Lectures on Gravitational Wave Data Analysis

John T. Whelan

Lectures given at IUCAA, Pune, 2014 January 16-17

This document contains four sets of lecture notes, each with its own page and section numbering:

1. The Geometry of Gravitational Wave Detection
2. Probability, Statistics, Fourier Analysis and Signal Processing
3. Continuous Wave Data Analysis: Fully Coherent Methods (extra notes)
4. Searches for a Stochastic Gravitational-Wave Background (extra notes)

This document has been assigned Document Control Center number LIGO-T1400017-v1. This material is based upon work supported by the National Science Foundation under Grant Numbers PHY-0855494 and PHY-1207010. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

The Geometry of Gravitational Wave Detection

John T. Whelan

Lecture given at IUCAA, Pune, 2014 January 16

Contents

0 Preview

Most important formula in lecture: strain measured by detector

$$h = \overleftrightarrow{h} : \overleftrightarrow{d} = h_+ \underbrace{\overleftrightarrow{e}_+ : \overleftrightarrow{d}}_{F_+} + h_\times \underbrace{\overleftrightarrow{e}_\times : \overleftrightarrow{d}}_{F_\times} \tag{0.1}$$

where

$$\overleftrightarrow{e}_+ = \vec{\ell} \otimes \vec{\ell} - \vec{m} \otimes \vec{m} \tag{0.2a}$$

$$\overleftrightarrow{e}_\times = \vec{\ell} \otimes \vec{m} + \vec{m} \otimes \vec{\ell} \tag{0.2b}$$

and

$$\overleftrightarrow{d} = \frac{\vec{u} \otimes \vec{u} - \vec{v} \otimes \vec{v}}{2} \tag{0.3}$$

Note $\vec{\imath}, \vec{\jmath}, \vec{k}, \vec{\ell}, \vec{m}, \vec{u}, \vec{v}$ are all unit vectors.

1 Propagating Gravitational Waves

1.0 Reminders from General Relativity

Given a spacetime on which you've defined some coördinates $\{x^\mu\}$ where $\mu \in \{0, 1, 2, 3\}$, the metric tensor can be written in terms of its components $\{g_{\mu\nu}\}$, and the spacetime interval is[1]

$$ds^2 = g_{\mu\nu}\, dx^\mu\, dx^\nu \tag{1.1}$$

$\{g_{\mu\nu}\}$ are the components of a tensor field, which means they are functions of the spacetime coördinates $\{x^\mu\}$. We can consider a different set of coördinates $\{x^{\bar{\mu}}\}$, and as with any tensor, the metric tensor has a new set of components $\{g_{\bar{\mu}\bar{\nu}}\}$ associated with the corresponding basis, defined by

$$g_{\bar{\mu}\bar{\nu}} = g_{\mu\nu} \frac{\partial x^\mu}{\partial x^{\bar{\mu}}} \frac{\partial x^\nu}{\partial x^{\bar{\nu}}} \tag{1.2}$$

which ensures that

$$ds^2 = g_{\mu\nu}\, dx^\mu\, dx^\nu = g_{\bar{\mu}\bar{\nu}}\, dx^{\bar{\mu}}\, dx^{\bar{\nu}} \tag{1.3}$$

is the same no matter which coördinate system we use. Note that this notation considers $\bar{\mu}$ and μ to be different indices, and stresses that there are different sets of coördinates, so that there's no real connection between x^1 and $x^{\bar{1}}$. (They might for example be Cartesian coördinates $\{t, x, y, z\}$ and double-null coördinates $\{u, v, \theta, \phi\}$.) This does conceal, however, that as functions of four variables, $\{g_{\mu\nu}\}$ and $\{g_{\bar{\mu}\bar{\nu}}\}$ take different sets of arguments to describe the geometry at the same spacetime point. So we could write, more completely,

$$\bar{g}_{\alpha\beta}(\{\bar{x}^\gamma\}) = g_{\mu\nu}(\{x^\lambda\}) \frac{\partial x^\mu}{\partial \bar{x}^\alpha} \frac{\partial x^\nu}{\partial \bar{x}^\beta} \tag{1.4}$$

This form is especially useful when considering an infinitesimal coördinate transformation, in which $\bar{x}^\mu = x^\mu + \xi^\mu$ where ξ^μ is in some sense small. Then, to first order in $\{\xi^\mu\}$, the change in the components of the metric tensor can be shown to be

$$\bar{g}_{\mu\nu}(\{x^\lambda\}) - g_{\mu\nu}(\{x^\lambda\}) = -\nabla_\mu \xi_\nu - \nabla_\nu \xi_\mu \tag{1.5}$$

where

$$\nabla_\mu \xi_\nu = \frac{\partial \xi_\nu}{\partial x_\mu} - \Gamma^\lambda_{\mu\nu} \xi_\lambda \tag{1.6}$$

is the usual covariant derivative defined in terms of the Christoffel symbols

$$\Gamma^\lambda_{\mu\nu} = \frac{g^{\lambda\rho}}{2} \left(\frac{\partial g_{\mu\nu}}{\partial \rho} - \frac{\partial g_{\rho\nu}}{\partial \mu} - \frac{\partial g_{\mu\rho}}{\partial \nu} \right) \tag{1.7}$$

and $\{g^{\mu\nu}\}$ is the matrix inverse of $\{g_{\mu\nu}\}$, so that $g^{\mu\lambda} g_{\lambda\nu} = \delta^\mu_\nu$.

The spacetime interval of special relativity is associated with the Minkowski metric, which can be written as

$$\eta_{\mu\nu}\, dx^\mu\, dx^\nu = -c^2\, dt^2 + \delta_{ij}\, dx^i\, dx^j \tag{1.8}$$

the linearized theory of gravity assumes that the metric tensor can be written as some background metric plus a small perturbation. Choosing Minkowski as the background metric, we have

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu} \tag{1.9}$$

we can make a gauge transformation (small coördinate change)

$$h_{\mu\nu} \to h_{\mu\nu} - \partial_\mu \xi_\nu - \partial_\nu \xi_\mu \tag{1.10}$$

[1] We're using the Einstein summation convention: there is an implicit sum over repeated $\mu, \nu, \lambda, \ldots$ indices, e.g., $\sum_{\mu=0}^{3}$. In this linearized 3-space language, there's no strong distinction between subscripted and superscripted indices, but we should have one "upstairs" and one "downstairs" for the implicit sum, and have the same "unbalanced" indices in the same places on both sides of an equation.

and represent the same physical spacetime in slightly different coördinates. This gauge transformation is analogous to the transformation that allows us to change $\{A^\mu\} = \{\varphi, \vec{A}\}$ in electromagnetism and not change the physical electric and magnetic fields. One convenient gauge condition that we can enforce is the so-called transverse-traceless-temporal gauge, in which $h_{0\mu} = 0$, $\eta^{\mu\lambda}\partial_\lambda h_{\mu\nu} = 0$, and $\eta^{\mu\nu}h_{\mu\nu} = 0$. The temporal part of the gauge condition means that we can just talk about the spatial components of the metric perturbation, and in fact we won't need to talk about spacetime indices outside of this introductory review. Instead, we have spatial components $\{h_{ij}\}$ where $\delta^{ij}h_{ij} = 0$ and $\delta^{ik}\partial_k h_{ij} = 0$.

1.1 The polarization decomposition

If we describe linearized GR in the transverse-traceless-temporal gauge, the spacetime interval is replaced by[2]

$$ds^2 = -c^2\,dt^2 + (\delta_{ij} + h_{ij})\,dx^i\,dx^j \tag{1.11}$$

where

$$\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \tag{1.12}$$

is the Kronecker delta and $\{h_{ij}\}$ are small perturbations. In this gauge, the components $\{h_{ij}\}$ all obey the wave equation

$$\left(-\frac{1}{c^2}\frac{\partial^2}{\partial t^2} + \nabla^2\right) h_{ij} = 0 \tag{1.13}$$

Because the metric (1.11) is the Minkowski metric plus a small spatial perturbation, we can do all of the important calculations for GW detection in the notation of vectors in a three-dimensional Euclidean space.

A wave coming from a single distant source can be treated as a plane wave propagating along a unit vector $\vec{k}$ which points from the source to the observer. If we choose our coördinate axes so that this unit vector has components

$$\{k_i\} \equiv \boldsymbol{k} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \tag{1.14}$$

the components of the metric tensor perturbation are

$$\{h_{ij}\} \equiv \mathbf{h} = \begin{pmatrix} h_+ & h_\times & 0 \\ h_\times & -h_+ & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{1.15}$$

where the two independent degrees of freedom h_+ and $h_\times$ are functions of $t - \vec{k}\cdot\vec{r}/c$. We can also write this as

$$h_{ij} = h_+ e_{+\,ij} + h_\times e_{\times\,ij} \tag{1.16}$$

or

$$\mathbf{h} = h_+\mathbf{e}_+ + h_\times\mathbf{e}_\times \tag{1.17}$$

in terms of the matrices

$$\{e_{+\,ij}\} \equiv \mathbf{e}_+ = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{and} \quad \{e_{\times\,ij}\} \equiv \mathbf{e}_\times = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{1.18}$$

It's useful, however, to be able to write things without referring to a specific coördinate system.[3] We think of a vector $\vec{v}$ as a physical object with magnitude and direction, not just as the collection of

[2] We're again using the Einstein summation convention: there is an implicit sum over repeated $i, j, k, \ldots$ indices, e.g., $\sum_{i=1}^{3}$.

[3] Our "coördinate-free" notation will only be covariant in the three-dimensional sense; since it assumes we're still in the TT gauge.

three numbers $\{v_i\}$. In fact, we can resolve the vector $\vec{v}$ in different bases, e.g., $\{\vec{e}_i\}$ and $\{\vec{e}'_i\}$. There will be different components $\{v_i\}$ and $\{v_{i'}\}$ defined by

$$v_i = \vec{e}_i \cdot \vec{v} \qquad \text{vs} \qquad v_{i'} = \vec{e}'_i \cdot \vec{v} \tag{1.19}$$

and we could collect them into different 3×1 matrices (column vectors)

$$\{v_i\} \equiv \boldsymbol{v} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \qquad \text{vs} \qquad \{v_{i'}\} \equiv \boldsymbol{v}' = \begin{pmatrix} v_{1'} \\ v_{2'} \\ v_{3'} \end{pmatrix} \tag{1.20}$$

The matrices $\boldsymbol{v}$ and $\boldsymbol{v}'$ contain different triples of numbers, but they are just different ways of describing the same object $\vec{v}$. The vector $\vec{v}$ is what's fundamental, because "physics is about things."[4] Sometimes we'll want to refer to a complicated object in a coördinate-independent way. For that purpose it's useful to introduce what's known as *abstract index notation*, using Latin indices from the front of the alphabet. So if I write v_a, this is the same as writing $\vec{v}$. It refers not to a particular component or set of components, but to the object which has components $\{v_i\}$ when resolved in the basis $\vec{e}_i$.

Now we return to consideration of equation (1.16) which describes the metric perturbations $\{h_{ij}\}$, or the corresponding matrix equation (1.17). We'd like to describe this in an abstract way which doesn't rely on a particular set of basis vectors. We can do this by considering the construction of the matrices $\mathbf{e}_+$ and $\mathbf{e}_\times$ from the components of unit vectors $\vec{\ell}$ and $\vec{m}$ which form an orthonormal triple with $\vec{k}$. In the coördinate system we're working in so far those vectors have components

$$\{\ell_i\} \equiv \boldsymbol{\ell} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \qquad \text{and} \qquad \{m_i\} \equiv \boldsymbol{m} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \tag{1.21}$$

and we can write the plus and cross polarization matrices as

$$\mathbf{e}_+ = \boldsymbol{\ell}\boldsymbol{\ell}^{\mathrm{T}} - \boldsymbol{m}\boldsymbol{m}^{\mathrm{T}} \tag{1.22a}$$
$$\mathbf{e}_\times = \boldsymbol{\ell}\boldsymbol{m}^{\mathrm{T}} + \boldsymbol{m}\boldsymbol{\ell}^{\mathrm{T}} \tag{1.22b}$$

Or, in terms of components,

$$e_{+\,ij} = \ell_i \ell_j - m_i m_j \tag{1.23a}$$
$$e_{\times\,ij} = \ell_i m_j + m_i \ell_j \tag{1.23b}$$

We know how to talk about a vector like $\vec{\ell}$ (in arrow notation) or ℓ^a (in abstract index notation). If we think about the matrix $\{e_{+\,ij}\}$ as making up the components, in a particular basis, of a tensor, we can describe that tensor in abstract index notation as $e_{+\,ab}$ and define it and its counterpart $e_{\times\,ab}$ by

$$e_{+\,ab} = \ell_a \ell_b - m_a m_b \tag{1.24a}$$
$$e_{\times\,ab} = \ell_a m_b + m_a \ell_b \tag{1.24b}$$

If we like to use arrow notation, we can define these two basis tensors equivalently using the tensor (dyad) product as

$$\overleftrightarrow{e}_+ = \vec{\ell} \otimes \vec{\ell} - \vec{m} \otimes \vec{m} \tag{1.25a}$$
$$\overleftrightarrow{e}_\times = \vec{\ell} \otimes \vec{m} + \vec{m} \otimes \vec{\ell} \tag{1.25b}$$

That then allows us to write the general plane wave propagating along $\vec{k}$ in covariant tensor notation as

$$\overleftrightarrow{h} = h_+ \overleftrightarrow{e}_+ + h_\times \overleftrightarrow{e}_\times \tag{1.26}$$

[4] T. Dray and C. A. Manogue, *The Geometry of Vector Calculus*, http://www.math.oregonstate.edu/BridgeBook/book/revper/things

where $\overleftrightarrow{e}_+$ and $\overleftrightarrow{e}_\times$ are defined according to (1.25) from the orthonormal basis vectors $\vec{\ell}$ and $\vec{m}$ perpendicular to $\vec{k}$, and h_+ and $h_\times$ are functions of $t - \vec{k}\cdot\vec{r}/c$.

Examples of other tensors you may have seen are the inertia tensor $\overleftrightarrow{I}$ from rigid body motion, the stress tensor $\overleftrightarrow{T}$ from statics or electromagnetism, and the quadrupole moment tensor $\overleftrightarrow{Q}$ from a multipole expansion. These are all symmetric, second-rank tensors like $\overleftrightarrow{h}$.

1.2 Exercise: Normalization of Basis Tensors

Dot products involving tensors can be defined in straightforward ways using the abstract index notation as a guide. For example, $\overleftrightarrow{T}\cdot\vec{v}$ is the vector which can be written in abstract index notation[5] as $[\overleftrightarrow{T}\cdot\vec{v}]_a = T_{ab}v^b$, $\overleftrightarrow{S}\cdot\overleftrightarrow{T}$ is the tensor $[\overleftrightarrow{S}\cdot\overleftrightarrow{T}]_{ab} = S_{ac}T^c{}_b$ and the double dot product $\overleftrightarrow{S}:\overleftrightarrow{T}$ is the scalar which is the "trace" of this: $\overleftrightarrow{S}:\overleftrightarrow{T} = S_{ab}T^{ba}$

1. Use the abstract index notation to show that the double dot product of two dyads is $(\vec{u}\otimes\vec{v}):(\vec{a}\otimes\vec{b}) = (\vec{v}\cdot\vec{a})(\vec{b}\cdot\vec{w})$

2. Calculate the four double dot products $\overleftrightarrow{e}_A:\overleftrightarrow{e}_B$, where A and B can each be $+$ or $\times$. (I.e., calculate $\overleftrightarrow{e}_+:\overleftrightarrow{e}_+$, $\overleftrightarrow{e}_+:\overleftrightarrow{e}_\times$, etc.)

1.3 Change of Basis

The propagation direction $\vec{k}$ does not uniquely specify the construction of basis tensors $\overleftrightarrow{e}_+$ and $\overleftrightarrow{e}_\times$; we also need to choose a vector $\vec{\ell}$ in the plane perpendicular to $\vec{k}$. (This then uniquely determines $\vec{m} = \vec{k}\times\vec{\ell}$.) For different types of sources and analyses, there may be a choice of polarization basis which is particularly convenient. It may also be desirable to convert between a convenient polarization basis and some canonical reference basis constructed only from the propagation direction and some absolute reference directions.

For example, suppose that we specify the source location in equatorial coördinates in terms of its right ascension α and declination δ. This is equivalent to specifying $\vec{k}$; we can assign a reference basis to each sky position by producing a prescription for defining additional unit vectors $\vec{\imath}$ and $\vec{\jmath}$ which, together with $\vec{k}$, form a an orthonormal set. One prescription is to require $\vec{\imath}$ to be parallel to the celestial equator, i.e., perpendicular to the direction of the Earth's axis. We choose $\vec{\imath}$ to point in the direction of *decreasing* right ascension, so that the third vector $\vec{\jmath} = \vec{k}\times\vec{\imath}$ points into the Northern celestial hemisphere. This is illustrated in figure 1. From these unit vectors $\vec{\imath}$ and $\vec{\jmath}$ we can construct a reference polarization basis for traceless symmetric tensors transverse to $\vec{k}$:

$$\overleftrightarrow{\varepsilon}_+ = \vec{\imath}\otimes\vec{\imath} - \vec{\jmath}\otimes\vec{\jmath} \tag{1.27a}$$
$$\overleftrightarrow{\varepsilon}_\times = \vec{\imath}\otimes\vec{\jmath} + \vec{\jmath}\otimes\vec{\imath} \tag{1.27b}$$

The basis vectors $\vec{\ell}$ and $\vec{m}$, from which the source's natural polarization basis is constructed, lie in the same plane as $\vec{\imath}$ and $\vec{\jmath}$, since they're all perpendicular to the propagation direction $\vec{k}$. The natural basis can be located relative to the reference basis by the angle from $\vec{\imath}$ to $\vec{\ell}$, measured counter-clockwise around $\vec{k}$, as shown in figure 2. As in the usual rotation of basis vectors, we can resolve $\vec{\ell}$ and $\vec{m}$ in terms of $\vec{\imath}$ and $\vec{\jmath}$:

$$\vec{\ell} = \vec{\imath}\cos\psi + \vec{\jmath}\sin\psi \tag{1.28a}$$
$$\vec{m} = -\vec{\imath}\sin\psi + \vec{\jmath}\cos\psi \tag{1.28b}$$

We can substitute (1.28) into (1.25) to get $\overleftrightarrow{e}_+$ and $\overleftrightarrow{e}_\times$ in terms of $\overleftrightarrow{\varepsilon}_+$ and $\overleftrightarrow{\varepsilon}_\times$. The one tricky thing is the tensor product, which is

[5] In an expression like $\vec{u}\cdot\vec{v} = u^a v_a = u_a v^a$, it doesn't matter which abstract index is written "upstairs" and which "downstairs", as long as there is one in each position.

