HOW EXTRACTING INFORMATION FROM DATA HIGHPASS FILTERS ITS ADDITIVE NOISE

Victor S. Reinhardt
Raytheon Space and Airborne Systems
El Segundo, CA, USA

Abstract

This paper examines the characteristics of three types of random error measures in the presence of negative power law (neg-p) noise: (a) the observable residual error after removing an estimate of an information containing causal function from data, (b) the jitter, the residual error with additional highpass (HP) filtering, and (c) M^{th}-order difference (Δ) variances, such as the Allan variance (1^{st}-order Δ-variance of the fractional frequency error y(t)) and the Hadamard-Picinbono variance (2^{nd}-order Δ-variance of y(t)). Measures (b) and (c) are used to mitigate perceived divergence problems in the mean square (MS) of Measure (a) due to the presence of neg-p noise. This paper proves that this perception is wrong; it shows that the MS of Measure (a) converges in the presence of neg-p noise by demonstrating that extracting a statistically optimal estimate of the causal behavior from data HP filters the noise in the measure. It is further shown that the order of this noise HP filtering increases with the complexity of the model function used to estimate the causal behavior in the data. Thus, if one is free to choose the complexity of the model function, the MS observable residual error is guaranteed to converge for any negative power in the noise PSD. Because of this, it is shown that the jitter can be defined simply as the observable residual error without additional HP filtering, making the jitter and residual error the same error measure. This paper finally shows that an M^{th}-order Δ-variance is also a measure of the MS of the observable residual error for any number of data samples when the model function is an $(M-1)^{th}$-order polynomial. This completes the equivalence, showing that Measures (a), (b), and (c) all measure the same kind of error when the model function for the causal behavior is a polynomial. The consequences of this equivalence are then explored. Among these is a physical explanation for the fact that the Allan variance is sensitive to frequency drift, while the Hadamard-Picinbono variance is not.

INTRODUCTION

Different approaches for specifying random time, phase, and frequency (TPF) error are used across the electrical engineering (EE) community. As shown in Figure 1, these approaches are as follows.

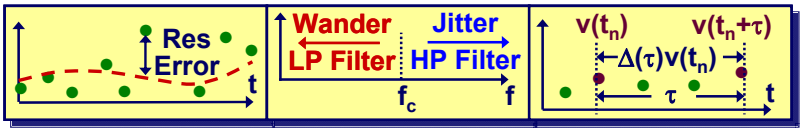

Figure 1. Residual error (left), jitter and wander (middle), and difference variances (right).

OBSERVABLE RESIDUAL ERROR

We will define $\sigma^2_{v-j}(N,M)$, the mean square (MS) observable residual error, as the MS of the difference between a set of N data samples $v(t_n)$ (over an observation time T) and $v_{w,M}(t_n,\mathbf{A})$ an estimate of a true causal function $v_c(t)$ imbedded in the data **[1]**. The parameter M in $\sigma^2_{v-j}(N,M)$ relates to the M in $v_{w,M}(t_n,\mathbf{A})$ and will be explained later. $\sigma^2_{v-j}(N,M)$ is an important random error statistic in system specification, because it relates directly to primary performance measures in many systems. These measures include the signal-to-noise ratio (SNR), the bit or symbol error rate (BER), the noise power ratio (NPR), the effective number of bits (ENOB), and the multiplicative noise ratio (MNR) (signal processing noise) **[2-8]**. In many treatments, the terms standard, sample **[9]**, and MS (RMS for deviate) error are used when $\sigma^2_{v-j}(N,M)$ is intended, but we will use the neutral term residual error, because it does not have the alternate definitions or connotations associated with these other terms. Note that we are explicit in stating that it is $v_{w,M}(t_n,\mathbf{A})$, the estimate of the causal function, that is removed from the data, not $v_c(t)$, the true causal function. The distinction between the estimated and true causal functions is often glossed over in treatments of residual error, but this distinction will be important later in our discussion. Thus, only the observable residual error, based on causal behavior estimated from the data set itself, is directly measurable or observable from a set of data, even if the desired error measure for system specification is the true residual error **[1]**.

$v_{w,M}(t_n,\mathbf{A})$, the model function used to estimate $v_c(t)$, will be considered a function of M parameters represented by the column vector $\mathbf{A}=(a_0,a_1,...a_{M-1})'$ (' is the matrix transpose) as well as the observation or sample time t_n. Thus, the M in $v_{w,M}(t,\mathbf{A})$ is the source of the M in $\sigma^2_{v-j}(N,M)$. In this paper, we will adjust $\mathbf{A}$ in $v_{w,M}(t_n,\mathbf{A})$ to obtain a statistically optimal estimate of $v_c(t)$ by utilizing a least-squares fit (LSQF) over the N samples of data $v(t_n)$ **[1]**, where we will assume the $v(t_n)$ are evenly spaced over the observation interval T. For Gaussian random (though divergent neg-p) noise, such an LSQF is equivalent to other maximum likelihood methods **[1,10]**, especially if we allow the LSQF to be weighted **[1]**. Estimating the M parameters in $\mathbf{A}$ or the function $v_{w,M}(t,\mathbf{A})$ from the data is equivalent to the extraction of causal information from the data, hence the title of the paper. This is true whether the information is a desired product of the system or is just another error parameter that impacts system performance, such as the frequency aging of an oscillator.

JITTER AND WANDER

Jitter and wander **[4,11,12]** have been introduced to deal with perceived divergence problems in the residual error associated with the presence of negative power law (neg-p) noise **[13]**. Time jitter and wander are currently defined as brick-wall highpass (HP) and lowpass (LP) filtered variations in the time error x(t) or the time interval error TIE with a crossover frequency f_c excluding causal frequency offsets and drifts (and implicitly causal time offsets) **[4,11,12]**. Thus, jitter and wander are effectively HP and LP filtered x(t) residual errors after removal of the 2nd-order causal behavior in x(t). The brick-wall HP filtering in the jitter ensures convergence for neg-p noise and dumps the convergence problem into the wander, which is usually ignored in discussions of jitter.

For the purposes of the discussion that follows, we note that power law (neg-p) noise for a general variable v(t) is wide-sense stationary noise with a single sideband power spectral density (PSD) $L_v(t) \propto f^p$ **[13]**. Here, we are departing from **[13]** in utilizing the SSB PSD $L_v(t)$ rather than the double

sideband or single-sided PSD $S_v(t)$, hence the use of $L_v(t)$ to avoid confusion with the DSB form $S_v(t)$. Neg-p noise will designate such power law noise when $p<0$ (usually $-1,-2,-3,-4$ **[13]**), and the $p=0$ case is generally called white v-noise. We also note that v(t) will be generally considered a reference to TPF variables, such as x(t) the time error, $\phi(t)$ the phase error, or y(t) the fractional frequency error **[4,13]**, but results in this paper in terms of v(t) will apply any variable without limitation.

The problem associated with the use of jitter and wander is that the relationship of f_c to natural filtering parameters in the system under consideration is often unclear. The ITU arbitrarily defines f_c as 10 Hz **[4]**. This is helpful in standardizing producers of TPF equipment, but bears only an accidental relationship to parameters in general user systems. The IEEE Broadcast Technology Society (BTS) **[11]** and the Society of Motion Picture and Television Engineers (SMPTE) **[12]** relate f_c to the loop bandwidth of a phase locked loop (PLL) in a user system. This is helpful for users with PLLs in their systems, but leave users without appropriate PLLs in their systems in doubt. We will address this f_c relationship problem later in this paper.

M[TH]-ORDER DIFFERENCE (Δ) VARIANCES

$\sigma^2_{v,M}(\tau)$, the M[th]-order difference (Δ) variances **[14]**, is a generalization of the Allan or two-sample variance **[13]**, which is the 1[st]-order Δ-variance of y(t) and the zero-dead-time Hadamard variance **[15-18]** or Picinbono variance **[19]**, which is the 2[nd]-order Δ-variance of y(t). These variances are considered stability measures and are used because of their excellent convergence properties in the presence of neg-p noise **[13,14,16-19]**. In fact, one can show, for any M, that the M[th]-order Δ-variance of v(t) will HP filter $L_v(f)$ with a 2M[th] zero at f=0 **[14]**. As indicated in Figure 2, $\sigma^2_{v,M}(\tau)$ is given by **[14]**

$$\sigma^2_{v,M}(\tau) = \lambda_M^{-1} MS\{\Delta(\tau)^M v(t_n)\} \tag{1}$$

where $\Delta(\tau)$ is the forward difference operator over the separation interval τ defined by **[14]**

$$\Delta(\tau)v(t) = v(t+\tau) - v(t) \tag{2}$$

and the normalization constant is defined as **[14]**

$$\lambda_M = \sum_{m=0}^{M}\left(\frac{M!}{m!(M-m)!}\right)^2. \tag{3}$$

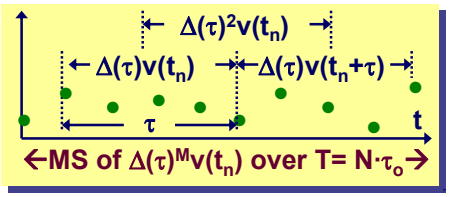

Figure 2. M[th]-order difference variances.

