

Application of matched-filtering to analysis of nearly periodic gravitational wave signals

Andrzej Królak^{1,2}

¹Institute of Mathematics, Polish Academy of Sciences
00-656 Warsaw, Poland

²National Center for Nuclear Research
05-400 Świerk-Otwock, Poland

Multi-Messenger CW workshop, Amsterdam 11-13 July 2023

- A. D. Whalen, Detection of Signals in Noise (Academic Press, 1971).
- M. H. A. Davies, in Gravitational Wave Data Analysis, edited by B. F. Schutz (Kluwer, Dordrecht, 1989), pp. 73-94.
- P. Jaranowski, A. Królak, B. F. Schutz, Phys. Rev. D 58, 063001 (1998): the \mathcal{F} - statistic.
- P. Jaranowski and A. Królak, Analysis of Gravitational-Wave Data (Cambridge University Press, Cambridge, 2009).

Signal detection as hypothesis testing

Suppose we have random data x and we want to know whether a signal s is present in the data.

- Null hypothesis H_0 : signal is absent, data have distribution $p_0(x)$,
- Alternative hypothesis H_1 : signal is present, data have distribution $p_1(x)$.

A hypothesis test (or decision rule) δ is a partition of the observation set into two subsets R and its complement R' . If data are in R we accept null hypothesis otherwise we reject it.

Neyman-Pearson approach

There are two kind of errors that we can make. Type I error is choosing hypothesis H_1 when H_0 is true and type II error is choosing H_0 when H_1 is true.

- Type I error - false alarm probability: $P_F = \int_{R'} p_0(x) dx$
- Type II error - false dismissal probability: $P_S = 1 - \int_R p_1(x) dx$

Neyman-Pearson lemma:

Test that maximizes probability of detection subject to fixed false alarm probability is the likelihood ratio test.

Thus the test is to compare the likelihood ratio Λ

$$\Lambda = \frac{p_1(x)}{p_0(x)}$$

to a threshold calculated from the accepted false alarm probability.

- Very often we know the waveform of the signal that we are searching for in the data in terms of a finite number of unknown parameters.
- We would like to find optimal procedures of estimating these parameters.
- An estimator of a parameter θ is a function $\hat{\theta}(x)$ that assigns to each data the "best" guess of the true value of θ .
- Note that because $\hat{\theta}(x)$ depends on the random data an estimator is always a random variable.

Parameter estimation cont.

The simplest, although an *ad hoc* method, is the maximum likelihood estimation.

We introduce the *likelihood function*:

$$l(\theta; \mathbf{x}) := p(\mathbf{x}; \theta)$$

Maximum likelihood (ML) estimators are obtained by solving the equation

$$\frac{\partial \ln l(\theta, \mathbf{x})}{\partial \theta} = 0.$$

Asymptotically (i.e., when signal-to-noise ratio tends to infinity) the ML estimators are unbiased, normally distributed and their covariance matrix is equal to the inverse of the Fisher matrix Γ defined by

$$\Gamma_{ij} := E \left[\frac{\partial \ln l(\mathbf{x}; \theta)}{\partial \theta_i} \frac{\partial \ln l(\mathbf{x}; \theta)}{\partial \theta_j} \right], \quad i, j = 1, \dots, k,$$

where k is the number of parameters.

- The Cramèr-Rao bound states that for *unbiased* estimators the covariance matrix of the estimators $C \geq \Gamma^{-1}$ where the inequality $A \geq B$ means that the matrix $A - B$ is non negative definite.
- Thus calculating the Fisher matrix we can get lower bound on the variances of the estimators.
- There are more sophisticated bounds but there are cumbersome to calculate.

Gaussian case

Assume Gaussian, stationary, mean zero noise n and additive signal ($x = n + s$).

By Cameron-Martin formula the log likelihood ratio is given by

$$\ln \Lambda[x] = (x|s) - \frac{1}{2}(s|s),$$

scalar product $(\cdot | \cdot)$ is defined by (\Re denotes the real part, $S_h(f)$ is one-sided spectral density)

$$(x|y) := 4\Re \int_0^\infty \frac{\tilde{x}(f)\tilde{y}^*(f)}{S_h(f)} df.$$

Likelihood ratio test is equivalent to correlating the data x with the expected signal s and comparing the correlation to a threshold G_o .

