

Affine-invariant ensemble sampling with applications to LIGO data

Candidate Number: 8233R
Supervisor: Dr. Will Handley

May 17, 2017

Abstract

The Adaptive Directional Sampling (ADS) algorithm is presented and compared against both traditional Metropolis-Hastings methods and alternative affine-invariant ensemble samplers such as emcee. Computational tests showed ADS requires fewer likelihood calls to converge to target likelihoods compared to the alternative ensemble algorithms in unbound Euclidean parameter spaces. Emcee remains the most effective sampler once convergence has been achieved. Adaptations of ensemble methods to likelihoods on non-Euclidean parameter spaces were introduced. Non-Euclidean samplers outperformed their traditional counterparts when sampling from simulated LIGO likelihoods and other test distributions. Adaptations of the Adaptive Directional Sampler using principles of parallel transport along geodesics were shown to most effectively sample toroidal and spherical parameter spaces.

1 Introduction

In February 2016, LIGO announced the first detection of gravitational waves (GW) originating from a binary black hole (BH) merger [1]. This experimental observation, which validates Einstein's prediction of GWs, opens a new era of astronomy where gravitational interferometry can be used to detect and measure astronomical bodies.

A Bayesian analysis of the experiment's posterior probability distribution function (PDF) was conducted to obtain estimates for the black holes' physical parameters such as their mass and spin. However, in astronomical research being conducted at LIGO and the Planck Collaboration [2], exploration of the multivariate posteriors is extremely computationally demanding. The parameter space being explored becomes exponentially larger as the dimensionality is increased. Direct evaluation of the entire posterior thus becomes impossible. The posterior is therefore represented by a set of samples from the original distribution.

Several sampling processes have been developed with the aim of analysing large data sets in an efficient manner. Many, including those used by LIGO [4–6], are based on Markov-Chain Monte-Carlo (MCMC) and Metropolis-Hastings (MH) methods. However, such methods risk giving misleading results if applied without proper consideration. Sampling effec-

tiveness is dependent on the shape of the distribution and the initial parametrisation of the sampling process.

The posteriors shown in Figure 1 are highly correlated and so are not optimally explored by conventional sampling methods described above. The creation of affine-invariant and ensemble sampling procedures attempt to address some of these limitations.

The aim of this project is to implement and test novel affine-invariant sampling techniques and compare their effectiveness against alternative MCMC procedures. One of the contributions of this research was to modify traditional samplers to enable effective sampling on non-Euclidean parameter spaces, such as those defined by spherical angular coordinates.

The structure of this report is as follows. Section 2 outlines the background theory behind the detection and Bayesian analysis of GW and reviews the state of MCMC methods. Section 3 introduces novel affine invariant methods which are assessed and compared against traditional samplers in Section 4. Section 5 discusses the adaptation and tests of affine samplers to non-Euclidean parameter spaces. The application of samplers to simulated LIGO data is examined in Section 6. Possible areas of future research are explored in Section 7 and conclusions are drawn in Section 8.

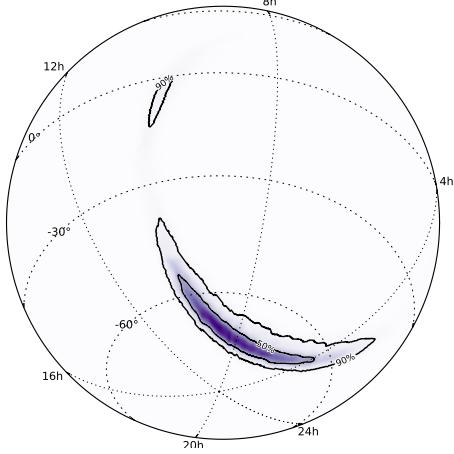


Figure 1: An orthographic projection of the posterior PDF for the sky location. Note the high degree of correlation. Traditional methods would not be effective at sampling from this distribution without calibration. Figure reproduced from [1].

2 Background

Sections 2.1 and 2.2 provide an overview of the key aspects in GW theory relevant to this report. A more detailed account of the theory can be found in [7–11].

2.1 Origin of Gravitational Waves

Assume an observer lying in a static region described by the metric $g_{\mu\nu} = \eta_{\mu\nu}$ (i.e. the Minkowski metric). Fluctuations in the metric can be described by a linear perturbation:

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \quad (1)$$

where the space-time fluctuations are assumed to be small, $|h_{\nu\mu}| \ll 1$. By solving the Gravitational Field Equations for the metric in Equation (1), Einstein showed that the tensor which represents the gravitational field:

$$\tilde{h}_{\mu\nu} = h_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}h_{\alpha}^{\alpha},$$

satisfies the wave equation in vacuum, and hence has a solution:

$$\tilde{h}_{\mu\nu} = A_{\mu\nu} \exp(ik^{\alpha}x_{\alpha}),$$

where $A_{\mu\nu}$ describes the wave's polarisation and amplitude. By setting an appropriate gauge (the Transverse-Traceless gauge), it can be shown that waves propagating in the z direction, $A_{\mu\nu}$ are defined by an orthonormal basis

$$h_+ = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad h_{\times} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \quad (2)$$

The tensor h_+ simultaneously causes expansion in the x axis and contraction in the y axis (and vice versa), and h_{\times} acts similarly in the direction 45° to the x, y axis.

2.2 GW Detection and Analysis

LIGO detects the passing of GW by measuring the differential arm length (ΔL) between the perpendicular arms of two interferometers. ΔL is proportional to the gravitational strain h which describes the fractional change in proper space caused by a gravitational perturbation. h can be written as a linear combination of the polarisation tensors in Equation (2):

$$h(t) = F_+h_+(t) + F_{\times}h_{\times}(t),$$

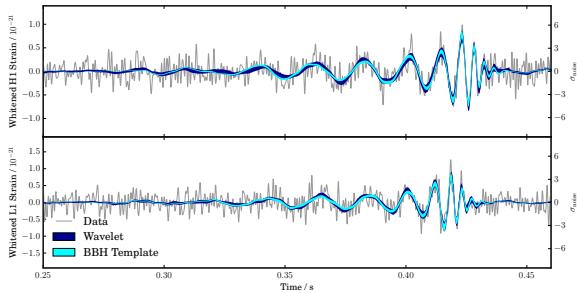
where F_+, F_{\times} are functions dependent on the detector's geometry and the BH Merger.

Exact solutions of the Einstein's field equations have not yet been found, leading to the development of analytic approximation schemes which depend on the system under consideration. One such scheme is the Post-Newtonian approximation (PN) which gives analytic solutions of the field equations in increasing orders of v/c .

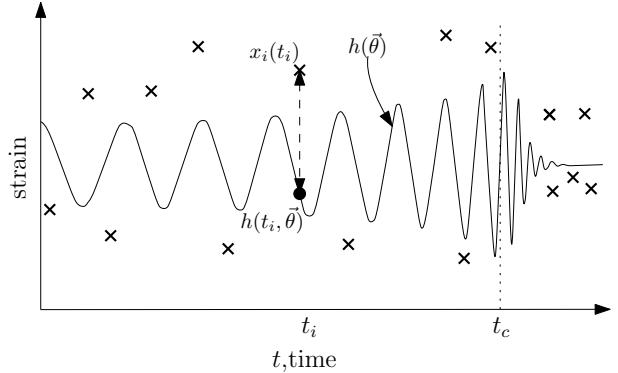
The likelihood analysis carried out in Section 6 uses results from a simulation based on a 2nd order PN expansion for inspiralling BH binaries. It generates a series of discrete theoretical values of the strain and adds Gaussian uncertainty to each data point. The likelihood of the data set is given by:

$$\mathcal{L}(\boldsymbol{\theta}) = P(\{x_i(t_i)\}|\boldsymbol{\theta}) = \prod_{i=1}^{N_D} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x_i(t_i) - h_i(t_i, \boldsymbol{\theta}))^2}{2\sigma^2}\right), \quad (3)$$

where $x_i(t_i)$ is the measured data and $h_i(t_i, \boldsymbol{\theta})$ is the theoretical strain at time t_i for the BH PN model parameter set $\boldsymbol{\theta}$ (see Figure 2b). Finding the likelihood for



(a) Trace of the GW strain measured by the two LIGO interferometers. The measured data (grey) and reconstructed waveforms (cyan) are shown. Taken from [1].



(b) Example of discrete data set $x_i(t_i)$ against the theoretical trace $h(t, \vec{\theta})$ for a parameter set $\vec{\theta}$. The parameter set $\vec{\theta}$ which minimises the total square distance $\sum_i (x_i(t_i) - h(t_i, \vec{\theta}))^2$ is the maximum likelihood point. t_c is the coalescence time of the BH merger.

Figure 2: Examples of GW traces, comparing the measured data and theoretical values.

all values of $\boldsymbol{\theta}$ is computationally impossible. Sampling methods such as those discussed in this section are used to draw representative samples from the desired distribution.

System parameters could in principle be chosen using the maximum likelihood technique. However, this approach ignores any prior knowledge about the system. Figures 3a and 3b show the likelihood for a merger at position $(\theta, \phi) = (\pi/2, \pi)$. The likelihood in Figure 3a is conditional on the parameter set (except angular location) taking their true values. Figure 3b the angular likelihood when marginalised over the parameter set. The likelihood in Figure 3a is unimodal and centred tightly about the true angular coordinates. The likelihood in Figure 3b is not maximal near the correct parameter set, showing that prior knowledge of the other parameters must be considered when estimating a set. This is achieved via Bayesian analysis.

2.3 Bayesian Analysis

It is appropriate at this stage to motivate the search for posterior distributions, such as those shown in Figure 1.

