Abstract. The computational complexity of MCMC methods for the exploration of complex probability measures is a challenging and important problem. A challenge of particular importance arises in Bayesian inverse problems where the target distribution may be supported on an infinite dimensional space. In practice this involves the approximation of measures defined on sequences of spaces of increasing dimension. Motivated by an elliptic inverse problem with non-Gaussian prior, we study the design of proposal chains for the Metropolis-Hastings algorithm with dimension independent performance. Dimension-independent bounds on the Monte-Carlo error of MCMC sampling for Gaussian prior measures have already been established. In this paper we provide a simple recipe to obtain these bounds for non-Gaussian prior measures. To illustrate the theory we consider an elliptic inverse problem arising in groundwater flow. We explicitly construct an efficient Metropolis-Hastings proposal based on local proposals, and we provide numerical evidence which supports the theory.

Key words. MCMC, inverse problems, Bayesian, spectral gaps, non-Gaussian

AMS subject classifications. 65C40, 60J22, 60J05, 35R30, 62F15

1. Introduction. In many applications in science and technology the main unknowns cannot be observed directly or a direct observation would be destructive. A prime example for this is computed tomography where the aim is the reconstruction of the properties of a human body given measurements at the rim of the X-ray tube. Often it is possible to model the data as the output of a mathematical model taking the unknowns as parameters. The area of inverse problems is concerned with the reconstruction of these parameters from data. Classically, this is achieved by choosing the parameter which minimises a regularized least squares functional. Whereas it is difficult to quantify the uncertainty for this method, it is straightforward in the Bayesian approach to inverse problems. The Bayesian method is based on the idea that not all parameter choices are a priori equally likely. Instead, the a priori knowledge about these parameters is modelled as a probability distribution - called the prior. By specifying the distribution of the noise, the parameters and the observed data can then be treated as jointly distributed random variables. Under certain conditions on the prior, model and noise, there exists a unique conditional distribution of the parameters given the data. This distribution is called the posterior and is an update of the prior using the data. In this way uncertainty can be quantified using the posterior variance or the posterior probability of a set in the parameter space. Usually, the posterior is only expressed implicitly as an unnormalised density with respect to the prior. For this reason sampling algorithms are used to approximate posterior expectations by the sample average. The most famous sampling algorithms are those of Metropolis-Hastings type which were introduced by N. Metropolis [39] and generalised by W. K. Hastings [23]. The idea of the Metropolis-Hastings algorithm is to add an accept-reject step to a Markov chain proposal so that the resulting Markov chain converges to the target measure. For a recent review of Markov Chain Monte Carlo (MCMC) algorithms we refer the reader to [8]. In this article, we consider target measures arising from Bayesian inverse problems. In this case the underlying mathematical models are often based on PDEs or integral operators that have to be approximated for computations. Appropriate reviews are contained in [27] and

*University of Warwick, Mathematics Institute, Zeeman Building, Coventry CV4 7AL
The former is a key reference as the Bayesian approach is applied to real world applications using MCMC and optimisation techniques. This reference shows that the resulting methods can compete with state-of-the regularisation techniques in, for example, dental X-ray imaging. Whereas this reference applies the Bayesian method to a discretised version of the inverse problem, the survey article [27] concerns the Bayesian approach to the full infinite dimensional problem which was originally developed in [11]. This approach was also taken independently in [33, 30]. We design efficient proposals for Metropolis-Hastings algorithms for target measures arising from the Bayesian approach to inverse problems.

The overall error in estimating posterior expectation using MCMC can be reduced by a better approximation to the underlying continuum model or by increasing the number of samples used in the average. Under limited computing power, this results in a trade-off between approximation and Monte-Carlo error which has quantitatively been investigated in [24]. This trade-off is influenced by the fact that for many sampling algorithms the Monte-Carlo error increases with the dimension of the state space. Thus, even if the number of samples stays fixed, a finer approximation can lead to a worse Monte-Carlo error. For the Bayesian approach to inverse problems, the prior, the posterior and some Metropolis-Hastings algorithms can be formulated on appropriate function spaces. The Monte-Carlo error of these algorithms for a fixed number of MCMC steps is only effected up to a point by the dimensionality of the state space. This insight was first properly stated by A. M. Stuart, J. Voss and P. Wiberg in [51]. These three authors pointed out the need for dimension independent sampling algorithms and constructed such methods for conditioned diffusions in the additive noise setting. This aim has been achieved by constructing well-defined sampling algorithms for measures defined via a density with respect to Gaussian measures on function spaces. This subject has then been developed further, both for conditioned diffusion and Bayesian inverse problems, and is surveyed in [5] and [12]. Recently, we have made this insight rigorous for the preconditioned Crank-Nicolson sampling algorithm by considering the convergence rate of underlying Markov chains in terms of spectral gaps [21]. In the inverse problem setting, this corresponds to posteriors arising from priors given by a density with respect to a Gaussian measure. However, having a density with respect to a Gaussian measure is not a natural assumption for all applications. In image processing for example, Bayesian methods are used to recover and reconstruct images and Gaussian priors that tend to remove or blur the edges which are supposed to be recovered. This effect is described in detail in [7]. Recovering the sharp interfaces between different rocks is also important for applications in geophysics. This has lead to the investigation of non-Gaussian priors for example in [14, 22, 31, 32]. In this paper we extend the idea of dimension independent sampling to Bayesian inverse problems with non-Gaussian priors.

In general, we assume that the target measure \( \mu \) is a Borel measure on the Banach space \( X \). It is given by

\[
\mu \propto L \mu_0
\]

where \( \mu_0 \) is the reference Borel probability measure on \( X \). The main result in this article is based on the observation that the proposal kernels for function space sampling algorithms in [12] are all reversible and have an \( L^2 \)-spectral gap when applied to the Gaussian reference measure. In this case the Monte-Carlo error can be bounded above for all \( L^2 \)-functions in terms of a lower bound on the \( L^2 \)-spectral gap. This motivates our main result which can be summarized as follows: theorem[Main Result] Suppose
that $\mu \propto \mathcal{L}_{\mu_0}$, $\mathcal{L}$ is bounded above and away from zero and the proposal Markov kernel $Q$ has an $L^2_{\mu_0}$-spectral gap. Then the Metropolis-Hastings Markov kernel has an $L^2_{\mu}$-spectral gap.

Thus, our strategy for Bayesian inverse problems will be to design proposals that are reversible with respect to the prior and that have an $L^2_{\mu_0}$-spectral gap. The Metropolis-Hastings algorithm will then perform an accept-reject step according to the likelihood in order to produce samples from the posterior. This result should be viewed in context of our recent results in [21] demonstrating that the $L^2$-spectral gap of the preconditioned Crank-Nicolson (pCN) algorithm with respect to the Gaussian reference measure is preserved for the corresponding Metropolis-Hastings kernel. In the same article, we used the Ornstein-Uhlenbeck proposal and assumed that $\mathcal{L}$ in (1.1) is log-Lipschitz. However, no global bounds on $\mathcal{L}$ were needed in order to prove the preservation of the $L^2_{\mu}$-spectral gap. In this way the main result here can be viewed as an extension to a much larger class of proposals and reference measures under partially stronger assumptions.

A related result has been proved for the Gibbs sampler applied to perturbations of Gaussian measures by Y. Amit [1]. However, it is not clear how it could be generalized for arbitrary reference measure.

Our main result stated above is proved using the method of conductance in Section 3.2. The conductance of Markov processes for its invariant distribution is the infimum of the probability of a transition from a set to its complement divided by the invariant probability of the subset. The conductance is a very important tool because Cheeger’s inequality bounds the $L^2$-spectral gap from above and below in terms of conductance. This result can be found in [36] and is related to Cheeger’s inequality in differential geometry [9]. Conductance has mainly been used for discrete settings, examples can be found in [48] and [17]. In the discrete setting, it is also possible to combine bounds on the conductances over different subsets of a graph [37]. This has been extended to sampling techniques on $\mathbb{R}$ in [25] and on $\mathbb{R}^d$ for mixtures of log-concave measures in [19].