- The correlation $G = (x|s)$ is called the **matched filter**.
- Using the relation

$$E[(n|s)(n|s)] = (s|s),$$

we have

$$E[G] = (s|s), \text{ Var}[G] = (s|s)$$

- Signal-to-noise ratio ρ is defined as

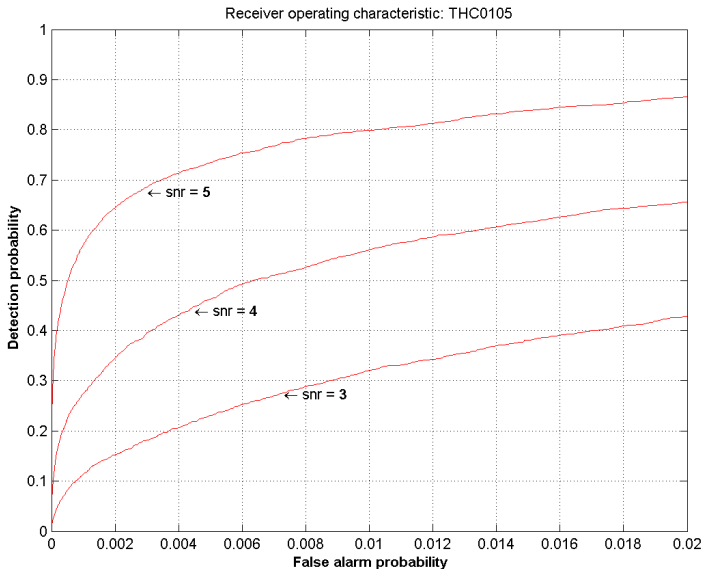
$$\rho = \sqrt{(s|s)}.$$

- Probabilities of false alarm and detection

$$P_F(G_o) = \text{erfc}\left(\frac{G_o}{\rho}\right), P_D(G_o) = 1 - \text{erfc}\left(\frac{G_o}{\rho} - \rho\right)$$

depend of one parameter - *signal-to-noise ratio*.

Receiver operating characteristic (ROC)



We introduce the generalized likelihood ratio

$$\mathcal{L} = \ln \left[\max_{\theta} \frac{p_1(\theta; x)}{p_0(x)} \right] = \ln \left[\frac{p_1(\hat{\theta}; x)}{p_0(x)} \right].$$

and we compare it to a threshold.

- Often GW signal s can be represented as a linear combination of certain time dependent functions of usually unknown m parameters η :

$$s(a, \eta) = \sum_{k=1}^n a_k h_k(t; \eta) = a \cdot h.$$

- Then

$$\ln \Lambda(a, \eta) = a^T \cdot N - \frac{1}{2} a^T \cdot M \cdot a.$$

where

$$N = (x|h), \quad M = (h|h).$$

Gaussian case and GW signals cont.

- ML estimators \hat{a} of amplitudes are found by solving the equations

$$\frac{\partial \ln l}{\partial a} = N = M \cdot a = 0$$

and are obtained in a closed analytic form

$$\hat{a} = M^{-1} \cdot N.$$

- Substituting \hat{a} for a we have

$$\mathcal{L}(\eta) = \frac{1}{2} N^T \cdot M^{-1} \cdot N.$$

This is the \mathcal{L} - statistic.

\mathcal{L} depends only on m parameters η (called *intrinsic*), the n amplitude parameters a (called *extrinsic*) are eliminated.

False alarm and detection probabilities

- $2 \times \mathcal{L}$ - statistic has χ^2 distribution with n degrees of freedom when signal is absent and non-central χ^2 distribution with non-centrality parameter ρ^2 when signal is present.
- In a search we evaluate \mathcal{L} over a grid in m -dimensional parameter space.
- To estimate false alarm probability we need to estimate number of *independent* trials N_c .
- False alarm probability is the probability $P_F^T(L_o)$ that \mathcal{L} exceeds the threshold L_o in one or more cells.

$$P_F^T(L_o) = 1 - [1 - P_F(L_o)]^{N_c},$$

where $P_F(L_o)$ is the probability of exceeding threshold L_o from χ^2 distribution.