Bayesian analysis treats unknown model parameters as random variables. The uncertainties of these parameters are specified by associated probability distributions. Central to any discussion on Bayesian

analysis is Bayes' theorem. It relates the posterior $P(X|D, M)$ to the likelihood, $\mathcal{L} = P(D|X, M)$, via the prior, $P(X|M)$, and evidence $P(D|M)$:

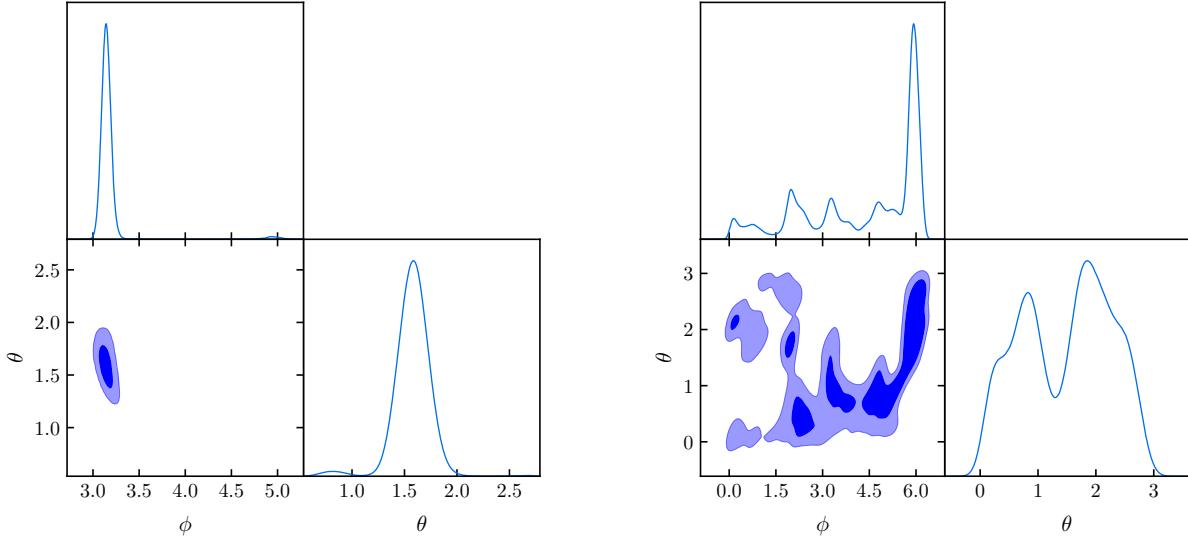
$$P(X|D, M) = \frac{P(D|X, M)P(X|M)}{P(D|M)}. \quad (4)$$

In the following discussion, X is the model's parameter space, and D is the data the model aims to describe [12].

The posterior distribution is often used to determine several features of the data expressed in terms of the posterior expectations:

$$\begin{aligned} E[f(X)|D] &= \int f(X)P(X|D, M)dX \\ &= \frac{\int f(X)P(D|X, M)P(X|M) dX}{\int P(D|X, M)P(X|M) dX}, \end{aligned} \quad (5)$$

where $f(X)$ is a function of the parameter set X [13]. In the context of the experiments conducted at LIGO, the masses and location of the BH merger can therefore be determined using the posterior shown in Figure 1.



(a) *Likelihood conditional on other parameters defining a BH merger.*

(b) *Likelihood marginalised over other parameters defining a BH merger.*

Figure 3: *BH location likelihoods for a merger at $(\theta = \pi/2, \phi = \pi)$.*

2.4 Monte Carlo Processes and Other Sampling Methods

For large probability distributions, the quantities in Equations (4) and (5) are impossible to calculate and are often estimated using Monte Carlo methods. The central idea is that the posterior expectation given by Equation (5) is estimated by calculating the mean value of $f(X_t)$ for a set of independent sample points X_t drawn from the posterior distribution [14].

MCMC is an adaptation of the Monte Carlo method which constrains the dependency of successive samples through a reversible Markov chain with an equilibrium distribution equal to the target distribution. This condition is satisfied if the Markov Chain is ergodic and has a transition probability $p(X, Y)$ which satisfies the ‘reversibility condition’:

$$\mathcal{L}(X)p(X, Y) = \mathcal{L}(Y)p(Y, X),$$

where $\mathcal{L}(\cdot)$ is the target distribution and $p(X, Y)$ is equal to the probability of the chain transitioning from $X \rightarrow Y$ [15].

2.5 Metropolis-Hastings Sampler

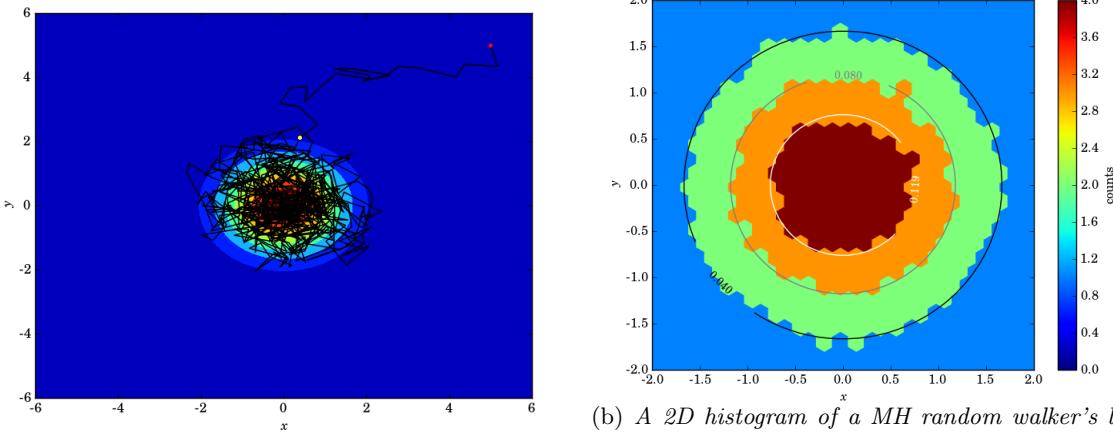
One of the most extensively used MCMC samplers is the Metropolis-Hastings algorithm. Its applicative flexibility makes it an increasingly popular sampling method compared to other methods such as the Gibbs Sampler [13, 15].

In MH processes, the sampling point X_{i+1} is drawn from the candidate generating probability distribution $q(\cdot|X_i)$. The general MH sampling process for a distribution \mathcal{L} can be summarised as follows:

1. Given a starting point X_i from a Markov series, select a trial point Y from the distribution $q(\cdot|X_i)$.
 2. Calculate the acceptance ratio $\alpha(X_i, Y)$:
- $$\alpha(X_i, Y) = \min \left\{ 1, \frac{\mathcal{L}(Y) \cdot q(Y|X_i)}{\mathcal{L}(X_i) \cdot q(X_i|Y)} \right\}.$$
3. Accept the step $X_{i+1} = Y$ with probability $\alpha(X_i, Y)$.
 4. Otherwise, set $X_{i+1} = X_i$.

MH methods must be applied with caution. One significant drawback is related to the correct number of

Figure 4: Plots illustrating (a) the random walk path taken by the MH walker and burn in period $X_0 - X_m$ (b) a 2D histogram for an MH sampling run with contours representing the normalised 75%, 50% and 25% probability lines of the target distribution. Note that the contours from the normal PDF occur near the transition of bin colour, showing that the MH iteration successfully mapped the target distribution.



(a) A MH walker path starting at X_0 (red dot) sampling from a 2D symmetric Gaussian with $\sigma_x = \sigma_y = 1$. The yellow dot X_m estimates when the walker starts sampling the stationary target distribution.

(b) A 2D histogram of a MH random walker's location used to sample a symmetric normal distribution. Hexagonal bins were coloured so that those coloured burgundy, orange, green and blue contained the top 25%, 25-50th percentile, 50-75th percentile, and the bottom 25% of steps within each bin.

iterations m (known as the *burn in*, shown in Figure 4a) that a chain will take to converge to equilibrium. One suggestion to mitigate against the variance due to the burn in period is to run an ensemble of Markov chains all sampling from the target distribution. The outputs of the random walker ensemble are used to give an estimate for the variance for the sampled distribution. The number of iterations can be increased until the estimated variance is at an acceptable level.

A second problem is related to the choice of $q(\cdot)$ upon which the sampling efficiency depends. The optimal choice for $q(\cdot)$ depends on the system under investigation [15]. Efficient sampling from the distributions calculated by LIGO poses difficulties that cannot be directly solved by a $q(\cdot)$ based approach. Conventional methods such as MH are variant to linear coordinate transformations of the parameter space. They are extremely slow at sampling highly correlated and non-spherically symmetric distributions (see Figure 1).

Affine-invariant MCMC processes have been shown to perform significantly faster than alternative methods when sampling highly correlated distributions. Ensemble sampling methods which enable commu-

nication between walkers have been extremely effective [16–18]. They open the door for the development of other affine-invariant sampling methods. The search for such algorithms is the main motivation behind this project.

2.6 Affine Invariant Ensembles Samplers

As discussed in the previous section, the inefficient sampling of highly correlated distributions using MH algorithms can be greatly improved by taking an affine transformation of the form $\mathbf{y} = \mathbf{Ax} + \mathbf{b}$. Alternatively the proposal distribution $q(\cdot|X_i)$ (normally a multivariate Gaussian) is calibrated so that it captures the symmetry of the target distribution. This involves the tuning of $\sim O(D^2)$ parameters where D is the dimension of the parameter space. The optimal calibration of such parameters is an extremely lengthy process and requires knowledge of the target distribution's shape. This motivates the search for sampling algorithms that are unaffected by linear transformations of the sampling space and involve fewer tunable parameters [16].

The samplers investigated in this report differ by calculating the trial steps using linear combinations of walker-to-walker displacements (see Figure 5). If the ensemble of N walkers set at $t = 0$, $\vec{\mathbf{X}}(t = 0) = (\mathbf{x}_1(0), \mathbf{x}_1(0), \dots, \mathbf{x}_N(0))$ is transformed by an affine transformation \mathbf{A} to give $\vec{\mathbf{X}}_{\mathbf{A}}(t = 0)$, the resulting step taken by any walker will be of the form

$$\sum_{t=0}^T \left(\sum_{v=1}^N \alpha_{v,t} \{ \mathbf{A} \mathbf{x}_v(t) \} \right) = \mathbf{A} \left(\sum_{t=0}^{T-1} \sum_{v=1}^N \alpha_{v,t} \mathbf{x}_v(t) \right),$$

where the sampler has run for $T + 1$ iterations and $\sum_{v=1}^N \alpha_{v,t} \mathbf{x}_v(t)$ is the step taken on the t^{th} sampling iteration. Therefore, the transformed ensemble at time $t = T$ is

$$\begin{aligned} \vec{\mathbf{X}}_{\mathbf{A}}(t = T) &= \mathbf{A} \vec{\mathbf{X}}(t = T) \\ \implies \vec{\mathbf{X}}(t = T) &= \mathbf{A}^{-1} \vec{\mathbf{X}}_{\mathbf{A}}(t = T), \end{aligned}$$

showing that the ensemble sampler is unaffected by affine transformations [18].

