Detecting and Characterizing Nonlinear Signal Components in VIRGO Data Channels

Ryan Goetz
Illinois Wesleyan University
Supervisor, Giancarlo Cella
Università di Pisa
INFN s. di Pisa

1 Motivation

The goal of the VIRGO Interferometer in Cascina, Italy is direct detection of gravitational radiation predicted by general relativity. Due to their characteristically weak interactions, observing gravitational waves requires an incredible level of instrumental sensitivity, and a relatively long period of integration. In this spirit, the apparatus has a design sensitivity of $3 \times 10^{-21}\text{Hz}^{-\frac{1}{2}}$ at 10Hz. As an instrument becomes increasingly sensitive, however, the collected data becomes more susceptible to meaningless noise signals that increase the necessary integration time or compromise the experiment altogether. It is imperative, then, that sources of error are fully understood and accounted for in analysis.

In an effort to identify noise, a number of detectors and probes have been integrated into the apparatus's control system. These give valuable measures of seismic activity, mechanical vibrations, thermal fluctuations, electrical, acoustic, and beam frequency noise, among others, at different locations across the entire experiment. Ideally, this data could identify the effects of noise in the main signal (the dark fringe). To do so, it is necessary to understand the coupling of channels to one another.

In many cases, certain channels might be linearly coupled to the dark fringe. This scenario is relatively simple to identify and understand. However, in more complicated cases we might observe nonlinear coupling that evades conventional techniques for analysis. These nonlinearities are the primary focus of our work; our intent being not only to identify them, but also to quantify them in as meaningful a manner as possible. It is expected that as sensitivity of the instrument increases, the vulnerability to nonlinear effects will increase. Physical sources of nonlinear signal components might be processes such as modulations, up conversions, or down conversions. The ultimate hope of this research is to

improve the understanding of how noise will manifest itself in the main signal and provide a foundation for increased resolution of the VIRGO experiment.

The interferometer is undergoing its second science run in the summer of 2009, called the VIRGO Science Run 2 (VSR2). This run will provide a great deal of data from both the main channel of interest, the dark fringe, as well as from auxiliary channels that are important for implementing the control system, and running data analysis.

In our analysis, we assume the existence of a fundamental set of independent, stochastic components from which all collected data can be constructed. For simplicity, it is often assumed that these components are Gaussian and zero-mean. This article assumes a cursory knowledge of probability and statistics. For a thorough address on these topics, the reader is referred to [1].

2 Defining Signal Nonlinearity

As of yet the use of the word "nonlinearity" has been ambiguous. Let us suppose, in accordance with the assumptions of our research, that all signals recorded in the VIRGO experiment can be expressed in terms of N fundamental (linear stochastic) components $\{p_i : 1 \leq i \leq N\}$. Then, in this article, a signal, s, is deemed linear if it can be written as:

$$s = \sum_{i=1}^{N} k_i p_i \tag{1}$$

, where $\{k_i : 1 \leq i \leq N\}$ is a set of scalars. Succinctly, a linear signal is defined as any signal that is a linear combination of a fundamental set of signal components. Generally, this kind of signal corresponds to relatively simple physical situations, and in practice we do not expect to see perfect linearity very often. There are infinitely many ways for a signal to be nonlinear, which we represent (maintaining reasonable generality) with the notation:

$$s = \sum_{i=1}^{N} \widehat{g_i}(p_1, \ \ldots \ , p_N) \tag{2}$$

, where $\{\widehat{g_i} : 1 \leq i \leq N\}$ is a set of operators with at least one nonlinear mapping. In general, $\widehat{g_i}$ can be arbitrarily complicated. One practical simplification might be the model:

$$s = \sum_{i=1}^{N} h_i p_i (\sum_{k=1}^{N} l_{i,k} p_k) \tag{3}$$

, where $\{h_i : 1 \leq i \leq N\}$ is a set of arbitrary scalar values and $\{l_{i,k} : 1 \leq i \leq N, 1 \leq k \leq N\}$ is a set of scalar values that can be determined by a function.

For our purposes, the model in (3) is acceptable, and we do not pursue the more general cases that are allowed for in (2). This constriction still leaves a great deal of relevant physical processes available, and, importantly, it affords us a method for resolving and quantifying nonlinearity. Note that if we let:

$$l_{i,k}(t) = \begin{cases} p_k(t)^{-1} & \text{, if } k = a \\ 0 & \text{, otherwise} \end{cases} \tag{4}$$

for any $1 \leq a \leq N$, we can recover a non-trivial linear model. (We do not worry about indeterminate forms since the probabilty that $p_k(t) = 0$ is zero). To the same end, we might let:

$$l_{i,k}(t) = \frac{1}{N p_k(t)} \tag{5}$$

. That this model (3) can be analytically reduced to a simple case (1) suggests that it might share certain behaviors with the linear model. In practice, we expect to find that many signals correspond to a model with:

$$l_{i,k} = \begin{cases} \delta & \text{, if } k = a_i \\ 0 & \text{, otherwise} \end{cases} \tag{6}$$

, for some $1 \leq a_i \leq N$, where δ is a scalar value. Note that we expect no time dependence for $l_{i,k}$; a condition necessary for our analysis. Without time-independence of $l_{i,k}$ the issue of quantifying nonlinearity becomes increasingly complicated, and we do not propose a technique for doing so in this article.