We note that this definition of λ_M makes all M orders of $\sigma^2_{v,M}(\tau)$ equal for uncorrelated white noise **[14]**. The MS operation in (1) is over N samples at $t_n = t_o + n\tau_o$ ($n=0$ *to* $N-1$), so that the samples again fall

within a total observation interval $T = N\tau_o$, where τ_o is the sampling interval. There is more than one way to take the MS over the data, and we will discuss this later in the Mathematical Niceties Section.

THE HIGHPASS FILTERING OF NOISE DUE TO INFORMATION EXTRACTION

The common wisdom is that the rigorous form of the MS residual is divergent in the presence of negative power law (neg-p) noise. This paper will show that this common wisdom is wrong, because the process of optimally estimating the causal function HP filters the noise in the residual error. We will also show that the order of this HP filtering increases with the complexity of the information extracted from the data, as expressed by the number of parameters M in $v_{w,M}(t,\mathbf{A})$. Thus, if one is free to choose the complexity of $v_{w,M}(t,\mathbf{A})$, the residual error is guaranteed to converge for any negative power in $L_v(f)$.

This HP filtering of the noise in the observable residual error allows one to eliminate additional filtering in the definitions of jitter and wander. Thus, we will redefine the jitter simply as the observable residual error without such additional filtering. Later in the paper, we will show that the needed HP filtering is supplied by a combination of the LSQF process and other system related filtering. The wander will then be redefined as the difference between the estimated and true causal functions, which we will show is LP filtered by the same system related filtering processes as the jitter. These new definitions solve the relationship problem between f_c and system parameters for general user systems. They also allow one to generalize the concepts of jitter and wander to include variations in any variable and the removal of any type of causal behavior.

M^{TH}-ORDER DIFFERENCE VARIANCES AS MEASURES OF RESIDUAL ERROR

The paper will also show that $\sigma^2_{v,M}(T/M)$ is a measure of $\sigma^2_{v-j}(N,M)$ for any N when $v_{a,M}(t,\mathbf{A})$ is an $(M-1)^{th}$ polynomial and the MS operation for $\sigma^2_{v-j}(N,M)$ uses the "unbiased" estimator; that is, it divides the sum in the MS operation by $N-M$. Unbiased is in quotes because it is the true unbiased estimator only for uncorrelated white noise **[20,21]**. This relationship between $\sigma^2_{v,M}(T/M)$ and $\sigma^2_{v-j}(N,M)$ is well- known for the Allan or two-sample variance $\sigma^2_y(\tau)$ [13][23] or $\sigma^2_{y-j}(2,1)$, which has been related to the N-sample variance $\sigma^2_y(N,\tau,\tau)$ **[23,24]** or $\sigma^2_{y-j}(N,1)$ using Allan-Barnes "bias" functions **[23,24]**. We will generalize this argument to relate $\sigma^2_{v,M}(T/M)$ to $\sigma^2_{v-j}(N,M)$ using a similar bias function concept, except that, in our case, we will hold the total observation time T constant as N varies. Because of this difference, the bias function relating $\sigma^2_{v,M}(T/M)$ to $\sigma^2_{v-j}(N,M)$ is much less dependent on p than in the Allan-Barnes case, and we will show that there is a simple approximate form for our bias functions independent of p.

This brings us full circle, showing that residual variances, jitter, and Δ-variances measure the same type of random error when a polynomial is used to model the causal behavior of the data. This relationship between $\sigma^2_{v,M}(T/M)$ and $\sigma^2_{v-j}(N,M)$ also yields a physical interpretation for the well-known insensitivities of $\sigma^2_{v,M}(\tau)$ to $(M-1)^{th}$-order or lower polynomial causal aging behavior **[25,26]**.

MATHEMATICAL NICETIES

DATA, NOISE, AND RESIDUAL ERROR MODEL

Figure 3 shows the model we will use for the N data samples $v(t_n)$ and the various types of residual errors we will be defining. Let us represent the total (continuous) data variable $v(t)$ as

$$v(t) = v_c(t) + v_p(t) \tag{4}$$

where $v_c(t)$ is the true causal function imbedded in the data and $v_p(t)$ is the true random noise. We will assume that $v(t)$ and, thus, $v_c(t)$ and $v_p(t)$, have been pre-filtered by a system response function written as $h_s(t)$ in the time domain and $H_s(f)$ in the frequency domain such that

$$v(t) = \int_{-\infty}^{+\infty} dt\, h_s(t-t')v_{in}(t') \tag{5}$$

where $v_{in}(t)$ is the variable prior to system filtering **[5,6,27]**. $H_s(f)$ describes the filtering action of the system on the variable over and above any filtering introduced by the LSQF. It is well-known that such a filter acting on the pre-filtered noise variable corresponding to $v_p(t)$ will produce an output PSD of $|H_s(f)|^2 L_v(t)$ when the pre-filtered PSD of the noise variable is $L_v(t)$ **[10]**. Thus, the PSD of the post-filtered $v_p(t)$ will be written as $|H_s(f)|^2 L_v(t)$ in this paper, so $|H_s(f)|^2$ explicitly appears in spectral formulas.

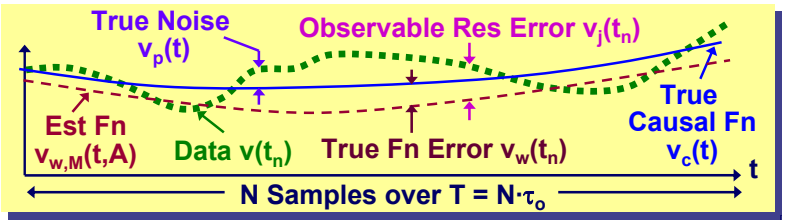

Figure 3. Data model and residual errors.

We will further assume the model function $v_{w,M}(t,\mathbf{A})$ is linear in A, so it can be represented by

$$v_{w,M}(t_n,\mathbf{A}) = \sum_{m=0}^{M-1} a_m u_m(t_n) \tag{6}$$

where the $u_m(t)$ are a set of (not necessarily orthogonal) basis functions. An important class of $v_{w,M}(t,\mathbf{A})$ that we will discuss consist of polynomials, for which $u_m(t_n) = t_n{}^m$.

The observable residual error $v_j(t_n)$ is thus

$$v_j(t_n) = v(t_n) - v_{w,M}(t_n,\mathbf{A}) \tag{7}$$

and its MS or variance is given by

$$\sigma_{v-j}^2 = \mathrm{MS}\{v_j(t_n)\}\,. \tag{8}$$

In this paper, we will define $v_j(t_n)$ above as the jitter. The true function error is given by

$$v_w(t_n) = v_{w,M}(t_n, \mathbf{A}) - v_c(t_n) \tag{9}$$

and its MS or variance is given by

$$\sigma^2_{v-w} = MS\{v_w(t_n)\}. \tag{10}$$

For this paper, we will define $v_w(t_n)$ above as the wander, which one can see is not directly observable variable, since one must know either $v_c(t)$ or $v_p(t)$ to generate it from the data. More will be said about this later.

We note, from (4) and (7) and $v_w(t_n)$, that we can write

$$v_p(t) = v_j(t) + v_w(t). \tag{11}$$

Thus, $v_j(t_n)$ and $v_w(t_n)$ sum together to form the total noise just as conventional jitter and wander do. Later in the paper, we will show that $v_j(t_n)$ and $v_w(t_n)$ have HP and LP properties similar to those in the old definitions **[4,11,12]**, but with their HP and LP properties completely determined by $H_s(f)$ and the LSQF estimation process.

M[TH]-ORDER DIFFERENCES

We note that $\Delta(\tau)^M v(t_n)$ can be written in expanded form as **[27]**

$$\Delta(\tau)^M v(t_n) = \sum_{m=0}^{M} c(M,m) v(t_n + m\tau) \tag{12}$$

where

$$c(M,m) = \frac{(-1)^{M-m} M!}{m!(M-m)!}. \tag{13}$$

Thus, λ_M in (3) can be written as

$$\lambda_M = \sum_{m=0}^{M} c(M,m)^2. \tag{14}$$

As an additional note, (3) and (14) correct a typographical error in **[27]** in which the upper limit of the sum was mistakenly written as $M-1$.

