Recent advances in the Gaussian Process formulation of GW searches on pulsar-timing data

Michele Vallisneri

Jet Propulsion Laboratory
California Institute of Technology

© 2016 California Institute of Technology

Scaling of likelihood

Marginalized likelihood: "Low" number of parameters. Computationally expensive.

$$L_m\left(\vec{\theta}\right) = \Pr\left(\vec{d} \mid \vec{\theta}\right)$$

$$\Pr\left(\vec{d} \mid \vec{\theta}\right) = \frac{\exp\left(-\frac{1}{2}\vec{d}^T\left(N^{-1} - N^{-1}T\overbrace{\left(T^T N^{-1} T + B^{-1}\right)^{-1}}^{\Sigma} T^T N^{-1}\right)\vec{d}\right)}{\sqrt{\det\left(2\pi N\right)\det\left(2\pi B\right)\det\left(2\pi\Sigma\right)}}$$

on 9-yr set:

$n_p \approx 400$

$t \approx 2\,\text{sec}$

Hierarchical likelihood: High number of parameters. Computationally cheap.

$$L_h\left(\vec{b},\vec{\theta}\right) = \Pr\left(\vec{d} \mid \vec{b}\right)\Pr\left(\vec{b} \mid \vec{\theta}\right)$$

$$\Pr\left(\vec{d} \mid \vec{b}\right) = \frac{\exp\left(-\frac{1}{2}\left(\vec{d} - T\vec{b}\right)^T N^{-1}\left(\vec{d} - T\vec{b}\right)\right)}{\sqrt{\det\left(2\pi N\right)}} \qquad \Pr\left(\vec{b} \mid \vec{\theta}\right) = \frac{\exp\left(-\frac{1}{2}\vec{b}^T B^{-1}\vec{b}\right)}{\sqrt{\det\left(2\pi B\right)}}$$

$n_p \approx 10000$

$t \approx 0.07\,\text{sec}$

Scaling of likelihood

$$L_m\left(\vec{\theta}\right) = \Pr\left(\vec{d} \mid \vec{\theta}\right)$$

Likelihood:
^3 in low-level parameters
^3 in pulsars
^1 in observations

Search parameters:
- Hyperparameters (~400)

MCMC:
- Nsteps ~ Npar^2

Total:
- Npar ~ Npsr
- Runtime ~ Npsr^5

$$L_h\left(\vec{b},\vec{\theta}\right) = \Pr\left(\vec{d} \mid \vec{b}\right)\Pr\left(\vec{b} \mid \vec{\theta}\right)$$

Likelihood:
^1 in low-level parameters
^1 in pulsars (up to 1000s of pulsars)
^1 in observations

Search parameters:
- Hyperparameters (~400)
- Low-level parameters (~10000)

MCMC:
- Nsteps ~ Npar^2

Total:
- Npar ~ Npsr
- Runtime ~ Npsr^3:

Can we do better?

Markov Chain Monte Carlo

Optimal acceptance ratio: 55% (2 dimensions), 24% (many dimensions)

Small steps: $\Delta L \approx \Sigma \frac{\partial L}{\partial \theta_i} \Delta \theta$

Random walk: $N_{Step} \propto N_p^2$

Hamiltonian Monte Carlo

Instead of random walk, integrate along Hamiltonian trajectory to propose new state.

Sample auxiliary variables (momenta).

Set potential to log likelihood (will need gradients).

Random walk: $N_{Step} \propto N_p^{5/4}$

Previous work went some way…

arXiv:1210.3578v2 [astro-ph.IM] 25 Jun 2013

A hyper–efficient model–independent bayesian method for the analysis of pulsar timing data.