3 Second-Order Statistical Analysis

3.1 Second-Order Testing

Second-order analysis is a staple of signal analysis. Every power spectrum, in fact, is a product of second-order analysis. For now, we will restrict our analysis to the time domain. In truth, any linear second-order correlation cannot determine nonlinear effects, and so we examine a direct generalization of second-order testing. There are several tests which are direct generalizations of second-order tests that we can perform to identify nonlinearity in a signal set. Let us define two signals in the following manner:

$$x \equiv a + b \tag{7}$$

$$y \equiv \alpha a c + c \tag{8}$$

, where a, b, and c are independent, stochastic, Gaussian signals each with zero mean, and α is a nonzero constant. In practice, we cannot record continuous measures with our instruments, and so we are forced to approximate the scenario with a discrete sampling of the signals:

$$x_n \equiv a_n + b_n \tag{9}$$

$$y_n \equiv \alpha a_n c_n + c_n \tag{10}$$

, where the subscripts are intended to denote a finite set of data. This signal set could be interpreted as a scenario in which a component of x, a, is modulated by the process c, where both b and c could be noise unique to their respective channels. We expect this process to have a non-Gaussian nature. If we evaluate the expectation value of $x_n y_n^2$ we find:

$$E[x_n y_n^2] = E[a_n c_n^2 + 2\alpha a_n^2 c_n^2 + \alpha^2 a_n^3 c_n^2 + b_n c_n^2 + 2\alpha b_n c_n^2 + \alpha^2 a_n^2 b_n c_n^2] \tag{11}$$

. Since the processes a, b, and c are Gaussian, we can ignore the first, third, fourth, and sixth terms, giving:

$$\begin{aligned} E[x_n y_n^2] &= E[2\alpha a_n^2 c_n^2 + 2\alpha a_n b_n c_n^2] \\ &= 2\alpha(E[a_n^2 c_n^2] + E[a_n b_n c_n^2]) \\ &= 2\alpha E[a_n^2] E[c_n^2] \end{aligned} \tag{12}$$

, which is guaranteed to be nonzero. In practice, we can approximate the expectation value by taking a mean, and so:

$$\overline{x_n y_n^2} \approx 2\alpha \bar{a_n^2} \bar{c_n^2} \tag{13}$$

. For simplicity let

$$\overline{x_n y_n^2} \equiv M_1 \tag{14}$$

$$\overline{a_n^2} \equiv A \tag{15}$$

$$\overline{c_n^2} \equiv C \tag{16}$$

. Rearranging (12) to get an explicit expression for α we have:

$$\alpha \approx \frac{M_1}{2AC} \tag{17}$$

. It is worth noting that M_1 can be directly measured from the data signals, whereas A and C must be estimated. The resulting estimated value of α is then a crude measure of the nonlinearity in our signal set. Though it will not always be the case, it is reasonable to assume that we have some knowledge of the processes a and c. This assumption allows us to extract some physical meaning from the value of α.

Of course, we can define another second-order test of nonlinearity. Let us use the same signal set x_n and y_n, and this time we evaluate:

$$E[x_n^2 y_n^2] - E[x_n^2]E[y_n^2]$$

. It is easiest to perform this calculation in parts:

$$E[x_n^2 y_n^2] = E[a_n^2 c_n^2 + 2a_n b_n c_n^2 + b_n^2 c_n^2 + 2\alpha a_n^3 c_n^2 + 4\alpha a_n^2 b_n c_n^2$$
$$+2\alpha a_n b_n^2 c_n^2 + \alpha^2 a_n^4 c_n^2 + 2\alpha^2 a_n^3 b_n c_n^2 + \alpha^2 a_n^2 b_n^2 c_n^2] \tag{18}$$

, which (thankfully) can be simplified to:

$$E[x_n^2 y_n^2] = E[a_n^2 c_n^2 + b_n^2 c_n^2 + \alpha^2 a_n^4 c_n^2 + \alpha^2 a_n^2 b_n^2 c_n^2]$$
$$= E[a_n^2]E[c_n^2] + E[b_n^2]E[c_n^2] + 3\alpha^2 E[a_n^2]^2 E[c_n^2] + \alpha^2 E[a_n^2]E[b_n^2]E[c_n^2] \tag{19}$$

. We then evaluate:

$$E[x_n^2]E[y_n^2] = E[a_n^2 + 2a_n b_n + b_n^2]E[\alpha^2 a_n^2 c_n^2 + 2\alpha a_n c_n^2 + c_n^2]$$
$$= E[a_n^2 + b_n^2]E[\alpha^2 a_n^2 c_n^2 + c_n^2]$$
$$= \alpha^2 E[a_n^2]^2 E[c_n^2] + \alpha^2 E[a_n^2]E[b_n^2]E[c_n^2] + E[a_n^2]E[c_n^2] + E[b_n^2]E[c_n^2] \tag{20}$$

