

Accelerating pulsar timing data analysis

Rutger van Haasteren[★]

Max-Planck-Institut für Gravitationsphysik (Albert-Einstein-Institut), D-30167 Hannover, Germany

Accepted 2012 October 29. Received 2012 October 22; in original form 2012 October 1

ABSTRACT

The analysis of pulsar timing data, especially in pulsar timing array (PTA) projects, has encountered practical difficulties: evaluating the likelihood and/or correlation-based statistics can become prohibitively computationally expensive for large data sets. In situations where a stochastic signal of interest has a power spectral density that dominates the noise in a limited bandwidth of the total frequency domain (e.g. the isotropic background of gravitational waves), a linear transformation exists that transforms the timing residuals to a basis in which virtually all the information about the stochastic signal of interest is contained in a small fraction of basis vectors. By only considering such a small subset of these ‘generalized residuals’, the dimensionality of the data analysis problem is greatly reduced, which can cause a large speedup in the evaluation of the likelihood: the ABC-method (Acceleration By Compression). The compression fidelity, calculable with crude estimates of the signal and noise, can be used to determine how far a data set can be compressed without significant loss of information. Both direct tests on the likelihood, and Bayesian analysis of mock data, show that the signal can be recovered as well as with an analysis of uncompressed data. In the analysis of International PTA Mock Data Challenge data sets, speedups of a factor of 3 orders of magnitude are demonstrated. For realistic PTA data sets the acceleration may become greater than six orders of magnitude due to the low signal-to-noise ratio.

Key words: gravitational waves – methods: data analysis – pulsars: general.

1 INTRODUCTION

In the past several decades, pulsar timing has been successfully used to study a wide range of science. Past successes include the confirmation of gravitational waves (GWs; Taylor & Weisberg 1982) and very accurate tests of general relativity (Kramer et al. 2006). The interesting science of these examples stems from the fact that accurate measurements of the times of arrival (TOAs) of the radio pulses allow for a precise determination of the trajectory of the pulsar relative to the Earth. This is possible because the TOAs can be accurately accounted for by current models of the pulsar trajectory, pulse propagation and pulsar spin evolution in relativistic gravity.

Among ongoing pulsar timing projects are pulsar timing arrays (PTAs), which are programmes designed to detect low-frequency (10^{-9} to 10^{-8} Hz) extragalactic GWs directly, by using a set of Galactic millisecond pulsars as nearly perfect Einstein clocks (Foster & Backer 1990). GWs perturb space–time between the pulsars and the Earth, and this creates detectable deviations from the strict periodicity in the TOAs (Estabrook & Wahlquist 1975; Sazhin 1978; Detweiler 1979). One of the main source candidates for PTAs

is an isotropic stochastic background of gravitational waves (GWB), thought to be generated by a large number of massive black hole binaries located at the centres of galaxies (Begelman, Blandford & Rees 1980; Phinney 2001; Jaffe & Backer 2003; Wyithe & Loeb 2003; Sesana, Vecchio & Colacino 2008), by relic GWs (Grishchuk 2005), or, more speculatively, by oscillating cosmic-string loops (Damour & Vilenkin 2005; Ölmez, Mandic & Siemens 2010; Sanidas, Battye & Stappers 2012).

The analysis of pulsar timing data, and even more so PTA data, can become prohibitively time consuming for large data sets. This is especially true for Bayesian data analysis methods, like the analysis of PTA data (van Haasteren et al. 2009, hereafter vHLML), and the correction for dispersion measure variations (Lee et al., in preparation). Typically, the computational cost scales as n^3 or n^2 , with n being the total number of observations; the computational difficulties will increase sharply over time.

In this work, one possible solution for the computational difficulties is explored in the case, the signal of interest being a time-correlated stochastic signal: the ABC-method (Acceleration By Compression). The ABC-method is based on lossy linear data compression. By significantly reducing the dimensionality of the problem, the evaluation of computationally expensive quantities can be greatly accelerated. We specifically focus on the

[★]E-mail: vhaasteren@gmail.com

International PTA (IPTA) Mock Data Challenge (released by M. Keith, K. J. Lee and F. A. Jenet¹), in which the GWB is a good example of a compressible stochastic signal.

The outline of the paper is as follows. In Section 2, we briefly review the relevant theory of pulsar timing observations, with a special attention to the likelihood in the presence of time-correlated stochastic signals. We introduce the ABC-method, and the compressibility of data sets, in Section 3. In Section 4, we look into some of the practicalities concerned with compression of PTA data, and investigate the computational demand of different terms in the evaluation of the likelihood. In that section, we provide and test a method based on cubic spline interpolation to estimate the compressed covariance matrix. This causes an extra speedup of a few orders of magnitude. Finally, we present our conclusions in Section 5.

2 PTA DATA ANALYSIS

The typical data processing pipeline for pulsar timing observations processes the raw baseband data in several data reduction steps, where at each data reduction step the data volume is drastically reduced. The data reduction steps condense the scientifically interesting information into a significantly smaller number of data points, sometimes mitigating noise in the process. At the end of the pipeline we are left with TOAs.

This work proposes a method to compress the TOA data even further to what we call generalized residuals. The data compression is based on the likelihood for the TOAs and the Fisher information, with information preserved only for a specific stochastic signal. To this end, we review the theory of TOAs, the likelihood, and inclusion of the timing model in this section.

2.1 The likelihood

We consider k pulsars, with n'_a TOAs for the a th pulsar, where the $n' = \sum_{a=1}^k n'_a$ TOAs are described as an addition of a deterministic and a stochastic part. In the observations, this distinction is blurred because we cannot fully separate the stochastic contributions from the deterministic contributions. In practice, we therefore work with timing residuals that are produced using first estimates β_{0i} of the m timing-model parameters β_i (i between 1 and m); this initial guess is usually assumed to be accurate enough to use a linear approximation of the timing model (Edwards, Hobbs & Manchester 2006). Here, $m = \sum_{a=1}^k m_a$ is the sum of the number of timing-model parameters of all the individual pulsars. In this linear approximation, the timing residuals depend on $\xi_i = \beta_i - \beta_{0i}$ as

$$\delta \mathbf{t}' = \delta \mathbf{t}^{\text{prf}} + \mathbf{M} \boldsymbol{\xi}, \quad (1)$$

where $\delta \mathbf{t}'$ are the timing residuals in the linear approximation to the timing model, $\delta \mathbf{t}^{\text{prf}}$ is the vector of pre-fitted timing residuals, $\boldsymbol{\xi}$ is the vector with timing-model parameters for all k pulsars and the $(n' \times m)$ matrix \mathbf{M} is the so-called design matrix (see e.g. section 15.4 of Press et al. 1992; vHLML), which describes how the timing residuals depend on the model parameters. As an example, for a simple timing model which only contains quadratic spindown, the matrix \mathbf{M} is a $(n' \times 3)$ matrix, with the j th column describing a $(j-1)$ th order polynomial. The elements of \mathbf{M} are then t_i^{j-1} , with t_i being the i th TOA.