DEFINITION OF MEAN SQUARE OPERATION

The MS operation for statistics of variances will be defined as

$$MS\{z(t_n)\} = \sum_{n=0}^{N-1} \xi_n \, | z(t_n) |^2. \tag{15}$$

This can represent various types of mean square operations depending on the values of $\xi_{u,n}$ and the form of $h_s(t)$. The unweighted biased MS is, thus, given by $\xi_n = N^{-1}$, and the "unbiased" MS is given by $\xi_n = (N-M)^{-1}$. One can also use various combinations of ξ_n and $h_s(t)$ to represent overlapping,

modified, or total averaging MS operations **[11,14,26]**. In theoretical variance representations of the MS operation (see Appendix A), we will assume the ensemble average or expectation operator $E\{\ldots\}$ has been applied in addition to the MS operation defined in (15).

THE HIGHPASS FILTERING OF NOISE DUE TO INFORMATION EXTRACTION

In this section, we will first explain intuitively and graphically how and why the LSQF estimation of $v_c(t)$ using $v_{w,M}(t,\mathbf{A})$ HP filters the noise in the residual error; then we will formally prove this assertion. For the intuitive explanation, consider Figure 4, where we show LQSF solutions for various power-law noise indices p. For $p=0$ (white noise), one can see that the solution behaves in the classical manner **[1]**, with $v_{w,M}(t,\mathbf{A})$ closely tracking $v_c(t)$. Using classical LSQF theory **[1]**, one can show that $v_{w,M}(t,\mathbf{A}) \to v_c(t)$ with $\sigma_{v-w} \to 0$ as $N \to \infty$ as long as the bandwidth of the data is such that the data samples remain uncorrelated for any value of N. For $p = -2$ and -4 (neg-p noise), however, one can see there are significant systematic long-term deviations in $v_{w,M}(t,\mathbf{A})$ from $v_c(t)$ for the large N case shown. This is due to the highly correlated nature of neg-p noise. In fact, using LSQF theory for correlated noise, one can show that these deviations will remain non-zero as $N \to \infty$, because $v_{w,M}(t,\mathbf{A})$ will track components in the noise with Fourier frequency f approximately equal or less than $f_T = 1/T$ (for an unweighted LSQF), regardless of the value of N. (It is noted that T is fixed as N is varied and, for a weighted LSQF, that $f_T = 1/T_{eff}$, where T_{eff} is determined by the weights ξ_n as well as the total time T.) This tracking of low-frequency (LF) noise arises because of the well-known inability of an LSQF to distinguish between causal behavior and noise that is correlated over the measurement interval **[1]**. It is this tracking that causes the noise to be HP filtered in $v_j(t)$ and σ^2_{v-j} with an HP cutoff knee at approximately f_T. One should point out that this LF tracking also occurs for white noise, but is only apparent for neg-p noise, because virtually all the power in the neg-p noise is in Fourier components with $f < 1/T$ (for any value of T).

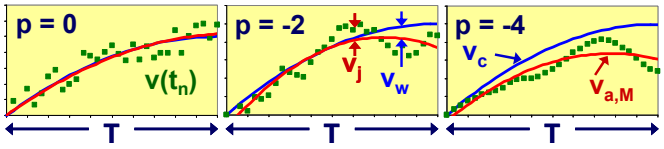

Figure 4. Simulated least squares solutions for various p values.

One can write a spectral representation for σ^2_{v-j} as

$$\sigma^2_{v\text{-}j} = 2\int_0^\infty L_v(f)|H_s(f)|^2 K_{v\text{-}j}(f)df + \sigma^2_{v\text{-}c}. \qquad (16)$$

The left term in (16) is a previously published spectral integral **[5,6,27]** that describes the $L_v(f)$-dependent part of σ^2_{v-j}. In this paper, a new term, $\sigma^2_{v\text{-}c}$, has been added in (16) to include the effects of model error. This term arises when the complexity of $v_{w,M}(t,\mathbf{A})$ is not sufficient to follow the variations in $v_c(t)$ over the observation interval T (See Appendix A for more detail.). In the integral part of (16),

there are two factors, $|H_s(f)|^2$ and $K_{v-j}(f)$, that can HP filter $L_v(f)$. As discussed in the Mathematical Niceties Section, $H_s(f)$ represents the explicit filtering action of the system under consideration on the data variable $v(t)$ **[5,6,27]**. $|H_s(f)|^2$ in (16) replaces the simple LP cut-off f_h used in previous formulations of the spectral integral to model the system **[8]**. $|H_s(f)|^2$ is a more accurate representation of a system's specific filtering properties than f_h and can be shown to have HP as well as LP behavior for many types of systems **[5,6,27]**. The importance of such HP filtering from $|H_s(f)|^2$ is that it helps σ^2_{v-j} converge in the presence of neg-p noise and, in fact, in and of itself can ensure the convergence of σ^2_{v-j} for some or all of the common neg-p values **[5,6,27]**.

A classic example of a system response function is the "delay" $H_s(f)$ shown in Figure 5. This well-known response function arises when one mixes a signal with a delayed version of itself, as in a delay line discriminator, radar system, or two-way ranging system **[27]**. One can show that $|H_s(f)|^2 = 4\sin^2(\pi f \tau_d)$ for such a system **[27]**. When $f << 1$, this $|H_s(f)|^2$ is proportional to f^2, which by itself allows σ^2_{v-j} to converge for neg-p noise with $p \geq -2$.

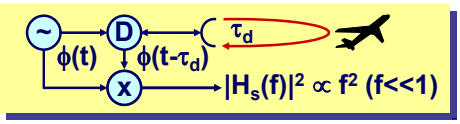

Figure 5. Delay system response function.

$K_{v-j}(f)$ is a spectral kernel that describes the spectral properties of the LSQF and MS generation process independent of the system response filtering **[6,27]**. In Appendix A, exact formulas are derived for $K_{v-j}(f)$ and an equivalent $K_{v-w}(f)$ for σ^2_{v-w} in terms of a spectral decomposition of the $v_{w,M}(t,\mathbf{A})$ basis functions, $|H_s(f)|^2$, and $L_v(f)$. Also derived in Appendix A is a similar kernel for σ^2_{v-c} in terms of a dual-frequency PSD for the nonstationary $v_c(t)$. In the next subsection, we will prove that

$$K_{v-j}(f) \propto f^{2M} \quad (fT << 1) \text{ when } v_{a,M}(t,\mathbf{A}) \text{ is an } (M\text{-}1)^{\text{th}}\text{-order polynomial} \tag{17}$$

$$K_{v-j}(f) \propto f^{p'} \quad (fT << 1 \text{ and } p' \geq 2) \text{ when } v_{a,M}(t,\mathbf{A}) \text{ is any function with a DC component.} \tag{18}$$

Figure 6 shows the results of a simulation verifying that (17) is indeed true for $M = 1$ to 5 and $N = 1000$. This verifies that $K_{v-j}(f)$ has the required HP filtering properties for $fT << 1$ to ensure that σ^2_{v-j} will converge for any neg-p value, if one can choose the form of $v_{w,M}(t,\mathbf{A})$.

Proof of Noise HP Filtering for Residual LSQF Error

Let us now prove assertions (17) and (18). We will prove (17) by decomposing the data $v(t_n)$ into components

$$v(t_n) = \sum_f v_f(t_n) + v_c(t_n) \tag{19}$$

where the $v_f(t) = V_f e^{j2\pi ft}$ are single-frequency noise components that sum to generate the total noise $v_p(t_n)$. Because of the linearity of $v_{a,M}(t,\mathbf{A})$ in $\mathbf{A}$ given by (6), we note that the LSQF solution $v_{w,M}(t,\mathbf{A})$ for the total input can be decomposed into the sum of LSQF solutions for the separate $v_f(t_n)$ and $v_c(t_n)$ inputs, or

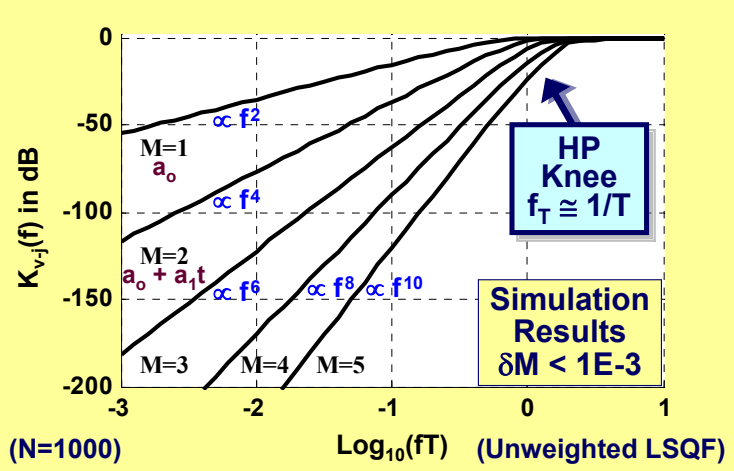

Figure 6. Simulation showing $K_{v-j}(f) \propto f^{2M}$ for $fT << 1$ for an $(M-1)^{th}$-order polynomial model function (and an unweighted LSQF).

