

Synchronized step multilevel Markov chain Monte Carlo

Sanjan C. Muchandimath^{a,*}, Alex A. Gorodetsky^a

^aUniversity of Michigan, Ann Arbor, 48105, MI, USA

Abstract

We propose SYNCE (synchronized step correlation enhancement), a new algorithm for coupling Markov chains within multilevel Markov chain Monte Carlo (ML-MCMC) estimators. We apply this algorithm to solve Bayesian inverse problems using multiple model fidelities. SYNCE is inspired by the concept of common random number coupling in Markov chain Monte Carlo sampling. Unlike state-of-the-art methods that rely on the overlap of level-wise posteriors, our approach enables effective coupling even when posteriors differ substantially. This overlap-independence generates significantly higher correlation between samples at different fidelity levels, improving variance reduction and computational efficiency in the ML-MCMC estimator. We prove that SYNCE admits a unique invariant probability measure and demonstrate that the coupled chains converge to this measure faster than existing overlap-dependent methods, particularly when models are dissimilar. Numerical experiments validate that SYNCE consistently outperforms current coupling strategies in terms of computational efficiency and scalability across varying model fidelities and problem dimensions.

Keywords: Markov chain Monte Carlo, Bayesian inverse problems, multilevel Monte Carlo, multi-fidelity, coupling
2000 MSC: 62F15, 62M05, 65C05, 65C40

1. Introduction

Bayesian inference is a well established technique for estimating parameters of complex mathematical models. In this stochastic framework, the parameters θ of a model $\mathcal{F}$ are modeled as random variables. By conditioning a prior distribution $\pi^0(\theta)$ on observed data $y_{\mathcal{D}}$, we obtain a posterior distribution $\pi(\theta | y_{\mathcal{D}})$ using Bayes' rule. Given this updated distribution over the parameters, application goals typically seek to estimate statistics of some output functional Q ; for example, to compute the expectation $\mathbb{E}_{\theta \sim \pi(\theta | y_{\mathcal{D}})}[Q(\theta)]$. In non-linear or non-Gaussian settings, these estimates are typically computed via sampling methods such as Markov chain Monte Carlo (MCMC) [1, 2].

However, unlike forward uncertainty quantification (UQ) problems where input parameter distributions are often known and can be sampled directly, Bayesian inverse problems (BIP) pose a greater computational challenge. MCMC algorithms struggle with sampling from posterior distributions that are high-dimensional and exhibit complex structures such as multi-modality or strong correlations. These struggles result in high autocorrelation and reduced effective sample size. Consequently, achieving accurate estimates of quantities of interest requires a large number of posterior samples N to overcome the slow $1/\sqrt{N}$ Monte Carlo convergence rate. This inefficiency is further exacerbated when the underlying simulations are computationally intensive – each posterior sample requires a forward model evaluation for the likelihood, and every sample drawn requires evaluating a possibly expensive functional Q . This simultaneous demand for high sample counts and expensive model evaluations makes Bayesian inverse problems particularly challenging.

Multi-fidelity strategies alleviate these computational costs by combining outputs from a hierarchy of models, leveraging cheaper, less accurate models to achieve computational speedups while preserving accuracy [3]. This paper seeks to improve these strategies to specifically target the high sampling costs inherent to BIPs. Implementing multi-fidelity methods requires two key components: (1) the availability of both low-fidelity and high-fidelity models,

*Corresponding author

Email addresses: sanjanmc@umich.edu (Sanjan C. Muchandimath), goroda@umich.edu (Alex A. Gorodetsky)

and (2) a model management strategy that distributes work/samples among these available models. Such hierarchies commonly arise in physical simulations – for example, through varying mesh resolutions, simplified physics, or surrogate modeling techniques. Historically, multi-fidelity model management strategies have been realized through control variate (CV) [4, 2, 5, 6], approximate control variate (ACV) [7, 8, 9], multilevel Monte Carlo (MLMC) [10, 11, 12, 13] and multilevel best linear unbiased estimator (MLBLUE) [14] techniques. The efficiency of these estimators relies on the *correlation* between the functional quantities of the low and high-fidelity models, with higher correlations leading to higher variance reduction.

This correlation requirement introduces unique challenges in the context of inverse problems. Unlike forward problems where input samples are drawn from a known distribution, inverse problems require sampling from posterior distributions that depend on the model fidelity. Hence, in the inverse context, we have an additional opportunity to reduce the computational cost of generating samples by appealing to *low-fidelity posteriors*. Rather than drawing samples from the high-fidelity posterior alone, we can mix and match samples from posteriors of different fidelities. However, this opportunity comes with some additional challenges. Since the posterior distributions change with model fidelity, we maintain correlation required for variance reduction by using *coupled* Markov chains. Coupling in this context refers to the idea of creating a joint Markov kernel that induces a Markov chain on the product space of the individual chains (e.g., a low-fidelity and high-fidelity chain). A fundamental requirement of this coupling is that the marginals of the joint chain target the distributions of the individual chains. Beyond this single requirement, there is significant freedom in how the joint kernel is constructed. This flexibility allows us to design clever coupling strategies that can increase the correlation between samples drawn from different fidelity posteriors.

Several related methodologies have been proposed in the literature for constructing such couplings within multilevel MCMC (ML-MCMC) estimators. Dodwell et al. [15] integrate a delayed acceptance strategy within ML-MCMC by proposing a point from a low-fidelity (coarse) MCMC chain as a candidate for the high-fidelity (fine) MCMC chain. The foundational delayed acceptance idea was formalized by Christen and Fox [16], who demonstrate that the method is effective only when the low-fidelity model is a good approximation of the high-fidelity model. Lykkegaard et al. [17] expand on the delayed acceptance proposal idea by creating ergodic chains using recursivity and leveraging an adaptive error model (AEM) to correct for dissimilar posteriors. Similarly, specialized proposals are explored in [18], which integrates delayed acceptance with dimension independent likelihood informed (DILI) MCMC [19]. Independent proposals are analyzed by Cianci et al. [20], where chains are coerced to accept the same proposed point from the independent sampler. However, the authors provide restrictive conditions on the independent proposal to ensure ergodicity. To enable state dependent proposals, Cianci and Pablo [21] used the maximal coupling technique in the ML-MCMC framework. Recently, other interesting applications of coupling have been applied to create unbiased [22, 23] and hybrid [24] estimators.

While these methodologies differ in their specific implementations, they all share the same strategy: couple chains by proposing the same sample for both the chains. This approach is similar to the logic successfully applied in forward UQ problems [25, 26]. In the forward setting, coupling methods rely on the idea that sharing samples between models (in a subspace) increases correlation between outputs, thereby reducing variance. However, this shared sample strategy has limitations in the inverse problem context. The effectiveness of these methods relies heavily on the overlap between the two posteriors targeted by the coupled chains – when models are dissimilar, a sample from a low-fidelity posterior may fall in a region of low probability for the high-fidelity posterior. This mismatch leads to frequent rejections, decoupling the chains and reducing the correlation between samples. This limitation necessitates a more robust approach — one that couples the evolution of the chains through synchronized steps rather than their states. In this context, we propose SYNCE (synchronized step correlation enhancement), a common random number (CRN) coupling algorithm that synchronizes the random steps of the coupled chains to maintain correlation even when the posterior distributions do not overlap.

The key contributions of this paper are:

1. We review existing methodologies for integrating MCMC with the MLMC estimator and provide insights into how these methods come under the framework of Markov chain coupling in Section 3.
2. We introduce SYNCE coupling, inspired by [28]. We prove the coupled chains admit a unique invariant probability measure and establish geometric ergodicity with explicit convergence rates in Theorem 4.1. Our analysis in Section 4.2 demonstrates that SYNCE achieves faster sampling convergence than existing methods when posteriors weakly overlap or exhibit mismatch.

3. We propose an adaptive version of our algorithm, SYNCE-AR in Sections 4.4 and 4.5 that works by adapting and resynchronizing the coupled Markov kernel based on insights from the theoretical analysis.
4. We validate that our proposed methodology is more efficient than existing couplings in the literature in Section 5, achieving gains in variance reduction by an order of two.

2. Background

In this section, we first set up the Bayesian inverse problem and review how it is solved using the standard Metropolis-Hastings MCMC (MH-MCMC) algorithm. We then focus on multi-fidelity methods and introduce the multilevel Markov chain Monte Carlo (ML-MCMC) framework for variance reduction.

2.1. Bayesian inverse problems and MCMC

Let $(X, \mathcal{X})$ be a measurable space with $X \subseteq \mathbb{R}^d$, and define the forward operator for the BIP as $\mathcal{F} : \Theta \rightarrow \mathcal{Y}$. Here, $\Theta \in X$ denotes the input parameter space and $\mathcal{Y} \subseteq \mathbb{R}^m$ denotes the output space of our model. For a particular parameter vector $\theta \in \Theta$, we denote the forward model as $y = \mathcal{F}(\theta)$. Assuming additive noise, the observation model is expressed as $y = \mathcal{F}(\theta) + \epsilon$, where $\epsilon \sim \mathcal{N}(0, \Sigma)$ is Gaussian noise with covariance Σ . Given observations $y_{\mathcal{D}} \in \mathcal{Y}$, BIPs seek the posterior distribution of the parameter set θ denoted by $\pi(\theta)$. This posterior distribution is given by Bayes' rule when assuming a prior distribution $\pi^0(\theta)$ over the parameters:

$$\pi(\theta) = \frac{\mathcal{L}(\theta)\pi^0(\theta)}{Z}, \quad (1)$$

where

$$\mathcal{L}(\theta) = \frac{1}{\sqrt{2\pi^d} \sqrt{\Sigma}} \exp\left(-\frac{1}{2} (y_{\mathcal{D}} - \mathcal{F}(\theta))^T \Sigma^{-1} (y_{\mathcal{D}} - \mathcal{F}(\theta))\right), \quad (2)$$

is the likelihood under Gaussian noise and Z is the normalizing constant termed the evidence.

The posterior has no closed form solution for general cases, and sampling approaches based on MCMC are used to generate samples distributed according to π . MCMC approaches create a sequence of random variables $\{\theta^i\}_{i \in \mathbb{Z}^+}$, where each θ^i is the state of the chain at iteration i and $\mathbb{Z}^+$ denotes the set of positive integers. The chain is constructed so that its stationary distribution is the target posterior $\pi(\theta)$. Among MCMC methods, the Metropolis-Hastings algorithm [29] is the most popular and widely used. The proposal distribution $q(\cdot | \cdot)$ and the acceptance probability $\alpha(\cdot, \cdot)$ define the algorithm. Samples are proposed iteratively from q and are accepted or rejected based on α . The method is shown in Algorithm 1 and induces a Markov transition kernel of the form:

$$p(\theta, A) = \int_A \alpha(\theta, \theta^*) q(\theta^* | d\theta) + \left(\int_X 1 - \alpha(\theta, \theta^*) q(\theta^* | d\theta) \right) \delta_{\theta}(A), \quad A \in \mathcal{X}, \quad (3)$$

where δ is the Dirac delta function. Existence of a stationary distribution is guaranteed if the Markov kernel above satisfies detailed balance, and convergence is guaranteed if the kernel is ergodic. Further proof and reading can be found in [29, 1]. Given the ability to draw samples from the posterior $\pi(\theta)$, we seek to model predictions by computing statistics of a quantity of interest (QoI) defined as $Q : \Theta \rightarrow \mathcal{Q}$, where $\mathcal{Q} \subseteq \mathbb{R}^k$ denotes the output space and $k \geq 1$ its dimensionality. In this work, we focus on the scalar case $k = 1$ but extensions to multi-output quantities follow directly from [9].

A standard Monte Carlo (MC) estimator for the posterior mean of Q is given by:

$$\hat{Q} = \mathbb{E}_{\Theta}[Q(\theta)] = \int_{\Theta} Q(\theta) \pi(\theta) d\theta \approx \frac{1}{N} \sum_{i=1}^N Q(\theta^i), \quad \theta^i \in \Theta, \theta^i \sim \pi, \quad (5)$$

where the notation $\hat{Q}$ denotes an estimator for the mean of Q obtained by generating N MCMC samples $\{\theta^i\}_{i=1}^N$ from the distribution π . The first equality in Eq. (5) comes from writing the estimate of the mean as an expectation, the

Algorithm 1 Metropolis-Hastings MCMC [29]

```

1: Input: Target density  $\pi$ , proposal distribution  $q(\cdot | \cdot)$ , number of samples  $N$ , initial guess  $\theta^0$ 
2: Output: Samples  $\{\theta^i\}_{i=1}^N$  from  $\pi$ 
3: for  $i = 0, 1, \dots, N - 1$  do
4:   Sample  $\theta^* \sim q(\cdot | \theta^i)$ 
5:   Sample  $u \sim \mathcal{U}(0, 1)$ 
6:   Set  $\theta^{i+1} = \theta^*$  if  $u < \alpha$  where

```

$$\alpha(\theta^i, \theta^*) = \min \left(1, \frac{\pi(\theta^*)q(\theta^i | \theta^*)}{\pi(\theta^i)q(\theta^* | \theta^i)} \right), \quad (4)$$

```

7:   Set  $\theta^{i+1} = \theta^i$  otherwise
8: end for
9: return  $\{\theta^i\}_{i=1}^N$ 

```

second equality comes from writing the expectation as an integral and the approximation comes from writing the integral as a Monte Carlo estimation with N samples. The estimator in Eq. (5) is unbiased with respect to the model used if the MCMC samples are independent and identically distributed. However, samples generated by MCMC are inherently correlated, and the estimator is only asymptotically unbiased. While recent work has developed unbiased estimators using couplings [22], in this work, we assume a sufficient burn-in period and disregard any initial bias. The error of the MC estimator is then measured by its variance, and when the variance of $Q(\theta)$ is finite, the variance of the estimator is inversely proportional to the number of samples N . In this paper, we do not consider the bias of the model with respect to representing some underlying truth model, and therefore our goal is to reduce the error of this procedure through variance reduction. Here, we employ multi-fidelity techniques.

2.2. Multi-fidelity methods for variance reduction

The Monte Carlo estimator in Eq. (5) is computationally expensive, with improvements to accuracy requiring orders of magnitude more samples. One strategy to address this challenge is to leverage multi-fidelity techniques that combine estimates from an ensemble of models to achieve a significantly lower variance. This ensemble typically includes a high-fidelity model and several low-fidelity models. In the context of a discretized PDE parameterized by ℓ , the low fidelity models are coarse grid approximations of the original discretization given by $\{\mathcal{F}_\ell\}_{\ell=0}^{L-1}$. Similarly, we obtain a hierarchy of output functionals $\{Q_\ell\}_{\ell=0}^{L-1}$. We assume that the evaluation of both $\mathcal{F}_\ell$ and Q_ℓ is a function of the discretization parameter $M_\ell = s^\ell M_0$. In the following, outputs with subscript L are assumed to be the highest fidelity with respect to which we seek variance reduction.

Although a plethora of multi-fidelity techniques have been explored in the past [3], in this work, we focus on creating *fusion* estimators of the type,

$$\hat{Q}_L^{\text{MF}} = f(\hat{Q}_0, \hat{Q}_1, \dots, \hat{Q}_L), \quad (6)$$

where $\hat{Q}_\ell$ denotes the estimator of the QoI at levels $\ell = 0, 1, \dots, L$ and $f(\cdot)$ is a function that combines these estimators. The goal of the fusion estimator is to reduce the variance of the estimator $\hat{Q}_L^{\text{MF}}$ by exploiting the correlation between the estimators $\hat{Q}_\ell$ at different levels. Both the choice of the function $f(\cdot)$ and the correlation between the multi-fidelity estimators play a crucial role in determining the efficiency of the estimator.