Estimations of the number of independent trials

- Consider the autocovariance function \mathcal{C} of \mathcal{L}

$$\mathcal{C}(\eta, \eta') = E[(\mathcal{L}[x; \eta] - m(\eta))(\mathcal{L}[x; \eta'] - m(\eta'))].$$

- Define an *elementary cell* as the region with the boundary where maximum of \mathcal{C} falls by $1/2$.
- Perform the Taylor expansion of $\mathcal{C}(\eta, \eta')$ around the maximum and keep the first order terms to find that elementary cell is approximately defined as

$$\sum_{k=1}^m \sum_{l=1}^m \tilde{\Gamma}_{kl}(\eta) \Delta\eta_k \Delta\eta_l \leq \frac{1}{2}.$$

- Define volume of the elementary cell as

$$V_c = \frac{(\pi/2)^{m/2}}{(m/2)! \sqrt{\det \tilde{\Gamma}}}.$$

and number of N_c of the elementary cells as

$$N_c = \frac{V}{V_c},$$

where V is the hypervolume of the parameter space.

Signal-to-noise ratio and Fisher matrix

Signal-to-noise ratio:

$$\rho^2 = (s|s) = a^T \cdot M \cdot a.$$

Components of the Fisher matrix:

$$\Gamma_{ij} = (h_i|h_j).$$

The ML estimators of the amplitudes are unbiased

$$E[\hat{a}] = a$$

and minimum variance. Covariance matrix \mathcal{C}

$$\mathcal{C} = M^{-1}$$

is exactly equal to the inverse of the Fisher matrix. Such estimators are called *efficient*.

Two amplitude case

- Signal

$$s = A_s h_s(t; \eta) + A_c h_c(t; \eta).$$

- Amplitude estimators

$$\hat{A}_s = \frac{C(x|h_s) - M(x|h_c)}{D}, \quad \hat{A}_c = \frac{S(x|h_c) - M(x|h_s)}{D},$$

where $C := (h_c|h_c)$, $S := (h_s|h_s)$, $M := (h_s|h_c)$, and $D := SC - M^2$.

- Statistic

$$\mathcal{L}(x; \eta) = \frac{S(x|h_c)^2 + C(x|h_s)^2 - 2M(x|h_s)(x|h_c)}{2D}.$$

Noise whitening

To take into account coloured noise perform whitening by dividing the Fourier transform of the data by the square root of the spectral density of noise then taking the inverse Fourier transform. Thus the whitened data $x_w(t)$ is given by

$$x_w(t) = \left(\mathcal{F}^{-1} \left\{ \frac{\tilde{x}}{\sqrt{S_h}} \right\} \right) (t),$$

where \mathcal{F}^{-1} is the inverse Fourier transform.

Matched filter s to the signal need also to be whitened.

By Parseval's theorem the scalar $(\cdot | \cdot)$ is then given by

$$(x|s) = 2 \int_{-\infty}^{\infty} x_w(t) s_w(t) dt.$$

Monochromatic signal in white noise

$$s = A_o \cos(\omega_o t - \phi_o).$$

Rewrite the signal (1) as

$$s = A_c \cos(\omega_o t) + A_s \sin(\omega_o t),$$

where

$$A_c = A_o \cos \phi_o, \quad A_s = A_o \sin \phi_o.$$

For white noise $S_h(f) = S_o = \text{const.}$

$$(x|s) = \frac{2}{S_o} \int_{-\infty}^{\infty} x(t)s(t) dt.$$

Assuming finite observation time T_o but very much larger than the signal period:

$$C \simeq S \simeq \frac{T_o}{S_o}, \quad M \simeq 0.$$

Monochromatic signal in white noise cont.

\mathcal{L} statistic is given by

$$\mathcal{L} = \frac{2}{T_o S_o} \left[\left(\int_0^{T_o} x(t) \cos(\omega_o t) dt \right)^2 + \left(\int_0^{T_o} x(t) \sin(\omega_o t) dt \right)^2 \right] \quad (1)$$

or

$$\mathcal{L} = \frac{2}{T_o S_o} |\tilde{F}(\omega_o)|^2, \quad (2)$$

where $\tilde{F}(\omega_o) = \int_0^{T_o} x(t) e^{-i\omega_o t} dt$ is the Fourier transform.

