# Searches for Gravitational Waves from Supermassive Black Hole Binaries using Hamiltonian Monte Carlo

Gabriel E. Freedman
*Department of Physics, University of Wisconsin-Milwaukee*
*Milwaukee, WI, USA*

## Abstract

Pulsar timing arrays (PTAs) can detect low-frequency gravitational waves by looking for correlated deviations in pulse arrival times. Current Bayesian searches using PTAs are hampered by the large number of parameters needed to be sampled concurrently with Markov Chain Monte Carlo methods. As the data span increases, this problem will only worsen. An alternative Monte Carlo sampling method, Hamiltonian Monte Carlo (HMC), utilizes Hamiltonian dynamics to produce sample proposals informed by first-order gradients of the model likelihood. This in turn allows it to converge faster to high dimensional distributions. We implement HMC as an alternative sampling method in our search for an isotropic stochastic gravitational wave background, and present the accuracy and efficiency of the algorithm for this analysis. We also discuss implications of tailoring this algorithm to additional gravitational wave searches.

## 1. Introduction

Pulsar timing arrays (PTAs; Sazhin 1978; Detweiler 1979; Foster & Backer 1990) seek to detect low-frequency gravitational waves (GWs) by looking for spatial correlations induced in the times of arrival (TOAs) pulses from millisecond pulsars. PTAs are most sensitive in the nanohertz frequency regime ($\sim$1–100 nHz), where the dominant source of GWs is expected to be a stochastic gravitational wave background (GWB) originating from a cosmic population of supermassive black hole binaries (SMBHBs; Sesana et al. 2005). The North American Nanohertz Observatory for Gravitational Waves (NANOGrav; Ransom et al. 2019) has been collecting pulsar TOA data since 2004.

Constraining the GWB shape and strength prove scientifically important on their own, as any information or detection of low-frequency GWs provides a valuable tool for studying parts of the dynamical universe not accessible through electromagnetic observations. Analyzing the GWB can also provide useful constraints on properties of the SMBHB population including the black hole-host galaxy scaling relations (Ravi et al. 2015; Sesana 2013) and the astrophysical environment of SMBHBs emitting GWs (Quinlan 1996). The data can also be used to model more speculative sources of the GWB such as primordial GWs from inflation (Grishchuk 1976; Lasky et al. 2016) and networks of cosmic strings (Siemens et al. 2007; Blanco-Pillado et al. 2014).

GW signals can be extracted as a common-process signal from pulsar timing data only after subtracting the pulsar's timing model and accounting for underlying sources of noise in both the pulsar and observing instruments. These analyses are frequently done using Bayesian techniques to be outlined in Sec. 2. In order to perform the Bayesian searches, NANOGrav makes use of the parallel-tempering Markov Chain Monte Carlo (MCMC) code `PTMCMCSampler` (Ellis & van Haasteren 2017).

MCMC methods work adequately for a large portion of statistical models, but simple MCMC algorithms such as random-walk Metropolis (Metropolis et al. 1953) or Gibbs sampling (Geman & Geman 1984) become considerably slow as the size and complexity of the model grows and takes considerably longer to converge. Both of the aforementioned methods use random-walk proposals to generate samples and explore the parameter space, which tends to be increasingly inefficient when the target distribution includes various correlations among the parameters (Neal 2011). Hamiltonian Monte Carlo (HMC; Duane et al. 1987; Neal 2011) removes the requirement to sample the model randomly, and replaces it with a simulation of Hamiltonian dynamics on the distribution itself. This scheme allows samples to be drawn at much further distances from one another, and explore the full parameter space in a

more efficient way. For a target distribution of dimension $d$, the cost of drawing an independent sample with HMC goes roughly as $O(d^{5/4})$, compared to $O(d^2)$ for random-walk Metropolis (Creutz 1988). The No-U-Turn Sampler (NUTS; Hoffman & Gelman 2011) algorithm provides a basis for performing analysis with HMC without pre-tuning the sampling.

This paper is organized as follows. In Sec. 2 we describe the signal model used in our analysis and summarize the HMC and NUTS algorithms. In Sec. 3 we present the results of a stochastic GWB search using HMC. In Sec. 4 we discuss these results and future directions of this research.

## 2. Methods

Many of the GW analyses carried out on PTAs employ Bayesian inference. It is a method of statistical inference in which one's knowledge of an event is updated with subsequent observations. It is rooted in the theory of Bayesian statistics, which defines the term "probability" as a degree of belief in an event. This stands in contrast to the frequentist definition of probability as a limit of the relative frequency of an event. At the heart of Bayesian inference is Bayes rule of conditional probabilities:

$$\mathcal{P}\left(\Theta|\mathcal{D},\mathcal{H}\right) = \frac{\mathcal{P}\left(\mathcal{D}|\Theta,\mathcal{H}\right)\mathcal{P}\left(\Theta|\mathcal{H}\right)}{\mathcal{P}\left(\mathcal{D}|\mathcal{H}\right)}, \tag{1}$$

where $\Theta$ denotes our model parameters, $\mathcal{D}$ our observational data, and $\mathcal{H}$ a particular choice of model. Our output is the posterior probability distribution of the parameters, $\mathcal{P}\left(\Theta|\mathcal{D},\mathcal{H}\right)$. We input the prior probability distribution $\mathcal{P}\left(\Theta|\mathcal{H}\right)$, the likelihood function $\mathcal{P}\left(\mathcal{D}|\Theta,\mathcal{H}\right)$, and the evidence $\mathcal{P}\left(\mathcal{D}|\mathcal{H}\right)$. The model evidence acts as a normalizing factor, and for the purpose of parameter estimation it is typically ignored. The likelihood is a function of the target parameters and describes the joint probability of the observed data.