Identical to vHLML and van Haasteren & Levin (2013, hereafter vHL), we model the stochastic contributions to the TOAs as a time-correlated stochastic signal, described by a random Gaussian process. The corresponding likelihood is

$$P(\delta \mathbf{t}' | \boldsymbol{\xi}, \boldsymbol{\phi}) = \frac{\exp\left[-\frac{1}{2} (\delta \mathbf{t}' - \mathbf{M} \boldsymbol{\xi})^T \mathbf{C}'^{-1} (\delta \mathbf{t}' - \mathbf{M} \boldsymbol{\xi})\right]}{\sqrt{(2\pi)^{n'} \det \mathbf{C}'}}, \quad (2)$$

where $\boldsymbol{\phi}$ is the vector describing all the stochastic model parameters and $\mathbf{C}' = \mathbf{C}'(\boldsymbol{\phi})$ is the covariance matrix of the sum of all stochastic signals. This includes the measurement uncertainties, the timing noise (red spin noise) and a possible GWB.

2.2 Marginalizing over the timing model

Using equation (2) is computationally not very efficient because of the large number of timing-model parameters. However, in the case of uniform priors (vHLML) and Gaussian priors (vHL) it is possible to analytically marginalize the posterior distribution over the timing-model parameters. In the remainder of this work, we assume no prior information about the timing-model parameters, and use uniform priors.

In their search for a simplified representation of the analytic marginalization procedure, vHL decomposed the design matrix into an orthogonal basis based on the singular value decomposition $\mathbf{M} = \mathbf{U} \boldsymbol{\Sigma} \mathbf{V}^*$, where \mathbf{U} and \mathbf{V} are $(n' \times n')$ and $(m \times m)$ orthogonal matrices, respectively, and $\boldsymbol{\Sigma}$ is an $(n' \times m)$ diagonal matrix. The first m columns of \mathbf{U} span the column space of \mathbf{M} , and the last $n' - m$ columns of \mathbf{U} span the complement. We denote these two subspace bases as \mathbf{F} and \mathbf{G} , respectively:

$$\mathbf{U} = (\mathbf{F} \ \mathbf{G}).$$

In Section 3.1, we show that \mathbf{G} is actually a lossless data compression matrix.

Now, integrating over $\boldsymbol{\xi}$, our marginalized likelihood becomes (vHLML)

$$\int d^m \boldsymbol{\xi} P(\delta \mathbf{t}' | \boldsymbol{\xi}, \boldsymbol{\phi}) = \frac{\sqrt{\det(\mathbf{F}^T \mathbf{C}'^{-1} \mathbf{F})}}{\sqrt{(2\pi)^{n'} \det \mathbf{C}'}} \times \exp\left(-\frac{1}{2} \delta \mathbf{t}'^T \mathbf{C}_p^{-1} \delta \mathbf{t}'\right), \quad (3)$$

with

$$\begin{aligned} \mathbf{C}_p^{-1} &= \mathbf{C}'^{-1} - \mathbf{C}'^{-1} \mathbf{F} (\mathbf{F}^T \mathbf{C}'^{-1} \mathbf{F})^{-1} \mathbf{F}^T \mathbf{C}'^{-1} \\ \mathbf{C}_p &= \mathbf{G} \mathbf{G}^T \mathbf{C}' \mathbf{G} \mathbf{G}^T, \end{aligned} \quad (4)$$

where the singular matrix \mathbf{C}_p^{-1} is the inverse of \mathbf{C}_p in the non-singular subspace of its basis. The singular matrix \mathbf{C}_p is the post-fit covariance matrix of the timing residuals (vHL; Demorest et al. 2012, hereafter D12). D12 use a pseudo-inverse based on a singular value decomposition of \mathbf{C}' to evaluate \mathbf{C}_p^{-1} in their evaluation of a GWB detection statistic; this is equivalent to marginalizing over the timing-model parameters (vHL).

3 THE ABC-METHOD

Data compression is the encoding of information in a smaller data volume than the original information data volume. This can be done without losing information (lossless) or with losing information (lossy) (Wade & Wade 1994). We would like to use data compression to reduce our data volume, with the aim of speeding up the computations that are necessary for the analysis of PTA data. In this

¹ http://www.ipta4gw.org/?page_id=214

work, we compress the data in such a way to retain the sensitivity to one stochastic signal (e.g. the isotropic GWB): the ‘ABC-method’.

In Section 3.1, we show that marginalization over the timing-model parameters is equivalent to lossless data compression. In Section 3.2, we expand the data compression formalism, and show how to construct a basis in which sensitivity to a particular signal is retained. We define the corresponding compression fidelity in Section 3.3. Finally, in Sections 3.4 and 3.5, we discuss how to interpret the compressed basis of generalized residuals, and how far a data set can be compressed without significant loss of information.

3.1 Marginalization: lossless data compression

vHL showed that equation (3) can be rewritten as

$$\int d^m \xi P(\delta \mathbf{t}' | \xi, \phi) = \frac{\exp \left[-\frac{1}{2} \delta \mathbf{t}'^T \mathbf{G} (\mathbf{G}^T \mathbf{C}' \mathbf{G})^{-1} \mathbf{G}^T \delta \mathbf{t}' \right]}{\sqrt{(2\pi)^n \det (\mathbf{G}^T \mathbf{C}' \mathbf{G})}}, \quad (5)$$

with notation as given in Section 2.2. This is an unmarginalized likelihood of a random Gaussian process in n dimensions, with data $\delta \mathbf{t} = \mathbf{G}^T \delta \mathbf{t}'$ and covariance matrix $\mathbf{C} = \mathbf{G}^T \mathbf{C}' \mathbf{G}$. The dimensionality of the data is reduced from n' to n due to the marginalization process. From here onwards, we start the convention that a prime superscript denotes that a vector or a covariance matrix lies in the larger unmarginalized space, whereas no prime denotes that either of them lies in the marginalized space. The vector $\delta \mathbf{t}$ contains all the information about all stochastic signals: marginalization over the timing-model parameters is the same as lossless linear data compression in this formalism. The matrix \mathbf{G} is our linear data compression matrix and $\delta \mathbf{t}$ is our vector of reduced data.

