LASER INTERFEROMETER GRAVITATIONAL WAVE OBSERVATORY
- LIGO -
CALIFORNIA INSTITUTE OF TECHNOLOGY
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Technical Note	**LIGO-T040125-00-Z**	Date: 2004/06/15 20:08:09 UTC

Cross-correlation of windowed, discretely-sampled data

J T. Whelan
Loyola University New Orleans

Distribution of this document:
All

California Institute of Technology
LIGO Project, MS 18-34
Pasadena, CA 91125
Phone (626) 395-2129
Fax (626) 304-9834
E-mail: info@ligo.caltech.edu

Massachusetts Institute of Technology
LIGO Project, Room NW17-161
Cambridge, MA 02139
Phone (617) 253-4824
Fax (617) 253-7014
E-mail: info@ligo.mit.edu

WWW: http://www.ligo.caltech.edu/

Abstract

The cross-correlation method used in stochastic gravitational wave searches is generalized to include non-trivial windowing of the data. Exact expressions are derived for the expected mean and variance of the statistic, along with approximations which are valid when the windows chosen are effective at controlling the behavior of the data in question.

1 Motivation

The primary technique to search for a stochastic background of gravitational waves in the output of a pair of gravitational wave detectors is to construct an optimally-filtered cross-correlation statistic.[1] The theoretical analysis[2, 3] describing the behavior of such a statistic has been carried out in a continuous, long-duration approximation. More recent work[4] has considered finite lengths of discretely sampled data, but has implicitly assumed rectangular windowing. If the data in question have a large dynamic range, the leakage of power across the frequency domain arising from the rectangular window can obscure the data at some frequencies. The present work considers the behavior of a cross-correlation statistic constructed from a pair of windowed sections of discretely-sampled data, to allow for the case where non-trivial windowing may be necessary.

2 Continuous Approximation

Consider the underlying data to be two continuous time series $g_1(t)$ and $g_2(t)$ (which may, for example, be the gravitational-wave strain associated with two detectors), and present a summary of the results obtained in [2, 3], expressed in slightly more general notation.

Given data from a time t_0 to a time $t_0 + T$, the cross-correlation statistic associated with a time-domain filter $Q(t_1 - t_2)$ is

$$Y_Q(t_0, t_0 + T) = \int_{t_0}^{t_0+T} dt_1 \int_{t_0}^{t_0+T} dt_2 \, g_1(t_1)^* \, Q(t_1 - t_2) \, g_2(t_1) \tag{2.1}$$

Defining the continuous Fourier transform

$$\widetilde{g}_{1,2}(f) = \int_{-\infty}^{\infty} dt \, e^{-i2\pi f(t-t_0)} \, g_{1,2}(t) \tag{2.2}$$

one can describe the statistical properties of the Gaussian random variables $g_{1,2}(t)$ in terms of expectation values

$$\langle \widetilde{g}_i(f)^* \widetilde{g}_j(f') \rangle = \delta(f - f') C_{ij}(f) \ . \tag{2.3}$$

The auto-correlation spectra $C_{11}(f)$ and $C_{22}(f)$ can be written in terms of the one-sided power spectral densities $P_1(f)$ and $P_2(f)$ as

$$C_{11}(f) = \frac{1}{2} P_1(|f|) \tag{2.4a}$$

$$C_{22}(f) = \frac{1}{2} P_2(|f|) \ . \tag{2.4b}$$

In [3], the expected mean value of the cross-correlation statistic is approximated as

$$\mu := \langle Y_Q(t_0, t_0 + T) \rangle \approx T \int_{-\infty}^{\infty} df\, \widetilde{Q}(f)\, C_{12}(f) \tag{2.5}$$

while the variance can be written, subject to the assumption that the auto-correlation spectra are much larger than the cross-correlation spectrum $C_{12}(f)$, as

$$\sigma^2 := \langle (Y_Q(t_0, t_0 + T) - \mu)^2 \rangle \approx \frac{T}{4} \int_{-\infty}^{\infty} df \left| \widetilde{Q}(f) \right|^2 P_1(|f|)\, P_2(|f|) \ . \tag{2.6}$$

3 Discretely Sampled Data

In a practical data analysis situation, the data are not continuous, but typically sampled at some constant frequency $(\delta t)^{-1}$, producing an N-point discrete time series:

$$g_{1,2}[j] = g_{1,2}(t_0 + j\,\delta t) \qquad j = 0, \ldots, N-1 \tag{3.1}$$

In this section, we deduce the properties of cross-correlations between these discrete time series from the average behavior of the underlying continuous quantities (*e.g.*, (2.3)).