We describe the data acquisition and preprocessing in Sec. 2.1. We define the likelihood function used for PTA analyses of the GWB in Sec. 2.2. We then outline the HMC algorithm in Sec. 2.3, and describe the No-U-Turn Sampler extensions to HMC in Sec. 2.4.

**2.1. Time of Arrival Data** Pulsars are observed using the 305-m Arecibo Observatory (AO) and 100-m Green Bank Telescope (GBT). At AO, observations are made using four receivers at bandwidths 327 MHz, 430 MHz, 1400 MHz, and 2100 MHz. At the GBT, two receivers are used: 800 MHz and 1400 MHz. The observing cadence for the set of pulsars is about once a month, which corresponds to a single observing "epoch". The data acquisition systems (Demorest 2007; DuPlain et al. 2008; Ford et al. 2010) take the incoming baseband data and perform coherent dedispersion, RFI excision, flux calibration, and polarization calibration. They return pulse profiles for the various frequency channels.

TOAs are generated from these pulse profiles using the `PSRCHIVE` package, which utilizes the Fourier domain algorithm of Taylor (1992) to move from frequency profiles to the time domain. The processed TOAs for each pulsar are then fit to a timing model that comprises parameters describing the specific pulsar such as sky locaiton, parallax, spin period and spin period derivative. For binary pulsars, this timing model also includes five Keplerian binary parameters. The timing model fits are done using the `TEMPO2` and `PINT` packages.

**2.2. PTA Likelihood** We now discuss the PTA likelihood function. Following the outline provided in Arzoumanian et al. (2016), we start by considering a single pulsar and its timing residual vector $\delta\mathbf{t}$ with length equal to the number of TOAs in our dataset, $N_{\rm TOA}$. This timing residual data can be decompose into individual components:

$$\delta\mathbf{t} = M\boldsymbol{\epsilon} + F\mathbf{a} + U\mathbf{j} + \mathbf{n}. \tag{2}$$

Each term describes a different inaccuracy or source of noise that contributes to the residual data. The term $M\boldsymbol{\epsilon}$ represents inaccuracies stemming from the subtraction of the pulsar's timing model, with $M$ the timing model design

matrix, and $\boldsymbol{\epsilon}$ the vector of timing model parameter offsets. $F\mathbf{a}$ encompasses effects due to low-frequency ("red") noise. We choose to define this in a rank-reduced basis where $F$ represents our matrix of basis functions, in this case alternating sine and cosine functions, and $\mathbf{a}$ a set of Fourier coefficients. $U\mathbf{j}$ describes noise that is completely uncorrelated in time but completely correlated across observations of a similar epoch. $U$ is the matrix that maps between $N_{\rm TOA}$ residual data and $N_{\rm epoch}$ observation sessions, and $\mathbf{j}$ accounts for the correlated noise in each epoch. The final term, $\mathbf{n}$, includes any other high-frequency ("white") noise that cannot be accounted for in the previous terms.

Previous Bayesian analysis schemes (van Haasteren et al. 2009; van Haasteren & Levin 2010; van Haasteren et al. 2011; Ellis et al. 2012; Ellis 2013; Ellis et al. 2013) have described the white noise with EFAC (constant multiplier to TOA uncertainties) and EQUAD (white noise added in quadrature to EFAC) parameters and employed a power-law model to describe the red noise. The sum of these white noise covariances we describe via a matrix $N$. The parameters describing $\boldsymbol{\epsilon}$, $\mathbf{a}$, and $\mathbf{j}$ we group as follows:

$$T = \begin{bmatrix} M & F & U \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} \boldsymbol{\epsilon} \\ \mathbf{a} \\ \mathbf{j} \end{bmatrix}. \tag{3}$$

We place a Gaussian prior on these parameters with covariance:

$$B = \begin{bmatrix} \infty & 0 & 0 \\ 0 & \varphi & 0 \\ 0 & 0 & \mathcal{J} \end{bmatrix}, \tag{4}$$

where $\infty$ represents a diagonal matrix of infinities corresponding to unconstrained uniform priors on all timing model parameters. The parameters that describe $\mathcal{J}$ we refer to as ECORR and correspond to the epoch-correlated white noise signals per receiving backend. The matrix $\varphi$ defines the parameters involving red noise signals, which includes low-frequency noise intrinsic to each pulsar, as well as the stochastic GWB. For this paper, we performed our analysis by modeling the GWB using a fiducial power-law spectrum of the characteristic GW strain and cross-power spectral density:

$$h_c(f) = A_{\rm gw}\left(\frac{f}{f_{yr}}\right)^{\alpha}, \qquad S_{ab}(f) = \Gamma_{ab}\frac{A_{\rm gw}^2}{12\pi^2}\left(\frac{f}{f_{\rm yr}}\right)^{-\gamma} f_{\rm yr}^{-3}, \tag{5}$$

where $\gamma = 3 - 2\alpha$. For a background generated by the GW emission from the evolution of a population of inspiraling SMBHBs in circular orbits, we have $\alpha = -2/3$, which implies $\gamma = 13/3$ (Phinney 2001). The function $\Gamma_{ab}$ is called the overlap reduction function (ORF) and describes the average correlations between any two pulsars $a$ and $b$ as a function of their angular separation. For an isotropic, stochastic GWB, this ORF is given by Hellings & Downs (1983).

