Accelerated MCMC combining metamodel-based independent proposals and delayed rejection

Jize Zhang
Graduate Student, Dept. of Civil & Environmental Engineering & Earth Sciences, University of Notre Dame, Notre Dame, USA

Alexandros A. Taflanidis
Associate Professor, Dept. of Civil & Environmental Engineering & Earth Sciences, University of Notre Dame, Notre Dame, USA

ABSTRACT: Markov Chain Monte Carlo (MCMC) simulation has significant computational burden when evaluation of the associated target probability density function (PDF) involves a complex numerical model. A novel framework to accelerate MCMC is developed here for such applications. It leverages a metamodel approximation of the target PDF to improve computational efficiency, while preserves convergence properties to the exact target PDF, avoiding potential accuracy problems introduced through the metamodel error. This approach relies on the delayed-rejection (DR) scheme to combine rapid exploring global (independent) proposals with robust random walk proposals. A Kriging metamodel-based density approximation is chosen as the global proposal to generate candidate samples in each MCMC step. For any rejected sample, DR allows an extra random walk, avoiding potential issues when Kriging offers a poor approximation (i.e., underestimates) to the actual target PDF and guaranteeing convergence. The overall computational efficiency is further improved through adaptive Kriging updating during the MCMC sampling phase, by systematically including candidate samples who can substantially enhance Kriging's accuracy into the training database. The computational efficiency and robustness of the established algorithm is demonstrated in an analytical benchmark problem and an engineering Bayesian inference problem.

1. INTRODUCTION

Probabilistic analysis of engineering systems often entails the simulation of samples from a target probability density function (PDF) related to the response of the system model, with these samples subsequently being used to estimate statistics of interest for the system performance. To formalize this problem, consider a system with uncertain model parameter vector $\boldsymbol{\theta} \in \Theta \subset \Re^{n_\theta}$ characterized by the probability density function $p(\boldsymbol{\theta})$ whose support is denoted by Θ. Let $\mathbf{z}(\boldsymbol{\theta}) \in \mathcal{Z} \subset \Re^{n_z}$ represent the response vector for the model parameter $\boldsymbol{\theta}$, assumed to be obtained through a time-consuming call to a deterministic numerical model (also called simulator). We are interested in simulating samples from the target density defined as follows:

$$\pi(\boldsymbol{\theta}) = \frac{h(\mathbf{z}\,|\,\boldsymbol{\theta})\,p(\boldsymbol{\theta})}{\int_{\Theta} h(\mathbf{z}\,|\,\boldsymbol{\theta})\,p(\boldsymbol{\theta})\,d\boldsymbol{\theta}} \propto h(\mathbf{z}\,|\,\boldsymbol{\theta})\,p(\boldsymbol{\theta}) \qquad (1)$$

where $\propto$ denotes proportionality and $h(\boldsymbol{\theta}\,|\,\mathbf{z}) : \Re^{n_\theta} \rightarrow \Re^{+}$ transforms the numerical system response $\mathbf{z}(\boldsymbol{\theta})$ to the quantity of interest (QoI) $h(.)$ for the specific uncertainty quantification (UQ) task. For instance, in Bayesian inference a common choice is to define the response $\mathbf{z}(\boldsymbol{\theta})$ as the log likelihood, i.e., $h(\boldsymbol{\theta}) = e^{z(\boldsymbol{\theta})}$. The notation $h(\boldsymbol{\theta})$ will be also used interchangeably with $h(\mathbf{z}\,|\,\boldsymbol{\theta})$ for clarity.

As directly simulating samples from $\pi(\boldsymbol{\theta})$ is impractical for applications with black-box numerical models, advanced Monte Carlo sampling approaches need to be utilized, such as

rejection sampling or Markov Chain Monte Carlo (MCMC) sampling (Robert and Casella 2013). In this work we focus on MCMC techniques, and particularly on one of its important classes: the Metropolis-Hastings (MH) algorithm (Metropolis et al. 1953).

