

Exact active subspace Metropolis–Hastings, with applications to the Lorenz-96 system

I. Schuster* P. G. Constantine† T. J. Sullivan‡

December 8, 2017

Abstract: We consider the application of active subspaces to inform a Metropolis–Hastings algorithm, thereby aggressively reducing the computational dimension of the sampling problem. We show that the original formulation, as proposed by Constantine, Kent, and Bui-Thanh (*SIAM J. Sci. Comput.*, 38(5):A2779–A2805, 2016), possesses asymptotic bias. Using pseudo-marginal arguments, we develop an asymptotically unbiased variant. Our algorithm is applied to a synthetic multimodal target distribution as well as a Bayesian formulation of a parameter inference problem for a Lorenz-96 system.

Keywords: Markov chain Monte Carlo, pseudo-marginal, active subspace, dimension reduction, Bayesian parameter inference.

2010 Mathematics Subject Classification: 65C40, 65C05, 62H25.

1 Introduction

Effective sampling of probability density functions on high-dimensional spaces is widely acknowledged as a challenging regime for Markov chain Monte Carlo (MCMC) techniques. The poor performance of many MCMC schemes in this setting can be attributed to the high dimension leading to both high variance and poor mixing properties, and it is therefore naturally appealing to attempt to improve the convergence properties of MCMC sampling by aggressively reducing the computational dimension to the greatest extent possible while still preserving statistically desirable properties.

Active subspaces (Constantine, 2015) are a dimensionality reduction scheme originating in applied mathematics and have recently been applied to Metropolis–Hastings algorithms (ASMH, Constantine et al., 2016). In its original formulation, ASMH was proved to sample from a density ρ_ϵ that is close to a desired target density ρ in the Hellinger distance. In this work, we use results from the theory of pseudo-marginal Metropolis–Hastings (Andrieu and Roberts, 2009) to

- establish that indeed ASMH in its original formulation results in a biased sample;
- derive a variant that samples *exactly* (in stationarity) from the target;
- make exact ASMH applicable when prior and likelihood violate the Gaussianity assumption.

When used in a pseudo-marginal setting, active subspaces play the role of a preprocessing technique for adapting the pseudo-marginal sampler to the structure of the posterior. After considering a synthetic bimodal Gaussian mixture model as a test case, we apply our technique to a Bayesian formulation of a parameter inference problem in the Lorenz-96 model.

*Institute of Mathematics, Free University of Berlin, and Zuse Institute Berlin, Takustrasse 7, 14195 Berlin, Germany, schuster@zib.de

†Computer Science, University of Colorado Boulder, Boulder, CO 80309, United States of America, paul.constantine@colorado.edu

‡Institute of Mathematics, Free University of Berlin, and Zuse Institute Berlin, Takustrasse 7, 14195 Berlin, Germany, sullivan@zib.de

Our motivation for the sampling of probability density functions on high-dimensional spaces stems from a desire to solve Bayesian inverse problems. That is, we have a density $\rho(x)$ that is given only proportionally by applying Bayes' theorem to a specified probability model and observed data that is assumed to be generated by the model. We assume the difficulty in this problem is aggravated by the fact that $x \in \mathbb{R}^m$ where m is large.

Indeed, MCMC sampling and its many variants are the computational workhorses of Bayesian inversion, but many problems of practical interest involve high- or even infinite-dimensional x , which is exactly the regime in which many MCMC schemes experience a dramatic deterioration in mixing or convergence properties or increase in autocorrelation and computational cost. Such problems can be attributed to multimodal and anisotropic structure, and the ease with which a high-dimensional random variable can have high variance: good mixing of the MCMC chain between well-separated modes is hard to ensure without a priori knowledge; some adaptation of isotropic proposals to anisotropic structure is essential for acceptable acceptance rates (Haario et al., 2001; Girolami and Calderhead, 2011); and variance controls the quality of MCMC estimates of integrals against ρ .

We propose to use active subspaces to deal with the problem of high dimension (Constantine et al., 2016). Concretely, we seek a reparametrization of the domain \mathbb{R}^m of ρ into $y \in \mathbb{R}^n, z \in \mathbb{R}^{m-n}$, such that $B_a y + B_i z = x \in \mathbb{R}^m$, $n \ll m$ and the marginal density $\rho(y)$ captures most of the variability of $\rho(x)$. Here $[B_a, B_i]$ is an orthonormal basis of \mathbb{R}^m and usually obtained by eigendecomposing a matrix informed by $\rho(x)$; further details are given in Section 2. The y are called the *active variables* and z the *inactive variables*. When considering integrals of the form $\int_{\mathbb{R}^m} f(x)\rho(x) dx$ that can not be solved analytically, we resort to Monte Carlo techniques. In this paper we propose to integrate over y with a Metropolis–Hastings-based MCMC algorithm, integrating out z with a nested numerical integration algorithm. This is the general method already proposed by Constantine et al. (2016).