As guiding application, we consider the posterior arising from the inverse problem of reconstructing the diffusion coefficient from noisy measurements of the pressure in a Darcy model of groundwater flow. The underlying continuum model then corresponds to a linear elliptic PDE in divergence form. The Bayesian approach to the inverse problem is taken by placing a prior based on a series expansion with uniformly distributed coefficients. In [47], well-definedness of the Bayesian inverse problem and a general Polynomial Chaos (gPC) method for approximating posterior expectations are established for this inverse problem. For a full comparison of the gPC method to MCMC algorithms, we refer the reader to [24]. In this research article, the Monte-Carlo error of the Metropolis-Hastings algorithms is bounded using convergence results for Markov chains from [10]. However, these results assume the Markov chain associated with the Metropolis-Hastings algorithm is $\phi$-irreducible. On function spaces, this condition seems only to be verifiable in special cases such as the independence sampler (IS) algorithm. The IS is an MCMC algorithm making independent proposals from one distribution. This choice of proposal leads to a poor performance especially if the posterior is concentrated. Our main result allows us to extend bounds on the Monte-Carlo error in [24] to a large class of locally moving algorithms. In particular we design the Reflection Random Walk Metropolis (RRWM) algorithm and show that it has the same asymptotic complexity as the IS algorithm using the main result of this article. Finally, we provide numerical evidence that the
RRWM and the IS algorithms are robust with respect to an increase in dimension. Furthermore, the simulations show that the RRWM algorithm is a substantial improvement over the IS algorithm especially for concentrated measures.

We give a brief exposition of Bayesian inverse problems and Metropolis-Hastings algorithms on general state spaces in Section 2. In Section 3 we introduce spectral gaps and the consequences for the sample average before we prove our main theorem. Section 4 focuses on elliptic inverse problems. We construct the RRWM sampling algorithm which satisfies the conditions of our main theorem for the previously introduced elliptic inverse problem. In Section 5 we compare the RRWM, the standard Random Walk Metropolis (RWM) and the IS algorithms using numerical simulations for the posterior arising from this particular inverse problem.

2. Review of Bayesian Inverse Problems and Metropolis-Hastings Algorithms.

This section is devoted to giving a brief summary on the relevant material on Bayesian inverse problems and to giving an introduction to Metropolis-Hastings algorithms on general state spaces. For more details we refer the reader to [49, 50] and [52, 8] respectively. The main idea of the Bayesian approach is to treat the parameters, the output of the mathematical model and the data as jointly distributed random variables. The randomness of the parameters is introduced artificially to subjectively model the uncertainty based on the a priori knowledge. The distribution of the parameters is called the prior. In the Bayesian framework the conditional probability distribution of the parameters given the noisy data is called the posterior. It is an update of the prior using the data and can be viewed as the solution to the inverse problem because it describes the a posteriori uncertainty about the parameters. The posterior is a very important tool because it can be used to

• obtain point estimates for the unknown in an inverse problem such as the posterior mean or the MAP estimator which can be related to the Tikhonov regularisation, see [15].
• quantify the uncertainty through the posterior variance or the posterior probability of a set in the parameter space.

We concentrate on the latter and note that both quantities can be written as posterior expectations. However, the calculation of posterior expectations is difficult to establish because the normalization constant is unknown. This is where Metropolis-Hastings algorithms come into the play. They can be used to approximate expectations without using the normalization constant because only ratios of the densities are needed. In order to implement a Metropolis-Hastings algorithm, the parameter space and the forward problem have to be discretised leading to a high dimensional state space. Therefore it is crucial that the algorithm performs well as the dimension increases which might be due to a a finer discretisation of the underlying continuum model. The performance of the algorithm can be measured by convergence of the underlying stochastic process to equilibrium. We survey different ways how the convergence rate is measured and provide them at the end of Section 2.2.

2.1. Bayesian Inverse Problems.

In the following we consider a general inverse problem for which the data is generated by

\[ y = G(a) + \eta \in Y. \]

Here \( \eta \) is the observational noise, \( a \in X \) is the input of the mathematical model, for example the initial condition or coefficients for a PDE, and \( G \) is the forward operator,
a mapping between the Hilbert spaces $X$ and $Y$. In this setting the inverse problem is concerned with the reconstruction of the input $a$ to the model $G$ given its noisy output, the data $y$. The problem has typically to be regularised in some way because $G$ can be non-injective and $\eta$ is unknown. Classically, this is done by choosing $a$ as the minimiser of a regularised least squares functional. Regularisation can also be approached by placing a prior $\mu_0(da)$ probability measure on $a$ containing all the a priori information. If, in addition, the forward operator $G$ and the distributions of $\eta$ is given, then $(a, y)$ can be treated as jointly varying random variables. Under mild assumptions, there exists a conditional probability measure on $a$ which is called the posterior, an update of the prior using the data. In contrast to the minimiser of a least squares functional, the posterior is continuous in the data with respect to the total variation and the Hellinger distance. The posterior is also continuous with respect to approximations of the forward model. For the precise statements of these results we refer the reader to the surveys [49] and [50]. Due to the latter result, it is possible to bound the difference between expectations calculated with respect to the posterior associated with the infinite dimensional and the discretised forward model. In Sections 2.2 and 3 we explain how the Metropolis-Hastings algorithm can be used to approximate expectations with respect to the posterior associated with the discretised forward model and how the resulting Monte-Carlo error can be bounded. In order to use Metropolis-Hastings algorithm we specify the posterior more explicitly. For finite dimensional distributions given as probability densities Bayes’ rule yields

\begin{equation}
\text{posterior} \propto \text{likelihood} \times \text{prior},
\end{equation}

more details for classical Bayesian statistics can be found in [3].

We consider a generalisation of Bayes’ rule to infinite dimension. In this article, we only consider finite dimensional data, that is $Y = \mathbb{R}^N$, but the results in [49] and [50] allow the data to be infinite dimensional as well. In the case of finite data, where the observational noise has a Lebesgue density $\rho$, the Bayesian framework can be summarised as follows

\begin{align}
\text{Prior} & \quad a \sim \mu_0 \\
\text{Noise} & \quad \eta \text{ with pdf } \rho(\eta) \\
\text{Likelihood} & \quad y|a \text{ r.v. with pdf } \rho(y - G(a)) \\
L(a) &= \rho(y - G(a)) \\
\text{Posterior} & \quad \frac{d\mu^y}{d\mu_0}(a) \propto L(a)
\end{align}

Subsequently, we drop the $y$ and hope that this does not cause any confusion for the reader. The important point to note here is that the Equations (2.1) and (2.2) are of the same form as the general target measure for the Metropolis-Hastings algorithm as in Equation (1.1) which will be reviewed in the next section.

**2.2. The Metropolis-Hastings Algorithm on General State Spaces.** The common idea of MCMC algorithms is to create a Markov chain with a prescribed invariant measure, called the target measure. Samples of this Markov chain under (mild) conditions satisfy a law of large numbers and can thus be used to approximate expectations with respect to the target measure. Under stronger conditions it is possible to control the resulting random error using a central limit theorem (CLT) or to establish bounds on the mean square error.
In this article, we consider the application of Metropolis-Hastings algorithms to Bayesian inverse problems previously introduced in Section 2.1. In order to implement Metropolis-Hastings algorithms to approximate posterior expectations such as the mean or the variance we have to discretise and approximate \( G \). Nevertheless some Metropolis-Hastings algorithms can be formulated on function spaces and it is conceivable that those perform better as the dimension of the approximation increases than those that cannot be formulated on function spaces. We have made this rigorous for the preconditioned Crank-Nicolson and the standard Random Walk Metropolis (RWM) algorithms for Gaussian prior in [21]. In the present article, we present this problem for non-Gaussian priors. For this reason we formulate the Metropolis-Hastings algorithm on general state spaces following [52].

The idea of the Metropolis-Hastings kernel is to add an independent accept-reject step to a proposal Markov kernel \( Q(x, dy) \) in order to produce a Markov kernel \( P(x, dy) \) with \( \mu \) as an invariant measure, that is

\[
\mu P = \int_X \mu(dx)P(x, dy) = \mu(dy).
\]

Subsequently, we will discuss a choice of the acceptance probability such that \( \mu \) is invariant for \( P \). Thereafter we consider the reversibility of both the proposal and Metropolis-Hastings kernel. This property is important because it yields error bounds on the sample average in combination with an \( L^2 \) spectral gap (c.f. Section 3). We will close this section by reviewing convergence results for Metropolis-Hastings algorithms.