The underlying foundation in MCMC is that trial samples from a different *proposal density* $q(\boldsymbol{\theta})$ are leveraged to draw samples from the target density $\pi(\boldsymbol{\theta})$. The choice of the proposal distribution critically controls MCMC efficiency, generally quantified by the degree of correlation between the Markov Chain samples, or, equivalently, by the number of effective independent samples when Markov Chain is used for statistical inference (Brooks et al. 2011). The optimal, yet impractical, choice for $q(\boldsymbol{\theta})$ is the target distribution itself, yielding independent identically distributed (i.i.d) samples. In common applications random walk proposals (RWs) are chosen (Robert and Casella 2013). Though RW are easy to implement and robust, their proper implementation requires tuning for the correct RW step size. Excessively large steps lead to frequent rejections and therefore high correlation. On the other hand, narrow steps facilitate moves only in the vicinity of the current sample and explore the support of the target PDF very slowly, yielding again high correlation. For many challenging target densities (e.g., nonlinear support or high skewness), RW can be quite inefficient.

The alternative to RW proposals are independent (also frequently referenced as global) proposals, which can be potentially highly efficient but do face their own challenges with respect to algorithmic robustness in guaranteeing convergence to the target PDF as well as proper selection to achieve high efficiency (Brooks et al. 2011). This paper investigates an MCMC sampling scheme that addresses these challenges by combining a surrogate model approximation of the target density as the global proposal density together with a RW component to improve sampling robustness in problematic regions, corresponding to regions with lower quality for the target density approximation. The formulation of the surrogate model based density approximation is discussed first (Section 2), followed (Section 3) by the new MCMC scheme termed Adaptive Kriging Delayed Rejection Adaptive Metropolis algorithm (AK-DRAM).

2. KRIGING-BASED DENSITY APPROXIMATION

For density approximation, we adopt the same setup in (Zhang and Taflanidis 2018a), with Kriging output corresponding to the response vector $\mathbf{z}$ and input chosen as the uncertain parameter vector $\boldsymbol{\theta}$. For forming the Kriging surrogate model, the output $\{z_i(\boldsymbol{\theta}^t),\ t=1,\ldots,n\}$ is observed at n distinct locations for the input $\{\boldsymbol{\theta}^t,\ t=1,\ldots,n\}$ called *training* (or *support*) *points* or *experiments*. Let $\boldsymbol{\Theta}=[\boldsymbol{\theta}^1\ldots\boldsymbol{\theta}^n]^T$ denote the input matrix, $\mathbf{Z}_i=[z_i^1\ \ldots\ z_i^n]^T$ the corresponding output vector. Based on the observations, Kriging approximates the response component z_i at $\boldsymbol{\theta}$ as a Gaussian distribution (Sacks 1989):

$$\tilde{z}_i(\boldsymbol{\theta}\,|\,\boldsymbol{\Theta},\mathbf{Z}_i)\sim\mathcal{N}\left(\mu_i(\boldsymbol{\theta}),\sigma_i^2(\boldsymbol{\theta})\right) \tag{2}$$

where $\mathcal{N}(a,b)$ stands for Gaussian distribution with mean a and variance b, $\sim$ is used herein to denote "distributed according to" and $\mu_i(\boldsymbol{\theta})$ and $\sigma_i^2(\boldsymbol{\theta})$ represent, respectively, the predictive mean and variance of the surrogate model. For simplicity we assume that it is feasible to develop a separate surrogate model for each response component z_i. Thus predictions for $\mathbf{z}$ are obtained by assembling all individual response component approximations, resulting in the predictive mean $\boldsymbol{\mu}(\boldsymbol{\theta})$, variance $\boldsymbol{\sigma}^2(\boldsymbol{\theta})$, vectors as well as the distribution $\tilde{\mathbf{z}}(\boldsymbol{\theta}\,|\,\boldsymbol{\Theta},\mathbf{Z})$.