However, departing from previous work on active subspace Metropolis–Hastings (ASMH), we develop a variant that both takes advantage of active subspaces and at the same time is asymptotically exact. Not introducing bias through the active subspace approximation while still taking advantage of the induced dimension reduction is of course preferable, especially as no additional computational cost is involved. To achieve this, we will make use of pseudo-marginal ideas developed by the Bayesian computation community (Beaumont, 2003; Andrieu and Roberts, 2009).

2 Active subspaces and their use for MCMC

For the purpose of this paper, we will consider a notion of active subspaces tailored to the problem of sampling. Assume that we are interested in the integral of some function $f: \mathbb{R}^m \rightarrow \mathbb{R}$ with respect to the target density $\rho: \mathbb{R}^m \rightarrow \mathbb{R}_+$, i.e. $\int_{\mathbb{R}^m} f(x)\rho(x) dx$. We further assume that this integral exists and is finite. We will construct matrices $B_a \in \mathbb{R}^{n \times m}, B_i \in \mathbb{R}^{(m-n) \times m}$ such that $[B_a^\top B_i^\top]^\top$ is an orthonormal basis of \mathbb{R}^m and such that either ρ or $|f|\rho$ vary mostly in the subspace spanned by B_a , while varying much less in the space spanned by B_i . Hence we call B_a the basis of the *active subspace*, and B_i the basis of the *inactive subspace*.

2.1 Constructing an active subspace

There are several methods of constructing an active subspace. We will go into detail about three of them, namely gradient covariance, posterior covariance and the linear regression method.

2.1.1 Covariance matrix of log-likelihood gradient

The method suggested by Constantine et al. (2016) for use with Metropolis–Hastings is to estimate the expected outer product of the log-likelihood gradient $\nabla \log(\rho_1(x))$ with respect to the prior ρ_p , resulting in the estimate

$$\hat{\Sigma} := \frac{1}{N} \sum_{i=1}^N \nabla \log(\rho_1(X_i)) \otimes \nabla \log(\rho_1(X_i)) \approx \int \nabla \log(\rho_1(x)) \otimes \nabla \log(\rho_1(x)) \rho_p(x) dx,$$

which can be acquired by ordinary Monte Carlo through sampling from the prior, $X_i \sim \rho_p$ for all i . Now one can eigendecompose $\widehat{\Sigma}$ and identify a spectral gap, i.e. a group of dominant eigenvalues well separated from the remainder. The eigenvectors with large eigenvalues are collected into the matrix B_a and are the basis of the active subspace. The rest of the eigenvalues are collected in matrix B_i , building the basis of the inactive subspace.

2.1.2 Posterior covariance

A method to construct an active subspace that can be used even when the gradient of the log-likelihood $\nabla \log(\rho_1(x))$ is unavailable or costly to compute involves estimating the covariance matrix of the posterior distribution $\int (x - \mu) \otimes (x - \mu) d\rho(x)$, where $\mu := \int x d\rho(x)$. Again, this can be estimated by using samples from the prior, $X_i \sim \rho_p$ for all i , using the importance sampling trick:

$$\widehat{\mu} := \frac{1}{\sum_{j=1}^N w_j} \sum_{i=1}^N w_i X_i,$$

where $w_i := \frac{\rho(X_i)}{\rho_p(X_i)}$, and

$$\widehat{\Sigma} := \frac{1}{\sum_{j=1}^N w_j} \sum_{i=1}^N w_i (X_i - \widehat{\mu}) \otimes (X_i - \widehat{\mu}).$$

Again we can eigendecompose $\widehat{\Sigma}$, identify a spectral gap and build the active subspace basis B_a using the eigenvectors with large eigenvalues, and the inactive subspace basis B_i using the eigenvectors with small eigenvalue.

The construction of Section 2.1.1 together with the Gaussian prior assumption was convenient for the original ASMH algorithm (Constantine et al., 2016). However, without increasing computational load, our pseudo-marginal approach allows for more freedom in constructing an active subspace for the purpose of using it in Metropolis–Hastings algorithms. The advantage of using the posterior covariance is that it is more closely related to the actual quantity of interest, namely the posterior distribution.