We analytically marginalize over the timing model parameters to reduce the overall dimensionality of our posterior (Lentati et al. 2013; van Haasteren & Vallisneri 2014) and are left with the form of the likelihood that is used for the analysis in this paper:

$$\mathcal{P}(\delta\mathbf{t}|\phi) \equiv L(\phi) = \frac{\exp\left(-\frac{1}{2}\delta\mathbf{t}^T C^{-1}\delta\mathbf{t}\right)}{\sqrt{\det 2\pi C}}, \tag{6}$$

where $C = N + TBT^T$.

**2.3. Hamiltonian Monte Carlo** We now provide a description of the HMC algorithm. In HMC (Duane et al. 1987; Neal 2011), we start by introducing an auxiliary momentum variable $p_i$ alongside each target parameter $q_i$. In most implementations, the momenta are chosen to be independent of the $q_i$ and follow a zero-mean Gaussian distribution, with a covariance matrix $M$ that is typically taken to be the identity. The log of the joint density of $\mathbf{p}$ and $\mathbf{q}$ defines our Hamiltonian:

$$H\left(\mathbf{p},\mathbf{q}\right) = U\left(\mathbf{q}\right) + K\left(\mathbf{p}\right) = -\mathcal{L}(\mathbf{q}) + \frac{1}{2}\mathbf{p}^T M^{-1}\mathbf{p}, \tag{7}$$

where $\mathcal{L}(\mathbf{q}) \equiv \log L(\mathbf{q})$ is the log of the likelihood function for the distribution of our target parameters $\mathbf{q}$. Analogous to Hamiltonian dynamics, we have a potential energy term $U(\mathbf{q})$ and a kinetic energy term $K(\mathbf{p})$. We then simulate the evolution of this system over time according to Hamilton's equations:

$$\frac{dq_i}{dt} = \frac{\partial H}{\partial p_i}, \qquad \frac{dp_i}{dt} = -\frac{\partial H}{\partial q_i}. \tag{8}$$

This can be solved numerically using a symplectic integrator such as a "leapfrop" method, which for an integration step size $\epsilon$ uses an update scheme:

$$\mathbf{p}^{t+\epsilon/2} = p^t + \left(\frac{\epsilon}{2}\right)\nabla_{\mathbf{q}}\mathcal{L}(\mathbf{q}^t), \quad \mathbf{q}^{t+\epsilon} = \mathbf{q}^t + \epsilon\mathbf{p}^{t+\epsilon/2}, \quad \mathbf{p}^{t+\epsilon} = p^{t+\epsilon/2} + \left(\frac{\epsilon}{2}\right)\nabla_{\mathbf{q}}\mathcal{L}(\mathbf{q}^{t+\epsilon}), \tag{9}$$

where superscripts denote the time at which the particular quantity is evaluated. The standard method for producing a chain of samples using HMC then proceeds as follows: We first resample our momenta distribution. Then for a set number of leapfrog steps $L$, we use Eq. (9) to evolve our system through time and propose some final position and momentum vectors $\tilde{\mathbf{q}}$ and $\tilde{\mathbf{p}}$. This proposal is accepted or rejected according to the Metropolis algorithm (Metropolis et al. 1953).

**2.4. No-U-Turn Sampler** The performance of the HMC algorithm is particularly sensitive to two user-defined parameters, the number of leapfrog steps $L$ and integration step size $\epsilon$, defined in the above section. If these parameters are not properly tuned, the algorithm may waste computation time or begin to exhibit unwanted random walk behavior and in some cases may not even be ergodic (Neal 2011). In general, tuning these parameters appropriately would require multiple preliminary runs.

The No-U-Turn Sampler (NUTS; Hoffman & Gelman 2011) offers an extension to the HMC algorithm that dynamically tunes the number of leapfrog steps $L$. NUTS uses a recursive doubling algorithm, similar to the one outlined in Neal (2000), to determine when the generated proposal trajectory begins to double back on itself, or make a "U-turn". The algorithm builds a binary tree, simulating Hamiltonian dynamics forwards and backwards randomly in time for $2^j$ steps, with $j$ the height of the full tree. If we define $\mathbf{q}^+$, $\mathbf{p}^+$ and $\mathbf{q}^-$, $\mathbf{p}^-$ as the position-momenta pairs of the left- and rightmost nodes of the bottom subtree, then the stopping condition for NUTS can be written as:

$$\left(\mathbf{q}^+ - \mathbf{q}^-\right)\cdot\mathbf{p}^- < 0 \quad \text{or} \quad \left(\mathbf{q}^+ - \mathbf{q}^-\right)\cdot\mathbf{p}^+ < 0. \tag{10}$$

The above procedure adaptively tunes the parameter $L$ for each iteration in the chain. The step size parameter $\epsilon$ in NUTS is set using the method of stochastic optimization with varying adaptation (Andrieu & Thoms 2008). In particular, Hoffman & Gelman utilize the primal-dual averaging algorithm proposed by Nesterov (2009). With $L$ and $\epsilon$ automatically tuned, NUTS can be run without any human intervention.

