Insights into searches for anisotropies in the nanohertz gravitational-wave background

Yacine Ali-Haïmoud[,1] Tristan L. Smith,[2] and Chiara M. F. Mingarelli[3,4]

[1]*Center for Cosmology and Particle Physics, Department of Physics, New York University, New York, New York 10003, USA*
[2]*Department of Physics and Astronomy, Swarthmore College, 500 College Ave., Swarthmore, Pennsylvania 19081, USA*
[3]*Department of Physics, University of Connecticut, 196 Auditorium Road, U-3046, Storrs, Connecticut 06269-3046, USA*
[4]*Center for Computational Astrophysics, Flatiron Institute, 162 Fifth Avenue, New York, New York 10010, USA*

Within the next several years pulsar timing arrays (PTAs) are positioned to detect the stochastic gravitational-wave background (GWB) likely produced by the collection of inspiralling supermassive black holes binaries, and potentially constrain some exotic physics. So far most of the pulsar timing data analysis has focused on the monopole of the GWB, assuming it is perfectly isotropic. The natural next step is to search for anisotropies in the GWB. In this paper, we use the recently developed PTA Fisher matrix to gain insights into optimal search strategies for GWB anisotropies. For concreteness, we apply our results to the European Pulsar Timing Array (EPTA) data, using realistic noise characteristics of its pulsars. We project the detectability of a GWB whose angular dependence is assumed to be a linear combination of predetermined maps, such as spherical harmonics or coarse pixels. We find that the GWB monopole is always statistically correlated with these maps, implying a loss of sensitivity to the monopole when searching simultaneously for anisotropies. We then derive the angular distributions of the GWB intensity to which a PTA is most sensitive, and illustrate how one may use these "principal maps" to approximately reconstruct the angular dependence of the GWB. Since the principal maps are neither perfectly anisotropic nor uncorrelated with the monopole, we also develop a frequentist criterion to specifically search for anisotropies in the GWB without any prior knowledge about their angular distribution. Lastly, we show how to recover existing EPTA results with our Fisher formalism, and clarify their meaning. The tools presented here will be valuable in guiding and optimizing the computationally demanding analyses of pulsar timing data.

I. INTRODUCTION

Pulsar timing arrays (PTAs; see, e.g., [1,2] for recent reviews) are poised to detect the nanohertz stochastic gravitational wave background (GWB) produced by the collection of inspiralling supermassive black hole binaries (SMBHB) within the next few years [3,4]. A detection of the GWB will provide a wealth of information about empirical scaling relations between galaxies and SMBH masses, and shed light on the dynamical processes that affect the evolution of SMBHBs after galaxy mergers (e.g., Refs. [5–7]). In addition, PTAs are sensitive to a range of low-frequency GWBs produced by more exotic but no less interesting processes, such as cosmic string networks, non-standard inflation, or phase transitions in the early Universe (see, e.g., Refs. [1,8,9]).

Three major PTAs have been operating for over a decade: the North American Nanohertz Observatory for Gravitational Waves (NANOGrav [10]), the European Pulsar Timing Array (EPTA [11]), and the Parkes Pulsar Timing Array (PPTA [12]). They moreover join effort within the International Pulsar Timing Array (IPTA [13]). These PTA collaborations have been setting increasingly stringent limits to the GWB amplitude, and refining their analysis methods as systematics are better understood. In the near future, the Square Kilometer Array (SKA) [14] is expected to detect hundreds of new low-noise millisecond pulsars [15], which will allow the construction a highly sensitive PTA [16]. With such large datasets, it is critical to have simple but robust and versatile forecasting tools, in order to understand the sensitivity of a PTA without having to run computationally-demanding simulations. In a previous paper [17] (hereafter, Paper I), we laid down the bases of a Fisher formalism to assess the sensitivity of a PTA to a GWB with generic frequency, angular, and polarization dependence. In

the present work, we explore in detail the angular dependence aspect, using realistic PTA noise properties.

Most PTA data analyses assume that the GWB is isotropic [18–22]. Although this is expected to be a good approximation, some amount of anisotropy should arise from the finite number of loud, nearby SMBHBs, as well as the non-uniform distribution of cosmological sources tracing large-scale structure [7,23]. Other sources of nanohertz GWs such as cosmic string networks [24–26] or domain walls [27] may also produce an anisotropic GWB. A handful of studies have looked into characterizing the sensitivity of PTAs to GWB anisotropies. References [23,28,29] computed the response of pulsar timing residual correlations to an anisotropic GWB, parametrized by its spherical-harmonic expansion. References [30,31] developed a Bayesian analysis pipeline and tested it against simulated GWBs. Lastly, Ref. [32] (hereafter TMG15) conducted a search for GWB anisotropies by performing Markov Chain Monte Carlo (MCMC) analyses of EPTA data. In order to keep computational costs manageable, they limited their analysis to a subset of 6 EPTA pulsars determined to be the most sensitive for *continuous* gravitational-wave searches [33]. TMG15 found this dataset was not informative enough to constrain anisotropies, as their upper limits saturate the physical prior which imposes a positive GWB intensity.

The main thrust of this paper is to build on the Fisher formalism derived in Paper I, in order to assess the information content of a real PTA, and in particular its sensitivity to the angular distribution of the GWB intensity. We start by laying out the basic formalism in Sec. II, where we derive the weak-signal-limit Fisher matrix for a GWB with a fixed frequency dependence, but arbitrary angular dependence. To construct this Fisher matrix for a specific PTA requires an accurate characterization of the noise properties of each pulsar. To provide a specific application of our approach as well as make contact with the analysis in TMG15, we focus on the EPTA in this paper. In Sec. III we describe how we infer the characteristic noise strains of the EPTA pulsars using publicly available timing data, and the public code HASASIA [34,35]. As a first application of the Fisher formalism, we show how one can rank-order pulsar pairs—rather than individual pulsars—by their contribution to the signal-to-noise ratio (SNR) of an isotropic GWB. In particular, in Fig. 3 we identify the best 44 EPTA pulsar pairs (out of 861 available) that contribute 90% of the SNR^2.

Furthermore we show how to use the Fisher matrix to forecast the detectability of GWB anisotropies using standard bases in Sec. IV. Specifically, we first consider a toy model of a stochastic "hot spot" in a known direction on top of an isotropic background, then forecast the sensitivity of the EPTA to the GWB amplitude in a finite number of coarse pixels, as well as to its spherical-harmonic coefficients. Importantly, we show that a PTA with $N_{\rm pair}$ pulsar pairs can constrain the projections of

the GWB on at most $N_{\rm pair}$ independent basis functions. For generic choices of bases (e.g., coarse pixels or spherical harmonics), these projections are, in general, statistically correlated with one another and with the amplitude of the GWB monopole. This implies that the sensitivity to each projection and to the monopole is systematically degraded as one expands the set of basis functions—e.g., by increasing the cutoff $\ell_{\rm max}$ in a spherical-harmonic expansion. In other words, the sensitivity to various GWB anisotropies strongly depends on the priors.

In Sec. V, we explicitly derive the $N_{\rm pair}$ statistically independent functions of direction that a PTA is most sensitive to. These "principal maps" are simply the eigenmaps of the Fisher matrix, and provide a model-independent and PTA-specific basis on which to decompose the GWB. They allow one to *search under the lamppost* provided by a given PTA. As an example application, we discuss how one may use principal maps to attempt to reconstruct the angular dependence of the GWB. We find, however, that a useful reconstruction likely requires a very large amplitude of the GWB, at least for the EPTA data we used. In Sec. VI, we propose an alternative approach to search for unknown anisotropies, in a model-independent fashion, using a frequentist criterion.

The hurried reader may want to skip ahead to Sec. VII, where we apply our formalism to the subset of 6 EPTA pulsars used in TMG15. We derive sensitivity estimates for the low-order spherical-harmonic amplitudes in good agreement with their upper limits. We also clarify the meaning of their pixel-by-pixel upper-limit map, as well as their limits on high-order spherical harmonic amplitudes, for which the number of independent parameters is larger than $N_{\rm pair}$. We prove that the parameters they constrain are not the pixel-by-pixel or spherical-harmonic amplitudes of the true GWB, but rather of its projection on the $N_{\rm pair}$-dimensional space of sky maps observable by the PTA. As a consequence, the derived limits get worse with more pulsars, and cannot be interpreted as true upper limits. We sharply demonstrate our point by closely reproducing their results using the Fisher matrix and its principal maps. This comparison with existing results ought to provide a clear demonstration of the usefulness of our Fisher formalism to forecasting the sensitivity of real PTAs to anisotropies.

We discuss our results and possible extensions in Sec. VIII. While we apply our formalism to the EPTA in this paper, we emphasize that our approach can be easily applied to any other PTA given the pulsar timing and parameter files. To this end, we will make a version of the PYTHON script used in this paper public.

II. BASIC CONVENTIONS AND FORMALISM

A. Map notation and dot product in map space

Throughout this paper, we denote by $\hat{\Omega}$ a direction in the sky (i.e., a unit-length vector). We refer to functions

$M : \hat{\Omega} \mapsto M(\hat{\Omega})$ as "maps" and denote them with bold symbols when referring to the function itself. We denote by $\mathbf{1} : \hat{\Omega} \mapsto 1$ the monopole map. Lastly, we define the dot product of two maps M_1, M_2 as follows:

$$\boldsymbol{M}_1 \cdot \boldsymbol{M}_2 \equiv \int \frac{d^2\hat{\Omega}}{4\pi} M_1(\hat{\Omega}) M_2(\hat{\Omega}). \tag{1}$$

With this convention, the monopole map has unit norm, and the average of a map M over directions is just $\mathbf{1} \cdot \boldsymbol{M}$.

B. Stochastic gravitational-wave background

We denote by $h_{ab}(f, \hat{\Omega})$ the GW strain at frequency f and with propagation direction $\hat{\Omega}$. We assume that it is a stationary Gaussian random field (as would be the case if it is generated by a large number of uncorrelated sources). It is therefore entirely determined by its rank-4 power spectrum, for which we give a complete geometric description in Paper I. Here we focus on the *total intensity* map $\mathcal{I}(f)$ of the GWB, i.e., we assume that the strain power spectrum takes the form

$$\langle h_{ab}(f, \hat{\Omega}) h_{cd}^*(f', \hat{\Omega}') \rangle = \frac{\delta_D(\hat{\Omega}', \hat{\Omega})}{4\pi} \frac{\delta_D(f' - f)}{2} \\ \times \mathcal{I}(f, \hat{\Omega}) \mathfrak{I}_{abcd}(\hat{\Omega}), \tag{2}$$

where the purely geometric (frequency-independent) rank-4 tensor $\mathfrak{I}_{abcd}(\hat{\Omega})$ is given by

$$\mathfrak{I}_{abcd}(\hat{\Omega}) \equiv \delta_{ac}^{\perp\hat{\Omega}} \delta_{bd}^{\perp\hat{\Omega}} + \delta_{ad}^{\perp\hat{\Omega}} \delta_{bc}^{\perp\hat{\Omega}} - \delta_{ab}^{\perp\hat{\Omega}} \delta_{cd}^{\perp\hat{\Omega}}, \tag{3}$$

$$\delta_{ab}^{\perp\hat{\Omega}} \equiv \delta_{ab} - \hat{\Omega}_a \hat{\Omega}_b. \tag{4}$$

We refer the reader to Paper I for the correspondence of our frame-independent expressions with the frame-dependent notation usually adopted in the literature, as well as the generalization of Eq. (2) to a linearly and circularly polarized GWB.

The GWB intensity is related to the characteristic strain $h_c(f)$ through

$$\mathbf{1} \cdot \mathcal{I} = \int \frac{d^2\hat{\Omega}}{4\pi} \mathcal{I}(f, \hat{\Omega}) = \frac{1}{2} \frac{h_c^2(f)}{f}. \tag{5}$$

Equivalently, we may write

$$\mathcal{I}(f, \hat{\Omega}) = \frac{1}{2} \frac{h_c^2(f)}{f} P(f, \hat{\Omega}), \tag{6}$$

where $\boldsymbol{P}(f)$ is the (frequency-dependent) angular distribution of GWB intensity, with unit angle average.

C. Pulsar timing residuals correlations

Consider a pulsar p, in the direction $\hat{p}$. The times of arrival (TOAs) of its pulses are a combination of a deterministic term due to intrinsic processes (such as the spin-down of the pulsar or variations in the position of the pulsar on the sky), a pulsar-specific intrinsic noise term N_p, uncorrelated between different pulsars (due to measurement noise, variations of the dispersion measure, as well as any intrinsic stochastic processes, such as pulsar "jitter"), and additional stochastic contributions correlated among pulsars. In this paper, we assume that the latter are exclusively generated by the GWB, and neglect additional sources of correlated noise, such as global clock and ephemeris errors [19,36,37]. After fitting for a deterministic timing model, the pulsar time residual R_p is thus

$$R_p = N_p + R_p^{\text{GW}}. \tag{7}$$

We define the (one-sided) cross-power spectrum $\mathcal{R}_{pq}(f)$ of the total time residual at pulsars p, q as follows:

$$\langle R_p(f) R_q^*(f') \rangle = \frac{\delta_D(f' - f)}{2} \mathcal{R}_{pq}(f), \tag{8}$$

and similarly define $\mathcal{R}_{pq}^{\text{GW}}(f)$ as the cross-power spectrum of the GW-induced timing residuals R_p^{GW}. The cross-power spectrum of the intrinsic noise is defined as

$$\langle N_p(f) N_q^*(f') \rangle = \frac{\delta_D(f' - f)}{2} \delta_{pq} \sigma_p^2(f). \tag{9}$$

In Paper I we derived the following simple compact form for the linear relation between $\mathcal{R}_{pq}^{\text{GW}}(f)$ and $\mathcal{I}(f)$:

$$\mathcal{R}_{pq}^{\text{GW}}(f) = \frac{1 + \delta_{pq}}{(4\pi f)^2} \boldsymbol{\gamma}_{\hat{p}\hat{q}} \cdot \mathcal{I}(f), \tag{10}$$

where the *pairwise timing response function*, $\boldsymbol{\gamma}_{\hat{p}\hat{q}}$, is the map with values

$$\gamma_{\hat{p}\hat{q}}(\hat{\Omega}) \equiv 2 \frac{(\hat{p} \cdot \hat{q} - (\hat{p} \cdot \hat{\Omega})(\hat{q} \cdot \hat{\Omega}))^2}{(1 + \hat{p} \cdot \hat{\Omega})(1 + \hat{q} \cdot \hat{\Omega})} \\ - (1 - \hat{p} \cdot \hat{\Omega})(1 - \hat{q} \cdot \hat{\Omega}). \tag{11}$$

Using Eq. (6), we may rewrite Eq. (10) as

$$\mathcal{R}_{pq}^{\text{GW}}(f) = \Gamma_{pq}(f) \frac{h_c^2(f)}{12\pi^2 f^3}, \tag{12}$$

where

$$\Gamma_{pq}(f) \equiv \frac{3}{8}(1 + \delta_{pq})\boldsymbol{\gamma}_{\hat{p}\hat{q}} \cdot \boldsymbol{P}(f) \tag{13}$$

is the frequency-dependent "overlap reduction function."

If the GWB is a pure monopole, the timing residual cross correlation is proportional to the Hellings and Downs function [38], and only depends on the dot product $\hat{p} \cdot \hat{q}$ through the function

$$\mathcal{H}(\hat{p} \cdot \hat{q}) \equiv \boldsymbol{\gamma}_{\hat{p}\hat{q}} \cdot \mathbf{1}. \tag{14}$$

Specifically, this function is given by, for $\mu \equiv \hat{p} \cdot \hat{q}$,

$$\mathcal{H}(\mu) \equiv \frac{3 + \mu}{3} + 2(1 - \mu) \ln\left(\frac{1 - \mu}{2}\right). \tag{15}$$

Note that we do not impose any normalization on the Hellings and Downs function $\mathcal{H}(\mu)$: it is simply defined as the response to an isotropic GWB intensity with unit amplitude, through Eq. (14). In particular, $\mathcal{H}(1) = 4/3$.

D. Gaussian likelihood for the GWB intensity

In what follows we use capital indices to label unique pairs of distinct pulsars, e.g., $I \equiv (p, q) = (q, p)$ for $p \neq q$. We then define $\mathcal{R}_I \equiv \mathcal{R}_{pq}$, $\gamma_I \equiv \gamma_{\hat{p}\hat{q}}$, and $\mathcal{H}_I \equiv \mathcal{H}(\hat{p} \cdot \hat{q})$.