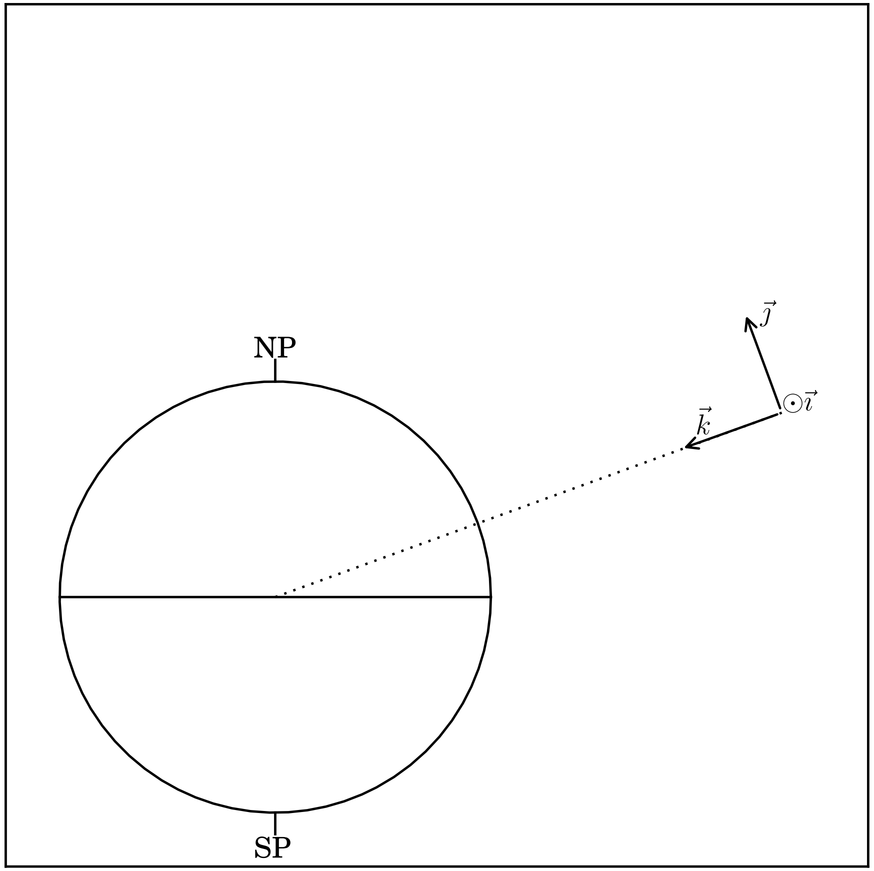

Figure 1: Definition of the unit vectors $\vec{\imath}$ and $\vec{\jmath}$, orthogonal to the propagation direction $\vec{k}$, used to define the reference polarization basis tensors $\overleftrightarrow{\varepsilon}_+$ and $\overleftrightarrow{\varepsilon}_\times$ via (1.27). The unit vector $\vec{\imath}$ is orthogonal both to $\vec{k}$ (which points from the source to the observer) and to the axis of the equatorial coördinate system. (I.e., it is parallel to the celestial equator.) The unit vector $\vec{\jmath}$ ($= \vec{k} \times \vec{\imath}$) points into the Northern hemisphere.

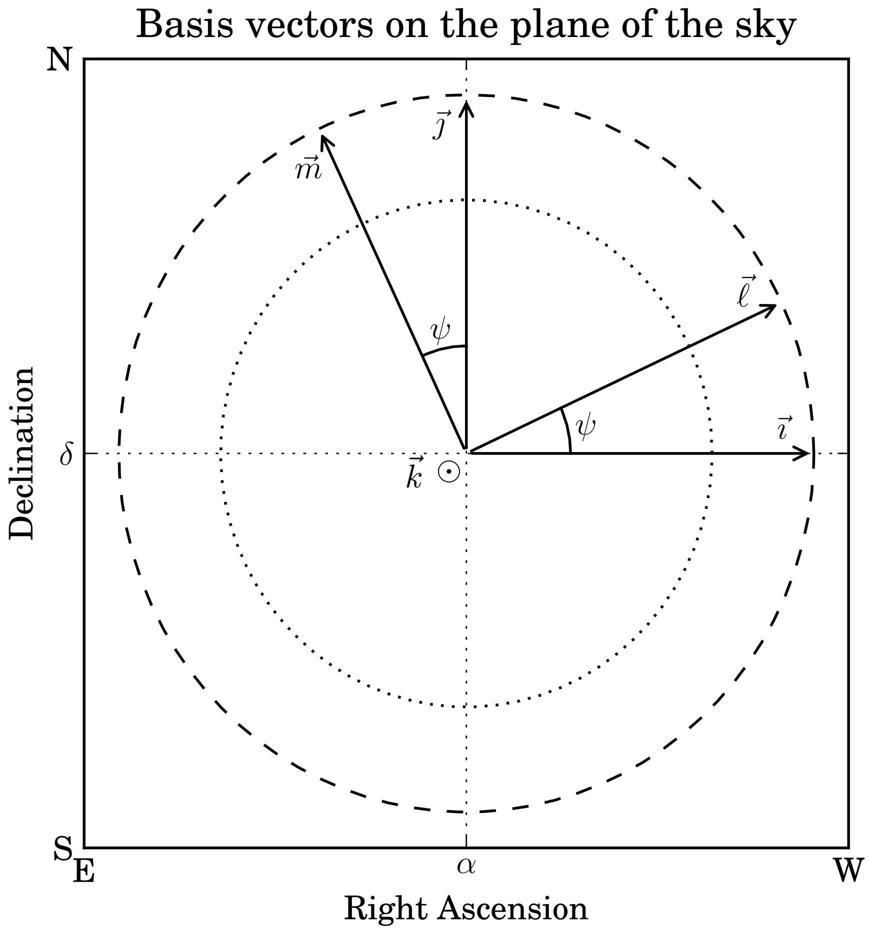

Figure 2: Rotation of basis. The natural polarization basis tensors $\overleftrightarrow{e}_+$ and $\overleftrightarrow{e}_\times$ are created from the unit vectors $\vec{\ell}$ and $\vec{m}$. The reference polarization basis tensors $\overleftrightarrow{\varepsilon}_+$ and $\overleftrightarrow{\varepsilon}_\times$ are created from the unit vectors $\vec{\imath}$ and $\vec{\jmath}$ via (1.25). The polarization angle ψ which completes the specification of $\vec{\ell}$ and therefore of the natural polarization basis, is measured from $\vec{\imath}$ to $\vec{\ell}$, counter-clockwise around $\vec{k}$. (For the example illustrated in this figure, ψ lies between 0 and $\pi/2$.)

not commutative. One example term is

$$\vec{\ell}\otimes\vec{m} = -\vec{\imath}\otimes\vec{\imath}\cos\psi\sin\psi+\vec{\imath}\otimes\vec{\jmath}\cos^2\psi+\vec{\jmath}\otimes\vec{\imath}\sin^2\psi+\vec{\jmath}\otimes\vec{\jmath}\sin\psi\cos\psi \tag{1.29}$$

1.4 Exercise: Change of Polarization Basis

Do the algebra and apply the double angle formulas to show that

$$\overleftrightarrow{e}_+ = \overleftrightarrow{\varepsilon}_+ \cos 2\psi + \overleftrightarrow{\varepsilon}_\times \sin 2\psi \tag{1.30a}$$
$$\overleftrightarrow{e}_\times = -\overleftrightarrow{\varepsilon}_+ \sin 2\psi + \overleftrightarrow{\varepsilon}_\times \cos 2\psi \tag{1.30b}$$

This shows that the specification of the polarization basis associated with a particular source requires three angles: the right ascension α and declination δ to specify the sky position and thus the propagation direction $\vec{k}$, and an additional *polarization angle* ψ to define the orientation of the preferred polarization basis $\{\overleftrightarrow{e}_+, \overleftrightarrow{e}_\times\}$ relative to some reference basis like $\{\overleftrightarrow{\varepsilon}_+, \overleftrightarrow{\varepsilon}_\times\}$. Note that, since (1.30) contains only trig functions of 2ψ, the polarization angle ψ can generally be taken to range over π rather than 2π. (Changing ψ to $\psi+\pi$ would flip both $\vec{\ell}$ and $\vec{m}$, and leave the polarization basis tensors unchanged.)

2 Interaction with a Detector

2.1 The Detector Tensor

The simplest description of an interferometric gravitational-wave detector (see figure 3) is to say it measures the difference between the lengths of its arms. That's a bit too simplistic, though, and it opens the trap of outsiders asking "how can a GW measure anything if both the spacetime and the detector are stretching?" So instead, let's note that it measures the difference in phase of

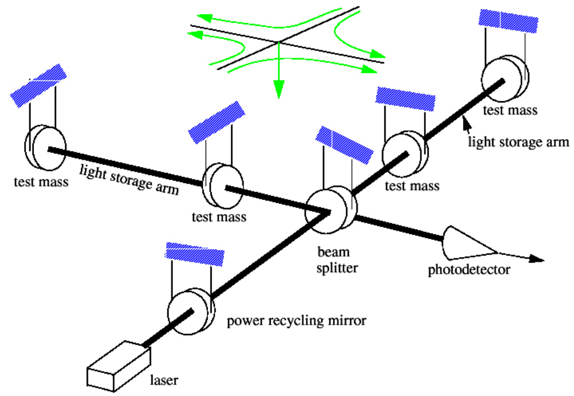

Figure 3: Schematic of an interferometric gravitational wave detector. Image by Ray Frey. We define the basis vectors $\vec{u}$ and $\vec{v}$ to lie parallel to the two arms.

light which has gone down and back one arm versus the other. This is equivalent to measuring the difference in the roundtrip travel time of photons down the two arms, a measurement which can be made entirely locally, and without worrying about the effects of the wave, since the time components of the spacetime metric are not changed in the transverse-traceless-temporal gauge. We also use the fact that points with constant coördinates in the TT gauge are in free fall, i.e., experience no non-gravitational forces, to note that if define coördinates so that the beam splitter is at the origin and the end mirror of an arm is at position $(x^1, x^2, x^3) = (L_0, 0, 0)$, those coördinates will not be changed by the passage of a gravitational wave. We can now consider the trajectory of a photon going down the arm from $(0,0,0)$ to $(L_0,0,0)$ and back. Its trajectory $x^1(t)$ will be given by solving the differential equation

$$ds^2 = -c^2\,dt^2 + (1+h_{11})(dx^1)^2 = 0 \tag{2.1}$$

or, working to first order in the perturbation h_{11},

$$dt = \frac{\sqrt{1+h_{11}}}{c}\left|dx^1\right| \approx \left(1+\frac{1}{2}h_{11}\right)\frac{\left|dx^1\right|}{c} \tag{2.2}$$

Now, in general, we have to worry about the fact that $h_{11}\left(t-\frac{\vec{k}\cdot\vec{r}}{c}\right)$ is a function of space and time, but if the travel time of the photon is short compared to the period of the gravitational wave, or equivalently if the gravitational wavelength is large compared to the length of the arm, we are in the so-called long-wavelength limit, and we can approximate h_{11} as a constant during the trajectory of the photon. Then the time the photon takes to go from $(0,0,0)$ to $(L_0,0,0)$ and back is

$$T_1 = \left(1+\frac{1}{2}h_{11}\right)\frac{2L_0}{c} =: \frac{2L_1}{c} \tag{2.3}$$

As usual, we would like to approach this problem in a coördinate-free way, so we define a unit vector $\vec{u}$ along the arm, which has components

$$\{u^i\} \equiv \boldsymbol{u} = \begin{pmatrix}1\\0\\0\end{pmatrix} \tag{2.4}$$

in the specialized coördinate system we've used to describe the detector arm. The metric perturbation component h_{11} appearing in (2.3) can be written as

$$h_{11} = u^i h_{ij} u^j = \boldsymbol{u}^{\mathrm{T}}\mathbf{h}\boldsymbol{u} \tag{2.5}$$

This quantity can be written in a basis independent way, using either abstract index or arrow notation, as

$$u^a h_{ab} u^b = \vec{u}\cdot\overleftrightarrow{h}\cdot\vec{u} \tag{2.6}$$

so that

$$L_{\vec{u}} = L_0\left(1+\frac{1}{2}h_{ab}u^a u^b\right) = L_0\left(1+\frac{1}{2}\vec{u}\cdot\overleftrightarrow{h}\cdot\vec{u}\right) \tag{2.7}$$

where we have written $L_{\vec{u}}$ to emphasize that the interferometer arm is parallel to the unit vector $\vec{u}$.

Now consider an interferometer with one arm along the unit vector $\vec{u}$ and the other along the unit vector $\vec{v}$.[6] The interferometer measures the difference in roundtrip times down the two arms; this divided by $2L_0/c$ is known as the gravitational wave strain $h(t)$:

$$\begin{aligned} h &= \frac{L_{\vec{u}}-L_{\vec{v}}}{L_0} = \frac{1}{2}\left(\vec{u}\cdot\overleftrightarrow{h}\cdot\vec{u}-\vec{v}\cdot\overleftrightarrow{h}\cdot\vec{v}\right) \\ &= h_{ab}\frac{u^a u^b - v^a v^b}{2} = h_{ab}d^{ab} = \overleftrightarrow{h} : \overleftrightarrow{d} \end{aligned} \tag{2.8}$$

where we have defined the detector tensor

$$d^{ab} = \frac{u^a u^b - v^a v^b}{2} \qquad \text{or} \qquad \overleftrightarrow{d} = \frac{\vec{u}\otimes\vec{u}-\vec{v}\otimes\vec{v}}{2} \tag{2.9}$$

Subject to the approximations of this section (primarily the long-wavelength limit), the detector response to a particular gravitational wave tensor $\overleftrightarrow{h}$ is determined by the detector tensor $\overleftrightarrow{d}$. (We also need to know the location of the detector, in order to tell what phase of the gravitational wave is hitting at what local time.)

2.2 Antenna response functions

If we recall the resolution (1.26) of $\overleftrightarrow{h}$ into a preferred polarization basis, the strain measured by a detector is

$$h = \overleftrightarrow{h} : \overleftrightarrow{d} = (h_+ \overleftrightarrow{e}_+ + h_\times \overleftrightarrow{e}_\times) : \overleftrightarrow{d} = h_+ F_+ + h_\times F_\times \tag{2.10}$$

[6]Note that in this approach we do *not* need to assume the arms are perpendicular, as would be the case if we did everything in a Cartesian coördinate system where $\vec{u}$ had components $(1,0,0)$ and $\vec{v}$ had components $(0,1,0)$.

where the *antenna pattern factors* are given by

$$F_+ = \overleftrightarrow{d} : \overleftrightarrow{e}_+ = d^{ab} e_{+\,ab} \tag{2.11a}$$

$$F_\times = \overleftrightarrow{d} : \overleftrightarrow{e}_\times = d^{ab} e_{\times\,ab} \tag{2.11b}$$

For a given detector at a given time (i.e., for a fixed detector tensor $\overleftrightarrow{d}$), F_+ and $F_\times$ will depend on the three angles defining the sky position and polarization basis with respect to some reference system. E.g., using equatorial coördinates, they will depend on right ascension, declination and the polarization angle ψ. Note that we don't need to define F_+ and $F_\times$ as explicit complicated functions of detector latitude, longitude, etc. The fundamental conceptual definition of the antenna pattern functions is (2.11), and all the rest is just working out the dot products. In particular, we can separate out the dependence on the polarization angle ψ; if we know the sky position of the source, we can construct the reference polarization basis $\{\overleftrightarrow{\varepsilon}_+, \overleftrightarrow{\varepsilon}_\times\}$, and for a given detector at a given sidereal time, we can construct the combinations

$$a = \overleftrightarrow{d} : \overleftrightarrow{\varepsilon}_+ = d^{ab} \varepsilon_{+\,ab} \tag{2.12a}$$

$$b = \overleftrightarrow{d} : \overleftrightarrow{\varepsilon}_\times = d^{ab} \varepsilon_{\times\,ab} \tag{2.12b}$$

from which we get

$$F_+(\alpha,\delta,\psi) = \quad a(\alpha,\delta)\,\cos 2\psi + b(\alpha,\delta)\,\sin 2\psi \tag{2.13a}$$

$$F_\times(\alpha,\delta,\psi) = -\,a(\alpha,\delta)\,\sin 2\psi + b(\alpha,\delta)\,\cos 2\psi \tag{2.13b}$$

2.3 Exercise: Invariant Combination

Consider the combination

$$F_+^2 + F_\times^2 = a^2 + b^2 \tag{2.14}$$

which is manifestly independent of the polarization angle ψ and therefore the same in any polarization basis. Define a transverse traceless projector

$$P^{\mathrm{TT}\vec{k}ab}{}_{cd} = \frac{1}{2} \sum_{A=+,\times} e_A{}^{ab} e_{A\,cd} = \frac{1}{2} \sum_{A=+,\times} \varepsilon_A{}^{ab} \varepsilon_{A\,cd} \tag{2.15}$$

and show that

$$F_+^2 + F_\times^2 = 2 d_{ab} P^{\mathrm{TT}\vec{k}ab}{}_{cd} d^{cd} \tag{2.16}$$

The projector $P^{\mathrm{TT}\vec{k}ab}{}_{cd}$ picks out the symmetric traceless tensor components transverse to $\vec{k}$. For the case of an interferometer with perpendicular arms, whose detector tensor is given by (2.9), the detector tensor is already traceless and transverse to a vector normal to the plane of the detector. Use this to calculate the maximum value of $F_+^2 + F_\times^2$, which occurs for waves coming from directly overhead or underfoot.