## 3. Results

We now present the results of a stochastic GWB search conducted using HMC via the NUTS algorithm, and compare against a similar search using traditional MCMC methods. In order to utilize HMC, we first must compute some form of the gradient $\nabla_{\phi_i}\mathcal{L}(\phi_i)$ of our PTA likelihood defined in Eq. (6). For complex models, numerical methods such as automatic differentiation (Griewank & Walther 2000) can make this task less burdensome on the user, but at the steep cost of computation speed. Analytic gradients are therefore preferred for optimized sampling. Taking the gradient with respect to all hyperparameters $\phi_i$ of our log likelihood gives:

$$\nabla_{\phi_i}\mathcal{L}(\phi_i) = -\frac{1}{2}\left[\mathbf{r}^T C^{-1}\frac{\partial \mathbf{r}}{\partial \phi_i} - \mathbf{r}^T C^{-1}\frac{\partial C}{\partial \phi_i}C^{-1}\mathbf{r} + \frac{\partial \mathbf{r}^T}{\partial \phi_i}C^{-1}\mathbf{r}\right] - \frac{1}{2}\mathrm{Tr}\left(C^{-1}\frac{\partial C}{\partial \phi_i}\right). \tag{11}$$

These derivatives can be computed exactly for all parameters defined in Sec. 2.2 for a stochastic GWB analysis. We note, however, that this is not true in general. Any additional signals added to our model are not guaranteed to have an analytic derivative of the likelihood. We make modifications to the PTA analysis code `enterprise` (Ellis et al. 2019) to accommodate this gradient, and use the NUTS algorithm provided in the code `piccard` (Vallisneri & van Haasteren 2017).

We run a stochastic GWB search using the NANOGrav 5-year dataset (Demorest et al. 2013) with both the MCMC and HMC algorithms. The dataset includes 17 pulsars. The intrinsic white noise parameters are fixed to their maximum of the posterior probability distribution gotten from individual pulsar noise analyses. This leaves 36 free parameters in our model (2 intrinsic red noise parameters per pulsar, plus 2 describing the background as a power law). The resulting posteriors for the background amplitude $A_{\rm gw}$ and spectral index $\gamma_{\rm gw}$ are shown in Fig. 1. The MCMC and HMC sampling procedures return similar results, with near identical 95% upper limits on the amplitude parameter ($A_{\rm gw,mcmc}, A_{\rm gw,hmc} < 3.7\times10^{-14}$).

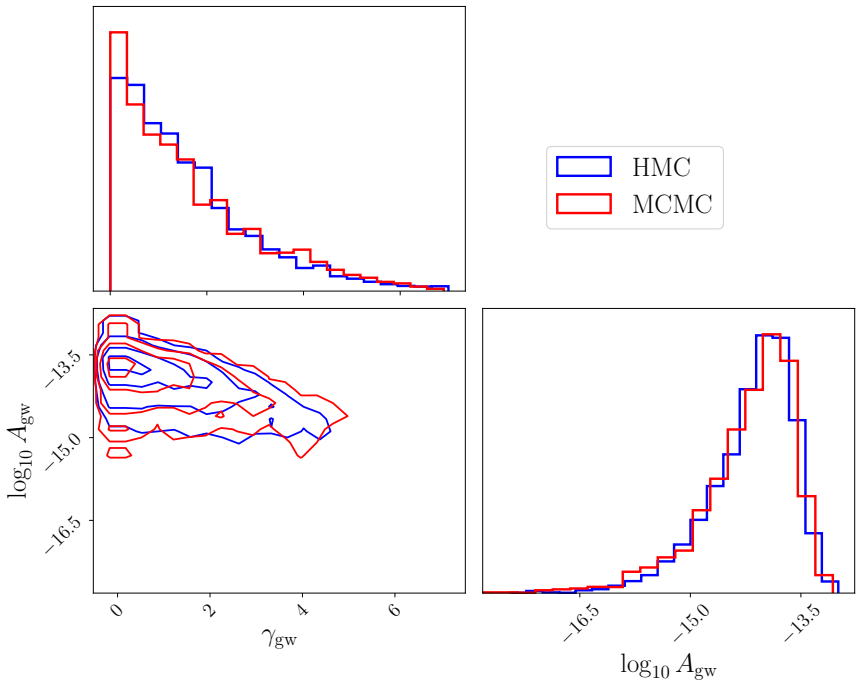


Figure 1: Bayesian posteriors for the amplitude $A_{\rm gw}$ and spectral index $\gamma_{\rm gw}$ of a common-process signal run using either MCMC or HMC as the primary sampling method. We observe that the two procedures produce similar results, with 95% upper limits on $A_{\rm gw}$ of $A_{\rm gw,mcmc} < 3.65\times10^{-14}$ and $A_{\rm gw,hmc} < 3.61\times10^{-14}$.
The MCMC routine was run for a total $M = 1,000,000$ samples, and the HMC routine was run for only $M = 4,000$.

The convergence of the two sets of Markov chains was confirmed using the Gelman-Rubin (Gelman & Rubin 1992) and Geweke (Geweke 1992) tests, indicating that reduction in sample number of 250x from MCMC was HMC is justifiable. We further compare the efficiency of HMC sampling by looking at the autocorrelation lengths of the two sets of chains, or measuring how far one must jump through the chain to find the next statistically significant sample. A truncated summary of the computed autocorrelation lengths for both sampling techniques is shown in Table 1, focusing on the two GWB parameters as well as a pair of individual pulsar red noise parameters for comparison. The autocorrelation lengths for the HMC chains are nearly all of order $10^0$, implying that every generated sample is statistically significant and no computation is wasted. This is in agreement with the principle of the HMC algorithm: to take larger, more-informed steps to avoid random walk-like behavior.