Let us define $\hat{\mathcal{R}}_I(f)$ to be unbiased estimators of the timing residual cross-power spectra at a given frequency f. Assuming that the only source of correlated timing residuals for distinct pulsars is the GWB, these estimators have mean value $\langle \hat{\mathcal{R}}_I \rangle = \mathcal{R}_I^{\mathrm{GW}}$. In practice, one cannot directly construct finite-variance estimators of the cross-power spectra, but only of finite bandpowers, from which one may then extract $\hat{\mathcal{R}}_I(f)$ for sufficiently narrow bands, see Paper I for details. Mathematically, the covariance matrix of the estimators $\mathcal{C}_{IJ} \equiv \mathrm{cov}(\hat{\mathcal{R}}_I(f), \hat{\mathcal{R}}_J(f))$ is inversely proportional to the bandwidth Δf used to estimate the bandpowers: we denote it as

$$\mathcal{C}_{IJ} = \frac{1}{\Delta f}\mathfrak{C}_{IJ}, \tag{16}$$

where both $\mathcal{C}_{IJ}$ and $\mathfrak{C}_{IJ}$ depend on frequency.

Estimators of $\mathcal{R}_I$ at frequencies separated by more than $\sim 1/T$ are uncorrelated, where T is the characteristic observation time. If the estimators are constructed from a sufficiently large number of uncorrelated data samples, we may approximate their probability distribution $\mathcal{P}$ as the product of uncorrelated multivariate Gaussians:

$$\mathcal{P}(\{\hat{\mathcal{R}}_I\}|\mathcal{I}) \propto \prod_f \exp\left[-\frac{1}{2}\Delta f \sum_{IJ}(\hat{\mathcal{R}}_I(f) - \mathcal{R}_I^{\mathrm{GW}}(f))\right.$$
$$\left. \times \mathfrak{C}_{IJ}^{-1}(\hat{\mathcal{R}}_J(f) - \mathcal{R}_J^{\mathrm{GW}}(f))\right].$$
$$\approx \exp\left[-\frac{1}{2}\int df \sum_{IJ}(\hat{\mathcal{R}}_I(f) - \mathcal{R}_I^{\mathrm{GW}}(f))\right.$$
$$\left. \times \mathfrak{C}_{IJ}^{-1}(\hat{\mathcal{R}}_J(f) - \mathcal{R}_J^{\mathrm{GW}}(f))\right], \tag{17}$$

where the second approximation holds provided the frequency bands are narrow, $\Delta f \ll f$.

As is standard in Bayesian data analysis, one may interpret this probability distribution as the *likelihood* $\mathcal{L}$ of the signal $\mathcal{I}(f)$ given the data (assuming flat priors):

$$\mathcal{L}(\mathcal{I}) \propto \mathcal{P}(\{\hat{\mathcal{R}}_I\}|\mathcal{I}). \tag{18}$$

It is useful to write this probability explicitly as a Gaussian distribution for $\mathcal{I}$. To do so, we first define the maps γ_I^*, as the unique linear combinations of the N_{pair} pairwise timing response functions that satisfy $\gamma_I^* \cdot \gamma_J = \delta_{IJ}$. These are the *dual maps* (not to be confused with complex conjugates) of the pairwise timing response functions. We then define the following estimator of the intensity map:

$$\hat{\mathcal{I}}(f) \equiv (4\pi f)^2 \sum_I \hat{\mathcal{R}}_I(f)\gamma_I^*. \tag{19}$$

We may then rewrite the likelihood (18) in the form

$$\mathcal{L}(\mathcal{I}) \propto \exp\left[-\frac{1}{2}\int df(\mathcal{I}(f) - \hat{\mathcal{I}}(f)) \cdot \mathcal{G} \cdot (\mathcal{I}(f) - \hat{\mathcal{I}}(f))\right], \tag{20}$$

where $\mathcal{G}(f, \hat{\Omega}, \hat{\Omega}')$ is a continuous inverse covariance "matrix," given by

$$\mathcal{G}(f, \hat{\Omega}, \hat{\Omega}') = \frac{1}{(4\pi f)^4} \sum_{I,J} \gamma_I(\hat{\Omega})\mathfrak{C}_{IJ}^{-1}\gamma_J(\hat{\Omega}'). \tag{21}$$

E. Weak-signal limit

In Paper I we derived the weak-signal limit of $\mathfrak{C}_{IJ}$:

$$\mathfrak{C}_{IJ} \approx \frac{1}{2T_{pq}} \frac{\sigma_p^2(f)\sigma_q^2(f)}{\mathcal{T}_p(f)\mathcal{T}_q(f)}\delta_{IJ}, \qquad I \equiv (p, q), \tag{22}$$

where $\mathcal{T}_p(f)$ is the transmission function, which accounts for the loss of information when fitting a deterministic timing model [39], and $T_{pq} \equiv \min(T_p, T_q)$ is the minimum observation time of the pulsar pair (p, q) (or more generally, the time of overlap between the two pulsar observations). This limit holds if the GWB-induced timing

residual is subdominant to the intrinsic timing noise for each pulsar, and at every frequency, i.e., if

$$\mathcal{I}(f) \ll (4\pi f)^2 \sigma_p^2(f), \quad \forall\, p, \quad \forall\, f. \qquad (23)$$

In what follows we adopt the compact notation of Ref. [39] and define the characteristic noise strain for pulsar p as

$$h_{c,p}^2(f) \equiv \frac{3}{4} \frac{(4\pi f)^2 f \sigma_p^2(f)}{\mathcal{T}_p(f)}. \qquad (24)$$

The factor of 3/4 arises from the Hellings and Downs normalization convention in Ref. [39]; we chose to use exactly the same definition for $h_{c,p}$ as in that paper to avoid any confusion. With this notation, we find the following expression for the weak-limit intensity inverse-covariance:

$$\mathcal{G}(f,\hat{\Omega},\hat{\Omega}') \approx \sum_I \mathcal{G}_I(f)\gamma_I(\hat{\Omega})\gamma_I(\hat{\Omega}'), \qquad (25)$$

where, for $I = (p,q)$, we have defined

$$\mathcal{G}_I(f) \equiv \frac{9}{8} \frac{f^2 T_{pq}}{h_{c,p}^2(f) h_{c,q}^2(f)}. \qquad (26)$$

F. Application to a GWB with a power-law frequency dependence

It is of course easier to search for signals with a specific frequency dependence. In the remainder of this paper, we will specialize to a GWB whose frequency and direction dependence factorize, and with a power-law frequency dependence, of the form

$$\mathcal{I}(f,\hat{\Omega}) = \frac{\mathcal{A}(\hat{\Omega})}{2f}(f/f_{\mathrm{yr}})^{-2\alpha}, \qquad \alpha > 0, \qquad (27)$$

where $f_{\mathrm{yr}} \equiv (1\ \mathrm{year})^{-1}$. With this convention, and as can be seen from Eq. (5) the characteristic strain takes the form $h_c(f) = A_h (f/f_{\mathrm{yr}})^{-\alpha}$, where $A_h^2 \equiv \mathbf{1} \cdot \mathcal{A}$ is the angle average of $\mathcal{A}$.

While we keep our expressions general, for numerical applications we will specialize to a power law $\alpha = 2/3$. Indeed, this is what is expected from a population of SMBHBs on circular orbits, shrinking exclusively due to GW emission [40]. The true frequency dependence may differ from this simple form, especially at low-frequencies if the binaries are eccentric or if environmental effects such as stellar hardening and/or gas torques are driving the binaries to merge instead of GWs [1,5,41]. Note that our formalism can readily be applied to an arbitrary frequency dependence.

We now want to rewrite Eq. (20) as a likelihood for $\mathcal{A}$. To do so, we define the dimensionless Fisher "matrix"

$$\mathcal{F}(\hat{\Omega},\hat{\Omega}') \equiv \sum_I \mathcal{F}_I \gamma_I(\hat{\Omega})\gamma_I(\hat{\Omega}'), \qquad (28)$$

$$\mathcal{F}_I \equiv \int df \frac{(f/f_{\mathrm{yr}})^{-4\alpha}}{4f^2} \mathcal{G}_I(f), \qquad (29)$$

as well as the inverse-variance-weighted estimator amplitude map

$$\hat{\mathcal{A}} \equiv \sum_I \hat{\mathcal{A}}_I \gamma_I^*, \qquad (30)$$

$$\hat{\mathcal{A}}_I \equiv \frac{1}{\mathcal{F}_I} \int df \frac{(f/f_{\mathrm{yr}})^{-2\alpha}}{2f} \mathcal{G}_I(f)(4\pi f)^2 \hat{\mathcal{R}}_I(f). \qquad (31)$$

The likelihood of the GWB amplitude $\mathcal{A}$ is then proportional to

$$\mathcal{L}(\mathcal{A}) \propto \exp\left[-\frac{1}{2}(\mathcal{A} - \hat{\mathcal{A}}) \cdot \mathcal{F} \cdot (\mathcal{A} - \hat{\mathcal{A}})\right]. \qquad (32)$$

Equations (28) and (32) are fundamental to all the forthcoming calculations in this work.

The total SNR^2 for an arbitrary GWB with amplitude $\mathcal{A}$ is given by

$$\mathrm{SNR}^2 = \mathcal{A} \cdot \mathcal{F} \cdot \mathcal{A} = \sum_I \mathcal{F}_I [\gamma_I \cdot \mathcal{A}]^2. \qquad (33)$$

We see that for each pulsar pair, the coefficient $\mathcal{F}_I$ characterizes its overall sensitivity to a power-law GWB, and the pairwise-timing response function γ_I characterizes its angular response.

Before moving on, let us remark that, given N_{pair} distinct pulsar pairs, there are at most N_{pair} independent detectable components of the GWB. In particular, a PTA is completely blind to the projection of the GWB perpendicular to all the pairwise timing response functions. Specifically, given an amplitude $\mathcal{A}$, we may always decompose it as a piece $\mathcal{A}_{\parallel}$ which is a linear combination of the γ_I's, and a piece $\mathcal{A}_{\perp}$ orthogonal to all of them; the latter is completely undetectable by the PTA. The estimator $\hat{\mathcal{A}}$ generalizes the "optimal statistic" $\hat{A}_{\mathrm{GWB}}^2$ [19,42] to an anisotropic GWB. It is an unbiased estimator of the *observable* piece of the signal $\mathcal{A}_{\parallel}$, but of course not of the full map $\mathcal{A}$. We can therefore equally write $\hat{\mathcal{A}} \equiv \hat{\mathcal{A}}_{\parallel}$.

G. Contribution of anisotropies to timing residual autocorrelations

The GWB also contributes to each pulsar's timing residual autocorrelation (or power spectrum). These

autocorrelations differ from the cross correlations in an essential aspect: one *does* expect a nonzero autocorrelation even in the absence of a GWB, due to intrinsic noise. In the analysis of real data, one must therefore search simultaneously for an appropriately parametrized intrinsic noise spectrum, for each pulsar, jointly with a GWB-induced noise common to all pulsars (a "common red noise"). Because of the unknown intrinsic noise, one therefore cannot hope to ever achieve a *detection* of the GWB with autocorrelations alone [43]. Still, autocorrelations do contain information about the GWB: at the very least, they can be used to set upper limits on its amplitude.

Here we expand on a point that we made in Paper I: a pulsar's autocorrelation term is sensitive to the GWB anisotropy. In other words, GWB anisotropies contribute a common red noise. While this fact can be readily inferred from the expressions for the overlap reduction functions given in Ref. [23], it is unclear whether this important point has been accounted for in previous searches for anisotropic GWB. Moreover, here we show that only the GWB monopole, dipole and quadrupole contribute to pulsars' autocorrelations.

From Eqs. (10) and (27), the GWB contribution to the single-pulsar timing residual power spectrum is proportional to $\gamma_{\hat{p}\hat{p}} \cdot \mathcal{A}$. As we saw in Paper I, the single-pulsar timing response function $\gamma_{\hat{p}\hat{p}}(\hat{\Omega})$ is a linear combination of a monopole, dipole projected along $\hat{p}$, and quadrupole twice projected onto $\hat{p}$:

$$\gamma_{\hat{p}\hat{p}}(\hat{\Omega}) = \frac{4}{3} - 2\hat{p} \cdot \hat{\Omega} + \hat{p}^i \hat{p}^j \left(\hat{\Omega}^i \hat{\Omega}^j - \frac{1}{3} \delta^{ij} \right). \quad (34)$$

Let us define the following geometric decomposition of the GWB dimensionless amplitude:

$$\mathcal{A}(\hat{\Omega}) = \mathcal{A}_0 + 3 \sum_{i=1}^{3} \mathcal{A}_1^i \hat{\Omega}^i + \frac{15}{2} \sum_{i,j=1}^{3} \mathcal{A}_2^{ij} \left(\hat{\Omega}^i \hat{\Omega}^j - \frac{1}{3} \delta^{ij} \right)$$
$$+ \Delta\mathcal{A}(\hat{\Omega}), \quad (35)$$

where the tensor $\mathcal{A}_2^{ij}$ is symmetric and trace-free (thus has 5 independent components), and $\Delta\mathcal{A}$ only contains contributions that are octupolar and higher order. The components $\mathcal{A}_0, \mathcal{A}_1^i$, and $\mathcal{A}_2^{ij}$ are uniquely defined through

$$\mathcal{A}_0 \equiv \mathcal{A} \cdot \mathbf{1}, \qquad \mathcal{A}_1^i \equiv \int \frac{d^2\hat{\Omega}}{4\pi} \mathcal{A}(\hat{\Omega}) \hat{\Omega}^i, \quad (36)$$

$$\mathcal{A}_2^{ij} \equiv \int \frac{d^2\hat{\Omega}}{4\pi} \mathcal{A}(\hat{\Omega}) \left(\hat{\Omega}^i \hat{\Omega}^j - \frac{1}{3} \delta^{ij} \right). \quad (37)$$

They are, respectively, the monopole, dipole, and quadrupole components of the GWB. With this decomposition, we obtain

$$\gamma_{\hat{p}\hat{p}} \cdot \mathcal{A} = \frac{4}{3} \mathcal{A}_0 - 2 \sum_{i=1}^{3} \hat{p}^i \mathcal{A}_1^i + \sum_{i,j=1}^{3} \hat{p}^i \hat{p}^j \mathcal{A}_2^{ij}. \quad (38)$$

Thus, as noted in Paper I, the autocorrelated power spectra constrain specific combinations of the GWB monopole, dipole, and quadrupole. Importantly, when allowing for anisotropies, one must self-consistently propagate their impact on the single-pulsar noises, i.e., one must account for "common red processes" which arise from the dipole and quadrupole pieces of the GWB, in addition to its monopole. If autocorrelations are used to set upper limits on a common red process, their dependence on dipole and quadrupole GWB anisotropies systematically degrades the inferred upper limits on the monopole.

One of our main objectives here is to explore the ability of PTAs to *detect* the GWB. While autocorrelations do contain information, they cannot be used to claim a detection [43]. As a consequence we focus on the correlations between different pulsars in the remainder of this paper.

III. DATASET USED IN THIS PAPER

Throughout most of this paper, to provide a concrete example, we will apply our formalism to the full EPTA dataset described in the last data release [44]. The EPTA is composed of the high-precision timing of 42 millisecond pulsars for up to 17 years with an overlap of 21 pulsars with the NANOGrav Nine-Year dataset [45]. We show the locations of the EPTA pulsars relative to the Galactic plane in Fig. 1.

Out of the 42 EPTA pulsars, six were determined to contribute 90% of the SNR^2 in simulated *continuous* GW searches [33], and were used for searches of anisotropies in the GWB in Ref. [32] (hereafter, TMG15). Specifically, these six pulsars are $J0613 - 0200$, $J1012 + 5307$, $J1600 - 3053$, $J1713 + 0747$, $J1744 - 1134$, and $J1909 - 3744$. Their
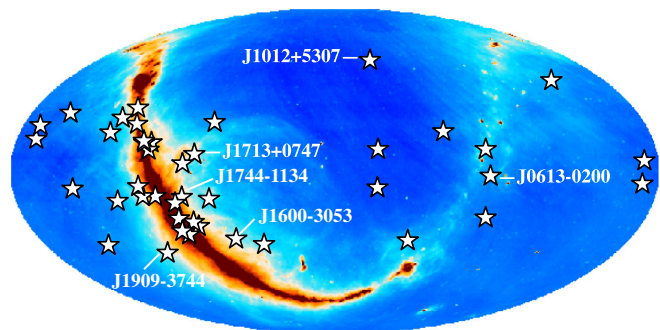

FIG. 1. Locations of the EPTA pulsars, in equatorial coordinates. We use the full EPTA dataset in Secs. III–VI. The subset of 6 highlighted pulsars is used when compare our results to those of Ref. [32] in Sec. VII. The background map is the Stockert and Villa-Elisa 1.4 GHz continuum map [46,47], which situates the positions of the pulsars relative to the Galactic plane.
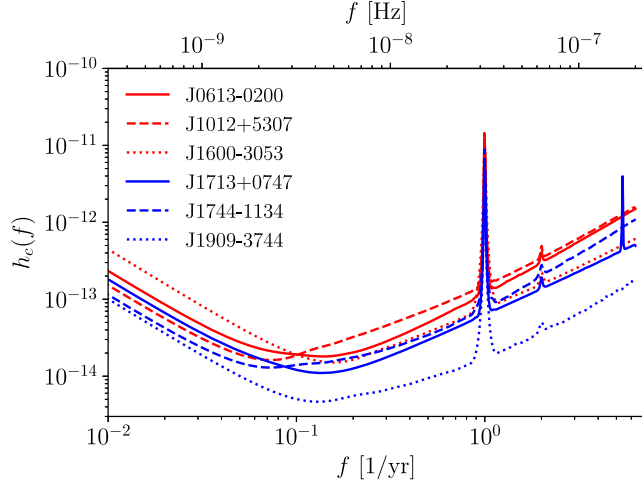
FIG. 2. Characteristic noise strains for 6 of the EPTA pulsars, produced using publicly available EPTA data [48] and HASASIA [34,35], see main text for details. The spikes at 1/yr and 2/yr arise from degeneracies with the Earth's orbital motion when fitting the pulsar position and distance. For J1713−0747, the spike at $f \approx 5.4$/yr is at the 67.8-day period of this binary system.

locations in the sky are shown in Fig. 1, and their characteristic noise strains in Fig. 2. We will use this subset of the EPTA in Sec. VII to compare our sensitivity estimates to those of TMG15.