3.2 Lossy linear data compression

We would like to compress the reduced data $\delta \mathbf{t}$ even further, without losing too much information about the stochastic signal of our interest. We expect this to be possible, since usually the signal and the noise differ in power spectral density. Only some parts of the spectrum are dominated by the signal; other parts are dominated by the noise. The data compression scheme in this work is based on throwing away the parts of the data that are dominated by the noise by using linear data compression: $\mathbf{x} = \mathbf{H}^T \delta \mathbf{t}$, with \mathbf{x} being the compressed data, or ‘generalized residuals’ as we will call them and \mathbf{H} being the *compression matrix*. Here, the number of columns of \mathbf{H} is less than the number of rows, where we define the compression to be the total number of timing residuals divided by the number of compressed generalized timing residuals. We derive one possible scheme to construct a suitable \mathbf{H} in this section.

In order to determine how much information about our signal of interest is in our data, we use the Fisher information. We acknowledge that formally the Fisher information does not completely quantify how well a parameter can be confined with a specific data set, especially in the case of a low signal-to-noise ratio (e.g. Vallisneri 2008), but in this exploratory work we consider the Fisher information as a sufficient first attempt. Denoting the log likelihood of equation (5) as Λ , we find for the Fisher information

$$\mathcal{I}_{\theta\phi} = \left\langle -\frac{\partial^2 \Lambda}{\partial \theta \partial \phi} \right\rangle = \frac{1}{2} \text{Tr} \left(\frac{\partial \mathbf{C}}{\partial \theta} \mathbf{C}^{-1} \frac{\partial \mathbf{C}}{\partial \phi} \mathbf{C}^{-1} \right), \quad (6)$$

where $\mathcal{I}_{\theta\phi}$ is the Fisher information, and ϕ and θ are model parameters that affect the signal power spectral density. Suppose that the stochastic processes in the reduced data $\delta \mathbf{t}$ are described by the

covariance matrix $\mathbf{C} = \Sigma + a^2 \mathbf{S}$, where Σ is the covariance matrix of the noise and \mathbf{S} is the covariance matrix of the signal of interest with amplitude a^2 . We would like to know which basis vectors have the largest contribution to the Fisher information, which would be easiest to determine if we could completely diagonalize the matrices in the trace of equation (6). This is possible with a non-orthogonal transformation. Even though the inner product is not preserved in such a transformation, the trace remains invariant. We use a square root of the noise matrix, $\Sigma_w^{-1/2} = \Sigma^{-1/2}$, to do this. For the moment, we assume that this estimate of Σ_w is indeed correct, but in Section 3.4 we argue that an inaccurate noise estimate still results in a usable compression. In this new basis, the whitened data and covariance become $\delta \mathbf{t}^w = \Sigma_w^{-1/2} \delta \mathbf{t}$ and $\mathbf{C}^w = \Sigma_w^{-1/2} \mathbf{C} \Sigma_w^{-1/2}$. The maximum sensitivity based on the Fisher information now has a simple form

$$\frac{a^2}{\text{Var}(a)} \leq a^2 \mathcal{I}_{aa} = 2 \sum_{i=1}^n \frac{a^2 \lambda_i^2}{(1 + a \lambda_i)^2}, \quad (7)$$

where λ_i is the i th eigenvalue of \mathbf{C}^w . The $a \lambda_i$ should be interpreted as signal-to-noise ratios. We can only evaluate equation (7) if we have complete knowledge of the signal \mathbf{S} , the signal amplitude a , and the noise Σ . However, the λ_i and the corresponding basis vectors do not depend on a , which means we can examine the sensitivity to a as a function of the number of λ_i we include. Here, we do assume to have knowledge of Σ and \mathbf{S} .

In the limit where a is large, the strong signal limit, we can neglect the one in the denominator of the sum of equation (7), which makes all terms in the sum equal. This means that all generalized residuals carry equal information as is expected in such a case: the noise is negligible compared to the signal, so no parts of the signal are buried under the noise. In the strong signal limit, data compression is therefore not possible. Note that the sensitivity is then proportional to the number of generalized residuals, as it should be.

In the limit that a is small, the low-signal limit (LSA), we can neglect all terms $a \lambda_i$ in the denominator of the sum, making the sensitivity equal to the sum of all $a^2 \lambda_i^2$. The distribution of values of the λ_i eigenvalues is determined by the power spectral density of the signal compared to the noise. If the signal spectrum is the same as the noise spectrum, all λ_i will be identical. However, if the signal has a different spectrum than the noise, λ_i can span a wide range of values, where the large λ_i correspond to basis vectors where the signal is relatively large compared to the noise. In this case, there are nearly redundant data points, and compression is possible.

3.3 The compression fidelity

We define the fidelity $\mathcal{F} \in [0, 1]$ to be the fraction of the total sensitivity we retain in our compressed data. We choose the number of generalized residuals that we keep, l , to be the smallest number such that

$$\frac{\sum_{i=1}^l \lambda_i^2 / (1 + a \lambda_i)^2}{\sum_{i=1}^n \lambda_i^2 / (1 + a \lambda_i)^2} \geq \mathcal{F}, \quad (8)$$

where we have ordered λ_i to have the largest values for the lowest indices. We typically work with $\mathcal{F} \geq 0.99$, which in favourable cases like the IPTA Mock Data Challenge allows for compressions greater than 10: less than 10 per cent of the original data volume is kept.

Computationally, we suggest to use a singular value decomposition to produce the eigenvalues and eigenvectors of \mathbf{C}^w , where our fidelity criterion, equation (8), keeps only l of the n generalized

residuals δt_i^w . We construct the $(n \times l)$ matrix \mathbf{W} as consisting of the columns of the l eigenvectors that belong to the selected eigenvalues. The data compression matrix is now $\mathbf{H} = \Sigma_w^{-1/2} \mathbf{W}$. Using equation (5), we now find for the likelihood of the compressed data $\mathbf{x} = \mathbf{H}^T \delta \mathbf{t}$:

$$P(\mathbf{x}|\phi) = \frac{\exp\left[-\frac{1}{2} \mathbf{x}^T (\mathbf{H}^T \mathbf{C} \mathbf{H})^{-1} \mathbf{x}\right]}{\sqrt{(2\pi)^l \det(\Sigma_w) \det(\mathbf{H}^T \mathbf{C} \mathbf{H})}}, \quad (9)$$

where the extra determinant of Σ_w comes from the whitening and can be ignored in practice as it is absorbed in the overall normalization constant. This equation is the basis of the ABC-method, as the computationally expensive inversion has been replaced with a lower dimensional one.