2.2.1. Multilevel Monte Carlo

The multilevel Monte Carlo (MLMC) estimator [11] is a popular multi-fidelity fusion technique that has been widely used in the context of forward uncertainty quantification. The MLMC estimator is based on the telescoping sum representation of the high-fidelity QoI:

$$\hat{Q}_L^{\text{MLMC}} = \hat{Q}_0 + \sum_{\ell=1}^L (\hat{Q}_\ell - \hat{Q}_{\ell-1}) = \sum_{\ell=0}^L \hat{Y}_\ell, \quad (7)$$

where the level wise estimator is defined as $\hat{Y}_\ell = \hat{Q}_\ell - \hat{Q}_{\ell-1}$ and $\hat{Y}_0 = \hat{Q}_0$. The MLMC estimator is unbiased and the variance of the estimator is dependent on the variance of each of the terms $\{\hat{Y}_\ell\}_{\ell=0}^L$ in Eq. (7). The key is then to create highly correlated estimators $\hat{Q}_\ell$ and $\hat{Q}_{\ell-1}$ by coupling samples that will result in $\mathbb{V}[Y_\ell] \rightarrow 0$ as $\ell \rightarrow \infty$. Consider the variance of the level wise outputs at a particular level ℓ given by:

$$\mathbb{V}[Y_\ell] = \mathbb{V}[Q_\ell] + \mathbb{V}[Q_{\ell-1}] - 2\text{Cov}[Q_\ell, Q_{\ell-1}].$$

Assuming $\mathbb{V}[Q_\ell] \approx \mathbb{V}[Q_{\ell-1}] = V$ and using the Pearson correlation coefficient,

$$\rho_\ell = \text{Cov}[Q_\ell, Q_{\ell-1}] / \sqrt{\mathbb{V}[Q_\ell]\mathbb{V}[Q_{\ell-1}]}, \quad (8)$$

we write the variance as $\mathbb{V}[Y_\ell] = 2V(1 - \rho_\ell)$. The variance of the MLMC estimator can be reduced by increasing the correlation between the *successive* level-dependent functional outputs Q_ℓ and $Q_{\ell-1}$. In standard MLMC applications, when exact sampling is possible for the forward problem, correlating successive level estimators is achieved by sharing samples of the highest fidelity L for both the estimators at a particular level. However, for inverse problems, the distributions of each level may be different and correspond to the inverse problem at each such level.

3. Inverse problem estimators and couplings

In the inverse case, where MCMC sampling is employed, achieving high correlations between the functional estimators becomes significantly complex. The dependency on the forward model introduced by MCMC methods complicates direct sharing of samples, making it *challenging to use the same input samples for each level, as is typical for MLMC approaches for forward problems*. These challenges necessitate studying *couplings* that can efficiently establish and exploit such correlations within the MCMC framework. In this section, we will formally introduce couplings and demonstrate their critical role in variance reduction. Then, we will extend our discussion to the integration of MLMC frameworks with MCMC methods to reduce the computational cost of the estimation procedure for inverse problems.

3.1. Couplings

Let $(X, \mathcal{X}, \mu)$ and $(Y, \mathcal{Y}, \nu)$ be two probability spaces. Coupling of two probability measures μ and ν is defined as a joint probability measure Γ on $(\Omega, \mathcal{O}, \Gamma)$ where $\Omega = X \times Y$ and $\mathcal{O} = \mathcal{X} \otimes \mathcal{Y}$, such that the marginals of Γ are μ and ν [30]. In other words, for all measurable sets $A \in \mathcal{X}$ and $B \in \mathcal{Y}$, we have $\Gamma[A \times Y] = \mu(A)$ and $\Gamma[X \times B] = \nu(B)$. Couplings are typically used to establish some distributional properties between the two measures μ and ν . For instance, they can be employed to analyze the distance between probability measures in terms of distance metrics [31], to facilitate transportation problems [30], or to prove convergence properties of MCMC methods [32, 33]. More recently, novel couplings have been proposed to exploit variance reduction and provide unbiased estimations [22, 23] in the context of MCMC.

Couplings can be broadly classified into two categories: *deterministic* and *non-deterministic* [30]. In deterministic couplings, the relationship between two random variables X and Y defined on the two probability spaces is described by a function. Formally, this translates to the existence of a measurable function $T : X \rightarrow Y$ such that $Y = T(X)$ and $T_\# \mu = \nu$, where $T_\#$ is called the pushforward operator. The coupling is then defined by the identity map $\mathbb{I}$ on X and the transport map T on Y , i.e., $\Gamma = (\mathbb{I}, T)_\# \mu$. An important deterministic coupling example is the optimal transport (OT) problem which finds the optimal transport map by introducing a cost function $c(x, y)$ on Ω that can be interpreted as the work needed to transport a unit of mass from $x \in X$ to $y \in Y$. Some applications of transport maps can be found in [34, 35]. Non-deterministic couplings, referred to simply as couplings, arise from invoking randomization during their construction and are described through a joint probability measure Γ on Ω . A notable example is the maximal coupling technique [31], where the joint distribution Γ is constructed such that the probability of the random variables X and Y being equal is maximized [31].

We now extend the concept of coupling to Markov chains following the approach of [31, 36]: suppose we have an ergodic MCMC kernel $p_X : X \times X \rightarrow [0, 1]$ with measure μ and another ergodic MCMC kernel $p_Y : Y \times Y \rightarrow [0, 1]$ with measure ν . A coupling of these two kernels is a joint kernel $p : \Omega \times \Omega \rightarrow [0, 1]$ with joint measure Γ satisfying $\int_Y p((x, y), A \times dy) = p_X(x, A)$ and $\int_X p((x, y), dx \times B) = p_Y(y, B)$. These equations imply that any Markov chain

$\{Z^i\}_{i \in \mathbb{Z}^+} = (\{X^i\}_{i \in \mathbb{Z}^+}, \{Y^i\}_{i \in \mathbb{Z}^+})$ constructed with the joint kernel p is composed such that the sequences $\{X^i\}_{i \in \mathbb{Z}^+}$ and $\{Y^i\}_{i \in \mathbb{Z}^+}$ follow the transition dynamics of their respective kernels p_X and p_Y . Crucially, while the marginals are fixed, the joint properties of the coupled chain can be designed to achieve specific goals, such as maximizing correlation between the two chains for variance reduction purposes.

In this work, we are dealing with MH-MCMC type methods for our inverse problem estimators, where the efficiency of the estimator depends on the quality of Markov chain coupling. As established in [36], the coupling of two MH-MCMC kernels can be represented as a combination of a proposal coupling (q_X and q_Y) followed by an acceptance coupling (α_X and α_Y). Specifically, to effectively couple two Markov chains together, we first couple the proposal distributions to produce a proposed state $Z^* = \{X^*, Y^*\}$ given the current state $Z^i = \{X^i, Y^i\}$. We then perform the acceptance coupling step by marginally accepting or rejecting the proposed state on both chains using the same uniform random number in the accept-reject function (step 6 in Algorithm 1). This decomposition highlights that the key to constructing efficient Markov chain couplings lies in designing effective *proposal couplings*.

3.2. Multilevel Markov chain Monte Carlo estimators

Next, we integrate the MLMC estimator with MCMC methods to reduce the computational cost of solving inverse problems. The core of the integration lies in generating correlated samples from coupled Markov chains at consecutive fidelity levels. For each level $\ell = 1, \dots, L$ in the MLMC estimator (Eq. (7)), we generate samples for two output functionals Q_ℓ and $Q_{\ell-1}$ to estimate $\mathbb{E}[Q_\ell - Q_{\ell-1}]$. The samples used for Q_ℓ are drawn from a *fine* chain at level ℓ , while the samples for $Q_{\ell-1}$ are drawn from a *coarse* chain at the same level. The strategy for defining the target distributions for these two chains is crucial for the efficiency of the ML-MCMC estimator. There are three potential routes: (1) First, both fine and coarse chains target the highest fidelity posterior π_L . While conceptually simple, this approach is computationally prohibitive because of the high cost of evaluating the likelihood at the finest level. (2) Second, samples from the coarse posteriors at each level approximate the fine posteriors at that level. While computationally cheaper, this approach introduces bias in the estimator. (3) Third, each chain targets its respective level-dependent posterior, i.e., the coarse chain at level ℓ targets $\pi_{\ell-1}$ while the fine chain targets π_ℓ . We adopt this strategy, as it is unbiased and computationally efficient, leaving only the challenge of constructing effective couplings between the chains.

To implement the level-dependent structure of the couplings, we introduce notation for the coupled samples. Let $\{\theta_\ell^i\}_{i \in \mathbb{Z}^+} \sim \pi_\ell$ denote the fine samples at level ℓ , and let $\{\vartheta_{\ell-1}^i\}_{i \in \mathbb{Z}^+} \sim \pi_{\ell-1}$ denote the coarse samples at the same level. The coupled chain at MLMC level ℓ thus produces pairs $\{(\theta_\ell^i, \vartheta_{\ell-1}^i)\}_{i \in \mathbb{Z}^+}$. We emphasize that the notation ϑ distinguishes the role of the sample in the telescoping sum, not just its distribution. For example, θ_0^i and ϑ_0^i are both drawn from the same posterior π_0 , but θ_0^i is used in the fine component of the MLMC estimator at level 0, while ϑ_0^i is used in the coarse component at level 1. Thus, the MLMC estimator in Eq. (7) specialized to the Monte Carlo estimate case becomes:

$$\hat{Q}^{\text{ML-MCMC}} = \frac{1}{N_0} \sum_{i=1}^{N_0} Q_0(\theta_0^i) + \sum_{\ell=1}^L \frac{1}{N_\ell} \sum_{i=1}^{N_\ell} (Q_\ell(\theta_\ell^i) - Q_{\ell-1}(\vartheta_{\ell-1}^i)), \quad \theta_\ell^i \sim \pi_\ell, \vartheta_{\ell-1}^i \sim \pi_{\ell-1}, \quad (9)$$

assuming that the Markov chains are sufficiently burned in.

The efficiency of the ML-MCMC estimator above depends on the correlation between the fine and coarse outputs at each level. This transforms the problem of variance reduction in inverse problems into one of constructing effective proposal couplings ($\Gamma_\ell = \{q_\ell, q_{\ell-1}\}$) of the level-dependent Markov chains as discussed in section 3.1. Achieving efficient proposal couplings facilitates high correlations between the chains, essential for reducing the variance of the level wise (difference) estimators in Eq. (9). Once the coupling is established, we can solve an optimization problem to find the optimal number of samples for each MLMC level depending on the covariance structure of the functional estimates [11].

We now provide a meta-algorithm for the ML-MCMC framework in Algorithm 2. The first step in the base algorithm is to sample from the lowest fidelity posterior distribution π_0 using any ergodic MCMC algorithm and a proposal distribution $\Gamma_0 = q_0$, for example, Algorithm 1. Then, for the successive levels, based on the coupling method for the joint distribution, we construct a coupled Markov chain $\{(\theta_\ell^i, \vartheta_{\ell-1}^i)\}_{i \in \mathbb{Z}^+}$ for which the marginal chains at two successive levels are highly correlated and sampled from their respective target distributions. This is done by

Algorithm 2 Meta algorithm for multilevel Markov chain Monte Carlo (ML-MCMC)

```

1: Input: List of target distributions, joint proposal distributions, number of samples and initial guesses for all
   MLMC levels  $\{\pi_\ell, \Gamma_\ell, N_\ell, \theta_\ell^0\}_{\ell=0}^L$ 
2: Output: Coupled chains  $\{\theta_0^i\}_{i=1}^{N_0}, \{(\theta_\ell^i, \vartheta_{\ell-1}^i)\}_{i=1}^{N_\ell}$  for  $\ell = 1, 2, \dots, L$ 
3: if  $\ell = 0$  then
4:    $\{\theta_0^i\}_{i=1}^{N_0} = \text{MH-MCMC}(\pi_0, \Gamma_0, N_0, \theta_0^0)$ 
5: end if
6: for  $\ell = 1, 2, \dots, L$  do
7:   for  $i = 0, 1, \dots, N_\ell - 1$  do
8:     Sample  $\theta_\ell^*, \vartheta_{\ell-1}^* \sim \Gamma_\ell((\cdot, \cdot) \mid (\theta_\ell^i, \vartheta_{\ell-1}^i))$ 
9:     Sample  $u \sim \mathcal{U}(0, 1)$ 
10:    Accept or reject  $\theta_\ell^{i+1} = \theta_\ell^*$  if  $u < \alpha_\ell$ , where  $\alpha_\ell$  is given by Eq. (4) for  $\pi_\ell$  and  $q_\ell$ 
11:    Accept or reject  $\vartheta_{\ell-1}^{i+1} = \vartheta_{\ell-1}^*$  if  $u < \alpha_{\ell-1}$ , where  $\alpha_{\ell-1}$  is given by Eq. (4) for  $\pi_{\ell-1}$  and  $q_{\ell-1}$ 
12:   end for
13:   Save the coupled chains  $\{\theta_\ell^i, \vartheta_{\ell-1}^i\}_{i=1}^{N_\ell}$ 
14: end for
15: return  $\{\theta_0^i\}_{i=1}^{N_0}, \{(\theta_\ell^i, \vartheta_{\ell-1}^i)\}_{i=1}^{N_\ell}$  for  $\ell = 1, 2, \dots, L$ 

```

proposing a joint state from the coupled Markov kernel followed by an accept-reject step based on the same uniform random number u as discussed in section 3.1. The coupled chains along with the coarsest level chain are saved and returned. The samples from all the chains are then used to compute the functional quantities used in the estimator in Eq. (9).

3.3. Existing coupling methods for ML-MCMC

We now review some existing non-deterministic coupling methodologies from the literature, each representing a specific choice for the proposal coupling Γ_ℓ in Algorithm 2

3.3.1. Delayed acceptance coupling [16]

A powerful strategy to construct the coupled proposal distribution Γ_ℓ is inspired by the delayed acceptance method [16]. Unlike standard MH couplings that separates the proposal and acceptance steps, the DA method integrates a two-stage acceptance procedure into the proposal generation itself. The key insight is that a proposal is first evaluated against a computationally cheaper coarse model before being evaluated against the more expensive fine model. This method requires less computational effort compared to standard MCMC because the fine level forward model is called only if the coarse level proposal is accepted, leading to the name delayed acceptance. This idea motivates using the *coarse level posterior distribution* $\pi_{\ell-1}$ as a *proposal for the fine level posterior distribution* π_ℓ . The general DA coupling strategy, which replaces Line 8 in Algorithm 2, is outlined below.

Algorithm 3 General delayed acceptance coupling [15, 17] (Line 8 in Algorithm 2)

```

1: Generate a coarse proposal  $\vartheta_{\ell-1}^*$  via a method-specific procedure.
2: Set the fine proposal  $\theta_\ell^* \leftarrow \vartheta_{\ell-1}^*$ 
3: return  $(\theta_\ell^*, \vartheta_{\ell-1}^*)$ 

```

The key design choice is how to generate the coarse proposal $\vartheta_{\ell-1}^*$ in step 1 of Algorithm 3. We review two implementations.