2.4 Earth-fixed basis vectors

A convenient basis for describing ground-based GW detectors is one fixed to the Earth: the unit vector $\vec{e}_3^{\,*}$ points parallel to the Earth's axis, from the center of the Earth to the North Pole. The unit vector $\vec{e}_1^{\,*}$ points from the center of the Earth to the point on the equator with 0° latitude and longitude. This then makes the remaining unit vector $\vec{e}_2^{\,*} = \vec{e}_3^{\,*} \times \vec{e}_1^{\,*}$ point from the center of the Earth to the point on the equator with latitude 0° and longitude 90°E. The asterisk has nothing to do with a complex conjugate, but rather stresses that the basis vectors are co-rotating with the Earth.

2.5 Exercise: Detector Tensor

Consider a detector located at a latitude of 30°N and a longitude of 90°W, with one arm pointing due West and the other arm pointing due South. (This is approximately the configuration of the LIGO Livingston detector, except that the angles are not quite so nice.) Write the components of the following along the co-rotating basis vectors $\{\vec{e}^*_i\}$:

1. The unit vector $\vec{u}$ along the West arm

2. The unit vector $\vec{v}$ along the South arm

Resolve the detector tensor $\overleftrightarrow{d} = \frac{\vec{u}\otimes\vec{u}-\vec{v}\otimes\vec{v}}{2}$ along the unit dyads $\{\vec{e}^*_i \otimes \vec{e}^*_j\}$.

2.6 Equatorial basis vectors

There is a corresponding basis, inertial with respect to the fixed stars: the unit vector $\vec{e}_3$ points to the Celestial North Pole (which means that $\vec{e}_3 = \vec{e}^*_3$). The unit vector $\vec{e}_1$ points to the point with declination $\delta = 0°$ and right ascension $\alpha = 0\,\text{hr}$, i.e., the intersection of the ecliptic with the celestial equator known as the Vernal (March) Equinox. The third unit vector $\vec{e}_2 = \vec{e}_3 \times \vec{e}_1$ thus points to the point with $\delta = 0°$ and right ascension $\alpha = 6\,\text{hr}$.

The relationship between these bases is shown in figure 4. In particular, if we define the angle γ to correspond to the Sidereal Time at the Greenwich Meridian (which increases by 24 hours, i.e., $360° = 2\pi$, every sidereal day of approximately 23 hours and 56 minutes, so that $\gamma = \Omega_\oplus(t - t_{\text{Gmid}})$, where $\Omega_\oplus$ is the Earth's rotation frequency and t_{Gmid} corresponds to sidereal midnight at 0° longitude), then

$$\vec{e}^*_1 = \vec{e}_1 \cos\gamma + \vec{e}_2 \sin\gamma \tag{2.17a}$$
$$\vec{e}^*_2 = -\vec{e}_1 \sin\gamma + \vec{e}_2 \cos\gamma \tag{2.17b}$$

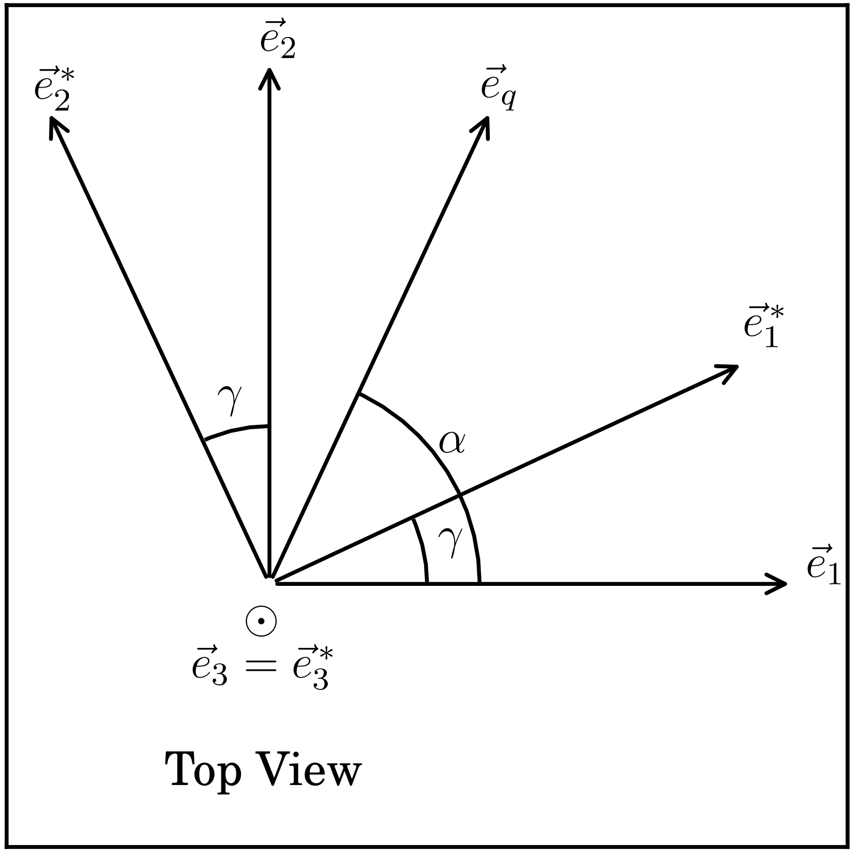

Figure 4: Relationship between the Earth-fixed and inertial bases, and illustration of right ascension. $\vec{e}_3 = \vec{e}^*_3$ points along the Earth's rotation axis towards the North Pole and the Celestial North Pole. $\vec{e}^*_1$ points to 0° latitude and longitude while $\vec{e}_1$ points to $\delta = 0°$ and right ascension $\alpha = 0\,\text{hr}$. As the Earth rotates, the starred unit vectors rotate relative the unstarred ones. $\vec{e}^*_1$ and $\vec{e}_1$ coincide when the sidereal time at Greenwich is midnight. We define the angle γ to be the Greenwich Sidereal Time (GST), which is the angle from $\vec{e}_1$ to $\vec{e}^*_1$, measured counterclockwise around $\vec{e}_3$. The unit vector $\vec{e}_q$ is the projection into the equatorial plane of the vector $\vec{k}$ from the observer to the source, as shown in figure 5.

2.7 Exercise: Basis associated with a sky position

Consider a potential source of gravitational waves with right ascension $\alpha = 0\,\text{hr}$ and declination $\delta = +60°$. Work out the components in the equatorial basis $\{\vec{e}_i\}$ of the propagation vector $\vec{k}$, as well as the vectors $\vec{\imath}$ and $\vec{\jmath}$ which point "West on the sky" and "North on the sky" at that point. Take the dot products $\{\vec{\imath}\cdot\vec{e}_i^{\,*}\}$ and $\{\vec{\jmath}\cdot\vec{e}_i^{\,*}\}$ to find the components of $\vec{\imath}$ and $\vec{\jmath}$ in the Earth-fixed basis, as a function of Greenwich sidereal time γ. These can be used to find the amplitude modulation coëfficients a and b for this source and e.g., the detector of exercise 2.5.

2.8 Sketch of General Calculation

2.8.1 Polarization basis from α, δ, and ψ

To get the antenna response functions for an arbitrary sky point, we need to do the following: calculate the components of $\vec{k}$ in some basis, given α and δ; do likewise for the perpendicular unit vectors $\vec{\imath}$ and $\vec{\jmath}$; find $\vec{\ell}$ and $\vec{m}$ for the given ψ, if desired. We can then construct $\overleftrightarrow{\varepsilon}_{+,\times}$ from $\vec{\imath}$ and $\vec{\jmath}$ or $\overleftrightarrow{e}_{+,\times}$ from $\vec{\ell}$ and $\vec{m}$, as needed.

The first step is to resolve the propagation direction $\vec{k}$ for a given right ascension α and declination δ. First, consider the plane containing $\vec{e}_3$ and $\vec{k}$, as shown in figure 5 Define $\vec{e}_q$ to be the unit vector pointing towards the point on the Celestial Equator with right ascension α, which also lies in the same plane. Since δ is the angle measured up from the Celestial equator to the sky position of the source, we can resolve

$$\vec{k} = -\vec{e}_q \cos\delta - \vec{e}_3 \sin\delta \tag{2.18}$$

To get the components of $\vec{e}_q$ in the equatorial basis, we look at the equatorial plane in figure 4. The right ascension is the angle

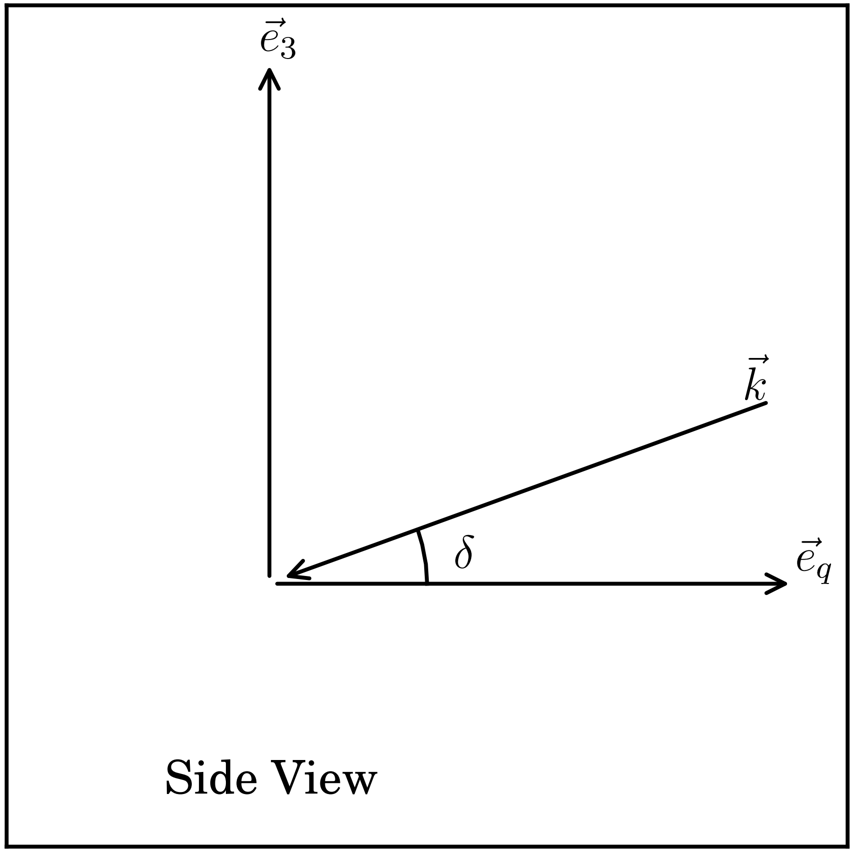

Figure 5: Illustration of declination, in the plane containing the propagation vector $\vec{k}$ and the unit vector $\vec{e}_3$ pointing along the Earth's rotation axis towards the Celestial North Pole. The unit vector $\vec{e}_q$ is the projection into the equatorial plane of the vector $\vec{k}$ from the observer to the source.

around from the Vernal Equinox ($\vec{e}_1$) to the sky position of the source, so

$$\vec{e}_q = \vec{e}_1 \cos\alpha + \vec{e}_2 \sin\alpha \tag{2.19}$$

and thus

$$\vec{k} = -\vec{e}_1 \cos\delta\cos\alpha - \vec{e}_2 \cos\delta\sin\alpha - \vec{e}_3 \sin\delta \tag{2.20}$$

The components along $\vec{e}_1$, $\vec{e}_2$, and $\vec{e}_3$ will be constant for a given source. To get the components along the starred unit vectors, we just need to note that the angle from $\vec{e}_1^*$ to $\vec{e}_q$ is $\alpha - \gamma$.[7] In terms of the starred basis,

$$\vec{e}_q^* = \vec{e}_1^* \cos(\alpha-\gamma) + \vec{e}_2^* \sin(\alpha-\gamma) \tag{2.21}$$

and thus

$$\vec{k} = -\vec{e}_1^* \cos\delta\cos(\alpha-\gamma) - \vec{e}_2^* \cos\delta\sin(\alpha-\gamma) - \vec{e}_3^* \sin\delta \tag{2.22}$$

The calculations of $\vec{\imath}$ and $\vec{\jmath}$ in either basis proceed along similar lines.

Note that to calculate $F_{+,\times} = \overleftrightarrow{d} : \overleftrightarrow{e}_{+,\times}$ it may actually be easier to work out

$$F_+ = \overleftrightarrow{d} : \overleftrightarrow{e}_+ = \vec{\ell}\cdot\overleftrightarrow{d}\cdot\vec{\ell} - \vec{m}\cdot\overleftrightarrow{d}\cdot\vec{m} \tag{2.23a}$$
$$F_\times = \overleftrightarrow{d} : \overleftrightarrow{e}_\times = \vec{\ell}\cdot\overleftrightarrow{d}\cdot\vec{m} + \vec{m}\cdot\overleftrightarrow{d}\cdot\vec{\ell} \tag{2.23b}$$

rather than constructing the matrix representations of $\overleftrightarrow{e}_{+,\times}$ in a particular basis.

[7]Note that in astronomy one defines a quantity called "hour angle" associated with a point on an observer's sky so that the hour angle increases with increasing sidereal time. Since γ is the Greenwich Sidereal Time, $\gamma - \alpha$ is the Greenwich Hour Angle of a source with right ascension α. The combination $\alpha - \gamma$ can thus be referred to as "Minus Greenwich Hour Angle".

2.8.2 Detector tensor from coördinates of detector

Given the coordinates of an Earth-bound detector (latitude, longitude and elevation) and some angles to identify the directions of its arms (usually and azimuth measured clockwise from local North and an altitude angle above or below the local horizontal plane), we can work out the components in the Earth-fixed basis $\{\vec{e}_i^*\}$ of the unit vectors $\vec{u}$ and $\vec{v}$ along its arms. (Note that we often neglect both the elevation in meters relative to some reference shape of the Earth–sphere or ellipsoid–as well as the altitude angles of the arms relative to the horizontal, since these are both small for practical detectors.) Since there are only a few detectors on the Earth, it's actually usually easiest just to work out the components of $\overleftrightarrow{d}$ in the Earth-fixed basis, once and for all.

The calculation can be done by working out the components of vectors and tensors in the Earth-fixed (starred) or the non-rotating (unstarred) inertial basis by going in two steps:

1. Express the unit vectors $\vec{u}$ and $\vec{v}$ in terms of a local basis $\{\vec{E}, \vec{N}, \vec{U}\}$ of unit vectors pointing East (along a parallel of latitude), North (along a meridian) and Up (normal to the local reference tangent plane, using the azimuth and possibly altitude angle of each arm.

2. Express the basis vectors $\{\vec{E}, \vec{N}, \vec{U}\}$, which correspond to a particular latitude and longitude, in terms of the starred or unstarred reference basis.

3 Preferred polarization basis

3.1 The Quadrupole Formula

Gravitational waves from a particular direction can be resolved in different polarization bases, but some make the calculations more convenient than others. For a stochastic background, we're looking at a superposition of different sources, so the result is an unpolarized signal which can be described equally well in any basis. For unmodelled bursts, there's nothing special about the source, but the detector network can pick out a preferred basis for some search methods. For modelled signals such as nearly periodic continuous wave signals, or compact binary inspirals, the geometry of the system provides us with a preferred polarization basis in which the signal description is simple.