2.6.1 Emcee: The MCMC Hammer

One affine invariant method, emcee, has emerged as an effective sampling algorithm, and has been widely implemented by the astrophysical community. A brief outline of the algorithm will be given below; for a more detailed discussion see Goodman & Weare [16].

The ‘stretch move’ used to obtain the trial step is summarised by the following algorithm:

1. For the i^{th} iteration, randomly select trial and complementary walkers ($X_{a,i}$ and $X_{b,i}$ respectively) from the ensemble.
2. Draw a random number Z_i from the distribution $g(Z = z)$ where

$$g(z) \propto \begin{cases} \frac{1}{\sqrt{z}} & \text{if } z \in \left[\frac{1}{a}, a \right] \\ 0 & \text{otherwise} \end{cases}$$

and $a > 1$.

3. Determine the trial step Y :

$$\begin{aligned} Y &= X_{b,i} + Z_i(X_{a,i} - X_{b,i}) \\ &= X_{a,i} + (1 - Z_i)(X_{b,i} - X_{a,i}). \end{aligned}$$

The choice of Z_i from the distribution $g(z)$ ensures that emcee satisfies the reversibility condition discussed in Section 2.4. Figure 5a illustrates how the trial step is calculated.

4. Accept the step $X_{a,i+1} = Y$ with probability $\alpha(X_{a,i}, Y)$ where

$$\alpha(X_{a,i}, Y) = \min \left\{ 1, Z_i^{N-1} \frac{\mathcal{L}(Y)}{\mathcal{L}(X_{a,i})} \right\},$$

where N is the dimension of the sample space.

5. Otherwise set $X_{a,i+1} = X_{a,i}$.

The key idea behind emcee is that the trial step will be in a region of significant probability density since the walker density is representative of the target distribution.

Emcee only has one tunable parameter to determine the step scaling a , which can be adjusted to improve the sampler’s performance.

3 ADS and Multidimensional Emcee Samplers

3.1 Adaptive Directional Sampling

The investigation of the affine-invariant parallel Adaptive Directional Sampler (ADS) proposed by Wilks, Roberts and George is the principal goal of this report. ADS qualitatively differs from emcee by drawing information from other regions of the distribution. ADS does not require the tuning of any parameters to optimise sampling efficiency unlike MH or emcee methods. Its algorithm is listed below:

1. For the i^{th} iteration, randomly select three walkers $X_{a,i}$, $X_{b,i}$ and $X_{c,i}$ from the ensemble.
2. Determine the trial step Y :

$$Y = X_{a,i} + (X_{b,i} - X_{c,i}).$$

3. Accept the step $X_{a,i+1} = Y$ with probability $\alpha(X_{a,i}, Y)$ where

$$\alpha(X_{a,i}, Y) = \min \left\{ 1, \frac{\mathcal{L}(Y)}{\mathcal{L}(X_{a,i})} \right\}.$$

4. Otherwise set $X_{a,i+1} = X_{a,i}$.

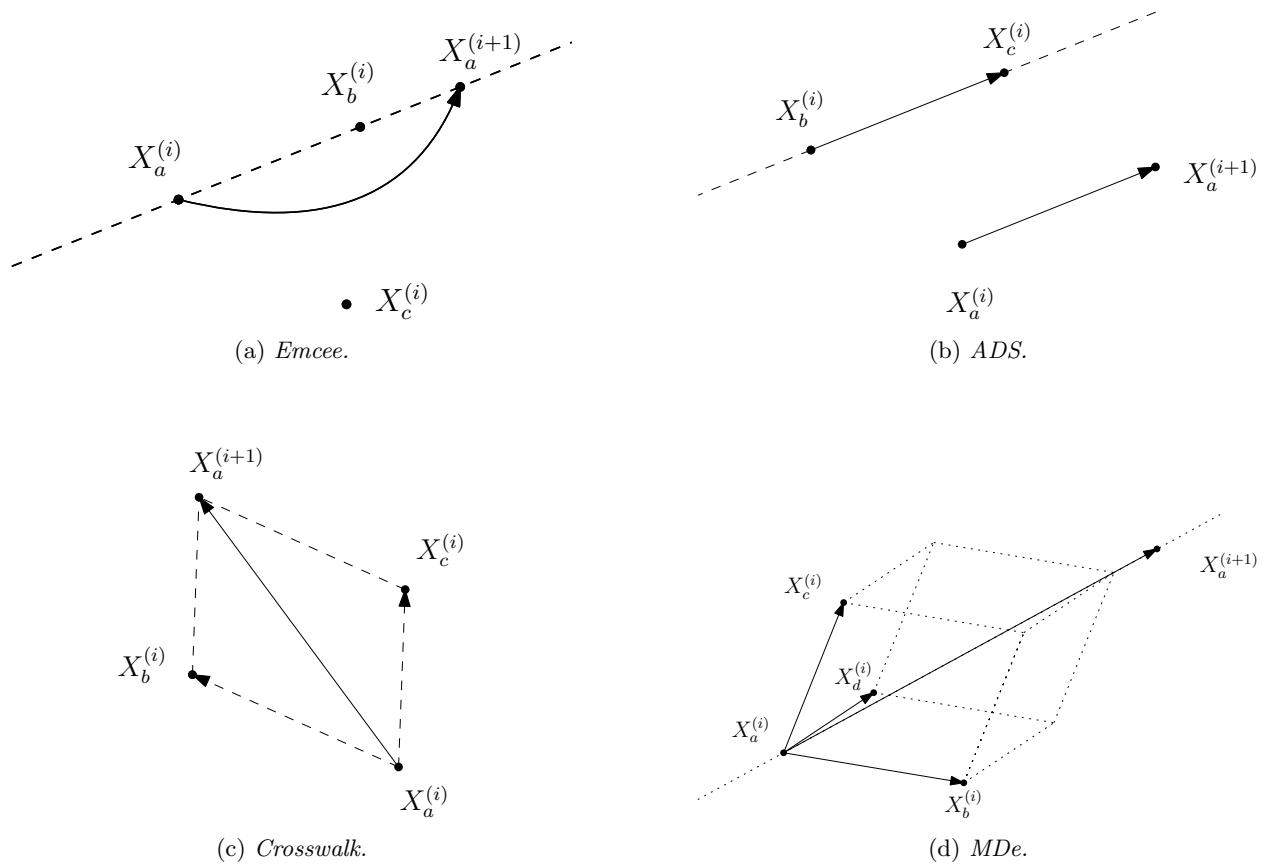


Figure 5: *Illustrations of ensemble sampler step calculations.*

3.2 Multidimensional Emcee

Another affine invariant method is the *crosswalk* sampling mechanism proposed by John Skilling (Figure 5c). A test walker X_a and two complementary points X_b, X_c are randomly selected from the ensemble. The trial point Y is obtained via the parallelogram rule:

$$Y = X_b + X_c - X_a.$$

This report proposes a variant method that considers aspects of both the crosswalk and emcee methods, defined in this report as the Multidimensional emcee (MDe). The proposed algorithm is illustrated in Figure 5d and is outlined below:

1. Randomly select N walkers $X_{1,i}, X_{2,i} \dots X_{N,i}$ where $N < S$. The walker to be moved is $X_{1,i}$.
2. Draw a number Z_i from the distribution $g(Z = z)$ defined in Section 2.6.1
3. Determine the trial step Y:

$$Y = X_{1,i} + \frac{(1 - Z_i)}{N - 1} \sum_{j=2}^N (X_{j,i} - X_{1,i}).$$

4. Accept the step $X_{1,i+1} = Y$ with probability $\alpha(X_{1,i}, Y)$ where

$$\alpha(X_{1,i}, Y) = \min \left\{ 1, Z_i^{N-1} \frac{\mathcal{L}(Y)}{\mathcal{L}(X_{1,i})} \right\},$$

where N is the dimension of the sample space.

5. Otherwise set $X_{1,i+1} = X_{1,i}$.

The goal of this method is to improve convergence of the sampler exploring parameter spaces of high dimensionality. Calculating a step using multiple walkers aims to generate a trial point in regions of greater probability density than emcee.

4 Assessment of Samplers on Euclidean Spaces

In this section, tests on the Euclidean sampling algorithms will be discussed. Sampler performance was assessed according to the following criteria:

1. **Accuracy of Likelihood estimate:** Density plots and comparisons of sample means and covariances against the true likelihood mean and covariance values.

2. Sampling efficiency: Efficient methods have fewer likelihood calls than their slower counterparts. This reduces the number of likelihood computations necessary to obtain a similar level of accuracy. Efficiency was assessed by comparing:

- The weights of samples (number of likelihood calls before a walker accepts a step) drawn using a particular algorithm.
- The autocorrelation of the sampling process, τ_{acor} . This gives a measure of the number of steps required for the MCMC sampler to reach its equilibrium state [19].
- The fraction of proposed steps that are accepted, a_f , known as the acceptance fraction. Gelman, Roberts and Gilks (1996) suggest that $0.2 < a_f < 0.5$ for efficient samplers [17].

3. Scaling with dimensionality: The efficiency of MCMC processes when sampling from distributions of increasing dimensionality.

Numerical tests were carried out on the sampling algorithms using multivariate Gaussian distributions and the Rosenbrock density:

$$\mathcal{L}(x_1, x_2) \propto \exp \left(-\frac{100(x_2 - x_1^2)^2 + (1 - x_1)^2}{20} \right).$$

The former will be used to compare sampling performance against increasingly correlated distributions. The highly correlated latter distribution cannot be transformed back into a spherically symmetric form by a single affine transformation, and provides an example where affine invariant methods may struggle [16]. Contours of the Rosenbrock Density are shown in Figure 6.