Equation (9) is completely general, and can be readily applied to realistic data sets. As in van Haasteren et al. (2011), all timing-model parameters and jumps can be included in the likelihood, and are therefore by design part of the data compression scheme. We therefore expect not to encounter any difficulties in applying data compression to realistic data sets, even though in this work we only test the effectiveness on the Mock Data Challenge.

3.4 Interpreting the compressed basis

As we have discussed in Section 3.1, marginalizing over the timing model is the same as linear data compressing the subspace of the original data $\delta \mathbf{t}'$ orthogonal to the columns of the design matrix \mathbf{M} . Similarly, the data compression we suggest in Section 3.2 is equivalent to marginalizing over vectors that lie in the subspace orthogonal to the column space of \mathbf{H} with uniform priors. By considering the data in the basis orthogonal to the column space of \mathbf{H} to be nuisance parameters with uniform priors, the resulting likelihood of the compressed generalized residuals becomes independent of the value of the data in the orthogonal complement (we have not found another prior with the same property). This interpretation of data compression in terms of marginalization ensures us that we are not introducing any biases or unwanted systematics in our analysis. The difference with the marginalization over the timing model is that we do not marginalize over physical nuisance parameters; we are throwing away information. The data compression matrix \mathbf{H} as constructed in Section 3.2 guarantees that we throw away as little information about the signal amplitude a as possible. This is not true for the other parameters this signal may also depend on: optimal sensitivity to those parameters possibly requires a different basis, construction of which is the subject of an ongoing follow-up research. We ignore this issue in the rest of this exploratory work, and assume that sensitivity to the signal amplitude is sufficient for our purposes.

The interpretation of data compression in terms of marginalizing over non-physical parameters ensures us that the likelihood of the compressed data in equation (9) is also valid if we do not provide good estimators for Σ_w and \mathbf{S} . We may be throwing away more information than we thought if our estimators are not accurate, but we do not introduce any bias or systematics in our likelihood. It is therefore not imperative to be thorough in the estimates of the signal and the noise; a reasonable guess may be sufficient for practical purposes.

It is instructive to inspect the compressed basis vectors $\mathbf{G}\mathbf{H}$ for highly compressible signals. We choose the IPTA Mock Data Challenge as an example, since the GWB signal strongly dominates the noise at the lowest frequencies in these data sets. In Fig. 1 we present the first three compressed basis vectors for J0030+0451

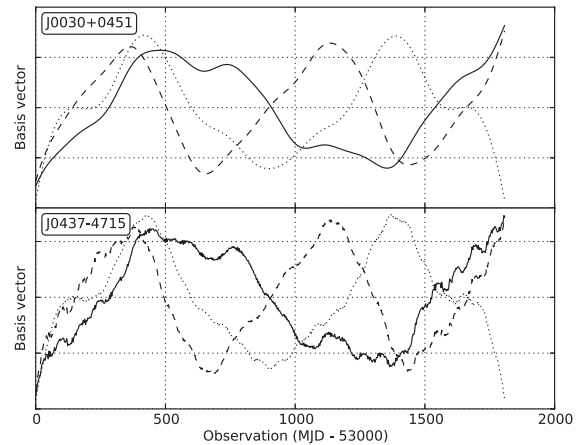


Figure 1. The first three compressed basis vectors of J0030+0451 and J0437–4715 of IPTA Mock Data Challenge open 1. These basis vectors are the first three columns of the matrix $\mathbf{G}\mathbf{H}$. The basis vectors are normalized, so we have ignored the scaling on the y-axis. The basis vectors of J0437–4715 have more high-frequency structure due to the fact that J0437–4715 is in a binary.

and J0437–4715 of Mock Data Challenge open 1. We observe that, roughly, the first basis vector corresponds to a third-order polynomial: start negative, then ascend to a maximum, descend to a minimum and finally end positive. The other two basis vectors display a similar behaviour with the order of the polynomial equal to the order of the basis vector +2. Note that the zeroth, first and second order are missing due to the removal of quadratics in our marginalization over the timing-model parameters.

We note that the compressed basis vectors for both pulsars in Fig. 1 are similar, except that those of J0437–4715 display more high-frequency behaviour. This is because J0437–4715 resides in a binary, and the timing model therefore includes parameters for binary motion.

3.5 Compressibility: how far can we go?

A natural question that arises in data compression is how much we can compress the data without losing a significant amount of information. To answer this question, we consider the fidelity as a function of the number of generalized residuals in the data set. For compressible data sets we expect the fidelity to stay close to 1, only to drop for high-compression rates. One possible measure of compressibility, which we use in our application of the ABC-method, is the maximum compression for which the fidelity stays above 0.99. This maximum compression depends on the signal amplitude and power spectral density compared to that of the noise.

As an example, we plot the fidelity of the mock data of J0030+0451 from the open Mock Data Challenge versus the compression in Fig. 2, where the signal of interest is the gravitational-wave background. In these data sets the noise is white. Open data set 3 does formally contain some extra (mildly) red noise which we do include in these plots, but the level of red noise is so low that it is negligible in practice. Because the signal is of such a different spectral shape than the noise, data compression is very efficient. In such a case, the higher the noise level compared to the signal, the more compressible the data set is.

In Fig. 2, we plot the fidelity for open data set 3 in the LSA. Data compression is most efficient in the LSA. As a comparison we show the fidelity for an incompressible signal in Fig. 2 as well. This

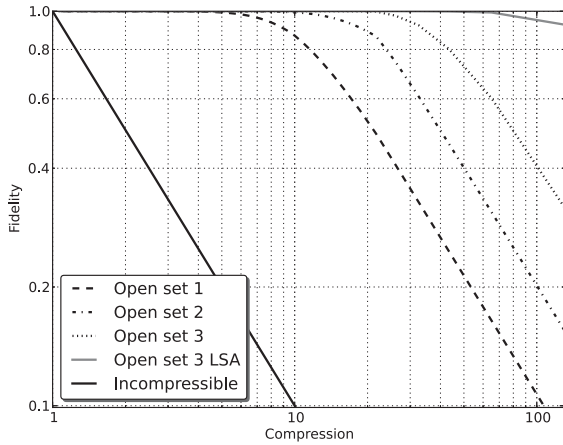


Figure 2. The fidelity of the GWB signal as a function of the compression for the pulsar J0030+0451 of all IPTA open data challenge sets. Open data set 3 contains more redundant information than the other two sets because the GWB is smaller, and the GWB signal is therefore buried under the noise at a larger portion of the spectrum. Open data set 2 contains more redundant information than open data set 1 because the error bars for J0030+0451 observations were set higher in open data set 2 than in open data set 1. We have also plotted the fidelity of the low-signal approximation (open three LSA) for open data set 3, and the fidelity of an incompressible signal (incompressible).

corresponds to the high-signal limit. We see that an increase in the compression results in an equal decrease in the fidelity.