Now, with respect to how to approximate the QoI (h), one can simply replace the response with its predictive mean, leading to $\tilde{h}(\boldsymbol{\theta})=h(\boldsymbol{\mu}\,|\,\boldsymbol{\theta})$. This however neglects information in the predictive variance (Zhang and Taflanidis 2018a). A common alternative is to consider the expected QoI, especially if this can be expressed in closed form with dependence on $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$. This ultimately leads to *predictive* QoI expressed as

$$\tilde{h}(\boldsymbol{\theta}) = \int_{\mathbb{R}^{n_z}} h(\tilde{z} \mid \boldsymbol{\theta}) \phi\left(\tilde{z}(\boldsymbol{\theta}); \mu(\boldsymbol{\theta}), \sigma^2(\boldsymbol{\theta})\right) d\tilde{z} \qquad (3)$$

representing the expectation under the distribution in Eq. (2), with $\phi(.;a,b)$ corresponding to a Gaussian PDF with mean a and variance b. It is assumed herein that Eq. (3) has a closed form solution for $\tilde{h}(\boldsymbol{\theta})$. Details on derivation of this solution are discussed in (Zhang and Taflanidis 2018a). Finally, utilizing the predictive QoI $\tilde{h}(\boldsymbol{\theta})$, the density approximation can be expressed as:

$$\tilde{\pi}(\boldsymbol{\theta}) = \frac{\tilde{h}(\boldsymbol{\theta}) p(\boldsymbol{\theta})}{\int_{\Theta} \tilde{h}(\boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta}} \propto \tilde{h}(\boldsymbol{\theta}) p(\boldsymbol{\theta}) \qquad (4)$$

With respect to quality of approximation $\tilde{\pi}(\boldsymbol{\theta})$ it should be stressed that the selection of the experiments (vector $\boldsymbol{\Theta}$), what is commonly referenced to as Design of Experiments (DoE), is critical (Sacks 1989). In this work, we adopt an iterative and adaptive DoE approach that is intended to maximize the computational efficiency, namely the AK-SSD (*adaptive Kriging stochastic sampling and density approximation*) (Zhang and Taflanidis 2018a), though the proposed sampling algorithm is compatible with any alternative scheme.

3. AK-DRAM METHODOLOGY

3.1. Independent Metropolis-Hastings

With an approximated density $\tilde{\pi}(\boldsymbol{\theta})$ that is expected to be close to the target $\pi(\boldsymbol{\theta})$, it is tempting to adopt $\tilde{\pi}(\boldsymbol{\theta})$ as global proposal for MH. This leads to the following independent MH: given current sample $\boldsymbol{\theta}^r$, a trial candidate $\boldsymbol{\theta}^*$ is drawn from distribution $\tilde{\pi}(\boldsymbol{\theta})$, and accepted (setting $\boldsymbol{\theta}^{r+1} = \boldsymbol{\theta}^*$) with probability:

$$\alpha(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^r) = \min\left[1, \frac{h(\boldsymbol{\theta}^*)\tilde{h}(\boldsymbol{\theta}^r)}{h(\boldsymbol{\theta}^r)\tilde{h}(\boldsymbol{\theta}^*)}\right] \qquad (5)$$

If rejected, the previous sample is repeated (setting $\boldsymbol{\theta}^{r+1} = \boldsymbol{\theta}^r$). As discussed in the introduction, independent MH can yield significant computational efficiency gain as long as the proposal resembles the target density well, with the best case being the proposal density equaling the target one.

Important challenges do exist though for selecting $\tilde{\pi}(\boldsymbol{\theta})$ as MH proposal density. First the generated Markov chain will lose irreducibility, leading to biased predictions, if the proposal density has narrower support that the target density. Furthermore, the performance of independent MH, quantified by the correlation among chain samples and the convergence rate to the stationary distribution, critically depends on the supremum of the ratio between the target and approximated densities:

$$M = \max_{\boldsymbol{\theta} \in \Theta} \left[\pi(\boldsymbol{\theta}) / \tilde{\pi}(\boldsymbol{\theta})\right] \qquad (6)$$

with larger values for M adversely impacting computational efficiency (Liu 1996). Even when $\tilde{\pi}(\boldsymbol{\theta})$ provides a good global fit to $\pi(\boldsymbol{\theta})$, the value of M can be still high as bounding the latter requires high pointwise approximation accuracy through the whole domain. Delayed rejection is utilized to address these challenges.