In this section, two MH walkers were used. The first uses a spherically symmetric Gaussian of $\sigma = 1$ as the proposal distribution (MH Sph.). The second uses a calibrated Gaussian distribution with $q(\cdot)$ equal to the target likelihood (MH Cal.). Each sampling process was run three times. The tables in this report the mean and variance across these three measurements.

A list of acronyms used in the following sections is provided in Table 1 for reference.

Table 1: A list of acronyms detailing the samplers which are referred to in the report. See Section 5 for a discussion of the spherical and toroidal samplers at the bottom of the list.

Sampling algorithm	Acronym
Adaptive Directional Sampler	ADS
Multidimensional Emcee	MDe
Calibrated Metropolis-Hastings	MH Cal.
Spherical Metropolis-Hastings	MH Sph.
Parallel Transport Spherical Sampler	PTSS
Euclidean Approximated Spherical Sampler	EASS
Spherical Emcee	SE
Toroidal Emcee	TE
Toroidal ADS	TADS

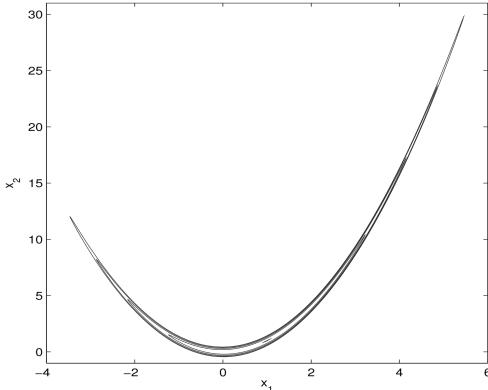


Figure 6: Contours from the Rosenbrock Density. Taken from [16].

4.1 Sampling Gaussian Distributions

A 5D multivariate Gaussian with mean $\mu = \mathbf{0}$ and a covariance matrix with on-diagonal variance elements $\Sigma_{ii} = 2$ was chosen as a test case. All covariance elements $\Sigma_{ij}(i \neq j)$ were set to the same value. The samplers' were tested against Gaussians of varying correlations. Four test runs of 5×10^6 iterations were carried out for $\Sigma_{ij} = \{1, 1.75, 1.9, 1.98\}$. An ensemble size of 100 walkers was chosen and the Goodman scaling factor was set to $a = 2.5$, which maximised the sampling efficiency. MDe steps were calculated using five walkers.

Table 2: Autocorrelation times (multiplied by 10^{-3}) and acceptance ratio for the Rosenbrock Density.

	τ_{acor}	a_f
ADS	57.4 ± 0.7	0.054
emcee	49.6 ± 6.2	0.224
MH Sph.	0.036 ± 0.001	0.0019
MDe	69.1 ± 6.9	0.19

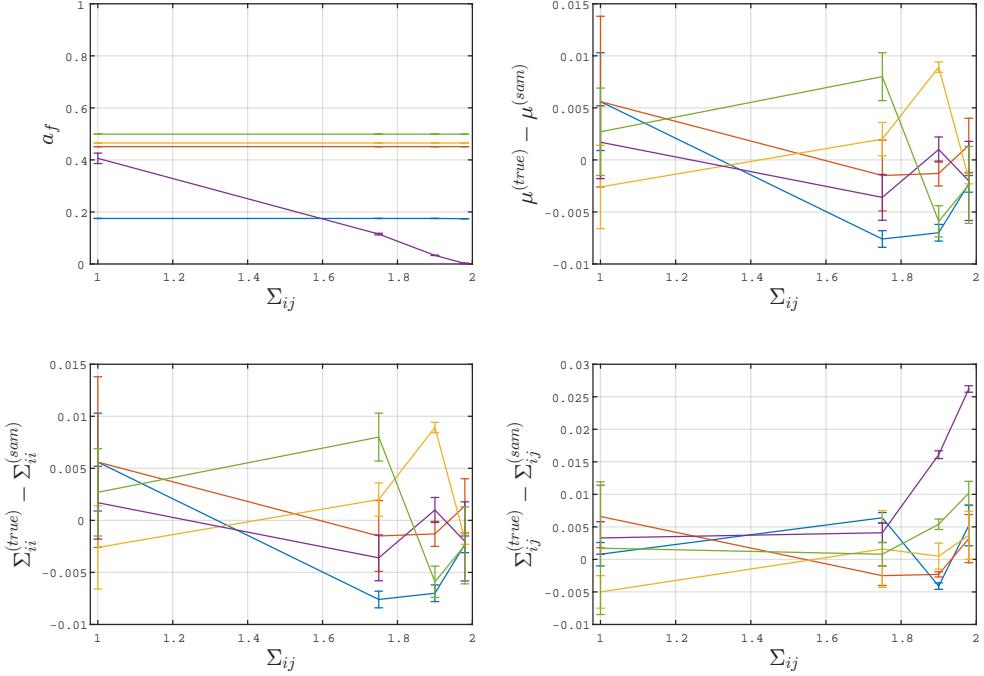
4.2 Sampling the Rosenbrock Density

The Rosenbrock function was sampled for $N = 10^7$ iterations. The scaling factor was set to $a = 2$ and the MH sample space was set to a spherical Gaussian with $\sigma = 1$. MDe used five walkers to determine the trial step.

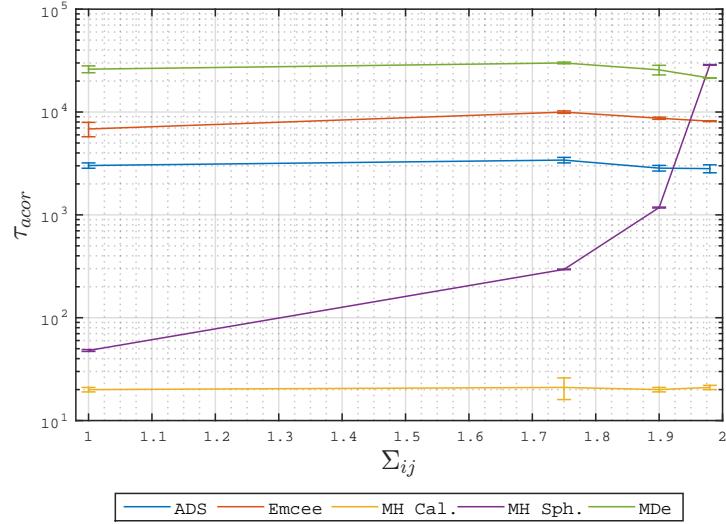
Table 2 shows of the autocorrelation times and acceptance ratios. Sample density plots and histograms of walker weights are shown in Figures 10a and 10b.

4.3 Scaling with Dimensionality

The algorithms' ability to scale with increasing likelihood dimensionality was assessed by sampling from multivariate Gaussians parametrised by $\Sigma_{ii}, \Sigma_{ij} = 2.0, 1.75$. For an ensemble size of 150, Gaussians of



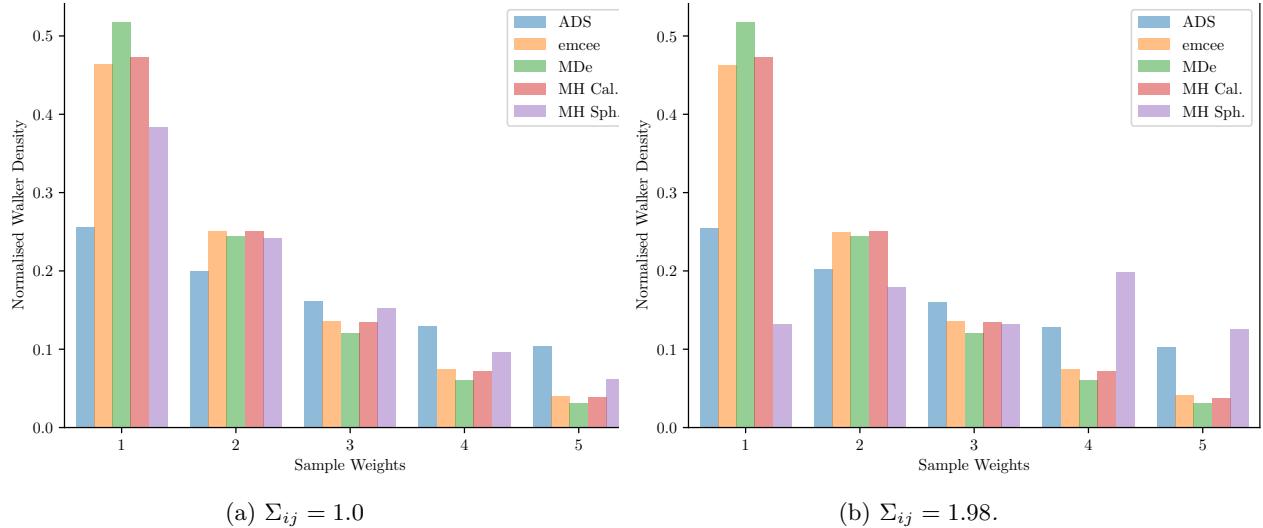
(a) Clockwise starting from the top-left: (i) Acceptance ratios (ii) Deviation of estimated mean from the true value (iii) Deviation of estimated on-diagonal covariance terms from the true value (iv) Deviation of estimated off-diagonal covariance terms from the true value.



(b) Autocorrelation times.

Figure 7: Comparison of algorithms sampling multivariate Gaussians of varying correlation. Note that the weights of the affine-invariant algorithms do not change when the correlation is varied.

Figure 8: Histograms of walker weights for Gaussians of different correlations.



dimensions $D = \{2, 10, 15, 20\}$ were used.

The number of walkers used by the multidimensional emcee was set to the dimensionality of the sampling space. Figure 9 shows the autocorrelation times and acceptance ratios from the sampling runs.

4.4 Discussion

Figures 7a and 7b show that the autocorrelation times and acceptance ratios do not fluctuate significantly for the affine invariant methods. This result strongly supports the discussion in Section 2.6 which stated that the performance of the ensemble methods are not affected by increasingly correlated distributions. At $a_f = 0.175$, ADS has a significantly lower acceptance ratio than emcee methods. Figure 7b shows that ADS has the lowest autocorrelation time out of the affine-invariant methods. While ADS draws samples at half the rate of emcee, this increased selectivity may result in a quicker convergence to the equilibrium distribution.