2.1.3 Linear regression for a one-dimensional active subspace

As an elementary tool for discovering *the* most important direction of variation, Constantine (2015) suggests using a method inspired by the notion of a *sufficient summary plot* in visual statistics (Cook, 1998). The idea is to find a linear regression model mapping the m -dimensional x approximately to $\rho(x)$ (or alternatively the optimal sampling distribution $|f(x)|\rho(x)$). Having acquired some points $\{X_i\}_{i=1}^N$, either by random sampling or with a deterministic mesh, the idea is to fit the vector B_a and scalar b in the linear regression model

$$\rho(x) = B_a x + b$$

using $\{(X_i, \rho(X_i))\}_{i=1}^N$ as input and accompanying output pairs. The fitted B_a now approximates the direction that explains most of the variation of the target density. To build the basis for the inactive subspace, we simply complete B_a to a basis of \mathbb{R}^m using the Gram–Schmidt method, resulting in an orthonormal basis. All of the basis vectors orthogonal to B_a build a basis for the inactive subspace, B_i .

2.2 Usage in Metropolis–Hastings

Starting from the general notion of active subspaces, the basic proposal of Constantine et al. (2016) is to run Metropolis–Hastings on the active subspace (y space) instead of the full space (x space), integrating out the inactive variables z using a nested numerical integration technique. For example, Constantine et al. (2016) used a nested Gauss–Hermite quadrature for the nested integral. When considering the move of the Markov chain from a current state y in the active subspace to a proposed state y' , Constantine et al. (2016) compute a new numerical estimate of both marginal densities $\rho(y)$ and $\rho(y')$ for use inside the MH accept/reject step. After changing the nested integration technique to importance sampling, the complete original algorithm is given in Algorithm 1 when setting the *originalVersion* variable to *True*. One of the advantages of using importance sampling is that the evaluation of the target density can be

fully parallelized. For details on the original active subspace Metropolis–Hastings (ASMH) algorithm, see Constantine et al. (2016).

3 Pseudo-marginal MCMC

Modern pseudo-marginal methods were developed in the context of population genetics by Beaumont (2003) and their convergence properties have been studied in depth by Andrieu and Roberts (2009). While pseudo-marginal methods were not developed with reparametrizations of the integration problem in mind, they split the full state space into variables of interest y and nuisance variables z . This split was originally motivated by inference problems in Bayesian models and the observation that sometimes ‘nuisance variables’ have to be introduced in the probability model solely for technical reasons.

In our setting we can only evaluate the joint density $\rho(y, z)$ (up to proportionality), but in pseudo-marginal methods we are interested in computing integrals with respect to the marginal density $\rho(y)$ instead. The idea, very similar to the original ASMH algorithm, is to get a statistical estimate $\hat{\rho}(y)$ of $\rho(y)$ and use the estimate in place of the actual density. If $\mathbb{E}[\hat{\rho}(y)] = \rho(y)$, i.e. the estimate is unbiased, then it is possible to construct a Metropolis–Hastings algorithm that will asymptotically produce samples *exactly* following $\hat{\rho}(y)$. Because $\hat{\rho}(y)$ only approximates the marginal density, but the algorithm is asymptotically exact, these types of algorithms have also been called *exact approximate* sampling algorithms. Concretely, the estimate will be used inside the Metropolis–Hastings acceptance probability:

$$\alpha = \min \left(1, \frac{\hat{\rho}(y')q_y(y'|y)}{\hat{\rho}(y)q_y(y|y')} \right),$$

where y' is the proposed new point in the active subspace and $q_y(y'|y)$ is the density of the proposal.

Whenever $\hat{\rho}(y')$ is computed only once per proposed point and recycled for later acceptance probability computations (in case the proposal is accepted), then under the condition that the estimate is unbiased and some further mild conditions, the resulting sampling algorithm was proved in Andrieu and Roberts (2009) to asymptotically sample from the desired marginal density over y . The algorithm class is called grouped independence Metropolis–Hastings (GIMH; Beaumont, 2003; Andrieu and Roberts, 2009).

If, however, we get a new estimate of the marginal density $\rho(y)$ of the current state for each new computation of the Metropolis–Hastings acceptance probability, then this will result in samples that are provably biased. This class of algorithms is called Monte-Carlo-within-Metropolis (MCwM; Beaumont, 2003; Andrieu and Roberts, 2009).

4 Exact ASMH

The original ASMH paper (Constantine et al., 2016) proved that the invariant distribution of its sampling algorithm, $\rho_\epsilon(y)$, is close in Hellinger distance to the marginal density $\rho(y)$. In fact, when considering an unbiased randomized method for the nested integration problem (over the inactive variables z), the original ASMH represents a particular instance of a MCwM algorithm and is thus biased¹. The pseudo-marginal analysis of the original ASMH using a deterministic nested integration technique is slightly less clear, as the question of whether the estimate is unbiased becomes meaningless in that case.