Lindley Lentati,* P. Alexander, and M. P. Hobson
Astrophysics Group, Cavendish Laboratory, JJ Thomson Avenue, Cambridge, CB3 0HE, UK

S. Taylor and J. Gair
Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge, CB3 0HA, UK

S. T. Balan
Astrophysics Group, Cavendish Laboratory, JJ Thomson Avenue, Cambridge, CB3 0HE, UK and Department of Physics and Astronomy,University College London, Gower Street, London, WC1E 6BT, UK

R. van Haasteren
Max-Planck-Institut fur Gravitationsphysik (Albert-Einstein-Institut), D-30167 Hannover, Germany
(Dated: June 26, 2013)

A new model–independent method is presented for the analysis of pulsar timing data and the estimation of the spectral properties of an isotropic gravitational wave background (GWB). Taking a Bayesian approach, we show that by rephrasing the likelihood we are able to eliminate the most costly aspects of computation normally associated with this type of data analysis. When applied to the IPTA Mock Data Challenge datasets this results in speedups of approximately two to three orders of magnitude compared to established methods, in the most extreme cases reducing the run time from several hours on the high performance computer 'DARWIN' to less than a minute on a normal work station. Due to the versatility of this approach we present three applications of the new likelihood. In the low signal–to–noise regime we sample directly from the power spectrum coefficients of the GWB signal realisation. In the high signal–to–noise regime, where the data can support a large number of coefficients, we sample from the joint probability density of the power spectrum coefficients for the individual pulsars and the GWB signal realisation using a 'Guided Hamiltonian Sampler' to sample efficiently from this high dimensional (~ 1000) space. Critically in both these cases we need make no assumptions about the form of the power spectrum of the GWB, or the individual pulsars. Finally where one wishes however, we show a power-law model can still be fitted at the point of sampling. We then apply this method to a more complex dataset designed to represent better a future IPTA or EPTA data release. We show that even in challenging cases where the data features large jumps of the order 5 years, with observations spanning between 4 and 18 years for different pulsars and including steep red noise processes we are able to parameterise the underlying GWB signal correctly. Finally we present a method for characterising the spatial correlation between pulsars on the sky, making no assumptions about the form of that correlation, therefore providing the only truly general Bayesian method of confirming a GWB detection from pulsar timing data.

I. INTRODUCTION

Millisecond pulsars (MSPs) have for some time been known to exhibit exceptional rotational stability, with decade long observations providing timing measurements with accuracies similar to atomic clocks (e.g. [23, 29]). Such stability lends itself well to the pursuit of a wide range of scientific goals, e.g. observations of the pulsar PSR B1913+16 showed a loss of energy at a rate consistent with that predicted for gravitational waves [39], whilst the double pulsar system PSR J0737-3039A/B has provided precise measurements of several 'post Keplerian' parameters allowing for additional stringent tests of general relativity [27].

By measuring the arrival times (TOAs) of the radio pulses to high precision it is possible to construct a timing model: a deterministic model that describes the physical properties of the pulsar e.g. its binary period and spin evolution, its trajectory, post-Keplerian terms and so on. A detailed description of this process is available in the Tempo2 series of papers [8, 17, 18]. The timing model can then be subtracted from the TOAs resulting in a set of residuals that contain within them any physical effects not correctly accounted for by the timing model.

In this paper we will be concerned with extracting information from these residuals that results from time-correlated stochastic signals. These can include additional red noise terms due to rotational irregularities in the neutron star [37] or correlated noise between the pulsars due to a stochastic gravitational wave background (GWB) generated by, for example, coalescing black holes

* ltl21@cam.ac.uk

Mon. Not. R. Astron. Soc. **000**, 000–000 (0000) Printed 20 October 2014 (MN LaTeX style file v2.2)

Low-rank approximations for large stationary covariance matrices, as used in the Bayesian and generalized-least-squares analysis of pulsar-timing data

Rutger van Haasteren[1⋆], Michele Vallisneri[1]
[1]*Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91109*

printed 20 October 2014

ABSTRACT
Many data-analysis problems involve large dense matrices that describe the covariance of wide-sense stationary noise processes; the computational cost of inverting these matrices, or equivalently of solving linear systems that contain them, is often a practical limit for the analysis. We describe two general, practical, and accurate methods to approximate stationary covariance matrices as low-rank matrix products featuring carefully chosen spectral components. These methods can be used to greatly accelerate data-analysis methods in many contexts, such as the Bayesian and generalized-least-squares analysis of pulsar-timing residuals.