Most of the gravitational waves seen by a distant observer, from a typical system, are in the form of quadrupole radiation. The metric perturbation is given by the quadrupole formula as[8]

$$h_{ab} = \frac{2G}{c^4 d} P^{\mathrm{TT}\vec{k}\,cd}_{\ ab} \ddot{I\!\!\!-}_{cd}(t - d/c) \tag{3.1}$$

where $\vec{k}$ is the direction from the source to the observer and d is the distance. We can accomplish the projection onto transverse traceless states by writing $\overleftrightarrow{h}$ in terms of its plus and cross components as usual:

$$\overleftrightarrow{h} = h_+ \overleftrightarrow{e}_+ + h_\times \overleftrightarrow{e}_\times \tag{3.2}$$

where (3.1) tells us that

$$h_A = \frac{2G}{c^4 d} \frac{\overleftrightarrow{e}_A}{2} : \frac{d^2}{dt^2} \overleftrightarrow{I\!\!\!-}(t - d/c) \qquad A = +, \times \tag{3.3}$$

The $t - d/c$ indicates that the observer is seeing the source as it was at a time in the past, when the gravitational waves now reaching the observer were emitted. Here $\overleftrightarrow{I\!\!\!-}$ is the reduced quadrupole moment defined in MTW equation (36.3):

$$\overleftrightarrow{I\!\!\!-} = \iiint \rho \left(\vec{r} \otimes \vec{r} - \overleftrightarrow{1} \frac{r^2}{3} \right) d^3V \tag{3.4}$$

[8]See for example equation (36.20) of Misner, Thorne and Wheeler, *Gravitation* (1973)

If we recall the standard definition of the moment of inertia tensor $\overleftrightarrow{I}$ from mechanics

$$\overleftrightarrow{I} = \iiint \rho \left(\overleftrightarrow{1} r^2 - \vec{r} \otimes \vec{r} \right) d^3V \tag{3.5}$$

we see that the two are related by

$$I\!\!\!-_{ab} = -P^{\mathrm{T}\,cd}_{\ ab} I_{cd} \tag{3.6}$$

(note the minus sign!), where $P^{\mathrm{T}\,cd}_{\ ab}$ is the projection operator onto traceless symmetric tensors:

$$P^{\mathrm{T}\,cd}_{\ ab} = \frac{1}{2} \left(\delta^c_a \delta^d_b + \delta^d_a \delta^c_b \right) - \frac{1}{3} \delta_{ab} \delta^{cd} \tag{3.7}$$

3.2 Geometry of a non-precessing quasiperiodic source

Consider a mass distribution which is rigidly rotating with constant angular velocity about one of its principal axes of inertia. This could be the nearly-periodic signal given off by a rotating neutron star, or the signal from a binary system where the inspiral is occurring slowly. We can expand the inertia tensor about its principal axes like this:

$$\overleftrightarrow{I} = I_1 \vec{u}_1 \otimes \vec{u}_1 + I_2 \vec{u}_2 \otimes \vec{u}_2 + I_3 \vec{u}_3 \otimes \vec{u}_3 \tag{3.8}$$

In this approximation, I_1, I_2, and I_3 are all constant; if it's rotating about $\vec{u}_3$ with angular speed Ω, then the principal axes can be written with respect to some non-rotating axes $\{\vec{u}^0_i\}$ as

$$\begin{aligned}
\vec{u}_1 &= \vec{u}^0_1 \cos\Omega(t - t_0) + \vec{u}^0_2 \sin\Omega(t - t_0) && \text{(3.9a)} \\
\vec{u}_2 &= -\vec{u}^0_1 \sin\Omega(t - t_0) + \vec{u}^0_2 \cos\Omega(t - t_0) && \text{(3.9b)} \\
\vec{u}_3 &= \vec{u}_3 && \text{(3.9c)}
\end{aligned}$$

It's not hard to work out the time derivatives of the basis vectors along the principal axes due to the rotation:

$$\frac{d\vec{u}_1}{dt} = \Omega\vec{u}_3 \times \vec{u}_1 = \Omega\vec{u}_2 \tag{3.10a}$$
$$\frac{d\vec{u}_2}{dt} = \Omega\vec{u}_3 \times \vec{u}_2 = -\Omega\vec{u}_1 \tag{3.10b}$$
$$\frac{d\vec{u}_3}{dt} = \Omega\vec{u}_3 \times \vec{u}_3 = \vec{0} \tag{3.10c}$$

This means that

$$\begin{aligned}\frac{d}{dt}\overleftrightarrow{I} &= \Omega I_1(\vec{u}_1 \otimes \vec{u}_2 + \vec{u}_2 \otimes \vec{u}_1) - \Omega I_2(\vec{u}_2 \otimes \vec{u}_1 + \vec{u}_1 \otimes \vec{u}_2) \\ &= \Omega(I_1 - I_2)(\vec{u}_1 \otimes \vec{u}_2 + \vec{u}_2 \otimes \vec{u}_1)\end{aligned} \tag{3.11}$$

and

$$\frac{d^2}{dt^2}\overleftrightarrow{I} = -2\Omega^2(I_1 - I_2)(\vec{u}_1 \otimes \vec{u}_1 - \vec{u}_2 \otimes \vec{u}_2) \tag{3.12}$$

Since this is already traceless, (3.6) tells us that

$$\frac{d^2}{dt^2}\overleftrightarrow{\not{I}} = 2\Omega^2(I_1 - I_2)(\vec{u}_1 \otimes \vec{u}_1 - \vec{u}_2 \otimes \vec{u}_2) \tag{3.13}$$

To get the explicit time dependence of (3.13), we could substitute the explicit time-dependent forms of $\vec{u}_1$ and $\vec{u}_2$ into (3.13), but it's easier to note that the combination $\vec{u}_1 \otimes \vec{u}_1 - \vec{u}_2 \otimes \vec{u}_2$ appearing in (3.13) is a traceless tensor transverse to $\vec{u}_3$ and so if we define basis tensors[9]

$$\overleftrightarrow{E}_+ = \vec{u}^0_1 \otimes \vec{u}^0_1 - \vec{u}^0_2 \otimes \vec{u}^0_2 \tag{3.14a}$$
$$\overleftrightarrow{E}_\times = \vec{u}^0_1 \otimes \vec{u}^0_2 + \vec{u}^0_2 \otimes \vec{u}^0_1 \tag{3.14b}$$

[9]Note that these are traceless tensors transverse to $\vec{u}_3$ and *not* to the propagation direction $\vec{k}$.

then, by analogy to the polarization rotation in section 1.3, we have

$$\vec{u}_1 \otimes \vec{u}_1 - \vec{u}_2 \otimes \vec{u}_2 = \overleftrightarrow{E}_+ \cos 2\Omega(t - t_0) + \overleftrightarrow{E}_\times \sin 2\Omega(t - t_0) \tag{3.15}$$

and

$$\frac{d^2}{dt^2}\overleftrightarrow{\not{I}} = 2\Omega^2(I_1 - I_2)\left(\overleftrightarrow{E}_+ \cos[\Phi(t) + \phi_0] + \overleftrightarrow{E}_\times \sin[\Phi(t) + \phi_0]\right) \tag{3.16}$$

where

$$\Phi(t) = 2\Omega(t - d/c) \tag{3.17}$$

and

$$\phi_0 = -2\Omega t_0 \tag{3.18}$$

This means that the polarization components of the gravitational wave travelling in a particular direction are

$$h_A = \frac{4G\Omega^2(I_1 - I_2)}{c^4 d}\left[\frac{\overleftrightarrow{e}_A : \overleftrightarrow{E}_+}{2}\cos[\Phi(t) + \phi_0] + \frac{\overleftrightarrow{e}_A : \overleftrightarrow{E}_\times}{2}\sin[\Phi(t) + \phi_0]\right] \tag{3.19}$$

So far, we haven't specified the non-rotating basis vectors $\vec{u}^0_1$ and $\vec{u}^0_2$, perpendicular to $\vec{u}_3 = \vec{u}^0_3$ (which, incidentally, determine ϕ_0), nor the basis vectors $\vec{\ell}$ and $\vec{m}$, perpendicular to $\vec{k}$, which define the polarization basis. We can do this by picking $\vec{\ell} = \vec{u}^0_1$ along the line of nodes, which is perpendicular to both the propagation direction $\vec{k}$ and the system angular momentum direction $\vec{u}_3$. If ι is the inclination angle between the angular momentum direction $\vec{u}$ and the propagation vector $\vec{k}$, as shown in figure 6, the dot products between the basis vectors defining $\{\overleftrightarrow{e}_A\}$ and $\{\overleftrightarrow{E}_A\}$ are

$$\vec{\ell} \cdot \vec{u}^0_1 = 1 \qquad \vec{m} \cdot \vec{u}^0_1 = 0 \qquad \vec{\ell} \cdot \vec{u}^0_2 = 0 \qquad \vec{m} \cdot \vec{u}^0_2 = \cos\iota \tag{3.20}$$

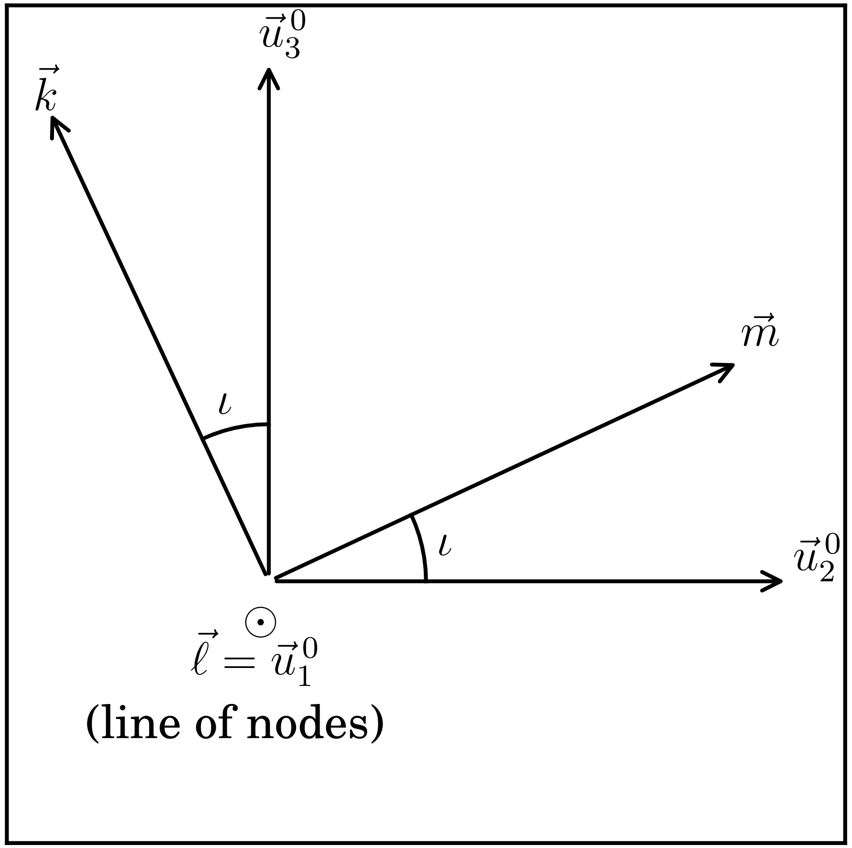

Figure 6: Illustration of bases associated with a rotating gravitational-wave source and its propagation. The unit vector $\vec{k}$ points from the source to the observer, and $\vec{u}_3^0$ points along the axis of rotation; the angle between these is the inclination ι. The preferred polarization basis (see figure 2) is constructed by choosing $\vec{\ell}$ to be along the line of nodes, perpendicular to both $\vec{k}$ and $\vec{u}_3^0$.

3.3 Exercise: Projection from source to propagation basis

Use the inner products (3.20) to show that

$$\overleftrightarrow{e}_+ : \overleftrightarrow{E}_+ = 1 + \cos^2\iota \qquad \text{and} \qquad \overleftrightarrow{e}_+ : \overleftrightarrow{E}_\times = 0 \tag{3.21a}$$

$$\overleftrightarrow{e}_\times : \overleftrightarrow{E}_+ = 0 \qquad \text{and} \qquad \overleftrightarrow{e}_\times : \overleftrightarrow{E}_\times = 2\cos\iota \tag{3.21b}$$

3.4 Waveform in preferred basis

This then means that, in the preferred basis,

$$h_+ = h_0 \frac{1+\cos^2\iota}{2}\cos[\Phi(t)+\phi_0] \tag{3.22a}$$

$$h_\times = h_0 \cos\iota \, \sin[\Phi(t)+\phi_0] \tag{3.22b}$$

where the GW amplitude is

$$h_0 = \frac{4G\Omega^2(I_1 - I_2)}{c^4 d} \tag{3.23}$$

Placing the basis vector $\vec{\ell}$ along the line of nodes means that this preferred polarization basis has ψ as the angle from "West on the sky" to the line of nodes, i.e., if we're talking about objects moving in circular orbits, this is the angle from "West on the sky" to the long axis of the projected orbit. The nice feature of the preferred basis is that h_+ and $h_\times$ are a quarter-cycle out of phase, as illustrated in `http://ccrg.rit.edu/~whelan/gwmovie/`

3.5 Exercise: slow inspiral

The waveform (3.22) also applies to the early stages of a binary inspiral. Consider two objects of mass m_1 and m_2, total mass M, and reduced mass $m_1 m_2/M$. If the trajectories are

$$\vec{r}_1(t) = \frac{m_2}{M}\vec{r}(t) \tag{3.24a}$$

$$\vec{r}_2(t) = -\frac{m_1}{M}\vec{r}(t) \tag{3.24b}$$

where

$$\vec{r}(t) = r(t)\cos\phi(t)\vec{u}_1^{\,0} + r(t)\sin\phi(t)\vec{u}_2^{\,0} \tag{3.25}$$

show that

$$\frac{d^2}{dt^2}\overleftrightarrow{I} \approx -\mu r^2\dot{\phi}^2\left(\vec{u}_1 \otimes \vec{u}_1 - \vec{u}_2 \otimes \vec{u}_2\right) \tag{3.26}$$

What conditions are needed on $\dot{r}$, $\ddot{r}$, and $\ddot{\phi}$ in order to make this approximation valid?

3.6 "Effective distance" for inspiral signals

If we compare the form (3.26) to (3.13), we see that we've just replaced $2\Omega^2(I_1 - I_2)$ with $-\mu r^2\dot{\phi}^2$ and therefore by comparison to (3.22) we can write down

$$h_+ \approx -\frac{A(t)}{d}\frac{1+\cos^2\iota}{2}\cos\Phi(t) \tag{3.27a}$$

$$h_\times \approx -\frac{A(t)}{d}\cos\iota\,\sin\Phi(t) \tag{3.27b}$$

where we've collected together the part of the amplitude

$$A(t) = \left(\frac{4G\mu[r(t-d/c)\dot{\phi}(t-d/c)]^2}{c^4}\right) \tag{3.28}$$

which depends only on properties of the source like masses and trajectory.[10] If we think about the signal generated in a detector with antenna pattern factors F_+ and $F_\times$, we get

$$\begin{aligned} h(t) &= \frac{A(t)}{d}\left[-F_+\frac{1+\cos^2\iota}{2}\cos\Phi(t) - F_\times\cos\iota\sin\Phi(t)\right] \\ &= \frac{A(t)}{d}\left(\sqrt{F_+^2\frac{(1+\cos^2\iota)^2}{4} + F_\times^2\cos^2\iota}\right)\cos[\Phi(t)-\Psi] \end{aligned} \tag{3.29}$$

where we have used the usual trick of rewriting

$$\alpha\cos\varphi + \beta\sin\varphi = \gamma\cos(\varphi - \psi) \tag{3.30}$$

where

$$\alpha = \gamma\cos\psi \qquad \text{and} \qquad \beta = \gamma\sin\psi \tag{3.31}$$

so

$$\gamma = \sqrt{\alpha^2 + \beta^2} \tag{3.32}$$

So we see that the overall amplitude is determined by the distance to the source, but in this slow-inspiral approximation, that distance appears together with the observing geometry in a combination known as *effective distance*:

$$d_{\text{eff}} = \frac{d}{\sqrt{F_+^2\frac{(1+\cos^2\iota)^2}{4} + F_\times^2\cos^2\iota}} \tag{3.33}$$

Note that the factor in the square root is a maximum (and therefore the effective distance corresponding to a given physical distance) when $|\cos\iota| = 1$, i.e., we are seeing the binary face on ($\iota = 0$ or $\iota = \pi$). Using the result of section 2.3, that $F_+^2 + F_\times^2 \leq 1$, we have

$$\left(\frac{d}{d_{\text{eff}}}\right)^2 = F_+^2\frac{(1+\cos^2\iota)^2}{4} + F_\times^2\cos^2\iota \leq F_+^2 + F_\times^2 \leq 1 \tag{3.34}$$

which means that $d_{\text{eff}} \geq d$. The effective distance equals the physical distance if the binary is seen face-on, and is at the detector's zenith or nadir.

[10] There is still some dependence on distance to the source, of course, but only in terms of when the signal arrives, not in the overall magnitude.

Probability, Statistics, Fourier Analysis and Signal Processing

John T. Whelan

Lecture given at IUCAA, Pune, 2014 January 17

Contents

1 Probability and Statistical Inference

1.1 Logic and Probability

There are numerous interpretations of probability, but one which applies well to observational science is that of an extended logic. Let A be a proposition which could be either true or false, e.g., "The orbital period of Mars is between 686 and 687 days," "John Whelan is an Indian citizen as of Jan 1, 2020," or "My detector will collect 427 photons in the next two hours." We may know, given the information at hand, that A is definitely true or definitely false, or we may be uncertain about the answer, either because our knowledge of the situation is incomplete, or because it refers to the outcome of an experiment with a random element, which has not occurred yet. The probability of the proposition A (which we also call an "event") is a number between 0 and 1 which quantifies our degree of certainty, given the information at hand. We write this as $P(A|I)$, where I represents some state of knowledge, to emphasize that the probability we assign always depends on the information we have, the assumption that a model is correct, etc. If A is definitely true, in the context of I, then $P(A|I) = 1$. If it's definitely false, $P(A|I) = 0$.

If A represents the outcome of an experiment which we could somehow arrange to repeat under identical circumstances, then $P(A|I)$ will be approximately equal to the long-term frequency of the event A. I.e., if we do some large number N of repetitions of the experiment, at the beginning of which we recreate the situation described by I, the approximate number of experiments in which A will turn out to be true is $N \times P(A|I)$. In the classical or "frequentist" approach to statistics, this is the only sort of event to which we're allowed to assign a probability, but in the more general "Bayesian" framework we are free to assign probabilities to any logical proposition.