The Metropolis-Hastings algorithm accepts a move from \( x \) to \( y \) proposed by the kernel \( Q(x, dy) \) with acceptance probability \( \alpha(x, y) \). Thus, the algorithm takes the following form

**Algorithm** Initialise \( X_0 \). For \( i=0, \ldots, n \) do:

Generate \( Y \sim Q(X_i, \cdot), U \sim U(0, \infty) \) independently and set

\[
X_{i+1} = \begin{cases} 
Y & \text{if } \alpha(X_i, Y) > U \\
X_i & \text{otherwise}
\end{cases}
\]

The transition kernel \( P(x, dy) \) associated with the Metropolis-Hastings algorithm can be written as

\[
P(x, dy) = \alpha(x, y)Q(x, dy) + \delta_x(dy) \left( 1 - \int_E Q(x, dy)\alpha(x, y) \right).
\]

If the Radon-Nikodym derivative \( \frac{d\mu(dy)Q(y,dx)}{d\mu(dx)Q(x,dy)} \) exists, then \( \mu \) is invariant for \( P \) for the choice

\[
\alpha(x, y) := \min \left( 1, \frac{d\mu(dy)Q(y,dx)}{d\mu(dx)Q(x,dy)} \right).
\]

In finite dimensions a common dominating measure is the Lebesgue measure. However, in infinite dimensions there is no equivalent of the Lebesgue measure. Furthermore, the Feldmann-Hajek theorem (c.f. [13]) implies that \( Q(x, dy) \) corresponding to a Gaussian random walk is mutually singular for different \( x \). Nevertheless, it is instructive to consider the case if there is a common dominating measure \( \lambda \), that is

\[
\mu \propto L\lambda \text{ and } Q(x, dy) = q(x, y)\lambda(dy),
\]

Subsequently, we will discuss a choice of the acceptance probability such that \( \mu \) is invariant for \( P \). Thereafter we consider the reversibility of both the proposal and Metropolis-Hastings kernel. This property is important because it yields error bounds on the sample average in combination with an \( L^2 \) spectral gap (c.f. Section 3). We will close this section by reviewing convergence results for Metropolis-Hastings algorithms.
then
\begin{equation}
\alpha(x, y) = \frac{L(y)q(y, x)}{L(y)q(x, y)} \wedge 1.
\end{equation}

Note that we can work with the unnormalised density because the acceptance probability is based on the ratio of \(L\) at \(x\) and \(y\). In fact, \(\mu\) is not only invariant for the Metropolis-Hastings kernel \(P\) but the kernel \(P\) is also reversible with respect to \(\mu\) (see [52]), which is defined subsequently.

**Definition 2.1.** A Markov kernel \(P\) is reversible with respect to a measure \(\mu\) if
\[
\mu(dx) P(x, dy) = \mu(dy) P(y, dx).
\]

If the proposal \(Q\) is reversible with respect to the prior \(\mu_0\), then (2.3) reduces to
\begin{equation}
\alpha(x, y) = \min \left(1, \frac{dL(y)\mu_0(dy)Q(y, dx)}{dL(x)\mu_0(dx)Q(x, dy)} \right) = \frac{L(y)}{L(x)} \wedge 1.
\end{equation}

The problem in designing (efficient) proposals on function spaces is that the Radon-Nikodym derivative in Equation (2.3) does often not exist. This follows from different almost sure properties of \(\mu_0(dx)\) and \(\int Q(y, dx)\mu(y)\) such as quadratic variation or regularity properties. The simplest proposal which preserves these properties is to pick \(\nu\) with the same almost sure properties and use the proposal kernel
\[Q(x, dy) = \nu(dy)\].

The resulting algorithm is called independence sampler (IS) because the proposal does not depend on the current state \(x\).

For Bayesian inverse problems it is natural to design proposals that are reversible for the prior because this preserves the almost sure properties and leads to a simple acceptance rule only involving the likelihood (c.f. Equation (2.5)). In particular this is the case for the IS algorithm with \(\nu = \mu_0\).

In general, Metropolis-Hastings algorithms are run in order to approximate \(\int \mu(dx)f(x)\) by
\begin{equation}
S_{n, n_0}(f) = \frac{1}{n} \sum_{i=n_0}^{n+n_0} f(X_i)
\end{equation}
where \(n_0\) is the burn-in corresponding to throwing away the first \(n_0\) samples in order to reduce the bias. The resulting error takes the form.
\[e_{n, n_0}(f) = \mu(f) - S_{n, n_0}(f)\].

The complexity of Metropolis-Hastings algorithms can be quantified as
\[
\text{number of necessary steps} \times \text{cost of one step}.
\]

The cost of one step is usually easy to quantify and depends on the problem at hand. The number of necessary steps depends on the prescribed error level (for example fixed width (asymptotic) confidence interval see [26], [34] and [15]) and the convergence properties of the Markov chain. If \(X_i\) were i.i.d. samples, the central
limit theorem yields that the error is of order $O(n^{-\frac{1}{2}})$. A large part of the literature is concerned with proving that this is still the case for the correlated samples of an algorithm or even bounding the leading constant in $O(n^{-\frac{1}{2}})$. The methods which are used in the literature are related to different notions of convergence of the Markov chain associated with a Metropolis-Hastings algorithm to its equilibrium. These can broadly be classified as follows [10, 15]:

1. For a metric $d$ on the space of measures, such as the total variation or the Wasserstein metric, the rate of convergence to equilibrium can be characterised through the decay of $d(\nu P^t, \mu)$ where $\nu$ is the initial distribution of the Markov chain.

2. For the Markov operator $P$ the convergence rate is given as the operator norm of $P$ on a space of functions from $X$ to $\mathbb{R}$ modulo constants. The most prominent example here is the $L^2$-spectral gap (see Section 3).

3. In the regeneration and the so-called split-chain approach the evolution of the algorithm is split into independent pieces. This can be used in order to prove central limit theorems.

The regeneration and total variation methods have been very successful in obtaining rates for finite dimensional problems. An excellent review of this is given in [12]. However, in that article it is assumed that the algorithm is $\psi$-irreducible, that is the existence of a positive measure $\phi$ such that

$$\phi(A) > 0 \Rightarrow P(x, A) > 0 \quad \forall x.$$ 

This property often fails for infinite dimensional problems because the transition probabilities tend to be mutually singular for different starting points (this is even the case for a Gaussian random walk). One exception is the IS algorithm [24].

Having introduced Bayesian inverse problems and Metropolis-Hastings algorithms on general state spaces, we are now in the position to formulate and prove the main result of this article.

3. Spectral Gap for Metropolis-Hastings algorithms. Metropolis-Hastings algorithms play an important role for the approximation of $\mu(f)$ by $S_{n,n_0}(f)$ given by Equation (2.6) with error $e_{n,n_0}(f)$. Therefore much theory is aimed at establishing (asymptotic) bounds on the error. We will first define $L^2$-spectral gaps and state the appropriate theorems from the literature that allow us to bound the error in terms of an $L^2$-spectral gap. Here lies the importance of our main theorem because it establishes an $L^2_{\mu_0}$-spectral gap for Metropolis-Hastings chains for the posterior in terms of the $L^2_{\mu_0}$-spectral gap of the corresponding proposal chain for the prior.

3.1. The $L^2$-Spectral Gap and its Implications. In order to define $L^2$-spectral gaps, we recall how a Markov kernel $P$ with invariant measure $\mu$ acts on $L^2_{\mu}(X)$. Recall that $L^2_{\mu}(X)$ is the set of all Borel-measurable functions on $X$ such that

$$\|f\|_{L^2_{\mu}}^2 = \int_X f^2(x)d\mu(x) < \infty.$$ 

The Markov kernel $P$ acts naturally on $L^2_{\mu}(X)$ as

$$Pf(x) = \int_X P(x,dy)f(y).$$ 

If $P$ is reversible, this implies that $P$ can be viewed as an operator on $L^2_{\mu}(X)$ (Jensen’s inequality) such that all its eigenvalues have modulus less than or equal to one.
The spectral gap is given by the difference between the largest eigenvalue (in terms of the modulus) which is always one because $P^1 = 1$ and the second largest eigenvalue of the operator $P$. The following definition is based on the variational characterisation of the largest eigenvalue of the linear operator $P$ modulo constants.

**Definition 3.1.** ($L^2_\mu$-spectral gap) A Markov operator $P$ with invariant measure $\mu$ has an $L^2_\mu$-spectral gap $1 - \beta$ if

$$\beta = \sup_{f \in L^2_\mu} \frac{\|Pf - \mu(f)\|_2}{\|f - \mu(f)\|_2} < 1.$$ 

The two main implications of an $L^2_\mu$-spectral gap are a CLT for $S_{n,n_0}(f)$ which implies an asymptotic bound on the error of size $O\left(\frac{1}{\sqrt{n}}\right)$ and a non-asymptotic bound on the mean square error. The latter yields non-asymptotic confidence intervals using Chebyshev’s inequality.