3.2. Enhancement through delayed rejection

The independent MH discussed in the previous section can be potentially highly efficient but lacks robustness in domains where the Kriging-based approximated density (serving as proposal density) does not match well the target density. To overcome this drawback, a RW proposal is integrated into the sampling scheme using a delayed rejection (DR) strategy (Tierney and Mira 1999). According to the DR scheme if the candidate sample $\boldsymbol{\theta}^*$ drawn from independent proposal $\tilde{\pi}(\boldsymbol{\theta}^*)$ is initially rejected based on acceptance ratio $\alpha(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^r)$ of Eq. (5), instead of repeating the current sample, a second-stage RW move is proposed from another distribution $\boldsymbol{\theta}^{**} \sim \mathcal{N}(\boldsymbol{\theta}^r, \boldsymbol{\Sigma})$. To maintain the detailed balance conditions, and guarantee MH convergence to target PDF, candidate sample is accepted (setting $\boldsymbol{\theta}^r = \boldsymbol{\theta}^{**}$) with probability:

$$\gamma(\boldsymbol{\theta}^{**} \mid \boldsymbol{\theta}^r, \boldsymbol{\theta}^*) = \min\left[1, \frac{h(\boldsymbol{\theta}^{**})[1-\alpha(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^{**})]}{h(\boldsymbol{\theta}^r)[1-\alpha(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^r)]}\right] \qquad (7)$$

If rejection still occurs, then the previous chain sample is finally repeated, setting $\boldsymbol{\theta}^{r+1} = \boldsymbol{\theta}^{r}$ .As DR introduces extra computational burden, as multiple acceptance ratios, and therefore expensive simulator evaluations, might be needed per candidate, it is important to ensure that the DR implementation is properly designed to yield greater computational efficiency.

For selecting the covariance matrix $\boldsymbol{\Sigma}$ for the RW proposal, an adaptive approach is adopted (Haario et al. 2001). This is accomplished by scaling the covariance matrix to mimic the optimal RW proposal in (Gelman et al. 1996) using the sample history of the Markov chain for guidance, leading:

$$\boldsymbol{\Sigma} = s_d \cdot \begin{cases} \boldsymbol{\Sigma}_0, & r \le N_0 \\ \hat{\boldsymbol{\Sigma}}_\pi, & r > N_0 \end{cases} \tag{8}$$

where s_d is the dimensional-adaptive scaling constant set as $s_d = 2.4^2 / n_\theta$, $\hat{\boldsymbol{\Sigma}}_\pi$ is the empirical covariance matrix estimated from the past samples in the Markov chain and $\boldsymbol{\Sigma}_0$ is the initial covariance matrix, used till the chain is sufficiently long (larger than N_o) to provide confidence for relying on the empirical covariance matrix. Here $\boldsymbol{\Sigma}_0$ is chosen as the covariance matrix of the approximated density $\tilde{\pi}(\boldsymbol{\theta})$. The non-adaption period is selected proportional to the number of Kriging experiments: $N_0 = 100n$.

The proposed AK-DRAM algorithm ultimately integrates independent MH and RW. It inherits the global exploration traits of the independent proposal, since that is used in the first stage. Simultaneously it also inherits the robustness features of the RW second-stage, circumventing challenges identified in the previous section for independent MH. Irreducibility is guaranteed through the second-stage RW component (Roberts and Rosenthal 2007) regardless of the support coverage of the first-stage proposal density, while convergence to stationary distribution $\pi(\boldsymbol{\theta})$ follows directly from adaptive DR (Haario et al. 2006). The computational efficiency is also not critically influenced by the supremum constant M due to the integration of RW second stage. This efficiency can be further improved by an adaptive refinement of the surrogate model to promote a better match between $\tilde{\pi}(\boldsymbol{\theta})$ and $\pi(\boldsymbol{\theta})$, using information from the sampling phase.

3.3. Adaptive Kriging approximation refinement

Instead of keeping the Kriging-based density approximation fixed after the preliminary surrogate model tuning phase [i.e., initial formulation of $\tilde{\pi}(\boldsymbol{\theta})$], information from the sampling phase can be used for promoting an adaptive refinement, since for any candidate sample $\boldsymbol{\theta}^c$ (including first stage $\boldsymbol{\theta}^*$ and second stage ones $\boldsymbol{\theta}^{**}$ with respect to the DR implementation), the response $\mathbf{z}(\boldsymbol{\theta}^c)$ evaluated through the simulator is available. Such a refinement can facilitate a better match between $\tilde{\pi}(\boldsymbol{\theta})$ and $\pi(\boldsymbol{\theta})$ which can yield substantial benefits for the MCMC sampling.