However, it is possible to construct an active-subspace-informed algorithm that will asymptotically sample from $\rho(y)$ exactly, using the pseudo-marginal approach of Beaumont (2003) and Andrieu and Roberts (2009). It would be possible to use, for example, Metropolis-within-Gibbs steps (Andrieu and Roberts, 2009) to marginalize over the inactive variables z , which however we leave to future work. Instead, we will investigate in detail using the importance sampling construction suggested originally by Beaumont (2003). An unbiased estimator $\hat{\rho}(y)$ of $\rho(y)$ is constructed using the importance sampling trick $\rho(y) = \int \frac{\rho(B_a y + B_i z)}{q_z(z)} q_z(z) dz$. This enables estimation $\rho(y)$ from a distribution with density q_z which can be chosen to be easy to sample from, under relatively mild assumptions on q_z . One obvious possibility would be to construct q_z utilizing B_i . Then the exact ASMH (eASMH) algorithm is given in Algorithm 1 when setting the *originalVersion* variable to *False*.

¹In fact, the algorithm as stated in Constantine et al. (2016) might also be read as an asymptotically unbiased GIMH algorithm (Beaumont, 2003). However, the implementation accompanying the paper and personal discussion with the

Algorithm 1 Active subspace Metropolis–Hastings

Input: $\rho(x)$, initial value x_0 , number of samples from active variables N , number of pseudo-marginal samples from inactive variables M , boolean flag *originalVersion*

Output: linear reparametrization B_a, B_i s.t. $x = B_a y + B_i z$, samples from $\rho(x)$

```
 $B_a, B_i = \text{ConstructActiveSubspace}(\rho)$ 
solve  $x_0 = B_a y_0 + B_i z_0$  for  $y_0, z_0$ 
generate  $z_{0,j} \sim q_z$  for  $j = 1, \dots, M$ 
 $w_{0,j} = \rho(B_a y_0 + B_i z_{0,j}) / q_z(z_{0,j})$  for  $j = 1, \dots, M$ 
 $d = \frac{1}{M} \sum_{j=1}^M w_{0,j}$ 
for  $i$  from  $1, \dots, N$  do
  generate  $y'_i \sim q_y(\cdot | y_{i-1})$ 
  generate  $z'_{i,j} \sim q_z$  for  $j = 1, \dots, M$ 
   $w'_{i,j} = \rho(B_a y'_i + B_i z'_{i,j}) / q_z(z'_{i,j})$  for  $j = 1, \dots, M$ 
   $d' = \frac{1}{M} \sum_{j=1}^M w'_{i,j}$ 
  if originalVersion then
    generate  $z_{i,j} \sim q_z$  for  $j = 1, \dots, M$ 
     $w_{i,j} = \rho(B_a y_i + B_i z_{i,j}) / q_z(z_{i,j})$  for  $j = 1, \dots, M$ 
     $d = \frac{1}{M} \sum_{j=1}^M w_{i,j}$ 
  end if
  generate  $b \sim \text{Bernoulli}(\min(1, d' q_y(y_{i-1} | y'_i) / d q_y(y'_i | y_{i-1})))$ 
  if  $b = 1$  then
    Set  $d = d', y_i = y'_i, z_{i,j} = z'_{i,j}, w_{i,j} = w'_{i,j}$ 
  else
    Set  $y_i = y_{i-1}, z_{i,j} = z_{i-1,j}, w_{i,j} = w_{i-1,j}$ 
  end if
end for
return  $x$ -samples  $B_a y_i + B_i z_{i,j}$  for  $i = 1, \dots, N$  and  $j = 1, \dots, M$ 
```

The main innovation of the new eASM algorithm, when compared with the original ASM algorithm (*originalVersion* is *True*), is that no new estimate of the marginal density $\rho(y)$ is acquired at each iteration. Rather, the previous estimate is recycled, while of course estimating the marginal density of the newly proposed point $\rho(y')$. Not only does this recycling result in usage of fewer computational resources, but it also leads to the algorithm being unbiased as proved in Andrieu and Roberts (2009).

Another difference, as discussed before, is that in eASM the estimate $\hat{\rho}(y)$ is computed using an unbiased randomized method, namely importance sampling. While the method for computing the estimate is not strictly fixed in the original ASM, for the experiments a deterministic integral estimator (Gauss–Hermite) was used. The decision to use importance sampling in eASM was taken for theoretical reasons, as in this case well-known pseudo-marginal assumptions and proofs continue to hold.