. We then combine the two together to find:

$$E[x_n^2 y_n^2] - E[x_n^2]E[y_n^2] = 2\alpha^2 E[a_n^2]^2 E[c_n^2] \tag{21}$$

. Again, we can approximate expectation values with means:

$$\overline{x_n^2 y_n^2} - \bar{x_n^2}\bar{y_n^2} \approx 2\alpha^2 \bar{a_n^2}^2 \bar{c_n^2} \tag{22}$$

. For simplicity, we introduce the notation:

$$\overline{x_n^2 y_n^2} \equiv M_2 \tag{23}$$
$$\bar{x_n^2}\bar{y_n^2} \equiv M_3 \tag{24}$$

, which allows us to solve (21) for α^2:

$$\alpha^2 \approx \frac{M_2 - M_3}{2A^2 C} \tag{25}$$

. This yields an interesting result if we note that dividing (25) by (17) gives α:

$$\alpha \approx \frac{M_2 - M_3}{2A^2 C} * \frac{2AC}{M_1} = \frac{M_2 - M_3}{AM_1} \tag{26}$$

. This tells us that in order to extract α only one parameter, A (which the reader will recognize as the variance of a_n), needs to be specified; all other information is contained within the discrete channels x_n and y_n. This information could be used in future work in quantification.

3.2 Limits of Second-Order Analysis

To illustrate the shortcoming of second-order tests, let us define a new discrete channel of a signal z as:

$$z_n \equiv (\alpha a_n + \beta b_n)c_n + c_n \tag{27}$$

, where β is a nonzero constant. Note that it is no longer meaningful to characterize the nonlinearity with a single parameter. This model might correspond to a modulation of an x signal that has been filtered in some fashion. That is, we could also write:

$$z_n = x_n^f c_n + c_n \tag{28}$$

, where x_n^f is some form of x_n that has passed through a filter. Here we use filter in the most general sense (for instance some mechanical process that amplifies or attenuates certain components of x), as most man-made filtering techniques would not produce x_n^f as we have defined it. Nevertheless, this is a completely plausible physical model worth consideration. If we were to run the same tests for x_n and z_n we would find:

$$\begin{aligned} E[x_n z_n^2] = E[&a_n c_n^2 + 2\alpha a_n^2 c_n^2 + 2\alpha a_n b_n c_n^2 + \alpha^2 a_n^3 c_n^2 + \alpha^2 a_n^2 b_n c_n^2 + 2\beta a_n b_n c_n^2 \\ &+2\beta b_n^2 c_n^2 + 2\alpha\beta a_n^2 b_n c_n^2 + 2\alpha\beta a_n b_n^2 c_n^2 + \beta^2 a_n b_n^2 c_n^2 + \beta^2 b_n^3 c_n^2] \end{aligned} \tag{29}$$

, which becomes:

$$E[x_n z_n^2] = 2\alpha E[a_n^2]E[c_n^2] + 2\beta E[b_n^2]E[c_n^2] \tag{30}$$

. One will note the similarity between (30) and (12). Introducing the mean value approximation we have:

$$\overline{x_n z_n^2} \approx 2\alpha \bar{a_n^2}\bar{c_n^2} + 2\beta \bar{b_n^2}\bar{c_n^2} \tag{31}$$

, and adopting the notation:

$$\overline{x_n z_n^2} \equiv \tilde{M}_1 \tag{32}$$

$$\overline{b_n^2} \equiv B \tag{33}$$

gives:

$$\tilde{M}_1 \approx 2\alpha AC + 2\beta BC \tag{34}$$

. Note that this now requires an estimate of three parameters A, B, and C. We then wish to solve for:

$$E[x_n^2 z_n^2] - E[x_n^2]E[z_n^2]$$

. It serves little good to include the great deal of algebra involved in solving for this, therefore we omit the steps and state that:

$$E[x_n^2 z_n^2] - E[x_n^2]E[z_n^2] = 2\alpha^2 E[a_n^2]^2 E[c_n^2] + 2\beta^2 E[b_n^2]^2 E[c_n^2] + 4\alpha\beta E[a_n^2]E[b_n^2]E[c_n^2] \tag{35}$$

. With the mean approximation and new notation:

$$\overline{x_n^2 z_n^2} \equiv \tilde{M}_2 \tag{36}$$

$$\bar{x_n^2}\bar{z_n^2} \equiv \tilde{M}_3 \tag{37}$$

, this becomes:

$$\tilde{M}_2 - \tilde{M}_3 \approx 2\alpha^2 A^2 C + 2\beta^2 B^2 C + 4\alpha\beta ABC \tag{38}$$

. We cannot combine (34) and (38) to solve for α or β in such a way as to eliminate any of the unknown parameters. In order to extract any reasonable estimate of α (or β), we would need knowledge of A, B, C, and β (or α). This might suggest that we need another test to provide any missing information, but note that the only second order test left would be:

$$E[x_n^2 z_n]$$

, which for our given signal set evaluates to:

$$\begin{aligned} E[x_n^2 z_n] = E[&a_n^2 c_n + 2a_n b_n c_n + b_n^2 c_n + \alpha a_n^3 c_n + 2\alpha a_n^2 b_n c_n + \alpha a_n b_n^2 c_n \\ &+ \beta a_n^2 b_n c_n + 2\beta a_n b_n^2 c_n + \beta b_n^3 c_n] \\ &= 0 \end{aligned} \tag{39}$$

. And so this last statistic will not provide any meaningful information about our signals. We can see that as the complexity of the nonlinearities increases, it becomes increasingly difficult to design worthwhile second-order tests.