Amplitude estimators are given by

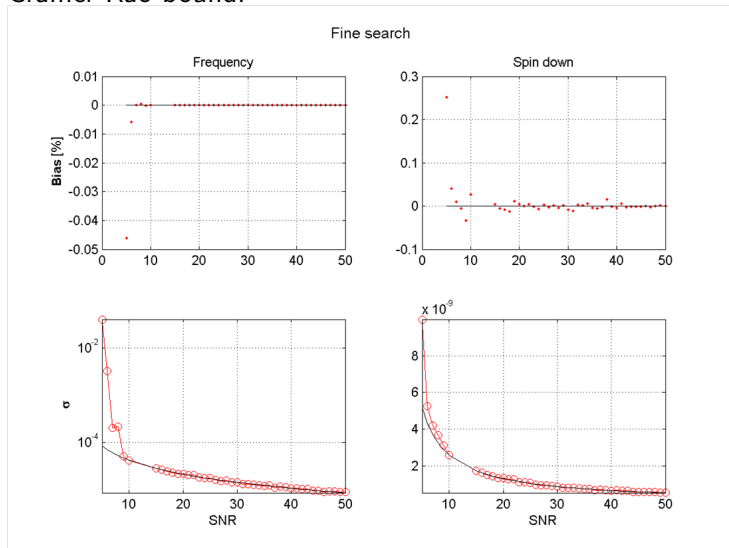
$$\hat{A}_s = \frac{2}{T_o} \int_0^{T_o} x(t) \cos(\omega_o t) dt, \quad \hat{A}_c = \frac{2}{T_o} \int_0^{T_o} x(t) \sin(\omega_o t) dt.$$

To test quality of parameter estimation we perform Monte Carlo simulation by injecting signals to Gaussian noise or to real data.

- Add signal to the data for a certain SNR.
- Calculate the statistic \mathcal{L} over a grid in the parameter space.
- Find maximum of \mathcal{L} over the grid.
- Do a fine search of the maximum with a maximization routine with initial parameters from the previous step.
- Repeat the procedure for an array of SNRs.

Parameter estimation - Monte Carlo simulations cont.

Threshold effect - below certain SNR large deviations from the Cramèr-Rao bound.



Deviations from the Cramèr-Rao bound

- Consider parameter space divided into elementary cells N_c . As a result of noise, maximum of the \mathcal{L} -statistic may fall in the cell where there is no signal. This is an *outlier*.
- Assume that outliers are uniformly distributed over the search interval of a parameter.

$$\sigma_{\text{out}}^2 = \frac{\Delta^2}{12},$$

where Δ is the length of the search interval for a given parameter.

- Total variance σ^2 of the estimator of a parameter is the weighted sum of the two errors

$$\sigma^2 = \sigma_{\text{out}}^2 q + \sigma_{\text{CR}}^2 (1 - q),$$

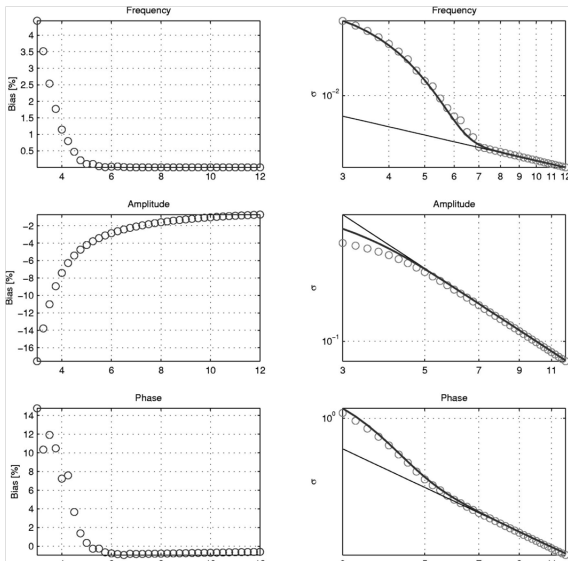
where q is the probability of the outlier σ_{CR} is the rms error from the Fisher matrix.

- The probability q can be approximated by the following formula:

$$q = 1 - \int_0^\infty p_1(\rho, \mathcal{L}) \left(\int_0^{\mathcal{L}} p_0(y) dy \right)^{N_c - 1} d\mathcal{L},$$

where p_0 and p_1 are probability density functions of respectively false alarm and detection.