First, we define a general cross-correlation statistic as

$$Y = \sum_{j=0}^{N-1} \delta t \sum_{k=0}^{N-1} \delta t\, w_1[j]\, g_1[j]^*\, Q[j-k]\, w_2[k]\, g_2[k] \tag{3.2}$$

(The factors of δt are to facilitate comparison between this expression and (2.1).) The optimal filter $Q[j-k]$ depends only on the difference between the two indices, consistent with the assumption that we're considering *stationary* random processes, but the introduction of the two N-point window functions $w_{1,2}[j]$ (assumed to be real) allows us to control edge effects by smoothing out the onset and ending of the analyzed data.

Since the sums over j and k in (3.2) both range from 0 to $N-1$, the argument of $Q[j-k]$ ranges from $-(N-1)$ to $N-1$, so a discrete Fourier transform (DFT) of Q will need to include at least $2N-1$ points. Since it is often more convenient to work with a $2N$-point DFT than a $2N-1$-point one (e.g., if N is a power of two or a product of small primes), we will in general zero-pad $Q[m]$ out to $M \geq 2N-1$ points, with the "extra" values (i.e., those with $N-1 < m \leq M-1$) defined by

$$Q[m] = \begin{cases} 0 & N-1 < m < M-(N-1) \\ Q[m-M] & M-(N-1) \leq m < M \end{cases} \ , \tag{3.3}$$

before defining the discrete Fourier transform

$$\widehat{Q}[\ell] = \sum_{m=0}^{M-1} e^{-i2\pi m\ell/M}\, Q[m] = \sum_{m=-(N-1)}^{N-1} e^{-i2\pi m\ell/M}\, Q[m]\ . \tag{3.4}$$

We can transform (3.2) into the frequency domain using the inverses of (2.2) and (3.4):

$$g_{1,2}(t) = \int_{-\infty}^{\infty} df\, e^{i2\pi f(t-t_0)}\, \widetilde{g}_{1,2}(f) \tag{3.5}$$

$$Q[m] = \frac{1}{M}\sum_{\ell=0}^{M-1} e^{i2\pi m\ell/M}\, \widehat{Q}[\ell]\ ; \tag{3.6}$$

the result is

$$Y = \frac{1}{M}\sum_{\ell=0}^{M-1} (\delta t)^2\, \widehat{Q}[\ell] \left(\int_{-\infty}^{\infty} df_1\, \widehat{W}_1([f_\ell - f_1]T)\, \widetilde{g}_1(f_1)\right)^* \int_{-\infty}^{\infty} df_2\, \widehat{W}_2([f_\ell - f_2]T)\, \widetilde{g}_2(f_2) \tag{3.7}$$

where $\delta f = (M\,\delta t)^{-1}$, $f_\ell = \ell\,\delta f$, and the transformed window

$$\widehat{W}_{1,2}(x) = \sum_{j=0}^{N-1} e^{-i2\pi xj/N}\, w_{1,2}[j] \tag{3.8}$$

is equivalent to an N-point discrete Fourier transform, but not limited to integer arguments. Note that by construction $\widehat{W}_{1,2}(x)$ is periodic with period N: $\widehat{W}_{1,2}(x+N) = \widehat{W}_{1,2}(x)$.