A. Evaluation of the characteristic noise strains

As described in the previous section, the primary quantity characterizing pulsar noise properties is the characteristic noise strain. Here, we explain how we estimate the characteristic noise strains of all 42 EPTA pulsars.

We model the intrinsic white noise in the standard way:

$$[\langle \delta R_p(t_i) \delta R_p(t_j) \rangle]_{\text{white}} = (\text{EFAC}_{p,b}^2 \sigma_{p,i}^2 + \text{EQUAD}_{p,b}^2)\delta_{ij},$$

(39)

where t_i is the ith time-of-arrival of pulsar p, $\sigma_{p,i}$ is the uncertainty in the ith timing residual of pulsar p, and the white noise parameters, $\text{EFAC}_{p,b}$ and $\text{EQUAD}_{p,b}$, are included for each observing system b (i.e., different telescope and/or backends). EFAC is dimensionless and EQUAD has dimensions of time (we use units of seconds for our analysis). Our priors on these parameters were a flat-linear prior on $\text{EFAC} \in [0.01, 10]$ and a flat-log prior on $\text{EQUAD} \in [10^{-8.5}, 10^{-3}]$ sec. The effects of variations in the dispersion measure (DM) and intrinsic red-noise (RN) were modeled with a power-law spectral density of the form

$$\sigma_{\text{p,X}}^2(f) = \frac{A_{\text{X}}^2}{12\pi^2}(f/f_{\text{yr}})^{-\gamma_{\text{X}}} \text{ yr}^3, \quad \text{X} = \text{RN or DM}, \quad (40)$$

where A_{RN} is independent of radio frequency ν whereas $A_{\text{DM}} \propto \nu^{-2}$. We use a flat-linear prior on the power-law indices, $\gamma_{\text{X}} \in [0, 7]$ and a flat-log prior on the dimensionless amplitudes $A_{\text{X}} \in [10^{-20}, 10^{-11}]$, for both RN and DM. The prior on the power-law indices have been chosen so that they range from white noise ($\gamma_{\text{X}} = 0$) to the steepest power-law for which the fit to the timing model removes any dependence on the functional form for these spectral densities at low frequencies ($f < 1/T_{\text{obs}}$) [49]. This range also covers the expected variation in the power-law index due to random walks in phase, period and period derivatives (which give $\gamma_{\text{RN}} = 2, 4, 6$, respectively) [19]. Note that we *do not* include a "common red noise" process in our analysis. This is consistent with our weak-signal assumption, and allows us to fit for each pulsar noise properties independently.

In order to extract values for these noise parameters we used the parallel tempering Markov-Chain Monte Carlo sampler PTMCMCSampler [50], the PTA software ENTERPRISE [51], and the TOA and timing model parameters used in the EPTA data release [48] from the EPTA repository [52]. We set the noise parameters to equal their median values from MCMC chains that contain 10^6 samples. We have confirmed that our noise parameters match well with the noise analysis described in [53].

After we have estimated the noise parameters for each pulsar we use HASASIA [34,35] to compute the total noise power spectrum $\sigma_p^2(f)$, transmission function $\mathcal{T}_p(f)$, and characteristic noise strain $h_{c,p}(f)$, for each one of the EPTA pulsars. As an illustration, we show the characteristic noise strains of 6 of the EPTA pulsars in Fig. 2. We will use these specific 6 pulsars when comparing of our results with those of TMG15 in Sec. VII.

B. Best pulsar pairs for isotropic GWB searches

In what follows we determine which pulsar pairs contribute most to the total SNR^2 from a cross-correlation analysis. We will see that only a small number of pulsar pairs are needed to get most of the SNR^2. A similar analysis was performed in Ref. [54] for the Parkes PTA; here we consider the EPTA and provide a much simpler, analytic calculation.

As can be seen from Eq. (33), for a general GWB, each pair I contributes $\mathcal{F}_I(\gamma_I \cdot \mathcal{A})^2$ to the SNR^2. The relative ranking of pulsar pairs thus depends on the angular dependence of the GWB, and there is no universal ranking that holds for an arbitrary GWB. Since we expect the GWB to be predominantly isotropic, it is sensible to rank pulsar pairs under this assumption.

For an isotropic GWB $\mathcal{A}(\hat{\Omega}) = A_h^2$, from Eq. (33) we see that the SNR^2 is given by

$$\text{SNR}^2 = A_h^4 \sum_I \mathcal{F}_I \mathcal{H}_I^2.$$

(41)

First, to check the soundness of our analysis, we estimate the 95% sensitivity to a monopole, $A_h^{95\%}$, such that $\mathrm{SNR}[A_h^{95\%}] = 2$. Using Eq. (41), we find

$$A_h^{95\%} = \sqrt{2}\left(\sum_I \mathcal{F}_I \mathcal{H}_I^2\right)^{-1/4}. \qquad (42)$$

Using the characteristic noise strains we computed for the EPTA dataset, we find $A_h^{95\%} \approx 2.5 \times 10^{-15}$. If we restrict ourselves to the 6 EPTA pulsars used in Refs. [20,32], we find instead $A_h^{95\%} \approx 3.4 \times 10^{-15}$. This lies between the 95% upper-limit values found in these references, of 3.0×10^{-15} [20] and 3.9×10^{-15} [32]. This gives us confidence that our characteristic noise strains provide a realistic description of the data.

From Eq. (41), we see that for an isotropic GWB, each pair $I = (p, q)$ contributes $A_h^4 \mathcal{F}_I \mathcal{H}_I^2$ to the SNR^2. We have ranked the 861 pulsar pairs of the full EPTA dataset in terms of their contributions to the SNR^2. In Fig. 3 we show the normalized cumulative contributions of the best 44 pulsar pairs, which contributed 90% of the SNR^2. While this is three times as many pairs as what can be constructed with the 6 pulsars used in Refs. [20,32], these 44 pairs represent 5% of the total number of pairs in the full dataset, and should still constitute a manageable collection of data. Moreover, we find that the 6 pulsars used in Refs. [20,32] only contribute 26% of the total SNR^2 for an isotropic GWB, while the best 15 pairs would amount to 68% of the SNR^2, as can be seen in Fig. 3.

This result is of significant importance to speed up future analyses of real data. It is well known that accounting for all the correlations between pulsar pairs in a full Bayesian analysis of timing residuals is computationally challenging. Here we see that it suffices to include a small, manageable number of pulsar pairs to recover most of the SNR^2. Our simple Fisher formalism allows us to efficiently determine which pairs to use for any given dataset. This can be done not only for an isotropic GWB, but also for any assumed angular dependence, if desired.

IV. SENSITIVITY TO GWB ANISOTROPIES USING STANDARD BASES

A. General approach

Suppose that we model the GWB amplitude to be a linear combination of known maps $\boldsymbol{M}_n$:

$$\mathcal{A} = \sum_n \mathcal{A}_n \boldsymbol{M}_n, \qquad (43)$$

where $\mathcal{A}_n$ are scalar amplitudes. The maps $\boldsymbol{M}_n$ could be, for instance, spherical harmonics, or Dirac functions centered at specific directions in the sky. They can be, in general, any set of linearly independent maps, and need not even be orthogonal nor normalized. We can always define the dual maps $\boldsymbol{M}_n^*$ (not to be confused with complex conjugates), such that $\boldsymbol{M}_n \cdot \boldsymbol{M}_m^* = \delta_{mn}$. Given the estimator map $\hat{\boldsymbol{A}}$, we may then define $\hat{\mathcal{A}}_n \equiv \boldsymbol{M}_n^* \cdot \hat{\boldsymbol{A}}$. The probability distribution $\mathcal{L}(\mathcal{A})$ given in Eq. (32) can then be interpreted as a joint
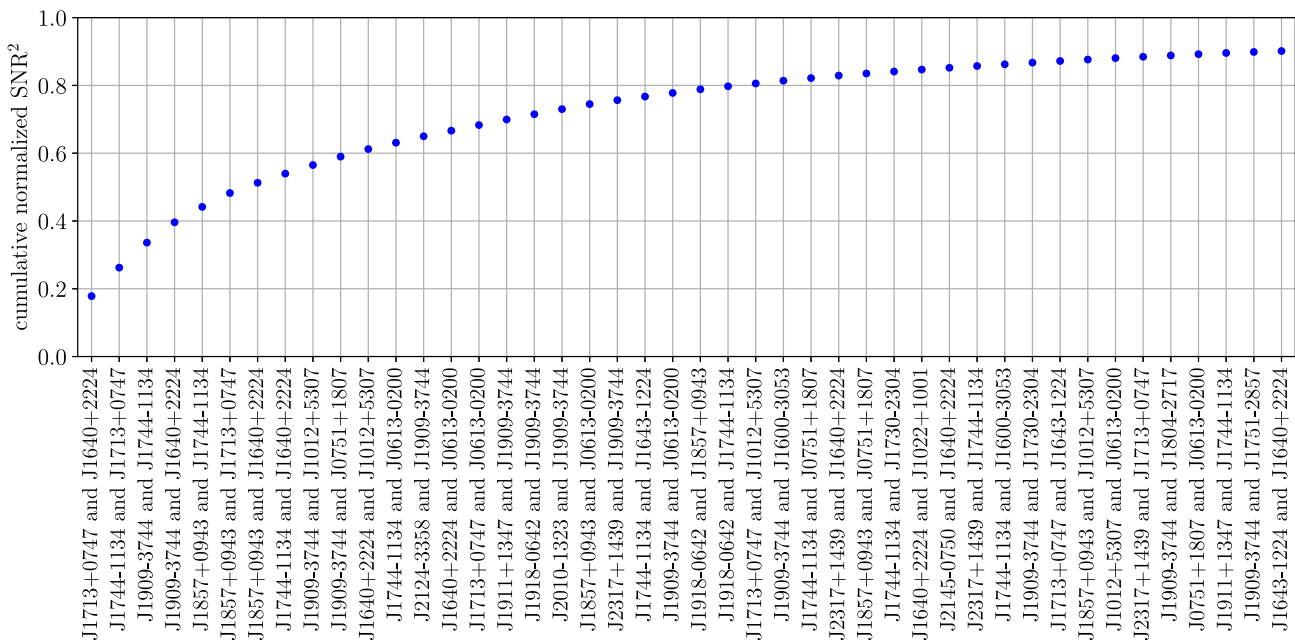

FIG. 3. Normalized cumulative SNR^2 for an isotropic GWB with characteristic strain $h_c(f) \propto f^{-2/3}$. The total SNR^2 is computed with the full EPTA array of 42 pulsars [44], corresponding to 861 distinct pairs. This figure only shows the contributions of the best 44 pairs, labeling the x-axis, required to reach 90% of the SNR^2.

probability distribution for the set of coefficients $\{\mathcal{A}_n\}$, and be recast in the form

$$\mathcal{P}(\{\mathcal{A}_n\}) \propto \exp\left[-\frac{1}{2}\sum_{mn}(\mathcal{A}_m - \hat{\mathcal{A}}_m)F_{mn}(\mathcal{A}_n - \hat{\mathcal{A}}_n)\right], \tag{44}$$

$$F_{mn} \equiv \boldsymbol{M}_m \cdot \boldsymbol{\mathcal{F}} \cdot \boldsymbol{M}_n. \tag{45}$$

In other words, the coefficients $\mathcal{A}_n$ are Gaussian-distributed, with mean $\hat{\mathcal{A}}_n$, and covariance

$$\mathrm{cov}(\mathcal{A}_n, \mathcal{A}_m) = (F^{-1})_{mn}. \tag{46}$$

In particular, upon marginalizing over all other parameters, the probability distribution of any single parameter $\mathcal{A}_n$ is a Gaussian with mean $\hat{\mathcal{A}}_n$ and variance $(F^{-1})_{nn}$. This means that, to be able to claim a 95%-confidence detection of a nonzero coefficient $\mathcal{A}_n$, the underlying true parameter needs to satisfy $\mathcal{A}_n \geq \mathcal{A}_n^{95\%}$, where

$$\mathcal{A}_n^{95\%} \equiv 2\sqrt{(F^{-1})_{nn}}. \tag{47}$$

This represents a threshold for *detection*, i.e., allows us to forecast the sensitivity of a given PTA. Note that $(F^{-1})_{nn} \geq 1/F_{nn}$, unless the map $\boldsymbol{M}_n$ is statistically uncorrelated with all other maps. Therefore the sensitivity to any given map amplitude $\mathcal{A}_n$ is always degraded if one simultaneously searches for other amplitudes.

It is important to note that at most N_{pair} map amplitudes can be simultaneously searched for. Indeed, the Fisher matrix $\boldsymbol{\mathcal{F}}$ only has rank N_{pair}, as can be seen by writing Eq. (28) formally as a sum of tensor products $\boldsymbol{\mathcal{F}} = \sum_I \mathcal{F}_I \gamma_I \otimes \gamma_I$. Therefore, the matrix F_{mn} is guaranteed to be singular if it has more than N_{pair} rows and columns. In other words, at most N_{pair} parameters can have finite variance. In practice, one can still set upper limits even with more than N_{pair} parameters, if additional priors are included (e.g., physical priors, or a prior based on the auto-correlation functions of individual pulsars). But in order to be able to claim a detection, at most N_{pair} parameters can be simultaneously searched for.

B. A toy problem: Monopole and hot spot with a known direction

As a first application of this formalism, let us consider a simple toy problem, with only two basis maps. Suppose that external considerations lead us to expect a GWB "hot spot" in known direction $\hat{\Omega}_*$, in a addition to a background monopole. Note that we are still considering a *stochastic* "hot spot", i.e., this cannot be generated by a single binary, but rather must originate from a concentration of many

binaries radiating incoherently. Both the hot spot and monopole have an unknown amplitude.

Mathematically, we assume that the GWB amplitude takes the form

$$\mathcal{A}(\hat{\Omega}) = \mathcal{A}_0 + 4\pi\mathcal{A}_1\delta_{\mathrm{D}}(\hat{\Omega};\hat{\Omega}_*). \tag{48}$$

Note that this is a different split between monopole and hot-spot than adopted in Ref. [29] and Paper I; with the convention (48), the amplitude of the hot spot is unbounded from above. One can interpret $\mathcal{A}_0$ as the amplitude of a background monopole, on top of which the hot spot sits. The total monopole amplitude is then $\mathcal{A}_0 + \mathcal{A}_1$.

In the notation of Sec. IV A, the maps $\boldsymbol{M}_0$, $\boldsymbol{M}_1$ are

$$\boldsymbol{M}_0 = \mathbf{1}, \qquad \boldsymbol{M}_1 : \hat{\Omega} \mapsto 4\pi\delta_{\mathrm{D}}(\hat{\Omega};\hat{\Omega}_*). \tag{49}$$

The two maps $\boldsymbol{M}_0$, $\boldsymbol{M}_1$ are not orthogonal, and more importantly, are in general not statistically independent, in the sense that $\boldsymbol{M}_0 \cdot \boldsymbol{\mathcal{F}} \cdot \boldsymbol{M}_1 \neq 0$.