$$v_{w,M}(t,\mathbf{A}) = \sum_f v_{w,M}(t,\mathbf{A}^{(f)}) + v_{w,M}(t,\mathbf{A}^{(c)}) \ . \tag{20}$$

Also, because the spectral noise components $v_f(t)$ for a wide-sense stationary noise process $v_p(t)$ are uncorrelated with each other **[10]** and with $v_c(t_n)$, we can use (20) to write σ^2_{v-j} as

$$\sigma^2_{v-j} = E\{MS\{|\, v_j(t_n)\,|^2\}\} = E\{MS\{|\, v_{j-c}(t_n)\,|^2\}\} + \sum_f E\{MS\{|\, v_{j-f}(t_n)\,|^2\}\} \tag{21}$$

where

$$v_{j-f}(t_n) = v_f(t_n) - v_{w,M}(t_n,\mathbf{A}^{(f)}) \quad v_{j-c}(t_n) = v_c(t_n) - v_{w,M}(t_n,\mathbf{A}^{(c)}) \tag{22}$$

and where $E\{..\}$ the expectation value has been added to the MS operation in order to generate σ^2_{v-j} in terms of the PSD $L_v(f)$ (see Appendix A). Thus, since (16) is just the infinitesimal limit of (21), the spectral properties of $K_{v-j}(f)$ in (16) can be determined by considering the LSQF properties of each $E\{MS\{|\, v_{j-f}(t_n)\,|^2\}\}$ term separately.

To do so, let us expand $V_f e^{j2\pi ft}$ using the well-known Taylor Theorem as

$$v_f(t) = V_f e^{j2\pi ft_0} \sum_{k=0}^{M-1} \frac{(j2\pi f(t-t_0))^k}{k!} + V_f \frac{(j2\pi f(t'-t_0))^M}{M!} \tag{23}$$

where t' is somewhere in $[t_0,t]$. We then note that the residual error $v_{j-f}(t_n)$ and, hence, $E\{MS\{|v_{j-f}(t_n)|^2\}\}$, would be zero, if $v_f(t)$ were given only by the right-hand finite sum term in (23). This is because the model function $v_{w,M}(t,\mathbf{A}^{(f)})$ and the finite sum term would then both be (M-1)th-order polynomials, so the fitted $v_{w,M}(t,\mathbf{A}^{(f)})$ would be exactly $v_f(t)$. Thus, when the Taylor series converges, the value of $E\{MS\{|v_{j-f}(t_n)|^2\}\}$ must be proportional to the square magnitude of the right-hand term in (23). This term is proportional to f^{2M} and, therefore, we must have $K_{v-j}(f) \propto f^{2M}$ for $f(t'-t_0) << 1$ or $fT << 1$, which is just (17).

To prove (18), we note that a DC component is a 0th-order polynomial. Therefore, by using (17) with M=1, $K_{v-j}(f)$ must be at least proportional to f^2 for $fT << 1$.

HP and LP Properties of Jitter and Wander Defined as Residual LSQF Errors

Now, let us discuss the HP and LP properties of jitter and wander defined as $v_j(t_n)$ and $v_w(t_n)$. The exact HP and LP properties of σ^2_{v-j} and σ^2_{w-j} can be derived from the formulas for $K_{v-j}(f)$ and $K_{v-w}(f)$ given in Appendix A. Here, we will discuss their overall nature. As discussed in the previous subsections, the LQSF causes $v_j(t_n)$ to be HP filtered with a knee frequency $f_T \cong 1/T$. Similarly, one can show that the LSQF causes $v_w(t_n)$ to be LP filtered with the same knee frequency. This LSQF filtering at f_T is shown in the left side of Figure 7. In addition, $H_s(f)$ filters both $v_j(t_n)$ and $v_w(t_n)$ equally, since $H_s(f)$ has the same effect on all variables. In the left side of Figure 7, this $H_s(f)$ filtering is shown parametrically using an HP knee f_l and a LP knee f_h. Thus, the brick-wall filtering properties of the jitter as $v_j(t_n)$ are determined by an HP knee f_c that is the higher of f_T and f_l and a LP knee given by f_h, which are purely functions of system parameters. We also note that the equivalents of conventional x-jitter and x-wander **[4,11,12]** are $x_j(t_n)$ and $x_w(t_n)$ with a 2nd-order time error polynomial removed. This guarantees that the x-jitter will converge for $p \geq -6$ in $L_x(f)$ without any help from $H_s(f)$, and thus guarantees the convergence of the x-jitter for all the neg-p noise components normally encountered.

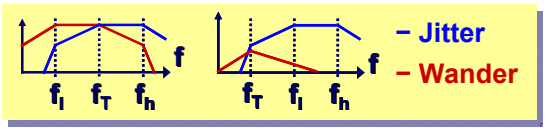

Figure 7. Jitter ($v_j(t_n)$) and wander ($v_w(t_n)$) HP & LP properties.

The right side of Figure 7 shows that the wander will disappear when $T \to \infty$ ($f_l >> f_T$) and all that will remain is the jitter, if the HP order of $H_s(f)$ is sufficient by itself to overcome the pole in $L_v(f)$ — That is, if $H_s(f)$ by itself guarantees the convergence of the wander. This case is the transition to stationary statistics, because the total $H_s(f)$ filtered noise variance must then be convergent for a finite bandwidth system. Furthermore, as mentioned previously, these definitions of jitter and wander as $v_j(t_n)$ and $v_w(t_n)$ allow these concepts to be generalized to any type of causal function removal and any variable.

WHAT TO DO WHEN THE RESIDUAL ERROR VARIANCE DOES DIVERGE

σ^2_{v-j} can diverge in the presence of neg-p noise when the problem being addressed fixes the form of the $v_{w,M}(t,\mathbf{A})$. In this paper, it is maintained: (a) that such a divergence is an indication of a real problem in the design, specification, or analysis of the system under consideration, and (b) that this real problem must be investigated and fixed, not sidestepped. From the discussions in this paper, it is obvious that the HP filtering properties of σ^2_{v-j} are fixed by $v_{w,M}(t,\mathbf{A})$, the σ^2_{v-j} measurement interval T, and $H_s(f)$ as given by the system specification (or problem definition) and design. Thus, such a divergence must indicate: (a) that something is essentially wrong with the system design or spec, or (b) that the system is okay, but faulty analysis generated a perceived (non-essential) divergence. For Case (a), the system itself has to be changed to correct the problem, and for Case (b), the system does not have to be changed to correct the problem, just the improper analysis.

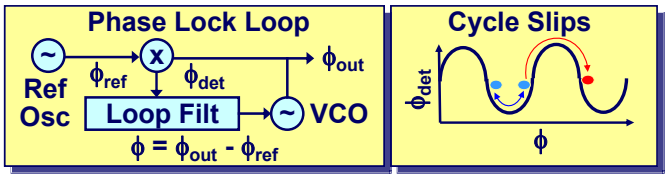

Figure 8. Cycle slips in a 1st-order phase lock loop due to f^{-3} noise.

The best way to understand how to deal with such divergences is to use the specific example of an essential divergence shown in Figure 8. Here, a 1st-order phase-lock loop (PLL) is operating using a reference oscillator with f^{-3} phase noise. As shown in the figure, such a PLL will cycle slip because of the f^{-3} noise **[27]**. An indication of these cycle slips appears in the linear loop analysis as a divergence in $\sigma^2_{\phi-j}$ for $M=0$ (no $v_{w,M}(t,\mathbf{A})$ removal), where ϕ is defined in Figure 8. $\sigma^2_{\phi-j}$ diverges in the linear analysis because $|H_s(f)|^2$ for a 1st-order PLL is proportional to f^2 for $f << 1$ and $K_{\phi-j}(f)=1$ for $M=0$, and this combination of HP orders are not insufficient to overcome the f^{-3} pole in $L_\phi(f)$. One could increase M in $\sigma^2_{\phi-j}$ to resolve the divergence in the linear model ($K_{\phi-j}(f) \propto f^2$ or higher), but this will not keep the PLL from cycle slipping! Thus, one can see that arbitrarily changing the error measure to eliminate the divergence in the analysis does nothing to solve the actual cycle-slipping problem.