In the following we present the the precise statement of the CLT due to Kipnis and Varadhan [28]. The following version is taken from [35].

**Proposition 3.2.** (Kipnis-Varadhan) Consider an ergodic Markov chain with transition operator $P$ which is reversible with respect to a probability measure $\mu$ and which has an $L^2_\mu$-spectral gap $1 - \beta$. For $f \in L^2$ we define

$$\sigma^2_{f,P} = \langle \frac{1 + P}{1 - P} f, f \rangle.$$ 

Then for $X_0 \sim \mu$ the expression $\sqrt{n}(S_n - \mu(f))$ converges weakly to $\mathcal{N}(0, \sigma^2_{f,P})$. Moreover, the following inequality holds

$$\sigma^2_{f,P} \leq \frac{2\mu((f^2 - \mu(f)^2))}{(1 - \beta)} < \infty.$$ 

The non-asymptotic bounds on the mean square error is due to Rudolf [45] and take the following form

**Proposition 3.3.** Suppose that we have a Markov chain with Markov operator $P$ having an $L^2_\mu$-spectral gap $1 - \beta$. For $p \in (2, \infty]$ let $n_0(p)$ be defined by

$$n_0(p) \geq \frac{1}{\log(\beta^{-1})} \begin{cases} \frac{p}{2(p-2)} \log^2 \left(\frac{32p}{p-2}\right) \left\| \frac{d\nu}{d\mu} - 1 \right\|_{\frac{p}{2(p-2)}} & \text{if } p \in (2, 4), \\ \log(64) \left\| \frac{d\nu}{d\mu} - 1 \right\|_{\frac{p}{2(p-2)}} & \text{if } p \in [4, \infty]. \end{cases}$$ 

Then for $S_{n,n_0}$ as in Equation (3.1) and $f \in L^2_\mu$,

$$\sup_{\|f\|_2 \leq 1} \mathbb{E} \left[ \left( \mu(f) - \frac{1}{n} \sum_{i=n_0}^{n_0+n} f(X_i) \right)^2 \right] \leq \frac{2}{n(1 - \beta)} + \frac{2}{n^2(1 - \beta)^2}.$$ 

If a Metropolis-Hastings algorithm has an $L^2_\mu$-spectral gap, then the two results above can be used to derive asymptotic and non-asymptotic confidence intervals and levels for the Monte-Carlo error $\epsilon_{n,n_0}(f) = \mu(f) - S_{n,n_0}(f)$. In the next section, we will prove our main theorem that yields an $L^2_\mu$-spectral gap for a large class of Metropolis-Hastings algorithms.
3.2. Main Result. The following theorem provides an explicit lower bound on the $L^2_\mu$-spectral gap of the Metropolis-Hastings chain in terms of the $L^2_{\mu_0}$-spectral gap of the proposal chain and in terms of the bounds on the density of the posterior with respect to the prior.

**Theorem 3.4 (Main Theorem).** Suppose that the proposal kernel $Q$ has an $L^2_{\mu_0}$-spectral gap $1 - \beta_{\text{prop}}$ and that the target measure takes the form

$$\mu = \frac{L \mu_0}{Z}$$

where $L_* := \inf L \leq L \leq \sup L =: L^*$. Then the $L^2_\mu$-spectral gap $1 - \beta$ of $P$ satisfies

$$\left( \frac{L_*}{L^*} \right)^4 \frac{(1 - \beta_{\text{prop}})^2}{8} \leq 1 - \beta \leq 2 \left( \frac{L_*}{L^*} \right)^2 \sqrt{1 - \beta_{\text{prop}}}.$$ 

**Proof.** This result is based on bounds of the $L^2$-spectral gap in terms of the conductance, which is defined by

\begin{equation}
C = \inf_{\mu(A) \leq \frac{1}{2}} \frac{\int_A P(x, A^c) d\mu(x)}{\mu(A)}
\end{equation}

to obtain bounds on the spectral gap by Cheeger’s inequality (see [33, 38])

\begin{equation}
\frac{C^2}{2} \leq 1 - \beta \leq 2C.
\end{equation}

In order to derive the appropriate inequality, we first replace $\mu$ by $\mu_0$ using the boundedness of $L$

$$\frac{\int_A P(x, A^c) d\mu(x)}{\mu(A)} \geq \frac{L_*}{L^*} \frac{\int_A P(x, A^c) d\mu_0(x)}{\mu_0(A)} \geq \left( \frac{L_*}{L^*} \right)^2 \frac{\int_A Q(x, A^c) d\mu_0(x)}{\mu_0(A)}.$$ 

The second step is justified because the acceptance probability is bounded below by $\frac{L_*}{L^*}$. Using Equation (3.3) for the proposal gives rise to

\begin{equation}
\frac{\int_A P(x, A^c) d\mu(x)}{\mu(A)} \geq \left( \frac{L_*}{L^*} \right)^2 \frac{(1 - \beta_{\text{prop}})}{2}.
\end{equation}

Taking the infimum and using Equation (3.3) again gives

\begin{equation}
\left( \frac{L_*}{L^*} \right)^4 \frac{(1 - \beta_{\text{prop}})^2}{8} \leq 1 - \beta.
\end{equation}

The other inequality can be derived similarly.

This result highlights the insight that the reference measure is crucial for designing efficient sampling algorithms on function spaces. A typical example would be the use of a Markov chain that has an $L^2_{\mu_0}$-spectral gap where $\mu_0$ is the prior of a Bayesian problem. If the likelihood is bounded, then the Metropolis-Hastings algorithm with this chain as the proposal has an $L^2_\mu$-spectral gap with $\mu$ being the posterior. However,
the result is not limited to this situation because \( \mu_0 \) and \( \mu \) can be arbitrary measures such that the density of \( \mu \) with respect to \( \mu_0 \) is bounded.

**Remark 1.** For a fixed target measure a larger \( L^2_{\mu_0} \)-spectral gap of \( Q \) implies a larger lower bound on the \( L^2_{\mu} \)-spectral gap of \( P \). In particular the largest lower bound is achieved for the IS algorithm. It is import to note that this does not imply that this choice leads to the largest spectral gap for \( P \). In fact, the simulations in Section 5 suggest otherwise.

**Remark 2.** The results obtained in [1] for the Gibbs sampler applied to a perturbation of a Gaussian measure suggest that the sharper inequalities

\[
\left( \frac{L_s}{L_s^*} \right) (1 - \beta_{\text{prop}}) \leq 1 - \beta \leq \left( \frac{L_s^*}{L_s} \right) (1 - \beta_{\text{prop}}).
\]

might hold. This seems an interesting question for further investigation.

4. Application to an Elliptic Inverse Problem. The theoretical result of the previous section was motivated by studying the reconstruction of the diffusion coefficient \( a \) given noisy observations of the pressure \( p \). We approach this inverse problem in the Bayesian framework by imposing a prior based on a series expansion with uniform coefficients.

Firstly, we will set up the forward problem and review the literature on the resulting inverse problem focusing on the Bayesian approach. Secondly, we will describe our prior, impose Gaussian observational noise and then show that the resulting posterior has a bounded density with respect to this prior. The rest of the section is devoted to constructing appropriate proposal kernels and proving a lower bound on their \( L^2_{\mu_0} \)-spectral gap. Thus, our main theorem implies a lower bound on the \( L^2_{\mu_0} \)-spectral gaps of the corresponding Metropolis-Hastings algorithms, in particular the RRWM algorithms. Whereas the simulations in Section 5 suggest that the RRWM outperforms the IS algorithm, our main result guarantees a lower bound on the spectral gap that is of the same order. Moreover, the construction of the RRWM is important in its own right because it constitutes an efficient sampling algorithm for elliptic inverse problems.