Applying the formalism of Sec. IV A, we must first compute the 2 by 2 inverse-covariance matrix F_{mn}. For any given $\hat{\Omega}_*$, its components are

$$F_{00} = \sum_I \mathcal{F}_I \mathcal{H}_I^2, \tag{50}$$

$$F_{01} = F_{10} = \sum_I \mathcal{F}_I \mathcal{H}_I \gamma_I(\hat{\Omega}_*), \tag{51}$$

$$F_{11} = \sum_I \mathcal{F}_I [\gamma_I(\hat{\Omega}_*)]^2. \tag{52}$$

The covariance matrix of the amplitudes $(\mathcal{A}_0, \mathcal{A}_1)$ is then the inverse of F. In Fig. 4, we show the minimum background monopole and hot-spot amplitudes detectable at the 95% confidence level, as a function of the direction of the hot spot. We see that the sensitivity to the background monopole amplitude $\sqrt{\mathcal{A}_0^{95\%}}$ varies from 2.5×10^{-15} to 3.3×10^{-15}, depending on the assumed location of the hot spot. This is to be compared with the estimated EPTA sensitivity to a monopole-only GWB, $A_h^{95\%} \approx 2.5 \times 10^{-15}$. We thus see that searching for a monopole simultaneously with another map, in this case a hot spot, systematically degrades the sensitivity to the monopole relative to the case where one assumes the GWB is a pure monopole. This is due to the statistical correlation (i.e., nonzero matrix element F_{01}) between the two maps.

It is important to emphasize that this analysis applies to a hot spot in a *known direction*. Without prior knowledge of the hot spot direction, it is not possible to set any constraint on its amplitude, as the number of free parameters would then be infinite, thus exceeding N_{pair}.
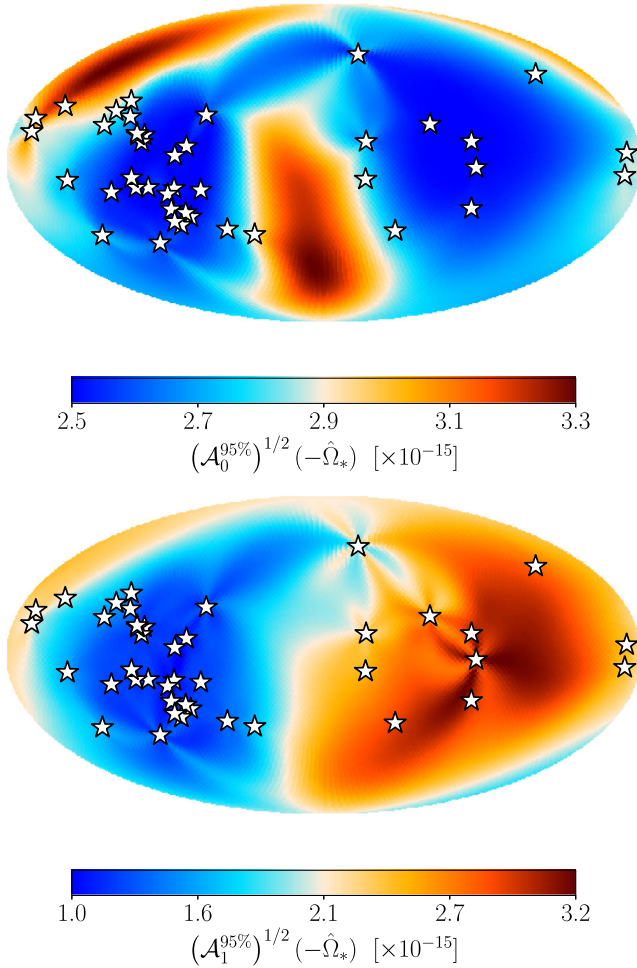
FIG. 4. Minimum amplitudes of a background monopole (top) and hot spot (bottom) detectable at the 95% confidence level by the EPTA, as a function of the hot spot direction $-\hat{\Omega}_*$. This assumes the GWB consists of a monopole and a single hot spot originating from a known direction.

As mentioned, this is mostly a toy problem, as in practice it is highly unlikely that one would expect only one single spot at a time. Next we consider a more realistic setup, where we simultaneously constrain the amplitudes of the GWB in multiple coarse pixels.

C. Detectability of coarsely pixelized anisotropies

As explained earlier, we may simultaneously search for at most $N_{\rm pair}$ independent amplitudes of the GWB. It is therefore not possible to simultaneously constrain the GWB in all pixels in the sky. We can, however, meaningfully analyze the amplitudes of GWB in $N_{\rm pix} \leq N_{\rm pair}$ coarse pixels. For this analysis, we will use the HEALPix pixelization [55] into $N_{\rm pix} = 12N_{\rm side}^2$ pixels of equal area $4\pi/N_{\rm pix}$. We define the $N_{\rm pix}$ coarse-pixel basis functions $\boldsymbol{P}_n$ such that $P_n(\hat{\Omega}) = 1$ if $\hat{\Omega}$ is within the nth pixel, and 0 otherwise, and search for coarse-grained anisotropies of the form

$$\mathcal{A} = \sum_{n=1}^{N_{\rm pix}} \mathcal{A}_n \boldsymbol{P}_n. \tag{53}$$

To compute the discrete Fisher matrix coefficients F_{nm}, we need to first computed the dot products $\boldsymbol{\gamma}_{\hat{p}\hat{q}} \cdot \boldsymbol{P}_n = \frac{1}{N_{\rm pix}} \langle \boldsymbol{\gamma}_{\hat{p}\hat{q}} \rangle_n$, where $\langle \boldsymbol{M} \rangle_n$ is the average of map $\boldsymbol{M}$ in the nth pixel. The latter is obtained by "degrading" the map to a resolution of $N_{\rm pix}$.

The 861 distinct pulsar pairs afforded by the EPTA allow us to use $N_{\rm side}$ as large as 8. We show the 95% sensitivity $A_{h,{\rm pix}}^{95\%} \equiv \sqrt{\mathcal{A}_n^{95\%}}$ in each coarse pixel in Fig. 5, for $N_{\rm side} = 1, 2, 4$, corresponding to $N_{\rm pix} = 12$, 48 and 192, respectively. For $N_{\rm side} = 8$, we found that the 768×768 coarse-pixel covariance matrix is numerically ill-conditioned, thus do not consider this case. We see that the pixel-by-pixel sensitivity degrades dramatically as $N_{\rm side}$ is increased (especially in the less well-constrained parts of the sky). Note also that the lower $N_{\rm pix}$ limits cannot be simply recovered by degrading the higher $N_{\rm pix}$ limits, as the order of map-degrading and inverting the Fisher matrix cannot be interchanged.

The average characteristic strain squared is given by

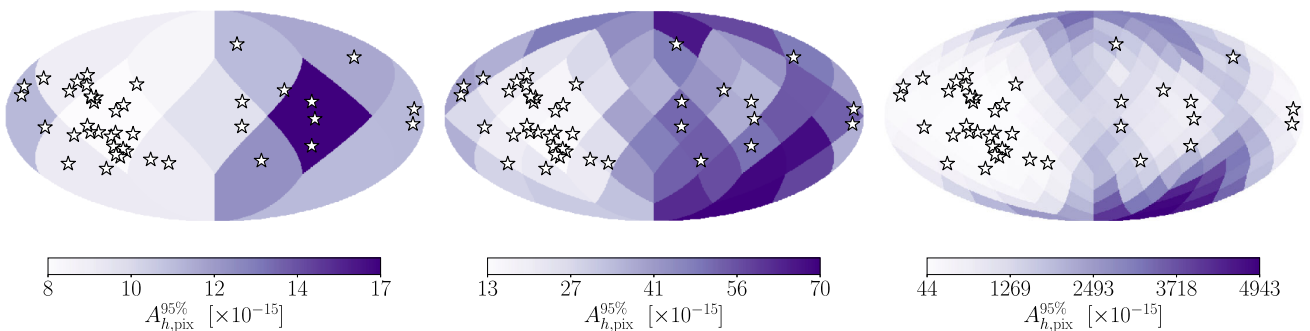$$A_h^2 = \frac{1}{N_{\rm pix}} \sum_{n=1}^{N_{\rm pix}} \mathcal{A}_n, \tag{54}$$

FIG. 5. Sensitivity of the EPTA to a coarsely pixelized GWB, for $N_{\rm pix} = 12$, 48 and 192 HEALPix pixels. Specifically, this shows the sensitivity in each coarse pixel after marginalizing over all other pixels.

which has variance

$$\text{var}(A_h^2) = \frac{1}{N_{\text{pix}}^2} \sum_{n,m} \text{cov}(\mathcal{A}_n, \mathcal{A}_m), \qquad (55)$$

translating to a 95% sensitivity $(A_h^{95\%})^2 = 2\sqrt{\text{var}(A_h^2)}$. For the EPTA, we find $A_h^{95\%} = (4.6, 5.0, 7.8) \times 10^{-15}$ for $N_{\text{side}} = 1, 2, 4$, respectively. This is to be contrasted with $A_h^{95\%} = 2.5 \times 10^{-15}$ if the GWB were known to be strictly isotropic. This highlights once again that the general form assumed for the GWB strongly affects the sensitivity to the monopole.

D. Detectability of spherical-harmonic amplitudes

1. Setup and physical prior

Following common practice [23,32], we may decompose GWB anisotropies on the basis of *real* spherical harmonics $\mathcal{Y}_{\ell m}(\hat{\Omega})$, which we normalize such that $\mathcal{Y}_{\ell m} \cdot \mathcal{Y}_{\ell' m'} = \delta_{\ell\ell'}\delta_{mm'}$ (i.e., with a factor $\sqrt{4\pi}$ larger than the standard definition of spherical harmonics). Explicitly, we assume the GWB amplitude takes the form

$$\mathcal{A} = \sum_{\ell=0}^{\ell_{\text{max}}} \sum_{m=-\ell}^{\ell} \mathcal{A}_{\ell m} \mathcal{Y}_{\ell m}. \qquad (56)$$

Note that this form assumes the rather stringent prior that harmonic coefficients for $\ell > \ell_{\text{max}}$ all vanish.

In order for $\mathcal{A}$ to represent a physical GWB, it needs to be positive everywhere. To our knowledge, there is no simple analytic expression for this physical prior. Nevertheless, we can derive an approximation as follows. We define the GWB anisotropy $\Delta\mathcal{A}(\hat{\Omega}) \equiv \mathcal{A}(\hat{\Omega}) - \mathcal{A}_{00}$. It has the following variance over directions in the sky:

$$\langle [\Delta\mathcal{A}]^2 \rangle_{\hat{\Omega}} \equiv \int \frac{d^2\hat{\Omega}}{4\pi} [\Delta\mathcal{A}(\hat{\Omega})]^2 = \Delta\mathcal{A} \cdot \Delta\mathcal{A}$$

$$= \sum_{\ell=1}^{\ell_{\text{max}}} \sum_{m=-\ell}^{\ell} \mathcal{A}_{\ell m}^2 = \sum_{\ell=1}^{\ell_{\text{max}}} (2\ell + 1)\mathcal{C}_\ell, \qquad (57)$$

where we have defined

$$\mathcal{C}_\ell \equiv \frac{1}{2\ell + 1} \sum_{m=-\ell}^{\ell} \mathcal{A}_{\ell m}^2. \qquad (58)$$

A simple approximate physical prior follows: we require that the monopole amplitude $\mathcal{A}_{00}$ is larger than twice the rms anisotropy, implying

$$\mathcal{C}_0 \geq 4 \sum_{\ell=1}^{\ell_{\text{max}}} (2\ell + 1)\mathcal{C}_\ell. \qquad (59)$$

This implies a conservative physical prior

$$\mathcal{C}_\ell \leq \frac{\mathcal{C}_0}{4(2\ell + 1)}, \qquad \ell \geq 1. \qquad (60)$$

This simple criterion reproduces quite accurately the numerical results of TMG15 (see inset in their Fig. 1). Our conventions map to those of TMG15 through $\mathcal{A}_{\ell m} = A_h^2 c_{\ell m}/\sqrt{4\pi}$, $\mathcal{C}_\ell = A_h^4 C_l/4\pi$. Note that the monopole coefficient is just the amplitude of the characteristic strain squared, $\mathcal{A}_{00} = A_h^2$, and that $\mathcal{C}_0 = \mathcal{A}_{00}^2 = A_h^4$.

2. Forecasted sensitivity of the EPTA

Let us start by pointing out that for a given maximum $\ell = \ell_{\text{max}}$, the decomposition (56) includes $(\ell_{\text{max}} + 1)^2$ independent coefficients. It thus follows that one may constrain the spherical-harmonic amplitudes from pulsar pair cross correlations only if $\ell_{\text{max}} \leq \sqrt{N_{\text{pair}}} - 1$. For larger values of ℓ_{max}, the $\mathcal{A}_{\ell m}$ are then only constrained by the physical prior. For the full EPTA, we may therefore use $\ell_{\text{max}} \leq 28$. For the subset of 6 EPTA pulsars, one is restricted to $\ell_{\text{max}} \leq 2$.

We compute the covariance of the amplitudes $\mathcal{A}_{\ell m}$ as described in Sec. IV A. Instead of showing the sensitivity to the individual $\mathcal{A}_{\ell m}$ coefficients, we estimate the detectability of the rotationally-invariant coefficients $\mathcal{C}_\ell$ defined in Eq. (58). Their mean and covariance matrix are given by

$$\langle \mathcal{C}_\ell \rangle = \frac{1}{2\ell + 1} \sum_{m=-\ell}^{\ell} \text{var}(\mathcal{A}_{\ell m}), \qquad (61)$$

$$\text{cov}(\mathcal{C}_\ell, \mathcal{C}_{\ell'}) = 2 \sum_{m=-\ell}^{\ell} \sum_{m'=-\ell'}^{\ell'} \frac{\text{cov}(\mathcal{A}_{\ell m}\mathcal{A}_{\ell' m'})^2}{(2\ell + 1)(2\ell' + 1)}. \qquad (62)$$

The $\mathcal{C}_\ell$ are not Gaussian-distributed (especially for small ℓ). Nevertheless, we may define an *approximate* 95% sensitivity estimate as follows:

$$\mathcal{C}_\ell^{95\%} \equiv \langle \mathcal{C}_\ell \rangle + 2\sqrt{\text{var}(\mathcal{C}_\ell)}. \qquad (63)$$

Note that even for $\ell = 0$, this gives $\mathcal{C}_0^{95\%} = (1 + 2\sqrt{2})\text{var}(\mathcal{A}_{00}) \approx 3.8\text{var}(\mathcal{A}_{00})$, which is very close to the correct $2 - \sigma$ sensitivity $\mathcal{C}_0^{95\%} = 4\text{var}(\mathcal{A}_{00})$.

We show $\mathcal{C}_\ell^{95\%}$ in Fig. 6, for several values of ℓ_{max}. We see that for any given coefficient $\mathcal{C}_\ell$, the sensitivity systematically degrades as ℓ_{max} is increased. In particular, as anticipated, the sensitivity to the monopole significantly worsens as ℓ_{max} is increased. This is because the monopole is correlated with other spherical harmonics, and as one enlarges the space of functions to be searched over, the uncertainty on the monopole amplitude increases.

Let us remark that the coefficients $\mathcal{C}_\ell$ are statistically correlated. We explicitly give their correlation coefficients
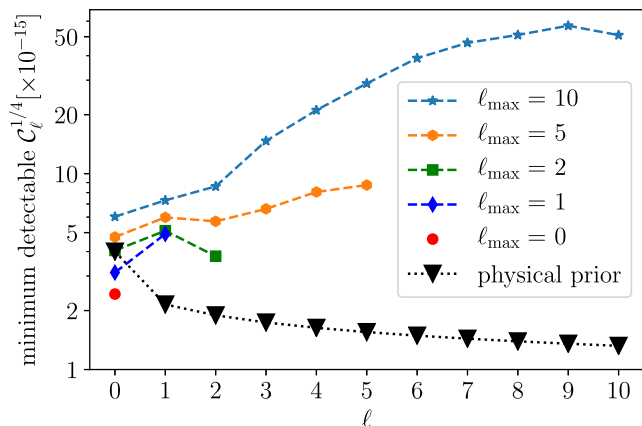
FIG. 6. Approximate 95% sensitivity of the full EPTA to the coefficients $\mathcal{C}_\ell \equiv \sum_m \mathcal{A}_{\ell m}^2/(2\ell+1)$, as a function of the cutoff $\ell_{\max}$, beyond which the coefficients are assumed to strictly vanish. The physical prior is computed for a monopole upper limit $A_h \leq 4 \times 10^{-15}$ [32], and is systematically lower than the minimum detectable anisotropy amplitudes.

in Table I for $\ell_{\max} = 2$ and 5. We see that these correlations coefficients are in general not small, and depend on the chosen $\ell_{\max}$.