Another important pitfall of the second-order analysis previously outlined is that it is only valid for Gaussian fundamental components. While it is certainly the case that many physical processes exhibit Gaussian nature, it is not necessarily the case that all components of the VIRGO data act in accordance with a normal distribution. In this sense, second-order statistical analysis is incomplete.

4 Higher-Order Statistical Analysis

4.1 Extending to Higher-Order Statistics

There are many ways to introduce higher-order statistical analysis, all of which have unique advantages and disadvantages. Seeing as this article is not intended to be a reference on statistical analysis, our approach is to give reasonable (often qualitative) arguments for how the need for higher-order statistics might arise. For a thorough treatment of cumulants and higher-order statistics, the reader is referred to [1], [2], and [3].

Let $\{h_m\}$ be a set of n random scalar values which define a plotted distribution H. Note that this distribution allows for no knowledge of any possible temporal structures, but often this is not critical. We can obtain some useful information about the nature of H with the first and second order statistics we have already worked with. For example, we can define a mean of H, which we denote $\bar{H}$, by evaluating:

$$\bar{H} \equiv \frac{1}{n} \sum_{m=1}^{n} h_m \tag{40}$$

. Qualitatively, this corresponds to an effective central point for the distribution. (We say that this is "effective" because H might be constructed in such a way as to make a central point essentially meaningless). It is often convenient to subtract this mean from all values in the set $\{h_m\}$:

$$h_m^0 = h_m - \bar{H} \tag{41}$$

, and we now call H^0 the distribution defined by the set $\{h_m^0\}$. It is easy to show that the mean of H^0 is:

$$\begin{aligned}
\bar{H^0} &= \frac{1}{n} \sum_{m=1}^{n} h_m^0 \\
&= \frac{1}{n} \sum_{m=1}^{n} (h_m - \bar{H}) \\
&= \frac{1}{n} (\sum_{m=1}^{n} h_m - \sum_{m=1}^{n} \bar{H}) \\
&= \frac{1}{n} \sum_{m=1}^{n} h_m - \frac{1}{n} \sum_{m=1}^{n} \bar{H} \\
&= \bar{H} - \bar{H} \\
&= 0
\end{aligned} \tag{42}$$

. We might wish to say that H^0 is centered about zero. Often, however, we are less concerned with a central point, and more concerned with the shape of H^0. For instance, we might want a measure of the relative spread of H^0 about its mean. A natural measure might be the variance of H^0:

$$\begin{aligned}
\sigma^2(H^0) &\equiv \frac{1}{n} \sum_{m=1}^{n} (h_m^0 - \bar{H^0})^2 \\
&= \frac{1}{n} \sum_{m=1}^{n} (h_m^0)^2
\end{aligned} \tag{43}$$

. Now suppose that we would like a measure of the degree to which H^0 is asymmetric, or unevenly dense, about its mean. A reasonable statistic might be something such as:

$$\mu^3(H^0) = \frac{1}{n}\sum_{m=1}^{n}(h_m^0 - \bar{H^0})^3$$

$$= \frac{1}{n}\sum_{m=1}^{n}(h_m^0)^3 \tag{44}$$

, though if we wish to compare this measure to that of other distributions, it is helpful to account for the variance of H^0, so we can revise our statistic:

$$S(H^0) = \frac{1}{n}\sum_{m=1}^{n}\left(\frac{h_m^0}{\sigma}\right)^3 \tag{45}$$

, which can be written as:

$$S = \frac{\mu^3}{\sigma^3} \tag{46}$$

. Formally, we call S the skewness of H^0. Note that S is easiest to interpret for distributions of zero mean. In general, distributions tend to be most dense near the mean and gradually tail off in either direction; often a marker of non-deterministic processes. Assuming H^0 follows this trend, it might be physically meaningful to examine the weight of the tail in relation to the denser region about the mean. The proper statistic would not favor either extreme and would give great weight to large values while diminishing the impact of smaller ones. A natural suggestion might be:

$$\mu^4(H^0) = \frac{1}{n}\sum_{m=1}^{n}(h_m^0 - \bar{H^0})^4$$

$$= \frac{1}{n}\sum_{m=1}^{n}(h_m^0)^4 \tag{47}$$

, which, as with the skewness, we should standardize for given variance:

$$K'(H^0) = \frac{1}{n}\sum_{m=1}^{n}\left(\frac{h_m^0}{\sigma}\right)^4 \tag{48}$$

, which we simply write as:

$$K' = \frac{\mu^4}{\sigma^4} \tag{49}$$

. One will notice that for a normal distribution $\mu^4 = 3\sigma^4$, and so K' evaluates to:

$$K' = \frac{3\sigma^4}{\sigma^4} = 3 \tag{50}$$

, and so we use the convention:

$$K = \frac{\mu^4}{\sigma^4} - 3 \tag{51}$$

. We call the statistic K the kurtosis of H^0. Not only does K give information about the weight of the tail of H^0, but it also serves as a de facto measure of peakedness about the mean. Generally, a higher kurtosis corresponds to a greater peakedness.