Several basic operations can be used to combine logical propositions:

- Negation. $\overline{A}$ is true if A is false, and vice-versa. In words, we can think of $\overline{A}$ as "not A". (Other notations include A' and $\neg A$.)

- Intersection. A, B is true if A and B are both true. In words, this is "A and B". (Other notations include $A \cap B$ and $A \wedge B$.) The advantage of the comma is that $P(A, B|I)$ is the probability that both A and B are true, given I.

- Union. $A + B$ is true if either A or B (or both) is true. In words, this is "A or B". (Other notations include $A \cup B$ and $A \vee B$.) Note the unfortunate aspect of this notation that $+$ is to be read as "or" rather than "and".

There are basic rules of probability corresponding to these logical operations:

- $P(A|I) + P(\overline{A}|I) = 1$

- The product rule: $P(A, B|I) = P(A|B, I)P(B|I)$

- The sum rule: if A and B are mutually exclusive, i.e., if $P(A, B|I) = 0$, then $P(A + B|I) = P(A|I) + P(B|I)$.

Note that in this approach, where all probabilities are conditional, the product rule is really what's fundamental. Classical approaches to probability instead define the conditional probability as $P(A|B) = \frac{P(A,B)}{P(B)}$, and therefore only entertain consideration of the conditional probability $P(A|B)$ if B is not only something to which they're allowed assign a probability, but for which that probability is nonzero.

Because the logical "and" and "or" operations are symmetrical, i.e., A, B is equivalent to B, A and $A + B$ is equivalent to $B + A$, we can write the product rule in two different ways:

$$P(A, B|I) = P(A|B, I)P(B|I) = P(B|A, I)P(A|I) \tag{1.1}$$

this can be rearranged into *Bayes's Theorem*, which says that

$$P(A|B, I) = \frac{P(B|A, I)P(A|I)}{P(B|I)} \tag{1.2}$$

which is incredibly useful when you naturally know $P(B|A, I)$ but would like to know $P(A|B, I)$. For instance, suppose A refers to "I have terrible-disease-of-the-year (TDY)", B refers to "I test positive for TDY", and I represents the information that I had no extra risk factors or symptoms for TDY but was routinely tested, 0.1% of people in such a group have TDY, the test has a 2% false positive rate (2% of people without TDY will test positive for it) and a 1% false negative rate (1% of people with TDY will test negative for it). This information tells us that:

- $P(A|I) = 0.001$ so $P(\overline{A}|I) = 0.999$.

- $P(\overline{B}|A, I) = 0.01$ so $P(B|A, I) = 0.99$.

- $P(B|\overline{A}, I) = 0.02$ so $P(\overline{B}|\overline{A}, I) = 0.98$.

Additionally, since $B = B, A + B, \overline{A}$,

$$\begin{aligned} P(B|I) &= P(B, A|I) + P(B, \overline{A}|I) \\ &= P(B|A, I)P(A|I) + P(B|\overline{A}, I)P(\overline{A}|I) \\ &= 0.99 \times 0.001 + 0.02 \times 0.999 = 0.00099 + 0.01998 \\ &= 0.02097 \end{aligned} \tag{1.3}$$

We can then use Bayes's theorem to show that

$$P(A|B,I) = \frac{0.00099}{0.02097} \approx 0.04721 \tag{1.4}$$

I.e., if I test positive for TDY, I have about a 4.7% chance of actually having the disease. This is a lot less than $P(B|A,I)$, which is 99%!

In the context of observational science, Bayes's theorem is most commonly applied to a situation where H is a hypothesis which I'd like to evaluate and D is a particular set of data I've collected. It's usually straightforward to work out $P(D|H,I)$, the probability of observing a particular set of data values given a model, but I generally want to answer the question, what is my degree of belief in the hypothesis H after the observation. The answer, according to Bayes's Theorem, is

$$P(H|D,I) = \frac{P(D|H,I)P(H|I)}{P(D|I)} \tag{1.5}$$

1.2 Probability Distributions

In what follows, we will often suppress the explicit mention of the background information I on which all of our probabilities are conditional. The logical propositions to which we often assign probabilities involve the values of some random or otherwise unknown quantities. So for example, $N_{\text{counts}} = 37$ or $70\,\text{km/s/Mpc} < H_0 < 75\,\text{km/s/Mpc}$. Sometimes the notation gets a bit confused between a quantity and its value, and you'll see things like X for a "random variable" and x for a value it can take on. You'd like to be able to specify the probability that $X = x$, as a function of x. In practice, this is slightly complicated by whether we think of X as taking on only discrete values, or if it can take on any value in a continuous range.

If X is discrete, we can talk about its *probability mass function* $p_X(x) = P(X = x)$. This is often just written $p(x)$ or $P(x)$. For instance, if X is the number of events in a particular interval from a stationary process in which the events are independent of one another, and the average number of events expected in the interval given the long-term event rate is μ it is described by the Poisson distribution

$$p(x) = P(X = x) = \begin{cases} \frac{\mu^x}{x!}\, e^{-\mu} & x = 0, 1, 2, \ldots \\ 0 & \text{otherwise} \end{cases} \tag{1.6}$$

However, it often happens that X is continuous, so that it is vanishingly unlikely that it takes on one specific value. For instance, the height of a randomly chosen person will not be exactly 175 cm. If you measure it to more significant figures, it will turn out to be 175.25 cm or 175.24732 cm etc. So instead we want to talk about the probability for X to be in a small interval, which we call the *probability density function*

$$f(x) = \lim_{dx \to 0} \frac{P(x - \frac{dx}{2} < X < x + \frac{dx}{2})}{dx} \tag{1.7}$$

so that

$$P(a < X < b) = \int_a^b f(x)\, dx \tag{1.8}$$

The pdf might be called pdf(x) or even $P(x)$. A useful notation for the pdf is $\frac{dP}{dx}$, which tends to make the impact of changes of variables more obvious. In the end, it's a bit hopeless to try to stick to one letter, since you might want to talk about the joint probability distribution associated with some discrete and some continuous random variables. To give a concrete example, a common probability distribution is the Gaussian distribution with parameters μ and σ, which has pdf

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-(x-\mu)^2/2\sigma^2}\ , \qquad -\infty < x < \infty \tag{1.9}$$

In either case, you can define an operation known as the *expectation value*

$$E\left[g(X)\right] = \begin{cases} \sum_x g(x)\, p(x) & X \text{ discrete} \\ \int_{-\infty}^{\infty} g(x)\, f(x)\, dx & X \text{ continuous} \end{cases} \tag{1.10}$$

with the mean $\mu_X = E\left[X\right]$ as a special case, and also the variance

$$E\left[(X-\mu_X)^2\right] = E\left[X^2\right] - \mu_X^2 \tag{1.11}$$

To have a sensible probability distribution, we should satisfy a normalization condition $\sum_x p(x) = 1$ or $\int_{-\infty}^{\infty} f(x)\, dx = 1$.

1.3 Some Basic Statistical Inference

1.3.1 Bayesian Methods

Broadly speaking, the kinds of questions we'd like to answer using observational data are:[1]

- Hypothesis testing and model selection: given some observed data $\mathbf{x}$, what can we say about the relative plausibility of models $\mathcal{H}_1$ and $\mathcal{H}_2$?

- Parameter estimation: if our model $\mathcal{H}$ depends on some parameters $\boldsymbol{\theta}$, how do the data affect our judgments about the plausible values of $\boldsymbol{\theta}$?

In principle, parameter estimation is a special case of hypothesis testing, since we could consider different hypotheses corresponding to different parameter values, but in practice the notation is slightly different.

In any case, the easiest thing to construct is the probability distribution for the data given the model and any parameters: $P(\mathbf{x}|\mathcal{H},\boldsymbol{\theta},I)$. This is often used to compare models or parameter values, and as such is considered a function of $\mathcal{H}$ and/or $\boldsymbol{\theta}$ and called the *likelihood function*. It can be related to probabilities for $\mathcal{H}$ and/or $\boldsymbol{\theta}$ using Bayes's theorem.

Putting aside the question of parameters, a comparison between models $\mathcal{H}_1$ and $\mathcal{H}_2$ would be to compare $P(\mathcal{H}_1|\mathbf{x},I)$ to $P(\mathcal{H}_2|\mathbf{x},I)$, where Bayes's theorem tells us

$$P(\mathcal{H}|\mathbf{x},I) = \frac{P(\mathbf{x}|\mathcal{H},I)P(\mathcal{H}|I)}{P(\mathbf{x}|I)} \tag{1.12}$$

if we take the ratio of these two probabilities, we get

$$\frac{P(\mathcal{H}_1|\mathbf{x},I)}{P(\mathcal{H}_2|\mathbf{x},I)} = \frac{P(\mathbf{x}|\mathcal{H}_1,I)}{P(\mathbf{x}|\mathcal{H}_2,I)}\frac{P(\mathcal{H}_1|I)}{P(\mathcal{H}_2|I)} \tag{1.13}$$

To get the actual ratio, we'd need to know the ratio $\frac{P(\mathcal{H}_1|I)}{P(\mathcal{H}_2|I)}$ of probabilities that we'd assign in the absence of the data, but it's more unambiguous just to quote the factor by which the ratio changed, which is the *Bayes factor*

$$\frac{P(\mathbf{x}|\mathcal{H}_1,I)}{P(\mathbf{x}|\mathcal{H}_2,I)} \tag{1.14}$$

which is just the likelihood ratio.

If we want to assume a particular hypothesis and make a statement about the parameters in light of the observed data, we again use Bayes's theorem to construct

$$P(\boldsymbol{\theta}|\mathbf{x},\mathcal{H}) = \frac{P(\mathbf{x}|\boldsymbol{\theta},\mathcal{H})\, P(\boldsymbol{\theta}|\mathcal{H})}{P(\mathbf{x}|\mathcal{H})} \tag{1.15}$$

It is conventional to call $P(\boldsymbol{\theta}|\mathbf{x},\mathcal{H})$ the *posterior probability distribution* on $\boldsymbol{\theta}$ and $P(\boldsymbol{\theta}|\mathcal{H})$ the *prior probability distribution*. This is in some sense an artificial distinction, but in reflects the fact that the latter is adjusted in light of the data to give the former.

[1] We write both $\mathbf{x}$ and $\boldsymbol{\theta}$ as vectors to emphasize the fact that there will in general be multiple data points and multiple parameters.

Note, also, that for a hypothesis $\mathcal{H}$ concerning a model involving parameters $\boldsymbol{\theta}$ to be complete, it must also specify a prior probability distribution $P(\boldsymbol{\theta}|\mathcal{H})$ for the parameters themselves. Finally, note that if we have the numerator of (1.15) as a function of $\boldsymbol{\theta}$, we can automatically calculate the denominator by a process called marginalization, which is basically a version of the sum rule:

$$P(\mathbf{x}|\mathcal{H}) = \int P(\mathbf{x},\boldsymbol{\theta}|\mathcal{H})\,d\boldsymbol{\theta} = \int P(\mathbf{x}|\boldsymbol{\theta},\mathcal{H})\,P(\boldsymbol{\theta}|\mathcal{H})\,d\boldsymbol{\theta} \tag{1.16}$$

1.3.2 Frequentist and Pseudo-frequentist Methods

In the frequentist formalism, you can't define things like $P(\mathcal{H}|\mathbf{x})$; instead you have to construct probabilistic statements about the random observed data $\mathbf{X}$, and then evaluate them in light of the actual observation $\mathbf{X} = \mathbf{x}$. It's usually necessary to distill the data into a single number known as a *statistic* $y(\mathbf{X})$ (or perhaps a few statistics). For example, if you want to choose between two hypotheses $\mathcal{H}_1$ and $\mathcal{H}_0$, you construct some statistic $y(\mathbf{X})$, and if it's above some threshold value y_c, you prefer $\mathcal{H}_1$, while if it's below, you prefer $\mathcal{H}_0$. Due to the randomness in the experiment, this test of the validity of $\mathcal{H}_1$ will not be perfect, i.e., there will be some chance you picked the wrong hypothesis. This is expressed by

$$\text{false alarm probability} = P(y(\mathbf{X}) > y_c|\mathcal{H}_0) \tag{1.17a}$$

$$\text{false dismissal probability} = P(y(\mathbf{X}) < y_c|\mathcal{H}_1) \tag{1.17b}$$

If you increase the threshold, you will decrease the false alarm probability, but increase the false dismissal probability (decrease the efficiency of the test). Of course, many statistics are possible, but you'd prefer to have one which minimizes the false dismissal probability for each value of the false dismissal probability. There is a result called the NeymanPearson lemma[2] which shows that this "most powerful test" is acheived by using as your detection statistic the likelihood ratio:

$$y(\mathbf{x}) = \frac{P(\mathbf{x}|\mathcal{H}_1)}{P(\mathbf{x}|\mathcal{H}_0)} \tag{1.18}$$

It may happen, though, that you have other reasons for wanting to use a sub-optimal detection statistic $y(\mathbf{x})$. Perhaps the likelihood ratio is too expensive or difficult to compute, or perhaps $\mathcal{H}_1$ is a *composite hypothesis* with some unknown parameters $\boldsymbol{\theta}$, and so all you have access to is $P(\mathbf{x}|\mathcal{H}_1,\boldsymbol{\theta})$. (Recall that in the frequentist approach you can't even define something like $P(\boldsymbol{\theta}|\mathcal{H}_1)$, so you can't marginalize over $\boldsymbol{\theta}$ to get $P(\mathbf{x}|\mathcal{H}_1)$.) Then you can go ahead and apply the sub-optimal frequentist test.

Suppose, though, that after calculating your statistic $y(\mathbf{x})$, you decide you want to interpret things in a Bayesian way after all. (This can happen in parameter estimation especially. There may be times when *no* physically possible set of parameters produces a high enough detection statistic, and you don't really want to do something like set a negative upper limit on an event rate or energy density.) You can go ahead and do Bayesian inference using the information available, which is now just $y(\mathbf{x})$, and construct something like $P(\mathcal{H}|y(\mathbf{x}),I)$ or $P(\boldsymbol{\theta}|y(\mathbf{x}),\mathcal{H},I)$. I like to think of it as Bayesian analysis of an experiment, where the experimental data are the output of a frequentist experiment. Of course if $y(\mathbf{x})$ was chosen well, it may be that $P(\mathcal{H}|y(\mathbf{x}),I) = P(\mathcal{H}|\mathbf{x},I)$ and you haven't lost any information by calclulating $y(\mathbf{x})$ and discarding the rest of your data $\mathbf{x}$.

[2] J. Neyman and E. S. Pearson, *Philosophical Transactions of the Royal Society A* **231**, 694 (1933)

1.4 Exercise: estimation with known Gaussian errors

Suppose we are making a series of n measurements $\{x_i\}$ of some unknown quantity θ, each of which has a known Gaussian error of standard deviation σ_i associated with it. This means that the pdf for $\mathbf{x}$ is

$$P(\mathbf{x}|\theta,\mathcal{H},I) = \prod_{i=1}^{n} \frac{1}{\sigma_i\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left[\frac{x_i-\theta}{\sigma_i}\right]^2\right) \tag{1.19}$$

where the hypothesis $\mathcal{H}$ includes the values of the σ_i. Show that

$$\chi^2(\mathbf{x},\theta) = \sum_{i=1}^{n}\left(\frac{x_i-\theta}{\sigma_i}\right)^2 = \left(\frac{\theta-\theta_0(\mathbf{x})}{\sigma_\theta}\right)^2 + \chi_0^2(\mathbf{x}) \tag{1.20}$$

where

$$\sigma_\theta^{-2} = \sum_{i=1}^{n}\sigma_i^{-2} \tag{1.21a}$$

$$\theta_0(\mathbf{x}) = \sigma_\theta^2\sum_{i=1}^{n}\sigma_i^{-2}x_i \tag{1.21b}$$

$$\chi_0^2(\mathbf{x}) = \sum_{i=1}^{n}\sigma_i^{-2}x_i^2 - \sigma_\theta^{-2}\theta_0(\mathbf{x})^2 \tag{1.21c}$$

This means that

$$P(\mathbf{x}|\theta,\mathcal{H},I) = \frac{1}{\prod_{i=1}^{n}\sigma_i\sqrt{2\pi}} \exp\left(-\frac{[\theta-\theta_0(\mathbf{x})]^2}{2\sigma_\theta^2} + \frac{\chi_0^2(\mathbf{x})}{2}\right) \tag{1.22}$$

Show that we can construct the posterior $P(\theta|\mathbf{x},\mathcal{H},I)$ corresponding to a specified $P(\theta|\mathcal{H},I)$ using only the statistic $\theta_0(\mathbf{x})$, and that we can construct the "evidence" $P(\mathcal{H}|\mathbf{x},I)$ corresponding to a specified prior probability $P(\mathcal{H}|I)$ (suitable for the Bayes factor construction) using only the statistic $\chi_0^2(\mathbf{x})$.