All samplers reconstruct the expected mean and covariance to a high degree of accuracy (within 10^{-5} of the true value). MH Sph. and MDe struggle to reconstruct highly correlated distributions.

MH Cal. outperforms the other algorithms' sampling efficiency, and reconstructs the likelihood mean and covariance to the same accuracy as the affine-invariant methods. If the proposal distribution $q(\cdot)$

is adjusted to match the symmetry of the target likelihood, MH remains a reliable method for likelihood sampling, even for highly correlated likelihoods. However, the results for MH Sph. shows that a poor choice of $q(\cdot)$ may result in inaccurate and inefficient sampling. This highlights one of the main strengths of affine-invariant methods: they do not require any calibration to sample effectively.

The results shown in Figure 9 and Figures 11a and 11b show that the sampling efficiency of emcee and MDe are less affected by increasing dimensionality than the alternative samplers. The ratio between the 2D and 20D acceptance ratios for ADS is of order 10^2 . MDe has the highest acceptance ratio as dimensionality is increased, although emcee is not significantly lower. Emcee also has a lower autocorrelation time than MDe and requires fewer likelihood calls to converge onto the target distribution.

Emcee outperformed all other algorithms when sampling the Rosenbrock density (see Table 2 and figure 10b). Although values of τ_{acor} have similar orders of magnitude for all ensemble methods, Figure 10b shows that emcee explored a much greater region of likelihood space than ADS. MH sampled from a significantly reduced region, showing that while it converges quicker than ensemble methods, it poorly samples highly correlated, non-symmetric likelihoods. MDe performed worse than emcee on both autocorrelation time and acceptance ratios showing that using more walkers to calculate the trial step may in fact

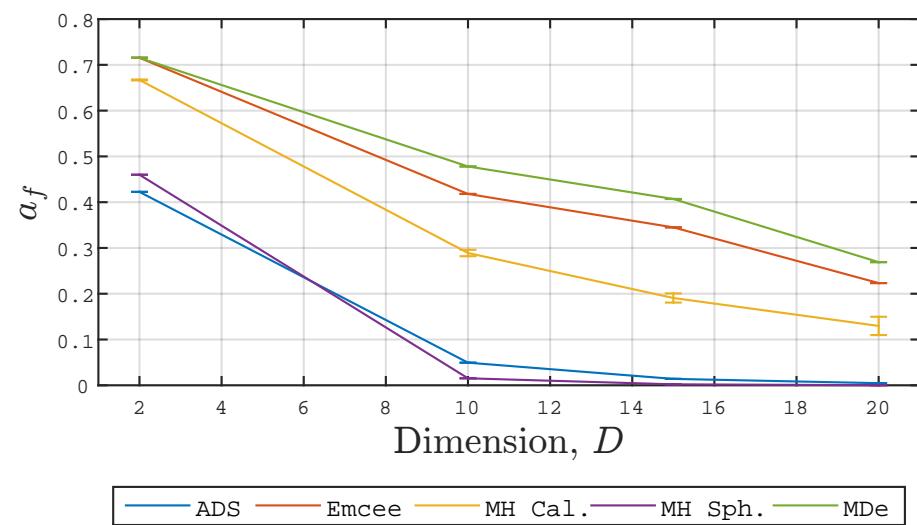
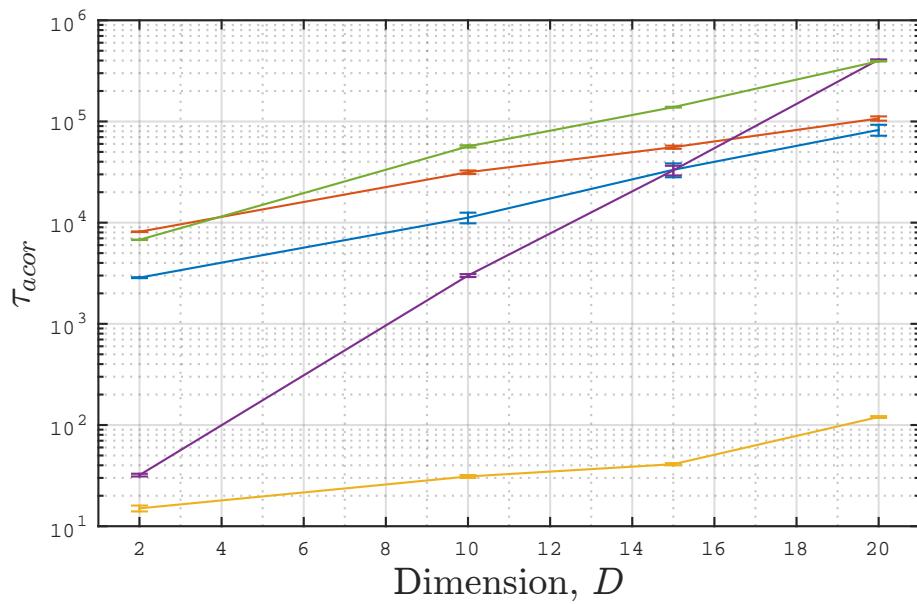


Figure 9: Autocorrelation times (top) and acceptance ratios (bottom) for Gaussians of increasing dimensionality.

Figure 10: Histogram and density plots from Rosenbrock sampling.

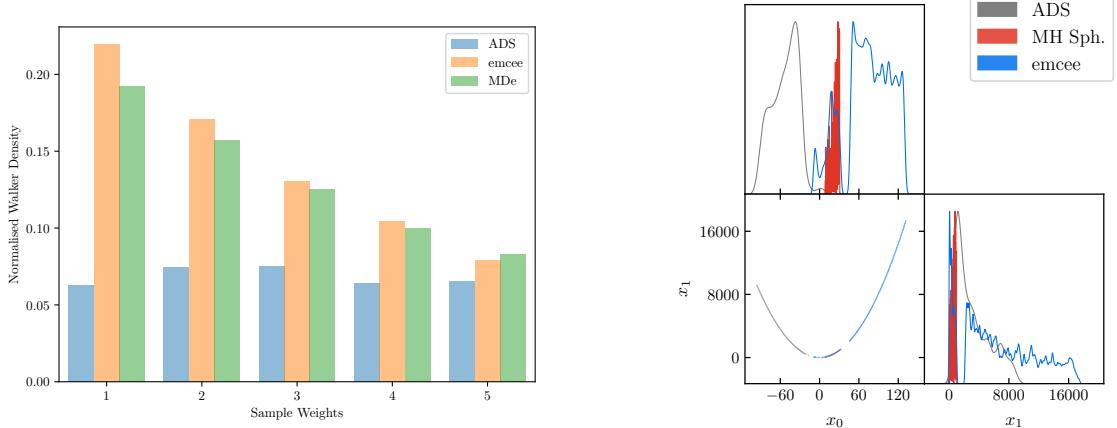
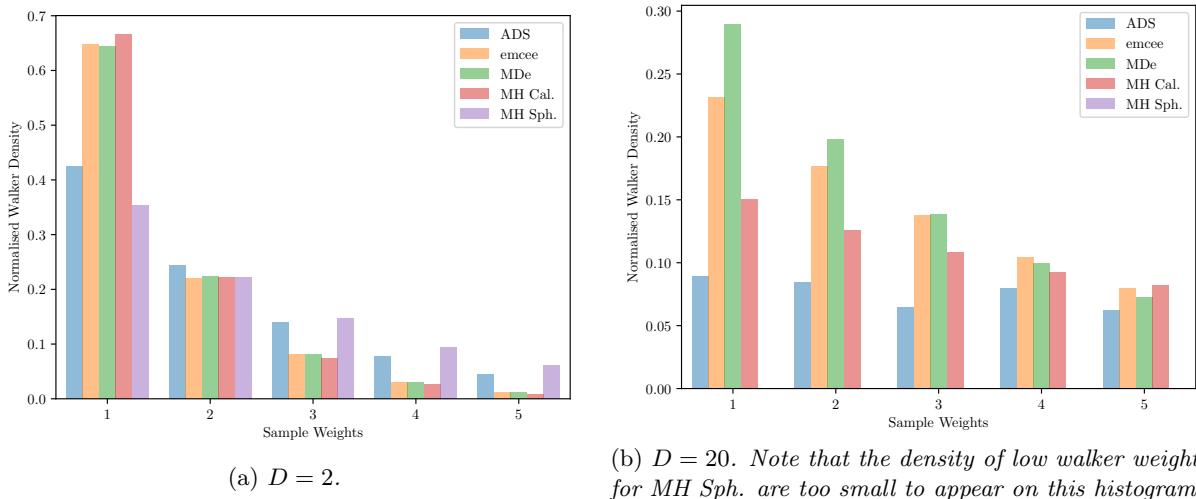
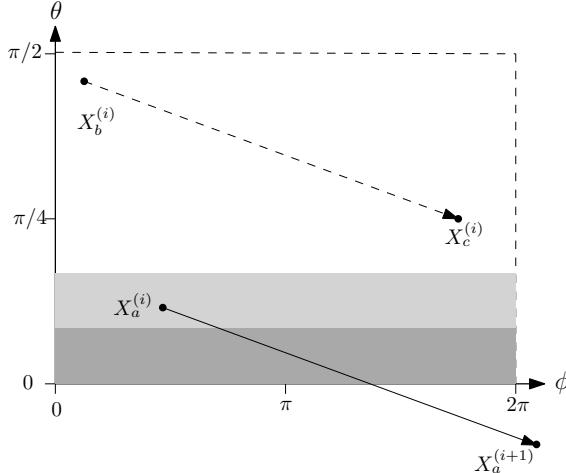
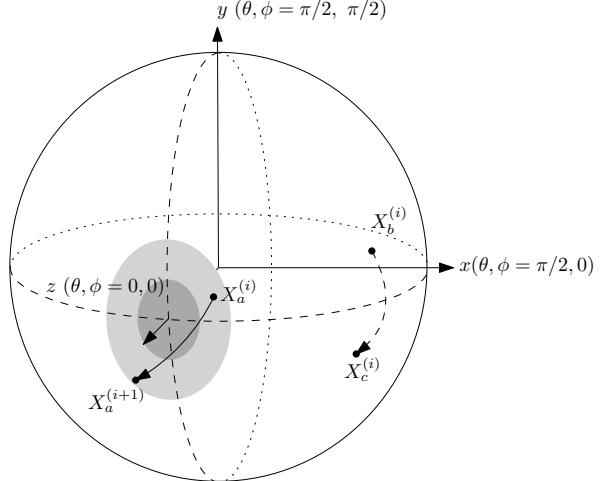


Figure 11: Histograms of walker weights for Gaussians of different dimensions.