Statistics beyond fourth-order have less concrete physical interpretations and are not introduced in this article. Furthermore, higher order statistics become increasingly difficult to accurately estimate from data. That is not to say that there is no potential utility in even higher order statistics, but this pursuit is beyond the scope of this work.

4.2 Frequency Domain

In a physical system, it is often convenient to think of signal behavior as the sum of independent frequency components. This approach is intuitively cohesive with our assertion of fundamental signal components. The extension to the frequency domain, as enacted through the application of Fourier's Theorem, is pivotal to our search for nonlinearity. In order to properly explain our particular use of the frequency domain, it is necessary to introduce several statistics and statistical concepts outlined in the remainder of this subsection. For this subsection we have adopted notation similar to that of [2].

Cumulants are a generalized statistic from which many useful measures (such as those defined in the previous subsection) can be constructed. As it is not necessary for one to have a strong intuition for cumulants in our work, discussion on cumulants is limited in this article to articulation rather than analysis. The second-order cumulant of two discrete temporal processes, x and y, is defined as:

$$C_{xy}(h) = E[x^*(n)y(n+h)] \tag{52}$$

, where the asterisk is meant to denote the complex conjugate. We define the cross-power spectrum of the two processes x and y as:

$$\Gamma_{xy}(\omega) = \sum_{h=-\infty}^{\infty} C_{x,y}(h)e^{-2\pi i\omega h} \tag{53}$$

, which we often write as:

$$\Gamma_{xy}(\omega) = E[X^*(\omega)Y(\omega)] \tag{54}$$

, where X and Y are the Fourier transforms of x and y respectively. The idea of a spectrum allows us to extract a kind of energy distribution, and proves useful in identifying correlations when generalized to higher orders. Introducing another discrete temporal process z, we can extend (52) to a third-order cumulant:

$$C_{xyz}(h, j) = E[x^*(n)y(n + h)z(n + j)] \tag{55}$$

, and define a cross-bispectrum of the processes x, y, and z as:

$$\Gamma_{xyz}(\omega_1, \omega_2) = \sum_{j=-\infty}^{\infty} \sum_{h=-\infty}^{\infty} C_{xyz}(h, j)e^{-2\pi i\omega_1 h}e^{-2\pi i\omega_2 j} \tag{56}$$

, which is written short-form as:

$$\Gamma_{xyz}(\omega_1, \omega_2) = E[X^*(\omega_1 + \omega_2)Y(\omega_1)Z(\omega_2)] \tag{57}$$

, where Z is the Fourier transform of z. An analysis of such extensions from higher order cumulants as in (54) and (57) is the subject of a field called Higher-Order Spectral Analysis (HOSA).

5 Independent Component Analysis

5.1 Principle of ICA

Blind source separation (BSS) is a term for any technique that aims to filter out source signals from a mixture with very little knowledge of the mixing process or the source signals themselves. Independent component analysis (ICA) is an approach to BSS in cases where the source signals are expected to be non-Gaussian and there are as many source mixtures as there are sources. In accordance with the Central Limit Theorem (CLT), mixtures of the source signals are better approximations of Gaussian signals than the sources themselves ([1] and [4]). In general, ICA techniques take advantage of this property (or an expansion of this property) to split mixtures into components.

ICA operates on two assumptions: the source signals are statistically independent (necessary condition for the CLT), and the mixing process is linear. For an in-depth explanation for these conditions the reader is referred to [4]. In **5.1.1** and **5.1.2** we proceed to define the two conditions.

5.1.1 Statistical Independence

Let us define two signals s_1 and s_2, then, in broad terms, we would say that s_1 and s_2 are statistically independent if, for any time t, the value $s_1(t)$ provided

no information about the value $s_2(t)$, and vice versa. Mathematically, this can be expressed as:

$$\forall p, q \quad E[s_1^p s_2^q] = E[s_1^p]E[s_2^q] \tag{58}$$

. Often it is enough that s_1 and s_2 are uncorrelated (when $p = q = 1$).

5.1.2 Linear Mixing

Let us define a signal mixture m_k of a set of N fundamental components $\{s_i : 1 \leq i \leq N\}$, then m_k is said to be a linear mixture if:

$$m_k = \sum_{i=1}^{N} \phi_{k,i} s_i \tag{59}$$

, for some set of scalars $\{\phi_{k,i} : 1 \leq i \leq N\}$. As each fundamental component is approximated as a time series, it makes sense to write s_i as a row vector $\vec{s_i}$, where each consecutive column corresponds to a consecutive measurement. Likewise, we represent the mixture m_k as a vector $\vec{m_k}$ in a similar fashion, so that we can now write:

$$\vec{m_k} = \sum_{i=1}^{N} \phi_{k,i} \vec{s_i} \tag{60}$$

. In practice, we need as many signal mixtures as there are fundamental components, and thus we must concern ourselves with the set $\{\vec{m_k} : 1 \leq k \leq N, \vec{m_i} \neq \vec{m_j} \text{ if } i \neq j\}$. Therefore, it is convenient to define the three matrices $\mathbf{M}$, $\mathbf{\Phi}$, and $\mathbf{S}$ such that:

$$\mathbf{M}_{ij} \equiv \vec{m}_{ij} \tag{61}$$

$$\mathbf{\Phi}_{ij} \equiv \phi_{i,j} \tag{62}$$

$$\mathbf{S}_{ij} \equiv \vec{s_i}_j \tag{63}$$

. With this new notation, (59) can be expanded as:

$$\mathbf{M} = \mathbf{\Phi S} \tag{64}$$

, where we call $\mathbf{\Phi}$ the mixing matrix. For our purposes, we need only to consider the case when $\mathbf{\Phi}$ is a square matrix. Equation (64) is representative of the general environment in which BSS is intended to work.

5.2 ICA and Other BSS Methods

The intent of BSS, as previously stated, is to recover source signals from a set of mixtures. From a theoretical standpoint, we can do this by defining a new matrix $\mathbf{\Phi}^{-1}$ such that:

$$\mathbf{\Phi}^{-1}\mathbf{\Phi} = \mathbf{I} \tag{65}$$

, where $\mathbf{I}$ is the identity matrix, and then applying this to the mixture matrix $\mathbf{M}$:

$$\mathbf{\Phi}^{-1}\mathbf{M} = \mathbf{\Phi}^{-1}\mathbf{\Phi}\mathbf{S} = \mathbf{S} \tag{66}$$

. Accordingly, BSS is analagous to a pursuit of the matrix $\mathbf{\Phi}^{-1}$ (commonly referred to as the unmixing matrix). In practice, little, if anything, is known about the mixing matrix $\mathbf{\Phi}$, and so an analytical approach to constructing $\mathbf{\Phi}^{-1}$ is impossible. Instead, BSS methods must take an indirect approach, and optimize an estimate of the unmixing matrix. Denote the estimate of the unmixing as $\tilde{\mathbf{\Phi}}^{-1}$, then in practice an application of the matrix would be as such:

$$\tilde{\mathbf{\Phi}}^{-1}\mathbf{M} = \tilde{\mathbf{S}} \tag{67}$$

, where $\tilde{\mathbf{S}}$ is an estimate of the signal matrix $\mathbf{S}$. In order to obtain a best estimate of $\mathbf{S}$, we must find a measure which when maximized or minimized indicates a low degree of mixing (minimizing Gaussianity is a common test). Define an arbitrary satisfactory measure d: $R^{n\times n} \rightarrow R$, and assume, without loss of generality, that maximized d corresponds to minimal mixing. And so we then aim to find a $\tilde{\mathbf{\Phi}}^{-1}$ such that $d(\tilde{\mathbf{S}})$, or equivalently $d(\tilde{\mathbf{\Phi}}^{-1}\mathbf{M})$, is a maximum. This is a maximization problem with n^2 parameters, but we can envision this as an n parameter problem if we let $\vec{\mu_i}$ represent the i^{th} row of the matrix $\tilde{\mathbf{\Phi}}^{-1}$. It is convenient to think of d as a mapping from the set $\{\vec{\mu_i}\}$ to the reals. This new domain treats the maximization problem as a search for optimal n-dimensional spacial orientations. It is helpful to affix a subscript to the measure's notation, d_M, to denote that the measure is implicitly dependent upon the particular mixture matrix $\mathbf{M}$. We state without argument that it is reasonable to assume d_M is differentiable with respect to each $\vec{\mu_i}$, thus the problem becomes finding a set $\{\vec{\mu_i}\}$ such that:

$$\forall i, \quad \frac{\partial d_M}{\partial \vec{\mu_i}} = 0 \tag{68}$$

. Though the partial derivative is somewhat abstract, not mathematically rigorous, and generally not indicative of the actual approach taken by a solving algrithm, it elucidates the broader structure of the approach which is not apparent if the technique is articulated as a maximization subject to n^2 individual parameters. In practice, there may be several solution sets (corresponding to global extrema and possible local extrema), and it then becomes imperative to evaluate d_M. The solution set $\{\vec{\mu_i}\}$ that maximizes the value of d constructs the best estimate of the unmixing matrix $\mathbf{\Phi}^{-1}$. This ensures that the extracted

$\tilde{\mathbf{S}}$ is the best estimate of $\mathbf{S}$ for the given measure d_M. To optimize results, the choice of d_M will generally depend upon assumptions about the nature of the fundamental signals.