Deviations from the Cramèr-Rao bound cont.



circles - MC simulation, line - CR bound, thick line - model

GW signal from a rotating star

The detector's response function s to a gravitational wave from a rotating neutron star has four amplitude parameters,

$$s(t) = \sum_{k=1}^4 A_k h_k(t; \eta).$$

The extrinsic parameters are the four constant amplitudes A_k given by

$$A_1 = h_{0+} \cos 2\psi \cos \phi_0 - h_{0\times} \sin 2\psi \sin \phi_0,$$

$$A_2 = h_{0+} \sin 2\psi \cos \phi_0 + h_{0\times} \cos 2\psi \sin \phi_0,$$

$$A_3 = -h_{0+} \cos 2\psi \sin \phi_0 - h_{0\times} \sin 2\psi \cos \phi_0,$$

$$A_4 = -h_{0+} \sin 2\psi \sin \phi_0 + h_{0\times} \cos 2\psi \cos \phi_0,$$

where ψ is the polarization angle and ϕ_0 is the constant phase and

$$h_{0+} = \frac{1}{2} h_0 (1 + \cos^2 \iota), \quad h_{0\times} = h_0 \cos \iota,$$

where ι is the angle between the angular momentum vector of the rotating neutron star and the line of sight.

GW signal from a rotating star cont.

For the simplest model of a rotating neutron star as a triaxial ellipsoid

$$h_0 = \frac{4\pi^2 G}{c^4} Q \frac{f_{GW}^2}{r},$$

where Q is the quadrupole moment of the star.

The four functions of time h_k depend on sky position and frequency:

$$h_1(t; \eta) = a(t; \delta, \alpha) \cos \phi(t; \mathbf{f}, \delta, \alpha),$$

$$h_2(t; \eta) = b(t; \delta, \alpha) \cos \phi(t; \mathbf{f}, \delta, \alpha),$$

$$h_3(t; \eta) = a(t; \delta, \alpha) \sin \phi(t; \mathbf{f}, \delta, \alpha),$$

$$h_4(t; \eta) = b(t; \delta, \alpha) \sin \phi(t; \mathbf{f}, \delta, \alpha).$$

We can approximate the phase $\phi(t)$ by

$$\phi(t; \eta) \cong 2\pi \sum_{k=0}^s f_k \frac{t^{k+1}}{(k+1)!} + \frac{2\pi}{c} \mathbf{n}_0(\delta, \alpha) \cdot \mathbf{r}_d(t) \sum_{k=0}^s f_k \frac{t^k}{k!}.$$

GW signal from a rotating star cont.

We can assume that over the bandwidth of the signal $S_h(f)$ is nearly constant and equal to S_o

$$\langle x|s \rangle = \frac{2}{S_o} \langle x(t) s(t) \rangle dt.$$

where

$$\langle \cdot \rangle = \int_0^{T_o} \cdot dt.$$

With approximation of a long observation time w.r.t. wave period we have

$$\langle h_1 h_3 \rangle \cong 0, \quad \langle h_1 h_4 \rangle \cong 0, \quad \langle h_2 h_3 \rangle \cong 0, \quad \langle h_2 h_4 \rangle \cong 0,$$

$$\langle h_1 h_1 \rangle \cong \langle h_3 h_3 \rangle \cong \frac{1}{2}A, \quad \langle h_2 h_2 \rangle \cong \langle h_4 h_4 \rangle \cong \frac{1}{2}B, \quad \langle h_1 h_2 \rangle \cong \langle h_3 h_4 \rangle \cong \frac{1}{2}C,$$

$$A = \langle a^2 \rangle, \quad B = \langle b^2 \rangle, \quad C = \langle ab \rangle.$$

GW signal from a rotating star cont.

Maximum likelihood estimators \hat{A}_k of the amplitudes A_k are given by

$$\hat{A}_1 \cong 2 \frac{B\langle xh_1 \rangle - C\langle xh_2 \rangle}{D}, \hat{A}_2 \cong 2 \frac{A\langle xh_2 \rangle - C\langle xh_1 \rangle}{D},$$
$$\hat{A}_3 \cong 2 \frac{B\langle xh_3 \rangle - C\langle xh_4 \rangle}{D}, \hat{A}_4 \cong 2 \frac{A\langle xh_4 \rangle - C\langle xh_3 \rangle}{D},$$

where $D = AB - C^2$.