(a) Parallel ADS sampling from a Euclidean mapping of angular coordinates.



(b) ADS sampling on a 2-sphere.

Figure 12: Comparison of sampling methods on Euclidean and spherical geometries. The same probability distribution (grey) is mapped onto each likelihood geometry.

hinder the performance of an algorithm.

5 Non Euclidean Sampling Methods

Many variables that parametrise a likelihood define specific geometries which emerge from their physical properties. For example, the spherical angular coordinates defining the location of the merger (θ, ϕ) parametrise a non-Euclidean space, the 2-sphere. The algorithms discussed in Sections 2.5 and 2.6 are unable to effectively sample such distributions as trial steps outside the angular domain are automatically rejected, limiting the efficiency of the sampler. Figure 12 shows an automatically rejected trial step on a Euclidean map of angular coordinates that may be accepted on an equivalent 2-sphere parameter space.

One of the goals of this report is to investigate three adaptations of the methods described in Section 2.6. The crucial contribution here is the determination of the trial step on a spherical surface; all other steps in the algorithm are identical to the Euclidean alternative. In the following discussion the stepping vector from A to B ($\hat{\mathbf{t}}_{ab}$) lies along the geodesic defined by the walkers.

5.1 Spherical emcee

An adaptation of emcee to a 2-sphere is presented below. Instead of choosing a trial point Y along the straight path connecting two randomly selected walkers X_a and X_b , spherical emcee will select a trial point along the geodesic connecting X_a and X_b .

5.2 Spherical ADS

Two variants of ADS were considered. The first calculates the stepping direction using a Euclidean approximation of the parameter space:

1. Calculate the tangent vectors on the 2-sphere surface pointing from X_a to X_b ($\hat{\mathbf{t}}_{ab}$) and X_a to X_c ($\hat{\mathbf{t}}_{ac}$).
2. Rotate X_a by θ_{bc} (the angular separation between X_b to X_c) on the great circle defined by $\overrightarrow{OX_a}$ and the vector $\hat{\mathbf{t}}_{walk}$ where

$$\hat{\mathbf{t}}_{walk} \propto \theta_{ab} \hat{\mathbf{t}}_{ab} + \theta_{ac} \hat{\mathbf{t}}_{ac}.$$

This method is referred to in what follows as the Euclidean Approximated Spherical Sampler (EASS).

A second method parallel transports the stepping vector along geodesics of the sphere. It is referred to

in this report as Parallel Transport Spherical Sampler (PTSS), and is outlined below:

1. Parallel transport the tangent vector $\hat{\mathbf{t}}_{cb}$ pointing from X_C to X_B along the great circle connecting X_C and X_A
2. Rotate X_a by θ_{bc} (the angle separation between X_b to X_c) on the great circle defined by $\overrightarrow{OX_a}$ and the parallel transported vector $\hat{\mathbf{t}}_{cb}'$.

None of the three methods are strictly affine invariant and such a concept does not generalise straightforwardly to curved spaces. This report does not look to prove whether sampling algorithms can be generalised to any curved space, but aims to investigate whether the stepping methods described in Section 2.6 can be adapted to non-Euclidean spaces.

5.3 Toroidal Samplers

ADS and emcee algorithms were adapted to sampling toroidal spaces. This report will only consider variables which have angular periodicity of 2π . Examples of such variables are the phase of the BH merger and the Earth's latitude locating the merger. As in the spherical samplers, only the step calculation process differs from the original algorithm.

The stepping vector on a torus was chosen to be the shortest distance between two walkers X_b, X_c on the torus. The Toroidal ADS (TADS) uses that vector to move walker X_a to its trial position. Toroidal emcee (TE) performs the stretch move along the direction defined by the stepping vector.

5.4 Non-Euclidean Sampler Tests

5.4.1 Test for Spherical Samplers

The spherical samplers described in Section 5 were compared against their Euclidean counter parts by sampling from a unimodal Gaussian distribution centred at $(\phi = 0, \theta = \pi/2)$, which should present the greatest challenge for non-Euclidean samplers. An ensemble of $S = 150$ walkers was chosen.

5.4.2 Tests for Toroidal Samplers

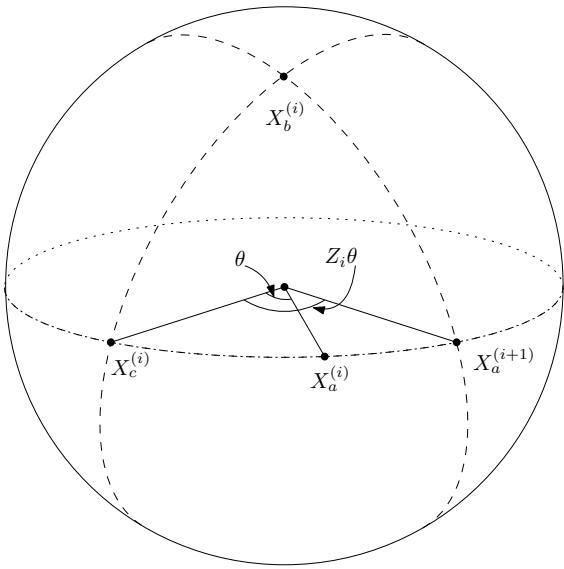
Toroidal samplers were compared against traditional ADS and emcee methods by sampling from unimodal distribution centred at the torus's origin, $(\theta_1 = 0, \theta_2 = 0)$. Such a distribution is mapped onto an Euclidean space as 4 regions of high probability density located at the corners of the parameter space. Traditional samplers would struggle to accurately sample this distribution. Ensembles of $S = 150$ were run for 1×10^6 iterations. The acceptance ratios and autocorrelation times for the samplers are presented in Table 4. The sampled means and covariances are compared with the true distribution in Table 6.

5.5 Discussion

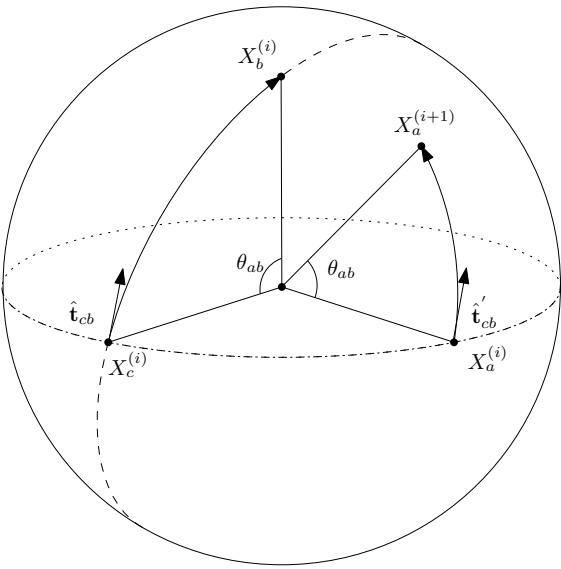
Table 5 suggests that of the three spherical samplers, PTSS most accurately maps the target likelihood. It outperforms the two Euclidean methods on both accuracy and efficiency, showing a clear advantage when sampling distributions centred near the edge of a 2-sphere Euclidean map. EASS has an extremely high acceptance ratio but has limited accuracy in its reconstruction of the target likelihood. It is possible that its lack of selectivity leads the acceptance of steps which hinder accurate reconstruction of the target covariance. Spherical emcee has the lowest auto correlation time, closely followed by PTSS (see Table 3). However, PTSS's increased accuracy of likelihood reconstruction makes it a more attractive choice as a sampling algorithm.

The two Euclidean methods reasonably reconstruct the spherical likelihood. ADS more accurately reconstructs the covariance whereas emcee better estimates the mean. Compared to emcee, ADS has a greater acceptance ratio and τ_{acor} , making it a more efficient sampler in bounded likelihoods. One suggestion for this is as follows. The scaling factor for emcee steps means that trial steps will often be greater than the original length between walkers. ADS step sizes are unscaled, and often smaller than emcee steps. This may cause a greater proportion of emcee trial steps to go out of bounds, leading to a slower convergence rate and acceptance ratio.

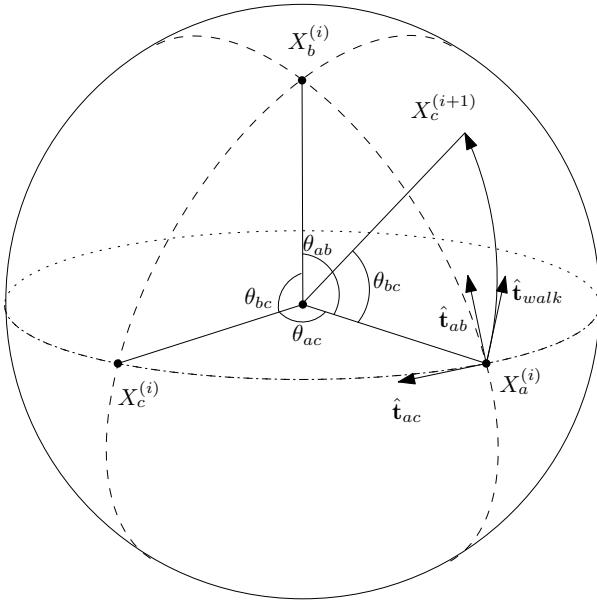
Tables 4 and 6 show that the toroidal samplers outperform traditional algorithms both in efficiency and accuracy. TE and TADS better estimated the mean and covariance than traditional methods. Since the



(a) *Spherical emcee (SE)*.



(b) *Parallel Transport Spherical Sampler (PTSS)*.



(c) *Euclidean Approximated Spherical Sampler (EASS)*.

