H}}$ corresponding to example 6.4.2 with $\kappa = 5$, for sample size and increasing dimension (columns).	131
6.7.	iCEred with refinement for example 6.4.2: number of LSF and gradient calls for different sample size, curvature values (rows) and dimensions (columns). The $\widehat{c\mathbf{v}}$ of the failure probability estimate is also shown (in red).	131
6.8.	Evolution of the rank for different simulation runs and $N = \{100, 250, 500, 1000\}$. 1st row: rank and smoothing parameters. 2nd row: iCEred iterations and rank.	133
6.9.	Evolution of the rank with the iCEred iterations for several simulation runs and $N \in \{100, 250, 500\}$ (columns). The star marks the final value of the rank at the given simulation run.	134
6.10.	Probability of failure estimated by SuS for increasing number of terms in the KL expansion, and probability of failure estimated by iCEred at 868 KL terms.	135
A.1.	Joint, marginal and conditional densities. Samples from a banana-shaped distribution are generated to represent the joint PDF.	149
A.2.	Simple illustration of the conditional expectation as the best least squares projector in the Hilbert space of second-order random variables.	149
B.1.	Boundary and body forces acting on a body in equilibrium. Global and local domains of a four-noded finite element.	151
B.2.	Gradient of the LSF $\nabla g(\boldsymbol{\theta})$ (6.38): comparison between finite differences (blue) and the adjoint method (red) for 6 different samples $\boldsymbol{\theta}$. We plot the first 50 components of the gradient vector.	159

List of Tables

2.1.	Truncation order of the KL expansion k that complies with 95% of the spatially averaged random field variability	26
2.2.	Truncation order of the KL expansion k that complies with 95% of the Young's modulus field variability.	31
3.1.	RMSE of failure probability estimates obtained by each method (average of $N_{\text{sim}} = 100$). For SIS, extra $0.1N$ samples are used due to burn-in.	48
3.2.	Statistics of the failure probability estimates using $\ell = 0.1$ obtained by each method.	50
3.3.	Statistics of the failure probability estimates using $\ell = 0.9$ obtained by each method.	51
4.1.	Prior and posterior probability of failure using $\ell = \{0.02, 0.16, 0.30\}$ as correlation lengths of the prior Young's modulus. The natural logarithm of the model evidence computed by BUS-SuS is also reported.	83
5.1.	Efficiency metrics (5.30) of the dimension parameter k for adaptive and standard tBUS-SuS.	104
6.1.	Summary of some approaches for dimension reduction in Bayesian inverse problems.	118
6.2.	Failure probability estimates by iCEred of the example in subsection 6.4.3: the iCEred results are shown as an average of 50 simulations.	132
6.3.	Failure probability estimates for the example 6.4.4.	134

List of Abbreviations

BUS	Bayesian Updating with Structural reliability methods
BUS-SuS	BUS with subset simulation
CLT	central limit theorem
CDR	certified dimension reduction
cv	coefficient of variation
CE	cross-entropy
CI	credible interval
CDF	cumulative distribution function
dof	degrees of freedom
FIS	failure-informed subspace
FE	finite elements
GL	Gauss–Legendre
iCE	improved cross-entropy
iCEred	improved cross-entropy with dimension reduction
IS	importance sampling
i.i.d.	independent and identically distributed
KL	Karhunen–Loève
KLD	Kullback–Leibler divergence
LIS	likelihood-informed subspace
MCMC	Markov chain Monte Carlo
MAP	maximum a posteriori
MLE	maximum likelihood estimator
MSE	mean-squared error
MH	Metropolis–Hastings
MwG	Metropolis-within-Gibbs
MC	Monte Carlo

ODE ordinary differential equation
PDE partial differential equation
PC penalized complexity
PM posterior mean
pCN preconditioned Crank–Nicolson
PDF probability density function
QoI quantity of interest
RJMCMC reversible jump MCMC
RMSE root-mean-squared error
SIS sequential importance sampling
SLLN strong law of large numbers
SuS subset simulation
tBUS trans-dimensional BUS
tBUS-SuS tBUS with subset simulation
UQ uncertainty quantification