3.6 Compressing realistic data sets: a prescription

For realistic data sets we generally do not know the details of the signal and the noise. The noise typically has to be characterized from the data, and we may not even be certain of the presence of a signal of interest. Since the fidelity depends on estimates of the signal and the noise, it is not clear how far exactly we can compress the data set without losing information from the signal. Here, we therefore recommend a conservative approach when preparing the ABC-method.

$a\lambda_i$ in the denominator of equation (8) represents the signal relative to the noise. The larger it is relative to 1, the less likely we will discard that generalized residual. Therefore, if we are sure not to overestimate the noise, and if we are sure not to underestimate the signal, the compression fidelity will not be overestimated. Specifically, we recommend to calculate the fidelity as follows.

- 1) Construct the noise covariance estimate Σ_w such that it only consists of the TOA uncertainties.
- 2) Choose a suitable spectral form for the signal of interest. For example, this consists of fixing the spectral index $\gamma = 13/3$ for the GWB.
- 3) Use the estimates of vHL (equations 22 and 24 of vHL) to estimate the signal amplitude. For a GWB signal, this is

$$\sigma_{\text{GWB}} = 1.37 \times 10^{-9} \left(\frac{A_h}{10^{-15}} \right) \left(\frac{T}{\text{yr}} \right)^{5/3}, \quad (10)$$

where T is the duration of the experiment, A_h is the dimensionless GWB amplitude and σ_{GWB} is the rms residual due to the GWB in the data. For other power spectral densities a similar calculation to vHL is required.

By completely ignoring other effects like red spin noise in these estimates, we are ensured that we do not throw away more informa-

tion than we should. Indeed, more noise in this calculation would mean a higher compression. This conservative approach is therefore also guaranteed to work in the presence of (strong) red noise.

We note that this approach can overestimate the fidelity if the TOA uncertainties have been overestimated, or when the shape of the signal power spectral density has been estimated incorrectly with, for instance, an incorrect spectral index. The TOA uncertainties depend on complex details of the data reduction pipeline prior to the formation of the TOAs and of the cross-correlation of the pulse profile with a template (Taylor 1992). However, underestimation of the TOA uncertainty is uncommon in practice. How to choose a suitable basis to be sensitive to the spectral index is a subject of ongoing follow-up research. Here, we assume that we know the spectral index of the signal of interest.

In the open Mock Data Challenge, shown in Fig. 2, high compressions of over 50 still yield a fidelity close to $\mathcal{F} = 1$ in the LSA. Since realistic data sets are expected to be in the LSA – we have not detected a GWB yet – we expect high compressions in realistic data sets to be possible as well. However, realistic data sets can have far more TOAs per pulsar than the 130 TOAs per pulsar in the Mock Data Challenge. Since in the low-signal limit only a few generalized residuals per pulsar are enough to reach $\mathcal{F} \geq 0.99$, we expect very high compressions, possibly up to $c = 1000$ depending on the size of the data set, to be realistic for initial PTA applications.

4 LINEAR DATA COMPRESSION IN PRACTICE

Although the raw data of pulsar observations can be quite voluminous, the pulsar TOA data files are typically several kilobytes in size. It seems quite unlikely that data volume at this stage of the analysis is ever going to be a problem; the only reason to resort to data compression is that it can greatly accelerate the analysis of pulsar timing data. In this section, we discuss the computational costs of evaluating the likelihood function with the ABC-method, and we present some computational shortcuts. A straightforward application is a Bayesian analysis (e.g. vHML), but other analysis methods described in the time domain are expected to see an equally large acceleration (e.g. D12). Special attention is given to power-law signals, for which we present a convenient approximation of the compressed covariance matrix, thereby maximizing the effectiveness of data compression.

4.1 Computational demand

The computational demand of equation (5) scales as n^3 (vHL) due to the inversion operation of an $(n \times n)$ matrix. With linear data compression, we have decreased the size of the inversion matrix, which will therefore also decrease the computational demands. The computational demand of the inversion in equation (9) scales as l^3 . Depending on the compression, this l^3 operation may or may not be the computational bottleneck. For large enough compression factors, the computational bottleneck will either be the computation of \mathbf{C} (n^2 operation), or the multiplication $\mathbf{H}^T \mathbf{C} \mathbf{H}$ (ln^2 operation). In the case of an array of pulsars, the matrix \mathbf{H} will be block diagonal if the data compression has been done per individual pulsar. Then, the computation of $\mathbf{H}^T \mathbf{C} \mathbf{H}$ can be accelerated with a factor of the number of pulsars by blockwise multiplication (vHL).

In this assessment of computational demand, we have neglected the construction of the data compression matrix \mathbf{H} . A computationally expensive singular value decomposition of a full covariance matrix is required for this. However, this only needs to be done once: we do not change the compressed basis during subsequent

likelihood evaluations, even if we vary the noise/signal parameters during a Markov chain Monte Carlo simulation. Since the compressed basis can be calculated for each pulsar individually, we therefore do not expect the construction of the data compression matrix \mathbf{H} to be a computational bottleneck in the foreseeable future.

4.2 Testing the acceleration of the likelihood

We test the performance of the ABC-method on the IPTA Mock Data Challenge: all challenges consist of 130 observations per pulsar, with 36 pulsars. Our likelihood contains the following deterministic and stochastic signal contributions:

- (i) the TEMPO2 (Hobbs, Edwards & Manchester 2006) timing-model parameters,
- (ii) error bars for every TOA,
- (iii) power-law red timing noise for every pulsar and
- (iv) a correlated GWB.

Evaluation of the likelihood of equation (3) took on average 38.3 s^2 , where most of that time comes from inverting the full covariance matrix.

We compare the efficiency of equation (3) to that of the data compression likelihood of equation (9), where the latter equation becomes equation (5) when the compression is 1. In the evaluation of the compressed likelihood, three terms take up the majority of the computational cost.

- (i) \mathbf{C}^{GW} , the evaluation of the $(n \times n)$ elements of the covariance matrix of the GWB.
- (ii) $\mathbf{H}^T \mathbf{C}^{\text{GW}} \mathbf{H}$, the matrix multiplication to obtain the compressed covariance matrix.
- (iii) $(\mathbf{H}^T \mathbf{C} \mathbf{H})^{-1}$, inversion of the compressed covariance matrix.