Using (2.3), we can find exact expressions for the expected mean and variance of Y. The mean is

$$\mu = \langle Y\rangle = \frac{1}{M}\sum_{\ell=0}^{M-1} (\delta t)^2\, \widehat{Q}[\ell] \int_{-\infty}^{\infty} df\, \widehat{W}_1([f_\ell - f]T)^*\, \widehat{W}_2([f_\ell - f]T)\, C_{12}(f) \tag{3.9}$$

while the variance, subject to the simplifying assumption that $\sigma \gg \mu$ and $P_1(f), P_2(f) \gg C_{12}(f)$, is

$$\begin{aligned}\sigma^2 \approx \langle Y^2\rangle = \frac{1}{4}\frac{1}{M^2}\sum_{\ell=0}^{M-1}\sum_{m=0}^{M-1}\int_{-\infty}^{\infty} df_1 \int_{-\infty}^{\infty} df_2\\ (\delta t)^2\, \widehat{Q}[\ell]^*\, \widehat{W}_1([f_\ell - f_1]T)\, P_1(|f_1|)\, \widehat{W}_1([f_m - f_1]T)^*\, \widehat{W}_2([f_m - f_2]T)\\ P_2(|f_2|)\, \widehat{W}_2([f_\ell - f_2]T)^*\, (\delta t)^2\, \widehat{Q}[m]\end{aligned} \tag{3.10}$$

The expressions (3.9) and (3.10) can be simplified by defining

$$\mathcal{C}_{12}[\ell] = \int_{-\infty}^{\infty} df\, \widehat{W}_1([f_\ell - f]T)^*\, \widehat{W}_2([f_\ell - f]T)\, C_{12}(f) \tag{3.11}$$

and

$$\begin{aligned}\mathcal{K}[\ell, m] = \frac{1}{4}\int_{-\infty}^{\infty} df_1 \int_{-\infty}^{\infty} df_2\, P_1(|f_1|)\, P_2(|f_2|)\\ \widehat{W}_1([f_\ell - f_1]T)\, \widehat{W}_1([f_m - f_1]T)^*\, \widehat{W}_2([f_m - f_2]T)\, \widehat{W}_2([f_m - f_2]T)^*\end{aligned} \tag{3.12}$$

so that the expressions for the mean and variance of Y are written in the form of matrix multiplication:

$$\mu = \frac{1}{M} \sum_{\ell=0}^{M} (\delta t)^2 \, \widehat{Q}[\ell] \, \mathcal{C}_{12}[\ell] \tag{3.13a}$$

$$\sigma^2 = \frac{1}{M^2} \sum_{\ell=0}^{M} \sum_{m=0}^{M} (\delta t)^2 \, \widehat{Q}[\ell]^* \mathcal{K}[\ell, m] (\delta t)^2 \, \widehat{Q}[m] \tag{3.13b}$$

Given these expressions, it is straightforward to show (analogous to the construction in [4]) that the ratio μ/σ is maximized when

$$\widehat{Q}[\ell] \propto \sum_{m=0}^{M-1} \mathcal{K}^{-1}[\ell, m] \mathcal{C}_{12}[m] \tag{3.14}$$

where $\mathcal{K}^{-1}[\ell, m]$ is the matrix inverse of $\mathcal{K}[\ell, m]$:

$$\sum_{m=0}^{M-1} \mathcal{K}^{-1}[\ell, m] \mathcal{K}[m, \ell'] = \delta_{\ell\ell'} \tag{3.15}$$

4 Approximations in the Presence of Effective Windowing

Since inverting the matrix $\mathcal{K}[\ell, m]$ is a computationally-intensive operation, it is ordinarily preferable to work in a regime in which the exact expressions (3.9) and (3.10) reduce to discrete approximations of their continuous counterparts (2.5) and (2.6). Reaching such a regime is exactly the purpose of choosing "good" windowing functions $w_{1,2}[j]$.

Before describing the effective-windowing condition, we note that if (as is the case for any sensible discrete-sampling situation) the data $g_{1,2}(t)$ have had an anti-aliasing filter applied to them so that $\widetilde{g}_{1,2}(f)$ is negligible whenever $|f| \geq 1/(2\,\delta t)$, we can change the limits of all our frequency integrals from $(-\infty, \infty)$ to $(-\frac{1}{2\,\delta t}, \frac{1}{2\,\delta t})$.