Key words: matrices: rank reduction – methods: data analysis – gravitational waves – pulsar timing

1 INTRODUCTION

For various applications in the mathematical theory of Gaussian processes, in the field of machine learning, and (for these authors especially) in the data analysis of timing residuals from pulsars, one needs to take a matrix inverse of the form $(N+K)^{-1}$, where N is a very large diagonal matrix and K is a very large dense matrix obtained by evaluating a wide-sense stationary correlation function k at a set of observation times t_i:

$$K_{ij} = k(t_i, t_j) = C(t_i - t_j) = C(\tau). \quad (1)$$

In the case of pulsar timing, the diagonal of the matrix N would contain the measurement errors for the times of arrival of the pulses, and $C(\tau)$ would describe correlated timing noise with a power-law spectrum $S(f)$, which is present in many pulsar-timing datasets (Shannon & Cordes 2010); by way of the Wiener–Khinchin theorem,

$$C(\tau) = \int_0^\infty \mathrm{d}f\, S(f) \cos(2\pi f \tau). \quad (2)$$

In the Bayesian analysis of pulsar-timing *residuals* (the times of arrival minus their theoretical model; see van Haasteren et al. 2009; Arzoumanian et al. 2014; van Haasteren & Vallisneri 2014, and references therein) the inverse $(N+K)^{-1}$ is required to evaluate the likelihood of the data against a Gaussian-process model of its noise-like components (van Haasteren & Vallisneri 2014); in the generalized–least-squares analysis of residuals (Coles et al. 2011), the inverse enters the normal equations of the least-squares problem.

Taking the inverse of an $n \times n$ matrix is an $O(n^3)$ operation, which becomes forbidding for large n. A popular technique to speed up the inverse is to approximate the dense matrix K as a low-rank product $F\Phi F^T$, where F is $n \times m$ (with $m \ll n$) and Φ is diagonal (and $m \times m$). The Woodbury matrix identity (Hager 1989) then yields