(1) *Hierarchical ML-MCMC with subsampling* [15]. In this method, the coarse proposal $\vartheta_{\ell-1}^*$ is generated by approximating an independent draw from the coarse posterior $\pi_{\ell-1}$. This approximation is achieved through a recursive subsampling procedure. At $\ell = 0$, a standard long MCMC chain is run and then subsampled to produce approximately independent samples from π_0 . These samples are then used as proposals for the next level $\ell = 1$. The proposals are

then accepted or rejected based on the delayed acceptance framework [16]. These samples are then subsampled again to produce approximately independent samples from π_1 which are then used as proposals for level $\ell = 2$. This process is repeated until the finest level L is reached. This recursive independent subsampling procedure is shown in Algorithm 4. A major limitation of this recursive method is the factorial growth in computational cost with the number of

Algorithm 4 Recursive independence sampling [15] (Line 1 in Algorithm 3)

- 1: **If** $\ell = 0$: Run a long MCMC chain targeting π_0 and subsample it to produce a set of nearly independent samples.
 - 2: **For each level** $k = 1, 2, \dots, \ell - 1$ (recursive step):
 - 3: Use nearly independent samples from level $k - 1$ as proposals in a DA MCMC chain targeting π_k [16].
 - 4: Subsample the resulting chain to produce nearly independent samples from π_k .
 - 5: Set $\vartheta_{\ell-1}^*$ as the final subsampled state from $\pi_{\ell-1}$.
 - 6: **return** $\vartheta_{\ell-1}^*$
-

levels. To produce one independent sample $\vartheta_{\ell-1}^i$ on level $\ell - 1$, we need to compute $T_k := \prod_{k'=k}^{\ell-1} t_{k'}$ samples on each of the levels $k = 0, 1, \dots, \ell - 1$, where t_k is the subsampling rate for level k . Furthermore, because the method relies on finite subsampling to approximate independent samples from the coarse posterior distributions, it introduces bias into the estimation procedure, and the coupled chains cannot be guaranteed to be ergodic [15, 20].

(2) *Multilevel delayed acceptance (MLDA) sampler* [17]. The MLDA sampler resolves the ergodic limitations of the hierarchical sampling approach by constructing proposals via recursive finite-length MCMC subchains. Specifically, a proposal candidate $\vartheta_{\ell-1}^*$ at level ℓ is generated by running a short MCMC subchain targeting $\pi_{\ell-1}$, which itself may be generated by running a short MCMC subchain targeting $\pi_{\ell-2}$, and so on, until the coarsest level $\ell = 0$ is reached, where a standard MCMC chain is run. This procedure is summarized in Algorithm 5. Crucially, the Metropolis-

Algorithm 5 Recursive subchain using MLDA [17] (Line 1 in Algorithm 3)

- 1: **If** $\ell = 1$: Generate subchain targeting π_0 using any ergodic MCMC algorithm and return proposal ϑ_0^* for level 1.
 - 2: **If** $\ell > 1$: Generate subchain by recursively calling MLDA targeting $\pi_{\ell-1}$.
 - 3: Set the final state $\vartheta_{\ell-1}^*$ from the subchain targeting $\pi_{\ell-1}$ as the coarse proposal.
 - 4: **return** $\vartheta_{\ell-1}^*$
-

Hastings acceptance probability is correctly formulated to account for the exact transition kernel of the generating sub-chain, rather than assuming an independent proposal. This preserves the detailed balance condition and ensures that the resulting coupled ML-MCMC sampler is provably unbiased and ergodic.

Both the hierarchical ML-MCMC with subsampling and the MLDA sampler effectively utilize the delayed acceptance coupling strategy [16] to generate proposals for the fine level chain from the coarse level posterior distribution. In both algorithms, we screen the fine level proposals $\theta_\ell^* \leftarrow \vartheta_{\ell-1}^*$ through a sequence of cheaper coarse level acceptances, which can significantly reduce the computational cost when the coarse model is a good approximation of the fine model. However, both methods suffer from a factorial growth in computational cost with the number of levels. Every proposal θ_ℓ^* at level ℓ requires running subchains at all coarser levels, leading to an exponential increase in the total number of model evaluations as the number of levels increases. Furthermore, both methods rely on the assumption that the coarse and fine level posterior distributions are sufficiently similar to achieve high acceptance rates and effective correlation between the chains. When the posteriors are dissimilar, the acceptance rates drop significantly, leading to poor mixing and reduced variance reduction benefits.

To mitigate this issue of dissimilar posteriors, the authors in [17] propose an adaptive error model (AEM). The AEM approximates the discrepancy between the coarse and fine level likelihoods using a Gaussian error model. The mean and covariance of this error model are learned adaptively using Haario’s adaptive MCMC algorithm [37] during the sampling process. The learned error model is then used to correct the coarse level likelihood, effectively aligning it more closely with the fine level likelihood. This correction improves the similarity between the coarse and fine posteriors, leading to higher acceptance rates and better sampling efficiency. However, the AEM works only if all the levels of the hierarchy use the same data. If the data or likelihood function changes between levels, the discrepancy

between model outputs is no longer a function of the model error alone, but also of the data inconsistency. In such cases, we cannot learn or apply a valid error model, limiting the applicability of this approach. We demonstrate these challenges in Section 5.3, where we compare AEM performance against our method on a prey-predator model. In summary, while both delayed acceptance coupling methods provide effective strategies for coupling ML-MCMC chains, they face challenges related to computational cost growth and sensitivity to posterior similarity.

3.3.2. Independent proposal coupling [20]

The next method uses an independent proposal as the joint proposal distribution in Algorithm 2. The authors in [20] *couple the two chains with a proposal that is independent of the current state of either chain followed by an accept-reject step based on the same uniform random number*. This approach ensures the generation of marginally true chains that are correlated, assuming the chains exhibit a good overall acceptance rate and demonstrate effective mixing. They also show that under some conditions on the proposal and posterior densities, there exists a unique invariant probability measure for the coupled chain. However, achieving ergodicity requires a strong theoretical condition: the proposal distribution q_ℓ^* must have tails that decay more slowly than those of both posterior distributions π_ℓ and $\pi_{\ell-1}$ (Assumption A.1. in [20]). If the proposal decays faster than the posteriors, the coupled chain may fail to explore the tails, breaking ergodicity.

Algorithm 6 Independent proposal coupling [20] (Line 8 in Algorithm 2)

- 1: Construct independent proposal q_ℓ^*
 - 2: Sample $\theta^* \sim q_\ell^*$ from the independent proposal
 - 3: Set $\theta_\ell^*, \theta_{\ell-1}^* = \theta^*$
 - 4: **return** $(\theta_\ell^*, \theta_{\ell-1}^*)$
-

The delayed acceptance coupling method in section 3.3.1 is a special case of the independent proposal coupling method where the proposal distribution is chosen to be the coarse level posterior distribution ($q_\ell^* = \pi_{\ell-1}$). Therefore, the restrictive assumption on the tail decay of the proposal distribution in the independent proposal coupling method also applies to the delayed acceptance coupling method. The authors in [20] clearly show that the coupled chain obtained using the framework in [15] will not be ergodic unless the proposal distribution satisfies the condition in A.1., which is usually the case in practice. However, the independent proposal coupling method has its own challenges, particularly in selecting an appropriate proposal distribution that satisfies the tail decay assumption while maintaining sufficient correlation between the chains. Given that the posterior distributions of the coupled chains are unknown, determining a suitable proposal distribution for the tail decay assumption becomes a non-trivial task. The authors in [20] recommend utilizing density approximation techniques such as Kernel Density Estimation (KDE), flow-based generative models, or Laplace Approximations of the posterior distributions as viable proposal candidates. Nevertheless, these methods introduce additional computational complexity to the problem.

3.3.3. Maximal coupling [21]

Maximal coupling is a classic coupling technique that aims to maximize the probability of two coupled random variables being equal. This concept has been primarily used to prove convergence properties of MCMC chains [31]. The authors in [21] adapt this idea to the ML-MCMC framework by *maximally coupling chains at successive levels in the ML-MCMC framework to sample $\{\theta_\ell^*, \theta_{\ell-1}^*\}$ from a maximal coupling of $\{q_\ell, q_{\ell-1}\}$ such that $\theta_\ell^* \sim q_\ell(\cdot \mid \theta_{\ell-1}^*)$ and $\theta_{\ell-1}^* \sim q_{\ell-1}(\cdot \mid \theta_\ell^*)$, with $\mathbb{P}(\theta_\ell^* \neq \theta_{\ell-1}^*) = \|q_\ell(\cdot \mid \theta_\ell^*) - q_{\ell-1}(\cdot \mid \theta_{\ell-1}^*)\|_{TV}$, where $\|\cdot\|_{TV}$ is the total variation distance*.

Note that even though we term this method as a maximal coupling of the two chains, the proposal distributions are maximally coupled and a subsequent accept-reject step is performed on the proposals $\{\theta_\ell^*, \theta_{\ell-1}^*\}$ with a common uniform random number. A procedure to sample from such a maximal coupling Γ^* is taken from [31] called γ -coupling and is described in Algorithm 7. The readers are referred to [22, 36] for more such sampling techniques to sample from maximally coupled proposal distributions.

Although this method allows for state-dependent proposals in contrast to the independent proposal coupling method, and is relatively inexpensive to implement, the correlation between the chains depends on the total variation distance between the two proposal distributions, with smaller distances encouraging tighter coupling. Because

Algorithm 7 Maximal Coupling [21, 22, 31] (Line 8 in Algorithm 2)

- 1: Sample $\theta_\ell^* \sim q_\ell(\cdot \mid \theta_\ell^i)$ and $\psi \mid \theta_\ell^* \sim \mathcal{U}(0, q_\ell(\theta_\ell^* \mid \theta_\ell^i))$
 - 2: **If** $\psi \leq q_{\ell-1}(\theta_\ell^* \mid \vartheta_{\ell-1}^i)$, set $\vartheta_{\ell-1}^* \leftarrow \theta_\ell^*$
 - 3: **Else**, sample $\vartheta_{\ell-1}^* \sim q_{\ell-1}(\cdot \mid \vartheta_{\ell-1}^i)$ and $\phi \mid \vartheta_{\ell-1}^* \sim \mathcal{U}(0, q_{\ell-1}(\vartheta_{\ell-1}^* \mid \vartheta_{\ell-1}^i))$ until $\phi > q_\ell(\vartheta_{\ell-1}^* \mid \theta_\ell^i)$
 - 4: **return** $(\theta_\ell^*, \vartheta_{\ell-1}^*)$
-

the proposals are state-dependent, we want samples from each chain to be close to each other. However, achieving this proximity is infeasible in high dimensions or when the posterior distributions are dissimilar, which frequently occurs in practical applications. In those settings, the resulting weak coupling leads to poor correlation between the chains, limiting the variance reduction benefits of the ML-MCMC estimator.

4. Proposed coupling methodology

In this section, we introduce our SYNchronized step Correlation Enhancement (SYNCE) coupling method. We demonstrate how our approach overcomes the posterior-overlap limitations of the existing coupling methods, and provide a detailed framework for its implementation within the ML-MCMC estimator.

4.1. SYNCE coupling

Our proposed method is inspired from Pinto and Neal [28], where they couple Markov chains using common random number couplings. They couple two chains — the chain of an unknown target distribution, and a Gaussian approximation of the unknown target distribution to reduce the variance of their estimator. They state that *two chains are coupled when their transitions are determined by the same random numbers*. We extend this idea to our framework by using common random numbers for the two chains that sample from the ℓ 'th and $(\ell - 1)$ 'th level posterior distributions. Specifically, we use the same Gaussian random draw $\eta \sim \mathcal{N}(0, C_\ell)$ for proposal generation and the same uniform random variable $u \sim \mathcal{U}(0, 1)$ for the accept-reject step at both levels. Using the same random numbers η and u for both the chains ensures that the two chains are coupled and that the samples from the two chains are highly correlated. The simplest form of our proposed approach is given in Algorithm 8.

Algorithm 8 Baseline SYNCE coupling (Line 8 in Algorithm 2)

- 1: Sample $\eta \sim \mathcal{N}(0, C_\ell)$
 - 2: Set $\theta_\ell^* = \theta_\ell^i + \eta$ and $\vartheta_{\ell-1}^* = \vartheta_{\ell-1}^i + \eta$
 - 3: **return** $(\theta_\ell^*, \vartheta_{\ell-1}^*)$
-

This same-step strategy is fundamentally different from the same-sample strategy employed by the existing coupling methods discussed in section 3.3. Our approach ensures that the samples from both chains remain highly correlated, even when the posterior distributions are substantially different. In Fig. 1, we provide a visual comparison of the four coupling methods when the two posterior distributions π_ℓ and $\pi_{\ell-1}$ are substantially different.

In Fig. 1a, the fine chain uses the last accepted point from the coarse chain as its proposal. Since the posteriors are different, most of the proposals will be rejected leading to poor mixing and low correlation. In Fig. 1b, a common sample is proposed independently for both the chains. Proposals are either accepted by both chains (green dots), rejected by both chains (red dots) or accepted by one chain and rejected by the other (green/red dots). When the posteriors have no overlapping region, most of the proposals will be rejected by both chains leading to poor correlation and mixing. Fig. 1c depicts maximal coupling, where the probability of proposing the same sample via maximal coupling depends directly on the overlap between the two proposal distributions, which is measured by their total variation distance. When the posteriors are far apart, this overlap shrinks, reducing the probability that the chains propose identical points. Consequently, maximal coupling becomes ineffective for non-overlapping posterior distributions, reducing the coupling to independent product of the marginals. Finally, Fig. 1d illustrates the SYNCE coupling method, where the proposals are forced to have the same magnitude and direction of change. This approach produces high correlation by design and ensures good mixing even when the posteriors are different.

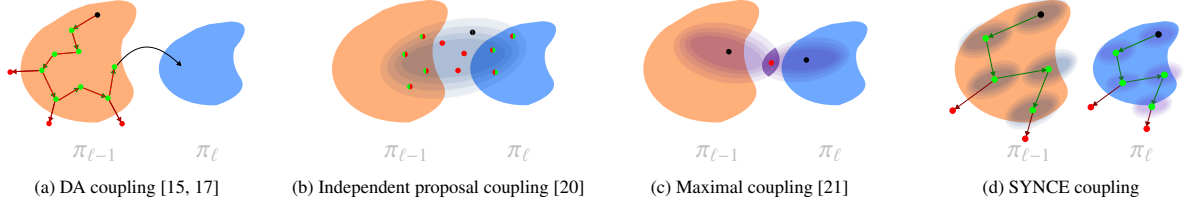


Figure 1: Comparison of the four coupling methods for two posteriors at levels $\ell - 1$ and ℓ that are different and do not overlap.

4.2. Theoretical analysis

In this section, we provide a theoretical analysis of the proposed SYNCE coupling method. We first define the coupled Markov kernel p_{ℓ}^{SYNCE} induced by the SYNCE coupling method in Algorithm 8. We then prove that the coupled kernel admits an invariant probability measure and establish geometric ergodicity to this invariant measure. Specifically, we obtain quantitative bounds on the rate of convergence of the coupled kernel to its invariant measure.

Let $p_{\ell}^{\text{SYNCE}} : X^2 \times X^2 \rightarrow [0, 1]$ denote the Markov transition kernel induced by SYNCE coupling in Algorithm 8. The kernel uses the coupled proposal to propose new states $(\theta_{\ell}^*, \vartheta_{\ell-1}^*) \sim \Gamma_{\ell}((\cdot, \cdot) | (\theta_{\ell}^i, \vartheta_{\ell-1}^i))$ for the two chains at level ℓ and $\ell - 1$. Following the notation $\alpha_{\ell}^i := \alpha_{\ell}(\theta_{\ell}^i, \theta_{\ell}^*)$, $\alpha_{\ell-1}^i := \alpha_{\ell-1}(\vartheta_{\ell-1}^i, \vartheta_{\ell-1}^*)$, $\Gamma_{\ell}^i := \Gamma_{\ell}(\theta_{\ell}^i | d\theta_{\ell}^*)$ and for any $A \in X^2$, we define the joint kernel as

$$\begin{aligned}
 p_{\ell}^{\text{SYNCE}}(\theta_{\ell}^i, A) &= \int_{X^2} \min(\alpha_{\ell}^i, \alpha_{\ell-1}^i) \Gamma_{\ell}^i \delta_{\theta_{\ell}^*, \vartheta_{\ell-1}^*}(A) + \int_{X^2} (\alpha_{\ell}^i - \alpha_{\ell-1}^i)^+ \Gamma_{\ell}^i \delta_{\theta_{\ell}^*, \vartheta_{\ell-1}^i}(A) \\
 &\quad + \int_{X^2} (\alpha_{\ell-1}^i - \alpha_{\ell}^i)^+ \Gamma_{\ell}^i \delta_{\theta_{\ell}^i, \vartheta_{\ell-1}^*}(A) + \left(1 - \int_{X^2} \max(\alpha_{\ell}^i, \alpha_{\ell-1}^i) \Gamma_{\ell}^i\right) \delta_{\theta_{\ell}^i, \vartheta_{\ell-1}^i}(A).
 \end{aligned} \tag{10}$$

where $x^+ := (x + |x|)/2$, $\forall x \in \mathbb{R}$. Each of the four terms above correspond to the four possible transitions in the coupled chain. The first term corresponds to the case where both the chains accept their respectively proposed point, the second and third terms correspond to the cases where only one of the chains accepts their proposed points and the last term corresponds to the case when both the chains reject their proposed points.