| Parameter | MCMC | HMC |
|---|---|---|
| J1713+0747 $\log_{10} A$ | 205.70 | 1.66 |
| J1713+0747 $\gamma$ | 121.16 | 0.99 |
| J1909-3744 $\log_{10} A$ | 258.51 | 1.37 |
| J1909-3744 $\gamma$ | 207.87 | 1.00 |
| $\log_{10} A_{gw}$ | 229.60 | 2.04 |
| $\gamma_{gw}$ | 160.36 | 1.20 |

Table 1: Autocorrelation lengths for a selection of parameters for a 17-pulsar stochastic GWB analysis performed with both MCMC and HMC methods. Parameters shown include the two parameters describing the background as well as individual red noise parameters for two pulsars

This increase in efficiency for drawing HMC samples does not come for free. Computing the full gradient vector alongside the likelihood will necessarily slow down the speed of sample generation. We profile the speed of the current implementation of the gradient calculation of Eq. (11) to compare to the existing evaluation of Eq. (6). The results for the dataset used in this analysis are shown in Fig. 2. Note that the jump between pulsars 9 and 10 represent the inclusion into the dataset of pulsar J1713+0747, which has considerably more TOAs than its counterparts.

The gradient calculation is in its current state two orders of magnitude of slower than the single likelihood evaluation. Complicating the matter further is the fact that generating steps in HMC often requires multiple gradient evaluations to find a proposal a far enough distance away from the current state. We also note that prior to running HMC, we whiten the parameter vector via a coordinate transformation, which requires a one-time calculation of the likelihood Hessian matrix. This is a computationally expensive step that occurs before any sampling can take place.

## 4. Discussion

The preliminary comparison given in this paper shows consistency between the MCMC and HMC sampling techniques for performing a stochastic GWB search using PTAs. We find that HMC sampling provides orders of magnitude improvement in the efficiency of generated independent samples for a model including 17 pulsars and 36 free parameters. In coming years, PTA datasets will continue to grow in size both in the number of pulsars and the span of the data, and HMC could prove valuable in taming the added complexity.

We recognize that the positive benefits of HMC may not seem to outweigh the stark increase in runtime. The computational cost of the gradient calculation in HMC can be improved but not entirely overruled. We reiterate that one evaluation of the gradient does not equal one HMC sample. For a single step in the NUTS algorithm, the construction of the height $j$ binary tree requires $O(2^j)$ gradient evaluations, a high cost if the optimal number of leapfrog steps $L$ is large. There are methods available for faster inversion of the matrices involved in Eqs. (6) and (11) (for example, Woodbury 1950).

Aside from the increase in efficiency, HMC also finds value in sampling parameters that are strongly correlated or with high uncertainties that would otherwise fail to converge in a reasonable amount of time using MCMC methods. This benefit could be used to add to an existing model, such as adding parameters that account for differences in solar system ephemerides (Vallisneri et al. 2020). It could also be useful when performing entirely different GW analyses. One such example is searching for GWs originating from individual SMBHBs (for example, Aggarwal et al. 2019), where the model of interest now centers on parameters describing a single binary black hole system.

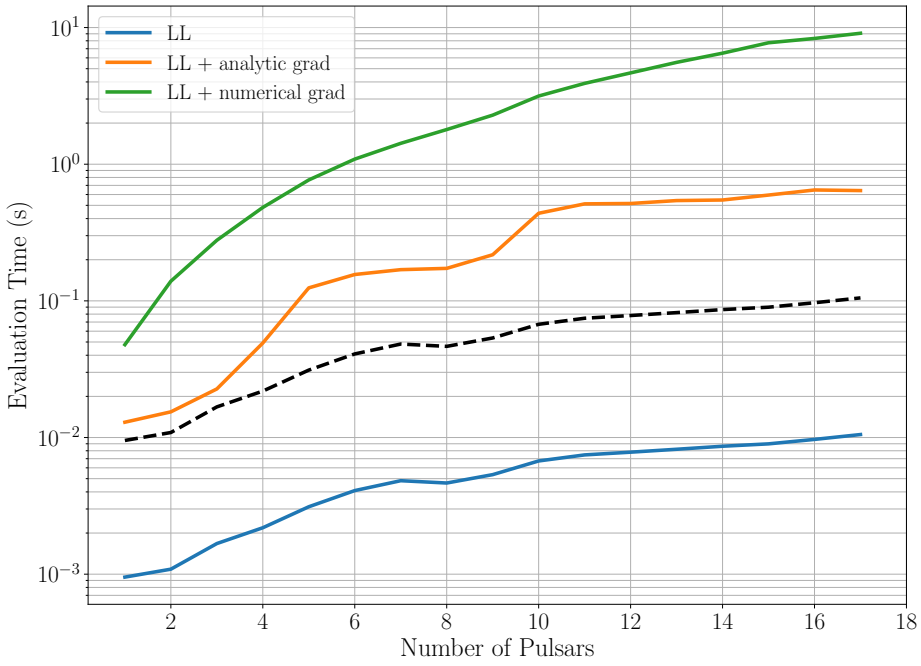


Figure 2: Evaluation times for three methods of computing Eqs. (6) and (11): the log likelihood calculation only (solid blue line), the log likelihood calculation and gradient with all derivatives calculated analytically (solid orange line), and the log likelihood with gradient calculated entirely using numerical methods (solid green line). The dashed black line represents an order of magnitude increase on the base log likelihood evaluation time. These speed profiles are plotted as a function of the number of pulsars added into the dataset. As not all pulsars have been observed for the same length of time, the bumps in the plot represent the addition of large data span pulsars into the calculations.