4.1. The Underlying PDE and Well-Definedness of the Forward Model. The forward problem is based on the relation between \( p \) and \( a \) modelled by the following elliptic PDE with Dirichlet boundary conditions

\[
\begin{align*}
-\nabla \cdot (a \nabla p) &= g(x) \quad \text{in } D \\
p &= 0 \quad \text{on } \partial D
\end{align*}
\]

where \( D \) is a bounded domain in \( \mathbb{R}^d \) and \( p \) and \( a \) are scalar functions on \( D \). We assume that \( a^* \geq a(x) \geq a_* > 0 \) for all almost every \( x \in D \). The subset of \( L^\infty(D) \)-functions that satisfy this condition is denoted by

\[
L^\infty_+: = \{ u \in L^\infty \mid \text{ess inf } u > 0 \}.
\]

If, additionally, \( g \) is in the Sobolev space \( H^{-1} \), then the solution operator \( p(x; a) : L^\infty_+ \to H^1 \), mapping to the unique weak solution of \( (4.1) \), is well-defined (for details we refer the reader to [50]). We suppose that the forward operator \( \mathcal{G} \), giving rise to the data, is based on the solution operator as follows

\[
\mathcal{G}(a) = \mathcal{O}(p(\cdot; a))
\]
where $\mathcal{O}$ is called the observation operator. Additionally, we suppose that is is equal to $\mathcal{O} = (l_1, \ldots, l_N)$ with $l_i \in H^{-1}$.

The inverse problem associated with the above forward problem is well-known and it is particularly relevant in oil reservoir simulations and the modelling of groundwater flow, see for example [38]. A survey of classical least squares approaches to this inverse problem can be found in [29] for which recently error estimates have been obtained in [53]. A rigorous Bayesian formulation of this inverse problem with log-Gaussian priors and Besov priors is given in [16] and [14] respectively, both are reviewed in [50]. There is also an extensive literature in the uncertainty quantification community studying how uncertainty propagates through the forward model. This can be investigated by considering different realisations of the input. This approach can be combined with the finite element [18] and Galerkin methods [2] used to approximate the underlying equation. For the elliptic problem under consideration, this has been studied in [10]. In fact, it can be more efficient to use generalised Polynomial Chaos (gPC) [46] instead of Monte Carlo methods. Recently, gPC methods also have been applied to the elliptic inverse problem considered in this article [47, 24]. Since gPC often suffers from a large constant and has only been developed for a few inverse problems, it is important to construct efficient samplers tailored for the prior and likelihood at hand. Moreover, also MCMC can be speeded up using the multi level approach. The expectation of interest is written as difference corresponding to a finer and finer discretization such that more MCMC samples are used for coarser discretisations [24].

### 4.2. Prior on an expansion of the Diffusion Coefficient

Following [24, 47] we choose a prior on the coefficients $(u_1, \ldots, u_J)$ for $J \in \mathbb{N} \cup \{\infty\}$ giving rise to the diffusion coefficient

$$a(u)(x) = \bar{a}(x) + \sum_{j \in J} \gamma_j u_j \psi_j(x)$$

(4.3)

where $\|\psi_i\|_{L^\infty} = 1$. We suppose that $u_i \overset{i.i.d.}{\sim} U(-1, 1)$ which corresponds to a prior given by

$$\mu_0^J = \bigotimes_{j=1}^J U(-1, 1).$$

Additionally, the choice of $\gamma_i$ is supposed to satisfy $a_* = \inf \bar{a} - \sum_{j=1}^J \gamma_j > 0$ for all choices of $J$. In particular $\{\gamma_i\}$ have to be summable, $a \in L^\infty_+ \mu_0$-a.s. and the solution operator $p$ is well-defined for $\mu_0$ almost every $a(u)$.

We would like to note that similar probability measures have been studied for the propagation of uncertainty in [10].

### 4.3. Bounds on the Density of the Posterior

We suppose the data is given by

$$y = \mathcal{G}(a(u)) + \eta$$

where $\eta \sim \mathcal{N}(0, \Gamma)$. The well-definedness of the corresponding posterior for $J \in \mathbb{N} \cup \{\infty\}$ has been proven in [17] and [50]. It takes the form

$$\frac{d\mu}{d\mu_0} \propto \exp \left( -\frac{1}{2} \|y - \mathcal{G}(a)\|_{\Gamma}^2 \right).$$
We also know that
\[
\|G(a)\|_F \leq \|\Gamma\|_2 N \max_i \|l_i\|_{H^{-1}} \sup_{-\frac{1}{n_i} \leq 1} \|p(a)\|_{H^1} \leq C \|\Gamma\|_2 N \max_i \|l_i\|_{H^{-1}} a_*^{-1}
\]
where \(a_* = \text{ess inf } a\). Note that \(C\) depends on \(N\) (see Equation (4.2)) but can be chosen uniformly in \(J\). This gives rise to the following upper and lower bounds on the likelihood
\[
L^* = 1
\]
\[
L_* = \exp \left( -2C^2 \|\Gamma^{-1}\|_2^2 N^2 \left( \max_i \|l_i\|_{H^{-1}} \right)^2 a_*^{-2} \right).
\]

### 4.4. Spectral Gaps for the Prior and the Posterior.

In order to apply our main result, we have to choose a proposal kernel \(Q\) that is reversible and has an \(L^2_{\mu_0}\)-spectral gap with respect to \(\mu_0 = U(-1,1)^J\).

Given any kernel that has an \(L^2_{U(-1,1)}\)-spectral gap we may apply the tensorisation property of \(L^2\)-spectral gaps (see e.g. [3, 20]) to conclude that applying this kernel to each component yields a kernel with the same spectral gap for \(U(-1,1)^J\). Whereas we construct the one dimensional proposal distributions explicitly below, it is worth pointing out that it is possible to obtain an appropriate one-dimensional proposal using the Metropolis-Hastings kernel for \(U(-1,1)\) with a one-dimensional proposal distribution. Then the resulting Markov kernel is uniformly ergodic under mild assumptions [42] implying an \(L^2_{U(-1,1)}\)-spectral gap [44]. Note that the resulting proposal on \([-1,1]^J\) can be accepted even if some of the one-dimensional Metropolis-Hastings algorithms have rejected their proposal.

Alternatively, a Markov kernel with \(L^2_{U(-1,1)}\)-spectral gap can be obtained by considering a random walk with symmetric proposal
\[
Q_{\text{RW}}(x,dy) = q(x-y)dy
\]
\[
Q_{\text{RW}}(x,dy) = \mathcal{L}(x + \xi), \text{ where } \xi \sim q
\]

and repeatedly reflecting \(y\) at the boundaries \(-1\) and \(1\). The reflection can be represented according to the following function
\[
R(x) = \begin{cases} 
 y & y \leq 1 \\
 2 - y & 1 < y < 3 \\
 -4 + y & 3 \leq y \leq 4
\end{cases}
\]

where \(y = x \mod 4\).

We call the Metropolis-Hastings algorithm based on tensorisations of this proposal Reflection Random Walk Metropolis (RRWM) algorithm. In this way we can write the proposal kernel as
\[
Q^{\text{RRWM}}(x,dy) = \mathcal{L}(R(x + \xi))
\]

where \(\xi \sim q\). Its density with respect to the Lebesgue measure takes the following form
\[
Q^{\text{RRWM}}(x,dy) = \sum_{k \in \mathbb{Z}} q(x - y + 4k) + q(x + y + 4k + 2).
\]
The proposal kernel $Q_{RRWM}$ is reversible with respect to $U(-1, 1)$ because $q^{RRWM}(x, y) = q^{RRWM}(y, x)$. In the following we consider the RRWM with uniform random walk ($\xi \sim U(-\epsilon, \epsilon)$) and with standard random walk ($\xi \sim N(0, \epsilon^2)$) which we call Reflection Uniform Random Walk Metropolis (RURWM) and Reflection Standard Random Walk Metropolis (RSRWM), respectively. In contrast to the RSRWM, the proposal of the RURWM has a density $q^{RURWM}$ with closed form. For $\epsilon < 1$ it is given by

$$
q^{RURWM}_\epsilon(x, y) \propto \begin{cases} 
1 & -1 \leq x, y \leq 1, |x - y| \leq \epsilon, y > -x - 2 + \epsilon, y < -x + 2 - \epsilon \\
2 & -1 \leq x, y \leq 1, y \leq -x - 2 + \epsilon \text{ or } y \geq -x + 2 - \epsilon \\
0 & \text{otherwise}
\end{cases}.
$$

The following result shows that the RRWM has an $L^2$-spectral of order $\epsilon$.