1.5 Further Reading

- Jaynes, E. T., *Probability Theory: The Logic of Science* (Cambridge, 2003)
- Sivia, D. S., *Data Analysis: A Bayesian Tutorial*, 2nd edition (Oxford, 2006)
- Gregory, P., *Bayesian Logical Data Analysis for the Physical Sciences* (Cambridge, 2005)

2 Fourier Analysis

2.1 Continuous Fourier Transforms

You're probably familiar with the continuous Fourier transform

$$\widetilde{x}(f) = \int_{-\infty}^{\infty} dt\, x(t)\, e^{-i2\pi f(t-t_0)} \tag{2.1}$$

and its inverse

$$x(t) = \int_{-\infty}^{\infty} df\, \widetilde{x}(f)\, e^{i2\pi f(t-t_0)} \tag{2.2}$$

Notes:

- Lots of conventions, but note using f instead of ω gets rid of annoying 2π normalizations.
- If $x(t)$ is really a function of time, the origin/epoch t_0 is arbitrary and has no physical meaning. If it's a function of time *difference*, then $t_0 = 0$ makes sense.

The identity

$$\int_{-\infty}^{\infty} df\, e^{i2\pi f(t-t')} = \delta(t-t') \tag{2.3}$$

is useful for proving properties of continuous Fourier transforms.

2.2 Discrete Fourier Transforms

Real data is neither continuous nor infinite in duration. Consider discretely-sampled time series data of duration $T = N\delta t$:

$$x_j = x(t_j) = x(t + j\delta t) \qquad j = 0, 1, \ldots, N-1 \tag{2.4}$$

Its discrete Fourier transform is

$$\widehat{x}_k = \sum_{j=0}^{N-1} x_j \, e^{-i2\pi f_k (t_j - t_0)} = \sum_{j=0}^{N-1} x_j \, e^{-i2\pi jk/N} \tag{2.5}$$

where $f_k = k\delta f$, and

$$\delta f \, \delta t = \frac{\delta t}{T} = \frac{1}{N} \; . \tag{2.6}$$

We can define $\widehat{x}_k$ for any integer k, but there are only N independent values, thanks to the identifications

$$\widehat{x}_{N+k} = \widehat{x}_k \qquad \text{always} \tag{2.7a}$$

$$\widehat{x}_{-k} = \widehat{x}_k^* \qquad \text{if } \{x_j\} \text{ real} \tag{2.7b}$$

This means, for a real time series $\{x_j\}$, the N real numbers in the Fourier domain are (assuming N even)

- 1 real value x_0
- $\frac{N}{2} - 2$ complex values $\{x_k | k = 1, \ldots \frac{N}{2} - 1\}$
- 1 real value $x_{-N/2} = x_{N/2}$

The identity

$$\sum_{k=0}^{N-1} e^{i2\pi(j-\ell)k/N} = N \, \delta_{j,\ell \bmod N} \tag{2.8}$$

shows us the inverse transform

$$x_j = \frac{1}{N} \sum_{k=0}^{N-1} \widehat{x}_k \, e^{i2\pi jk/N} = \frac{1}{N} \sum_{k=-N/2}^{N/2-1} \widehat{x}_k \, e^{i2\pi jk/N} \tag{2.9}$$

If we consider (2.5) to be an approximation of the integral in (2.1), we'd identify

$$\delta t \, \widehat{x}_k \sim \widetilde{x}(f_k) \tag{2.10}$$

If we plug (2.2) into (2.5) we can get the actual formula

$$\delta t \, \widehat{x}_k = \int_{-\infty}^{\infty} df \, \delta_{N,\delta t}(f_k - f) \widetilde{h}(f) \tag{2.11}$$

with a kernel

$$\delta_{N,\delta t}(x) = \delta t \sum_{j=0}^{N-1} e^{-i2\pi j \delta t \, x} \tag{2.12}$$

this is not quite a Dirac delta function for two reasons:

1. It is periodic with period $\frac{1}{\delta t}$, so it's peaked at $x = 0$, $x = \frac{1}{\delta t}$, $x = -\frac{1}{\delta t}$, etc.
2. It has an oscillatory "ringing" behavior around its peaks.

The second point is related to an issue known as spectral leakage which we won't go into; the first is known as aliasing, and it means that actually $\delta t \, \widehat{x}_k$ is a sum of not only $\widetilde{h}(f_k)$ but also $\widetilde{h}(f_k + \frac{1}{\delta t})$, $\widetilde{h}(f_k - \frac{1}{\delta t}) = \widetilde{h}^*(\frac{1}{\delta t} - f_k)$, etc. This means that to avoid confusion of different frequency components, the original time series $h(t)$ should have undergone some analog processing so that $\widetilde{h}(f)$ is negligible unless

$$-\frac{1}{2\,\delta t} < f < \frac{1}{2\,\delta t} \tag{2.13}$$

which defines the Nyquist frequency $f_{\mathrm{Ny}} = \frac{1}{2\,\delta t}$ which is half the sampling rate $\frac{1}{\delta t}$.

3 Random Data

We'll often be interested in cases where the data $\{x_i\}$ are random with some mean and variance defined by the expectation values

$$E[x_j] = \mu_j \tag{3.1}$$

$$E[(x_j - \mu_j)(x_\ell - \mu_\ell)] = \sigma^2_{j\ell} \tag{3.2}$$

If the data are Gaussian, these are enough to define a probability density[3]

$$P(\mathbf{x}) = (\det 2\pi\boldsymbol{\sigma}^2)^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\sigma}^{-2}(\mathbf{x}-\boldsymbol{\mu})\right) \tag{3.3}$$

where $\mathbf{x}$ and $\boldsymbol{\mu}$ are column vectors made up out of $\{x_j\}$ and $\{\mu_j\}$, respectively, $\boldsymbol{\sigma}^2$ is a matrix made of $\{\sigma_{j\ell}\}$ and $\boldsymbol{\sigma}^{-2}$ is its inverse. For simplicity we'll assume the data have zero mean. We'll also start in the continuous picture; the random process associated with $x(t)$ is stationary if

$$E\left[x(t), x(t')\right] = K_x(t-t') \tag{3.4}$$

which defines the autocorrelation function $K_x(t-t')$ (in general it would have to be written $K_x(t,t')$). The Fourier transform of the autocorrelation function is the two-sided power spectral density

$$S_x^{\text{2-sided}}(f) = \int_{-\infty}^{\infty} d\tau\, K_x(\tau)\, e^{-i2\pi f\tau} \tag{3.5}$$

We can use (2.3) to show that, formally,

$$E\left[\widetilde{x}(f')^*\,\widetilde{x}(f)\right] = \delta(f-f')\, S_x^{\text{2-sided}}(f) \tag{3.6}$$

[3] We'll call this $P(\mathbf{x})$ rather than $f(\mathbf{x})$ to avoid confusion with the frequency.

Since $S_x^{\text{2-sided}}(f) = S_x^{\text{2-sided}}(-f)$, for real $x(t)$, define one-sided PSD

$$S_x(f) = \begin{cases} S_x^{\text{2-sided}}(0) & f = 0 \\ S_x^{\text{2-sided}}(-f) + S_x^{\text{2-sided}}(f) & f > 0 \end{cases} \tag{3.7}$$

Unfortunately (?) this is what most GW observers mean by PSD, so formulas have an extra factor of two ($S_x(f) = 2S_x^{\text{2-sided}}(f)$), e.g.,

$$E\left[\widetilde{x}(f)^*\,\widetilde{x}(f)\right] = \delta(f-f')\,\frac{S_x(f)}{2} \tag{3.8}$$

We can translate this into a discrete Fourier transform; just as $\widehat{x}_k \sim \widetilde{x}(f_k)$, we can show

$$E\left[|\widehat{x}_k|^2\right] \sim \frac{N}{2\delta t} S_x(f_k) \tag{3.9}$$

with the usual caveats about leakage and aliasing. Now consider the case of zero-mean Gaussian data: let $\widehat{x}_k = \xi_k + i\eta_k$ and treat ξ_0, $\{\xi_k, \eta_k | k = 1 \ldots \frac{N}{2} - 1\}$, $\xi_{N/2}$ as independent and Gaussian with

$$E\left[\xi_k^2\right] = E\left[\eta_k^2\right] = \sigma_k^2 = \frac{N}{4\delta t} S_x(f_k) \tag{3.10}$$

so probability density is

$$\begin{aligned} P(\{\xi_k, \eta_k | k = 1 \ldots \frac{N}{2} - 1\}) &= \prod_{k=1}^{N/2-1} \frac{1}{2\pi\sigma_k^2} \exp\left(-\frac{\xi_k^2}{2\sigma_k^2} - \frac{\eta_k^2}{2\sigma_k^2}\right) \\ &\propto \exp(\Lambda) \end{aligned} \tag{3.11}$$

with log-likelihood

$$\Lambda \sim -\sum_{k=1}^{N/2-1} \frac{2\delta t}{N} \frac{|\widehat{x}_k|^2}{S_x(f_k)} \sim -\sum_{k=1}^{N/2-1} 2\delta f \frac{|\widetilde{x}(f_k)|^2}{S_x(f_k)} \sim -2\int_0^\infty df\, \frac{|\widetilde{x}(f)|^2}{S_x(f)} \tag{3.12}$$

This means

$$P(x) \propto e^{-\frac{1}{2}\langle x|x\rangle} \tag{3.13}$$

where the inner product is

$$\langle y|z\rangle = 4\,\mathrm{Re}\int_0^\infty df\,\frac{\widetilde{y}^*(f)\,\widetilde{z}(f)}{S_x(f)} \tag{3.14}$$

The unfamiliar factor of 4 is one factor of 2 because the integral is only over positive frequencies and one because of the use of the one-sided power spectral density.

If the data vary slowly over the observation time, it may be useful to divide it into pieces of length T and Fourier transform each of them

$$\widetilde{x}_I(f) = \int_{t_{I0}}^{t_{I0}+T} dt\,x(t)\,e^{-i2\pi f(t-t_{I0})} \tag{3.15}$$

In principle, the statistical properties of different segments will be related because of the autocorrelation function $K(t-t')$. But if the correlation length–the time over which $K(\tau)$ is non-negiligible–is small compared to T, we can neglect this, and the log likelihood function would become $P(x) \propto e^{\Lambda(x)}$ with

$$\Lambda = -2\,\mathrm{Re}\sum_I\int_0^{f_{\mathrm{Ny}}} df\,\frac{|\widetilde{x}_I(f)|^2}{S_I(f)} \tag{3.16}$$

Continuous Wave Data Analysis: Fully Coherent Methods

John T. Whelan

Extra Notes for IUCAA, Pune, 2014 January 17

Contents

1 Signal Model

1.1 GWs from rotating neutron star

We showed yesterday that the gravitational wave signal arriving at a time τ from an object a distance d away rotating with angular frequency Ω is

$$\overleftrightarrow{h}(\tau) = A_+ \cos[\Phi(\tau) + \phi_0] \overleftrightarrow{e}_+ + A_\times \sin[\Phi(\tau) + \phi_0] \overleftrightarrow{e}_\times \tag{1.1}$$

where

$$A_+ = h_0 \frac{1 + \cos^2 \iota}{2} \tag{1.2a}$$

$$A_\times = h_0 \cos \iota \tag{1.2b}$$

where the GW amplitude is

$$h_0 = \frac{4G\Omega^2(I_1 - I_2)}{c^4 d} \tag{1.3}$$

and the phase evolution is

$$\Phi(\tau) = 2\Omega(\tau - \tau_0) \tag{1.4}$$

Some notes:

- I_1, I_2 and I_3 are the moments of inertia about the principal axes. Rotational oblateness means that $I_3 > I_1, I_2$. For a perfect spheroid, I_1 and I_2 would be equal and there would be no gravitational radiation. To get a "triaxial" neutron star, you need some sort of deformation.

- The signal was derived for an object rotating at a constant velocity. Since real neutron stars generally spin down gradually, we can modify this by Taylor expanding the phase.

$$\Phi(\tau) = f_0(\tau - \tau_0) + \frac{1}{2} f_1(\tau - \tau_0)^2 + \frac{1}{3!} f_2(\tau - \tau_0)^3 + \ldots \tag{1.5}$$

The frequency f_0 and spindowns f_1, f_2, etc are parameters of the signal.

- The time τ is the time that the waves arrive at some inertial point, usually taken to be the solar system barycenter (SSB). Since we're interested in the signal that arrives at time t at the location of the detector, we have to use the time delay

$$t = \tau + \frac{\vec{k} \cdot (\vec{r}_{\text{det}} - \vec{r}_{\text{ssb}})}{c} + \text{relativistic corrections} \tag{1.6}$$

which translates into a Doppler shift which depends on the sky position, as represented by the propagation direction $\vec{k}$.

- The signal may also be Doppler-shifted by the proper motion of the neutron star. If this motion is inertial (constant velocity), this is just a constant Doppler shift which we fold into the definition of f_0. If the neutron star is in a binary system, we need to include the binary orbital parameters in the definition of the signal.

- The signal (1.1) is written in the preferred polarization basis $\overleftrightarrow{e}_{+,\times}$ constructed from unit vectors $\vec{\ell}$ and $\vec{m}$ orthogonal to each other and the propagation direction $\vec{k}$, with $\vec{\ell}$ lying in the neutron star's equatorial plane. This means they depend on the orientation of the neutron star's rotation axis, which is a parameter of the system. We can relate them to the canonical basis $\overleftrightarrow{\varepsilon}_{+,\times}$ constructed out of the vectors $\vec{\imath}$ and $\vec{\jmath}$ pointing West and North on the sky:

$$\overleftrightarrow{e}_+ = \overleftrightarrow{\varepsilon}_+ \cos 2\psi + \overleftrightarrow{\varepsilon}_\times \sin 2\psi \tag{1.7a}$$
$$\overleftrightarrow{e}_\times = -\overleftrightarrow{\varepsilon}_+ \sin 2\psi + \overleftrightarrow{\varepsilon}_\times \cos 2\psi \tag{1.7b}$$

1.2 Exercise: JKS decomposition

By using (1.7) and the angle sum formulas, expand out (1.1) so that its ϕ_0 and ψ dependence is explicit. Show that it can be written

$$\overleftrightarrow{h}(\tau) = \mathcal{A}^\mu \overleftrightarrow{h}_\mu(\tau) \tag{1.8}$$

(using the Einstein summation convention that repeated upper and lower indices are summed over, now with $\sum_{\mu=1}^4$) where

$$\mathcal{A}^1 = A_+ \cos 2\psi \cos\phi_0 - A_\times \sin 2\psi \sin\phi_0 \tag{1.9a}$$
$$\mathcal{A}^2 = A_+ \sin 2\psi \cos\phi_0 + A_\times \cos 2\psi \sin\phi_0 \tag{1.9b}$$
$$\mathcal{A}^3 = -A_+ \cos 2\psi \sin\phi_0 - A_\times \sin 2\psi \cos\phi_0 \tag{1.9c}$$
$$\mathcal{A}^4 = -A_+ \sin 2\psi \sin\phi_0 + A_\times \cos 2\psi \cos\phi_0 \; . \tag{1.9d}$$

and

$$\overleftrightarrow{h}_1(\tau) = \overleftrightarrow{\varepsilon}_+ \cos\Phi(\tau) \tag{1.10a}$$
$$\overleftrightarrow{h}_2(\tau) = \overleftrightarrow{\varepsilon}_\times \cos\Phi(\tau) \tag{1.10b}$$
$$\overleftrightarrow{h}_3(\tau) = \overleftrightarrow{\varepsilon}_+ \sin\Phi(\tau) \tag{1.10c}$$
$$\overleftrightarrow{h}_4(\tau) = \overleftrightarrow{\varepsilon}_\times \sin\Phi(\tau) \tag{1.10d}$$

2 Data Analysis Method

2.1 Likelihood function

The combinations (1.9), first discovered by Jaranowski, Królak and Schutz[1] mean that the parameters h_0, ι, ψ and ϕ_0 can be treated

[1] P. Jaranowski, A. Królak, and B. F. Schutz, *Phys Rev D* **58**, 062001 (1998), hereafter JKS. Note that the decomposition is not unique. J. T. Whelan, R. Prix, C. J. Cutler, and J. L. Willis, arXiv:1311.0065 [to appear in *CQG*] exhibit an alternative decomposition which is more closely connected to the physical amplitude parameters.

differently from the other signal parameters. They're known as *amplitude parameters*, as compared with the phase parameters (sky position, frequency, spindowns, binary orbital parameters, etc), also known as Doppler parameters, because most of them are associated with the Doppler modulation of the signal. We refer to them collectively as λ.

The signal in a detector at time t is then

$$h(t) = \overleftrightarrow{h}(\tau(t)) : \overleftrightarrow{d} = \mathcal{A}^{\mu} h_{\mu}(t;\lambda) \tag{2.1}$$

where the four signal waveforms are

$$\begin{aligned} h_1(t;\lambda) &= a(t;\lambda)\,\cos\Phi(t;\lambda) && \text{(2.2a)}\\ h_2(t;\lambda) &= b(t;\lambda)\,\cos\Phi(t;\lambda) && \text{(2.2b)}\\ h_3(t;\lambda) &= a(t;\lambda)\,\sin\Phi(t;\lambda) && \text{(2.2c)}\\ h_4(t;\lambda) &= b(t;\lambda)\,\sin\Phi(t;\lambda) && \text{(2.2d)} \end{aligned}$$

where $a = \overleftrightarrow{\varepsilon}_{+} : \overleftrightarrow{d}$ and $b = \overleftrightarrow{\varepsilon}_{\times} : \overleftrightarrow{d}$ are the amplitude modulation coëfficients for that detector and sky position. They also change slowly with time, since the rotation of the Earth changes the geometry of the double-dot products $\overleftrightarrow{\varepsilon}_{+,\times} : \overleftrightarrow{d}$. (The detector tensor $\overleftrightarrow{d}$ has constant components in a basis co-rotating with the Earth, and the polarization basis tensors $\overleftrightarrow{\varepsilon}_{+,\times}$ have constant components in an inertial basis.)