One can fix the real problem in two ways. First, one can change the design and eliminate the cycle slips altogether. This is accomplished simply by changing to a 2nd-order PLL, for which $|H_s(f)|^2 \propto f^4$ $(f << 1)$ **[27]**. Second, one can allow occasional cycle slips, because the system users can tolerate them. However, one must then change the system spec so the phase noise without the slips can be properly measured. This is accomplished by specifying that $\sigma^2_{\phi-j}$ is to be measured excluding data containing cycle slips, which effectively changes $K_{\phi-j}(f)$. In addition, one should also include a mean time to cycle slip requirement in the spec to ensure that the cycle slips don't become a nuisance. An example of a non-essential divergence is simply the failure to recognize the HP filtering due to causal extraction in σ^2_{v-j}.

RELATING $\sigma^2_{v,M}(\tau)$ TO σ^2_{v-j} WHEN $v_{a,M}(t,\mathbf{A})$ IS A POLYNOMIAL

For $N = M+1$ data points, Appendix B shows that

$$\sigma^2_{v-j}(N = M+1, M) = \sigma^2_{v,M}(T/M) \qquad (24)$$

when $\sigma^2_{v-j}(N,M)$ uses the unweighted "unbiased" MS ($\xi_n = (N-M)^{-1} = 1$) and $v_{w,M}(t,\mathbf{A})$ is an $(M-1)^{th}$-order polynomial. For M=1, (24) is just the well-known statement that the Allan variance of $y(t)$ is the two-sample variance when $h_s(t)$ is a box-car average over τ **[13,23]**, that is $\sigma^2_{y-j}(2,1)$. One can also analytically demonstrate that (24) is true for the M=2 case. Thus, the Hadamard-Picinbono variance is equal to $\sigma^2_{y-j}(3,2)$ (when $h_s(t)$ is a box-car average over τ), that is, when frequency offset and drift are modeled in the causal behavior.

As shown in Figure 9, one can extend (24) to any number of samples by investigating the behavior of LSQF simulations as N is varied (while T and M remain fixed). In the figure, we've plotted the biased form of the residual error deviate indicated by $RMS\{v_j\}$ ($\xi_n = N^{-1}$) versus the number of samples N. One can see from the figure that $RMS\{v_j\}$ does not vary much as N increases from $N = M+1$ to very large values, especially for neg-p noise. Using this approximate invariability of $RMS\{v_j\}$ as N is changed, one can then write

$$\sigma^2_{v-j}(N,M) \cong \frac{N}{N-M}\sigma^2_{v,M}(T/M) \text{ (any N)} \qquad (25)$$

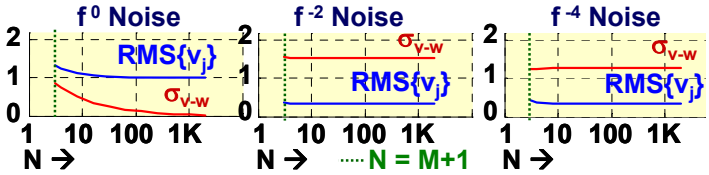

Figure 9. Errors in LSQF residuals as N is varied (fixed T and M = 2).

We also note that one can obtain exact expressions similar to (25) for any specific p by deriving M^{th}-order "bias" function relationships in a fashion similar to those derived by Allan **[19]** and Barnes **[20]** for the $M = 1$ case. One should note, however, for $M = 1$, that our bias functions have inherently different behavior than those of Allan and Barnes. This is because we fix both T and τ ($T = M\tau$) as N is varied, while Allan and Barnes fix τ as N is varied and let the total observation interval ($N\tau$) change with N

(Allan and Barnes define T as the time between successive t_n samples, not the total observation interval.). Thus, our bias functions are very close to $N/(N-M)$ for all N and p, unlike the Allan-Barnes bias functions, which vary widely with p.

From the above, one can see that $\sigma^2_{v,M}(T/M)$ can be interpreted as a measure of the MS residual error $\sigma^2_{v-j}(N,M)$ for any N when $v_{w,M}(t,\mathbf{A})$ is an $(M-1)^{th}$-order polynomial. This has several important consequences, which are discussed as follows:

(a) $\sigma^2_{v,M}(\tau)$ can be used to determine $\sigma^2_{v-j}(N,M)$ in residual error problems when $v_{w,M}(t,\mathbf{A})$ is an $(M-1)^{th}$-order polynomial. Thus, $\sigma^2_{v,M}(\tau)$ and $\sigma^2_{v-j}(N,M)$ can be viewed as equivalent error measures for these problems.

(b) In such residual error problems, there is a great advantage in using $\sigma^2_{v,M}(T/M)$ for $\sigma^2_{v-j}(N,M)$, because one need not perform the LQSF and remove $v_{w,M}(t,\mathbf{A})$ from the raw data in order to generate $\sigma^2_{v,M}(T/M)$. This is because of the well-known insensitivity of $\sigma^2_{v,M}(\tau)$ to $(M-1)^{th}$- or lower-order polynomial behavior **[14,25]**. We further note that $\sigma^2_{v,M}(T/M)$ without $v_{w,M}(t,\mathbf{A})$ removal remains equivalent to $\sigma^2_{v-j}(N,M)$ with such removal even when there is model error. This is because model error equally effects $\sigma^2_{v,M}(T/M)$ and $\sigma^2_{v-j}(N,M)$.

(c) (25) provides guidance about which orders of $\sigma^2_{v,M}(\tau)$ are appropriate when it is meant to measure purely random error. The interpretation of $\sigma^2_{v,M}(\tau)$ as a measure of $\sigma^2_{v-j}(N,M)$ shows, if one wants $\sigma^2_{v,M}(\tau)$ to measure only random error when $v_{w,M}(t,\mathbf{A})$ can't follow the variations in $v_c(t)$ over T, that one must remove such causal behavior first. This equivalence of $\sigma^2_{v,M}(\tau)$ and $\sigma^2_{v-j}(N,M)$ is only strictly true only when $\tau = T/M$. However, for τ decoupled from T/M, one can assume that the same approximate sensitivity applies. Thus, this interpretation explains the well-known insensitivity of the Hadamard-Picinbono variance to frequency drift and the sensitivity of the Allan variance to such drift.

(d) Conventional x-jitter as defined in **[4-6]** (but without an ad-hoc f_c) is equivalent to the Hadamard-Picinbono variance of x(t) ($\sigma^2_{x,3}(\tau)$).

CONCLUSIONS

We have demonstrated that $\sigma^2_{v,M}(\tau)$, $\sigma^2_{v-j}(N,M)$, and the MS jitter (without ad-hoc filtering) can all be viewed as essentially measuring the same type of error. The key to this is the demonstration of three major facts: (a) that a statistically optimal removal of the causal behavior in the data HP filters the noise in the observable residual error, (b) that the order of the noise HP filtering is a function of the complexity of the model function used to estimate the causal behavior, and (c) that $\sigma^2_{v,M}(T/M)$ is a measure of $\sigma^2_{v-j}(N,M)$ when the model function used to estimate the causal behavior is an $(M-1)^{th}$-order polynomial.

Table 1 shows the consequences of interpreting M = 0, 1, 2, and 3 $\sigma^2_{x,M}(\tau)$ and $\sigma^2_{y,M}(\tau)$ as σ^2_{x-j} and σ^2_{y-j} with aging removed. An important conclusion shown in the table is that low-order variances such as $\sigma^2_{x,0}(\tau)$, $\sigma^2_{x,1}(\tau)$, and $\sigma^2_{y,0}(\tau)$ are appropriate for characterizing "random" error in coherent frequency synthesizers and coherent time and frequency distribution equipment (excluding the frequency reference). This is because uncontrolled time or frequency offsets that are fixed over the measurement interval are part of the "random" error that must be considered in specifying such devices; that is, these devices are not supposed to have such uncontrolled, but fixed, offsets. On the other hand, such fixed offsets are generally not considered part of the "random" instabilities in oscillators, where higher-order difference variances are used, but are modeled as causal error or compensated for using PLLs or other similar techniques. Therefore, the precision oscillator community uses higher-order difference variances as measures of "random" error. This difference in application explains the dichotomy in the use of variances between some producers of time and frequency distribution equipment **[4]** and producers of precision oscillators **[13]**.

As a final note, consider the difference between σ_{v-w} and $\mathrm{RMS}\{v_j\}$ in Figure 9. The simulations in this case utilized an $H_s(f)$ with $f_l << f_T$, so the theoretical wander is effectively divergent. We see from the figure that $\sigma_{v-w} > \mathrm{RMS}\{v_j\} \cong \sigma_{v-j}$ (for large N, using (25)). Thus, the observable error is underestimating the true function error, as is expected from correlated LSQF theory. In fact, in running multiple Figure 9 simulation sessions, one sees σ_{v-w} vary widely from run to run, while $\mathrm{RMS}\{v_j\}$ or σ_{v-j} remains stable.