Establishing ergodicity of the coupled kernel in Eq. (10) is not trivial because the joint invariant measure $\mu_{\ell, \ell-1}$ is not explicitly defined a-priori. Consequently, its existence and invariance must be derived rather than assumed. Our proof strategy builds upon the classical Meyn-Tweedie framework [42]: we demonstrate that the coupled kernel satisfies ψ -irreducibility, aperiodicity, minorization, and a Foster-Lyapunov drift condition. These properties collectively ensure the existence of a unique invariant probability measure and guarantee geometric convergence to this measure. To facilitate our analysis, we introduce the following assumptions regarding the target densities and proposal distributions used in the marginal MH-MCMC kernels at levels ℓ and $\ell - 1$.

Assumption 4.1. For each marginal level $j \in \{\ell, \ell - 1\}$ targeting measure μ_j with density π_j (i.e., either π_{ℓ} or $\pi_{\ell-1}$) with a random walk proposal $q(\cdot)$,

(A.1) [Cf. Theorem 2.1 in [44]] The target density π_j is continuous and bounded away from 0 and ∞ on compact sets S , i.e.,

$$0 < \pi_{\min} \leq \pi_j(x) \leq \pi_{\max} < \infty, \quad \forall x \in S,$$

where $\pi_{\min} = \min_{x \in S} \pi_j(x)$ and $\pi_{\max} = \max_{x \in S} \pi_j(x)$.

(A.2) [Cf. Eq. (25) in [44]] The target density π_j is super-exponentially light, i.e.,

$$\limsup_{|x| \rightarrow \infty} \left\langle \frac{x}{|x|}, \nabla \log \pi_j(x) \right\rangle = -\infty.$$

(A.3) [Cf. Eq. (8) in [44]] The proposal density $q(x)$ of the random walk increment is bounded away from zero in some neighborhood of the origin, i.e., there exists $\delta_q > 0$ and $\epsilon_q > 0$ such that

$$q(x) \geq \epsilon_q, \quad \forall |x| \leq \delta_q.$$

A standard result in random walk Metropolis theory states that if the proposal is locally positive (assumption (A.3)) and the target is continuous and bounded (assumption (A.1)), then every compact set is a small set [43, 44]. A small set is a region where the chain can forget its past and regenerate with some positive probability. When the chain strays away from these compact small sets, the super-exponential tail condition (assumption (A.2)) ensures that there exists a drift function V_j that pulls the chain back toward these small sets. The local mixing property (minorization) combined with the global drift condition guarantees geometric ergodicity of the chain. We formalize these ideas in the following lemma, which provides necessary ingredients for our subsequent analysis of the coupled kernel.

Lemma 4.1. Let $p_j, j \in \{\ell, \ell-1\}$ be the marginal MH-MCMC kernel (Eq. (3)) targeting π_j with random walk proposal $q(\cdot)$ satisfying Assumption 4.1. Then, there exists a small set $S_j \subset X$ and Lyapunov function V_j such that:

(B.1) A one step minorization condition holds on S_j ; that is, there exists a probability measure ν_j on $(X, \mathcal{X})$ and a (minorizing) constant $\delta_j \in (0, 1)$ such that

$$p_j(x, A) \geq \delta_j \nu_j(A), \quad \forall x \in S_j, A \in \mathcal{X}.$$

(B.2) The Lyapunov function $V_j : X \rightarrow [1, \infty)$ satisfies the drift condition:

$$(p_j V_j)(x) \leq \Lambda_j V_j(x) + b_j \mathbb{1}_{x \in S_j}, \quad \forall x \in X.$$

Here, $(p_j V_j)(x) := \mathbb{E}_{p_j(x, \cdot)} [V_j(X^{n+1}) | X^n = x]$ denotes the expected value of V_j after one step of the chain starting from x , $\Lambda_j \in (0, 1)$, $b_j \in \mathbb{R}_+$ are constants, and $S_j := \{x \in X : V_j(x) \leq \Delta_j\}$ for some $\Delta_j > 0$.

PROOF (LEMMA 4.1). The proof is provided in [44] for item (B.1) and [43] for item (B.2).

□

With the marginal properties now established, we now present the main theorem of this section that characterizes the ergodicity of the SYNCE coupled kernel.

Theorem 4.1 (Geometric ergodicity of the SYNCE coupled kernel). Suppose that Assumption 4.1 holds for the two target densities $\pi_\ell, \pi_{\ell-1}$ and the Gaussian random walk proposal $q = N(0, C_\ell)$. Then, for any ML-MCMC level $\ell = 1, \dots, L$,

1. The joint kernel p_ℓ^{SYNCE} in Eq. (10) admits a unique invariant probability measure $\mu_{\ell, \ell-1}$ on $(X^2, \mathcal{X}^2)$ with marginals π_ℓ and $\pi_{\ell-1}$.
2. The joint chain induced by the SYNCE coupled kernel is geometrically ergodic; that is, there exist a joint Lyapunov function $\hat{V}_\ell : X^2 \rightarrow [1, \infty)$, a constant $R_\ell < \infty$, and a convergence rate $\rho \in (0, 1)$ such that

$$\left\| \left(p_\ell^{\text{SYNCE}} \right)^n (x, \cdot) - \mu_{\ell, \ell-1}(\cdot) \right\|_{TV} \leq R_\ell \hat{V}_\ell(x) \rho^n, \quad \forall n \geq 0, x \in X^2.$$

3. The convergence rate ρ satisfies the explicit bound:

$$\rho = \inf_{r \in (0, 1)} \max \left\{ (1 - \delta)^r (2\hat{b} + 1)^{1-r}, \left(\hat{\Lambda}_{\text{SYNCE}} + \frac{2\hat{b} + 1 - \hat{\Lambda}_{\text{SYNCE}}}{\Delta + 1} \right)^{1-r} \right\}, \quad (11)$$

where the constants depend on the joint small set $S = S_\ell \times S_{\ell-1}$ constructed in Lemma A.3 as follows:

- $\delta \propto \epsilon_q \frac{\pi_{\min}}{\pi_{\max}}$ is the minorization constant from Lemma A.2. Recall that ϵ_q is the lower bound on the proposal density from assumption (A.3), and $\pi_{\min}, \pi_{\max}$ are the posterior bounds from assumption (A.1) within the joint small set S .
- $\hat{\Lambda}_{\text{SYNCE}} = \max(\Lambda_\ell, \Lambda_{\ell-1}) + \max\left(\frac{b_\ell}{1+\Delta_{\ell-1}}, \frac{b_{\ell-1}}{1+\Delta_\ell}\right)$ is the joint drift rate from Lemma A.3.
- $\hat{b} = \frac{1}{2}(b_\ell + b_{\ell-1})$ is the drift constant on the small set S from Lemma A.3.
- Δ is the threshold defining the joint small set $S := \theta \in X^2 : \hat{V}_\ell(\theta) \leq \Delta$ from Lemma A.3.

PROOF (THEOREM 4.1). The proof follows by verifying the conditions of the Meyn-Tweedie framework (see Theorem 15.0.1 in [42]). Since p_ℓ^{SYNCE} is a coupling of standard MH-MCMC kernels, its marginal invariance is guaranteed by construction. Geometric ergodicity then follows from establishing ψ -irreducibility and aperiodicity (Lemma A.1), a minorization condition (Lemma A.2), and a joint drift condition (Lemma A.3). An explicit analytical bound for the convergence rate ρ is provided in Eq. (11) based on Theorem 1.3.5 in [46].

□

Remark 1. The geometric convergence rate ρ for the SYNCE coupled kernel established in Theorem 4.1 depends exclusively on the marginal posterior bounds, tail properties, and the lower bound on the proposal. Consequently, the rate remains unaffected by the degree of overlap between the two posterior distributions π_ℓ and $\pi_{\ell-1}$.

4.3. Convergence analysis and comparison

Theorem 4.1 establishes that the SYNCE algorithm generates samples that converge asymptotically to the unique joint invariant measure $\mu_{\ell, \ell-1}$, preserving the correct marginals π_ℓ and $\pi_{\ell-1}$. The property of geometric ergodicity guarantees that convergence to this invariant measure occurs at an exponential rate. As evident from Eq. (11), this rate is determined by two key components: the minorization constant $\delta \propto \epsilon_q \pi_{\min} / \pi_{\max}$ and the drift constant $\hat{\Lambda}_{\text{SYNCE}}$. These constants encapsulate the local mixing behavior within small sets and the global drift behavior outside these sets, respectively. This implies that the rate depends exclusively on the marginal posterior bounds, tail behavior, and the proposal bound, independent of the overlap between the posterior or proposal distributions. Even in scenarios where the two posterior distributions π_ℓ and $\pi_{\ell-1}$ are dissimilar or disjoint, the SYNCE coupling method maintains a robust convergence rate determined solely by the efficiency of the marginal random walk exploration. This robustness stems from the methodology of synchronizing the steps of both chains instead of attempting to force them to accept the same sample as seen in Fig. 1.

The theoretical analysis also provides insights towards improving its performance further through the lens of adaptation and resynchronization. To achieve fast convergence rates, we need to minimize the constant ρ . This requires us to maximize the proposal bound ϵ_q and minimize the drift constant $\hat{\Lambda}_{\text{SYNCE}}$. The constant ϵ_q is a property of the proposal distribution q , representing the minimum probability density for making small moves. A larger ϵ_q can be achieved by choosing a proposal with an appropriate step size (covariance)—not too large, not too small – that encourages efficient local exploration. The drift constant $\hat{\Lambda}_{\text{SYNCE}}$ reflects how strongly the chain is pulled back from the tails of the distribution. The pull is strongest when the proposal distribution q is well-matched to the geometry of the target posterior. A proposal that suggests moves aligned with the target’s shape and scale will lead to higher acceptance rates and a stronger, faster drift back to the high-probability region, resulting in a smaller $\hat{\Lambda}_{\text{SYNCE}}$. This motivates the need for adaptation and resynchronization strategies, which we discuss in sections 4.4 and 4.5.

We compare the convergence rates of the SYNCE kernel with the three existing coupling methods. While explicitly comparing the convergence rates is challenging due to the complexity of the constants involved, we analyze the limiting factors of the convergence rate in the worst-case scenario – when the two posterior distributions π_ℓ and $\pi_{\ell-1}$ are dissimilar and have little to no overlap. This regime is particularly important because it corresponds to cases with distinct model fidelities or changing data between levels.

Independent proposal/Delayed acceptance coupling. For the independent proposal coupling method in Algorithm 6, the uniform convergence rate bound derived in [20] is given by

$$\|p^n(x, \cdot) - \mu(\cdot)\|_{TV} \leq (1 - \rho)^n, \quad \forall n \geq 0, \forall x \in X, \rho \in (0, 1),$$

with $\rho = c(1 - \|\pi_\ell - \pi_{\ell-1}\|_{TV})$. The constant $c \in (0, 1)$ is related to the proposal distribution $c \leq \text{ess inf}_{x \in X} (q_\ell^*(x)/\pi_j(x))$. Here, $q_\ell^*(x)$ is the independent proposal distribution in Algorithm 6 and $j \in \{\ell, \ell-1\}$. The delayed acceptance method in Algorithm 3 is a special case with $q_\ell^* = \pi_{\ell-1}$. If the two posterior distributions $\pi_\ell, \pi_{\ell-1}$ do not overlap, then the TV distance $\|\pi_\ell - \pi_{\ell-1}\|_{TV}$ will be close to 1 and the convergence will be slow because ρ will be close to 0. This dependence on posterior overlap is a significant drawback of the independent proposal coupling method and the delayed acceptance coupling method.

Maximal coupling. The maximal coupling method in Algorithm 7 is designed to maximize the probability that the two chains propose the same state, $\theta_\ell^* = \vartheta_{\ell-1}^*$. As a random walk algorithm, its convergence is governed by the same theoretical framework as the SYNCE kernel established in Theorem 4.1. However, a critical difference lies in the minorization constant δ . For maximal coupling, the construction of the minorizing measure ν relies on the event that the chains successfully couple (see Lemma 6.3.2 in [21]). Specifically, the minorization constant δ for this method is bounded by the integral of the overlap between the proposal densities, $\int \min(q_\ell(\cdot | \theta_\ell^i), q_{\ell-1}(\cdot | \vartheta_{\ell-1}^i)) dx = 1 - \|q_\ell(\cdot | \theta_\ell^i) - q_{\ell-1}(\cdot | \vartheta_{\ell-1}^i)\|_{TV}$. If the posteriors π_ℓ and $\pi_{\ell-1}$ are dissimilar, the state-dependent proposals q_ℓ and $q_{\ell-1}$ will also have small overlap. In this limit, the minorization constant $\delta \rightarrow 0$, causing the convergence rate ρ to approach 1 (see Theorem 1.3.5 in [46]). While this method improves upon the independent proposal coupling method by utilizing state-dependent proposals, it still suffers from slow convergence when the posteriors are dissimilar.

The distinct advantages and disadvantages of each coupling strategy are summarized in Table 1, which highlights how the SYNCE coupling method effectively addresses the limitations of existing approaches by ensuring high correlation between chains for any two posterior distributions while being computationally efficient and straightforward to implement.

Method	Pros	Cons
DA ML-MCMC [15]	Good correlation for close posteriors	Factorial cost, non-ergodic
MLDA [17]	Good correlation and ergodic	Factorial cost, requires consistent data
Independent proposal [20]	Good correlation and ergodic	Requires good independent proposal
Maximal coupling [21]	State dependent proposals	Requires overlap between proposals
SYNCE coupling [this paper]	State dependent proposals, high correlation, cheap, easy to implement	Requires proposal covariance tuning

Table 1: Advantages and disadvantages of all the methods discussed. SYNCE overcomes the limitations of existing methods by ensuring high correlation between chains for any two posterior distributions while being computationally inexpensive and easy to implement.

Remark 2. While Theorem 4.1 establishes geometric ergodicity of the SYNCE coupled kernel, obtaining rigorous theoretical guarantees for the overall cost and convergence rates of the ML-MCMC estimator requires verifying a set of technical assumptions [12, 15, 20]. Providing a theoretical proof that these assumptions hold in general for the SYNCE coupling method is beyond the scope of this paper. Instead, we provide numerical evidence in section 5 that the assumptions are satisfied for a subsurface flow problem taken directly from [20]. For completeness, we present the full set of assumptions and the resulting ϵ -cost complexity theorem in the supplementary material section S1.