6 Integrating ICA and HOSA

Our stated purpose is to identify and quantify nonlinear correlations between environmental channels and the dark fringe. For a more complete explanation of the integration technique, the reader is referred to [5]

We define a main channel of interest, z, and two mixture groups x and y such that:

$$x = \sum_{i=0}^{n-1} c_{1,i} s_i \tag{69}$$

$$y = \sum_{i=0}^{n-1} c_{2,i} s_i \tag{70}$$

, where $\{c_{1,i} : 0 \leq i \leq n-1\}$ and $\{c_{2,i} : 0 \leq i \leq n-1\}$ are sets of scalars and $\{s_i : 0 \leq i \leq n-1\}$ is the set of available data channels. It is also convenient at times to use vector representation, so that we write:

$$\vec{x} = \vec{c_1}\mathbf{S} \tag{71}$$

$$\vec{y} = \vec{c_2}\mathbf{S} \tag{72}$$

, where $\vec{x}$, $\vec{y}$, $\vec{c_1}$, and $\vec{c_2}$ are row vectors and $\mathbf{S}$ is a matrix defined in a similar sense as that in (63). For simplicity, let $\vec{z_k}$, $\vec{x_k}$, $\vec{y_k}$ be the k^{th} component of the vectors $\vec{z}$, $\vec{x}$, and $\vec{y}$ respectively. We then denote the Fast Fourier Transform (FFT) for these series as:

$$Z(\omega) = \sum_{k=0}^{n-1} \vec{z_k} e^{-2\pi i \omega \frac{k}{n}} \tag{73}$$

$$X(\omega) = \sum_{k=0}^{n-1} \vec{x_k} e^{-2\pi i \omega \frac{k}{n}} \tag{74}$$

$$Y(\omega) = \sum_{k=0}^{n-1} \vec{y_k} e^{-2\pi i \omega \frac{k}{n}} \tag{75}$$

, where the argument ω is a particular frequency. We are then interested in the quantity:

$$\Gamma_{zxy}(\omega_1, \omega_2) = E[Z^*(\omega_1 + \omega_2) X(\omega_1) Y(\omega_2)] \tag{76}$$

. We hope to extremize:

$$\Lambda' = \left|\Gamma_{zxy}(\omega_1, \omega_2)\right|^2 \tag{77}$$

, but we immediately observe that Λ' can be made arbitrarily large by making the scalar values $\|\vec{c_1}\|$ and $\|\vec{c_2}\|$ arbitrarily large. Hence, we must subject our extremization to a reasonable condition that prevents arbitrarily large Λ. We choose to restrict the the bispectra of x and y such that:

$$\Gamma_{xx}(\omega_1) = \Gamma_{yy}(\omega_2) = 1 \tag{78}$$

. This condition will place sufficient bounds upon $\vec{c_1}$ and $\vec{c_2}$. We employ the method of Langrangian multipliers and construct a new quantity that we wish to extremize:

$$\Lambda = \left|\Gamma_{zxy}(\omega_1, \omega_2)\right|^2 + \lambda_1\Gamma_{xx}(\omega_1) + \lambda_2\Gamma_{yy}(\omega_2) \tag{79}$$

. So our problem is to find vectors $\vec{c_1}$, $\vec{c_2}$ that maximize Λ. The mathematics are quite lengthy, and will not be included here, but it can be shown [5] that the problem simplifies to a search for the largest eigenvalue of:

$$U_X^{-1^T}\Gamma_{zxy}U_Y^{-1}U_Y^{-1^+}\Gamma_{zxy}^+U_X^{-1^*} \tag{80}$$

, where:

$$\Gamma_{xx} = U_X^+U_X \tag{81}$$

$$\Gamma_{yy} = U_Y^+U_Y \tag{82}$$

for a specified combination of frequencies. From this eigenvalue we can optimize the group mixing coefficients $c_{1,i}$ and $c_{2,i}$. The cross-bispectrum is then evaluated for the groups and main channel and plotted. This value tells us whether a modulation between a frequency of one group and that of another are present in the main channel.

The product of our evaluation will be a visual that allows for interpretation, we choose this product to be the cross-bispectrum of two groups of channels and one main channel. The expectation is that the dark fringe channel V1:Pr_B1_ACp will serve as the main channel. The script allows for frequency window specification with respect to both groups, and also allows the user to set a threshold percentage for the results. A time window is selected and the channels are transformed into frequency domain series; the appropriate calculations ensue. It is possible to specify a given number of successive time intervals over which the results are averaged.

7 Applying Technique to VIRGO Data Channels

Though time for actual testing of the code was extremely limited, certain observations and explications are worth including.

7.1 Simulated Data Testing

When testing an algorithm, it is often useful to use inputs that are designed with expected, well understood outputs, which can then be juxtaposed with actual outputs. Unfortunately, there was not sufficient time to perform an extensive test of the script with simulated data, and the coding seemed to have a hard time handling modifications of the simulated data forms. This would suggest that the parameters for creating the simulated data are not limited in such a way that is fully compatible with the script. Future work with simulated data is a possibility, but is not of primary concern.

7.2 VIRGO Data Testing

The VIRGO data tests illustrated several distinct problems with the algorithm. The most obvious issue is the existence of unwanted artifact structures which visually suggest correlations that do not exist. The prevalence of the phantom structures is troubling, and probably arises from the frequency resolution with which the algorithm is implemented. If this is indeed the case, longer periods of averaging will not necessarily hide these artifacts.

More cause for concern are the inconsistencies in the visual between evaluations for different frequency windows. Oftentimes structures will appear at certain locations in particular evaluations, only to be completely absent, or of different nature, when the frequency window is adjusted (see presentation slides for figures). The reason for this is less clear than that of the artifact issue. It is possible that this is actually an extension of the artifact problem, and that different frequency windows produce different artifacts.