All other operations are negligible compared to these three. In Fig. 3 we present the computational cost of these three terms, together with the sum of the three, in the bottom panel. The uncompressed likelihood is given as a single point. We see that the inversion of the compressed covariance matrix is the dominant term for low-compression factors: if roughly 70 or more generalized residuals per pulsar are kept. For higher compression factors, the evaluation of \mathbf{C}^{GW} is the most time-consuming part of the evaluation of the likelihood. Because this is an n^2 operation that does not depend on the compression, compressing the data to less than 50 generalized residuals per pulsar does not gain us any computational efficiency in this configuration.

4.3 Signals with unknown amplitude

As explained in the previous section, in the case where a data set is highly compressible, the computational bottleneck becomes evaluating \mathbf{C} , which contains \mathbf{C}^{GW} in the example of Section 4.2, at each step of the likelihood function. If we label the contributions to the compressed covariance matrix as $\mathbf{H}^T \Sigma_i \mathbf{H}$, then in some cases it is possible to greatly accelerate the evaluation of $\mathbf{H}^T \Sigma_i \mathbf{H}$. The simplest type of stochastic signal is the type where the power spectral density shape is known completely, but the amplitude N_i is an unknown model parameter. Examples of signals of this type include the stochastic behaviour due to TOA uncertainties (with an

² All computations in this work are performed on a single workstation, code linked with an Automatically Tuned Linear Algebra System (ATLAS) library that came with the GNU/Linux distribution.

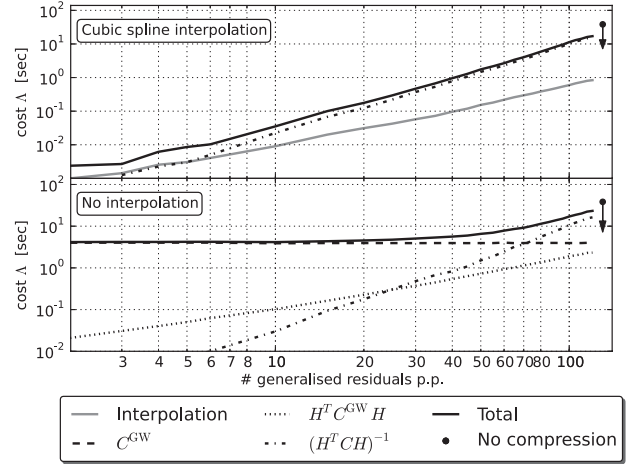


Figure 3. The computational cost of the dominating terms in the compressed likelihood, as a function of the number of compressed generalized residuals per pulsar. An array of 36 pulsars was used, with 130 observations per pulsar. The dominating terms are as follows. (1) \mathbf{C}^{GW} , (dashed line), the evaluation of the $(n \times n)$ elements of the covariance matrix of the GWB. Only present in the lower panel. (2) $\mathbf{H}^T \mathbf{C}^{\text{GW}} \mathbf{H}$, (dotted line), the matrix multiplication to obtain the compressed covariance matrix. Only present in the lower panel. (3) Interpolation, (grey solid line), the construction of the $(l \times l)$ compressed covariance matrix $\mathbf{H}^T \mathbf{C} \mathbf{H}$ by cubic spline interpolation. Only present in the upper panel. (4) $(\mathbf{H}^T \mathbf{C} \mathbf{H})^{-1}$, (dash-dotted line), inversion of the compressed covariance matrix. The total computational cost is shown as a solid line, and the uncompressed likelihood of equation (3) is shown as an upper limit at 130 generalized residuals per pulsar. In the lower panel, these terms are evaluated for the compressed likelihood of equation (9), without any computational shortcuts. For high-compression factors (low number of compressed generalized residuals), the evaluation of \mathbf{C}^{GW} is dominant, which means that further compression does not buy one more computational time. In the upper panel, the compressed likelihood is evaluated, where the cubic spline interpolation method of Section 4.4 is used to evaluate \mathbf{C}^{GW} . In this case, the inversion $(\mathbf{H}^T \mathbf{C} \mathbf{H})^{-1}$ is always the dominant term, and data compression is most efficient. Note how the line for $(\mathbf{H}^T \mathbf{C} \mathbf{H})^{-1}$ is (nearly) identical in both panels.

unknown scaling, or ‘EFAC’, parameter), pulse phase jitter (e.g. Cordes & Shannon 2010), or a GWB with a known spectral index. For these types of signal we can evaluate $\mathbf{H}^T \Sigma_i \mathbf{H}$ just once for unit amplitude, and store this in memory. Then, each time we need to evaluate the likelihood function, we can multiply this stored matrix with the amplitude N_i to obtain the compressed covariance matrix without having to recalculate such matrices every time. Especially when $l \ll n$, this greatly reduces the time necessary to evaluate $\mathbf{H}^T \Sigma_i \mathbf{H}$.

4.4 Power-law signals

Most stochastic signal models have more free parameters than only an amplitude, and the acceleration method of Section 4.3 is not applicable. In this section, we present a practical solution for signals with two free parameters: an amplitude, and some other parameter. We focus only on signals with a power-law power spectral density, but we expect that the method is also appropriate for other signals with a parametrized power spectral density.

Power-law signals are used in various ways in pulsar timing, both as a model for noise sources (i.e. red spin noise; Cordes & Shannon 2010; Shannon & Cordes 2010), and as signal sources (i.e. the

istotropic GWB; Phinney 2001; vHLML. We use the following definition for the power spectral density of a power-law signal:

$$S(f) = a^2 \left(\frac{1}{\text{yr}^{-1}} \right) \left(\frac{f}{\text{yr}^{-1}} \right)^{-\gamma}, \quad (11)$$

where f is the signal frequency, a is the signal amplitude and γ is the spectral index that describes the steepness of the spectrum. The rms in the timing residuals of such a signal is given by $\sigma_{\text{rms}}^2 = \int_0^\infty df S(f)$. Because this is an unphysical power spectrum that diverges at the low frequencies, in practice a third parameter is used to describe a power-law signal that represents a lowest frequency f_L below which the signal is assumed to be zero. The reduced data δt and therefore also the compressed data x are not affected by f_L (vHL; Blandford, Romani & Narayan 1984; Lee et al. 2012).