Table 1. Difference variances of y(t) and x(t) as residual error variances with aging removed.

M	Δ Var of y	Aging Excluded Application	Δ Var of x	Aging Excl Application
0	MS{y}	None Synthesizers & Rel time dist equip	MS{x}	None Abs time dist equip
1	Allan y	y offset Oscillators (y drift in instability)	TIErms²/2 MS{TIE}/2	x offset Synth & rel time dist
2	Hadamard Picinbono	y ofs & drift Oscillators (y drift not in instability)	Allan x Jitter 2 [28]	x & y offset Osc (y drift in instab)
3	$\sigma_{x,3}^2(\tau)$ is equivalent to MS of x or TIE jitter with time & freq offset & freq drift removed		Hadamard Picinbono	x,y ofs y drift Osc (y drift not in instab)

Thus, in order to get a reliable estimate of σ_{v-w} over T, one must take data over periods much longer than T and determine the power law structure of the true noise $v_p(t)$ (except for the case where $H_s(f)$ alone guarantees convergence and $f_l > f_T$). Then one can use the formulas in Appendix A to estimate the true expectation average of σ_{v-w} using $L_v(f)$ accurately determined from the data over the period >> T. This is where the classic techniques involving the use of the modified Allan variance or direct spectral measurements are invaluable **[13]**. Thus, the final conclusion of this paper is the well-known fact in the PTTI community that one must perform careful analysis and measurements over periods >> T to determine good estimates of the true causal function error or true residual error when neg-p noise is

involved. The exception for this is again the case where $H_s(f)$ alone guarantees convergence and $f_l > f_T$. In this case, the wander is not divergent and the $H_s(f)$ is such that the estimate of the causal function from data over T is a good one.

APPENDIX A. DERIVATION OF THE SPECTRAL KERNELS

In this section, we will derive the kernels $K_{v-j}(f)$ and $K_{v-w}(f)$ for σ^2_{v-j} and σ^2_{v-w} and a dual-frequency kernel $K_{v-c}(f_g, f)$ for σ^2_{v-c}. For generality, we will let $v(t)$ and $v_c(t)$ be complex and the weights ξ_n be arbitrary. To generate these kernels, we will first minimize χ^2 the sum of the squares weighted by the ξ_n, which can be written in matrix notation as

$$\chi^2 = << (v(t') * - \mathbf{A}^\dagger \mathbf{U}^\dagger(t'))(v(t') - \mathbf{U}(t')\mathbf{A}) >> . \tag{A.1}$$

In the above:

(a) $\mathbf{A}^\dagger$ is the complex conjugate transpose of the M-element column vector $\mathbf{A} = (a_0, a_1, ...a_{M-1})'$ (' is the transpose).

(b) $\mathbf{U}(t') = (u_0(t'), u_1(t'), ...u_{M-1}(t'))$ is an M-element row vector representing the M basis functions $u_m(t)$ for $v_{w,M}(t, \mathbf{A})$ in (6), and we note that $v_{w,M}(t, \mathbf{A})$ in this notation is

$$v_{w,M}(t, \mathbf{A}) = \mathbf{U}(t)\mathbf{A} \ . \tag{A.2}$$

(c) <<…>> is the weighted average over the data samples (denoted by the dummy time index t'), which, for the purposes of relating the LSQF to continuous Fourier transform PSDs, <<…>> is defined as

$$<< z(t') >> = \int_{-\infty}^{+\infty} dt' \rho(t') z(t') \tag{A.3}$$

where the density function $\rho(t)$ is given by

$$\rho(t) = \sum_{n=0}^{N-1} \xi_n \delta(t - t_n) . \tag{A.4}$$

(d) $v(t)$, $v_p(t)$, and $v_c(t)$ are all assumed to be filtered by the system response function $h_s(t)$ as in (5).

In the well-known manner, we differentiate χ^2 with respect to $\mathbf{A}^\dagger$ to obtain the following LQSF solution

$$\mathbf{A} = \mathbf{Q} << \mathbf{U}(t')^\dagger v(t') >> \tag{A.5}$$

$$v_{w,M}(t, \mathbf{A}) = \mathbf{U}(t)\mathbf{Q} << \mathbf{U}(t')^\dagger v(t') >> \tag{A.6}$$

where

$$\mathbf{Q} = << (\mathbf{U}(t')^\dagger \mathbf{U}(t')) >>^{-1} . \tag{A.7}$$

Using (A.3) and (A.4), one can write (A.6) as

$$v_{w,M}(t, \mathbf{A}) = \int_{-\infty}^{+\infty} dt' g_w(t, t') v(t') = \int_{-\infty}^{+\infty} df \, G_w(t, f) V(f) H_s(f) \tag{A.8}$$

where: (a) the Green's function $g_w(t,t')$ is

$$g_w(t,t') = \mathbf{U}(t)\mathbf{Q}\mathbf{U}(t')^\dagger \rho(t'). \tag{A.9}$$

(b) $G_w(t,f)$ is the complex conjugate of the Wigner-Ville spectrum **[29]** of $g_w(t,t')$ is

$$G_w(t,f) = \Im_{-f,t'}\{g_w(t,t')\} \equiv \int_{-\infty}^{+\infty} dt\, e^{j2\pi ft} g_w(t,t') \tag{A.10}$$

and (c) $V(f)H_s(f)$ is the Fourier transform of $v(t)$ is

$$V(f)H_s(f) = \Im_{f,t}\{v(t)\} \equiv \int_{-\infty}^{+\infty} dt\, e^{-j2\pi ft} v(t) \tag{A.11}$$

Note that $V(f)$ alone is the Fourier transform of the time domain variable before being convoluted with $h_s(t)$, so that $H_s(f)$ is explicitly shown in the right side of (A.8).

We note from (A.2) that $v_{w,M}(t,\mathbf{A})$ is linear in $\mathbf{A}$, and thus the LSQF solutions (A.5) and (A.6) are linear in $v(t)$. Therefore, we can write

$$v_{w,M}(t,\mathbf{A}) = v_{w,M}(t,\mathbf{A}^{(p)}) + v_{w,M}(t,\mathbf{A}^{(c)}) \tag{A.12}$$

where

$$\mathbf{A}^{(p)} = \mathbf{Q} << \mathbf{U}(t')^\dagger v_p(t') >> \quad \mathbf{A}^{(c)} = \mathbf{Q} << \mathbf{U}(t')^\dagger v_c(t') >>. \tag{A.13}$$

Using this linear separation, (A.8), and (A.12), we can write

$$\begin{aligned} &v_j(t) = v_{j,p}(t) - v_{j,c}(t) \quad v_w(t) = v_{w,p}(t) + v_{j,c}(t) \\ &v_{w,p}(t) + v_{j,p}(t) = v_w(t) + v_j(t) = v_P(t) \end{aligned} \tag{A.14}$$

where

$$g_j(t,t') = \delta(t-t') - g_w(t,t') \quad G_j(t,f) = \Im_{-f,t'}\{g_j(t,t')\} = e^{j\omega t} - G_w(t,f) \tag{A.15}$$

$$\begin{aligned} v_{w,p}(t) &= v_{w,M}(t,\mathbf{A}^{(p)}) = \int_{-\infty}^{+\infty} dt' g_w(t,t') v_p(t') = \int_{-\infty}^{+\infty} df\, G_w(t,f) V_p(f) H_s(f) \\ v_{j,p}(t) &= v_p(t) - v_{w,M}(t,\mathbf{A}^{(p)}) = \int_{-\infty}^{+\infty} dt' g_j(t,t') v_p(t') = \int_{-\infty}^{+\infty} df\, G_j(t,f) V_p(f) H_s(f) \\ v_{j,c}(t) &= v_{w,M}(t,\mathbf{A}^{(c)}) - v_c(t) = -\int_{-\infty}^{+\infty} dt' g_j(t,t') v_c(t') = -\int_{-\infty}^{+\infty} df\, G_j(t,f) V_c(f) H_s(f) \end{aligned} \tag{A.16}$$

and where $V_p(f)$ and $V_c(f)$ are defined in similar fashion as the definition of $V(f)$ in (A.11) (that is, before $h_s(t)$ is applied). We note from (A.14) and (A.16) that $v_{j,c}(t)$ is the model residual and $v_{j,p}(t)$ and $v_{w,p}(t)$ are the jitter and wander solely due to $v_p(t)$, which add together to produce $v_p(t)$ just like the total jitter and wander $v_j(t)$ and $v_w(t)$.