We would like to define an effective windowing function as one whose transformed window is sufficiently sharply peaked about zero argument to overcome the dynamic range of the data in question. We would then be free to treat transformed windows like delta functions in the sense that their arguments can be replaced with zero when they appear elsewhere in the same integral. However, we note that the periodicity of $\widehat{W}_{1,2}(x)$ means that the best we can do is insist that $\widehat{W}_{1,2}(x)$ is negligible unless $x \approx 0 \mod N$. In expressions like

$$\mathcal{C}_{12}[\ell] = \int_{-1/(2\,\delta t)}^{1/(2\,\delta t)} df \, \widehat{W}_1([f_\ell - f]T)^* \, \widehat{W}_2([f_\ell - f]T) \, C_{12}(f) \tag{4.1}$$

This means $f \approx f_\ell \mod N/T = \ell\,\delta f \mod 1/(\delta t)$. But only one such frequency will lie in the range $[-\frac{1}{2\delta t}, \frac{1}{2\delta t}$, namely

$$\mathrm{f}_\ell := \begin{cases} \ell\,\delta f & \ell < M/2 \\ \ell\,\delta f - 1/\delta t & \ell > M/2 \end{cases} \tag{4.2}$$

($\ell = M/2$ corresponds to the Nyquist frequency, at which the anti-aliasing filter should suppress the frequency-domain data anyway). The windowing thus allows us to replace $C_{12}(f)$ with $C_{12}(\mathrm{f}_\ell)$ in (3.9) and obtain

$$\mu \approx \frac{1}{M}\sum_{\ell=0}^{M-1}(\delta t)^2\,\widehat{Q}[\ell]\,C_{12}(\mathrm{f}_\ell)\int_{-1/(2\,\delta t)}^{1/(2\,\delta t)} df\,\widehat{W}_1([f_\ell - f]T)^*\,\widehat{W}_2([f_\ell - f]T) \tag{4.3}$$

The integral can be evaluated [using (3.8)] as

$$\begin{aligned}\int_{-1/(2\,\delta t)}^{1/(2\,\delta t)} df\,\widehat{W}_1([f_\ell - f]T)^*\,\widehat{W}_2([f_\ell - f]T) &= \sum_{j=0}^{N-1}\sum_{k=0}^{N-1} w_1[j]w_2[k]\int_{-1/(2\,\delta t)}^{1/(2\,\delta t)} df\; e^{i2\pi(f_\ell - f)(j-k)\delta t} \\ &= \frac{1}{\delta t}\sum_{j=0}^{N-1}\sum_{k=0}^{N-1}\delta_{jk}w_1[j]w_2[k] = \frac{1}{\delta t}N\overline{w_1w_2}\end{aligned} \tag{4.4}$$

where the overbar indicates an average over the N points of the window:

$$\overline{w_1w_2} = \frac{1}{N}\sum_{j=0}^{N-1} w_1[j]w_2[k] \tag{4.5}$$

This then tells us

$$\mu \approx \overline{w_1w_2}\,T\sum_{\ell=0}^{M-1}\delta f\,(\delta t\,\widehat{Q}[\ell])\,C_{12}(\mathrm{f}_\ell) \tag{4.6}$$

where we have used again the definition $\delta f = 1/(M\,\delta t)$.

We can identify (4.6) (up to the factor of $\overline{w_1w_2}$) as a discrete approximation to (2.5) if we note that (3.4) relates the discrete and continuous Fourier transforms according to

$$\delta t\,\widehat{Q}[\ell] \approx \widetilde{Q}(\mathrm{f}_\ell) \tag{4.7}$$

Similarly, applying the effective windowing assumptions to (3.10) allows us to replace f_1 and f_2 with either f_ℓ or f_m (and hence m with ℓ) and obtain

$$\begin{aligned}\sigma^2 \approx \frac{(\delta t)^4}{4M^2}&\sum_{\ell=0}^{M-1}\sum_{m=0}^{M-1} P_1(|\mathrm{f}_\ell|)\,P_1(|\mathrm{f}_\ell|)\,\widehat{Q}[\ell]^*\,\widehat{Q}[\ell] \\ &\int_{-1/(2\,\delta t)}^{1/(2\,\delta t)} df_1\,\widehat{W}_1([f_\ell - f_1]T)\,\widehat{W}_1([f_m - f_1]T)^*\int_{-1/(2\,\delta t)}^{1/(2\,\delta t)} df_2\,\widehat{W}_2([f_m - f_2]T)\,\widehat{W}_2([f_\ell - f_2]T)^*\end{aligned} \tag{4.8}$$