**Theorem 4.1.** The $L^{2n}$-spectral gap $1-\beta_\epsilon$ of the $Q^{RURWM}_\epsilon$ with respect to $U(-1, 1)$ satisfies

$$
1 - \beta_\epsilon \geq 1 - \frac{4}{5} |\frac{1}{2}| \geq \frac{4}{25}.
$$

**Proof.** We will first prove that for $n = \left\lceil \frac{3}{\epsilon} \right\rceil$ there is an $\epsilon$-independent lower bound on the spectral gap of $(Q^{RURWM}_\epsilon)^n$. This is achieved by showing that $[-1, 1]$ is a small set for $(Q^{RURWM}_\epsilon)^n$. The resulting lower bound on the spectral gap is then transferred from to $(Q^{RURWM}_\epsilon)^n$ to $Q^{RURWM}_\epsilon$ using the spectral theorem.

By $\tilde{q}_\epsilon(x, y) = 1_{\{x-\epsilon \leq y \leq x+\epsilon\}}$ we denote the density of the unreflected random walk. An induction shows that the density $\tilde{q}_{\epsilon,n}$ of the $n$-step transition takes the form

$$
\tilde{q}_{\epsilon,n}(x, y) = (n(n+1)/2)^{-1} \sum_{k=0}^{n-1} (n-k) \left( 1_{\{k+1, k\epsilon\}}(y) + 1_{\{k\epsilon, (k+1)\epsilon\}}(y) \right).
$$

Hence in particular

$$
\tilde{q}_{\epsilon,n}(x, y) \geq (n(n+1)/2)^{-1} 1_{\{x-\frac{\epsilon}{2} \leq y \leq x+\frac{\epsilon}{2}\}} \frac{n}{2}.
$$

Choosing $n = \left\lceil \frac{3}{\epsilon} \right\rceil$, we note that

$$
\tilde{q}_{\epsilon,n}(x, y) \geq (n(n+1)/2)^{-1} \frac{n}{2} \geq \frac{4}{2(\frac{3}{\epsilon} + 1)^2} \frac{\epsilon}{2} \geq \frac{1}{10} \quad \forall y \in [-1, 1].
$$

The density $q^{RURWM}_{\epsilon,n}$ of $(Q^{RURWM}_\epsilon)^n$ is pointwise larger than $\tilde{q}_{\epsilon,n}$ because each $y$ might have several preimages under $R$ (c.f. Equation (4.4)). Therefore $q^{RURWM}_{\epsilon,n}$ also satisfies

$$
q^{RURWM}_{\epsilon,n}(x, y) \geq \tilde{q}_{\epsilon,n}(x, y) \geq \frac{1}{10} \quad \forall y \in [-1, 1].
$$

Thus, the state space $[-1, 1]$ is a small set and Theorem 8 in [42] implies that

$$
\left\| (Q^{RURWM}_\epsilon)^n (d, dy) - U(-1, 1) \right\|_{TV} \leq \left( \frac{4}{5} \right)^k.
$$
For reversible Markov processes uniform ergodicity implies a spectral gap of the corresponding size, see for example [45]. Hence \( (Q\text{_{RURWM}})^n \) has a spectral gap of size \( 1 - \hat{\beta} \) with \( \hat{\beta} = \frac{4}{\pi} \). The spectral theorem for self-adjoint operators now implies that \( Q\text{_{RURWM}} \) has a spectral gap of size

\[
1 - \beta = 1 - \hat{\beta}^\frac{1}{n}.
\]

The last inequality in this theorem is due to the mean value theorem.

Theorem 4.1 provides a lower bound on the spectral gap for the proposal of the RURWM algorithm. In a similar manner it can also be shown that the proposal of the RSRWM has an \( L^2_{\mu_0} \)-spectral gap of order \( \epsilon \). The lower bound on the spectral gap of the resulting Metropolis-Hastings algorithm follows now from our main theorem.

Corollary 4.2. Let \( Q \) be a Markov kernel that has an \( L^2_{\mu_{(\sigma,1)}} \)-spectral gap

\[
1 - \beta_{\text{prop}} \in \mathbb{N} \cup \{\infty\} \quad \text{and} \quad Q_J = \bigotimes_{j=1}^J Q(a_j, d\tilde{a}_j).
\]

Then the Metropolis-Hastings transition kernel \( P_J \) for \( \mu_J \) with proposal \( Q_J \) has an \( L^2_{\mu_J} \)-spectral gap \( 1 - \beta_J \) and there is a \( J \)-independent lower bound of the form

\[
1 - \beta \geq \exp \left( -8C^2 \| \Gamma^{-1} \| N^2 \left( \max_i \| l_i \|_{H^{-1}} \right)^2 a_\ast^{-2} \right) (1 - \beta_{\text{prop}})^2.
\]

In this section, we have constructed the RRWM algorithm for the elliptic inverse problem with prior based on a series expansion with uniformly distributed coefficients. In the next section, we will compare this algorithm to the IS and RWM algorithms using simulations.

5. Numerical Comparison of Different MCMC Algorithms for a particular Elliptic Inverse Problem. In this section, we apply the Random Walk Metropolis (RWM) algorithm, the Importance Sampling (IS) and the Reflection Random Walk Metropolis (RRWM) algorithms to the posterior arising from the elliptic inverse problem considered in Section 4. We use the resulting simulations to illustrate the following two aspects:

- On the one hand the acceptance probability of the standard RWM algorithm decreases quickly as the dimension of the state space increases. On the other hand, the relation between the step size and the acceptance probability of the RRWM algorithm is not affected by the dimension.
- The performance of the IS algorithm is only affected up to a point by the dimension \( J \) of the state space. However, it does not perform well for concentrated target measures. However, choosing an appropriate step size for the RRWM algorithm leads to a good performance.

We first describe the implementation of the forward model, the choice of prior and the implementation of the IS, the RWM and the RRWM algorithms. The remaining part of the section is then divided into presenting the dependence of the relationship between step size and acceptance rate on the dimension as well as the decay of the autocorrelation.

5.1. The Setup. We consider the elliptic inverse problem as described in Section 4 on the domain \( D = [0, 1] \). In this case there is an explicit formula linking the pressure \( p \) and the diffusion coefficient \( a \) which has been implemented using a trapezoidal rule. We choose the prior as in Equation (4.3) on the coefficients \( u_i \), that is

\[
\mu_0 J = \bigotimes_{j=0}^J \mathcal{U}(-1, 1).
\]
These coefficients give rise to the diffusion coefficient

\[ a(u)(x) = \bar{a}(x) + \sum_{j=0}^{J} \gamma_j u_j \psi_j(x) \] where \( a_j \sim \mathcal{U}(-1, 1) \).

For our simulations we set

\[ \bar{a}(x) = 4.38. \]

\[ \psi_{2j-1}(x) = \cos(2\pi j x), \quad \gamma_{2j} = \frac{1}{j^2}, \quad K \geq j \geq 1 \]

\[ \psi_{2j}(x) = \sin(2\pi j x), \quad \gamma_{2j-1} = \frac{1}{j^2}, \quad K \geq j \geq 1 \]

\[ \psi_0(x) = 1, \quad \gamma_0 = 1. \]

Note that the lower bound \( a(x) \geq 1 \) is independent of \( J = 2K \). Data corresponds to evaluations of the pressure uniformly spaced at distance \( d \) apart

\[ y = G(a^\uparrow) + \eta = (p(id) + \eta_i)_{i=0}^{1/d} \]

where \( \eta \sim \mathcal{N}(0, \sigma^2 I) \) and \( a^\uparrow \) is fixed input which is generated a draw from the prior.

Subsequently, we consider the IS, RWM, RURWM and RSRWM algorithms with the following proposal kernels

\begin{align}
(5.1) & \quad Q^{\text{IS}}(x, dy) = \mu_0(dy) \\
(5.2) & \quad Q^\epsilon_{\text{RWM}}(x, dy) = \mathcal{N}(x, \epsilon I_{d \times d}) (dy) \\
(5.3) & \quad Q^\epsilon_{\text{RURWM}}(x, dy) = \otimes_{i=1}^{d} L(R(x + \epsilon \xi)), \xi \sim \mathcal{U}(-1,1) \\
(4.4) & \quad Q^\epsilon_{\text{RSRWM}}(x, dy) = \otimes_{i=1}^{d} L(R(x + \epsilon \xi)), \xi \sim \mathcal{N}(0,1).
\end{align}

Note that the Metropolis-Hastings acceptance ratio, as described in Section 2, implies that the RWM algorithm simply rejects any proposal outside the unit cube.