4.4. Adaptation

The performance of SYNCE is highly dependent on the choice of the covariance C_ℓ for the same-step η in Algorithm 8. In Fig. 1d, when steps of high magnitude are proposed, the proposals are likely to be rejected, especially if the posteriors are different. This leads to low acceptance rates, suboptimal mixing and small effective sample sizes. On the other extreme, when the steps are too small, samples are accepted quite often and the correlation is quite high, but the chains do not explore the posterior space effectively. To balance exploration, mixing, and correlation, we employ

adaptive MCMC techniques [37, 38, 40]. These works provide a framework to adapt the proposal distribution such that optimality in the sense of some target measure is reached.

Standard adaptation algorithms update a scalar λ that scales the empirical covariance of the samples collected so far to ensure that the acceptance rate of the chain is close to some target value α^* . The updates are given by (Algorithm 4 in [40]):

$$\begin{aligned} \log(\lambda^{i+1}) &= \log(\lambda^i) + \gamma^{i+1} (\alpha(\theta^i, \theta^*) - \alpha^*), \\ \mu^{i+1} &= \mu^i + \gamma^{i+1} (\theta^{i+1} - \mu^i), \\ \Sigma^{i+1} &= \Sigma^i + \gamma^{i+1} ((\theta^{i+1} - \mu^i)(\theta^{i+1} - \mu^i)^T - \Sigma^i), \end{aligned} \quad (12)$$

where γ^{i+1} 's are chosen to satisfy the diminishing adaptation condition [40]. In our coupled framework, we update the scaling factor λ , empirical mean μ , and covariance Σ individually for each chain, while maintaining coupling through a shared underlying random number. Specifically, we generate a common random number η from the standard normal distribution and multiply it with the square root of the adapted covariance and the scaling factor for each chain separately. The coupled proposals are now given by:

$$\eta \sim \mathcal{N}(0, I_d), \quad \theta_\ell^* = \theta_\ell^i + \lambda_\ell^i \sqrt{\Sigma_\ell^i} \eta, \quad \vartheta_{\ell-1}^* = \vartheta_{\ell-1}^i + \lambda_{\ell-1}^i \sqrt{\Sigma_{\ell-1}^i} \eta, \quad (13)$$

where d is the dimension of the parameter space. This formulation ensures that the chains are coupled, and the samples are highly correlated while exploring the individual posterior spaces effectively. This specific form has ties to the optimal transport equation for two Gaussians [30], specifically for our case, transporting the sample from the standard normal Gaussian to the level-dependent marginal.

4.5. Resynchronization

Covariance adaptation enables SYNCE to improve correlation between chains targeting dissimilar posteriors. However, in high-dimensional problems, we may encounter stages where the proposed step is accepted by one chain and rejected by the other. Over many iterations, this can cause the absolute states of the chains to become decorrelated, even though the underlying proposal mechanism remains synchronized. This phenomenon exposes a fundamental trade-off: SYNCE correlates the change in state, which is robust, but it does not directly force the states themselves to be correlated, which depends on the relative acceptance behavior of the two chains. To counteract this drift, we periodically resynchronize the chains to the same state. Specifically, we construct a mixture kernel [21]: the SYNCE coupling kernel p_ℓ^{SYNCE} and another kernel p_ℓ^* that proposes the same sample for both the chains, such as the independent proposal kernel or the delayed acceptance kernel. The transition kernel $p_\ell^{\text{SYNCE-R}}$ is then given by:

$$p_\ell^{\text{SYNCE-R}} = \omega_\ell p_\ell^* + (1 - \omega_\ell) p_\ell^{\text{SYNCE}}, \quad (14)$$

where $\omega_\ell \in (0, 1)$ determines the frequency of resynchronization. Crucially, resynchronization is only beneficial when the posteriors exhibit sufficient overlap. If the distributions are disjoint, resynchronization will likely lead to both chains rejecting the proposed point, resulting in no state change and wasted computational effort. Therefore, we set ω_ℓ to 0 for small ℓ values and increase it as ℓ increases. This ensures that when the posteriors are far off from each other, no resynchronization takes place and the chains are coupled using the SYNCE coupling kernel. When the posteriors become closer with higher levels, the chains are resynchronized using both the SYNCE coupling kernel and the resynchronization kernel. We note that the optimal choice of ω_ℓ is non-trivial. In practice, we estimate the degree of overlap using samples from the burn-in period and tune ω_ℓ accordingly. While an adaptive strategy for ω_ℓ could be developed, we leave this for future work.

The final adaptive proposed coupling method is given in Algorithm 9. The algorithm improves upon Algorithm 8 by adapting and scaling the covariance for both the chains separately and simultaneously, and resynchronizing the chains when needed for better correlation. Using resynchronization only for the fine levels, the adaptive parameters λ, μ and Σ are used to set the mean and covariance of the independent proposal distribution in Algorithm 6: $\mu_\ell^{\text{IMH}} = 1/2 (\mu_\ell^i + \mu_{\ell-1}^i)$ and $\Sigma_\ell^{\text{IMH}} = 1/2 (\Sigma_\ell^i + \Sigma_{\ell-1}^i)$ for $\ell = 1, 2, \dots, L$. This ensures that for close and similar shaped posteriors, we obtain highly correlated chains with good mixing properties.

Algorithm 9 SYNCE with adaptation and resynchronization (SYNCE-AR) (Line 8 in Algorithm 2)

```

1: Input: Target densities  $\pi_\ell, \pi_{\ell-1}$ ; current states  $\theta_\ell^i, \vartheta_{\ell-1}^i$ ; adaptive parameters  $\lambda_\ell^i, \Sigma_\ell^i, \lambda_{\ell-1}^i, \Sigma_{\ell-1}^i$ ; resync weight  $\omega_\ell$ 
2: Output: Next states  $\{\theta_\ell^*, \vartheta_{\ell-1}^*\}$ 
3: Sample  $w \sim \mathcal{U}(0, 1)$ 
4: Sample  $\eta \sim \mathcal{N}(0, \mathcal{I}_d)$ 
5: if  $w \leq \omega_\ell$  then
6:    $(\theta_\ell^*, \vartheta_{\ell-1}^*) = \text{Algorithm 6/Algorithm 3}$   $\triangleright \text{resynchronize}$ 
7: else if  $w > \omega_\ell$  then
8:   Set  $\theta_\ell^* = \theta_\ell^i + \lambda_\ell^i \sqrt{\Sigma_\ell^i} \eta$  and  $\vartheta_{\ell-1}^* = \vartheta_{\ell-1}^i + \lambda_{\ell-1}^i \sqrt{\Sigma_{\ell-1}^i} \eta$   $\triangleright \text{SYNCE-A}$ 
9: end if
10: return  $(\theta_\ell^*, \vartheta_{\ell-1}^*)$ 

```

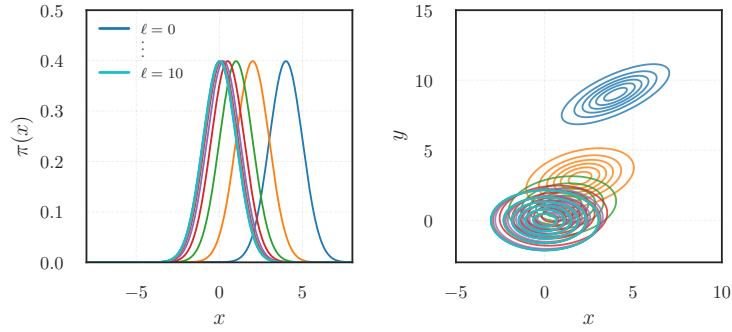


Figure 2: Posterior distributions for the Gaussian examples at different levels. The figure on the left corresponds to section 5.1 and the figure on the right corresponds to section 5.2.

5. Experimental results

In this section, we validate the performance of the SYNCE coupling method through four numerical examples. The first example is a simple Gaussian taken from [20] to illustrate the effectiveness and sanity of the proposed methodology. The second example is a two-dimensional extension to the first Gaussian problem that highlights the need for adaptation and resynchronization in the SYNCE coupling method. These two examples are not BIP involving a likelihood function, but are useful to illustrate the performance of the proposed coupling method. The third example is a prey-predator model governed by a system of ordinary differential equations (ODEs) taken from [17] to compare the proposed method with the AEM enhanced MLDA coupling method described in section 3.3.1. The final example involves uncertainty quantification for Darcy flow [20, 15]. This example is a well known benchmark and will be used to show that the proposed methodology satisfies the assumptions needed for the ϵ -cost complexity theorem in the supplementary material section S1.

5.1. Shifting Gaussian

In this example, we sample from shifting Gaussians parameterized as $\pi_\ell = \mathcal{N}(2^{-\ell+2}, 1)$ for $\ell = 0, 1, \dots, L$. The target posteriors are plotted on the left in Fig. 2. This setup serves as a robust test of coupling efficiency, particularly at coarser levels (e.g., $\ell = 1$ and $\ell = 2$) where the distributions are significantly shifted despite sharing similar shapes. We compare four methods: the delayed acceptance coupling with hierarchical subsampling (Algorithm 4), the independent proposal coupling method (Algorithm 6), the maximal coupling method (Algorithm 7) and the proposed SYNCE coupling method (Algorithm 8). We fix $L = 6$ and use the same Gaussian proposal $q_\ell = \mathcal{N}(2, 3)$ across all levels for the relevant algorithms. The proposal distribution for the coarsest $\ell = 0$ posterior is set to $q_0 = \mathcal{N}(\cdot, 1)$ and 50,000 samples are used for all levels according to [20]. We use the same proposal covariance for SYNCE with $C_\ell = 3$ to provide a fair comparison.

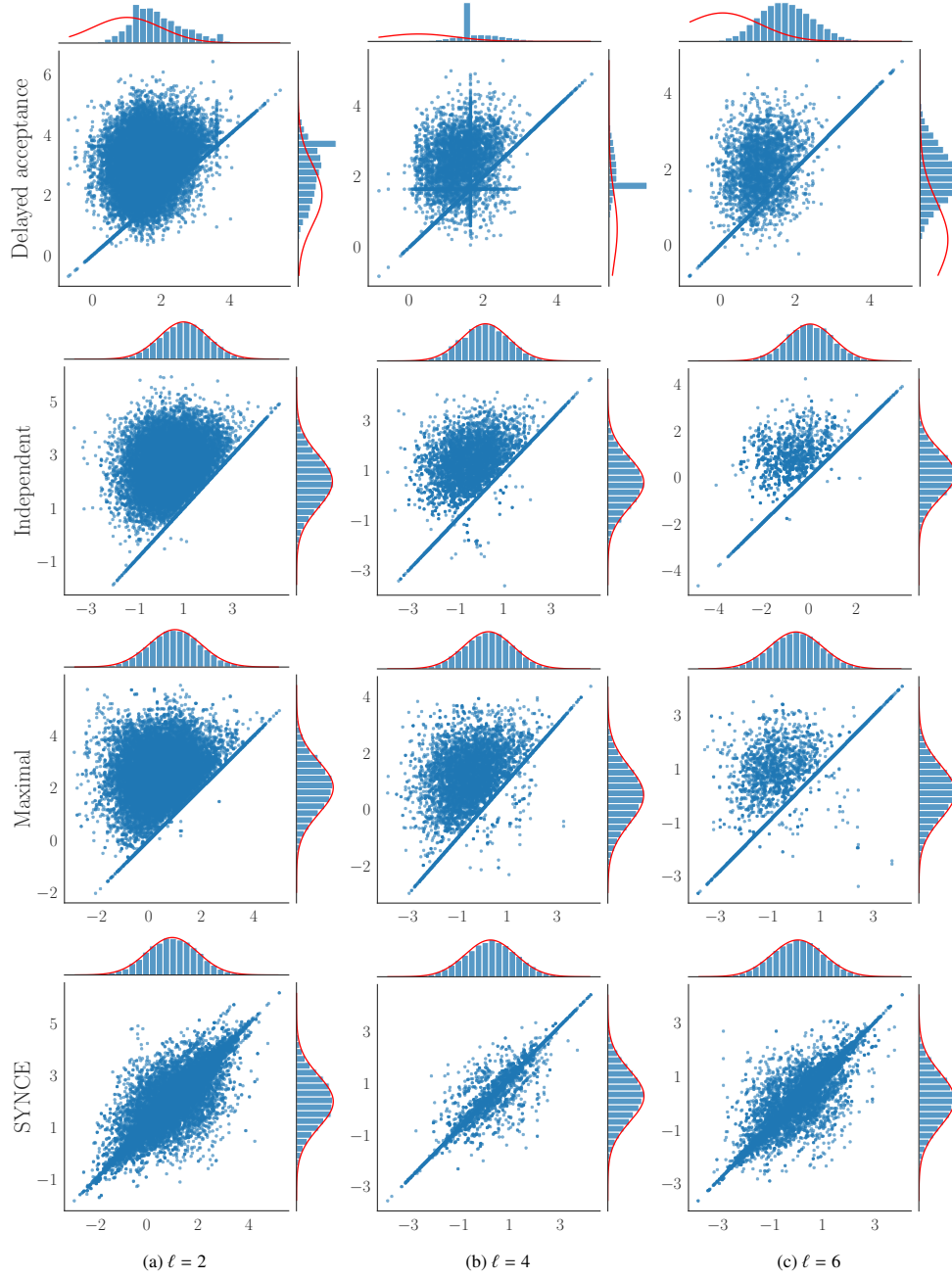


Figure 3: Scatter plots and histograms of samples from the different coupling algorithms for the shifting Gaussian example. For each level, the X-axis represents the samples at the ℓ 'th level and the Y-axis represents the samples at the $(\ell - 1)$ 'th level. The red line on the histogram plots represents the true posterior distribution at the respective levels. Existing methods either fail to correctly sample from the marginals (row 1), produce diffused scatter plots with weak correlation despite correct marginals (row 2), or exhibit degrading diagonal concentration as posteriors diverge (row 3). SYNCE (row 4) consistently produces tightly clustered points along a dominant diagonal across all levels – indicating robust correlation. The adaptive diagonal slope is not constrained to one, reflecting the relative shift between the posteriors while maintaining right marginals.

Fig. 3 provides validation if the coupling algorithms are able to sample from the correct marginal distributions and demonstrate the joint distribution of levels $\ell = \{2, 4, 6\}$ for the four methods. In the first row of Fig. 3, the DA

with hierarchical subsampling method fails to correctly sample from the marginal distributions. The DA method fails because the hierarchical subsampling scheme violates the tail decay assumption required for ergodicity, as discussed in section 3.3.2. The other three methods all correctly recover the true marginals, indicating that they are ergodic. However, the joint scatter plots differ significantly between the proposed SYNCE coupling method and the existing methods. The samples from the three existing methods are either perfectly coupled (lying on the line with slope one) or uncorrelated (scattered points). This behavior confirms that these methods only induce correlation when the samples fall into the shared overlapping region of the posteriors. Outside the shared region, the samples become uncorrelated, leading to scattered joint distributions above the diagonal (posteriors shifting from right to left). This scattering effect is more pronounced at coarser levels where the posteriors are more dissimilar. In contrast, the SYNCE coupling method produces a tightly clustered diagonal line with a slope that is not necessarily equal to one. This behavior indicates that SYNCE does not care about the overlap between the posteriors, effectively capturing the shift between them without any prior knowledge about their distributions.

In Fig. 4, we show how well the samples are correlated between two successive MLMC levels. We plot the Pearson correlation coefficient in Eq. (8) between the samples at the ℓ 'th level and the $(\ell - 1)$ 'th level for the different coupling algorithms. We see that the delayed acceptance, independent proposal and the maximal coupling methods produce samples that are not well correlated at the coarser levels. This weak correlation is a direct consequence of the fact that the posteriors at these coarse levels have minimal overlapping posterior regions. However, our method does not depend on this overlap and rather tries to couple (different) samples by proposing the same noise in the Gaussian random walk. This is why the SYNCE coupling method achieves the largest correlation, particularly for coarser levels.