The straightforward approach for this refinement is to augment the current experiment set using the information from all new candidate samples. However, this unnecessarily increases the computational complexity for Kriging predictions (which scales cubically with the number of experiments n). In addition, the utility as potential experiments is not the same for all candidate samples, with some of them belonging in domains in Θ where the surrogate model accuracy is already high.

Instead, the selection of samples to be considered as experiments should be carefully made. Particularly, we are more interested into preventing under-approximation of target densities [$\tilde{\pi}(\boldsymbol{\theta}) << \pi(\boldsymbol{\theta})$] rather than over-approximation [$\tilde{\pi}(\boldsymbol{\theta}) >> \pi(\boldsymbol{\theta})$], as the former is primarily responsible for all independent proposal challenges, including under-coverage of support, large supremum ratio M and poor mixing. With this in mind, we examine the utility of each trial sample by quantifying the ratio of the target density to its approximation, $m(\boldsymbol{\theta}) = \pi(\boldsymbol{\theta}) / \tilde{\pi}(\boldsymbol{\theta})$. Larger values of $m(\boldsymbol{\theta})$ mean that the trial represents severely under-approximated region,

and therefore its inclusion into existing experiments should be highly favorable for improving the density approximation in a problematic domain. This inclusion ultimately decreases first-stage rejection rates. Exact calculation, though, of $m(\boldsymbol{\theta})$ is infeasible since the target density and/or its approximation might be only known up to a normalization constant. The remedy is to consider the ratio of non-normalized densities, referenced herein as value of information (VoI) for surrogate refinement:

$$\bar{m}(\boldsymbol{\theta}) = h(\boldsymbol{\theta}) / \tilde{h}(\boldsymbol{\theta}) \tag{9}$$

which only depends on available information for each candidate sample, and is proportional to the aforementioned relative utility of each sample $m(\boldsymbol{\theta})$. We adaptively choose the threshold for determining if a candidate sample should be considered as an experiment, by comparing its VoI to all existing experiments. The sample is added in the database only if it offers relatively higher utility (based on VoI) than available experiments. For performing this comparison, an adjustment is required for the VoI for existing experiments, since for them by default $\{\bar{m}(\boldsymbol{\theta}^t) = 1;\ t = 1,...,n\}$ due to Kriging's interpolation characteristics. Leave-one-out (LOO) predictions are used instead, calculating the predictive QoI $\tilde{h}(\boldsymbol{\theta}^t)$ for each experiment using the entire database excluding itself. Such use of LOO statistics is supported by the readily-available closed-form solutions for the LOO Kriging predictive mean and variance (Dubrule 1983), allowing estimation of the predictive QoI without the need to actually reconstruct the surrogate model leaving sequentially each experiment out. The prediction for the response z_i for the t-th experiment $\boldsymbol{\theta}^t$ using the entire database apart from this experiment remains Gaussian with predictive mean $\mu_i^{-t}(\boldsymbol{\theta}^t)$ and variance $\sigma_i^{-t}(\boldsymbol{\theta}^t)^2$. This ultimately leads to LOO predictive QoI $\tilde{h}^{-t}(\boldsymbol{\theta}^t) = \tilde{h}(\boldsymbol{\mu}^{-t}, \boldsymbol{\sigma}^{-t} \mid \boldsymbol{\theta}^t)$ and to LOO VoI:

$$\bar{m}_{LOO}(\boldsymbol{\theta}^t) = h(\boldsymbol{\theta}^t) / \tilde{h}^{-t}(\boldsymbol{\theta}^t); t = 1,...,n \tag{10}$$

The threshold for whether a candidate sample should be included as an experiments can be then selected as ζ-quantile (for example, ζ = 80%) of the LOO VOI among existing experiments.

To avoid numerical challenges with large datasets, a maximum number of experiments n_{max} should be set, avoiding any further surrogate model refinement when $n>n_{max}$. This choice also has a side benefit of ensuring the overall ergodicity for the first-stage proposal density as it limits for problematic applications the continuous updating of the global proposal density.