5 Experiments

5.1 Mixture of Gaussians

In a first experiment, we used an artificial target density, namely a mixture of two Gaussians in two dimensions, with means and covariances given by

$$\mu_1 = -\mu_2 := \begin{bmatrix} 2 \\ 2 \end{bmatrix}, \quad \Sigma_1 = \Sigma_2 := \begin{bmatrix} 1 & -0.9 \\ -0.9 & 1 \end{bmatrix},$$

and the density itself by

$$0.5 \mathcal{N}(\mu_1, \Sigma_1) + 0.5 \mathcal{N}(\mu_2, \Sigma_2).$$

authors clarified that this is not the case.

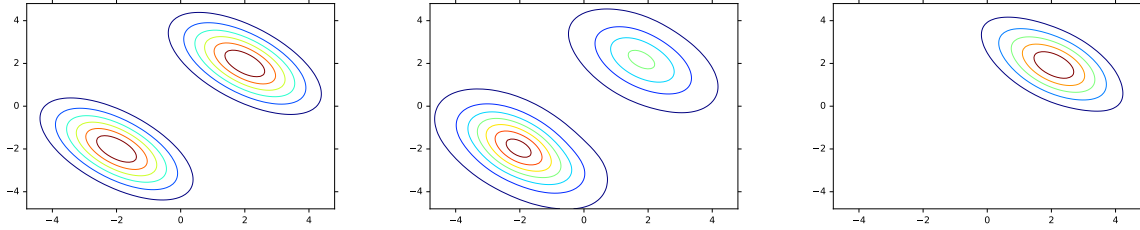


Figure 5.1: Convolution of target density with Gaussian density used for the random walk proposal (left), estimate of the same convolution with exact active subspace Metropolis–Hastings (middle) and with vanilla Metropolis–Hastings (right).

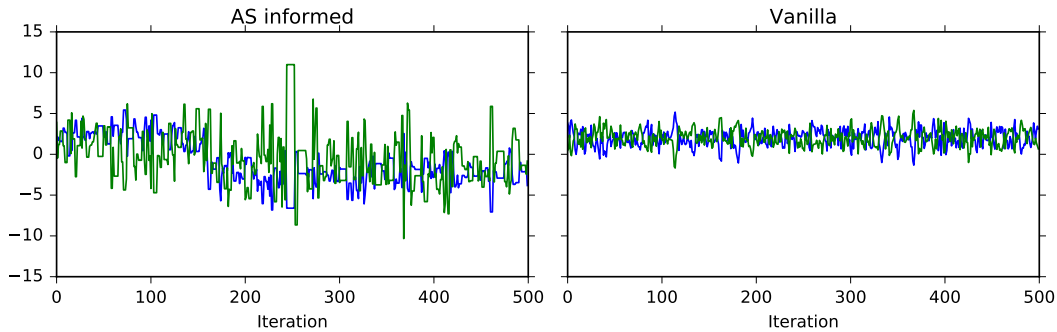


Figure 5.2: The MH trace plot shows that eASMH explores both modes of the target 2-dimensional Gaussian mixture distribution while vanilla MH becomes stuck in one mode.

The resulting bimodal structure is visualized in Figure 5.1 (left). For the vanilla MH algorithm we used a Gaussian random walk proposal with identity covariance matrix, started at the origin, and ran it for 5500 iterations, discarding the first 500 as burn-in.

For the pseudo-marginal active subspace algorithm, we acquired an initial 500 samples from an isotropic Gaussian centered at the origin with variance 10. From these, we built an active subspace approximation using the linear regression method described in Section 2.1.3. The algorithm was then run on the active variable for 500 iterations with the inactive variable integrated out using 10 nested samples per proposed point. This resulted in 5500 overall target evaluations, just as for the vanilla MH implementation. The acceptance rate of the vanilla MH algorithm was 32%, while the pseudo-marginal active subspace algorithm eASMH accepted 49% of the time, which is close to the optimal acceptance rate (Roberts and Rosenthal, 2001). Thus, while the amounts of overall computation for vanilla MH and eASMH are comparable, the eASMH algorithm demonstrates improved dynamics as quantified by the acceptance rate.

Furthermore, the resulting kernel density estimates are plotted in Figure 5.1 and can be compared with the ground truth target density (which has also convolved with the (Gaussian) density kernel). Clearly, the eASMH algorithm (middle) is able to recover both modes of the target distribution, which is not the case for the vanilla MH algorithm. This is also clear from the trace plot in Figure 5.2, which shows how the active subspace algorithm explores all relevant regions of the state space, while the vanilla algorithm gets stuck in one mode. Quantitatively, this results in lower autocorrelation values for the pseudo-marginal active subspace algorithm (Figure 5.3).