Finally, note that this analysis specifically estimates the minimum amplitudes necessary for a *detection*, through pulsar timing cross correlations. In addition, the monopole, dipole and quadrupole can be constrained by pulsar autocorrelations, as discussed in Sec. II G. In the limit that autocorrelations constrain the $\ell = 0, 1, 2$ harmonic coefficients much more tightly than the cross correlations, the variance of the remaining coefficients should be obtained by inverting the Fisher matrix restricted to these

TABLE I. Dimensionless correlation coefficients $\text{cov}(\mathcal{C}_\ell, \mathcal{C}_{\ell'})/\sqrt{\text{var}(\mathcal{C}_\ell)\text{var}(\mathcal{C}_{\ell'})}$, for $\ell_{\max} = 2$ and $\ell_{\max} = 5$, using the EPTA pulsars. The ellipses indicate that the tables are symmetric. This illustrates that the $\mathcal{C}_\ell$'s are statistically correlated and that for a given pair ℓ, ℓ', the correlation coefficient depends on the assumed $\ell_{\max}$.

$\ell_{\max} = 2$			
	$\mathcal{C}_0$	$\mathcal{C}_1$	$\mathcal{C}_2$
$\mathcal{C}_0$	1	0.28	0.33
$\mathcal{C}_1$	$\cdots$	1	0.12
$\mathcal{C}_2$	$\cdots$	$\cdots$	1

$\ell_{\max} = 5$						
	$\mathcal{C}_0$	$\mathcal{C}_1$	$\mathcal{C}_2$	$\mathcal{C}_3$	$\mathcal{C}_4$	$\mathcal{C}_5$
$\mathcal{C}_0$	1	0.43	0.45	0.15	0.06	0.03
$\mathcal{C}_1$	$\cdots$	1	0.33	0.20	0.07	0.05
$\mathcal{C}_2$	$\cdots$	$\cdots$	1	0.46	0.20	0.17
$\mathcal{C}_3$	$\cdots$	$\cdots$	$\cdots$	1	0.39	0.27
$\mathcal{C}_4$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	1	0.52
$\mathcal{C}_5$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	1

coefficients, leading to a lower noise for the $\mathcal{C}_\ell$ with $\ell \geq 3$. We have checked explicitly that this lowers the noise by no more than ~ 10–20%.

In conclusion, we have demonstrated that our Fisher formalism allows to forecast the sensitivity of a PTA to the spherical-harmonic amplitudes of the GWB, for a given cutoff $\ell_{\max}$. We have applied this specifically for the EPTA, and shown that the minimum detectable amplitudes are systematically larger than allowed by the physical prior, given current upper limits on the monopole. This seems to indicate that spherical harmonics are a suboptimal choice of basis for anisotropy searches. We will specifically compare our results with those of TMG15 in Sec. VII.

V. PRINCIPAL MAPS OF A PTA

A. Motivation and formalism

As highlighted in the previous section, one of the challenges when using standard bases to decompose the GWB intensity is that the amplitudes of the basis maps are statistically correlated. Moreover, their covariance matrix depends on the number of maps considered. As a result, one cannot easily set model-independent limits on the map amplitudes.

If one is completely agnostic regarding the angular dependence of the GWB, it is best to search under the lamppost, i.e., look for the amplitudes of maps which are best-measured by a given PTA, and uncorrelated with one another. To do so, we construct the N_{pair} unit-norm *principal maps* $\{\mathcal{M}_n\}$. They are defined to extremize $\text{SNR}^2 = \mathcal{M}_n \cdot \mathcal{F} \cdot \mathcal{M}_n$, under the normalization constraint $\mathcal{M}_n \cdot \mathcal{M}_n = 1$. The solutions of this constrained optimization problem are simply the eigenmaps of the Fisher matrix, i.e., maps that satisfy

$$\mathcal{F} \cdot \mathcal{M}_n = \frac{1}{\Sigma_n^2} \mathcal{M}_n, \tag{64}$$

where we denoted the corresponding eigenvalues by $1/\Sigma_n^2$. Since $\mathcal{F}$ is symmetric, its eigenmaps are orthogonal to one another and, importantly, uncorrelated with one another:

$$\mathcal{M}_n \cdot \mathcal{M}_m = \delta_{mn}, \qquad \mathcal{M}_n \cdot \mathcal{F} \cdot \mathcal{M}_m = \frac{\delta_{mn}}{\Sigma_n^2}. \tag{65}$$

We may rewrite the Fisher matrix as a direct sum of tensor products of the principal maps:

$$\mathcal{F} = \sum_n \frac{1}{\Sigma_n^2} \mathcal{M}_n \otimes \mathcal{M}_n. \tag{66}$$

Any GWB amplitude $\mathcal{A}(\hat{\Omega})$ can always be written as a linear combination of the N_{pair} eigenmaps and a piece orthogonal to all of them:

$$\mathcal{A} = \mathcal{A}_\parallel + \mathcal{A}_\perp, \tag{67}$$

$$\mathcal{A}_\parallel \equiv \sum_{n=1}^{N_{\text{pair}}} \mathcal{A}_n \mathcal{M}_n, \qquad \mathcal{A}_\perp \cdot \mathcal{M}_n = 0 \quad \forall \ n. \tag{68}$$
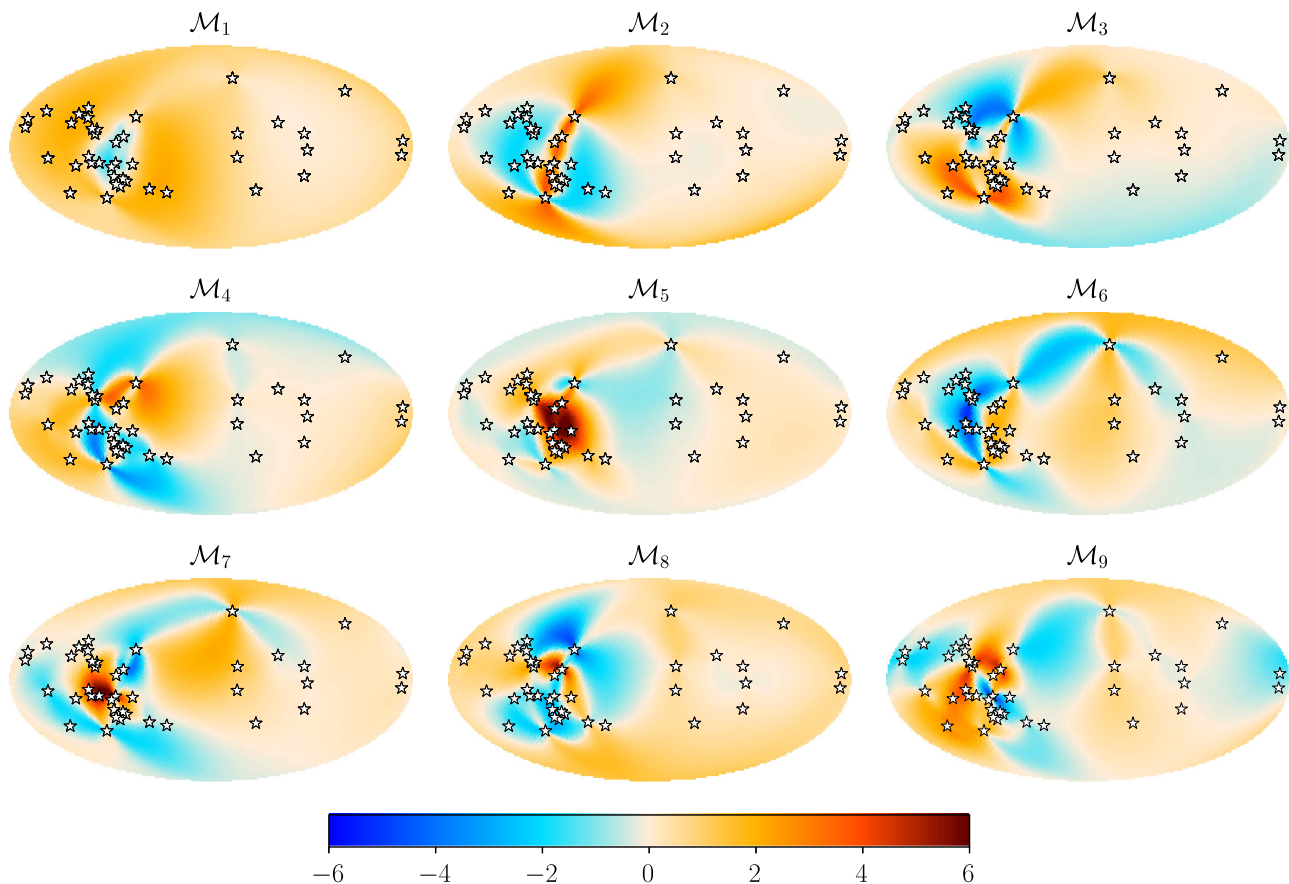
FIG. 7. First 9 principal maps of the EPTA, ordered by increasing noise. The maps have unit norm and their sign was chosen so that their average value (i.e., projection on the monopole) is positive, i.e., $\mathcal{M}_n \cdot \mathbf{1} > 0$.

In that case, the total SNR^2 is given by

$$\text{SNR}^2[\mathcal{A}(\hat{\Omega})] = \sum_{n=1}^{N_{\text{pair}}} \frac{\mathcal{A}_n^2}{\Sigma_n^2}. \qquad (69)$$

From Eq. (69), we see that the minimum amplitude of the eigenmap $\mathcal{M}_n$ detectable with $\text{SNR} = 2$ is $\mathcal{A}_n = 2\Sigma_n$.

The eigenmaps for an orthonormal basis of the N_{pair}-dimensional space of *observable* maps for a given PTA. Maps $\mathcal{A}_\perp$ which are orthogonal to this space are completely unobservable, i.e., have infinite noise.

To compute the eigenmaps in practice, we first write them in the form of linear combinations of the pairwise timing response functions:

$$\mathcal{M}_n(\hat{\Omega}) = \sum_I \mathcal{M}_n^I \sqrt{\mathcal{F}_I} \gamma_I(\hat{\Omega}). \qquad (70)$$

The eigenvalue problem (64) is then equivalent to the finite-dimensional eigenvalue-problem

$$\sum_J F_{IJ} \mathcal{M}_n^J = \frac{1}{\Sigma_n^2} \mathcal{M}_n^I, \qquad (71)$$

where the symmetric $N_{\text{pair}} \times N_{\text{pair}}$ matrix F_{IJ} has elements

$$F_{IJ} \equiv \sqrt{\mathcal{F}_I \mathcal{F}_J} \gamma_I \cdot \gamma_J. \qquad (72)$$

We show the first 9 eigenmaps of the EPTA in Fig. 7. These maps look very different from any known analytic basis functions (such as spherical harmonics): indeed, they
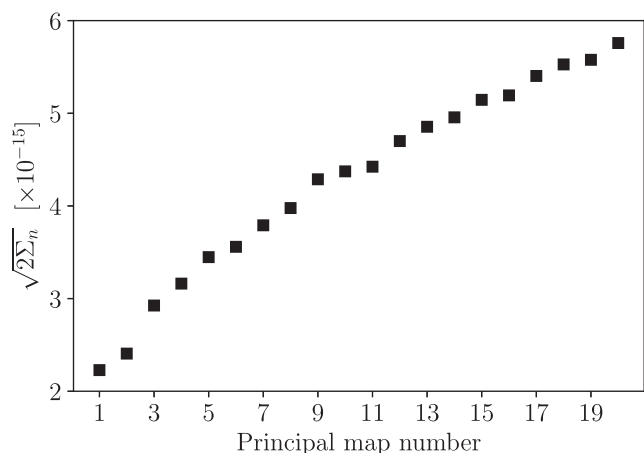

FIG. 8. Noise eigenvalues of the first 20 principal maps of the EPTA. Specifically, the principal maps are normalized to unity, and here we show the minimum amplitude of the coefficients multiplying them (with dimensions of characteristic strain) required for a 95%-confidence detection.

reflect the properties of a particular PTA. We show the eigenvalues (translated into a detectability threshold in characteristic strain) of the first 20 eigenmaps in Fig. 8. Note that beyond the first ∼20 eigenmaps, the noise eigenvalues Σ_n start increasing exponentially with n.

One can use the principal maps as a basis to decompose the observable GWB, and follow the same formalism as in Sec. IV. The advantage of using this basis is that the map amplitudes are statistically uncorrelated. As a consequence, upper limits obtained on any given map amplitude do not depend on the cutoff. In particular, one could start by searching for the amplitude of the first principal map alone, then search simultaneously for the amplitudes of the first two maps, etc…, and not change any of the sensitivity estimates.

B. Application: Map reconstruction

Several map-making methods for the GWB have been proposed, from a potential high-frequency GWB in the LIGO-band [56–60], to millihertz LISA band [57,61–66], to the nanohertz PTA band [23,28–30,42]. There are also studies exploring methods to extract maps of continuous gravitational waves [67,68]. Here we propose a new reconstruction method for the GWB, relying on the principal maps.

In principle, given N_{pair} estimated timing residual cross-spectra $\hat{\mathcal{R}}_I(f)$, one can build an estimate of the underlying GWB intensity from Eqs. (30)–(31). However, in the weak-signal limit, all the estimators $\hat{\mathcal{A}}_I$ are individually noise-dominated. Collectively, however, a PTA may still measure some independent pieces of information with high signal-to-noise ratio: these are precisely the amplitudes of the lowest-noise principal maps, $\hat{\mathcal{A}}_n \equiv \mathcal{M}_n \cdot \hat{\mathcal{A}}$. Concretely, using this expression with Eq. (31), we see that the $\hat{\mathcal{A}}_n$ are linear combinations of the timing residual cross-spectra $\hat{\mathcal{R}}_I(f)$, appropriately integrated over frequencies.

We can then use the principal maps to attempt to "reconstruct" the GWB angular dependence. Provided some of the individual principal map amplitudes $\hat{\mathcal{A}}_n$ are measured with sufficiently high individual $\text{SNR}_n \equiv \hat{\mathcal{A}}_n/\Sigma_n$ (say, $\text{SNR}_n > 3$), we can define the reconstructed map

$$\mathcal{A}_{\text{recon}} \equiv \sum_{n;\text{SNR}_n>3} \hat{\mathcal{A}}_n \mathcal{M}_n. \qquad (73)$$

This procedure is analogous to the production of "dirty map" in radio interferometry, i.e., including the contributions of observed "visibilities," and setting the remaining (unobserved) piece to zero. Unlike interferometry, however, the ability to reconstruct a map is a strong function of its amplitude. Indeed, for a single interferometer, different visibilities typically have comparable noise, and as a consequence all contribute to the dirty map once any one of them is detected with sufficiently high SNR. In

contrast, the noise of principal maps of a PTA increases steeply with the principal map number, see Fig. 8 for the principal maps of the EPTA. Note that this property is *not* a result of unequal pulsar noises: even for an array of equal pulsars, densely and isotropically distributed on the sky, the eigenvalues of the Fisher matrix are steep function of principal map index (see Fig. 4 in Paper I).

We illustrate the map reconstruction technique with the EPTA in Fig. 9. In the top row, we show the reconstructed maps obtained if the underlying GWB is a pure monopole, with amplitudes $A_h = 10^{-14}, \sqrt{3} \times 10^{-14}, 3 \times 10^{-14}$, from left to right. The bottom row shows the reconstructed maps obtained if the underlying GWB is proportional to the map shown in Fig. 10, with the same monopole amplitude. We see that even with these very large amplitudes, the reconstructed maps have little resemblance with the underlying GWB intensity distribution. Note, also, that the overall amplitude of these maps (corresponding to total SNR of 30, 100 and 300, respectively, from left to right), is so large that they are most likely inconsistent with the weak-signal limit.

C. Possible extensions of principal maps singling out the monopole

The principal maps provide a useful basis to decompose the GWB if one is fully agnostic about its angular distribution. In practice, however, this is not the case: one does expect a significant monopole component in the GWB. In other words, physical considerations isolate a preferred map, which should always be included in GWB searches. This map is in general different from the principal maps, which are entirely based on the noise and geometric properties of the PTA, rather than external physical considerations.

Figure 11 shows the monopole components of the first 20 EPTA principal maps, as well as their correlation coefficient with the monopole. Even though the monopole has the largest projection on the first principal map, we see that it still has significant projections on and correlations with some of the higher-order eigenmaps. In other words, higher-order eigenmaps are (i) not anisotropic and (ii) statistically correlated with the monopole.

In order to alleviate issue (i), one could try to construct a set of "principal anisotropies," $\{\mathcal{A}_n\}$, which are orthogonal to the monopole, i.e., satisfy $\mathcal{A}_n \cdot \mathbf{1} = 0$. One would then find the unit-norm maps extremizing $\mathcal{A}_n \cdot \mathcal{F} \cdot \mathcal{A}_n$ under this additional constraint. This optimization problem admits $N_{\text{pair}} - 1$ solutions, which, in addition to the monopole, form a basis of N_{pair} maps which can be used to search for a monopole and anisotropies. However, the principal anisotropies derived in this fashion are in general statistically correlated with the monopole, $\mathcal{A}_n \cdot \mathcal{F} \cdot \mathbf{1} \neq 0$, thus would inflate the error bar on the monopole when included in a search.