- σ -algebra, 145
- across-model simulation, 87
- active subspace, 117
- Bayes' theorem, 54, 86
- Borel sets, 145
- BUS formulation, 59, 64
- Cauchy–Navier equations, 28, 152
- central limit theorem, 35
- certified dimension reduction, 118
- chain rule of probability, 40
- coefficient of variation, 35
- conditional expectation, 149
- correlation, 12, 146
 - coefficient, 12
 - ellipsoidal, 13
- correlation length, 17
- covariance, 12, 16, 146
- Crank–Nicolson scheme, 58
- credible interval, 56
- cross-entropy method, 41
 - failure-informed, 121
 - improved, 44
- cumulative distribution function, 146
- deformation field, 154
- detailed balance condition, 56, 89
- dimension matching condition, 89
- dimension reduction, 115
- displacement field, 152, 154
- effective sample size, 37
- elliptic equation, 48
- Euler–Bernoulli equation, 21
- expectation, 12, 146
- failure-informed subspace, 119
- finite-dimensional distribution, 10
- force vector, 156
- forward response operator, 54
- Hilbert–Schmidt kernel, 16
- Hooke's law, 152, 155
- importance sampling, 36
- independence, 146
- indicator function
 - smooth, 34
- integral equations
 - Fredholm integral equation, 19
- inverse problem, 53, 54
- isoparametric formulation, 153
- Jacobian matrix, 157
- Karhunen–Loève expansion, 18
- Kullback–Leibler divergence, 42, 88
- Lebesgue measure, 145, 147
- likelihood-informed subspace, 116
- limit-state function, 33
- Markov chain Monte Carlo, 38, 56
 - trans-dimensional, 87, 89, 92
- Matérn kernel, 16
 - exponential, 17
 - squared exponential, 17
- maximum a posteriori estimator, 56
- mean-squared function, 12
- Metropolis–Hastings, 57
- Metropolis-within-Gibbs sampler, 92
- model choice/selection, 85, 86
- model evidence, 55, 73
- model mixing/averaging, 86
- Monte Carlo, 35
- natural coordinates, 153
- nested model, 92
- overdamped Langevin dynamic equation, 58
- Poisson ratio, 152
- posterior mean, 55
- posterior predictive, 56
- potential, 55
- preconditioned Crank–Nicolson, 59
- probability density function, 10, 14, 147
 - biasing, 36, 41
 - likelihood, 54
 - model posterior, 86

- optimal biasing, 37
- penalized complexity prior, 88
- posterior, 54, 86
- posterior-failure, 34, 119
- prior, 54, 86
- probability measure, 145
- probability of failure, 33
 - updated/posterior, 64
- probability space, 9, 145
- profile function, 148

- Radon–Nikodym derivative, 147
- random field, 9
 - equivalence, 13
 - anisotropic, 12
 - continuous, 13
 - Gaussian, 14
 - homogeneous, 12
 - isotropic, 12
 - lognormal, 15
 - mean-squared continuous, 13
 - mean-squared differentiable, 14
 - non-Gaussian, 15
 - non-homogeneous, 13
 - smoothness, 13
 - weakly homogeneous, 12
- random variable, 145
- random vector, 148
- rare event simulation, 115
- rejection sampling, 59
- reliability analysis, 33, 115
- reliability updating, 63
- resampling-moving scheme, 38
- reversible jump MCMC, 89
- root-mean-squared error, 35, 101

- sample space, 145
- saturated space MCMC, 92
- sequential importance sampling, 38
- shape function, 153
- shear modulus, 152
- spatial variation, 9
- standard deviation, 12, 146
- step-wise sampler, 94
- stiffness matrix, 156
- stress field, 155
- strong law of large numbers, 35
- subset simulation, 40, 61

- tBUS formulation, 96
- total probability theorem, 86

- uncertainty
 - aleatory, 1
 - epistemic, 1
- uncertainty quantification, 1

- variable selection model, 92
- variance, 146
- virtual work principle, 152, 156

- within-model simulation, 87

- Young’s modulus, 152