We ran the same algorithm for a 10-dimensional Gaussian mixture of two components with means coinciding except for the first dimension, which is 2 or -2 . In this case we also the trace plot clearly shows better mixing for the eASMH algorithm as compared to the vanilla MH algorithm (Figure 5.4). This is reflected in the autocorrelation, which drops more quickly for eASMH than for vanilla MH (Figure 5.5).

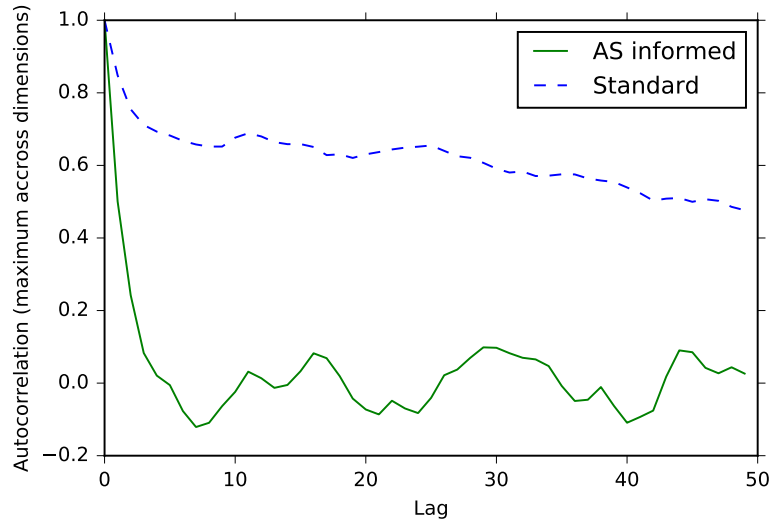


Figure 5.3: Autocorrelation using eASMH and vanilla MH up to lag 50 for the 2-dimensional Gaussian mixture, computed as the maximum autocorrelation across dimensions, and in each dimension the expected value over the complete chain.

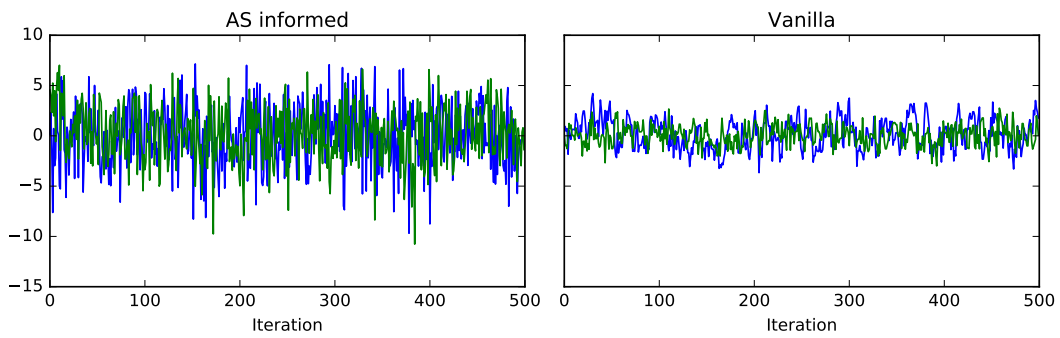


Figure 5.4: The MH trace plot shows that eASMH explores both modes of the target 10-dimensional Gaussian mixture distribution while vanilla MH becomes stuck in one mode.

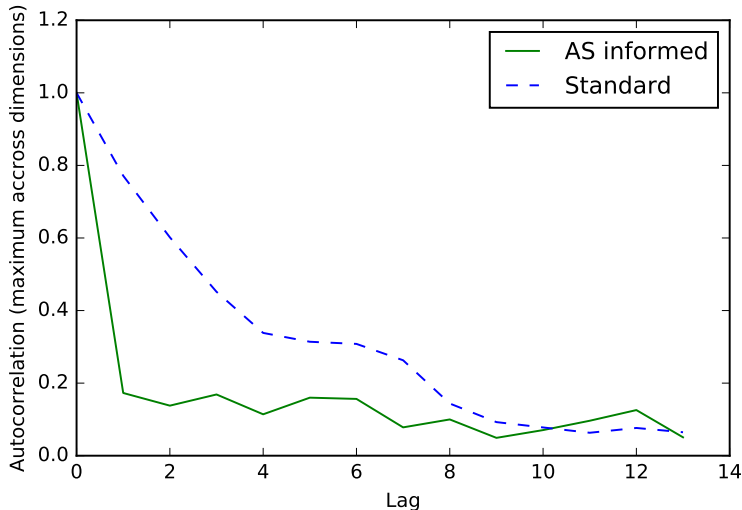


Figure 5.5: Autocorrelation using eASMH and vanilla MH up to lag 13 for the 10-dimensional Gaussian mixture, computed as the maximum autocorrelation across dimensions, and in each dimension the expected value over the complete chain.