To alleviate issue (ii), one could construct a set of "monopole-uncorrelated maps," $\{\mathcal{B}_n\}$, which are uncorrelated with the monopole, i.e., satisfy $\mathcal{B}_n \cdot \mathcal{F} \cdot \mathbf{1} = 0$. One would then find the unit-norm maps extremizing $\mathcal{B}_n \cdot \mathcal{F} \cdot \mathcal{B}_n$ under this additional constraint. This optimization problem also admits $N_{\text{pair}} - 1$ solutions, which, in addition to the monopole, form a basis of N_{pair} maps which can be used to search for a monopole and uncorrelated maps. However, the maps $\mathcal{B}_n$ derived in this fashion in general have a nonzero projection on the monopole, $\mathcal{B}_n \cdot \mathbf{1} \neq 0$. Thus, detecting a non-zero amplitude for these maps would not imply that one has detected anisotropies in the data.

One could try and alleviate both issues simultaneously by searching for a set of normalized maps $\mathcal{C}_n$ which are anisotropic *and* uncorrelated with the monopole, with extremal SNR^2. As long as $\mathcal{F} \cdot \mathbf{1}$ is not collinear with $\mathbf{1}$ (which is the case if the monopole is not an eigenmap of the Fisher matrix), the two constraints are independent. The resulting optimization problem thus admits $N_{\text{pair}} - 2$ solutions. To span the N_{pair}-dimensional set of observable maps, one must supplement these monopole-uncorrelated principal anisotropies with the monopole *and* $\mathcal{F} \cdot \mathbf{1} - (\mathcal{F} \cdot \mathbf{1})\mathbf{1}$ (properly normalized). Indeed, this additional map is orthogonal to the monopole and to all the $\mathcal{C}_n$ maps (which stems from the condition that the $\mathcal{C}_n$ are orthogonal to and uncorrelated with the monopole), thus linearly independent from all of them. However, this additional map is in general not statistically independent from the monopole nor from the maps $\mathcal{C}_n$. Therefore, this construction still does not alleviate the issue.

To conclude, unless the monopole is an eigenmap of the Fisher matrix, there is no good strategy to *agnostically* search for anisotropies within the standard Bayesian setup. In Sec. VI, we explore a frequentist approach to this problem.

VI. FREQUENTIST APPROACH TO AGNOSTIC ANISOTROPY SEARCHES

In the previous sections we have highlighted the difficulties in carrying a standard Bayesian search for anisotropies of pre-determined shapes alongside a monopole. Here we take a different approach, and derive a criterion to assess the presence of anisotropies in the data, regardless of their specific shape, without requiring a basis of maps on which to decompose the GWB.

A. Derivation of the frequentist criterion

Given the data translated into an estimator $\hat{\mathcal{A}}$ for the GWB amplitude, one may seek the monopole amplitude that minimizes the χ^2, given by

$$\chi^2(\mathcal{A}_0) = (\mathcal{A}_0 \mathbf{1} - \hat{\mathcal{A}}) \cdot \mathcal{F} \cdot (\mathcal{A}_0 \mathbf{1} - \hat{\mathcal{A}}). \quad (74)$$

The best-fit monopole amplitude is then simply

$$\mathcal{A}_0^{\text{bf}} = \frac{\mathbf{1} \cdot \mathcal{F} \cdot \hat{\mathcal{A}}}{\mathbf{1} \cdot \mathcal{F} \cdot \mathbf{1}}. \quad (75)$$
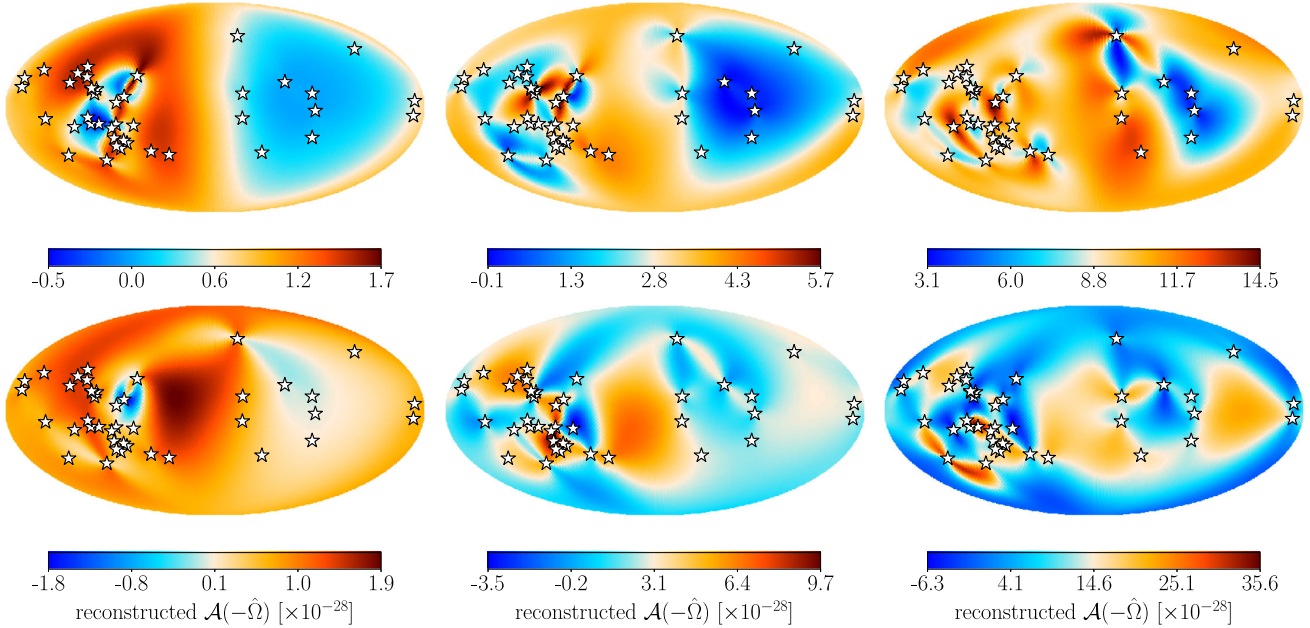

FIG. 9. Reconstructions of a purely isotropic GWB (top row) and of the "GWB" map shown in Fig. 10 (bottom row) with the EPTA. In each column, from left to right, the input map has monopole amplitude $A_h^2 = 10^{-28}, 3 \times 10^{-28}, 10^{-27}$, respectively. For the isotropic input map, these reconstructed maps are built from 3, 9, and 21 principal maps detected with individual SNR > 3. For the "GWB" map, the reconstructed maps are built from 3, 13, and 26 principal maps.

The SNR of the best-fit monopole is then $\mathrm{SNR}^{\mathrm{bf}} = |\mathcal{A}_0^{\mathrm{bf}}|\sqrt{\mathbf{1} \cdot \mathcal{F} \cdot \mathbf{1}}$. This quantifies how well the best-fit monopole is detected relative to statistical noise. However, it does not quantify whether a pure monopole is a good fit to the data or not.

To quantify the goodness of fit, one can examine the χ^2 at this best-fit value. After simplification, one obtains

$$\chi_{\mathrm{bf}}^2 \equiv \chi^2(\mathcal{A}_0^{\mathrm{bf}}) = \hat{\mathcal{A}} \cdot \mathcal{F} \cdot \hat{\mathcal{A}} - \frac{(\mathbf{1} \cdot \mathcal{F} \cdot \hat{\mathcal{A}})^2}{\mathbf{1} \cdot \mathcal{F} \cdot \mathbf{1}}$$
$$= \hat{\mathcal{A}} \cdot \tilde{\mathcal{F}} \cdot \hat{\mathcal{A}}, \tag{76}$$

where we have defined the projected Fisher matrix

$$\tilde{\mathcal{F}} \equiv \mathcal{F} - \frac{(\mathcal{F} \cdot \mathbf{1}) \otimes (\mathbf{1} \cdot \mathcal{F})}{\mathbf{1} \cdot \mathcal{F} \cdot \mathbf{1}}, \tag{77}$$

which satisfies $\mathbf{1} \cdot \tilde{\mathcal{F}} = \tilde{\mathcal{F}} \cdot \mathbf{1} = 0$. This implies that any monopole contribution to $\hat{\mathcal{A}}$ drops out of χ_{bf}^2.

If the underlying GWB is purely isotropic, so that $\hat{\mathcal{A}}$ is a monopole plus pure statistical noise, only the latter contributes to χ_{bf}^2, which has a chi-square distribution with $N_{\mathrm{pair}} - 1$ degrees of freedom. In particular, it has mean $N_{\mathrm{pair}} - 1$ and variance $2(N_{\mathrm{pair}} - 1)$.

In the limit that $N_{\mathrm{pair}} \gg 1$, χ_{bf}^2 is approximately Gaussian-distributed. A simple criterion for the detection of a anisotropy with 95% confidence (i.e., for the presence of a non-pure-statistical noise in $\hat{\mathcal{A}}$) is then

$$|\chi_{\mathrm{bf}}^2 - (N_{\mathrm{pair}} - 1)| > 2\sqrt{2}\sqrt{N_{\mathrm{pair}} - 1}. \tag{78}$$

Note that this criterion does not provide any information about the nature of the anisotropy (besides the fact that it must have some non-zero projections on one of the $N_{\mathrm{pair}} - 1$ eigenmaps of $\tilde{\mathcal{F}}$). It merely asserts that the data cannot be well described purely by a monopole.

Consider a map $\mathcal{A} = \mathcal{A}_1 M$, where $M \neq \mathbf{1}$ is a dimensionless map. The minimum amplitude $\mathcal{A}_1$ required for an anisotropy to be confidently detected in the data (without knowledge of its precise form) is thus

$$\mathcal{A}_1^{\mathrm{min}} = \left(\frac{2\sqrt{2}\sqrt{N_{\mathrm{pair}} - 1}}{M \cdot \tilde{\mathcal{F}} \cdot M} \right)^{1/2}. \tag{79}$$

This corresponds to a best-fit monopole $\mathcal{A}_0^{\mathrm{bf,min}} = \mathcal{A}_1^{\mathrm{min}}(\mathbf{1} \cdot \mathcal{F} \cdot M)/(\mathbf{1} \cdot \mathcal{F} \cdot \mathbf{1})$, hence to a minimum SNR for the best-fit monopole

$$\mathrm{SNR}^{\mathrm{min}} = \frac{(2\sqrt{2}\sqrt{N_{\mathrm{pair}} - 1})^{1/2}|\mathbf{1} \cdot \mathcal{F} \cdot M|}{\sqrt{(M \cdot \tilde{\mathcal{F}} \cdot M)(\mathbf{1} \cdot \mathcal{F} \cdot \mathbf{1})}}. \tag{80}$$
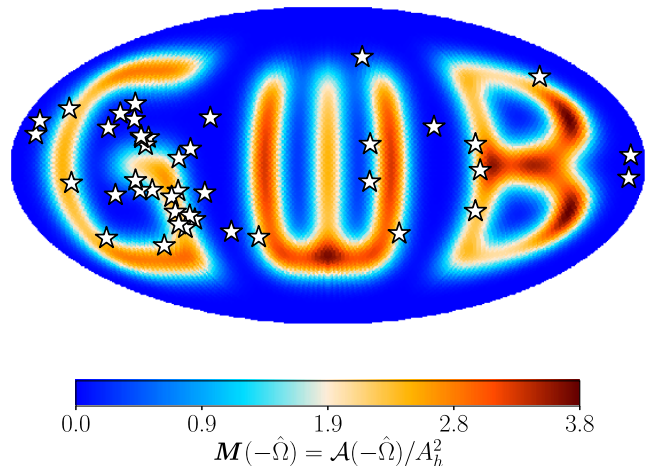

FIG. 10. Dimensionless GWB intensity map used to illustrate the map reconstruction technique in Sec. V B, as well as the frequentist approach in Sec. VI. This map is normalized to a unit monopole.

B. Illustration with the EPTA

We illustrate this frequentist approach to the map M shown in Fig. 10, which is normalized to have a unit monopole. First, we compute the best-fit monopole amplitude for this map, $\mathcal{A}_0^{\mathrm{bf}} \equiv \mathcal{A}_1(\mathbf{1} \cdot \mathcal{F} \cdot M)/(\mathbf{1} \cdot \mathcal{F} \cdot \mathbf{1})$. We find $\mathcal{A}_0^{\mathrm{bf}} = 0.96\mathcal{A}_1$: the best-fit monopole amplitude turns out to be rather close to the actual value. This is because the projections of M and 1 on the first few principal maps happen to be similar in this case. This need not be always the case for other maps or PTAs. From Eq. (79), and using the EPTA Fisher matrix, we find $\sqrt{\mathcal{A}_1^{\mathrm{min}}} \approx 1.2 \times 10^{-14}$ (this quantity has dimensions of characteristic strain). This corresponds to an SNR for the best-fit monopole amplitude $0.96 \times \mathcal{A}_1^{\mathrm{min}}\sqrt{\mathbf{1} \cdot \mathcal{F} \cdot \mathbf{1}} \approx 46$. In other words, if the GWB has the angular dependence shown in Fig. 10, the full EPTA would only be able to establish the presence of anisotropies with 95% confidence (without being able to specify their precise form) after it detects the best-fit monopole with SNR $\gtrsim 46$.

VII. REPRODUCING TMG15'S RESULTS

To our knowledge, TMG15 is the only study of anisotropies in the GWB with real PTA data. In this section, we show how to use our Fisher formalism to reproduce their results. We moreover clarify the meaning of some of their upper limits. Indeed, as discussed earlier, a PTA is sensitive to N_{pair} independent maps at most. This means that it is not possible to simultaneously constrain the GWB amplitude in more than N_{pair} pixels in the sky, or constrain more than N_{pair} of its spherical harmonic amplitudes. Yet, TMG15 present a 95% upper-limit map on the GWB in 12288 pixels, using only the 15 pairs afforded by 6 EPTA pulsars. They also present upper limits on spherical harmonic
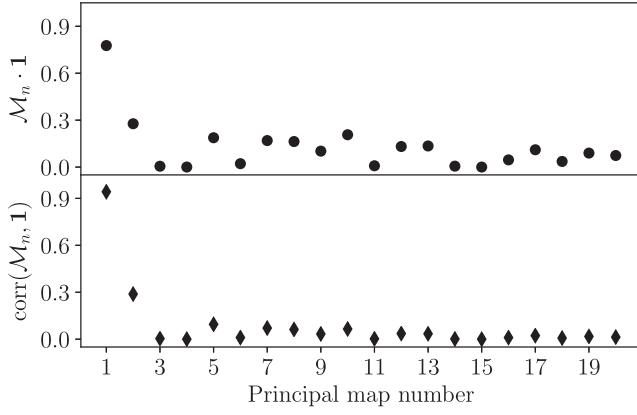
FIG. 11. Monopole component (upper panel) and correlation coefficient for the first 20 EPTA principal maps. The correlation coefficient is defined as $\mathrm{corr}(\mathcal{M}_n, \mathbf{1}) \equiv \mathcal{M}_n \cdot \mathcal{F} \cdot \mathbf{1}/[(\mathcal{M}_n \cdot \mathcal{F} \cdot \mathcal{M}_n)(\mathbf{1} \cdot \mathcal{F} \cdot \mathbf{1})]^{1/2}$.

coefficients with $\ell_{\max}$ as large as 10, corresponding to 121 independent coefficients. Clearly, these cannot be true upper limits. In this section we shall understand precisely what these limits are, and reproduce them approximately using the Fisher matrix and its eigenmap decomposition.

A. Pixel-by-pixel upper limit maps

Following TMG15, we define $\boldsymbol{P} \equiv \mathcal{A}/A_h^2$, which satisfies $\int d^2\hat{\Omega} P(\hat{\Omega}) = 4\pi$. With our notation, the N_{pair} overlap reduction functions are given by $\Gamma_I \equiv \frac{3}{8}\boldsymbol{\gamma}_I \cdot \boldsymbol{P}$. We now discretize the sky into N_{pix} pixels centered at $\hat{\Omega}_i$, with equal area $\Delta\Omega$. As in TMG15, we define $c_i \equiv P(\hat{\Omega}_i)$. The overlap reduction functions are then

$$\Gamma_I = \frac{3}{8} \int \frac{d^2\hat{\Omega}}{4\pi} \gamma_I(\hat{\Omega}) P(\hat{\Omega}) \approx \sum_{i=1}^{N_{\mathrm{pix}}} H_{Ii} c_i, \quad (81)$$

where the $N_{\mathrm{pair}} \times N_{\mathrm{pix}}$ matrix $\boldsymbol{H}$ has elements

$$H_{Ii} = \frac{3}{8} \gamma_I(\hat{\Omega}_i) \frac{\Delta\Omega}{4\pi}. \quad (82)$$

We can thus rewrite the relationship between the overlap-reduction functions and the power distribution $\boldsymbol{P}(\hat{\Omega})$ in matrix form, as $\vec{\Gamma} = \boldsymbol{H}\vec{c}$, where $\vec{\Gamma}$ is the N_{pair}-dimensional vector with components Γ_I and $\vec{c}$ is the N_{pix}-dimensional vector with components c_i.