### 5.2. Acceptance Probabilities for the RWM and RURWM Algorithms.

In Figure 5.1 we have plotted the acceptance probability against the step size for the RWM, RURWM and RSRWM algorithms for different choices of \( K \). The target for both is the posterior arising from 33 measurements with \( \sigma = 0.05 \) with artificial data. The step size parameter \( \epsilon \) affects the performance of all three algorithm. On the one hand large step sizes are beneficial because the algorithm can explore the state space more quickly whereas they lead to a small acceptance ratio (see Figure 5.1). On the other hand small step sizes lead to a high acceptance ratio but to highly correlated samples. The IS algorithm does not have a step size parameter and its average acceptance probability does not depend on the dimension. For this choice of parameters it is approximately 4.4%.

Figure 5.1 clearly illustrates that the acceptance probability of the RWM algorithm for a fixed step size deteriorates as the dimension increases. One reason for the decay of the acceptance probability of the RWM algorithm is that the probability of the proposal lying outside \([0,1]^d\) increases to 1 as \( d \to \infty \). Moreover, there is no visible impact of the dimension on the acceptance probability for the RURWM and RSRWM algorithms.
Figure 5.1: Dependence of the acceptance probability on the dimension

(a) Acceptance rate vs. step size for the RWM algorithm

(b) Acceptance rate vs. step size for the RURWM algorithm

(c) Acceptance rate vs. step size for the RSRWM algorithm
5.3. Autocorrelation of the IS, the RWM, the RURWM and the RSRWM Algorithms. Even though our lower bound on the spectral gap is smaller for the RRWM algorithms than for the IS algorithm (cf. Remark [1]), the numerical results in this section suggest that the RRWM algorithms outperforms the IS algorithm especially if \( \mu \) is peaked. The peakedness of \( \mu \) is achieved by observing \( p \) on a fine mesh with small noise \((dx = 0.03 \text{ and } \sigma = 0.03)\).

The computational cost of both algorithms is nearly the same because the cost of computing the likelihood is more expensive than generating the proposal, which is slightly more expensive for the RRWM algorithm. Subsequently, we compare the RWM, the IS, the RURWM and the RSRWM algorithm by plotting the autocorrelation. We consider \( K = 25 \) \((K = 250)\) corresponding to an expansion with 25 sine and 25 (250) cosine coefficients and a constant term thus giving rise to a 51 \((501)\) dimensional problem.

In order to compare the RWM and RRWM algorithms in a fair way we choose the step size \( \epsilon \) for in a way to get an acceptance rate close to 0.135. This is motivated for the RWM algorithm by the optimal scaling results in [41]. The optimality of this acceptance rate is indicated by proving that the properly rescaled samples converge to a Langevin diffusion whose time scale depends on the acceptance rate of the RWM algorithm. An acceptance rate of 0.135 corresponds to the largest time scale and thus to a quicker convergence to equilibrium of the Langevin diffusion. For the RRWM algorithms the acceptance rate is not affected by the choice of \( J \). However, it is reasonable to choose a step size with acceptance probability bounded away from one and zero.

For the RRWM algorithms we know that the spectral gap is bound below and thus asymptotic variance of the CLT (c.f. Proposition [3.2]) bound above. The asymptotic variance is equal to the integrated autocorrelation [40].

This is confirmed by the simulations for \( d = 0.1 \) and \( \sigma = 0.1 \) presented in Figure 5.2 which shows that the autocorrelation of the RURWM, the RSRWM and the IS algorithm is only affected up to a point by the dimension of the state space. In contrast, the autocorrelation of the RWM decays much slower for the 501 dimensional state space as for the 51 dimensional state space. In Figure 5.3 we consider the decay of the autocorrelation of the IS, the RWM, the RURWM and the RSRWM algorithms for more observations and lower observational noise \((d = 0.04 \text{ and } \sigma = 0.03)\). This has the effect that the measure concentrates in smaller regions of the state space making it harder to sample from. Figure 5.3 illustrates that the RURWM and the RSRWM algorithms can be tuned to work well for concentrated target measures such as this measure whereas the IS algorithm, even though dimension independent, behaves poorly.

For a fixed step size the RWM algorithm deteriorates as the dimension increases because the probability that one component steps outside \([-1, 1]\) converges to one. If the step size is scaled to zero appropriately the performance of the RWM algorithm deteriorates slower but for a large enough state space even the IS algorithm outperforms the RWM algorithm in any case. The reason for this is that Corollary 4.2 yields a dimension independent lower bound on the performance of the IS, RURWM and RSRWM algorithms.

6. Conclusion and Avenues of Further Research. In this article, we have shown that it is possible to transfer \( L^2 \)-spectral gaps from the proposal Markov kernel to the Metropolis-Hastings Markov kernel. This yields theoretical bounds for a large class of proposals for non-Gaussian measures on function spaces. Our main as-
Our main result justifies the use of sampling methods other than the IS algorithm for the Bayesian elliptic inverse problem considered above. However, our bounds do not show that locally moving algorithms, as the RURWM and RSRWM algorithms designed in Section 4, are asymptotically better than the IS algorithm. Comparing two sampling algorithms is difficult since it depends on the specific target. Moreover, the performance also depends on the choice of the parameters for example the step size of the algorithms. Nonetheless, rigorously showing that the RURWM and
RSRWM algorithms outperforms the IS algorithm, even in a special case, would be an interesting result.

Moreover, the range of the posterior density goes to infinity as the variance of the noise goes to zero. This suggests that sampling methods perform worse and worse as the observational noise goes to zero. Getting precise asymptotics of this behaviour would lead to a better understanding of the performance of sampling algorithms for Bayesian inverse problems.

As mentioned in Section 4, the proposal kernels of the RURWM and RSRWM algorithms are based on a tensorisation of Markov kernels for the uniform distribution on $[-1,1]$. It is also interesting to consider tensorisation of Metropolis-Hastings kernels for the uniform distribution on $[-1,1]$. Whereas we used the explicit structure of the prior, an interesting direction for more complicated priors is to use Metropolis-Hastings chains or combinations, such as tensorisation. This can lead to good proposals for another Metropolis-Hastings chain. Note that even if some of the Metropolis-Hastings algorithms in the tensorisation reject, the overall proposal can still be accepted. A deeper investigation of this approach can lead to a better understanding and guidelines for the design of efficient proposals. An interesting special case are MCMC algorithms for Bayesian inverse problems formulated on the coefficients of a Fourier series expansion. Usually the coefficients corresponding to high frequencies have only little impact on forward problem and hence the inverse problem. Developing proposals that exploit this phenomenon should also be pursued.

In this article, we considered the application of Metropolis-Hastings algorithms to the Bayesian approach to an elliptic inverse problem. A particular interesting extension would be to consider a multi-scale diffusion coefficient because there is interest in the fine and coarse scale properties of the permeability for example in subsurface geophysics. Homogenization results imply that different combinations of fine and coarse scales lead to effectively the same homogenized problem thus leading to a lack of identifiability. This also seems to be a very interesting idea.

Acknowledgments. The author would like to thank Professor Martin Hairer and Professor Andrew Stuart. SJV is grateful for the support of an ERC scholarship.

REFERENCES

[1] Y. Amit. Convergence properties of the Gibbs sampler for perturbations of Gaussians. *Ann. Statist.*, 24(1):122–140, 1996.
[2] I. Babuska, R. Tempone, and G. E. Zouraris. Galerkin finite element approximations of stochastic elliptic partial differential equations. *SIAM J. Numer. Anal.*, 42(2):800–825, 2004.
[3] D. Bakry. Functional inequalities for Markov semigroups. *Probability measures on groups: recent directions and trends*, pages 91–147, 2006.
[4] J.M. Bernardo and A.F.M. Smith. *Bayesian Theory*. Wiley, 1994.
[5] Alexandros Beskos, Konstantinos Kalogeropoulos, and Erik Pazos. Advanced MCMC methods for sampling on diffusion pathspace. *Stochastic Process. Appl.*, 123(4):1415–1453, 2013.
[6] N. Bou-Rabee and M. Hairer. Non-asymptotic mixing of the MALA algorithm. *IMA J. Numer. Anal.*, (33):pp. 80–110, 2013.
[7] C. Bouman and K. Sauer. A generalized Gaussian image model for edge-preserving MAP estimation. *IEEE Trans. Image Process.*, 2(3):296–310, 1993.
[8] Steve Brooks, Andrew Gelman, Galin Jones, and Xiao-Li Meng. *Handbook of Markov Chain Monte Carlo*. Chapman and Hall/CRC, 2011.
[9] J. Cheeger. A lower bound for the smallest eigenvalue of the Laplacian. In *Problems in analysis (Papers dedicated to Salomon Bochner, 1969)*, pages 195–199. Princeton Univ. Press, Princeton, N. J., 1970.
[10] A. Cohen, R. A. Devore, and C. Schwab. Convergence rates of best $N$-term Galerkin approximations for a class of elliptic sPDEs. *Found. Comput. Math.*, 10(6):615–646, 2010.
[11] S. L. Cotter, M. Dashti, J. C. Robinson, and A. M. Stuart. Bayesian inverse problems for functions and applications to fluid mechanics. *Inverse Probl.*, 25(11):115008, 43, 2009.