If we divide time up into shorter intervals of duration T, things are a little easier to describe; within a short timeframe, the signals are approximately monochromatic, and the AM coëfficients are approximately constant. Notationally, we add a label X to refer to the detector and I to refer to the time interval. If the data in instrument X during time interval I is

$$x^X_{Ij} = x^X(t_{I0} + j\delta t) \tag{2.3}$$

then we write its Fourier transform (also known as a "short Fourier transform" or SFT) as

$$\sum_{j=0}^{N-1} \delta t\, x^X_{Ij}\, e^{-i2\pi jk/N} \sim \int_{t_{I0}}^{t_{I0}+T} dt\, x(t)\, e^{-i2\pi f_k(t-t_{I0})} \equiv \widetilde{x}^X_I(f_k) \tag{2.4}$$

where $T = N\,\delta t$, $f_k = k\,\delta f$, $\delta f = \frac{N}{\delta t}$, etc. The signal contribution to the Ith SFT in detector X is

$$\widetilde{h}^X_I(f) = \mathcal{A}^{\mu} \widetilde{h}^X_{I,\mu}(f;\lambda) \tag{2.5}$$

where

$$\begin{aligned} \widetilde{h}^X_{I,1}(f;\lambda) &= a^X_I(\lambda)\,\widetilde{\cos\Phi^X_I}(f;\lambda) && \text{(2.6a)}\\ \widetilde{h}^X_{I,2}(f;\lambda) &= b^X_I(\lambda)\,\widetilde{\cos\Phi^X_I}(f;\lambda) && \text{(2.6b)}\\ \widetilde{h}^X_{I,3}(f;\lambda) &= a^X_I(\lambda)\,\widetilde{\sin\Phi^X_I}(f;\lambda) && \text{(2.6c)}\\ \widetilde{h}^X_{I,4}(f;\lambda) &= b^X_I(\lambda)\,\widetilde{\sin\Phi^X_I}(f;\lambda) && \text{(2.6d)} \end{aligned}$$

If the data in each detector consist of Gaussian noise $n^X_I(t)$ with a one-sided power spectral density $S^X_I(f)$, plus the signal,

$$\widetilde{x}^X_I(f) = \widetilde{n}^X_I(f) + \mathcal{A}^{\mu}\widetilde{h}^X_{I,\mu}(f;\lambda) \tag{2.7}$$

then the probability density of the data can be written as

$$P(x|\mathcal{A},\lambda) \propto e^{\Lambda(x|\mathcal{A},\lambda)} \tag{2.8}$$

where the log-likelihood is

$$\Lambda(x|\mathcal{A},\lambda) = -2\sum_I\sum_X \int_0^{f_{\mathrm{Ny}}} df\, \frac{\left|\widetilde{n}^X_I(f)\right|^2}{S^X_I(f)} = -\frac{1}{2}\langle x-h|x-h\rangle \tag{2.9}$$

and we've defined the inner product

$$\langle y|z\rangle = 4\,\mathrm{Re}\sum_I\sum_X \int_0^{f_{\mathrm{Ny}}} df\, \frac{\widetilde{y}^X_I(f)^*\,\widetilde{z}^X_I(f)}{S^X_I(f)} \tag{2.10}$$

On the other hand, if the data consist only of noise, then the probability density is

$$P(x|\mathcal{H}_N) \propto e^{-\frac{1}{2}\langle x|x\rangle} \tag{2.11}$$

That makes the log-likelihood ratio

$$\begin{aligned}
\ln\frac{P(x|\mathcal{A},\lambda)}{P(x|\mathcal{H}_N)} &= \frac{1}{2}\left[\langle x|h(\mathcal{A},\lambda)\rangle + \langle h(\mathcal{A},\lambda)|x\rangle - \langle h(\mathcal{A},\lambda)|h(\mathcal{A},\lambda)\rangle\right] \\
&= \mathcal{A}^\mu\langle h_\mu(\lambda)|x\rangle - \frac{1}{2}\mathcal{A}^\mu\langle h_\mu(\lambda)|h_\nu(\lambda)\rangle\mathcal{A}^\mu \\
&= \mathcal{A}^\mu x_\mu(\lambda) - \frac{1}{2}\mathcal{A}^\mu\mathcal{M}_{\mu\nu}(\lambda)\mathcal{A}^\nu
\end{aligned} \tag{2.12}$$

in terms of a "metric"

$$\mathcal{M}_{\mu\nu}(\lambda) = \langle h_\mu(\lambda)|h_\nu(\lambda)\rangle = 4\,\mathrm{Re}\sum_I\sum_X\int_0^{f_{\mathrm{Ny}}} df\,\frac{\widetilde{h}^X_{\mu I}(f;\lambda)^*\,\widetilde{h}^X_{\nu I}(f;\lambda)}{S^X_I(f)} \tag{2.13}$$

on amplitude parameter space and a data-vector

$$x_\mu(\lambda) = \langle h_\mu(\lambda)|x\rangle = 4\,\mathrm{Re}\sum_I\sum_X\int_0^{f_{\mathrm{Ny}}} df\,\frac{\widetilde{h}^X_{\mu I}(f;\lambda)^*\,\widetilde{x}^X_I(f)}{S^X_I(f)} \tag{2.14}$$

2.2 Detection Statistic

To search for gravitational waves, we need to have a way to decide between a signal hypothesis $\mathcal{H}_S$ and the noise hypothesis $\mathcal{H}_N$. The Neyman-Pearson lemma says that the best detection statistic we can use is a monotonic function of the likelihood ratio

$$\frac{P(x|\mathcal{H}_S)}{P(x|\mathcal{H}_N)} \tag{2.15}$$

In general, the signal hypothesis $\mathcal{H}_S$ represents a family of signals with different values for the signal parameters λ and $\mathcal{A}$, and the likelihood ratio involves marginalizing over those parameters:

$$\int d\mathcal{A}\,d\lambda\,P(\mathcal{A},\lambda|\mathcal{H}_S)\,\frac{P(x|\mathcal{A},\lambda)}{P(x|\mathcal{H}_N)} \tag{2.16}$$

This has been considered recently[2], but choosing a physical prior on the signal parameters leads to a more mathematically involved integration problem. So we'll focus for now on a method developed by JKS, where the likelihood ratio is analytically maximized over the amplitude parameters to produce a maximum-likelihood statistic

$$\mathcal{F}(\lambda) = \max_{\mathcal{A}}\ln\frac{P(x|\mathcal{A},\lambda)}{P(x|\mathcal{H}_N)} = \frac{1}{2}\widehat{\mathcal{A}}^\mu\mathcal{M}_{\mu\nu}\widehat{\mathcal{A}}^\nu \tag{2.17}$$

where, if we define $\mathcal{M}^{\mu\nu}$ as the matrix inverse of $\mathcal{M}_{\mu\nu}$, the maximum-likelihood values for the amplitude parameters are

$$\widehat{\mathcal{A}}^\mu = \mathcal{M}^{\mu\nu}x_\nu \tag{2.18}$$

and the $\mathcal{F}$ statistic is

$$\mathcal{F}(\lambda) = \frac{1}{2}x_\mu(\lambda)\mathcal{M}^{\mu\nu}(\lambda)x_\nu(\lambda) \tag{2.19}$$

Note that even constructing this statistic means choosing values for the phase parameters λ: sky position, frequency, spindowns, etc. In the case of a targeted search for gravitational waves from a neutron star seen as a radio or X-ray pulsar, this is no big deal; we know the sky position and the spin, including its evolution. When looking for unknown objects, we have to try a bunch of different points in parameter space. How far off you can be in parameter space before the detection statistic becomes ineffective

[2]R. Prix and B. Krishnan, *CQG* **26**, 204013 (2009)

depends on the coherent observing time: how far apart the earliest and latest observations are. For a fully coherent search, the number of points needed to cover parameter space grows rapidly with observing time, until the coherent search becomes impossible. Instead so-called semi-coherent methods, which are the topic of a different lecture, must be used.

2.3 Statistical properties of the $\mathcal{F}$-statistic

Focus now on the case where the phase parameters are known. If the unknown amplitude parameters are $\{\mathcal{A}^\mu\}$, we can show that the data vector has expectation value

$$E[x_\mu] = \mathcal{M}_{\mu\nu}\mathcal{A}^\nu = \mu_\mu \tag{2.20}$$

and variance

$$E[(x_\mu - \mu_\mu)(x_\nu - \mu_\nu)] = \mathcal{M}_{\mu\nu} \tag{2.21}$$

Since it's a linear combination of Gaussian data, it is itself Gaussian, i.e., its pdf is

$$P(\{x_\mu\}|\mathcal{A}) = [\det(2\pi\mathcal{M}_{\mu\nu})]^{-1/2} \exp\left(\frac{1}{2}(x_\mu - \mu_\mu)\mathcal{M}^{\mu\nu}(x_\nu - \mu_\nu)\right) \tag{2.22}$$

or equivalently, the pdf of the maximum likelihood estimates $\{\widehat{\mathcal{A}}^\mu\}$ is

$$P(\{\widehat{\mathcal{A}}^\mu\}|\mathcal{A}) = [\det(2\pi\mathcal{M}^{\mu\nu})]^{-1/2} \exp\left(\frac{1}{2}(\widehat{\mathcal{A}}^\mu - \mathcal{A}^\mu)\mathcal{M}_{\mu\nu}(\widehat{\mathcal{A}}^\nu - \mathcal{A}^\nu)\right) \tag{2.23}$$

$$2\mathcal{F} = x_\mu\mathcal{M}^{\mu\nu}x_\nu = \widehat{\mathcal{A}}^\mu\mathcal{M}_{\mu\nu}\widehat{\mathcal{A}}^\nu \tag{2.24}$$

obeys what's known as a non-central chi-squared distribution with four degrees of freedom and non-centrality parameter $\mathcal{A}^\mu\mathcal{M}_{\mu\nu}\mathcal{A}^\nu$. This means

$$E[2\mathcal{F}] = 4 + \mathcal{A}^\mu\mathcal{M}_{\mu\nu}\mathcal{A}^\nu \tag{2.25}$$

Searches for a Stochastic Gravitational-Wave Background

John T. Whelan

Extra Notes for IUCAA, Pune, 2014 January 17

Contents

1 Stochastic GW Backgrounds

Consider a superposition of many weak gravitational-wave sources. This may be of cosmological origin, associated with events in the early universe (inflation, phase transition, primordial gravitons, ...), or some unresolved astrophysical source, such as millions of white-dwarf binaries in our galaxy. The individual signals may not be detectable, but their combined effect would produce a random signal in gravitational wave detectors, analogous to the cosmic microwave background first observed by Penzias and Wilson.[1] Unlike Penzias and Wilson, we can't "point our detectors away from the sky", but we can distinguish a random GW signal from random instrumental noise, because of the correlations it would produce between the outputs of different detectors.

We can describe any superposition of gravitational waves by expanding it as a superposition of plane waves along each propagation vector $\vec{k}$:

$$\overleftrightarrow{h}(\vec{r},t) = \sum_{A=+,\times} \int_{-\infty}^{\infty} df \iint d^2\Omega_{\vec{k}}\, h_A(f,\vec{k})\, \overleftrightarrow{e}_A(\vec{k}) \exp\left(i2\pi f\left[t - \frac{\vec{k}\cdot\vec{r}}{c}\right]\right) \tag{1.1}$$

The statistical properties of $h_A(f,\vec{k})$ describe the nature of the stochastic background. Since it's a superposition of many individual signals, it's reasonable to assume it's Gaussian, stationary, and unpolarized, so that it's defined by its mean

$$E\left[h_A(f,\vec{k})\right] = 0 \tag{1.2}$$

and variance

$$E\left[h^*_A(f,\vec{k})\, h_{A'}(f',\vec{k}')\right] = \delta^2(\vec{k},\vec{k}')\, \delta_{AA'}\, \delta(f-f')\, H(f,\vec{k}) \tag{1.3}$$

[1] A. A. Penzias and R. W. Wilson, *ApJ* **142**, 419 (1965)

Specifically,

$$P(h) \propto \exp\left(-\frac{1}{2}\sum_{A=+,\times}\int_{-\infty}^{\infty} df \iint d^2\Omega_{\vec{k}}\, \frac{|h_A(f,\vec{k})|^2}{H(f,\vec{k})}\right) \tag{1.4}$$

If we consider the stochastic signal $h^X(t) = \overleftrightarrow{h}(t) : \overleftrightarrow{d}^X$ appearing in a detector X, it will also be Gaussian with zero mean; the covariance between data in detectors X and Y will be

$$\begin{aligned} E\left[\widetilde{h}^X(f')^*\widetilde{h}^Y(f)\right] = \sum_{A=+,\times}\iint d^2\Omega_{\vec{k}}\, H(f,\vec{k})\, F_A^X(\vec{k})F_A^Y(\vec{k}) \\ \times \exp\left(i2\pi f\left[\frac{-\vec{k}\cdot(\vec{r}^X-\vec{r}^Y)}{c}\right]\right)\delta(f-f') \end{aligned} \tag{1.5}$$

We can gain some insight into this if we write

$$\begin{aligned} \sum_{A=+,\times} F_A^X(\vec{k})F_A^Y(\vec{k}) &= \sum_{A=+,\times} d^X_{ab}e_A^{ab}(\vec{k})\, e_A^{cd}(\vec{k})\, d^Y_{cd} \\ &= 2d^X_{ab}\, P^{\mathrm{TT}\vec{k}ab}{}_{cd}\, d^{Y\,cd} \end{aligned} \tag{1.6}$$

where

$$P^{\mathrm{TT}\vec{k}ab}{}_{cd} = \frac{1}{2}\sum_{A=+,\times} e_A^{ab}(\vec{k})\, e_{A\,cd}(\vec{k}) \tag{1.7}$$

is an operator which projects onto the subspace of traceless, symmetric tensors transverse to the unit vector $\vec{k}$.

1.1 Exercise

Show that this is a projection operator, i.e., $P^{\mathrm{TT}\vec{k}ab}{}_{ef}P^{\mathrm{TT}\vec{k}ef}{}_{cd} = P^{\mathrm{TT}\vec{k}ab}{}_{cd}$, by using the normalization $e_A^{ab}(\vec{k})\, e_{B\,ab}(\vec{k}) = 2\delta_{AB}$ of the standard polarization basis tensors (see yesterday's lecture). What is the trace $P^{\mathrm{TT}\vec{k}ab}{}_{ab}$?

1.2 Spatial Distributions

The simplest signal geometry for a stochastic background is isotropic, so that $H(f,\vec{k}) = H(f)$. A slightly less specific assumption (although not fully general) is that we're interested in a background distributed in some way across the sky, whose spectrum is the same in each direction, i.e., $H(f,\vec{k}) = H(f)\mathcal{P}(\vec{k})$. There are several different strategies that are taken to address the direction dependence of a stochastic background:

- Search only for an isotropic background.
- Search for a background with a specified sky distribution, e.g., spread across a nearby galaxy cluster, or concentrated at a point.
- Attempt to reconstruct a sky map, e.g., by measuring the power in different spherical harmonics.

In this lecture we'll focus on the isotropic case, although the notes will include some formulas involving $\mathcal{P}(\vec{k})$.

If the spectrum factors, the correlation between different detectors is

$$E\left[\widetilde{h}^X(f')^*\widetilde{h}^Y(f)\right] = \gamma^{XY}(f)\frac{S_{\mathrm{gw}}(f)}{2}\delta(f-f') \tag{1.8}$$

where

$$\gamma^{XY}(f) = d^X_{ab}\, d^{Y\,cd}\, \frac{5}{4\pi}\iint d^2\Omega_{\vec{k}}\, \mathcal{P}(\vec{k})\, P^{\mathrm{TT}\vec{k}ab}{}_{cd}\, e^{-i2\pi f\vec{k}\cdot(\vec{r}^X-\vec{r}^Y)/c} \tag{1.9}$$

is known as the overlap reduction function, and

$$S_{\mathrm{gw}}(f) = \frac{16\pi}{5}H(f) \tag{1.10}$$

This normalization is chosen because, in the isotropic case where

$$\gamma^{XY}(f) = d^X_{ab}\, d^{Y\,cd}\, \frac{5}{4\pi} \iint d^2\Omega_{\vec{k}}\, P^{\mathrm{TT}\vec{k}ab}{}_{cd}\, e^{-i2\pi f\vec{k}\cdot(\vec{r}^X - \vec{r}^Y)/c} \tag{1.11}$$

the overlap reduction function for a detector with itself is

$$\gamma^{XX}(f) = d^X_{ab}\, d^{X\,cd}\, \frac{5}{4\pi} \iint d^2\Omega_{\vec{k}}\, P^{\mathrm{TT}\vec{k}ab}{}_{cd} = 2 d^X_{ab}\, d^{X\,ab} \tag{1.12}$$

For a standard interferometer with perpendicular arms, this is just 1, and so $S_{\mathrm{gw}}(f)$ is the contribution to the power spectrum in the detector due to the stochastic gravitational wave background.