For highly compressed data, the compressed covariance matrix $\mathbf{H}^T \mathbf{C} \mathbf{H}$ contains far less elements than \mathbf{C} : the number of unique elements for this matrix is $l(l+1)/2$. For a single pulsar power-law noise covariance matrix, this is typically only of the order of a hundred elements, depending on the number of observations and the compression. We propose to use an interpolation approximation for each element of the matrix $\mathbf{H}^T \mathbf{C} \mathbf{H}$ as a function of γ , with $1 < \gamma < 7$. The elements of the covariance matrix diverge at both ends of the interval. In the case of a single pulsar, this means we have $l(l+1)/2$ functions on the interval $1 < \gamma < 7$ that we want to write an interpolation approximation for. We choose a cubic spline interpolation method for this, where the domain of the function is divided in subintervals in which the function is approximated by a third-order polynomial. We construct all polynomials such that their values and derivatives match at the edges. The only free parameter in this approach is the number of cubics used in total. This number needs to be tuned for performance.

In Fig. 4 we show the difference between the true value and the interpolated value of an arbitrary element of $\mathbf{H}^T \mathbf{C} \mathbf{H}$ as a function of γ for J0030+0451 of Mock Data Challenge open 2. These results are typical; we find a similar plot for every element, where the difference between the true value and the interpolated value always inflates near the boundaries of the interval. We also show the difference between the accompanying log-likelihood Λ as a function of γ for the same data set. Here, we also see that the difference inflates near the boundaries. The precision of the interpolation depends on the number of cubic splines used in the interpolation. For lower numbers of splines in the approximation, we saw that the accuracy quickly decrease near the boundaries. This caused the compressed covariance matrix to become non-positive definite or singular close to the boundaries. In our simulations, 100 equally spaced cubic splines were enough on a slightly reduced interval $1.09 < \gamma < 6.91$ to not run into numerical issues.

The cubic spline interpolation removes the necessity to calculate the total covariance matrix \mathbf{C} . In the top panel of Fig. 3, we present the computational cost of the computationally dominant terms in the compressed likelihood, in the case where we use cubic spline interpolation for the elements of $\mathbf{H}^T \mathbf{C} \mathbf{H}$. The computationally dominant term is the inversion $(\mathbf{H}^T \mathbf{C} \mathbf{H})^{-1}$ for the whole range of possible compressions, which means that data compression is maximally efficient. We almost reached full capacity of Random Access Memory (RAM) of our workstation for very low compressions. For large data sets with an incompressible signal, this may cause problems for the cubic spline interpolation method. However, for current applications, we do not believe this to be an issue. For the Mock Data Challenge, the total typical speedup at 99 per cent fidelity is almost three orders of magnitude.

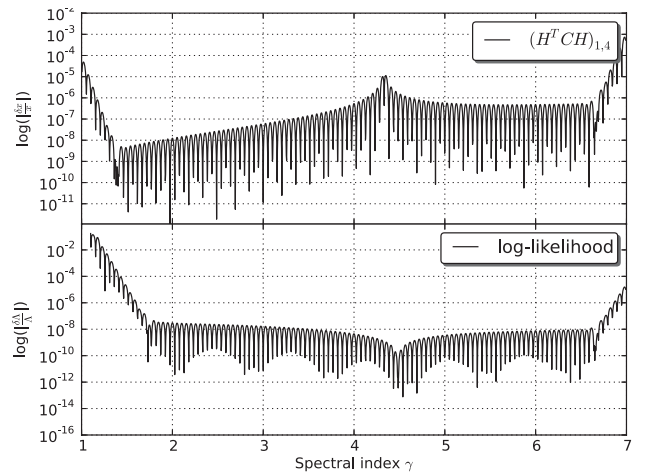


Figure 4. The likelihood and the covariance matrix $\mathbf{H}^T \mathbf{C} \mathbf{H}$ as a function of γ . For the pulsar J0030+0451, with data as in IPTA Mock Data Challenge open 2, we used the interpolation technique of Section 4.4 to approximate the elements of the compressed covariance matrix $\mathbf{H}^T \mathbf{C} \mathbf{H}$. In the upper panel, we have plotted $\log|\delta x/x|$ of element x (row 1, column 4) of the compressed covariance matrix as a function of the spectral index γ . Here, x is the true value of the element of $\mathbf{H}^T \mathbf{C} \mathbf{H}$, and δx is the difference between the true value of x , and the interpolated value. This plot looked similar for all elements. In the bottom panel, we have plotted the corresponding quantity for the log likelihood: $\log|\delta \Lambda/\Lambda|$, with Λ being the log likelihood, and $\delta \Lambda$ being the difference between the true and interpolated value. We initialized the cubic spline interpolation with 100 points, evenly distributed on the interval (1, 7). We see that the discrepancy between the interpolated and the true values grows steeply near the boundaries of the interval. At the boundaries, the elements of the compressed covariance matrix diverge.

4.5 Tests on the IPTA Mock Data Challenge

We test the ABC-method with the cubic spline interpolation technique on the open Mock Data Challenge. We present the results here of Mock Data Challenge open 1 because the noise level was the same for all pulsars in that challenge. That makes it easier to compare the results we see here with the fidelity levels of Fig. 2: they are approximately the same for all pulsars. In Fig. 2 we see that for a compression of 6, we start to approach $\mathcal{F} \approx 0.99$. This corresponds to 22 compressed generalized residuals per pulsar. In Fig. 5 we present the likelihood credible regions for Mock Data Challenge open 1 both for the full array of pulsars and for pulsar J0030+0451, with different compression levels. We see that with 22 generalized residuals per pulsar, the compressed likelihood is practically equal to the uncompressed likelihood, as predicted by Fig. 2. With less than 22 generalized residuals per pulsar, the likelihood credible regions are broader, with significant covariance between the GWB amplitude and the spectral index. This covariance may partially be a result of the compressed basis being optimal only for the injected value of the spectral index $\gamma = 4.33$; this dependence is the subject of follow-up work.

The results of this section hold for all three of the open Mock Data Challenge data sets: when the fidelity $\mathcal{F} \geq 0.99$, the likelihood credible regions were almost indistinguishable from the uncompressed likelihood credible regions. With a compression such that the fidelity is significantly less than that, the credible regions were broader, with a covariance between the amplitude and spectral index.