4. ILLUSTRATIVE EXAMPLES

4.1. Setup

To explore the benefit from different features of AK-DRAM, we compare to benchmark approaches: the adaptive Metropolis (AM) (Haario et al. 2001), and the delayed rejection adaptive Metropolis (DRAM) (Haario et al. 2006). In addition, a variant of the AK-DRAM is examined, denoted by K-DRAM, fixing the Kriging model and independent proposal in the sampling process. This version is used to examine the utility of the adaptive Kriging refinement discussed in Section 3.2.

For performance comparisons we first consider statistical efficiencies, which depends mainly on the correlations between the Markov Chain samples, since high correlation increases the estimator variance when these samples are used for statistical inference. This variance is proportional to the integrated autocorrelation time τ (Brooks et al. 2011). Here τ report the maximum (worst) integrated τ over all input dimensions. Another efficiency measure is the mean squared jump distance (d^2) between successive components of the chain (Gelman et al. 1996). Higher efficiency corresponds to lower τ (less correlation) and higher d^2 (bigger jumps). Since the applications of interest are computationally expensive, both τ and d^2 are adjusted per simulator evaluation instead of per sample, leading to the computational efficiency measures τ_c and d_c^2 (subscript c stands for

'computational'). The total number of simulations includes simulations required to obtain the initial Kriging experiments n_{init} (density approximation stage) and the simulation for the first-stage acceptance check N (this equals the chain length) and for the second-stage acceptance check N_{dr} when DR is adopted.

In all instances, the ζ-quantile for LOO VoI threshold is set to 80%. The max number of experiments n_{max} is set to 1000. Ten independent runs of each algorithm are performed. We always discard the first 1% samples of each chain as burn-in period to allow chain reach stationarity. For each run, the chains corresponding to the different examined approaches are all initialized at the same point. To ensure convergence of statistics, for each chain a large Markov chain length of $N = 20000$ after burn-in is used.

4.2. Analytical Example

The analytical example considers the two-dimensional banana shaped target density:

$$\pi(\boldsymbol{\theta}) = \phi\left(\begin{bmatrix} \theta_1 \\ \theta_2 + 0.03\theta_1^2 - 3 \end{bmatrix}; \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 100 & 0 \\ 0 & 1 \end{bmatrix}\right) \quad (11)$$

Given its mild non-linearity and low dimensionality, we obtain 18 space-filling experiments for formulating the Kriging surrogate model at the density approximation stage.

Results for the target density and its approximation are first presented in Figure 1. These results clearly demonstrate that the initial approximation [part (b)] encounters significant difficulties in the tail regions using space-filling experiments (red x) only. The first-stage adaption enables an automatic improvement of problematic regions, as the new experiments from MCMC trials (marked by black +) are concentrated on problematic tail regions offering finally a good approximation of the target density [part (c)]. Conducting refinements in the sampling phase without DR [part (d)] fails to provide significant improvements as potential experiments in the truly problematic tail regions are never identified, due to the fact that candidate trials are constrained only in the initial approximated density domain (no availability of DR step with RW).

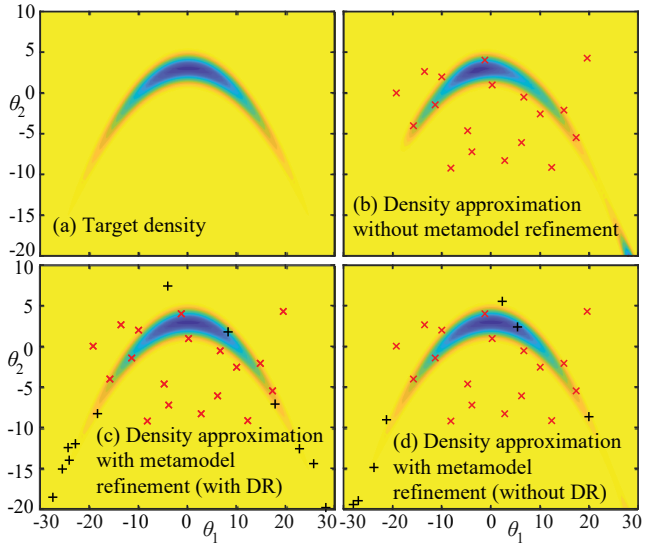

Figure 1. Contours of the (a) the true target density; (b) the approximated density before adaption (only with space-filling experiments in red x); (c) the approximated density after adaption (refinement experiments in black +); (d) the approximated density after adaption (refinement experiments in black +) but with AK-DRAM relying only on independent MH