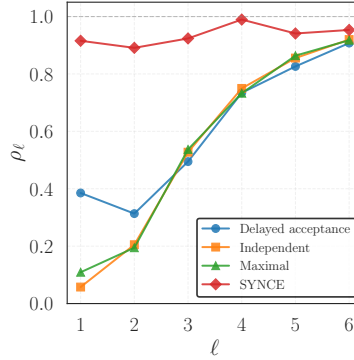


Figure 4: Pearson correlation coefficient between samples of all levels for the shifting Gaussian example. The coefficient is a linear measure of how well samples are correlated. Higher correlations yield greater variance reduction and we see that the SYNCE coupling algorithm achieves the greatest correlation, particularly for the coarser levels of the hierarchy.

5.2. Rotating-shifting Gaussian

Next, we consider a 2D family of Gaussians to demonstrate advantages of the SYNCE coupling method under varying covariance structures. The posterior densities are given by:

$$\pi_\ell = \mathcal{N}\left(\mu'_\ell = \begin{bmatrix} 2^{-\ell+2} \\ 3^{-\ell+2} \end{bmatrix}, \Sigma'_\ell = \begin{bmatrix} 2 & 2^{-\ell} \\ 2^{-\ell} & 1 \end{bmatrix}\right), \quad (15)$$

for $\ell = 0, 1, \dots, L$. The posterior distributions are visualized in Fig. 2. This example extends the first problem by introducing rotation in addition to the shift. This added complexity allows us to isolate and highlight the importance of adaptation and resynchronization discussed in sections 4.4 and 4.5. We run two different experiments with the number of levels set to $L = 6$ and the number of samples in each level set to 50,000. We discard the first 20,000 samples as burn in and apply any adaptation in this period to get a good estimate of the adaptation parameters. In the first experiment, we compare the baseline SYNCE coupling method without adaptation and resynchronization

(Algorithm 8) with the maximal coupling method (Algorithm 7) for Gaussian random walk proposals. In the second experiment, we compare three algorithms: SYNCE-A, SYNCE-AR (Algorithm 9) and the independent proposal (Algorithm 6) coupling method. SYNCE-A refers to the proposed SYNCE coupling method with only adaptation i.e., setting $\omega_\ell = 0$ for $\ell = 1, 2, \dots, L$ in Algorithm 9.

5.2.1. Experiment 1: Baseline SYNCE vs maximal

In this experiment, we set a fixed proposal distribution $q_\ell = \mathcal{N}(\cdot, [[3, 0], [0, 3]])$ for all the levels $\ell = 0, 1, \dots, L$. This proposal is chosen to ensure a 44% acceptance rate for the chains at all levels, the optimal value for Gaussian random walk proposals according to [39]. The consistent choice of proposal distribution also ensures a fair comparison between the two algorithms.

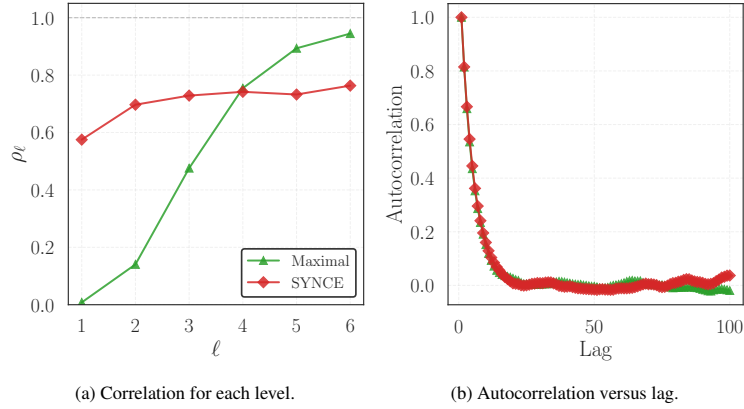


Figure 5: Comparison of the maximal coupling method and baseline SYNCE on the rotating-shifting Gaussian example (first dimension). (a) Level-wise correlation showing SYNCE's advantage at lower levels but disadvantage at finer levels. (b) Autocorrelation at the finest level, confirming comparable sampling efficiency because of similar proposal distributions.

In Fig. 5a, we plot the correlation between the samples at all levels for both the methods for the first dimension. The maximal coupling method produces samples that are highly correlated at finer levels and very poorly correlated at the coarser levels. The dissimilarity between the posteriors at the coarse levels results in a high total variation distance for the coupled kernel, leading to a low probability of a common sample being proposed for the two chains. Low probability of proposing the same sample results in poor correlation between the samples for the coarse levels as seen in the figure. SYNCE however, does produce samples that are well correlated at all levels, and, there are two important observations to be made. First, there is a noticeable increasing trend in the correlation, which appears to saturate to some fixed value as we move from coarse to fine levels. At the coarsest level $\ell = 1$, the correlation is significantly smaller compared to the other levels. This low correlation is attributed to the distinct shape (covariance structure) between the two posteriors π_0 and π_1 as seen in Fig. 2. Second, the correlation at the finest level is not as high as the maximal coupling method. Even though the posteriors are quite close in shape and distance, the two chains are not exactly in sync as our method does not propose the same candidate sample for the two chains. While taking the same step is beneficial at the coarser levels, it presents drawbacks at the finer levels. In Fig. 5b, we plot the autocorrelation of the samples obtained from the two methods for the finest level. As expected, both the methods show similar decays in the autocorrelation with the same covariance of the proposal distribution. The moderate correlation at the finer levels suggest that we need modifications in our SYNCE algorithm. In the next experiment, we show how adaptation and resynchronization serve this purpose.

5.2.2. Experiment 2: SYNCE-A vs SYNCE-AR vs independent proposal coupling

In this experiment, we compare the SYNCE coupling method with only adaptation (SYNCE-A), the SYNCE coupling method with adaptation and resynchronization (SYNCE-AR), and the independent proposal coupling method.

For SYNCE-A, we set the resynchronization weight $\omega_\ell = 0$ for all levels $\ell = 1, 2, \dots, L$ in Algorithm 9. For SYNCE-AR, we set $\omega = [0, 0, 0, 0.2, 0.3, 0.5]$ for their respective levels and choose the level wise independent proposal distribution $q_\ell = \mathcal{N}(\mu_\ell^{\text{IMH}}, [[3, 0], [0, 3]])$ with $\mu_\ell^{\text{IMH}} = (\mu'_\ell + \mu'_{\ell-1})/2$ as the resynchronization proposal. This proposal is chosen to satisfy the tail decay assumption and to achieve optimal mixing properties for both the coupled chains. Finally, we set the target acceptance rate to 44% for all levels.

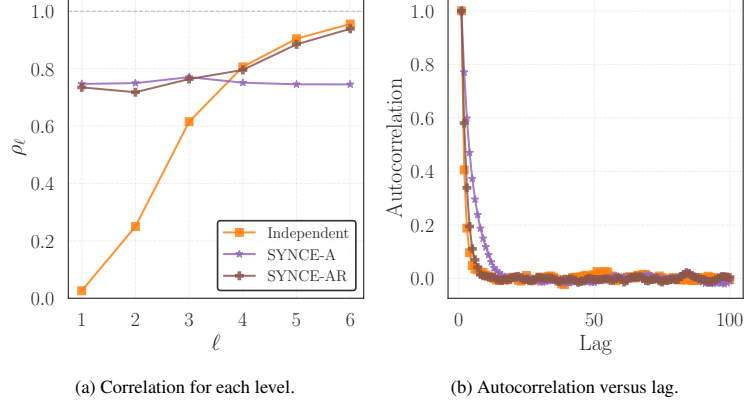


Figure 6: Comparison of independent, SYNCE-A, and SYNCE-AR on the rotating-shifting Gaussian example (first dimension). (a) Level-wise correlation highlighting improved correlation compared to Fig. 5a because of adaptation in the lower levels and resynchronization in the higher levels. (b) Autocorrelation at the finest level, confirming improved mixing with resynchronization.

In Fig. 6a, we provide correlation plots as before. The effect of adaptation at the coarser levels is clear when comparing this figure with Fig. 5a. Rescaling each individual chain’s proposal distribution to account for the respective posterior’s covariance structure ensures that the chains stay in sync as much as possible while exploring the posterior space. However, at the finer levels, we are still limited by the fact that the same sample is not proposed for the two chains. Introducing resynchronization solves this limitation, promoting better syncing of the chains, especially at the finest levels where we expect the posteriors to be quite close to each other. In Fig. 6b, we plot the autocorrelation of the samples for the finest level for the first dimension. Note the improvement in the autocorrelation decay when we introduce resynchronization to the SYNCE coupling method because it promotes better mixing in fine levels. Overall, adaptation is crucial for improving correlation at coarser levels, while resynchronization enhances correlation and mixing at finer levels.

5.3. Prey-predator problem

In this BIP, we adapt the Lotka-Volterra prey-predator model problem from [17] to compare the proposed SYNCE coupling method with the AEM enhanced MLDA method described in section 3.3.1. The original problem in [17] is not directly suitable for this comparison for two main reasons. First, the fidelity levels in the original setup are obtained by varying the time domain over which the model is solved. This choice of fidelity creates inconsistent data across levels, making it unsuitable for the AEM. Second, the resulting posterior distributions in the original problem across levels are quite similar to each other, which does not pose a challenge to multi-fidelity sampling.

To address these issues, we propose a modified version of the prey-predator problem. In this version, we maintain the same time domain of $\mathcal{T} = [0, 12]$ across all fidelity levels but introduce model discrepancy by adding some additional terms to the governing equations at coarser levels. This approach creates dissimilar posterior distributions across levels, providing a more rigorous test case (see supplementary material S2). The modification allows for comparison between the SYNCE method and the AEM enhanced MLDA method, implemented in the `tinyDA` library.

The Lotka-Volterra model describes the interaction between populations of prey (N) and predator (P) species over

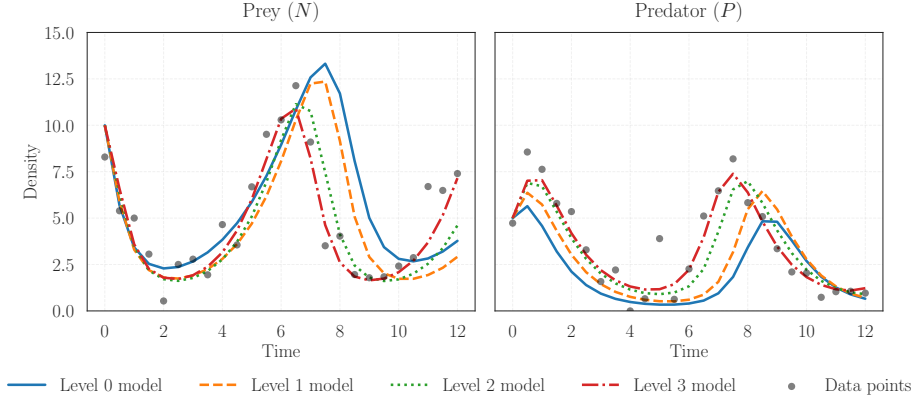


Figure 7: Model evaluations and data for the four fidelity levels of the prey-predator problem for the true parameters. Introducing model discrepancy terms creates differences between the model predictions at different fidelity levels.

time [17]. The governing equations of the *general* model are given by a system of ODEs as:

$$\begin{aligned} \frac{dN}{dt} &= aN - bNP - \epsilon_N N + \delta_N, \\ \frac{dP}{dt} &= cNP - dP - \epsilon_P P, \end{aligned} \quad (16)$$

where a, b, c, d are the model parameters and $\epsilon_N, \epsilon_P, \delta_N$ represent model discrepancy terms. The initial populations of prey and predator species, denoted by N_0 and P_0 respectively, are also considered as uncertain parameters. Thus, the complete set of uncertain parameters is given by $\theta = [N_0, P_0, a, b, c, d] \in \mathbb{R}^6$. We consider $L = 3$ levels and describe each level by the values of the model discrepancy terms as shown in the table below. For the BIP, we assume Gaussian

Table 2: Model discrepancy terms for different fidelity levels

Level (ℓ)	ϵ_N	ϵ_P	δ_N
3	0.0	0.0	0.0
2	0.1	0.0	0.0
1	0.3	0.1	0.0
0	0.5	0.2	0.3

prior distributions for all the parameters with means and variances as given in [17]. The models are solved using the SciPy library's `solve_ivp` function with the RK45 method. The observed data is generated by solving the finest level model with a true set of parameters $\theta_{\text{true}} = [10, 5, 1.0, 0.3, 0.2, 1.0]$ and adding Gaussian noise with zero mean and variance $\sigma_{\text{noise}}^2 = 1.0^2$ at 25 equally spaced time points in the time domain $\mathcal{T}$. The four fidelity models along with the observed data are shown in Fig. 7.

We run Algorithm 2 with both the coupling methods for 12,000 samples for the finest level and discard the first 2,000 samples as burn in. For the other three levels, we run the same 12,000 samples for SYNCE coupling method and set the subsampling rates $J_1 = J_2 = J_3 = 5$ for the AEM enhanced MLDA method [17]. This setting results in 60,000, 300,000, and 1,500,000 samples for the coarser fidelities. The proposal distributions for both the methods are set to Gaussian random walks with covariance matrices tuned to achieve an acceptance rate of approximately 44% for all levels. Further, we only use adaptation and no resynchronization for this problem for the SYNCE algorithm. We run 5 independent runs of the ML-MCMC algorithm for both the coupling methods and assume that the cost of evaluating each model follows the decay law: $C_\ell = 0.001 \times 2^\ell$ with $\gamma = 1$ (see supplementary material Assumption S1.1). The reported minimum correlation ρ_L is the median (over 5 runs) of the minimum correlation between the L and $L-1$ level

chains across all 6 parameter dimensions. The minimum ESS is computed similarly using the finest level samples. Total cost is $C_{\text{tot}} = \sum_{\ell=0}^L N_{\ell} C_{\ell}$. The comparisons are summarized in Table 3 for two, three and four levels.

Table 3: Comparison of the SYNCE coupling method with the AEM enhanced MLDA method for the prey-predator problem. While MLDA-AEM achieves high correlation and ESS for few levels, its exponential cost scaling results in poor efficiency per unit cost as the number of levels increase. In contrast, SYNCE demonstrates superior scalability, achieving twice the sampling efficiency in the four-level case.

Method	C_{tot}	Min. ρ_L	Min. ESS	Min. ρ_L/cost	Min. ESS/cost
<i>Two-level comparison</i>					
MLDA-AEM	84	0.75	753	0.0089	8.96
SYNCE	36	0.60	226	0.0167	6.27
<i>Three-level comparison</i>					
MLDA-AEM	468	0.50	971	0.0010	2.07
SYNCE	108	0.61	326	0.0056	3.02
<i>Four-level comparison</i>					
MLDA-AEM	2436	0.40	1374	0.0001	0.56
SYNCE	252	0.60	299	0.0023	1.18

MLDA with AEM consistently achieves high raw ESS across all level settings, demonstrating that the AEM effectively mitigates posterior discrepancies. However, this gain in sampling efficiency comes by recursive subsampling, forcing the coarse chains to run significantly long chains to serve as good proposals for the finer levels. As shown in Table 3, the total computational cost for MLDA-AEM grows exponentially with the number of levels, with the four-level case being nearly an order of magnitude more expensive than SYNCE at the same level. In contrast, SYNCE relies solely on the pairwise coupling of adjacent fidelity levels, avoiding the dependence on long coarser chains. This decoupling isolates the correlation performance from the number of levels, resulting in linear cost scaling and consistent correlation across levels.