We can now evaluate the integrals as

$$\int_{-1/(2\,\delta t)}^{1/(2\,\delta t)} df\,\widehat{W}_1([f_\ell - f_1]T)\,\widehat{W}_1([f_m - f_1]T)^* = \sum_{j=0}^{N-1}\sum_{k=0}^{N-1}\int_{-1/(2\,\delta t)}^{1/(2\,\delta t)} df\, w_1[j]w_1[k]e^{i2\pi[jf_\ell - kf_m + (k-j)f_1]\delta t}$$
$$= \frac{1}{\delta t}\sum_{j=0}^{N-1}\sum_{k=0}^{N-1}\delta_{jk}w_1[j]w_1[k]e^{i2\pi(jf_\ell - kf_m)\delta t} = \frac{1}{\delta t}\sum_{j=0}^{N-1}(w_1[j])^2 e^{i2\pi j(f_\ell - f_m)\delta t} \tag{4.9}$$

and similarly

$$\int_{-1/(2\,\delta t)}^{1/(2\,\delta t)} df_2\,\widehat{W}_2([f_m - f_2]T)\,\widehat{W}_2([f_\ell - f_2]T)^* = \frac{1}{\delta t}\sum_{k=0}^{N-1}(w_2[k])^2 e^{i2\pi k(f_m - f_\ell)\delta t} \tag{4.10}$$

so

$$\sigma^2 \approx \frac{(\delta t)^2}{4M}\sum_{\ell=0}^{M-1} P_1(|\mathrm{f}_\ell|)\,P_2(|\mathrm{f}_\ell|)\,\widehat{Q}[\ell]^*\,\widehat{Q}[\ell]$$
$$\sum_{j=0}^{N-1}\sum_{k=0}^{N-1}(w_1[j])^2(w_2[k])^2 e^{i2\pi(j-k)f_\ell\delta t}\underbrace{\frac{1}{M}\sum_{m=0}^{M-1}e^{i2\pi f_m(k-j)\delta t}}_{\delta_{jk}} \tag{4.11}$$
$$= \frac{T}{4}\,\overline{w_1^2 w_2^2}\sum_{\ell=0}^{M-1}\delta f\,\left|(\delta t\,\widehat{Q}[\ell])\right|^2 P_1(|\mathrm{f}_\ell|)\,P_2(|\mathrm{f}_\ell|)$$

Again, we note that this differs from a discrete approximation to (2.6) only by the extra factor

$$\overline{w_1^2 w_2^2} = \frac{1}{N}\sum_{j=0}^{N-1} w_1[j]^2 w_2[k]^2 \tag{4.12}$$

Note that the windowing replaces the time T in the usual signal-to-noise expressions with

$$\frac{(\overline{w_1 w_2})^2}{\overline{w_1^2 w_2^2}}\,T \tag{4.13}$$

For a pair of Hann windows, the ratio is 18/35.

Acknowledgments

This work was supported by the National Science Foundation under grants PHY-9981795 (UTB) and PHY-0300609 (Loyola), and by NASA contract JPL1219731 (UTB).

References

[1] Christensen N 1992 *Phys. Rev.* **D46** 5250.

[2] Allen B 1997 in *Proceedings of the Les Houches School on Astrophysical Sources of Gravitational Waves*, eds Marck J A and Lasota J P, Cambridge, 373; e-Print: gr-qc/9604033.

[3] Allen B and Romano J D 1999 *Phys. Rev.* **D59** 102001; e-Print: gr-qc/9710117.

[4] Finn L S and Romano J D 2001 *Preprint* UTBRG-2001-04.