The limits of the script quickly become apparent when running evaluations. The length of signal time intervals, which will theoretically increase the accuracy of the script output as it increases, is restricted by limited computing power, and so a proper implementation of the algorithm will likely require a better computer. Even within reasonable limits, the evaluations often take upwards of fifteen minutes, which severely limits the efficacy of the script as a real-time indicator.

7.3 Conclusions

At this point there is no basis for drawing any conclusions about the actual VIRGO apparatus or noise within. Any conclusion must be an evaluation of the

methodology and intent of the work rather than of the work itself. Confidence in the script's ability to identify nonlinear behavior can be separated into two components: theory and practice.

In theory, it appears that the algorithm is capable of detecting nonlinear correlations between channels. The cross-bispectrum is a legitimate measure of nonlinearities, and the optimization technique has proven to be effective (though further exploration should be done). Justifications for the method can be found in [5], and are well defended.

In practice, the script has not reached the point where it could be utilized by monitors in the control room. It is not yet ready for real-time analysis, and it is not immune to artifacts of the algorithm. Any further conclusions about the script would be unwarranted.

8 Future Work

The work and analysis outlined in previous sections does not constitute a successful realization of the goals stated in Section 1. As such, it is imperative that we work to improve our approach to the matter in the future. Though decreased proximity and access to data strings of interest will be significantly limited, there are still several projects that are reasonable to undertake. This section considers certain projects worth pursuing.

8.1 Algorithm Refinement

The algorithm itself is central to any future work, and it would be worthwhile to fix or refine any issues there might be. Notably, the script is very slow to evaluate; a truth that has hampered serious testing. It would merit the effort to cut down on the computational load of the algorithm. Other improvements not related to speed or efficiency are listed and briefly outlined in this subsection.

8.1.1 Nonlinear Indicator

One approach to improvement might be to reconsider the measure of nonlinearity that the algorithm employs. In order to determine the utility of such an approach, the physical interpretation of the current measure requires further exploration. Referral to a physical meaning allows us to choose a statistic that best amplifies nonlinearities of the nature that are most expected in the experiment. In the time domain, kurtosis is a very popular indicator, however the frequency domain is largely unchartered with respect to our maximization problem.

8.1.2 Signal Pre-Processing

The script output appears to be very noisy, especially for small frequency windows, and it could possibly benefit from pre-processing VIRGO data channels (windowing, whitening, etc.). A more rigorous analysis of the script may yield more insight into this issue, and we might possibly be able to develop a successful signal cleaning method.

8.1.3 Unknown Bug Testing

It is very likely that all the issues with the script have not been uncovered as of yet, and more testing with both simulated data and actual VIRGO data needs to be done to identify such problems. Any work with actual data from any of the VIRGO sensors and probes would be restricted by the author's limited access to the VIRGO network, though that is not to say that such a pursuit is not possible in the future. One known issue is the window inconsistency, the source of this issue, however is not clear and will require more work with the script.

8.2 Channel Selection

There is a tremendous amount of data available from probes and sensors throughout the apparatus and surrounding area which we could potentially analyze. The computational requirements of each evaluation, along with the large number of possible group arrangements makes a comprehensive approach unreasonable. Even given a set group length, testing all, or even most possible group arrangements is beyond our capability. Because of this, it is important that we are able to focus our efforts on channels that present the greatest potential for containing nonlinearities. This selection requires knowledge of possible sources of nonlinear processes in the experiment that is currently not available or not organized.

8.3 Quantification and Interpretation

As of yet, we do not have a structured approach to interpreting the visual output of the script. To develop such an approach will require extensive testing of the script, certainly involving more simulated data. As was mentioned earlier in this article, more research into statistics that might provide meaningful quantifying values could act as a possibility, though it would be well beyond any results reached in this article.

References and Other Readings

[1] M B Priestly, *Spectral Analysis and Time Series. Volumes I and II in 1 book*, Academic Press, 1983

[2] A Swami, J M Mendel, and C L Nikias, *Higher-Order Spectral Analysis Toolbox Users Guide*, United Signals & Systems Inc., 1998

[3] C Rose and M Smith, *Mathematical Statistics with Mathematica*, Springer Publishing Company, 2002

[4] J V Stone, *Independent Component Analysis: A Tutorial Introduction*, MIT Press, 2004

[5] G Cella, *Detection of Nonlinear Correlations*, VIRGO, July, 2009

E Cuoco and A Viceré, *Mathematical Tools for Noise Analysis*, VIRGO, February, 1999

G Cella, *Coding Notes*, VIRGO, August, 2009

A Hyvärinen, J Karhunen, and E Oja, *Independent Component Analysis*, Wiley-Interscience, 2001

A Hyvärinen and E Oja, *Independent Component Analysis: Algorithms and Applications*, Neural Networks, Volume 13, pp. 411-430, 2000

S M Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*, Prentice-Hall, Inc., 1998

J V Stone, *Independent Component Analysis*, Encyclopedia of Statistics in Behavioral Science, Volume 2, pp. 907-912, 2005

The Fundamentals of Signal Analysis: Application Note 243, Agilent Technologies, 2000