It is important to distinguish between mixing efficiency and coupling quality. The lower ESS observed in SYNCE is primarily due to the choice of proposal distributions rather than the coupling strategy itself. Since SYNCE’s mixing properties depend primarily on the proposal distributions (see Section 4.3), its sampling efficiency can be improved by integrating advanced proposal mechanisms. For instance, employing transport map proposals [35] could significantly enhance mixing while retaining strong correlation. We leave this exploration for future work.

Ultimately, when evaluating efficiency per unit cost, SYNCE demonstrates superior scalability compared to MLDA-AEM as the number of fidelity levels increase. In the two level case, MDLA-AEM achieves higher correlation and ESS, but at double the computational cost, rendering per unit cost metrics similar for both the methods. However, at the four-level setting, the computational overhead and posterior mismatch lead to pronounced efficiency degradation for MLDA-AEM. While MLDA-AEM excels for few closely related models, SYNCE is more suitable for problems with a large hierarchy of models and significant posterior dissimilarities. A detailed visualization of the posterior distributions for both methods is provided in S2.

5.4. Groundwater problem

To assess whether SYNCE’s parameter coupling strategy achieves effective variance reduction in forward uncertainty quantification, we consider the 2D Darcy’s subsurface flow equation. This is the model problem explored in great detail in [12, 15, 20]. The exact problem is taken from [20] and the governing equation of the problem is given by:

$$-\nabla_x \cdot (\kappa(x, \theta) \nabla_x u(x, \theta)) = 1, \quad x = (x_1, x_2) \in \Omega = [0, 1] \times [0, 1], \quad (17)$$

where κ represents the permeability, u represents the pressure head and $\theta \in \mathbb{R}^d, d = 4$ represents the uncertain parameters. We close the equations by providing boundary conditions on the domain as:

$$u|_{x_1=0} = 0, \quad u|_{x_1=1} = 0, \quad \partial_n u|_{x_2=0} = 0, \quad \partial_n u|_{x_2=1} = 0, \quad (18)$$

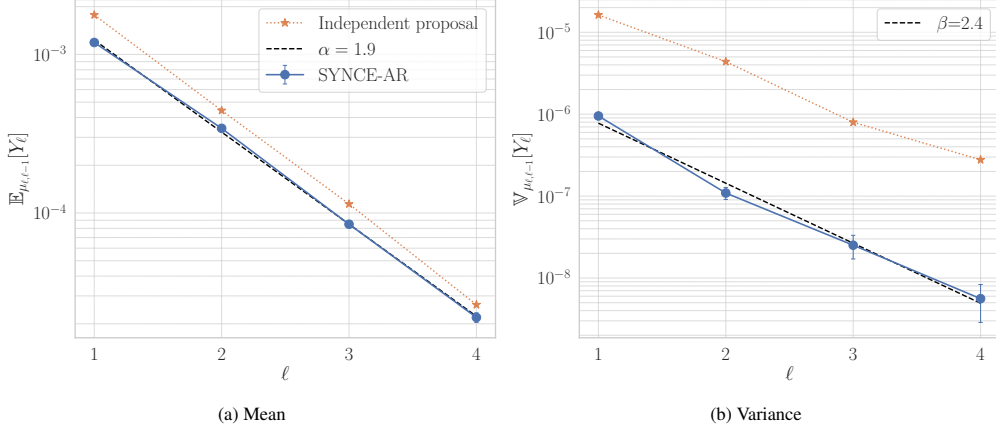


Figure 8: Decay of error plots for the groundwater problem. The independent proposal coupling and the SYNCE-AR methods are compared. Our method leads to almost an order of two gain in variance reduction compared to the result from [20].

where ∂_n is the normal derivative with n pointing outwards from the domain. The first two conditions are the Dirichlet boundary conditions applied on the left and right boundary of the domain, and the last two conditions are the Neumann boundary conditions applied on the bottom and top boundary of the domain. The permeability field is given by [20]:

$$\kappa(x, \theta) = \exp\left(\theta_1 \cos(\pi x) + \frac{\theta_2}{2} \sin(\pi x) + \frac{\theta_3}{3} \cos(2\pi x) + \frac{\theta_4}{4} \sin(2\pi x)\right) \quad (19)$$

The inverse problem is to infer the uncertain parameters $\theta = [\theta_1, \theta_2, \theta_3, \theta_4]$ given some observed pressure head data. We generate this data by solving the forward problem Eq. (17) on the finest grid with a random sample of $\theta_{\text{true}} \sim \mathcal{N}(0_d, \mathcal{I}_d)$ and observing the pressure head corrupted by a Gaussian noise at 4×4 equally spaced points inside the domain. Gaussian noise is added to the observed data with a zero mean and variance $\sigma_{\text{noise}}^2 = 0.01^2$. We set the number of levels $L = 4$ and for each level ℓ , the solution of Eq. (17) is obtained using the finite element method with a mesh size of $8 * 2^\ell \times 8 * 2^\ell$ triangular elements, using the FEniCS library [41].

The number of levels in the original problem in [20] is set to $L = 3$ with the coarsest mesh at 16×16 triangular elements. We set the number of levels to $L = 4$ with the coarsest mesh at 8×8 triangular elements as we can leverage the SYNCE coupling method to correlate samples even at very coarse levels. This is a direct consequence of the fact that the SYNCE coupling method does not depend on the closeness of the posteriors to correlate samples.

The quantity of interest (QoI) is the average pressure over the domain given by:

$$Q_\ell(\theta_\ell) = \int_{x \in \Omega} u_\ell(x, \theta_\ell) dx, \quad (20)$$

where Ω is the domain of the problem.

To empirically verify that SYNCE satisfies the decay rate assumptions required for ML-MCMC estimator convergence (see supplementary material Assumption S1.1), we run Algorithm 2 with the proposed states obtained from Algorithm 9. We set the number of samples in each level to 10,000 and consider the first 4,000 samples as burn-in. We run the algorithm for 10 independent runs and plot the decay of errors. For the coarsest chain ($\ell = 0$), we set the target acceptance ratio as 44% and run an adaptive MH-MCMC algorithm for that level. The two level wise adaptation parameters, the target acceptance ratio and the weight for resynchronization are set to 70% for each level ($\ell > 0$) and $\omega = [0.0, 0.3, 0.5, 0.7]$. These parameters were chosen to ensure low lag and high correlation between the samples at each level.

Fig. 8 compares our proposed methodology to an existing result using the independent proposal coupling method [20] for the same problem. The results for the independent proposal coupling method were obtained by digitizing Fig. 12 in [20]. We observe that the SYNCE-AR method leads to greater variance reduction for all the level wise difference estimators $\hat{Y}_\ell$. Even though we use a coarser level compared to the original work, we are able to achieve almost an order of two gain in variance reduction at all levels. This is a direct consequence of the proposed coupling method that

is able to correlate samples even at very coarse levels. The correlation values for the output functional at the different levels are: 0.84, 0.97, 0.99, 1.0 for $\ell = 1, 2, 3, 4$.

6. Conclusions

We introduced SYNCE, a new coupling methodology for multilevel Markov chain Monte Carlo estimators that achieves superior variance reduction compared to existing coupling approaches. Variance reduction in these estimators depends on the correlation between the samples at successive fidelity levels, which in turn depends on the coupling strategy used to construct the joint proposal distribution for the two chains. Existing coupling strategies depend on some form of overlap or similarity between the posteriors at different levels to correlate samples. This dependence limits the effectiveness of these coupling strategies when the posteriors are dissimilar or non-overlapping, which is often the case in practical applications. SYNCE however, does not depend on this overlap and couples chains by proposing the same noise in the Gaussian random walk proposals for both chains. We rigorously proved the existence of a joint invariant measure for the SYNCE kernel and established geometric ergodicity under standard assumptions. We also showed that the convergence rate of SYNCE is superior to existing coupling approaches when the posteriors are dissimilar or non-overlapping.

By utilizing adaptation and resynchronization, SYNCE adaptively rescales the proposal distributions and resynchronizes the chains to enhance correlation. Adaptation helps in accounting for the differences in the covariance structure of the posteriors at different levels, while resynchronization helps in promoting better mixing properties of the chains when posteriors are quite close. These enhancements enable the use of coarser model fidelities in the estimation procedure, yielding substantial computational savings. Numerical experiments on synthetic and real-world problems validate the effectiveness of the proposed methodology.

The SYNCE coupling method is the baseline realization of the proposed coupling framework we have set up. Future work will include exploring more general and nonlinear couplings. Transport maps [35, 34] may help with mixing properties of the fine chain by learning from the coarse chain. Another possible extension to this work is to explore the use of the SYNCE coupling methodology in the context of approximate control variates. These estimators provide additional performance improvements by using a control variate weight that is estimated adaptively. Additionally, the proposed methodology can be extended in the unbiased estimation context to eliminate the burn in bias [22, 23].

Author Contributions

Sanjan Muchandimath: Conceptualization, Methodology, Software, Validation, Writing - original draft. **Alex Gorodetsky:** Conceptualization, Methodology, Writing - Review & Editing, Supervision, Funding acquisition.

Acknowledgements

This work was funded by the Aeronautics Research Mission Directorate at NASA through the Transformative Aeronautical Concepts Program (TACP) and the D.2 Transformational Tools and Technologies Project (TTT) under contract no. 80NSSC23M0215.

Data Availability

Available soon here

Appendix A. Proof of Theorem 4.1

We prove Theorem 4.1 by establishing that the coupled kernel in Eq. (10) is ψ -irreducible and aperiodic, and that it satisfies minorization and drift conditions.

Lemma A.1 (ψ -irreducibility). *Under Assumption 4.1, the coupled kernel p_ℓ^{SYNCE} in Eq. (10) is ψ -irreducible.*

PROOF. Let ψ be the Lebesgue measure on X^2 . There exists a rectangle $K^2 = K \times K \in A$ with positive measure for any set A with $\psi(A) > 0$. Here, $K = \{x : |x| \leq R_K\}$ for some $R_K > 0$, a ball of radius R_K . For any starting pair $\theta_\ell = (\theta_\ell, \vartheta_{\ell-1}) \in X^2$,

$$\begin{aligned} p_\ell^{\text{SYNCE}}(\theta_\ell, A) &\geq p_\ell^{\text{SYNCE}}(\theta_\ell, K^2), \\ &\geq \int_{X^2} \min(\alpha_\ell(\theta_\ell, \theta_\ell^*), \alpha_{\ell-1}(\vartheta_{\ell-1}, \vartheta_{\ell-1}^*)) \Gamma_\ell(\theta_\ell | d\theta_\ell^*) \delta_{\theta_\ell^*}(K^2), \\ &\geq \int_X \min(\alpha_\ell(\theta_\ell, \theta_\ell + \eta), \alpha_{\ell-1}(\vartheta_{\ell-1}, \vartheta_{\ell-1} + \eta)) \mathbb{1}_{K^2}(\theta_\ell + \eta, \vartheta_{\ell-1} + \eta) q(\eta) d\eta. \end{aligned}$$

Recall from Algorithm 8 that the proposed samples are generated by $\theta_\ell^* = \theta_\ell + \eta$ and $\vartheta_{\ell-1}^* = \vartheta_{\ell-1} + \eta$ where $\eta \sim q$, $q = \mathcal{N}(0, C_\ell)$. The third inequality arises by restricting on the event that both marginals propose into the set K i.e. $\theta_\ell + \eta \in K$ and $\vartheta_{\ell-1} + \eta \in K$. For this condition to hold, the increment η must satisfy $\eta \in B = \{x : |x| \leq R_K + \max\{|\theta_\ell|, |\vartheta_{\ell-1}|\}\}$. By assumption (A.3), $q(\eta)$ is bounded below on the set B . Also, for positive and continuous $\pi_\ell, \pi_{\ell-1}$ as stated in assumption (A.2), $\min(\alpha_\ell(\theta_\ell, \theta_\ell^*), \alpha_{\ell-1}(\vartheta_{\ell-1}, \vartheta_{\ell-1}^*)) > 0$. Therefore $p_\ell^{\text{SYNCE}}(\theta_\ell, K^2) > 0$. Because the above holds for any set A with positive Lebesgue measure and any starting point θ_ℓ , the kernel is ψ -irreducible for $n = 1$. Furthermore, aperiodicity follows because the kernel is ψ -irreducible and the chain can stay in the same state with positive probability. □

Lemma A.2 (Minorization condition). *Under Assumption 4.1, the coupled kernel p_ℓ^{SYNCE} in Eq. (10) satisfies a one step minorization condition on any compact set $S \in \mathcal{X}^2$.*

PROOF. Let $C_\ell, C_{\ell-1} \in X$ be compact sets, both contained in $B(0, R)$ and define $C = C_\ell \times C_{\ell-1}$ as the set of starting points. Let B be the set from assumption (A.3) i.e. $B := \{\eta : |\eta| \leq \delta_q\}$. Define the intersection set D as the larger compact set containing all possible proposals $D := \{(\theta_\ell + \eta, \vartheta_{\ell-1} + \eta) : \eta \in B\}$. Then, for any starting point $\theta_\ell \in C$ and any set $A \in \mathcal{X}^2$, we have

$$\begin{aligned} p_\ell^{\text{SYNCE}}(\theta_\ell, A) &\geq \int_{X^2} \min(\alpha_\ell(\theta_\ell, \theta_\ell^*), \alpha_{\ell-1}(\vartheta_{\ell-1}, \vartheta_{\ell-1}^*)) \Gamma_\ell(\theta_\ell | d\theta_\ell^*) \delta_{\theta_\ell^*}(A). \\ &\geq \int_B \min(\alpha_\ell(\theta_\ell, \theta_\ell + \eta), \alpha_{\ell-1}(\vartheta_{\ell-1}, \vartheta_{\ell-1} + \eta)) \mathbb{1}_A(\theta_\ell + \eta, \vartheta_{\ell-1} + \eta) q(\eta) d\eta, \end{aligned}$$

From assumption (A.2), the posterior densities $\pi_\ell, \pi_{\ell-1}$ are bounded above and below away from zero on any compact set. Therefore,

$$\min(\alpha_\ell(\theta_\ell, \theta_\ell + \eta), \alpha_{\ell-1}(\vartheta_{\ell-1}, \vartheta_{\ell-1} + \eta)) \geq \frac{\pi_{\min}}{\pi_{\max}}.$$

We also use assumption (A.3) to bound $q(\eta)$ below by $\epsilon_q > 0$ on the set B . Therefore,

$$p_\ell^{\text{SYNCE}}(\theta_\ell, A) \geq \epsilon_q \frac{\pi_{\min}}{\pi_{\max}} \int_B \mathbb{1}_A(\theta_\ell + \eta, \vartheta_{\ell-1} + \eta) d\eta.$$

The integral term represents the volume of the set A reachable by the synchronized proposal from θ_ℓ . Because we are proposing the same increment η to both chains, this volume is non-zero only when A contains points along the line defined by $\{(\theta_\ell + \eta, \vartheta_{\ell-1} + \eta) : \eta \in B\}$, which is not the case in general. Hence this term does not define a non-trivial probability measure. However, since the chain is ψ -irreducible (Lemma A.1) and possesses a continuous density, we

can utilize Theorem 6.2.5 from [42] to guarantee that any compact set C is small. Intuitively, this theorem states that as long as the kernel has a continuous density (the same step move) and is ψ -irreducible (eventually visit any state), then it guarantees that minorization holds on any compact set. Therefore, this theorem guarantees the existence of a non-trivial measure $\nu(A)$ and a constant $\delta > 0$ for the minorization condition:

$$p_\ell^{\text{SYNCE}}(\theta_\ell, A) \geq \delta \nu(A). \quad (\text{A.1})$$

Furthermore, the minorization constant scales as $\delta \propto \epsilon_q \pi_{\min} / \pi_{\max}$, where the proportionality constant depends on the overlap of the proposal distribution with the compact set C [44].