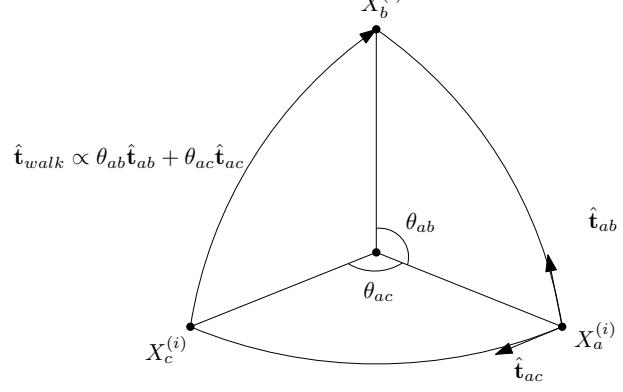


Figure 13: Illustration of ADS and emcee adaptations on a 2-sphere.

Table 3: Autocorrelation times and acceptance probabilities for spherical samplers ($S = 150$).

	$\tau_a/10^3$	a_f
PTSS	1.99 ± 0.59	0.420
SE	1.20 ± 0.14	0.215
EASS	4.31 ± 0.34	0.906
Euc. Emcee	17.9 ± 1.42	0.045
Euc. ADS	2.89 ± 0.67	0.2811

Table 4: Autocorrelation times and acceptance probabilities for toroidal samplers ($S = 150$).

	$\tau_a/10^3$	a_f
TADS	0.94 ± 0.11	0.420
TE	1.02 ± 0.32	0.415
Euc. Emcee	4.73 ± 1.70	0.017
Euc. ADS	1.65 ± 0.45	0.179

Table 5: Estimated means and covariances for Spherical sampling algorithms ($S = 150$).

	μ_ϕ	μ_θ	$\Sigma_{\phi\phi}$	$\Sigma_{\theta\phi}/10^{-3}$	$\Sigma_{\theta\theta}/10^{-1}$
True Samples	3.142	1.570	7.423	-0.0971	1.991
PTSS	3.141 ± 0.006	1.570 ± 0.010	7.166 ± 0.194	-0.147 ± 0.290	1.709 ± 0.053
SE	3.110 ± 0.017	1.570 ± 0.003	5.314 ± 0.034	-1.190 ± 0.185	3.003 ± 0.252
EASS	3.093 ± 0.021	1.571 ± 0.054	9.011 ± 0.726	2.161 ± 0.006	28.31 ± 0.35
Euc. Emcee	3.174 ± 0.147	1.569 ± 0.091	7.437 ± 0.509	102.8 ± 3.892	6.941 ± 0.857
Euc. ADS	3.231 ± 0.097	1.701 ± 0.162	6.493 ± 0.277	0.799 ± 0.733	3.000 ± 0.076

Table 6: Estimated means and covariances for Toroidal Samplers ($S = 150$).

	μ_ϕ	μ_θ	$\Sigma_{\phi\phi}$	$\Sigma_{\theta\phi}/10^{-4}$	$\Sigma_{\theta\theta}/10^{-1}$
True Samples	3.142	3.142	8.378	0.001	8.378
TADS	3.139 ± 0.009	3.154 ± 0.007	8.375 ± 0.012	-8.698 ± 4.122	8.388 ± 0.007
TE	3.178 ± 0.159	3.127 ± 0.138	8.308 ± 0.075	-28.25 ± 31.91	8.363 ± 0.049
Euc. Emcee	2.952 ± 0.798	3.155 ± 0.237	7.345 ± 0.518	108.2 ± 66.81	7.327 ± 0.884
Euc. ADS	3.018 ± 0.271	3.109 ± 0.147	8.485 ± 0.133	-101.4 ± 49.7	7.972 ± 0.238

Figure 14: Histogram and density plots for unimodal distribution on a sphere.

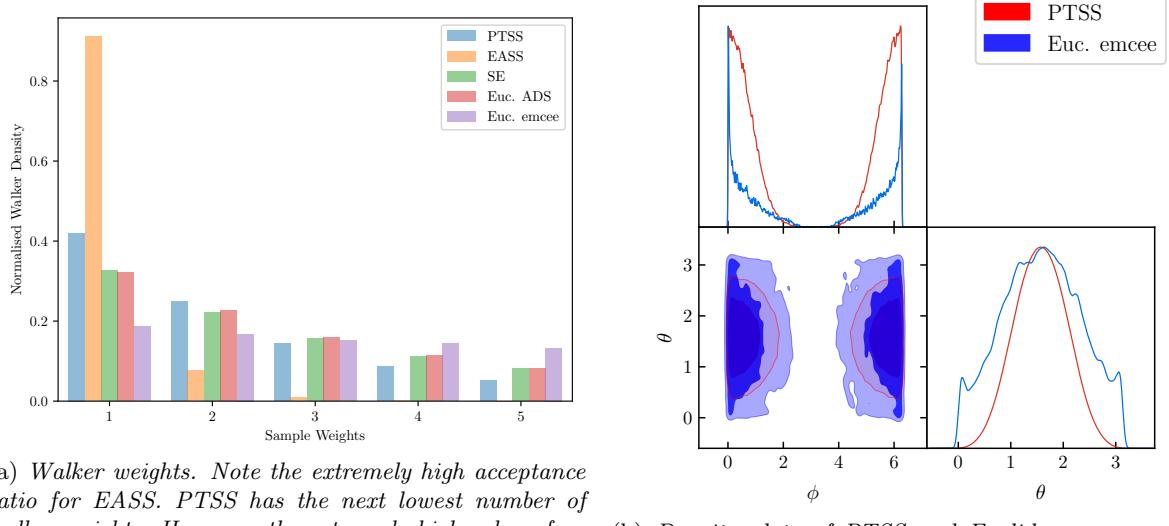


Figure 15: Histogram and density plots for unimodal distribution on a torus.

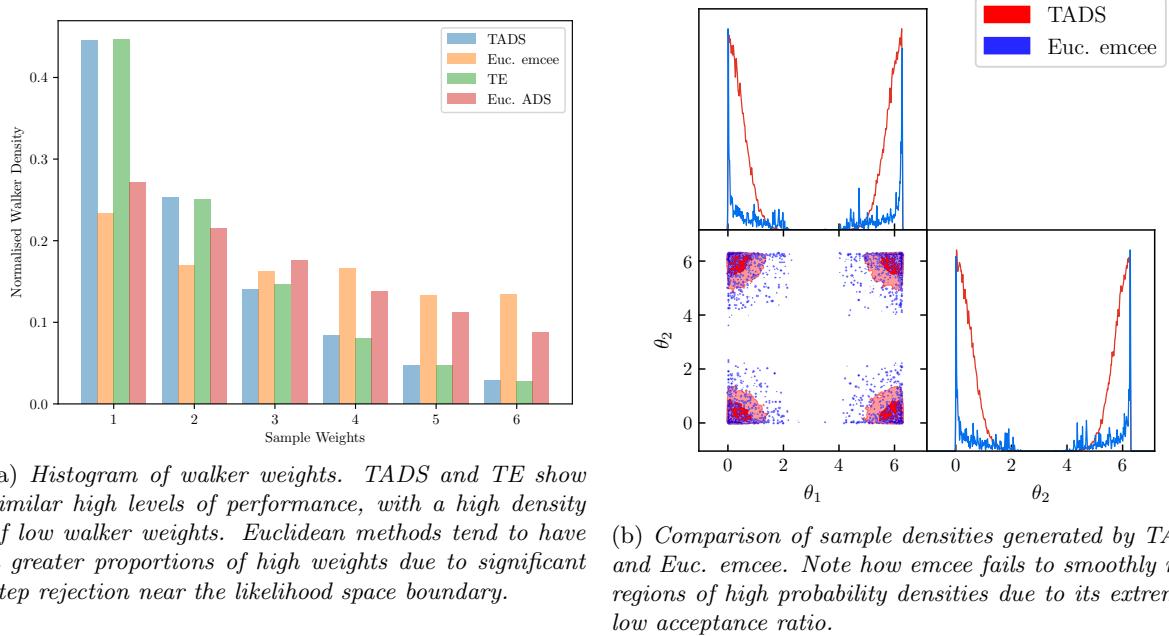


Table 8: Autocorrelation times and acceptance probabilities for LIGO sampling ($\theta = \pi/2, \phi = \pi$).

	$\tau_a/10^3$	a_f
PTSS	2.40 ± 0.65	0.369
SE	1.51 ± 0.40	0.050
EASS	3.41 ± 1.27	0.838
Euc. Emcee	4.46 ± 0.15	0.062
Euc. ADS	1.81 ± 0.24	0.382

Euclidean samplers treat the likelihood as four distinct modes, there is a significant probability that walkers will be unevenly distributed across the distribution. Certain probability modes may be sampled more frequently than others causing a high variance in the mean and covariance estimates.

6 LIGO Simulation

A final test of the samplers was based on the generation of samples from a simulated likelihood similar to that detected by LIGO¹. A total of nine physical quantities describe GW emission from a BH merger in the PN scheme. The Ensemble algorithms discussed in this report sampled the angular location of the merger defined on a 2-sphere (θ, ϕ) . The simulated angular coordinates were set to $(\theta = \pi/2, \phi = \pi)$. Samplers were run for 10^5 iterations with ensemble sizes of $S = 150$.

6.1 Discussion

Tables 7 and 8 highlight the strength of spherical samplers, in particular PTSS method. It has one of the lowest τ_{acor} and most accurately estimates the covariance and mean of the target likelihood. The SE and EASS samplers are able to accurately estimate the target mean, but are extremely weak at estimating the likelihood's covariance.

The accuracy of the μ, Σ estimates and the values of τ_{acor}, a_f for EADS match closely with PTSS. Emcee values for a_f, τ_{acor} show it to be an inefficient algorithm for spherical surfaces. The μ, Σ estimates

differ considerably from the expected values. Figure 16b presents a justification for this deviation. Emcee has not converged on the high density region centred at $(\theta, \phi) = (\pi/2, \pi)$ and has detected a high density region around $(\theta, \phi) \simeq (0.8, 5.4)$. This physically corresponds to the BH merging from ‘behind’ the LIGO detectors as opposed to from the front. The weaknesses of emcee as a sampler revealed qualitative information about the likelihood. The inability of LIGO to completely triangulate the BH merger’s locations was not revealed by the other samplers. Using slowly converging algorithms for an initial sampling run could be used to obtain a qualitative picture of the likelihood. Such samplers may prove useful in detecting weak secondary modes in multi-modal distributions.