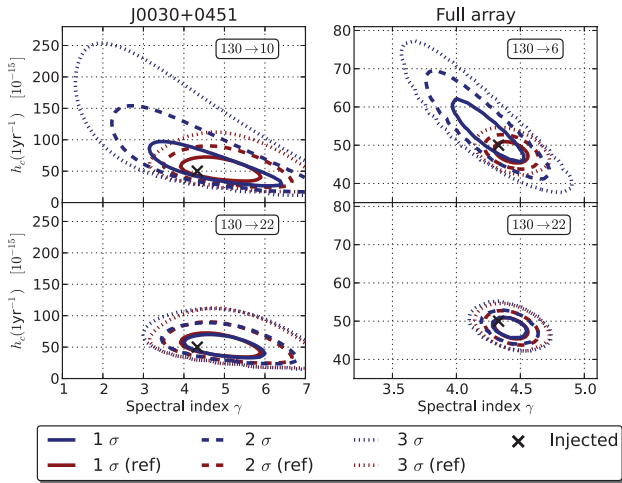


Figure 5. The compressed likelihood as a function of the GWB amplitude and spectral index γ for IPTA Mock Data Challenge open 1. No parameters are numerically marginalized over. The timing-model parameters are analytically marginalized over as part of the data compression. On the left-hand panel, the likelihood is plotted for only pulsar J0030+0451, with compression to 10 generalized residuals (top) and compression to 22 generalized residuals (bottom). On the right the likelihood is plotted for the full array of pulsars, with compression to 6 generalized residuals per pulsar (top) and compression to 22 generalized residuals per pulsar (bottom). In each panel, the blue lines represent the credible regions of the compressed likelihood, the red lines, labelled ‘ref’, represent the reference credible regions of the uncompressed likelihood of equation (5). The contours represent the 1σ (68 per cent), 2σ (95 per cent) and 3σ (99.7 per cent) credible regions. The injected values are marked with a ‘x’.

5 CONCLUSIONS

We investigate the acceleration of the analysis of pulsar timing data by compressing the data with a linear transformation, without losing a significant amount of information of a particular stochastic signal of interest: the ABC-method. In this formalism, marginalization over the timing-model parameters is equivalent to lossless linear data compression. We show that when the stochastic signal of interest has a significantly different spectrum than the noise, the data is highly compressible. The ABC-method is most efficient in the LSA, where the signal is buried under the noise over most of the frequency range. Data compression is not possible in the strong-signal limit, where the signal dominates the noise in the whole frequency range. The likelihood function of the compressed signal is computationally more efficient and unbiased.

We introduce the concepts of compression and compression fidelity, where the compression is the total number of timing residuals divided by the number of generalized timing residuals that are kept in the compression, and the fidelity is a measure of the amount of information about the signal of interest that is kept in the compression. For the IPTA Mock Data Challenge, we show that the compression is of the order of 10, at a fidelity $\mathcal{F} = 0.99$, if one is interested in the isotropic stochastic GWB.

When applied to highly compressible data sets, computational shortcuts are required to optimally accelerate the evaluation of the compressed likelihood. We present a practical method based on cubic spline interpolation of the compressed covariance matrix. When this interpolation approximation is used, the total acceleration of the evaluation of the compressed likelihood is c^3 , with c being

the compression. We test the cubic spline interpolation method, and conclude that it works well for the purposes of the IPTA Mock Data Challenge. The total acceleration is about three orders of magnitude for a compression of 10, with results almost identical to an analysis without the ABC-method.

The ABC-method can be readily applied to realistic data sets, without any adjustments. Realistic data sets of current PTAs are expected to reside in the low-signal approximation: no stochastic gravitational-wave background has been detected as of now. Therefore, a high compression factor of several hundred is realistic for such data sets, which yields a total acceleration of over six orders of magnitude. We expect linear data compression to become one of the key solutions for the issues related to computational cost in PTA data analysis.

ACKNOWLEDGMENTS

The author thanks K. J. Lee for some insightful discussions and Alberto Sesana for providing the inspiration for this work. Yuri Levin and Michele Vallisneri are thanked for their extensive comments on this manuscript. The research in this paper is conducted as part of the efforts of the European Pulsar Timing Array (EPTA).

REFERENCES

- Begelman M. C., Blandford R. D., Rees M. J., 1980, *Nat*, 287, 307
 Blandford R., Romani R. W., Narayan R., 1984, *JA&A*, 5, 369
 Cordes J. M., Shannon R. M., 2010, preprint (arXiv:e-prints)
 Damour T., Vilenkin A., 2005, *Phys. Rev. D*, 71, 063510
 Demorest P. B., Ferdman R. D., Gonzalez M. E., Nice D., Ransom S., Stairs I. H., Arzoumanian Z., Brazier A., 2012, preprint (arXiv:e-prints) (D12)
 Detweiler S., 1979, *ApJ*, 234, 1100
 Edwards R. T., Hobbs G. B., Manchester R. N., 2006, *MNRAS*, 372, 1549
 Estabrook F., Wahlquist H., 1975, *Gen. Relativ. Grav.*, 6, 439
 Foster R., Backer D., 1990, *ApJ*, 361, 300
 Grishchuk L. P., 2005, *Uspekhi Fizicheskikh Nauk*, 48, 1235
 Hobbs G. B., Edwards R. T., Manchester R. N., 2006, *MNRAS*, 369, 655
 Jaffe A., Backer D., 2003, *ApJ*, 583, 616
 Kramer M. et al., 2006, *Sci*, 314, 97
 Lee K. J., Bassa C. G., Janssen G. H., Karuppusamy R., Kramer M., Smits R., Stappers B. W., 2012, *MNRAS*, 423, 2642
 Ölmez S., Mandic V., Siemens X., 2010, *Phys. Rev. D*, 81, 104028
 Phinney E. S., 2001, preprint (arXiv:e-prints)
 Press W., Teukolsky S., Vetterling W., Flannery B., 1992, *Numerical Recipes in C*, 2nd edn. Cambridge Univ. Press, Cambridge
 Sanidas S. A., Battye R. A., Stappers B. W., 2012, *Phys. Rev. D*, 85, 122003
 Sazhin M. V., 1978, *SvA*, 22, 36
 Sesana A., Vecchio A., Colacino C. N., 2008, *MNRAS*, 390, 192
 Shannon R. M., Cordes J. M., 2010, *ApJ*, 725, 1607
 Taylor J. H., 1992, *Philos. Trans. R. Soc. Lond.*, 341, 117
 Taylor J. H., Weisberg J. M., 1982, *ApJ*, 253, 908
 Vallisneri M., 2008, *Phys. Rev. D*, 77, 042001
 van Haasteren R., Levin Y., 2013, *MNRAS*, 428, 1147 (vHL)
 van Haasteren R., Levin Y., McDonald P., Lu T., 2009, *MNRAS*, 395, 1005 (vHLML)
 van Haasteren R. et al., 2011, *MNRAS*, 414, 3117
 Wade G., Wade J. G., 1994, *Signal Coding & Processing*, 2nd edn. Cambridge Univ. Press, New York, NY
 Wyithe J., Loeb A., 2003, *ApJ*, 595, 614

This paper has been typeset from a $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.