## Acknowledgments

We thank Sarah Vigeland for helpful discussions and for providing useful comments on the manuscript. The NANOGrav project receives support from National Science Foundation (NSF) Physics Frontiers Center award number 1430284 and 2020265. This material is based upon work supported by NASA under Award No. RFP21_3.0 issued through Wisconsin Space Grant Consortium and the National Space Grant College and Fellowship Program, and any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Aeronautics and Space Administration.

## References

Aggarwal, K; Arzoumanian, Z; Baker, PT; Brazier, A; Brinson, MR; Brook, PR; Burke-Spolaor, S; Chatterjee, S; Cordes, JM; Cornish, NJ; Crawford, F; Crowter, K; Cromartie, HT; DeCesar, M; Demorest, PB; Dolch, T; Ellis, JA; Ferdman, RD; Ferrara, E; Fonseca, E; Garver-Daniels, N; Gentile, P; Hazboun, JS; Holgado, AM; Huerta, EA; Islo, K; Jennings, R; Jones, G; Jones, ML; Kaiser, AR; Kaplan, DL; Kelley, LZ; Key, JS; Lam, MT; Lazio, TJW; Levin, L; Lorimer, DR; Luo, J; Lynch, RS; Madison, DR; McLaughlin, MA; McWilliams, ST; Mingarelli, CMF; Ng, C; Nice, DJ; Pennucci, TT; Pol, NS; Ransom, SM; Ray, PS; Siemens, X; Simon, J; Spiewak, R; Stairs, IH; Stinebring, DR; Stovall, K; Swiggum, J; Taylor, SR; Turner, JE; Vallisneri, M; van Haasteren, R; Vigeland, SJ; Witt, CA; Zhu, WW; NANOGrav Collaboration. "The NANOGrav 11 yr Data Set: Limits on Gravitational Waves from Individual Supermassive Black Hole Binaries," *ApJ*, v. 880(2), 2019, p. 116. https://ui.adsabs.harvard.edu/abs/2019ApJ...880..116A

Andrieu, C; Thoms, JA. "A tutorial on adaptive mcmc," *Statistics and Computing*, v. 18, 2008, p. 343–373

Arzoumanian, Z; Brazier, A; Burke-Spolaor, S; Chamberlin, SJ; Chatterjee, S; Christy, B; Cordes, JM; Cornish, NJ; Crowter, K; Demorest, PB; Deng, X; Dolch, T; Ellis, JA; Ferdman, RD; Fonseca, E; Garver-Daniels, N;

Gonzalez, ME; Jenet, F; Jones, G; Jones, ML; Kaspi, VM; Koop, M; Lam, MT; Lazio, TJW; Levin, L; Lommen, AN; Lorimer, DR; Luo, J; Lynch, RS; Madison, DR; McLaughlin, MA; McWilliams, ST; Mingarelli, CMF; Nice, DJ; Palliyaguru, N; Pennucci, TT; Ransom, SM; Sampson, L; Sanidas, SA; Sesana, A; Siemens, X; Simon, J; Stairs, IH; Stinebring, DR; Stovall, K; Swiggum, J; Taylor, SR; Vallisneri, M; van Haasteren, R; Wang, Y; Zhu, WW; NANOGrav Collaboration. "The NANOGrav Nine-year Data Set: Limits on the Isotropic Stochastic Gravitational Wave Background," *ApJ*, v. 821(1), 2016, p. 13. https://ui.adsabs.harvard.edu/abs/2016ApJ...821...13A

Blanco-Pillado, JJ; Olum, KD; Shlaer, B. "Number of cosmic string loops," *Phys. Rev. D*, v. 89(2), 2014, p. 023512. https://ui.adsabs.harvard.edu/abs/2014PhRvD..89b3512B

Creutz, M. "Global monte carlo algorithms for many-fermion systems," *Phys. Rev. D*, v. 38, 1988, p. 1228–1238

Demorest, PB. Ph.D. thesis, University of California, Berkeley, 2007

Demorest, PB; Ferdman, RD; Gonzalez, ME; Nice, D; Ransom, S; Stairs, IH; Arzoumanian, Z; Brazier, A; Burke-Spolaor, S; Chamberlin, SJ; Cordes, JM; Ellis, J; Finn, LS; Freire, P; Giampanis, S; Jenet, F; Kaspi, VM; Lazio, J; Lommen, AN; McLaughlin, M; Palliyaguru, N; Perrodin, D; Shannon, RM; Siemens, X; Stinebring, D; Swiggum, J; Zhu, WW. "Limits on the Stochastic Gravitational Wave Background from the North American Nanohertz Observatory for Gravitational Waves," *ApJ*, v. 762(2), 2013, p. 94. https://ui.adsabs.harvard.edu/abs/2013ApJ...762...94D

Detweiler, S. "Pulsar timing measurements and the search for gravitational waves," *ApJ*, v. 234, 1979, p. 1100–1104. https://ui.adsabs.harvard.edu/abs/1979ApJ...234.1100D

Duane, S; Kennedy, AD; Pendleton, BJ; Roweth, D. "Hybrid Monte Carlo," *Physics Letters B*, v. 195(2), 1987, p. 216–222. https://ui.adsabs.harvard.edu/abs/1987PhLB..195..216D

DuPlain, R; Ransom, S; Demorest, PB; Brandt, P; Ford, J; A., S. v. 7019 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, 2008