5.2 Lorenz-96 parameter inference

The Lorenz-96 model is a dynamical system given by the ODEs

$$\begin{aligned} \frac{dX_k}{dt} &= -X_{k-1}(X_{k-2} - X_{k+1}) - X_k + F - \frac{hc}{b} \sum_{j=J(k-1)+1}^{kJ} Y_j \\ \frac{dY_j}{dt} &= -cY_{j+1}(Y_{j+2} - Y_{j-1}) - cY_j + \frac{hc}{b} X_{[(j-1)/J]+1} \end{aligned}$$

where the X_k variables are large-amplitude, low-frequency variables that are each coupled to many small-amplitude, high-frequency Y_j variables, and F is a forcing constant. We picked a system in 36 dimensions. We fixed a ground truth initial value and computed a forward solve with a forcing constant of $F = 8$, which is known to induce chaotic behavior. Data d was generated by computing a forward solve from the ground truth using a BDF method with a timestep of 0.01 between time 0 and 10 and adding independent centred Gaussian noise with variance 0.1 to the all points generated by the forward solve. We defined a prior density over all inference variables (initial values and the forcing constant) as an isotropic Gaussian centered at 0 with variance 4 in every dimension, resulting in a posterior on a 37-dimensional space. Given some point x , its likelihood is given by the density of a Gaussian centered at the forward solve $s(x)$ with isotropic variance 0.1, evaluated at d .

We collected 500 samples from the prior for constructing the active subspace, using the posterior covariance method outlined in Section 2.1.2. The spectral gap was found between the 4th and 5th largest eigenvalues, making the active subspace 4-dimensional. The eASMH algorithm was run for 500 samples in the active subspace, collecting 10 samples for each to estimate the marginal likelihood. This makes for 5500 posterior density evaluations overall, using the same number of forward solves of the ODEs. The vanilla MH algorithm was run for 5500 samples in the full state space, resulting in the same number of posterior density evaluations. In both cases, a simple Gaussian random walk proposal was used. The autocorrelation of every 10th full space sample is given as a function of lag time (i.e. number of MH steps) in Figure 5.7. Clearly, eASMH improves upon the short-lag autocorrelation of vanilla MH by a factor of around 0.84. A kernel density estimate of the marginal of the forcing constant and the last dimension computed from the eASMH samples is shown in Figure 5.6.

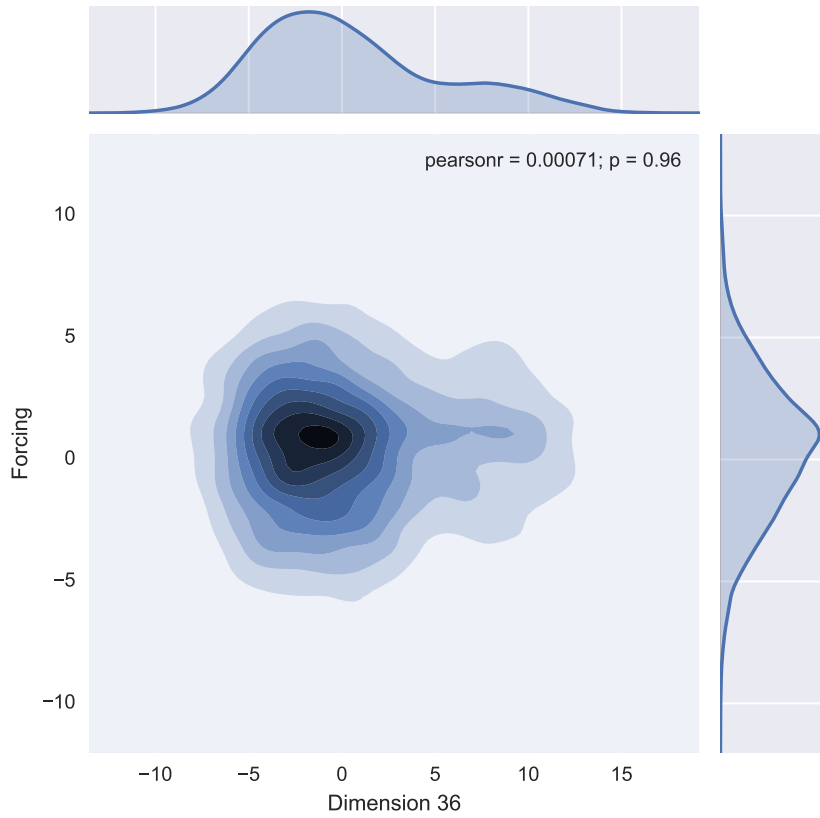


Figure 5.6: KDE estimate of marginal posterior density of forcing constant and last x component as computed using eASMH samples.