With the above approximations the statistic known as \mathcal{F} -statistic takes the following form.

$$\mathcal{F}[x; \eta] \cong \frac{2}{S_o} \frac{B|F_a|^2 + A|F_b|^2 - 2C\Re(F_a F_b^*)}{D}.$$

where F_a and F_b are given by:

$$F_a = \langle x(t)a(t) \exp[-i\phi(t)] \rangle,$$

$$F_b = \langle x(t)b(t) \exp[-i\phi(t)] \rangle.$$

$2 \times \mathcal{F}$ -statistic has a χ^2 distribution with 4 degrees of freedom.

We want to evaluate the \mathcal{F} - statistic efficiently using the FFT.
The phase $\phi(t)$ can be written as

$$\phi(t) = \omega_0[t + \phi_m(t)] + \phi_s(t),$$

where

$$\begin{aligned}\phi_m(t) &:= \frac{\mathbf{n}_0 \cdot \mathbf{r}_d(t)}{c}, \\ \phi_s(t) &:= \sum_{k=1}^s \omega_k \frac{t^{k+1}}{(k+1)!} + \frac{\mathbf{n}_0 \cdot \mathbf{r}_d(t)}{c} \sum_{k=1}^s \omega_k \frac{t^k}{k!},\end{aligned}$$

The functions $\phi_m(t)$ and $\phi_s(t)$ do not depend on the frequency ω_0 .

$$F_a = \int_0^{T_0} x(t) a(t) e^{-i\phi_s(t)} \exp \{ -i\omega_0[t + \phi_m(t)] \} dt.$$

We introduce a new time variable t_b , so called *barycentric time* ,

$$t_b := t + \phi_m(t; \mathbf{n}_0).$$

In the new time coordinate the integral is approximately given by

$$F_a \cong \int_0^{T_0} x[t(t_b)]a[t(t_b)]e^{-i\phi_s[t(t_b)]}e^{-i\omega_0 t_b} dt_b.$$

This is a Fourier transform that can be calculated efficiently with the FFT.

Conclusions

- Generalized likelihood ratio test, leading to \mathcal{F} -statistic, results in an efficient method to detect the signal and estimate its parameters.
- Main problem is that for searches of a large parameter space (like all-sky searches) it becomes computationally intractable. Usually coherent \mathcal{F} -statistic search becomes a part of a hierarchical search.
- Other data analysis algorithms are also widely used.
- For accurate parameter estimation usually *Bayesian estimation* is adopted.

see [Keith Riles, Living Reviews in Relativity 26, 3 \(2023\)](#) for a comprehensive review of CW analysis methods.

Effective banks of templates in detection of almost monochromatic gravitational waves

- [P. Jaranowski and A. Pisanski](#):
P. Jaranowski and A. Pisanski, arXiv:2302.05204 [gr-qc].
A. Pisanski, P. Jaranowski, and M. Pietka, Phys. Rev. D 83, 043001 (2011).
A. Pisanski and P. Jaranowski, Class. Quant. Grav. 32, 145014 (2015).
- [with early contributions by A. Królak and M. Pietka](#):
P. Astone, K. M. Borkowski, P. Jaranowski, M. Pietka, and A. Królak, Phys. Rev. D 82, 022005 (2010).
- [Template placement as covering problem, R. Prix](#):
R. Prix, Class. Quant. Grav. 24:S481-S490 (2007).
- [Fisher matrix as a metric for template placement, B. Owen](#):
B. Owen, Phys. Rev. D 53, 6749 (1996).