Ellis, J; van Haasteren, R. "jellis18/ptmcmcsampler: Official release", 2017

Ellis, JA. "A Bayesian analysis pipeline for continuous GW sources in the PTA band," *Classical and Quantum Gravity*, v. 30(22), 2013, p. 224004. https://ui.adsabs.harvard.edu/abs/2013CQGra..30v4004E

Ellis, JA; Siemens, X; Creighton, JDE. "Optimal Strategies for Continuous Gravitational Wave Detection in Pulsar Timing Arrays," *ApJ*, v. 756(2), 2012, p. 175. https://ui.adsabs.harvard.edu/abs/2012ApJ...756..175E

Ellis, JA; Siemens, X; van Haasteren, R. "An Efficient Approximation to the Likelihood for Gravitational Wave Stochastic Background Detection Using Pulsar Timing Data," *ApJ*, v. 769(1), 2013, p. 63. https://ui.adsabs.harvard.edu/abs/2013ApJ...769...63E

Ellis, JA; Vallisneri, M; Taylor, SR; Baker, PT. "ENTERPRISE: Enhanced Numerical Toolbox Enabling a Robust PulsaR Inference SuitE", 2019. https://ui.adsabs.harvard.edu/abs/2019ascl.soft12015E

Ford, JM; Demorest, PB; Ransom, S. v. 7740 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, 2010

Foster, RS; Backer, DC. "Constructing a Pulsar Timing Array," *ApJ*, v. 361, 1990, p. 300. https://ui.adsabs.harvard.edu/abs/1990ApJ...361..300F

Gelman, A; Rubin, DB. "Inference from iterative simulation using multiple sequences," *Statistical Science*, v. 7(4), 1992, p. 457–472

Geman, S; Geman, D. "Geman, d.: Stochastic relaxation, gibbs distribution, and the bayesian restoration of images. ieee trans. pattern anal. mach. intell. pami-6(6), 721-741," *IEEE Trans. Pattern Anal. Mach. Intell.*, v. 6, 1984, p. 721–741

Geweke, J. "Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments," In *IN BAYESIAN STATISTICS*, University Press, 1992, p. 169–193

Griewank, A; Walther, A. *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation*, SIAM, v. 19, 2000

Grishchuk, LP. "Primordial gravitons and possibility of their observation," *Soviet Journal of Experimental and Theoretical Physics Letters*, v. 23, 1976, p. 293. https://ui.adsabs.harvard.edu/abs/1976JETPL..23..293G

Hellings, RW; Downs, GS. "Upper limits on the isotropic gravitational radiation background from pulsar timing analysis.," *ApJ*, v. 265, 1983, p. L39–L42. https://ui.adsabs.harvard.edu/abs/1983ApJ...265L..39H

Hoffman, MD; Gelman, A. "The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo," *arXiv e-prints*, 2011, p. arXiv:1111.4246. https://ui.adsabs.harvard.edu/abs/2011arXiv1111.4246H

Lasky, PD; Mingarelli, CMF; Smith, TL; Giblin, JT; Thrane, E; Reardon, DJ; Caldwell, R; Bailes, M; Bhat, NDR; Burke-Spolaor, S; Dai, S; Dempsey, J; Hobbs, G; Kerr, M; Levin, Y; Manchester, RN; Osłowski, S; Ravi, V; Rosado, PA; Shannon, RM; Spiewak, R; van Straten, W; Toomey, L; Wang, J; Wen, L; You, X; Zhu, X. "Gravitational-Wave Cosmology across 29 Decades in Frequency," *Physical Review X*, v. 6(1), 2016, p. 011035. https://ui.adsabs.harvard.edu/abs/2016PhRvX...6a1035L

Lentati, L; Alexander, P; Hobson, MP; Taylor, S; Gair, J; Balan, ST; van Haasteren, R. "Hyper-efficient model-independent Bayesian method for the analysis of pulsar timing data," *Phys. Rev. D*, v. 87(10), 2013, p. 104021. https://ui.adsabs.harvard.edu/abs/2013PhRvD..87j4021L

Metropolis, N; Rosenbluth, AW; Rosenbluth, MN; Teller, AH; Teller, E. "Equation of State Calculations by Fast Computing Machines," *J. Chem. Phys.*, v. 21(6), 1953, p. 1087–1092. https://ui.adsabs.harvard.edu/abs/1953JChPh..21.1087M

Neal, R. *MCMC Using Hamiltonian Dynamics*, CRC Press, chap. 5, 2011, p. 113–162. https://ui.adsabs.harvard.edu/abs/2011hmcm.book..113N

Neal, RM. "Slice Sampling," *arXiv e-prints*, 2000, p. physics/0009028. https://ui.adsabs.harvard.edu/abs/2000physics...9028N

Nesterov, Y. "Primal-dual subgradient methods for convex problems," *Mathematical Programming*, v. 120, 2009, p. 221–259

Phinney, ES. "A Practical Theorem on Gravitational Wave Backgrounds," *arXiv e-prints*, 2001, p. astro–ph/0108028. https://ui.adsabs.harvard.edu/abs/2001astro.ph..8028P

Quinlan, GD. "The dynamical evolution of massive black hole binaries I. Hardening in a fixed stellar background," *New Ast.*, v. 1(1), 1996, p. 35–56. https://ui.adsabs.harvard.edu/abs/1996NewA....1...35Q