1.3 Exercise

Show that

$$\frac{5}{4\pi} \iint d^2\Omega_{\vec{k}} P^{\mathrm{TT}\vec{k}ab}{}_{cd} = 2P^{\mathrm{T}ab}{}_{cd} \tag{1.13}$$

where $P^{\mathrm{T}ab}{}_{cd}$ is a projector onto transverse symmetric tensors, in the following way

1. Argue by symmetry that it must be a constant times $P^{\mathrm{T}ab}{}_{cd}$.

2. Show that that constant has to be 2 by taking the trace of both sides of the equation

$$\frac{5}{4\pi} \iint d^2\Omega_{\vec{k}} P^{\mathrm{TT}\vec{k}ab}{}_{cd} \propto P^{\mathrm{T}ab}{}_{cd} \tag{1.14}$$

1.4 Aside: $\Omega_{\mathrm{gw}}(f)$

The spectrum of a gravitational wave background is often described in terms of the contribution to the cosmological parameter $\Omega = \rho/\rho_{\mathrm{crit}}$ where

$$\rho_{\mathrm{crit}} = \frac{3H_0^2}{8\pi G} \tag{1.15}$$

using the fact that the energy density in gravitational waves is

$$\begin{aligned}\rho_{\mathrm{gw}} &= \frac{c^2}{32\pi G}\langle \dot{h}_{ab}(t,\vec{r})\dot{h}^{ab}(t,\vec{r})\rangle = \frac{\pi c^2}{G}\int_0^\infty f^2 \iint d^2\Omega_{\vec{k}}\, H(f,\vec{k}) \\ &= \frac{4\pi^2 c^2}{G}\int_0^\infty f^2\, H(f)\, df\end{aligned} \tag{1.16}$$

The usual definition is the logarithmic energy density

$$\Omega_{\mathrm{gw}}(f) = \frac{f}{\rho_{\mathrm{crit}}}\frac{d\rho_{\mathrm{gw}}}{df} = \frac{32\pi^2}{3H_0^2} f^3 H(f) = \frac{10\pi}{3H_0^2} f^3 S_{\mathrm{gw}}(f) \tag{1.17}$$

This is of interest because some cosmological models (e.g., slow-roll inflation) predict a spectrum which corresponds to a constant $\Omega_{\mathrm{gw}}(f)$, but we will work in terms of $S_{\mathrm{gw}}(f)$ in this lecture because

1. The equations are simpler in terms of $S_{\mathrm{gw}}(f)$

2. $\Omega_{\mathrm{gw}}(f)$ depends on the (experimentally uncertain) value of the Hubble constant H_0

1.5 Overlap Reduction Function

Restricting attention now to an isotropic background, consider the overlap reduction function This normalization is chosen because, in the isotropic case where

$$\gamma^{XY}(f) = d^X_{ab}\, d^{X\,cd}\, \frac{5}{4\pi} \iint d^2\Omega_{\vec{k}}\, P^{\mathrm{TT}\vec{k}ab}{}_{cd}\, e^{-i2\pi f\vec{k}\cdot(\vec{r}^Y - \vec{r}^X)/c} \tag{1.18}$$

We've seen that it is equal to unity when X and Y refer to the same interferometric detector (as long as the arms are perpendicular). For any pair of detectors, it's a specific function of frequency, and in particular won't depend on when the observation is done. (Isotropy means that the rotation of the Earth doesn't change the observing geometry.) There are thus only $N_{\text{sites}}(N_{\text{sites}}-1)/2$ different functions to be worked out, where N_{sites} is the number of detector sites (e.g., in the initial detector era, $N_{\text{sites}}=4$: LIGO Hanford, LIGO Livingston, GEO and Virgo[2]; for the advanced detector era, we can add KAGRA and LIGO India.) It might seem like each of these just requires a numerical integration over the sky (propagation direction $\vec{k}$), but its even easier than that, since it's possible to work out explicit formulas for $\gamma^{XY}(f)$ in terms of the detector tensors and the separation vectors of the detectors.[3] Some examples are plotted in figure 1.

2 Data Analysis Method

2.1 Likelihood Ratio

We've described a signal model in which the signal contribution $\widetilde{h}^X(f)$ in the Fourier domain to detector X's output is Gaussian, with $E\left[\widetilde{h}(f)\right]=0$ and

$$E\left[\widetilde{h}^X(f')^*\widetilde{h}^Y(f)\right]=\gamma^{XY}(f)\frac{S_{\text{gw}}(f)}{2}\delta(f-f') \tag{2.1}$$

If we want to talk about breaking the data up into intervals of duration T labelled by I before Fourier transforming, this becomes

$$E\left[\widetilde{h}_I^X(f')^*\widetilde{h}_J^Y(f)\right]=\delta_{IJ}\gamma^{XY}(f)\frac{S_{\text{gw}}(f)}{2}\delta(f-f') \tag{2.2}$$

Note that for an isotropic background, the overlap reduction function does not change with time and therefore no additional I subscript is needed on $\gamma^{XY}(f)$.

[2]Also TAMA, depending on when you define the era. Resonant bar detectors can also be added to the picture, and were.

[3]See e.g., B. Allen and J. D. Romano *PRD* **59**, 102001 (1999) or J. T. Whelan *CQG* **23**, 1181 (2006).

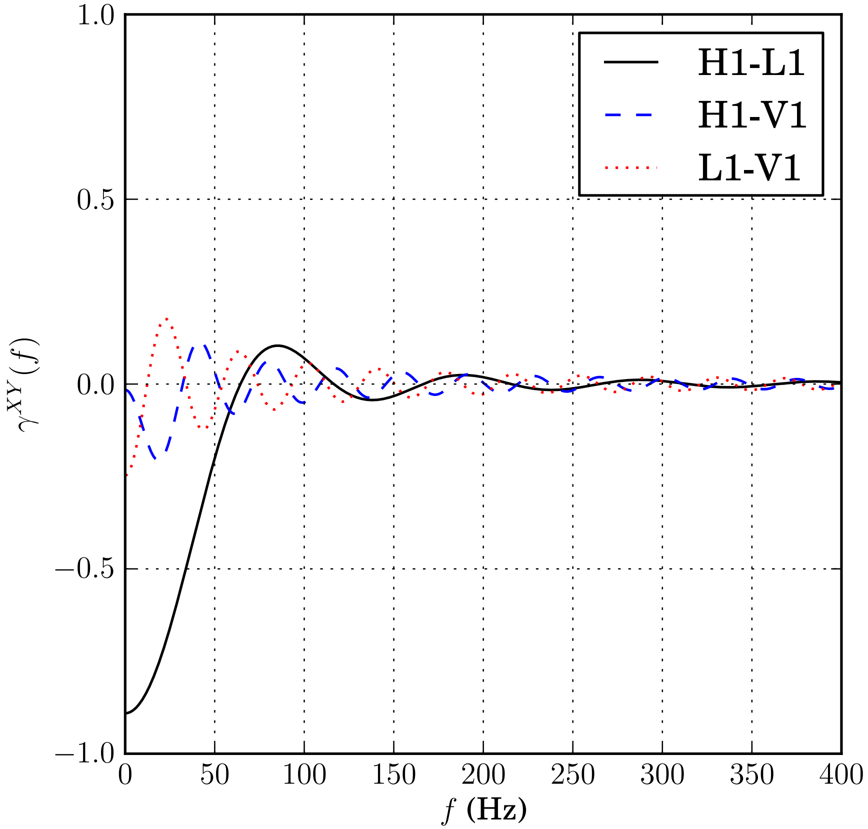

Figure 1: Plots of the isotropic overlap reduction function $\gamma^{XY}(f)$ for pairs of detectors among LIGO Hanford (H1), LIGO Livingston (L1) and Virgo. The overlap reduction function tends to oscillate and decrease in amplitude with increasing frequency, as correlations and anti-correlations of waves from different directions cancel. See Cella et al, *CQG*, **24**, S639 (2007) for more discussion.

If we assume that the data $\widetilde{x}_I^X(f)$ consist of this signal plus instrumental noise $\widetilde{n}_I^X(f)$ which is assumed to be zero-mean, independent between detectors and Gaussian with a one-sided power spectrum $S_I^X(f)$, the detector output will also be Gaussian, with $E\left[\widetilde{x}_I^X(f)\right] = 0$, and

$$E\left[\widetilde{x}_I^X(f')^* \widetilde{x}_J^Y(f)\right] = \delta_{IJ} \frac{S_I^{XY}(f)}{2} \delta(f - f') \tag{2.3}$$

with

$$S_I^{XY}(f) = \delta^{XY}\, S_I^X(f) + \gamma^{XY}(f)\, S_{\text{gw}}(f) \tag{2.4}$$

We'd like to construct a likelihood ratio $\frac{P(x|\mathcal{H}_s)}{P(x|\mathcal{H}_g)}$ between a model consisting of a stochastic signal plus noise to one with just Gaussian noise. For a given spectrum $S_{\text{gw}}(f)$, our assumptions tell us

$$P(x|\mathcal{H}_s, S_{\text{gw}}(f))) \propto \exp\left(-2 \int_0^{f_{\text{Ny}}} df \sum_I \sum_X \sum_Y \widetilde{x}_I^X(f)^*\, S^{-1}_{I\,XY}(f)\, \widetilde{x}_J^Y(f)\right) \tag{2.5}$$

where $S^{-1}_{I\,XY}(f)$ is the matrix inverse of $S_I^{XY}(f)$. If the spectrum in each detector is dominated by the noise, i.e., $S_{\text{gw}}(f) \ll S_I^X(f)$, we can use

$$S_I^{XY}(f) = \sqrt{S_I^X(f)} \left(\delta^{XY} + \frac{\gamma^{XY}(f)\, S_{\text{gw}}(f)}{\sqrt{S_I^X(f) S_I^Y(f)}}\right) \sqrt{S_I^Y(f)} \tag{2.6}$$

to approximate

$$\begin{aligned} S^{-1}_{I\,XY}(f) &\approx \frac{1}{\sqrt{S_I^X(f)}} \left(\delta^{XY} - \frac{\gamma^{XY}(f)\, S_{\text{gw}}(f)}{\sqrt{S_I^X(f) S_I^Y(f)}}\right) \frac{1}{\sqrt{S_I^Y(f)}} \\ &= \frac{\delta^{XY}}{S_I^X(f)} - \frac{\gamma^{XY}(f) S_{\text{gw}}(f)}{S_I^X(f) S_I^Y(f)} \end{aligned} \tag{2.7}$$

In this approximation, the logarithm of the likelihood ratio is

$$\ln \frac{P(x|\mathcal{H}_s, S_{\text{gw}}(f))}{P(x|\mathcal{H}_g)} \approx 2 \sum_I \sum_X \sum_Y \int_0^{f_{\text{Ny}}} df \frac{\gamma^{XY}(f) S_{\text{gw}}(f)}{S_I^X(f) S_I^Y(f)} \widetilde{x}_I^X(f)^* \, \widetilde{x}_I^Y(f) \tag{2.8}$$

We might want to consider a whole bunch of signal hypotheses with different spectra $S_{\text{gw}}(f)$, but for simplicity we know the shape of the spectrum so that $S_{\text{gw}}(f) = S_R\, \mathcal{S}(f)$ where $\mathcal{S}(f)$ is some known function over the frequencies of interest (e.g., constant or proportional to f^{-3}, the latter corresponding to constant $\Omega_{\text{gw}}(f)$). Then

$$\ln \frac{P(x|\mathcal{H}_s, S_R)}{P(x|\mathcal{H}_g)} \approx 2S_R \sum_I \sum_X \sum_Y \int_0^{f_{\text{Ny}}} df \frac{\gamma^{XY}(f) \mathcal{S}(f)}{S_I^X(f) S_I^Y(f)} \widetilde{x}_I^X(f)^* \, \widetilde{x}_I^Y(f) \tag{2.9}$$

2.2 Cross-Correlation Statistic

The standard stochastic background search doesn't quite use this as a detection statistic, though. That's because the log likelihood ratio includes terms with $X = Y$. These autocorrelation terms represent the contribution of the stochastic GW background to the power in each detector. While including them would make the search more sensitive if all of the assumptions that went into constructing (2.9) were true, it relies on the assumption that the instrumental noise in each detector is Gaussian. Rather than try to construct a more sophisticated statistic involving a more realistic noise hypothesis, in practice we just leave out those autocorrelation terms and construct a cross-correlation statistic

$$\mathcal{Y} = \sum_{X>Y} \sum_Y \sum_I \mathcal{Y}_{IXY} \tag{2.10}$$

with

$$\mathcal{Y}_{IXY} = 2 \int_0^{f_{\text{Ny}}} df \frac{\gamma^{XY}(f) \mathcal{S}(f)}{S_I^X(f) S_I^Y(f)} \widetilde{x}_I^X(f)^* \, \widetilde{x}_I^Y(f) \tag{2.11}$$

The analysis method then constructs the posterior pdf for the stochastic background strength S_R according to Bayes's theorem

$$P(S_R|\mathcal{Y}) = \frac{P(\mathcal{Y}|S_R)P(S_R)}{P(\mathcal{Y})} \tag{2.12}$$

where the prior $P(S_R)$ is taken to be uniform for $0 < S_R < S_{\max}$, with $S_{\max}$ large enough not to influence the construction, and the normalization $\frac{1}{P(\mathcal{Y})}$ calculated by requiring that $\int_0^{S_{\max}} P(S_R|\mathcal{Y}) = 1$.

Rather than worry about the actual distribution $P(\mathcal{Y}|S_R)$, we use the fact that $\mathcal{Y}$ is a sum of contributions from many times and frequencies to invoke the central limit theorem, and approximate it as Gaussian, so we just need its mean and variance. We calculate the mean using

$$E\left[\mathcal{Y}_{IXY}\right] = 2\int_0^{f_{\text{Ny}}} df\, \frac{\gamma^{XY}(f)\mathcal{S}(f)}{S_I^X(f)S_I^Y(f)}\, E\left[\widetilde{x}_I^X(f)^*\, \widetilde{x}_I^Y(f)\right] \tag{2.13}$$

This seems like a bit of a problem, because taking (2.3) literally (and recalling $X \neq Y$ would seem to indicate

$$E\left[\widetilde{x}_I^X(f)^*\, \widetilde{x}_I^Y(f)\right] = \gamma^{XY}(f)\, \frac{S_{\text{gw}}(f)}{2}\, \delta(0) \tag{2.14}$$

which is infinite. But a careful treatment notes that since $\widetilde{x}_I^X(f)$ is not literally a continuous Fourier transform, only a finite-time approximation of one, we really should have put in the finite-time approximation of the dirac delta, and we should replace $\delta(0)$ with the time baseline for the Fourier transform, $\frac{1}{\delta f} = T$. Thus

$$E\left[\mathcal{Y}_{IXY}\right] = T\, S_R \int_0^{f_{\text{Ny}}} df\, \frac{[\gamma^{XY}(f)\mathcal{S}(f)]^2}{S_I^X(f)S_I^Y(f)} \tag{2.15}$$

The calculation of the variance is simpler if we assume as usual that $S_{\text{gw}}(f) \ll S_I^X(f)$, leaving us with only

$$\begin{aligned} E\left[(\mathcal{Y}_{IXY})^2\right] &\approx T\int_0^{f_{\text{Ny}}} df\, \left(\frac{\gamma^{XY}(f)\mathcal{S}(f)}{S_I^X(f)S_I^Y(f)}\right)^2\, S_I^X(f)S_I^Y(f) \\ &= T\int_0^{f_{\text{Ny}}} df\, \frac{[\gamma^{XY}(f)\mathcal{S}(f)]^2}{S_I^X(f)S_I^Y(f)} \end{aligned} \tag{2.16}$$

where we have used the fact that

$$E\left[\widetilde{x}_I^X(f)^*\, \widetilde{x}_I^Y(f)\widetilde{x}_I^X(f')\, \widetilde{x}_I^Y(f')^*\right] = \frac{S_I^X(f)}{2}\delta(f-f')\frac{S_I^Y(f)}{2}\delta(f-f') \tag{2.17}$$

The mean and variance of the statistic $\mathcal{Y}$ are thus

$$E\left[\mathcal{Y}\right] = S_R\mathcal{I} \qquad E\left[\mathcal{Y}^2\right] = \mathcal{I} \tag{2.18}$$

where

$$\mathcal{I} = \sum_I \sum_{X>Y} \sum_Y T\int_0^{f_{\text{Ny}}} df\, \frac{[\gamma^{XY}(f)\mathcal{S}(f)]^2}{S_I^X(f)S_I^Y(f)} \tag{2.19}$$

This integral encapsulates the sensitivity of the search, since

$$P(\mathcal{Y}|S_R) = \frac{1}{\sqrt{2\pi\mathcal{I}}}\exp\left(-\frac{1}{2}\frac{(\mathcal{Y}-S_R\mathcal{I})^2}{\mathcal{I}}\right) \tag{2.20}$$

and the posterior becomes

$$P(S_R|\mathcal{Y}) = \frac{e^{-\mathcal{I}(S_R-\mathcal{Y}/\mathcal{I})^2/2}}{\int_0^\infty dS'_R\, e^{-\mathcal{I}(S'_R-\mathcal{Y}/\mathcal{I})^2/2}} \tag{2.21}$$

2.3 Exercise: Relation to standard formulas

This is not actually the way things are usually written; instead one normalizes a statistic $\mathcal{E}_{IXY} = \mathcal{N}_{IXY}\mathcal{Y}_{IXY}$ so that $E\left[\mathcal{E}_{IXY}\right] = S_R$ and $E\left[\mathcal{E}_{IXY}\right]^2 = \sigma_{IXY}^2$ and then constructs an optimal estimator

$$\mathcal{E} = \frac{\sum_I \sum_{X>Y} \sum_Y \sigma_{IXY}^{-2}\mathcal{E}_{IXY}}{\sum_I \sum_{X>Y} \sum_Y \sigma_{IXY}^{-2}} \tag{2.22}$$

Show that $\mathcal{E} = \mathcal{Y}/\mathcal{I}$, so that the two prescriptions are equivalent.