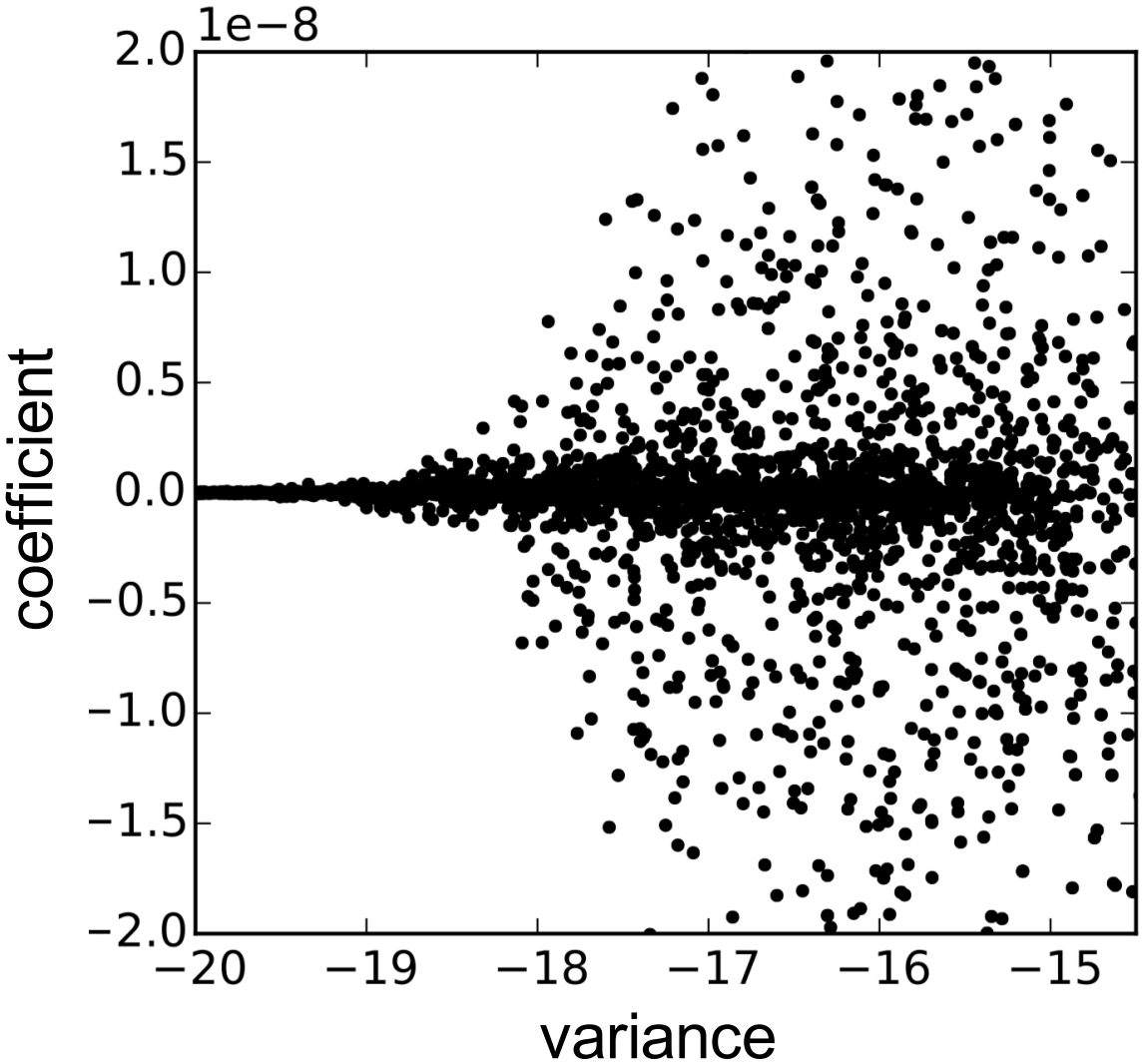
$$\begin{aligned}(N+K)^{-1} &\simeq (N + F\Phi F^T)^{-1} \\ &\simeq N^{-1} - N^{-1}F(\Phi^{-1} + F^T N^{-1} F)^{-1} F^T N^{-1},\end{aligned} \quad (3)$$

where the dominant cost is the $O(m^3)$ inversion of $\Phi^{-1} + F^T N^{-1} F$. The generalized least-squares approach (Coles et al. 2011) uses a Cholesky decomposition (Trefethen & Bau 1997) of $(N+K)$, which can also be obtained efficiently as a low-rank update (Goldfarb & Scheinberg 2004, 2005; Smola & Vishwanathan 2004). The problem then is to obtain an adequate approximation in the form $F\Phi F^T$. A straightforward solution would be the truncated singular-value decomposition (Trefethen & Bau 1997), which however is itself an $O(n^3)$ operation. Various other approximations have been discussed in the machine-learning literature (Rasmussen & Williams 2006). This paper discusses the limitations of the popular "Fourier-sum" prescription used in pulsar timing (Sec.