Ransom, S; Brazier, A; Chatterjee, S; Cohen, T; Cordes, JM; DeCesar, ME; Demorest, PB; Hazboun, JS; Lam, MT; Lynch, RS; McLaughlin, MA; Ransom, SM; Siemens, X; Taylor, SR; Vigeland, SJ. "The NANOGrav Program for Gravitational Waves and Fundamental Physics," In *Bulletin of the American Astronomical Society*, v. 51, 2019, p. 195. https://ui.adsabs.harvard.edu/abs/2019BAAS...51g.195R

Ravi, V; Wyithe, JSB; Shannon, RM; Hobbs, G. "Prospects for gravitational-wave detection and supermassive black hole astrophysics with pulsar timing arrays," *MNRAS*, v. 447(3), 2015, p. 2772–2783. https://ui.adsabs.harvard.edu/abs/2015MNRAS.447.2772R

Sazhin, MV. "Opportunities for detecting ultralong gravitational waves," *Soviet Ast.*, v. 22, 1978, p. 36–38. https://ui.adsabs.harvard.edu/abs/1978SvA....22...36S

Sesana, A. "Systematic investigation of the expected gravitational wave signal from supermassive black hole binaries in the pulsar timing band.," *MNRAS*, v. 433, 2013, p. L1–L5. https://ui.adsabs.harvard.edu/abs/2013MNRAS.433L...1S

Sesana, A; Haardt, F; Madau, P; Volonteri, M. "The Gravitational Wave Signal from Massive Black Hole Binaries and Its Contribution to the LISA Data Stream," *ApJ*, v. 623(1), 2005, p. 23–30. https://ui.adsabs.harvard.edu/abs/2005ApJ...623...23S

Siemens, X; Mandic, V; Creighton, J. "Gravitational-Wave Stochastic Background from Cosmic Strings," *PhRvL*, v. 98(11), 2007, p. 111101. https://ui.adsabs.harvard.edu/abs/2007PhRvL..98k1101S

Taylor, JH. "Pulsar Timing and Relativistic Gravity," *Philosophical Transactions of the Royal Society of London Series A*, v. 341(1660), 1992, p. 117–134. https://ui.adsabs.harvard.edu/abs/1992RSPTA.341..117T

Vallisneri, M; Taylor, SR; Simon, J; Folkner, WM; Park, RS; Cutler, C; Ellis, JA; Lazio, TJW; Vigeland, SJ; Aggarwal, K; Arzoumanian, Z; Baker, PT; Brazier, A; Brook, PR; Burke-Spolaor, S; Chatterjee, S; Cordes, JM; Cornish, NJ; Crawford, F; Cromartie, HT; Crowter, K; DeCesar, M; Demorest, PB; Dolch, T; Ferdman, RD; Ferrara, EC; Fonseca, E; Garver-Daniels, N; Gentile, P; Good, D; Hazboun, JS; Holgado, AM; Huerta, EA; Islo, K; Jennings, R; Jones, G; Jones, ML; Kaplan, DL; Kelley, LZ; Key, JS; Lam, MT; Levin, L; Lorimer, DR; Luo, J; Lynch, RS; Madison, DR; McLaughlin, MA; McWilliams, ST; Mingarelli, CMF; Ng, C; Nice, DJ; Pennucci, TT; Pol, NS; Ransom, SM; Ray, PS; Siemens, X; Spiewak, R; Stairs, IH; Stinebring, DR; Stovall, K; Swiggum, JK; van Haasteren, R; Witt, CA; Zhu, WW. "Modeling the Uncertainties of Solar System Ephemerides for Robust Gravitational-wave Searches with Pulsar-timing Arrays," *ApJ*, v. 893(2), 2020, p. 112. https://ui.adsabs.harvard.edu/abs/2020ApJ...893..112V

Vallisneri, M; van Haasteren, R. "Taming outliers in pulsar-timing data sets with hierarchical likelihoods and Hamiltonian sampling," *MNRAS*, v. 466(4), 2017, p. 4954–4959. https://ui.adsabs.harvard.edu/abs/2017MNRAS.466.4954V

van Haasteren, R; Levin, Y. "Gravitational-wave memory and pulsar timing arrays," *MNRAS*, v. 401(4), 2010, p. 2372–2378. https://ui.adsabs.harvard.edu/abs/2010MNRAS.401.2372V

van Haasteren, R; Levin, Y; Janssen, GH; Lazaridis, K; Kramer, M; Stappers, BW; Desvignes, G; Purver, MB; Lyne, AG; Ferdman, RD; Jessner, A; Cognard, I; Theureau, G; D'Amico, N; Possenti, A; Burgay, M; Corongiu, A; Hessels, JWT; Smits, R; Verbiest, JPW. "Placing limits on the stochastic gravitational-wave background using European Pulsar Timing Array data," *Monthly Notices of the Royal Astronomical Society*, v. 414(4), 2011, p. 3117–3128

van Haasteren, R; Levin, Y; McDonald, P; Lu, T. "On measuring the gravitational-wave background using Pulsar Timing Arrays," *MNRAS*, v. 395(2), 2009, p. 1005–1014. https://ui.adsabs.harvard.edu/abs/2009MNRAS.395.1005V

van Haasteren, R; Vallisneri, M. "New advances in the Gaussian-process approach to pulsar-timing data analysis," *Phys. Rev. D*, v. 90(10), 2014, p. 104012. https://ui.adsabs.harvard.edu/abs/2014PhRvD..90j4012V

Woodbury, MA. "Inverting modified matrices," In *Memorandum Rept. 42, Statistical Research Group*, Princeton Univ., 1950, p. 4