Table 1 presents all relevant performance statistics. These results show that both variants using independent Kriging proposals noticeably outperform AM and DRAM under all efficiency measures. This is of course anticipated: a good independent proposal that can traverse the whole domain rapidly should bring distinct benefits. AK-DRAM also substantially improves over its counterpart without Kriging adaption (K-DRAM): integrated autocorrelation times are reduced by factors of 28 (statistically) and 36 (computationally), while jump distances d^2 and d_c^2 are increased by 53% and 64%. Such an evident margin confirms the crucial importance of the adaptive surrogate model refinement. Despite the complexity of the approximated density, as evident by both the contours in Figure 1 (heavy tail regions) and the challenges encountered by the other approaches, AK-DRAM achieves almost independent sampling (τ close to 1), demonstrating the computational efficiency

proposed sampling scheme can offer. The comparison between AK-DRAM and K-DRAM can be further extended to their average first-stage acceptance rates, as higher rates infer a better independent proposal. The average acceptance rates are 97.5% for AK-DRAM and 82.8% for K-DRAM, further verifying the benefit from adaptively refining the approximated density during sampling. AK-DRAM also displays superior robustness when comparing the worst-case performances (shown in parenthesis).

Table 1. For the banana shaped distribution average (worst case in parenthesis) performance for all approaches. Bold marks the best perfomance.

	τ	τ_c	d^2	d_c^2
AM	63.9 (74.7)	63.9 (74.7)	0.11 (0.10)	0.11 (0.10)
DRAM	43.9 (55.3)	83.8 (105)	0.21 (0.20)	0.11 (0.09)
K-DRAM	29.5 (54.9)	38.9 (58.9)	1.83 (0.86)	1.68 (0.52)
AK-DRAM	**1.05 (1.12)**	**1.07 (1.17)**	**2.80 (2.73)**	**2.76 (2.63)**

This example is further leveraged to explore the benefits and the necessity of combining independent proposals, delayed rejection, and the surrogate model adaption strategy. We estimate the value of M, representing the supremum ratio between target and approximation, for the run plotted in Figure 1. This value is found to be excessively large (2.7×10^{10}) when $\tilde{\pi}(\boldsymbol{\theta})$ is given by the surrogate model only based on the initial set of experiments [part (b) of Figure 1]. As the value of M can be directly compared with τ to measure statistical efficiencies, the large identified M directly reveals the risk of using only independent proposals. Delayed rejection is capable of overcoming this challenge, as for K-DRAM the integrated autocorrelation times always stays below 30. The surrogate model adaption also helps in further mitigating this risk (compare AK-DRAM to K-DRAM). In addition, delayed rejection introduces clear benefit for producing candidate samples with greater utility for the surrogate model refinement, from the results of part (d) of Figure 1. Without the second-stage DR, the independent proposal do not move into problematic regions at all. Thus even the added experiments cannot improve the density approximation quality in the tail regions, contrary to the implementation with delayed rejection [part (c) of Figure 1]. The overall discussion stresses the importance of combining all ingredients of AK-DRAM: independent proposal with DR utilizing random walks, and adaptive refinement of the surrogate model approximation.

4.3. Engineering example

This problem considers Bayesian updating for the model of eight-story structure. Numerical details are included in (Zhang and Taflanidis 2018b). Problem pertains to updating stiffness characteristics of the finite element structural model using modal information (modal frequencies and modeshapes) for the first three modes. The dimensionality is n_θ=10 with eight parameters corresponding to building properties and remaining two to the error statistics for modeshapes and modal frequencies. The density approximation is established with the AK-SSD algorithm using a total of n_{init}=300 initial experiments.