7 Further Study

This project assesses the suitability of the ADS algorithm and non-Euclidean adaptations of other ensemble methods. The next step would be to test the algorithms with a full sampling run of the entire LIGO likelihood. In addition, investigating whether the combination of Euclidean and non-Euclidean sampling algorithms across a range of spaces significantly increases sampling performance is an interesting area to pursue

Sampling from multimodal distributions remains a significant challenge for ensemble samplers. Section 6 discussed the possibility of using slowly converging samplers to obtain a qualitative picture of likelihood modes. Hybridisation of different various algorithms to exploit their individual advantages could help resolve this issue. Another area to explore is based on the identification of walker clusters before full ensemble convergence occurs. This could identify different modes of the target distribution by dividing the clusters into independent ensembles.

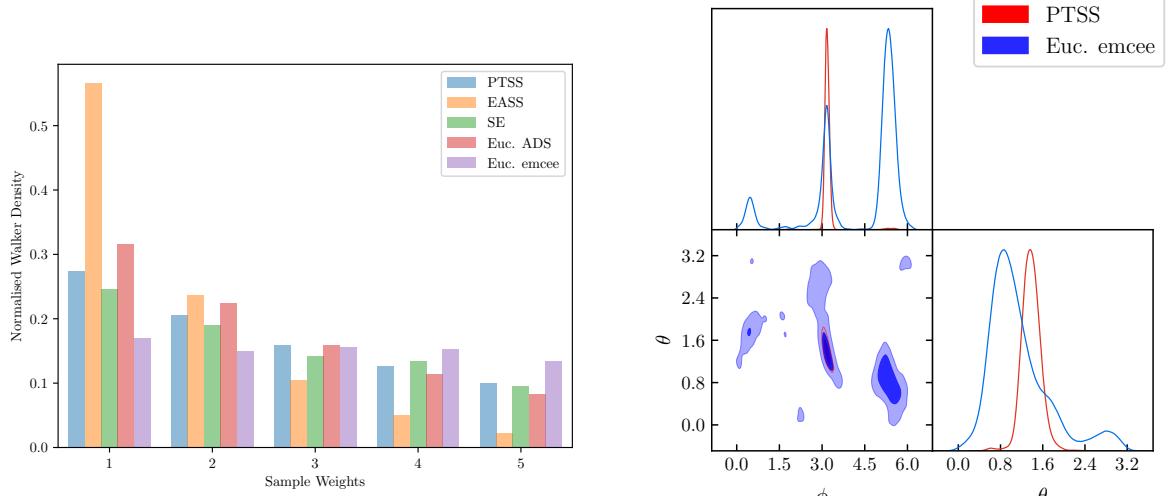
Whilst this report introduced the concept of non-Euclidean samplers, no mathematical investigation into what other algorithms exist that are invariant to other non-linear transformations. A possible topic for further work would be to investigate whether the methods discussed in this report can be generalised to any curved space.

¹The simulation was kindly provided by Dr. Will Handley.

Table 7: Estimated means and covariances for LIGO sampling ($\theta = \pi/2, \phi = \pi$).

	μ_ϕ	μ_θ	$\Sigma_{\phi\phi}$	$\Sigma_{\theta\phi}/10^{-3}$	$\Sigma_{\theta\theta}/10^{-1}$
True Samples	3.1401	1.5645	0.0020	-0.001	0.031
PTSS	3.108 ± 0.024	1.577 ± 0.009	0.076 ± 0.005	-0.003 ± 0.004	0.0437 ± 0.009
SE	3.094 ± 0.046	1.543 ± 0.011	1.138 ± 0.45	-0.581 ± 0.063	0.473 ± 0.210
EASS	3.055 ± 0.154	1.566 ± 0.102	0.009 ± 0.075	-0.006 ± 0.010	0.007 ± 0.005
Euc. Emcee	3.100 ± 0.240	1.835 ± 0.193	0.023 ± 0.007	0.019 ± 0.008	0.374 ± 0.076
Euc. ADS	3.085 ± 0.046	1.580 ± 0.045	0.080 ± 0.023	-0.027 ± 0.003	0.097 ± 0.004

Figure 16: Histogram and density plots for LIGO likelihoods for the angular location of the BH Merger. The true merger position is $\theta = \pi/2, \phi = \pi$.



(a) *Histogram of walker weights.* PTSS and ADS are more heavily distributed toward lower weights. EASS tends to over accept trial steps.

(b) *Likelihood density plot.* The spherical PTSS method has well sampled from the region of highest likelihood (around $(\theta = \pi/2, \phi = \pi)$). Emcee has not yet converged on the highest density, and shows a high density region at $(\phi \sim 5.4, \theta \sim 0.8)$.

8 Conclusions

The Adaptive Directional Sampling algorithm was presented and compared against traditional samplers such as Metropolis Hastings and emcee methods. Computational tests on a range of probability distributions assessed the accuracy and efficiency of likelihood reconstruction. The results showed that overall, ADS was the fastest ensemble sampler to converge onto the target likelihood in most unbounded Euclidean parameter spaces. Emcee outperformed alternative samplers post-convergence, and outperformed ADS when sampling distributions of high dimensionality. MH remains effective if an appropriate proposal distribution can be found. Depending on the target distribution, it outperforms affine-invariant alternatives.

Adaptations of ensemble methods to likelihoods on a 2-sphere and toroidal parameter space were introduced. Numerical tests on trial distributions and simulated LIGO likelihoods illustrated the strengths of such methods compared against their Euclidean alternatives. ADS adaptations using the principles of parallel transport along geodesics were shown to most effectively sample target likelihoods. Non-Euclidean adaptations of emcee were not as effective showing that samplers which work well on Euclidean spaces may not necessarily work as well in non-Euclidean geometries; the geometry of a likelihood space must be considered when selecting a effective sampling algorithm.

The results of this investigation demonstrate that ADS is extremely effective at sampling non-Euclidean spaces when compared to other alternative samplers. It hold promise for significantly improving the estimation of physical parameters in astrophysical applications.

Acknowledgements

I am exceptionally grateful to Dr. Will Handley for his invaluable guidance and support throughout this project.

References

- [1] B. P. Abbott et al. “Properties of the Binary Black Hole Merger GW150914”. In: *Physical Review Letters* 116.24 (2016), pp. 1–19. arXiv: 1602.03840.
- [2] N. Aghanim et al. “Planck 2015 results. XX. Constraints on inflation”. In: *Astronomy & Astrophysics* 10 (2015), pp. 1–28. arXiv: arXiv:1502.01594v1.
- [3] R. Abbott et al. “GW151226: Observation of Gravitational Waves from a 22-Solar-Mass Binary Black Hole Coalescence”. In: *Physical Review Letters* 116.24 (2016), pp. 241103–1,251103–14. doi: 10.1103/PhysRevLett.116.241103.
- [4] J. Veitch et al. “Parameter estimation for compact binaries with ground-based gravitational-wave observations using the LALInference software library”. In: *Physical Review D - Particles, Fields, Gravitation and Cosmology* 91.4 (2015). arXiv: 1409.7215.
- [5] Marc Van Der Sluys et al. “Parameter estimation of spinning binary inspirals using Markov-chain Monte Carlo”. In: (2008). arXiv: arXiv:0805.1689v1.
- [6] Christian Röver, Renate Meyer, and Nelson Christensen. “Bayesian inference on compact binary inspiral gravitational radiation signals in interferometric data”. In: (2006). arXiv: 0602067v2 [arXiv:gr-qc].
- [7] Michael Paul Hobson, George P Efstathiou, and Anthony N Lasenby. *General relativity: an introduction for physicists*. Cambridge University Press, 2006.
- [8] Kostas D Kokkotas. “Gravitational wave physics”. In: *Encyclopedia of Physical Science and Technology, 3rd Ed*, Vol. 7 7 (2002), pp. 67–85.
- [9] Toshifumi Futamase and Yousuke Itoh. “The post-Newtonian approximation for relativistic compact binaries”. In: *Living Reviews in Relativity* 10.1 (2007), p. 2.
- [10] Luc Blanchet et al. “Gravitational waveforms from inspiralling compact binaries to second-post-Newtonian order”. In: *Classical and Quantum Gravity* 13.4 (1996), p. 575.
- [11] Luc Blanchet. “Gravitational radiation from post-Newtonian sources and inspiralling compact binaries”. In: *Living Reviews in Relativity* 9.1 (2006), p. 4.
- [12] D. S. Sivia and J. (John) Skilling. *Data analysis : a Bayesian tutorial*. Oxford University Press, 2006.
- [13] Wendy L Martinez et al. *Computational StatisticsHandbook with MATLAB*. Chapman & Hall, 2002.
- [14] Christophe Andrieu. “An Introduction to MCMC for Machine Learning”. In: *Science* (2003), pp. 5–43. doi: 10.1023/A:1020281327116.
- [15] S. Chib and E. Greenberg. “Understanding the Metropolis-Hastings algorithm”. In: *The American Statistician* 49.4 (1995), pp. 327–335. arXiv: 1504.01896.
- [16] Jonathan Goodman and Jonathan Weare. “Ensemble Samplers with Affine Invariance”. In: *Applied Mathematical and Computational Science* 5.1 (2010), pp. 65–80.
- [17] Daniel Foreman-Mackey et al. “emcee : The MCMC Hammer”. In: *Publications of the Astronomical Society of the Pacific* 125.925 (2013), pp. 306–312. arXiv: 1202.3665.
- [18] W.R. Gilks, G.O. Roberts, and E.I. George. “Adaptive Direction Sampling”. In: *Journal of the Royal Statistical Society. Series D (The Statistician)* 43.1 (1994), pp. 179–189.
- [19] G.A. Pavliotis. “Applied Stochastic Processes”. In: *Imperial College London* (2009), pp. 1–140.