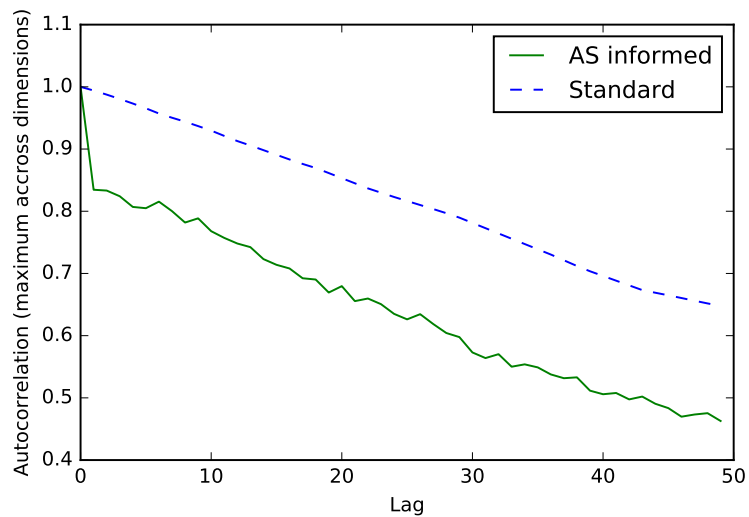


Figure 5.7: Autocorrelation using eASMH and vanilla MH up to lag 50 for the Lorenz-96 model, computed as the maximum autocorrelation across dimensions, and in each dimension the expected value over the complete chain.

6 Conclusions

In this work, we have used results from the theory of pseudo-marginal Metropolis–Hastings (Andrieu and Roberts, 2009) to conclude that active subspace Metropolis–Hastings in its original formulation results in a biased sample. We have derived an asymptotically unbiased variant that samples exactly from the target to eliminate this problem. Exact ASMH (eASMH) is applicable also when prior and likelihood are not Gaussian, which is an assumption made for the original ASMH algorithm. For eASMH, active subspaces are a preprocessing technique for adapting the pseudo-marginal sampler to the structure of the posterior. When using importance sampling as the nested Monte Carlo technique, target density evaluations can be computed in parallel, with noticeable speed gains compared to vanilla MH.

Our algorithm strongly decreased autocorrelation compared to vanilla MH when sampling from a mixture of two Gaussians. For a parameter inference problem in the Lorenz-96 Model, the results also clearly favour exact ASMH over vanilla MH. One possible area of further improvement is the construction of nonlinear reparametrizations of the full state space, so that nonlinear high posterior density regions can be accurately captured.

Acknowledgements

IS and TJS are supported by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) through grant CRC 1114 “Scaling Cascades in Complex Systems”, Project A06, and TJS is further supported by the Free University of Berlin within the Excellence Initiative of the DFG.

References

- C. Andrieu and G. O. Roberts. The pseudo-marginal approach for efficient Monte Carlo computations. *Ann. Statist.*, 37(2):697–725, 2009. doi:10.1214/07-AOS574.
- M. A. Beaumont. Estimation of population growth or decline in genetically monitored populations. *Genetics*, 164(3):1139–1160, 2003.
- P. G. Constantine. *Active Subspaces: Emerging Ideas in Dimension Reduction for Parameter Studies*, volume 2 of *SIAM Spotlights*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2015.
- P. G. Constantine, C. Kent, and T. Bui-Thanh. Accelerating Markov chain Monte Carlo with active subspaces. *SIAM J. Sci. Comput.*, 38(5):A2779–A2805, 2016. doi:10.1137/15M1042127.
- R. D. Cook. *Regression Graphics*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, 1998. doi:10.1002/9780470316931.
- M. Girolami and B. Calderhead. Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 73(2):123–214, 2011. doi:10.1111/j.1467-9868.2010.00765.x.
- H. Haario, E. Saksman, and J. Tamminen. An adaptive Metropolis algorithm. *Bernoulli*, 7(2):223–242, 2001. doi:10.2307/3318737.
- G. O. Roberts and J. S. Rosenthal. Optimal scaling for various Metropolis–Hastings algorithms. *Statist. Sci.*, 16(4):351–367, 2001. doi:10.1214/ss/1015346320.