By taking the ensemble average $E\{\ldots\}$ of the square of $v_j(t)$ and $v_w(t)$ from (A.14) and (A.16) and assuming $v_p(t)$ is uncorrelated with $v_c(t)$, one can write

$$\begin{aligned} E\{|\,v_j(t)\,|^2\} &= E\{|\,v_{j,p}(t)\,|^2\} + E\{|\,v_{j,c}(t)\,|^2\} \\ E\{|\,v_w(t)\,|^2\} &= E\{|\,v_{w,p}(t)\,|^2\} + E\{|\,v_{j,c}(t)\,|^2\} \end{aligned} \tag{A.17}$$

where

$$\begin{aligned} &E\{|\,v_{j,p}(t)\,|^2\} = \int_{-\infty}^{+\infty} df\,|\,G_j(t,f)H_s(f)\,|^2\,L_v(f) \quad E\{|\,v_{w,p}(t)\,|^2\} = \int_{-\infty}^{+\infty} df\,|\,G_w(t,f)H_s(f)\,|^2\,L_v(f) \\ &E\{|\,v_{j,c}(t)\,|^2 = \int_{-\infty}^{+\infty} df_g \int_{-\infty}^{+\infty} df\,L_c(f_g,f)G_j(t,f+0.5f_g)G_j^*(t,f-0.5f_g)H_s(f+0.5f_g)H_s^*(f-0.5f_g) \end{aligned} \tag{A.18}$$

with

$$L_v(f) = \Im_{f,\tau}\{R_v(\tau)\} \quad L_c(f_g,f) = \Im_{f_g,t_g}\{\Im_{f,\tau}\{R_c(t_g,\tau)\}\} \tag{A.19}$$

$$R_v(\tau) = E\{v_p(t_g+\tau/2)v_p^*(t_g-\tau/2)\} \quad R_c(t_g,\tau) = E\{v_c(t_g+\tau/2)v_c^*(t_g-\tau/2)\}\,. \tag{A.20}$$

In the above, is $L_c(f_g,f)$ is the rotated Loève spectrum **[29,30]** of $v_c(t)$ given by the double Fourier transform of the rotated double time autocorrelation function $R_c(t_g,\tau)$. The term rotated comes from writing the lag function $v_c(t_1)v_c^*(t_2)$ in terms of the "rotationally" transformed global time $t_g = 0.5(t_1+t_2)$ and local or differential time $\tau = (t_1 - t_1)$. When v(t) is real, we note that

$$\begin{aligned} &R_v(-\tau) = R_v(\tau) \qquad R_c(t_g,-\tau) = R_c(t_g,\tau) \\ &L_v(f) = L_v(-f) = L_v^*(f) \quad L_c(f_g,f)^* = L_c(-f_g,f) \quad L_c(f_g,-f) = L_c(f_g,f) \quad \text{[v(t) real]} \\ &G_w(t,f)^* = G_w(t,-f) \quad G_j(t,f)^* = G_j(t,-f) \end{aligned} \tag{A.21}$$

If we now integrate (A.18) over t from $-\infty$ to ∞, we obtain our final result

$$\begin{aligned} &\sigma_{v-j,p}^2 = \int_{-\infty}^{+\infty} df\,K_{v-j}(f)\,|\,H_s(f)\,|^2\,L_\beta(f) \qquad \sigma_{v-j}^2 = \sigma_{v-j,p}^2 + \sigma_{v-c}^2 \\ &\sigma_{v-w,p}^2 = \int_{-\infty}^{+\infty} df\,K_{v-w}(f)|\,H_s(f)\,|^2\,L_\beta(f) \qquad \sigma_{v-w}^2 = \sigma_{v-w,p}^2 + \sigma_{v-c}^2 \\ &\sigma_{v-c}^2 = \int_{-\infty}^{+\infty} df_g \int_{-\infty}^{+\infty} df\,K_{v-c}(f_g,f)H_s(f+0.5f_g)H_s^*(f-0.5f_g)\,L_c(f_g,f) \end{aligned} \tag{A.22}$$

Where our kernels are

$$\begin{aligned} &K_{v-j}(f) = \int_{-\infty}^{+\infty} dt\,|\,G_j(t,f)\,|^2 \quad K_{v-w}(f) = \int_{-\infty}^{+\infty} dt\,|\,G_w(t,f)\,|^2 \\ &K_{v-c}(f_g,f) = \int_{-\infty}^{+\infty} dt\,G_j(t,f+0.5f_g)G_j^*(t,f-0.5f_g) \end{aligned} \tag{A.23}$$

APPENDIX B. VERIFYING THAT $\sigma_{v-j}^2(M+1) = \sigma_{v,M}^2(T/M)$

In this appendix, we will verify that (25) is true when σ_{v-j}^2 is the "unbiased" uniformly weighted residual variance ($\xi_n = (N-M)^{-1}$). To do this, we will show that the stronger assertion

$$v_j(t_n) = [\Delta(T/M)^M v(t_o)]c(M,n)/\lambda_M \tag{B.1}$$

is true by Monte Carlo simulation. One can show (B.1) guarantees (25) by using (12) and (B.1) as follows

$$\sigma^2_{v-j}(M+1) = \sum_{n=0}^{M} |\, v_j(t_n)|^2 = |\, \Delta(T/M)^M v(t_o)\,|^2 \; \lambda_M{}^{-2} \sum_{n=0}^{M} |\, c(M,n)|^2$$
$$= \lambda_M{}^{-1} \,|\, \Delta(T/M)^M v(t_o)\,|^2 = \sigma^2_{v,M}(T/M) \tag{B.2}$$

(B.1) has been verified by Matlab simulation for multiple random data sets (runs) up to $M = 18$. Above this M value, the Matlab LSQF code used ran into precision difficulties in the computation of $v_{w,M}(t, \mathbf{A})$. An example Matlab simulation for $M = 5$ is shown in Table B-1. Thus, one can say that (25) is true for any M with an extremely high confidence because of the detailed nature of assertion (B.1).

To understand analytically why (B.1) is true, let us reformulate the notation of Appendix A into one more suited for an unweighted LSQF. Let us define the following $N = M + 1$ element column vectors ($n = 0$ to M): $\mathbf{C}_n = c(M,n)$, $\mathbf{V}_n = v(t_n)$, $\mathbf{V}_{j,n} = v_j(t_n)$, and $\mathbf{V}_w = v_{w,M}(t_n, \mathbf{A}) = \mathbf{UA}$, where $\mathbf{U}_{n,m} = t_n^m$ (M+1 rows for index n and M columns for index $m = 0$ to $M-1$). We also note that (12) can be written as $\Delta(T/M)^M v(t_o) = \mathbf{C}^\dagger \mathbf{V}$ and (B.1) becomes

$$\mathbf{V}_j = \mathbf{V} - \mathbf{V}_w = \mathbf{C}\mathbf{C}^\dagger \mathbf{V} / \lambda_M \tag{B.3}$$

where λ_M in (14) becomes

Table B-1. Monte Carlo verification of (B.1) for M = 5.

Run →	1	2	3	4	5	6
$\Delta(T/M)^M v(t_0)$	-4.3818	-20.6668	-15.3472	-11.062	-3.9762	7.6004
$\lambda_M v_i(t_n)/c(M,n)$						
t_0	-4.3818	-20.6668	-15.3472	-11.062	-3.9762	7.6004
t_1	-4.3818	-20.6668	-15.3472	-11.062	-3.9762	7.6004
t_2	-4.3818	-20.6668	-15.3472	-11.062	-3.9762	7.6004
t_3	-4.3818	-20.6668	-15.3472	-11.062	-3.9762	7.6004
t_4	-4.3818	-20.6668	-15.3472	-11.062	-3.9762	7.6004
t_5	-4.3818	-20.6668	-15.3472	-11.062	-3.9762	7.6004

$$\lambda_M = \mathbf{C}^\dagger \mathbf{C} \tag{B.4}$$

and χ^2 from (A.1) becomes

$$\chi^2 = \mathbf{V}_j^\dagger \mathbf{V}_j = (\mathbf{V}^\dagger - \mathbf{A}^\dagger \mathbf{U}^\dagger)(\mathbf{V} - \mathbf{UA}). \tag{B.5}$$

Regenerating the LSQF solution by taking the derivative of (B.5) with respect to $\mathbf{A}^\dagger$, we obtain

$$\mathbf{U}^\dagger \mathbf{V}_j = \mathbf{U}^\dagger (\mathbf{V} - \mathbf{UA}) = 0 \tag{B.6}$$

which yields the unweighted LSQF solution

$$\mathbf{V}_w = \mathbf{U}\mathbf{Q}\mathbf{U}^\dagger\mathbf{V} \tag{B.7}$$

where $\mathbf{Q}^{-1} = \mathbf{U}^\dagger\mathbf{U}$.