Unless $N_{\mathrm{pix}} = N_{\mathrm{pair}}$, the matrix $\boldsymbol{H}$ is not square, and *a fortiori* not invertible. One can however define its Moore-Penrose pseudo-inverse [69,70] $\boldsymbol{H}^+$. We may then define the N_{pix}-dimensional vector

$$\vec{\tilde{c}} \equiv \boldsymbol{H}^+ \vec{\Gamma} = (\boldsymbol{H}^+ \boldsymbol{H})\vec{c}. \quad (83)$$

Unless $N_{\mathrm{pix}} = N_{\mathrm{pair}}$, $\boldsymbol{H}^+ \neq \boldsymbol{H}^{-1}$ and $\vec{\tilde{c}} \neq \vec{c}$. Instead, as we demonstrate in Appendix, $\tilde{c}_i = P_{\|}(\hat{\Omega}_i)$ is the (discretized) power distribution *projected* on the N_{pair}-dimensional space spanned by the γ_I's. In other words, $\vec{\tilde{c}}$ corresponds to the observable component of the GWB power, i.e., the only piece of $\boldsymbol{P}(\hat{\Omega})$ that is measurable by a PTA. Importantly, it *cannot* be identified with $c_i = P(\hat{\Omega}_i)$ as done in TMG15. As a consequence, the map shown in TMG15 is only an upper-limit map for the *projected* (i.e., observable) GWB intensity. We refer to it as the Moore-Penrose pseudo-upper-limit map.

We now explicitly construct an approximation of this pseudo-upper-limit map. Given estimators for the timing residual correlations, one may construct an estimator for the projected GWB map $\hat{\mathcal{A}}_{\|} = \hat{\mathcal{A}}$ given by Eqs. (30)–(31). The pixel values of this estimator map are precisely $\hat{\mathcal{A}}_{\|}(\hat{\Omega}_i) = A_h^2 P_{\|}(\hat{\Omega}_i) = A_h^2 \tilde{c}_i$. In order to compute the variance of these values, it is best to first decompose it on the basis of principal maps $\mathcal{M}_n$:

$$\hat{\mathcal{A}}_{\|}(\hat{\Omega}_i) = \sum_{n=1}^{N_{\mathrm{pair}}} \hat{\mathcal{A}}_n \mathcal{M}_n(\hat{\Omega}_i), \quad \hat{\mathcal{A}}_n \equiv \mathcal{M}_n \cdot \hat{\mathcal{A}} \quad (84)$$

The N_{pair} amplitudes $\hat{\mathcal{A}}_n$ are uncorrelated Gaussian variables with covariance

$$\mathrm{cov}(\hat{\mathcal{A}}_n, \hat{\mathcal{A}}_m) = \delta_{nm}\Sigma_n^2. \quad (85)$$

We thus find

$$\mathrm{var}[\hat{\mathcal{A}}_{\|}(\hat{\Omega}_i)] = \sum_n (\Sigma_n \mathcal{M}_n(\hat{\Omega}_i))^2. \quad (86)$$

From this we can estimate a 95%-sensitivity to $\mathcal{A}_{\|}(\hat{\Omega}_i)$,

$$\mathcal{A}_{\|}^{95\%}(\hat{\Omega}_i) = 2\sqrt{\mathrm{var}[\hat{\mathcal{A}}_{\|}(\hat{\Omega}_i)]} \equiv [\tilde{A}_h^{95\%}(\hat{\Omega}_i)]^2, \quad (87)$$

where the last equality defines the characteristic strain $\tilde{A}_h^{95\%}(\hat{\Omega}_i)$.

We show $\tilde{A}_h^{95\%}(\hat{\Omega})$ in the top panel of Fig. 12. The qualitative *and* quantitative similarities with Fig. 2 of TMG15 are striking. The overall amplitude of our map is lower than that of TMG15, consistent with our lower 95% sensitivity estimate for the monopole amplitude $A_h^{95\%}$. Note, also, that we cannot hope to reproduce precisely the same map as TMG15, as their map is constructed from real data, i.e., a nonzero realization of the noise.

Let us point out that the pseudo-upper-limit map derived with this procedure gets *worse* if one includes *more* pulsars. Indeed, as can be seen from Eq. (86), the variance of $\hat{\mathcal{A}}_{\|}(\hat{\Omega}_i)$ is dominated by the noisiest eigenmaps. As more pulsars are added, and the number of eigenmaps grows, the

noisiest ones become increasingly noisy. Conversely, using only one pulsar pair results in a pseudo-upper-limit map with an overall amplitude lower by nearly one order of magnitude, as can be seen in the bottom panel of Fig. 12. In fact, in the right hemisphere, the map amplitude nearly vanishes, as does the pairwise-timing response function of the single pulsar pair. Clearly, adding more information should only tighten true upper limits, not make them weaker. This demonstrates once again that these Moore-Penrose pseudo-upper-limit maps cannot be interpreted as true pixel-by-pixel upper limits.
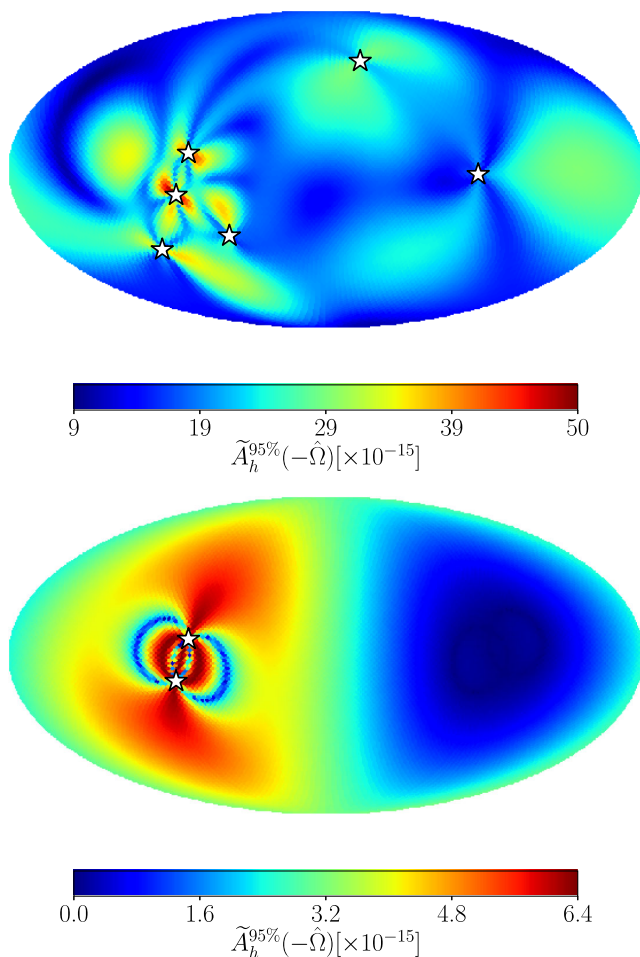

FIG. 12. Moore-Penrose pseudo-sensitivity maps produced with 6 EPTA pulsars (top) and one single EPTA pulsar pair (bottom). The qualitative and quantitative similarities of the top map with Fig. 2 of TMG15 are striking. The overall lower amplitude of our map relative to that of TMG15 is consistent with our lower 95% sensitivity to the monopole amplitude A_h. These maps *cannot* be interpreted as actual upper limits on the GWB intensity in each pixel, as a PTA cannot constrain more than N_{pair} independent components simultaneously. This is further demonstrated by the fact that the pseudo-sensitivity map produced with only two pulsars (hence less data) has lower amplitude than the one produced with 6 pulsars.

B. Constraints to spherical harmonic coefficients

We now apply the formalism of Sec. IV D to the subset of 6 EPTA pulsars used in TMG15, and compare our results to theirs. With the 15 pairs afforded by 6 pulsars, cross-correlation data can simultaneously constrain spherical-harmonic coefficients only if $\ell_{\text{max}} \leq 2$. For these values of ℓ_{max}, our 95% sensitivity estimates on $\mathcal{C}_1$ and $\mathcal{C}_2$, shown in Fig. 13, are remarkably close to their 95% upper limits in the absence of physical prior (see their Fig. 1, right panel, for $\ell_{\text{max}} = 1, 2$). However, we find that the 95% sensitivity to the monopole amplitude $\mathcal{C}_0$ gets significantly degraded as ℓ_{max} increases from 0 to 2, while the upper limit of TMG15 is virtually unaffected. We attribute this to their use of autocorrelations, which we do not include in our analysis. We note, however, that since the GWB dipole and quadrupole affect autocorrelations, increasing ℓ_{max} from 0 to 1 and then 2 should result in poorer constraints on the monopole. It is unclear why this is not what is found in TMG15.

Let us now understand TMG15's upper limits on the spherical-harmonic amplitudes for $\ell_{\text{max}} > 2$, for which there are more independent coefficients than independent data. In this case, TMG15 also define $\tilde{c}_{\ell m}$ given the overlap reduction function Γ_I through the Moore-Penrose pseudo inverse of the linear relation between Γ_I and $c_{\ell m}$. Again, unless the number of harmonic coefficients is equal to N_{pair}, $\tilde{c}_{\ell m} \neq c_{\ell m}$. For finite ℓ_{max}, there is no simple intuitive interpretation of the $\tilde{c}_{\ell m}$, but for $\ell_{\text{max}} \to \infty$, it is straightforward to show that the $\tilde{c}_{\ell m}$ are the spherical-harmonic coefficients of the *projected* GWB. Specifically, using our normalization conventions, we define $\tilde{\mathcal{A}}_{\ell m} \equiv A_h^2 \tilde{c}_{\ell m}/\sqrt{4\pi} = \mathcal{Y}_{\ell m} \cdot \hat{\mathcal{A}}_{\|}$. Using Eq. (84), we thus have
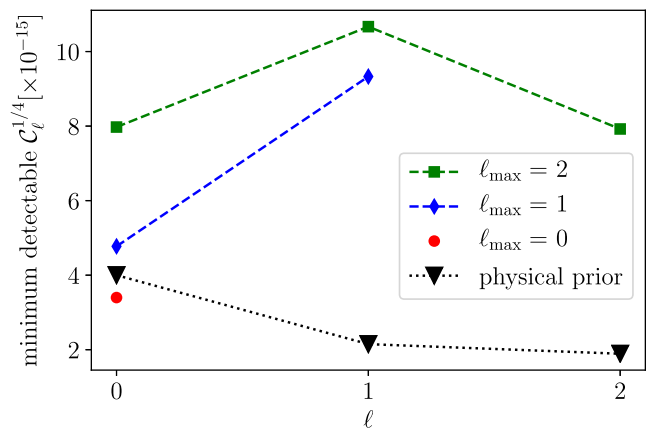

FIG. 13. 95% sensitivity estimates of the 6 EPTA pulsars to spherical harmonic amplitudes. Compare with Fig. 1 of TMG15, right panel, for $\ell_{\text{max}} \leq 2$. The physical prior assumes a monopole upper limit $A_h \leq 4 \times 10^{-15}$.
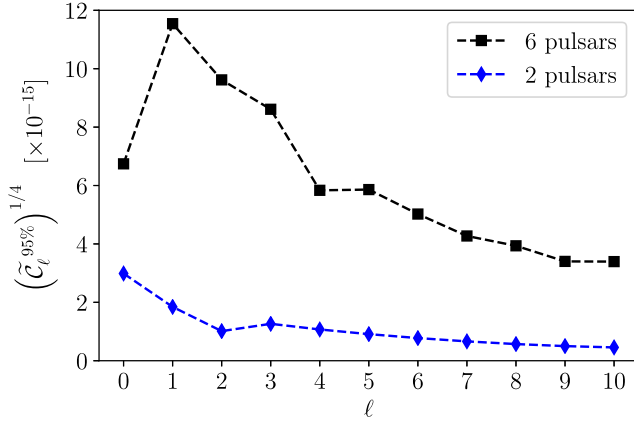
FIG. 14. 95% sensitivity to the Moore-Penrose pseudo $\mathcal{C}_\ell$'s in the limit $\ell_{\max} \to \infty$, for 6 EPTA pulsars (upper curve) and 2 EPTA pulsars (lower curve). Compare the upper curve with the right panel of Fig. 1 of TMG15, for $\ell_{\max} = 7, 10$. The $\tilde{\mathcal{C}}_\ell$ are obtained from the harmonic amplitudes of the *projection* of the GWB on the N_{pair}-dimensional space of observable maps, and therefore are *not* equal to the true harmonic amplitudes of the GWB, see text for details.

$$\tilde{\mathcal{A}}_{\ell m} = \sum_{n=1}^{N_{\text{pair}}} \hat{\mathcal{A}}_n \boldsymbol{\mathcal{Y}}_{\ell m} \cdot \boldsymbol{\mathcal{M}}_n. \tag{88}$$

Using Eq. (85), we thus obtain

$$\text{cov}(\tilde{\mathcal{A}}_{\ell m}, \tilde{\mathcal{A}}_{\ell' m'}) = \sum_{n=1}^{N_{\text{pair}}} \Sigma_n^2 (\boldsymbol{\mathcal{Y}}_{\ell m} \cdot \boldsymbol{\mathcal{M}}_n)(\boldsymbol{\mathcal{Y}}_{\ell' m'} \cdot \boldsymbol{\mathcal{M}}_n). \tag{89}$$

We then define $\tilde{\mathcal{C}}_\ell \equiv \sum_m \tilde{A}_{\ell m}^2 / (2\ell + 1)$. The mean and variance of these coefficients is given by Eqs. (61) and (62), with the substitution $\text{cov}(\mathcal{A}_{\ell m}, \mathcal{A}_{\ell' m'}) \to \text{cov}(\tilde{\mathcal{A}}_{\ell m}, \tilde{\mathcal{A}}_{\ell' m'})$, from which we can infer a 95% sensitivity $\tilde{\mathcal{C}}_\ell^{95\%}$ as in Eq. (63). We show $\tilde{\mathcal{C}}_\ell^{95\%}$ in Fig. 14. This can be compared to the right panel of Fig. 1 of TMG15. The similarity of their $\ell_{\max} = 7, 10$ results with ours (which holds in the limit $\ell_{\max} \to \infty$) is striking. Note again that we cannot expect to reproduce exactly TMG15's results, which are based on actual data, i.e., a particular realization of the noise. Here again, a sharp illustration of the fact that these Moore-Penrose pseudo-upper-limits cannot be interpreted as upper limits on the $\mathcal{C}_\ell$ is the fact that $\tilde{\mathcal{C}}_\ell^{95\%}$ decreases if one only includes 2 pulsars instead of 6, as can be seen in Fig. 14.

VIII. CONCLUSIONS

In this paper, we have built on the Fisher formalism we introduced in Paper I, and applied it to EPTA data, since this PTA is currently the only one with published limits on GWB anisotropy (TMG15). We showed through multiple examples how this framework can serve to understand the sensitivity of a PTA to a general, anisotropic GWB, and can be used to guide and optimize data analysis. We also showed how to recover existing EPTA results, and clarified their meaning, in the light of our improved understanding of the sensitivity of PTAs to GWB anisotropies. In particular, we pointed out that one cannot simultaneously detect—or constrain beyond physical priors—more than N_{pair} components of the GWB, be they pixel values or spherical harmonic amplitudes.

The fundamental tool of our formalism is the Fisher "matrix" for the direction-dependent GWB intensity. This matrix condenses the essential noise characteristics of a PTA in a compact form. It can be computed given the positions of the pulsars and their characteristic noise strains. The latter requires analyzing the timing data of individual pulsars, but not conducting any cross-correlation analysis with real data.

We started by using our formalism to find the pulsar pairs which are the most sensitive to an isotropic GWB with a power-law frequency dependence $h_c(f) \propto f^{-2/3}$. We identified the 44 best EPTA pairs (out of the 861 available) that provide 90% of the SNR^2. With the Fisher formalism, finding these best pairs for any PTA is a trivial exercise, and can be used to dramatically speed up real data analyses, at very little cost in sensitivity.

In the remainder of the paper we focused our attention to anisotropies in the GWB. We first considered the case where the GWB is assumed to be a linear combination of a finite number of known maps (no more than N_{pair}), such as spherical harmonics, coarse pixels, or hot spots in pre-determined directions. We showed that the sensitivity of a PTA to the component of the GWB on any of the basis maps systematically and quickly degrades as the number of basis functions is increased. This stems from the fact that the noise properties of standard basis maps are correlated; therefore, adding more functions to the search changes the overall covariance structure of the entire set of amplitudes. In particular, the ability to *detect* the monopole through pulsar cross correlations is systematically degraded as one searches simultaneously for anisotropies. We did not quantify the ability of a PTA to set upper limits through pulsar autocorrelations (or "common red noise" analyses), but pointed out that these are sensitive to the GWB monopole, dipole and quadrupole. Hence, upper limits on the monopole should also degrade as one accounts for anisotropies.