□

Lemma A.3 (Drift condition). *Under Assumption 4.1, the coupled kernel satisfies a drift condition with Lyapunov function $\hat{V}_\ell = 1/2 (V_\ell + V_{\ell-1})$, where V_j are the marginal Lyapunov functions for $j = \{\ell, \ell - 1\}$.*

PROOF. The proof adapts the strategy from [21]. Under Assumption 4.1, the marginal random walk chains with kernels $p_\ell, p_{\ell-1}$ satisfy drift conditions (Lemma 4.1 item (B.2)). Define the joint Lyapunov function as

$$\hat{V}_\ell(\theta_\ell) = \frac{1}{2} (V_\ell(\theta_\ell) + V_{\ell-1}(\vartheta_{\ell-1})), \quad (\text{A.2})$$

and $S = S_\ell \times S_{\ell-1}$ where S_ℓ and $S_{\ell-1}$ are the marginal small sets. Precisely, $S_\ell := \{\theta_\ell \in X : V_\ell(\theta_\ell) \leq \Delta_\ell\}$ and $S_{\ell-1} := \{\vartheta_{\ell-1} \in X : V_{\ell-1}(\vartheta_{\ell-1}) \leq \Delta_{\ell-1}\}$. The goal is to show that p_ℓ^{SYNCE} satisfies the drift condition with joint Lyapunov function $\hat{V}_\ell$ and joint small set S .

The state θ_ℓ can either lie in S or outside S . If $\theta_\ell \notin S$, then there are three possibilities — $\theta_\ell \in S_\ell, \vartheta_{\ell-1} \notin S_{\ell-1}$; $\theta_\ell \notin S_\ell, \vartheta_{\ell-1} \in S_{\ell-1}$ and $\theta_\ell \notin S_\ell, \vartheta_{\ell-1} \notin S_{\ell-1}$. In the first case, we have

$$p_\ell^{\text{SYNCE}} \hat{V}_\ell(\theta_\ell) \leq \frac{1}{2} (\Lambda_\ell V_\ell(\theta_\ell) + \Lambda_{\ell-1} V_{\ell-1}(\vartheta_{\ell-1})) + \frac{b_\ell}{2},$$

Since $\vartheta_{\ell-1} \notin S_{\ell-1}$, we have that $V_{\ell-1}(\vartheta_{\ell-1}) \geq \Delta_{\ell-1}$. Therefore,

$$\hat{V}_\ell(\theta_\ell) = \frac{1}{2} (V_\ell(\theta_\ell) + V_{\ell-1}(\vartheta_{\ell-1})) \geq \frac{1}{2} (1 + \Delta_{\ell-1}),$$

which implies that $1/2 \leq \hat{V}_\ell(\theta_\ell) / (1 + \Delta_{\ell-1})$. Using this in the previous inequality, we get

$$p_\ell^{\text{SYNCE}} \hat{V}_\ell(\theta_\ell) \leq \frac{1}{2} (\Lambda_\ell V_\ell(\theta_\ell) + \Lambda_{\ell-1} V_{\ell-1}(\vartheta_{\ell-1})) + \frac{b_\ell}{1 + \Delta_{\ell-1}} \hat{V}_\ell(\theta_\ell) \leq \left(\max(\Lambda_\ell, \Lambda_{\ell-1}) + \frac{b_\ell}{1 + \Delta_{\ell-1}} \right) \hat{V}_\ell(\theta_\ell).$$

Note that $\max(\Lambda_\ell, \Lambda_{\ell-1}) + b_\ell / (1 + \Delta_{\ell-1}) < 1$ because $\Lambda_j < 1$ and $0 < b_j < \infty$ for $j = \{\ell, \ell - 1\}$. The second case is similar and leads to a similar inequality,

$$p_\ell^{\text{SYNCE}} \hat{V}_\ell(\theta_\ell) \leq \left(\max(\Lambda_\ell, \Lambda_{\ell-1}) + \frac{b_{\ell-1}}{1 + \Delta_\ell} \right) \hat{V}_\ell(\theta_\ell).$$

The third case leads to

$$p_\ell^{\text{SYNCE}} \hat{V}_\ell(\theta_\ell) \leq (\max(\Lambda_\ell, \Lambda_{\ell-1})) \hat{V}_\ell(\theta_\ell).$$

Combining all three cases, we have that for $\theta_\ell \notin S$,

$$\begin{aligned} p_\ell^{\text{SYNCE}} \hat{V}_\ell(\theta_\ell) &\leq \hat{\Lambda} \hat{V}_\ell(\theta_\ell), \\ \hat{\Lambda}_{\text{SYNCE}} &= \max(\Lambda_\ell, \Lambda_{\ell-1}) + \max\left(\frac{b_\ell}{1 + \Delta_{\ell-1}}, \frac{b_{\ell-1}}{1 + \Delta_\ell}\right) < 1. \end{aligned} \quad (\text{A.3})$$

Finally, if $\theta_\ell \in S$, then

$$p_\ell^{\text{SYNCE}} \hat{V}_\ell(\theta_\ell) \leq \frac{1}{2} (\Lambda_\ell V_\ell(\theta_\ell) + \Lambda_{\ell-1} V_{\ell-1}(\theta_{\ell-1})) + \frac{1}{2} (b_\ell + b_{\ell-1}) \leq \hat{\Lambda}_{\text{SYNCE}} \hat{V}_\ell(\theta_\ell) + \hat{b},$$

where $\hat{b} = 1/2(b_\ell + b_{\ell-1})$. Combining Eqs. (A.3) and (A.4), we have that the coupled kernel p_ℓ^{SYNCE} satisfies a drift condition with Lyapunov function $\hat{V}_\ell$ given by

$$p_\ell^{\text{SYNCE}} \hat{V}_\ell(\theta_\ell) \leq \hat{\Lambda}_{\text{SYNCE}} \hat{V}_\ell(\theta_\ell) + \hat{b} \mathbb{1}_S(\theta_\ell). \quad (\text{A.4})$$

□

References

- [1] C. P. Robert, G. Casella, Monte Carlo Statistical Methods, Springer Texts in Statistics, Springer New York, 2004.
- [2] J. M. Hammersley, D. C. Handscomb, Monte carlo methods, in: Proceedings, volume 7, US Army Research Office., 1961, p. 17.
- [3] B. Peherstorfer, K. Willcox, M. Gunzburger, Survey of multifidelity methods in uncertainty propagation, inference, and optimization, SIAM Review 60 (2018) 550–591. doi:10.1137/16M1082469.
- [4] On control variate estimators, Computers & Operations Research 14 (1987) 219–225. doi:https://doi.org/10.1016/0305-0548(87)90024-4.
- [5] M. Emsermann, B. Simon, Improving simulation efficiency with quasi control variates, Stochastic Models 18 (2002) 425–448. doi:10.1081/STM-120014220.
- [6] G. Geraci, M. Eldred, G. Iaccarino, A multifidelity control variate approach for the multilevel monte carlo technique, Center for Turbulence Research Annual Research Briefs (2015) 169–181.
- [7] A. A. Gorodetsky, G. Geraci, M. S. Eldred, J. D. Jakeman, A generalized approximate control variate framework for multifidelity uncertainty quantification, Journal of Computational Physics 408 (2020) 109257. doi:https://doi.org/10.1016/j.jcp.2020.109257.
- [8] T. Pham, A. A. Gorodetsky, Ensemble approximate control variate estimators: Applications to multifidelity importance sampling, SIAM/ASA Journal on Uncertainty Quantification 10 (2022) 1250–1292. doi:10.1137/21M1390426.
- [9] T. O. Dixon, J. E. Warner, G. F. Bomarito, A. A. Gorodetsky, Covariance expressions for multifidelity sampling with multioutput, multistatistic estimators: Application to approximate control variates, SIAM/ASA Journal on Uncertainty Quantification 12 (2024) 1005–1049. doi:10.1137/23M1607994.
- [10] S. Heinrich, Multilevel monte carlo methods, in: S. Margenov, J. Waśniewski, P. Yalamov (Eds.), Large-Scale Scientific Computing, Springer Berlin Heidelberg, Berlin, Heidelberg, 2001, pp. 58–67.
- [11] M. B. Giles, Multilevel monte carlo methods, Acta Numerica 24 (2015) 259–328. doi:10.1017/S096249291500001X.
- [12] K. A. Cliffe, M. B. Giles, R. Scheichl, A. L. Teckentrup, Multilevel Monte Carlo methods and applications to elliptic PDEs with random coefficients, Comput. Visual Sci. 14 (2011) 3–15. doi:10.1007/s00791-011-0160-x.
- [13] F. Nobile, F. Tesei, A multi level monte carlo method with control variate for elliptic pdes with log-normal coefficients, Stochastic Partial Differential Equations: Analysis and Computations 3 (2015) 398–444. doi:10.1007/s40072-015-0055-9.

- [14] D. Schaden, E. Ullmann, On multilevel best linear unbiased estimators, *SIAM/ASA Journal on Uncertainty Quantification* 8 (2020) 601–635. doi:10.1137/19M1263534.
- [15] T. J. Dodwell, C. Ketelsen, R. Scheichl, A. L. Teckentrup, A hierarchical multilevel markov chain monte carlo algorithm with applications to uncertainty quantification in subsurface flow, *SIAM/ASA Journal on Uncertainty Quantification* 3 (2015) 1075–1108. doi:10.1137/130915005.
- [16] J. A. Christen, C. Fox, Markov chain monte carlo using an approximation, *Journal of Computational and Graphical Statistics* 14 (2005) 795–810. doi:10.1198/106186005X76983.
- [17] M. B. Lykkegaard, T. J. Dodwell, C. Fox, G. Mingas, R. Scheichl, Multilevel delayed acceptance mcmc, *SIAM/ASA Journal on Uncertainty Quantification* 11 (2023) 1–30. doi:10.1137/22M1476770.
- [18] T. Cui, G. Detommaso, R. Scheichl, Multilevel dimension-independent likelihood-informed mcmc for large-scale inverse problems, *Inverse Problems* 40 (2024) 035005. doi:10.1088/1361-6420/ad1e2c.
- [19] T. Cui, K. J. Law, Y. M. Marzouk, Dimension-independent likelihood-informed mcmc, *Journal of Computational Physics* 304 (2016) 109–137. doi:https://doi.org/10.1016/j.jcp.2015.10.008.
- [20] J. P. Madrigal-Cianci, F. Nobile, R. Tempone, Analysis of a class of multilevel markov chain monte carlo algorithms based on independent metropolis–hastings, *SIAM/ASA Journal on Uncertainty Quantification* 11 (2023) 91–138. doi:10.1137/21M1420927.
- [21] J. P. Madrigal Cianci, Hierarchical Markov chain Monte Carlo methods for Bayesian inverse problems, Ph.D. thesis, EPFL, Lausanne, 2022. doi:10.5075/epfl-thesis-8951.
- [22] P. E. Jacob, J. O’Leary, Y. F. Atchadé, Unbiased markov chain monte carlo methods with couplings, *Journal of the Royal Statistical Society Series B: Statistical Methodology* 82 (2020) 543–600. doi:10.1111/rssb.12336.
- [23] J. Heng, A. Jasra, K. J. H. Law, A. Tarakanov, On unbiased estimation for discretized models, *SIAM/ASA J. Uncertainty Quantification* 11 (2023) 616–645.
- [24] Y. Juntao, J. Adie, S. See, A. Gualandi, G. Mengaldo, A hybrid two-level mcmc framework to accelerate posterior mean estimation with deep learning surrogates for bayesian inverse problems, *Journal of Computational Physics* (2025) 114502. doi:https://doi.org/10.1016/j.jcp.2025.114502.
- [25] G. Geraci, M. S. Eldred, A. A. Gorodetsky, J. D. Jakeman, Leveraging active directions for efficient multifidelity uncertainty quantification, in: 6th European Conference on Computational Mechanics (ECCM 6), 2018, pp. 2735–2746.
- [26] R. R. Lam, O. Zahm, Y. M. Marzouk, K. E. Willcox, Multifidelity dimension reduction via active subspaces, *SIAM Journal on Scientific Computing* 42 (2020) A929–A956. doi:10.1137/18M1214123.
- [27] A. Chaudhuri, R. Lam, K. Willcox, Multifidelity uncertainty propagation via adaptive surrogates in coupled multidisciplinary systems, *AIAA Journal* 56 (2018) 235–249. doi:10.2514/1.J055678.
- [28] R. L. Pinto, R. M. Neal, Improving Markov chain Monte Carlo estimators by coupling to an approximating chain, Technical Report, Technical Report, 2001.
- [29] W. K. Hastings, Monte carlo sampling methods using markov chains and their applications, *Biometrika* 57 (1970) 97–109. doi:10.2307/2334940.
- [30] C. Villani, et al., Optimal transport: old and new, volume 338, Springer, 2009.
- [31] H. Thorisson, Coupling, Springer New York, New York, NY, 1998, pp. 319–339.
- [32] Y. F. Atchadé, J. S. Rosenthal, On adaptive markov chain monte carlo algorithms, *Bernoulli* 11 (2005) 815–828.

- [33] J. S. Rosenthal, Minorization conditions and convergence rates for markov chain monte carlo, *Journal of the American Statistical Association* 90 (1995) 558–566. doi:10.1080/01621459.1995.10476548.
- [34] B. Peherstorfer, Y. Marzouk, A transport-based multifidelity preconditioner for markov chain monte carlo, *Adv Comput Math* 45 (2019) 2321–2348. doi:10.1007/s10444-019-09711-y.
- [35] M. D. Parno, Y. M. Marzouk, Transport map accelerated markov chain monte carlo, *SIAM/ASA Journal on Uncertainty Quantification* 6 (2018) 645–682. doi:10.1137/17M1134640.
- [36] J. O’Leary, G. Wang, Metropolis–Hastings transition kernel couplings, *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques* 60 (2024) 1101 – 1124. doi:10.1214/22-AIHP1360.
- [37] H. Haario, E. Saksman, J. Tamminen, An adaptive Metropolis algorithm, *Bernoulli* 7 (2001) 223 – 242.
- [38] J. S. Rosenthal, et al., Optimal proposal distributions and adaptive mcmc, *Handbook of Markov Chain Monte Carlo* 4 (2011).
- [39] C. Pasarica, A. Gelman, Adaptively scaling the metropolis algorithm using expected squared jumped distance, *Statistica Sinica* 20 (2010) 343–364.
- [40] C. Andrieu, J. Thoms, A tutorial on adaptive MCMC, *Stat Comput* 18 (2008) 343–373. doi:10.1007/s11222-008-9110-y.
- [41] I. A. Baratta, J. P. Dean, J. S. Dokken, M. Habera, J. S. Hale, C. N. Richardson, M. E. Rognes, M. W. Scroggs, N. Sime, G. N. Wells, DOLFINx: The next generation FEniCS problem solving environment, 2023. doi:10.5281/zenodo.10447665.
- [42] S. P. Meyn, R. L. Tweedie, *Markov chains and stochastic stability*, Springer Science & Business Media, 2012. doi:10.1017/CB09780511626630.
- [43] G. O. Roberts, J. S. Rosenthal, General state space Markov chains and MCMC algorithms, *Probab. Surveys* 1 (2004). doi:10.1214/154957804100000024. arXiv:math/0404033.
- [44] S. F. Jarner, E. Hansen, Geometric ergodicity of Metropolis algorithms, *Stochastic Processes and their Applications* 85 (2000) 341–361. doi:10.1016/S0304-4149(99)00082-4.
- [45] K. L. Mengersen, R. L. Tweedie, Rates of convergence of the Hastings and Metropolis algorithms, *Ann. Statist.* 24 (1996). doi:10.1214/aos/1033066201.
- [46] Q. Qin, Convergence bounds for monte carlo markov chains, 2024. arXiv:2409.14656.