Linear model of the signal (2)

Simplified linear model of the signal for **all-sky** searches
(with s spindowns included):

$$h(t; h_0, \Psi_0, \xi) = h_0 \sin(\Phi(t; \xi) + \Phi_0), \quad (5a)$$

$$\Phi(t; \xi) = \sum_{k=0}^s \omega_k \left(\frac{t}{T_0}\right)^{k+1} + \alpha_1 \mu_1(t) + \alpha_2 \mu_2(t), \quad (5b)$$

the phase $\Phi(t; \xi)$ is a **linear function of the parameters ξ** ,

$$\xi := (\omega_0, \omega_1, \dots, \omega_s, \alpha_1, \alpha_2), \quad (6a)$$

$$\alpha_1 := 2\pi f_0 (\sin \alpha \cos \delta \cos \varepsilon + \sin \delta \sin \varepsilon), \quad \alpha_2 := 2\pi f_0 \cos \alpha \cos \delta. \quad (6b)$$

$\mu_1(t)$, $\mu_2(t)$ —ephemerides (known functions of time describing motion of the detector relatively to the barycentre of the Solar System), α is the right ascension and δ is the declination of the GW source, ε is the obliquity of the ecliptic.

Optimal covering of the parameter space

The boundary of the region defined by the inequality $C_0(\xi, \xi') = C_0(\tau) \geq C_{\min}$ is given by the isoheight of the function $C_0(\tau)$,

$$C_0(\tau) = C_{\min} \iff \sum_{i,j=1}^d \tilde{r}_{ij} \tau_i \tau_j = 1 - C_{\min}. \quad (16)$$

This is the equation of:

$d = 2$ — ellipse,

$d = 3$ — ellipsoid,

$d \geq 4$ — hyperellipsoid.

We want to find **optimal covering of \mathbb{R}^d by identical hyperellipsoids** [defined by Eq. (16)] (covering is defined by a grid of points being the centers of hyperellipsoids). Optimal covering consists of **possibly smallest number of hyperellipsoids**, we look for grids with the **smallest possible covering thicknesses**.

(The covering thickness is defined as the average number of hyperellipsoids that contain a point in the space.)

We impose two additional constraints on grids:

- 1 we want to speed up computation of the \mathcal{F} -statistic for all grid nodes by employing the FFT algorithm;
- 2 in all-sky searches we want efficiently resample of data to barycentric time—resampling is needed once per sky position for all the spindown values.

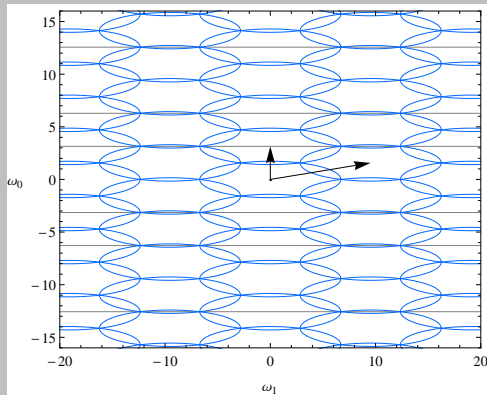
$d = 2$ case: directed searches with one spindown included

Two intrinsic parameters: (ω_0, ω_1) .

The possibility of using the FFT algorithm requires that the distance between the ω_0 -coordinates of the centers of ellipses is determined by the resolution of the FFT:

$$\Delta\omega_0 = 2\pi N/N_{\text{FFT}}. \quad (17)$$

The plot is made for $N = 344656$
(two sidereal days with $\Delta t_{\text{sample}} = 0.5$ s),
 $N_{\text{FFT}} = 2^{20}$ (then $\Delta\omega_0 \cong 2.06522$),
and $C_{\text{min}} = 0.75$.



Construction of the constrained optimal grids

- Replacing the problem of finding the optimal **covering of space by identical hyperellipsoids** by the problem of finding the optimal **covering of space by identical hyperspheres of unit radius**; this is done by finding a linear transformation $\tau = F\tau'$ which transforms the hyperellipsoid defined by the constant value C_{\min} of the autocovariance function into the hypersphere of unit radius. Also the constraints are translated into the new auxiliary parameter space. (F is found as the Cholesky decomposition of the matrix Γ .)
- **Construction of constrained grids in the auxiliary space**: the grids are simple deformations of the d -dimensional lattice coverings A_d^* of space \mathbb{R}^d by hyperspheres of unit radius.
The A_d^* lattices are the thinnest known lattices in dimensions $2 \leq d \leq 5$, and they are close to the thinnest known lattices in dimensions $6 \leq d \leq 15$. A_2^* is the hexagonal lattice, A_3^* is the body-centered cubic (bcc) lattice.
- **Transformation of the found grids into the original parameter space**, using the inverse linear transformation F^{-1} .