Having shown that standard basis maps are poorly adapted for agnostic GWB searches, we then constructed the "principal maps" of a PTA, which are the eigenmaps of the Fisher matrix. These maps form an orthonormal *and uncorrelated* basis of the N_{pair}-dimensional space of *observable* maps. In other words, they represent the N_{pair} statistically independent pieces of information accessible to a PTA, in the space of GWB maps. We showed how one

can in principle "reconstruct" the observable part of the GWB as a linear combination of principal maps detected with sufficiently high significance, a procedure similar to the making of "dirty maps" in radio interferometry. However, in contrast with a radio interferometer, for which different "visibilities" have similar noise levels, principal maps have highly unequal noise properties. As a consequence, the ability to reconstruct the GWB is strongly dependent on its overall amplitude. For the EPTA, we found that even detecting a small number of principal maps would typically require a GWB amplitude largely exceeding current bounds.

Principal maps are a useful basis if one is fully agnostic about the angular dependence of the GWB: they allow for an optimal search "under the lamppost" of a given PTA. However, the GWB is expected to be predominantly isotropic; physical considerations therefore mandate including the monopole as part of the search, and being able to identify anisotropic components of the GWB. We showed that the principal maps are in general correlated with the monopole and have a nonvanishing isotropic component, and are thus not adapted for such searches. We argued that one cannot construct a complete set of anisotropies which are uncorrelated among themselves and with the monopole, because the latter is not an eigenmap of the Fisher matrix.

In summary, there are two options when searching for a general, direction-dependent GWB within a standard Bayesian framework. On one hand, one can include the monopole in GWB searches, alongside anisotropic basis functions, but then suffer from the loss of sensitivity resulting from correlations with the monopole. On the other hand, one may search under the lamppost of PTAs with principal maps, being fully agnostic about the angular dependence of the GWB, but then give up the special status of an isotropic GWB. To circumvent these limitations, we proposed a frequentist criterion to determine whether the GWB is *consistent with isotropy*, even if one cannot determine the specific type of anisotropy. This criterion simply relies on computing the chi-squared statistics of the best-fit monopole amplitude, and comparing it with its statistical noise. We illustrated this with a fictitious, highly anisotropic GWB map, and showed that the EPTA would have to first detect its best-fit monopole with a SNR of 46 before being able to confidently assert that the data is inconsistent with a purely isotropic GWB.

The prospects for current PTAs to detect fully unknown anisotropies in the GWB appear to be somewhat limited, but the situation may improve dramatically with the several hundreds (or even thousands) of millisecond pulsars that SKA is expected to detect [15]. The Fisher formalism presented here provides a useful set of tools to address this question, under realistic assumptions for the properties of a future PTA built with SKA pulsars, and without requiring to run expensive simulations. We defer such forecasting analyses to future studies. In addition, there may be hope to detect anisotropy in the GWB through its *statistical*

properties, for instance through its correlations with tracers of large-scale structure [71]. We will explore these avenues in future work.

In closing, let us note that our Fisher formalism has far broader applications than what we considered in this work. First, we assumed an overall known frequency dependence, identical for the monopole and anisotropies. It would be straightforward to generalize our work to arbitrary frequency dependence, possibly dependent on the angular structure—something which would be difficult to do with a full-on data analysis. One could also include other sources of correlated noise, such as global clock or ephemeris errors [36,37], and go beyond current studies which ascribe them the same frequency dependence as the GWB. In addition, we could generalize our Fisher matrix beyond the weak-signal limit, and consider, for instance, the weak-anisotropy limit. Lastly, one could easily extend the formalism to include polarized GWBs. Overall, this formalism is a valuable tool that can be used to guide and cross-validate any full-blown data analysis. It will be especially important when SKA harvests hundreds of new millisecond pulsars, out of which PTAs with tens of thousands of independent pulsar pairs can be constructed. Once the GWB monopole is confidently detected by a PTA, the Fisher formalism will be extremely valuable in determining the next steps forward.

APPENDIX: PROOF THAT $\tilde{c}_i = P_\parallel(\hat{\Omega}_i)$

In this Appendix, completing Sec. VII, we prove that the vector $\vec{\tilde{c}} \equiv \boldsymbol{H}^+ \vec{\Gamma} = (\boldsymbol{H}^+ \boldsymbol{H})\vec{c}$ is the GWB angular distribution *projected* on the N_{pair}-dimensional space observable by a PTA. Thus, in general, $\vec{\tilde{c}} \neq \vec{c}$.

To see this, recall that the Moore-Penrose pseudo-inverse matrix is defined as $\boldsymbol{H}^+ = \boldsymbol{H}^{\text{T}}(\boldsymbol{H}\boldsymbol{H}^{\text{T}})^{-1}$. As a consequence, $\boldsymbol{H}^+\boldsymbol{H}$ is clearly a projector, since $[\boldsymbol{H}^+\boldsymbol{H}]^2 = \boldsymbol{H}^+\boldsymbol{H}$. Since $H_{Ii} \propto \gamma_I(\hat{\Omega}_i)$, $(\boldsymbol{H}^+\boldsymbol{H})\boldsymbol{M}_\perp = 0$ for any map $\boldsymbol{M}_\perp$ orthogonal to all the γ_I (in the limit of large N_{pix}, and replacing sums by integrals). Moreover, we have

$$(\boldsymbol{HH}^{\mathrm{T}})_{IJ} = \sum_i H_{Ij}H_{Ji}$$

$$= \left(\frac{3}{8}\frac{\Delta\Omega}{4\pi}\right)^2 \sum_i \gamma_I(\hat{\Omega}_i)\gamma_J(\hat{\Omega}_i)$$

$$\approx \frac{(3/8)^2\Delta\Omega}{4\pi}\boldsymbol{\gamma}_I \cdot \boldsymbol{\gamma}_J. \qquad (A1)$$

For a given pair index I, we define the N_{pix}-dimensional vector $\vec{\gamma}_I$ with elements $\gamma_I(\hat{\Omega}_i)$. The Jth component of the N_{pair}-dimensional vector $\boldsymbol{H}\vec{\gamma}_I$ is

$$[\boldsymbol{H}\vec{\gamma}_I]_J = \sum_i H_{Ji}\gamma_I(\hat{\Omega}_i) = \frac{3}{8}\frac{\Delta\Omega}{4\pi}\sum_i \gamma_J(\hat{\Omega}_i)\gamma_I(\hat{\Omega}_i)$$

$$\approx \frac{3}{8}\boldsymbol{\gamma}_J \cdot \boldsymbol{\gamma}_I = \left(\frac{3}{8}\frac{\Delta\Omega}{4\pi}\right)^{-1}(\boldsymbol{HH}^{\mathrm{T}})_{IJ}. \qquad (A2)$$

Therefore, the Kth component of the N_{pair}-dimensional vector $(\boldsymbol{HH}^{\mathrm{T}})^{-1}\boldsymbol{H}\vec{\gamma}_I$ is

$$[(\boldsymbol{HH}^{\mathrm{T}})^{-1}\boldsymbol{H}\vec{\gamma}_I]_K = \left(\frac{3}{8}\frac{\Delta\Omega}{4\pi}\right)^{-1}\delta_{KI}. \qquad (A3)$$

From this, we conclude that

$$(\boldsymbol{H}^+\boldsymbol{H})\vec{\gamma}_I = \boldsymbol{H}^{\mathrm{T}}(\boldsymbol{HH}^{\mathrm{T}})^{-1}\boldsymbol{H}\vec{\gamma}_I = \vec{\gamma}_I. \qquad (A4)$$

This achieves the proof that $\boldsymbol{H}^+\boldsymbol{H}$ is the projector on the space spanned by the N_{pair} maps $\boldsymbol{\gamma}_I$. Therefore, as announced, $\tilde{c}_i = P_{||}(\hat{\Omega}_i)$ is the GWB angular power projected onto that N_{pair}-dimensional space.

[1] S. Burke-Spolaor *et al.*, Astron. Astrophys. Rev. **27,** 5 (2019).
[2] C. M. F. Mingarelli, Nat. Astron. **3,** 8 (2019).
[3] S. R. Taylor, M. Vallisneri, J. A. Ellis, C. M. F. Mingarelli, T. J. W. Lazio, and R. van Haasteren, Astrophys. J. Lett. **819,** L6 (2016).
[4] L. Z. Kelley, L. Blecha, L. Hernquist, A. Sesana, and S. R. Taylor, Mon. Not. R. Astron. Soc. **471,** 4508 (2017).
[5] A. Sesana, Classical Quantum Gravity **30,** 224014 (2013).
[6] J. Simon and S. Burke-Spolaor, Astrophys. J. **826,** 11 (2016).
[7] C. M. F. Mingarelli, T. J. W. Lazio, A. Sesana, J. E. Greene, J. A. Ellis, C.-P. Ma, S. Croft, S. Burke-Spolaor, and S. R. Taylor, Nat. Astron. **1,** 886 (2017).
[8] X. Siemens, J. S. Hazboun, P. T. Baker, S. Burke-Spolaor, D. Madison, C. Mingarelli, J. Simon, and T. Smith, arXiv:1907.04960.
[9] P. D. Lasky *et al.*, Phys. Rev. X **6,** 011035 (2016).
[10] http://nanograv.org.
[11] http://www.epta.eu.org.
[12] https://www.atnf.csiro.au/research/pulsar/ppta/.
[13] http://ipta4gw.org.
[14] https://www.skatelescope.org.
[15] E. Keane *et al.*, Proc. Sci. AASKA14 (**2015**) 040.
[16] G. Janssen *et al.*, Proc. Sci. AASKA14 (**2015**) 037.
[17] Y. Ali-Haïmoud, T. L. Smith, and C. M. Mingarelli, Phys. Rev. D **102,** 122005 (2020).
[18] Z. Arzoumanian *et al.* (NANOGrav Collaboration), Astrophys. J. Lett. **905,** L34 (2020), https://iopscience.iop.org/article/10.3847/2041-8213/abd401/meta.
[19] Z. Arzoumanian *et al.* (NANOGrav Collaboration), Astrophys. J. **859,** 47 (2018).
[20] L. Lentati *et al.* (EPTA Collaboration), Mon. Not. R. Astron. Soc. **453,** 2576 (2015).
[21] R. M. Shannon *et al.*, Science **349,** 1522 (2015).
[22] J. P. W. Verbiest *et al.* (IPTA Collaboration), Mon. Not. R. Astron. Soc. **458,** 1267 (2016).
[23] C. M. Mingarelli, T. Sidery, I. Mandel, and A. Vecchio, Phys. Rev. D **88,** 062005 (2013).
[24] J. J. Blanco-Pillado and K. D. Olum, Phys. Rev. D **96,** 104046 (2017).
[25] S. Olmez, V. Mandic, and X. Siemens, J. Cosmol. Astropart. Phys. 07 (2012) 009.
[26] A. C. Jenkins and M. Sakellariadou, Phys. Rev. D **98,** 063509 (2018).
[27] J. Liu, R.-G. Cai, and Z.-K. Guo, arXiv:2010.03225 [Phys. Rev. Lett. (to be published)].
[28] J. Gair, J. D. Romano, S. Taylor, and C. M. F. Mingarelli, Phys. Rev. D **90,** 082001 (2014).
[29] S. C. Hotinli, M. Kamionkowski, and A. H. Jaffe, Open J. Astrophys. **2,** 8 (2019).
[30] S. R. Taylor and J. R. Gair, Phys. Rev. D **88,** 084001 (2013).
[31] S. R. Taylor, R. van Haasteren, and A. Sesana, Phys. Rev. D **102,** 084039 (2020).
[32] S. R. Taylor *et al.*, Phys. Rev. Lett. **115,** 041101 (2015).
[33] S. Babak *et al.*, Mon. Not. R. Astron. Soc. **455,** 1665 (2016).
[34] J. Hazboun, J. Romano, and T. Smith, J. Open Source Softw. **4,** 1775 (2019).
[35] J. S. Hazboun, J. D. Romano, and T. L. Smith, Phys. Rev. D **100,** 104028 (2019).
[36] C. Tiburzi, G. Hobbs, M. Kerr, W. A. Coles, S. Dai, R. N. Manchester, A. Possenti, R. M. Shannon, and X. P. You, Mon. Not. R. Astron. Soc. **455,** 4339 (2016).
[37] M. Vallisneri *et al.*, Astrophys. J. **893,** 112 (2020).
[38] R. w. Hellings and G. s. Downs, Astrophys. J. **265,** L39 (1983).
[39] J. S. Hazboun, J. D. Romano, and T. L. Smith, Phys. Rev. D **100,** 104028 (2019).
[40] E. S. Phinney, arXiv:astro-ph/0108028.

[41] S. Chen, A. Sesana, and W. Del Pozzo, Mon. Not. R. Astron. Soc. **470,** 1738 (2017).

[42] M. Anholm, S. Ballmer, J. D. E. Creighton, L. R. Price, and X. Siemens, Phys. Rev. D **79,** 084030 (2009).

[43] X. Siemens, J. Ellis, F. Jenet, and J. D. Romano, Classical Quantum Gravity **30,** 224015 (2013).

[44] G. Desvignes *et al.* (EPTA Collaboration), Mon. Not. R. Astron. Soc. **458,** 3341 (2016).

[45] Z. Arzoumanian *et al.* (NANOGrav Collaboration), Astrophys. J. **813,** 65 (2015).

[46] P. Reich and W. Reich, Astron. Astrophys. Suppl. Ser. **63,** 205 (1986), https://ui.adsabs.harvard.edu/abs/1986A% 26AS...63..205R/abstract.

[47] P. Reich, J. C. Testori, and W. Reich, Astron. Astrophys. **376,** 861 (2001).

[48] G. Desvignes *et al.*, Mon. Not. R. Astron. Soc. **458,** 3341 (2016).

[49] R. van Haasteren and Y. Levin, Mon. Not. R. Astron. Soc. **428,** 1147 (2013).

[50] J. A. Ellis and R. van Haasteren, https://doi.org/10.5281/ zenodo.1037579.

[51] J. A. Ellis, M. Vallisneri, S. R. Taylor, and P. T. Baker, http:// ascl.net/1912.015.

[52] http://www.epta.eu.org/aom.html.

[53] R. Caballero *et al.*, Mon. Not. R. Astron. Soc. **457,** 4421 (2016).

[54] D. R. B. Yardley *et al.*, Mon. Not. R. Astron. Soc. **414,** 1777 (2011).

[55] K. M. Górski, E. Hivon, A. J. Banday, B. D. Wand elt, F. K. Hansen, M. Reinecke, and M. Bartelmann, Astrophys. J. **622,** 759 (2005).

[56] B. Allen and A. C. Ottewill, Phys. Rev. D **56,** 545 (1997).

[57] N. J. Cornish, Classical Quantum Gravity **18,** 4277 (2001).

[58] S. Mitra, S. Dhurandhar, T. Souradeep, A. Lazzarini, V. Mandic, S. Bose, and S. Ballmer, Phys. Rev. D **77,** 042002 (2008).

[59] E. Thrane, S. Ballmer, J. D. Romano, S. Mitra, D. Talukder, S. Bose, and V. Mandic, Phys. Rev. D **80,** 122002 (2009).

[60] A. I. Renzini and C. R. Contaldi, Mon. Not. R. Astron. Soc. **481,** 4650 (2018).

[61] N. Cornish, Classical Quantum Gravity **19,** 1279 (2002).

[62] H. Kudoh and A. Taruya, Phys. Rev. D **71,** 024025 (2005).

[63] N. Seto and A. Cooray, Phys. Rev. D **70,** 123005 (2004).

[64] A. Taruya and H. Kudoh, Phys. Rev. D **72,** 104015 (2005).

[65] H. Kudoh, A. Taruya, T. Hiramatsu, and Y. Himemoto, Phys. Rev. D **73,** 064006 (2006).

[66] A. Taruya, Phys. Rev. D **74,** 104022 (2006).

[67] N. J. Cornish and R. van Haasteren, arXiv:1406.4511.

[68] J. D. Romano, S. R. Taylor, N. J. Cornish, J. Gair, C. M. Mingarelli, and R. van Haasteren, Phys. Rev. D **92,** 042003 (2015).

[69] E. H. Moore, Bull. Am. Math. Soc. **26,** 394 (1920).

[70] R. Penrose, Proc. Cambridge Philos. Soc. **51,** 406 (1955).

[71] D. Alonso, G. Cusin, P. G. Ferreira, and C. Pitrou, Phys. Rev. D **102,** 023002 (2020).