If we insert (B.3) into (B.6), we obtain

$$\mathbf{U}^\dagger\mathbf{C}\mathbf{C}^\dagger\mathbf{V} = 0 \tag{B.8}$$

and note that this must be true as a necessary condition for (B.3) to be the LSQF solution. To show this, we note from (12) that $\mathbf{U}^\dagger\mathbf{C} = \mathbf{0}$ as follows

$$\mathbf{U}^\dagger\mathbf{C} = \sum_{n=1}^{M+1} U_{n,m}\, c(M,n) = \sum_{n=1}^{M-1} c(M,n)\, t_{n-1}^{m-1} = \Delta(\tau)^M\, t_0^{m-1} = \mathbf{0} \tag{B.9}$$

because the M[th]-order difference of an (M-1)[th]-order or less power of t is zero **[14]**. Thus, (B.3) satisfies this necessary condition with any arbitrary λ_M. Multiplying (B.6) by $\mathbf{A}^\dagger$ yields the well-known orthogonality principle for an LSQF, which states that the residual error $\mathbf{V}_j$ is orthogonal to the estimated function $\mathbf{V}_w$. Thus, (B.9) states that $\mathbf{C}$ is also orthogonal to $\mathbf{V}_w$, which explains why (B.3) is an LSQF solution, since $\mathbf{C}\mathbf{C}^\dagger\mathbf{V}$ in (B.3) is the projection of $\mathbf{V}$ in the $\mathbf{C}$ direction. From (B.3) and (B.7), we also note that we can write

$$\mathbf{U}\mathbf{Q}\mathbf{U}^\dagger = \mathbf{I} - \mathbf{C}\mathbf{C}^\dagger/\lambda_M \tag{B.10}$$

which is useful in simplifying the calculations for the spectral kernels in Appendix A for the unweighted $N = M + 1$ LSQF case.

In trying to analytically prove that $\lambda_M = \mathbf{C}^\dagger\mathbf{C}$ must be the unique LQSF solution, some mathematical difficulties were encountered by the author. These must be resolved before developing a general analytical proof of (B.3). Therefore, for now, one must rely on Monte Carlo simulations to verify that (B.1) or (B.3) is true.

REFERENCES

[1] J. R. Wolberg, 1967, **Prediction Analysis** (D. Van Nostrand and Co., Princeton, N.J.)

[2] IEEE Std. 181-2003, *"IEEE Standard on Transitions, Pulses, and Related Waveforms*" (IEEE, Piscataway, New Jersey), 2003.

[3] IEEE Std. 1057-1994, *"Standard for Digitizing Waveform Recorders"* (IEEE, Piscataway, New Jersey), 1994.

[4] Recommendation ITU-T G.810, 1996, *"Definitions and Terminology for Synchronization Networks"* (International Telecommunications Union, Geneva, Switzerland).

[5] V. S. Reinhardt, 2004, "*The Calculation of Frequency Source Requirements for Digital Communications Systems,"* in Proceedings of the 2004 IEEE International Frequency Control Symposium and Exposition, 23-27 Aug, 2004, Montreal, Canada (IEEE Publication), pp. 151-157.

[6] V. S. Reinhardt, 2005, *"A Review of Time Jitter and Digital Systems,"* in Proceedings of the 2005 Joint IEEE International Frequency Control Symposium and Precise Time and Time Interval (PTTI) Systems and Applications Meeting, 29-31 August 2005, Vancouver, Canada (IEEE Publication 05CH37664C), pp. 38-45.

[7] *"Mixed-signal and DSP design techniques,"* Sec. 2, Sampled Data Systems, in **Analog Devices**, *http://www.analog.com/Analog_Root/static/pdf/dataConverters/MixedSignal_Sect2.pdf*, pp. 35-36.

[8] V. P. Tuzlukov, 2002, **Signal Processing Noise** (CRC Press, Boca Raton, Fla.).

[9] E. W. Weisstein, 2007, "*Variance*," in **MathWorld**, *http://mathworld.wolfram.com/Variance.html*

[10] W. B. Davenport and W. L. Root, 1987, **An Introduction to the Theory of Random Signals and Noise** (IEEE, Piscataway, New Jersey).

[11] IEEE Std. 1521-2003, 2003, *"IEEE Trial-Use Standard for Measurement of Video Jitter and Wander"* (IEEE Broadcast Technology Society, Piscataway, New Jersey).

[12] Standard SMPTE RP 184-2003, 2003, *"Measurement of Jitter in Bit-Serial Digital Systems"* (Society of Motion Picture and Television Engineers).

[13] IEEE Std. 1139-1999, *"Standard Definitions of Physical Quantities for Fundamental Frequency and Time Metrology—Random Instabilities"* (IEEE, Piscataway, New Jersey), 1999.

[14] V. S. Reinhardt, 2007, "*A Physical Interpretation of Difference Variances,"* in Proceedings of TimeNav'07, the 21st European Frequency and Time Forum (EFTF) Joint with 2007 IEEE International Frequency Control Symposium (IEEE-FCS), 29 May-1 June 2007, Geneva, Switzerland (IEEE Publication CH37839).

[15] W. J. Riley, *"The Hadamard Variance,"* http://www.wriley.com/paper4ht.htm.

[16] R. A. Baugh, 1971, *"Frequency Modulation Analysis with the Hadamard Variance,"* in Proceedings of the 25th Annual Frequency Control Symposium, 26-28 April 1971, Atlantic City, New Jersey, USA (IEEE, Piscataway, New Jersey), pp. 222–225.

[17] G. Sauvage and J. Rutman, 1973, *"Analyse spectrale du bruit de frequency des oscillateurs par la variance de Hadamard,"* **Annales des Telecommunications, 28,** 304-314.

[18] J. Rutman, 1978, "*Characterization of Phase and Frequency Instabilities in Precision Frequency Sources: Fifteen Years of Progress,*" **Proceedings of the IEEE, 66,** 1048–1075.

[19] J. J. Gagnepain, 1998, "*La Variance de B. Picinbono,"* **Traitement du Signal, 15,** 477-482.

[20] C. A. Greenhall, 1991, *"Recipes for Degrees of Freedom of Frequency Stability Estimators,"* **IEEE Transactions on Instrumentation and Measurement, IM-40,** 994-999.

[21] C. A. Greenhall and W. J. Riley, 2004, *"Uncertainty of Stability Variances Based on Finite Differences,"* in Proceedings of the 35th Annual Precise Time and Time Interval (PTTI) Systems and Applications Meeting, 2-4 December 2003, San Diego, California, USA (U.S. Naval Observatory, Washington, D.C.), pp. 267-279.

[22] J. A. Barnes, 1966, "*Atomic Timekeeping and the Statistics of Precision Signal Generators,"* **Proceedings of the IEEE, 54,** 207-220.

[23] D. W. Allan, 1966, *"Statistics of Atomic Frequency Standards,"* **Proceedings of the IEEE, 54,** 221-230.

[24] J. A. Barnes, 1971, *"Characterization of Frequency Stability,"* **IEEE Transactions on Instrumentation and Measurement, IM-20,** 105-120.

[25] F. Vernotte, 2001, *"Application of the Moment Condition to Noise Simulation and to Stability Analysis,"* in Proceedings of the 2001 IEEE International Frequency Control Symposium & PDA Exhibition, 6-8 June 2001, Seattle, Washington, USA (IEEE Publication 01CH37218), pp. 133-137.

[26] D. A. Howe, R. L. Beard, C. A. Greenhall, F. Vernotte, W. J. Riley, and T. K. Peppler, 2005, *"Enhancements to GPS Operations and Clock Evaluations Using a 'Total' Hadamard Deviation,"* **IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, UFFC-52,** 1253-1261.

[27] V. S. Reinhardt, 2006, *"The Properties of Time and Phase Variances in the Presence of Power Law Noise for Various Systems,"* in Proceedings of the 2006 IEEE International Frequency Control Symposium and Exposition, 5-7 June 2006, Miami, Florida, USA (IEEE Publication 06CH37752C), pp. 745-749.

[28] D. A. Howe and T. N. Tasset, "*Clock Error Estimation based on PM Noise Measurements*," in Proceedings of the 2003 IEEE International Frequency Control Symposium Jointly with the 17th European Frequency and Time Forum (EFTF), 5-8 May 2003, Montreal, Canada (IEEE Publication 03CH37409), pp. 541-546.

[29] L. L. Scharf, B. Firedlander, and D. J. Thomson, 1998, *"Covavariant Estimators of Time-Frequency Descriptors for Nonstationary Random Processes,"* in Conference Record of the 32nd Asilomar Conference on Signals, Systems, and Computers, 1-4 November 1998, Pacific Grove, California, USA (IEEE), **Vol. 1,** pp. 808-811.

[30] L. Cohen, 1995, **Time-Frequency Analysis** (Prentice-Hall, Englewood Cliffs, N.J.).