Results are reported in Table 2 and indicate once again that both AK-DRAM and K-DRAM yield significant efficiency gain. AK-DRAM is about one order of magnitudes more efficient than the best non-surrogate alternative in all measures. Kriging adaption seems to be very beneficial: AK-DRAM has significantly smaller integrated autocorrelation time (1/6 of K-DRAM) and nearly doubled jump distance. This problem creates greater challenges for the independent proposal to mimic the target one, as both surrogate-based approaches have relatively low first-stage acceptance rates (around 60% on average for AK-DRAM and 40% for K-DRAM). Yet the efficiency metrics reported in Table 2 for both approaches confirm their applicability even under imperfect independent proposals.

Table 2. For the Bayesian inference example average performancea. In parenthesis the worst case is also reported. Bold marks the best perfomance.

	τ	τ_c	d^2	d_c^2
AM	50.0	50.0	1.00	1.00
	(70.3)	(70.3)	(0.95)	(0.95)
DRAM	46.8	89.2	1.09	0.57
	(70.9)	(134.8)	(1.01)	(0.53)
K-DRAM	14.1	22.5	7.55	4.73
	(32.1)	(51.8)	(7.36)	(4.57)
AK-DRAM	**2.65**	**3.71**	**12.5**	**9.08**
	(2.86)	**(3.74)**	**(11.5)**	**(8.76)**

5. CONCLUSIONS

A new, computationally efficient MCMC scheme was discussed in this paper for generating samples from a target density whose evaluation involves a complex numerical model. This scheme leverages three popular MCMC features: independent Metropolis-Hastings (MHs) for rapid global exploration, adaptive random walk (RW) MH for robustness and delayed rejection (DR) for combining these two concepts. The efficiency gain is primarily facilitated through adopting a Kriging density approximation of the target density as global proposal. Challenges associated with adopting such a global proposal are addressed through DR using an adaptive local RW proposal that maintains convergence to the target density for the generated chain. The overall efficiency is further improved through a systematic refinement of the Kriging surrogate model and of the corresponding target density approximation. The proposed framework was applied to both an analytical problem and a realistic engineering problem. Both examples demonstrated the efficiency gain, robustness and convergence of the proposed sampling approach.

6. REFERENCES

Brooks, S., Gelman, A., Jones, G., and Meng, X.-L. (2011). *Handbook of markov chain monte carlo*, CRC press.

Dubrule, O. (1983). "Cross validation of kriging in a unique neighborhood." *Journal of the International Association for Mathematical Geology*, 15(6), 687-699.

Gelman, A., Roberts, G. O., and Gilks, W. R. (1996). "Efficient Metropolis jumping rules." *Bayesian Statistics*, 5(599-608), 42.

Haario, H., Laine, M., Mira, A., and Saksman, E. (2006). "DRAM: efficient adaptive MCMC." *Statistics and Computing*, 16(4), 339-354.

Haario, H., Saksman, E., and Tamminen, J. (2001). "An adaptive Metropolis algorithm." *Bernoulli*, 7(2), 223-242.

Liu, J. S. (1996). "Metropolized independent sampling with comparisons to rejection sampling and importance sampling." *Statistics and Computing*, 6(2), 113-119.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). "Equation of state calculations by fast computing machines." *The Journal of Chemical Physics*, 21(6), 1087-1092.

Robert, C., and Casella, G. (2013). *Monte Carlo statistical methods*, Springer Science & Business Media.

Roberts, G. O., and Rosenthal, J. S. (2007). "Coupling and ergodicity of adaptive Markov chain Monte Carlo algorithms." *Journal of Applied Probability*, 44(2), 458-475.

Sacks, J., Welch, W.J., Mitchell, T.J., Wynn, H.P. (1989). "Design and analysis of computer experiments." *Statistical Science*, 4(4), 409-435.

Tierney, L., and Mira, A. (1999). "Some adaptive Monte Carlo methods for Bayesian inference." *Statistics in Medicine*, 18(17-18), 2507-2515.

Zhang, J., and Taflanidis, A. A. (2018a). "Adaptive Kriging stochastic sampling and density approximation and its implementation to rare-event estimation." *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, 4(3), 04018021.

Zhang, J., and Taflanidis, A. A. (2018b). "Bayesian posterior sampling using a metamodel-based sequential approach." *Proc., The 19th Working Conference of the IFIP Working Group on Reliability and Optimization of Structural Systems*, June 26-29, 2018, Zurich, Switzerland.