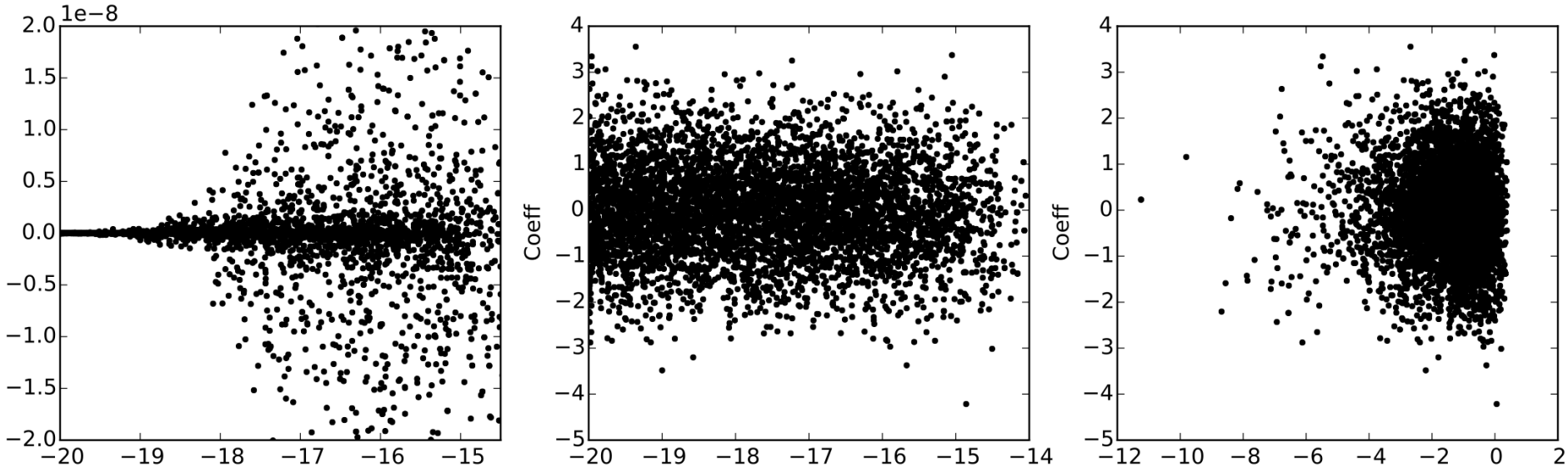
⋆ Email: vhaasteren@gmail.com

© 0000 RAS

Unfortunately, Neal's funnels kill us



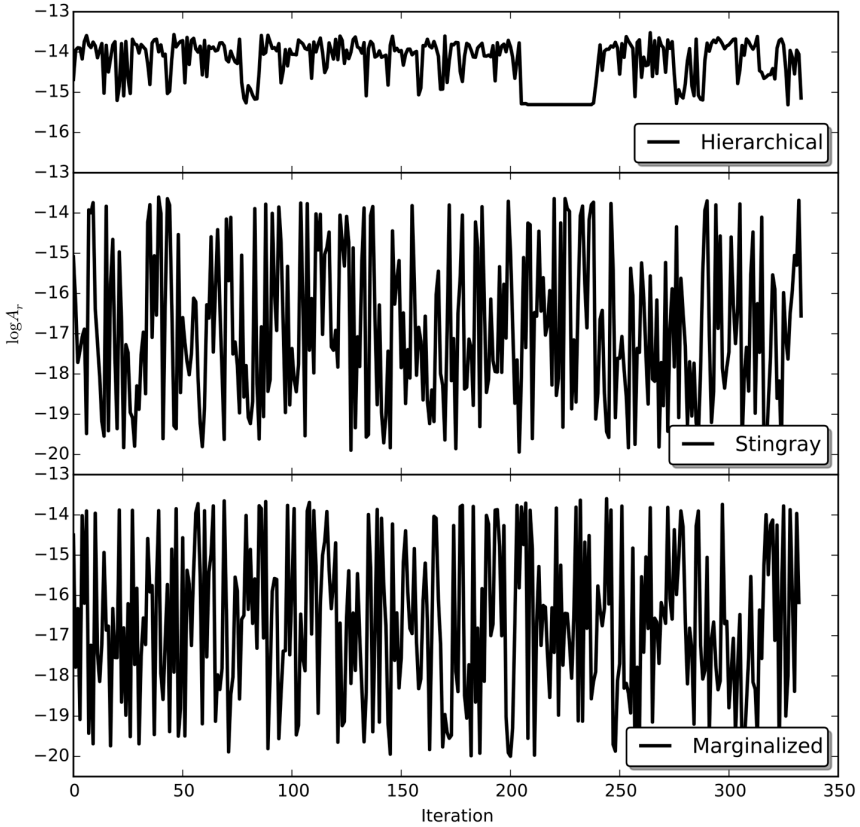
Our solution: develop *ad hoc* coordinate transformations