[12] S. L. Cotter, G. O. Roberts, A. M. Stuart, and D. White. MCMC methods for functions: Modifying old algorithms to make them faster. *ArXiv preprint 1202.0709*, 2011. to appear Stat. Sci.

[13] Giuseppe Da Prato and Jerzy Zabczyk. *Stochastic equations in infinite dimensions*, volume 44 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1992.

[14] M. Dashti, S. Harris, and A. M Stuart. Besov priors for Bayesian inverse problems. *Inverse Probl. Imaging*, 6:183–200, 2012.

[15] M. Dashti, K. J. H. Law, A. M. Stuart, and J. Voss. MAP Estimators and Posterior Consistency in Bayesian Nonparametric Inverse Problems. *ArXiv preprint 1303.4795*, 2013.

[16] M. Dashti and A. M. Stuart. Uncertainty quantification and weak approximation of an elliptic inverse problem. *SIAM J. Numer. Anal.*, 49:2524–2542, 2011.

[17] P. Diaconis and D. Stroock. Geometric bounds for eigenvalues of Markov chains. *Ann. Appl. Probab.*, 1(1):36–61, 1991.

[18] Roger G. Ghanem and Pol D. Spanos. *Stochastic finite elements: a spectral approach*. Courier Dover Publications, 2003.

[19] Y. Guan and S. M. Krone. Small world MCMC and convergence to multi-modal distributions: from slow mixing to fast mixing. *Ann. Appl. Probab.*, 17(1):284–304, 2007.

[20] Alice Guionnet and Boguslaw Zegarlinski. Lectures on logarithmic Sobolev inequalities, volume 1801 of *Séminaire de Probabilités, XXXVI*. Springer, 2002.

[21] M. Hairer, A. M. Stuart, and S. J. Vollmer. Spectral Gaps for a Metropolis-Hastings Algorithm in Infinite Dimensions. *ArXiv preprint 1112.1392*, 2011.

[22] T. M. Hansen, K. S. Cordua, and K. Mosegaard. Inverse problems with non-trivial priors: Efficient solution through sequential Gibbs sampling. *Comput. Geosci.*, pages 1–19, 2012.

[23] W. K. Hastings. Monte-Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97, 1970.

[24] V. H. Hoang, C. Schwab, and A. M. Stuart. Complexity Analysis of Accelerated MCMC Methods for Bayesian Inversion. *ArXiv e-prints*, July 2012.

[25] S.F. Jarner and W.K. Yuen. Conductance bounds on the \(L^2\) convergence rate of Metropolis algorithms on unbounded state spaces. *Adv. in Appl. Probab.*, 36(1):243–266, 2004.

[26] G. L. Jones, M. Haran, B. S. Caffo, and R. Neath. Fixed-width output analysis for Markov chain Monte Carlo. *J. Amer. Statist. Assoc.*, 101(476):1537–1547, 2006.

[27] K. Kunisch. Numerical methods for parameter estimation problems inverse problems in diffusion processes. In *Proc. GAMM-SIAM Symp. (Philadelphia, PA: SIAM)*, pages 199–216, 1995.

[28] S. Lasanen. Measurements and infinite-dimensional statistical inverse theory. *PAMM*, 1080102:1080101–1080102, 2007.

[29] S. Lasanen. Non-Gaussian statistical inverse problems. Part I: Posterior distributions. *Inverse Probl. Imaging*, 6(2):215–266, 2012.

[30] S. Lasanen. Discretizations of generalized random variables with applications to inverse problems. *Ann. Acad. Sci. Fenn. Math. Diss.*, (130):64, 2002. Dissertation, University of Oulu, Oulu, 2002.

[31] K. Kunisch. Numerical methods for parameter estimation problems inverse problems in diffusion processes. In *Proc. GAMM-SIAM Symp. (Philadelphia, PA: SIAM)*, pages 199–216, 1995.

[32] S. Lasanen. Non-Gaussian statistical inverse problems. Part II: Posterior convergence for approximated unknowns. *Inverse Probl. Imaging*, 6(2):267–287, 2012.

[33] S. Lasanen. Discretizations of generalized random variables with applications to inverse problems. *Ann. Acad. Sci. Fenn. Math. Diss.*, (130):64, 2002. Dissertation, University of Oulu, Oulu, 2002.

[34] K. Latuszyński and W. Niemiro. Rigorous confidence bounds for MCMC under a geometric drift condition. *J. Complexity*, 27(1):23–38, 2011.

[35] K. Latuszynski and G. O. Roberts. CLTs and asymptotic variance of time-sampled Markov chains. *Methodol. Comput. Appl. Probab.*, pages 1–11, 2011.

[36] G. F. Lawler and A. D. Sokal. Bounds on the \(L^2\) spectrum for Markov chains and Markov processes: a generalization of Cheeger’s inequality. *American Mathematical Society*, 309(2), 1988.

[37] Neal Madras and Dana Randall. Factoring graphs to bound mixing rates. In *37th Annual Symposium on Foundations of Computer Science (Burlington, VT, 1996)*, pages 194–203. IEEE Comput. Soc. Press, Los Alamitos, CA, 1996.

[38] D. McLaughlin and E. R. Townley. A reassessment of the groundwater inverse problem. *Water Resour. Res.*, 32(5):1131–1161, 1996.
[39] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, E. Teller, et al. Equation of state calculations by fast computing machines. *J. Chem. Phys.*, 21(6):1087, 1953.

[40] Sean Meyn and Richard L. Tweedie. *Markov Chains and Stochastic Stability*. Cambridge University Press, Cambridge, second edition, 2009. With a prologue by Peter W. Glynn.

[41] P. Neal, G. O. Roberts, and W. K. Yuen. Optimal scaling of random walk Metropolis algorithms with discontinuous target densities. *Ann. Appl. Probab.*, 22(5):1880–1927, 2012.

[42] G. O. Roberts and J. S. Rosenthal. General state space Markov chains and MCMC algorithms. *Probab. Surv.*, 1:20–71, 2004.

[43] G. O. Roberts and R. L. Tweedie. Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, pages 341–363, 1996.

[44] G. O. Roberts and J. S. Rosenthal. Geometric ergodicity and hybrid Markov chains. *Electron. Comm. Probab.*, 2:13–25, 1997.

[45] D. Rudolf. Explicit error bounds for Markov chain Monte Carlo. *Dissertationes Math. (Rozprawy Mat.*), 485:1–93, 2012.

[46] C. Schwab and C. J. Gittelson. Sparse tensor discretizations of high-dimensional parametric and stochastic PDEs. *Acta Numer.*, 20:291–467, 2011.

[47] C. Schwab and A. M. Stuart. Sparse deterministic approximation of Bayesian inverse problems. *Inverse Probl.*, 28(4):045003, 32, 2012.

[48] A. Sinclair and M. Jerrum. Approximate counting, uniform generation and rapidly mixing Markov chains. *Inform. and Comput.*, 82(1):93–133, 1989.

[49] A. M. Stuart. Inverse problems: a Bayesian perspective. *Acta Numer.*, 19:451–559, 2010.

[50] A. M. Stuart. The Bayesian Approach To Inverse Problems. *ArXiv preprint 1302.6989*, 2013.

[51] A. M. Stuart, P. P. Wiberg, and J. Voss. Conditional path sampling of SDEs and the Langevin MCMC method. *Commun. Math. Sci.*, 2:685–697, 2004.

[52] L. Tierney. A note on Metropolis-Hastings kernels for general state spaces. *Ann. Appl. Probab.*, 8(1):1–9, 1998.

[53] L. Wang and J. Zou. Error estimates of finite element methods for parameter identification problems in elliptic and parabolic systems. *Discrete Contin. Dyn. Syst. Ser. B*, 14:1641–1670, 2010.