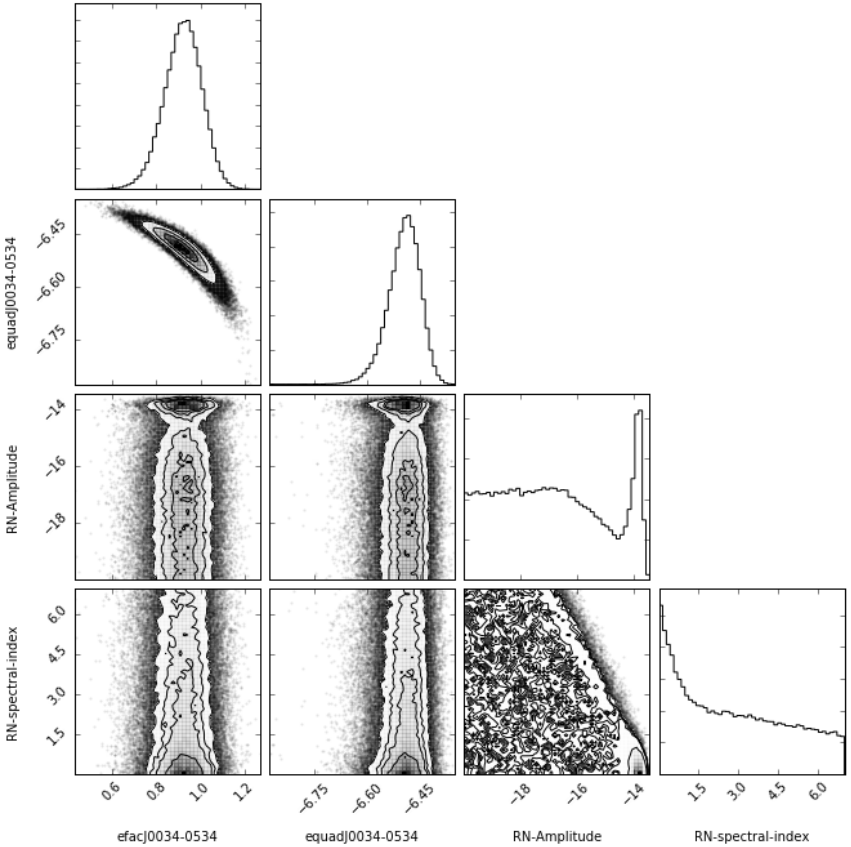


Stingray transform Interval transform

- The Stingray transform scales as Nb^3, per pulsar.
- To use HMC, transforms need to be propagated across parameter gradients (and ideally the Hessian)
- The Stingray transform involves a Cholesky decomposition. We developed a custom algorithm to compute the Cholesky derivative.

It works!


Hierarchical
Stingray
Marginalized
Iteration


equadJ0034-0534
RN-Amplitude
RN-spectral-index
efacJ0034-0534
equadJ0034-0534
RN-Amplitude
RN-spectral-index

In conclusion

- Hierarchical likelihood scales with Npsr (not Npsr^3)
- Hamiltonian Monte Carlo scales as Npar^(5/4) (not Npar^2)
- Total: Npsr^(9/4) instead of Npsr^5

- Fully working No-U-Turn-Sampler running on PTA datasets
- Requires *ad hoc* coordinate transformations:
  Regular pars > Stingray pars > Interval pars > Whitened pars

- Some NANOGrav timings (9-yr set: 37 psrs, 30 frequencies)

Hierarchical (mar jit):	88 ms	5173 dim
Transformed (mar jit):	238 ms	5173 dim
Marginalized (mar jit):	1670 ms	505 dim

- Still working on loss of efficiency when including GWB