# White Noise Rejection in a Deterministic Setting

Fernando Paganini
Electrical Engineering, M/S 116-81
California Institute of Technology
Pasadena, CA 91125
Email: fernando@hot.caltech.edu

## Abstract

This paper provides a framework for analyzing white noise disturbances in linear systems. Rather than the usual stochastic approach, noise signals are described as elements in sets and the disturbance rejection properties of the system are considered in a worst case setting. The description is based on constraints in signal space, directly verifiable on experimental data. These constraints can be given a representation compatible with standard robust control, allowing the formulation of white noise rejection problems in the presence of other sources of uncertainty. It is also shown how the framework can capture as a special case the usual stochastic approach, with equivalent results.

## 1   Introduction

Inaccuracies in mathematical models of physical systems are often characterized by the introduction of external disturbances, which account for phenomena which are too complex or unpredictable to be conveniently captured by the model. A model must then be accompanied by a description of the disturbance, and this implies a basic choice in the mathematical framework. The *deterministic* approach is to specify a *set* of allowable disturbances, and leads to *worst case* analysis over this set. Alternatively, the *stochastic* paradigm specifies a *measure* (probability distribution) in the disturbances, and leads naturally to analysis in the *average*.

1

The two relevant factors underlying this choice are mathematical convenience and the objective of obtaining a realistic and tight characterization of the empirical disturbance.

Probabilistic models have been the typical choice for *white noise* disturbances, which appear, for example, when considering the cumulative macroscopical effect of very high dimensional fluctuations at the microscopic level. Indeed, a stochastic process seems to provide a very good approximation to these phenomena, and in many cases this leads to tractable mathematics. An example of this in control theory is the classical $\mathcal{H}_2$ (LQG) *disturbance rejection* problem, namely the design of a feedback system which minimizes the (average) sensitivity of a linear system to stochastic noise.

The stochastic paradigm is less attractive, however, in practical control problems where disturbances coexist with other forms of uncertainty, such as unmodeled dynamics, which are described more naturally in a deterministic setting. In this *robust* control setting, mathematical convenience calls for deterministic disturbance rejection problems (e.g. $\mathcal{H}_\infty$, $\mathcal{L}_1$) where the disturbance is allowed to vary in a given set (e.g., the unit ball of $\mathcal{L}_2$, $\mathcal{L}_\infty$). Even though this leads to a successful robustness analysis methodology, some conservatism is involved in these deterministic classes, since the signals which give worst case performance (e.g. persisting sinusoids, in the $\mathcal{H}_\infty$ case) are often very unlikely disturbances in practical situations.

Attempts to combine deterministic uncertainty with stochastic white noise (the "Robust $\mathcal{H}_2$" problem, see for example [17]) face the difficulty of analyzing simultaneously the worst case effect of the uncertainty and the average effect of the disturbance, and have only resulted in upper bounds. Similar difficulties arise when attempting to establish closer connections

2

between classical system identification and robust control, since the former relies entirely in the stochastic paradigm for disturbances.

An important remark is that there is nothing inherently stochastic about white noise: it is known that deterministic chaotic systems can produce spectral effects indistinguishable from random noise.

This discussion leads us to the main objective of this paper, which is to obtain tight deterministic descriptions for white noise, suitable for robust control purposes. We now specify in more detail the objectives these set descriptions should meet:

1. They must allow a finite time horizon formulation; this is essential if these descriptions are to be used in practical problems involving data, such as system identification.

2. They must be rich enough to include "typical" instances of stochastic white noise signals.

3. They must be tight enough, so that *worst case* rejection properties of a system under disturbances of the set are essentially the same as *average* rejection properties under stochastic white noise.

4. They must allow for a mathematical formulation similar to other deterministic descriptions of uncertainty, to permit a simple formulation of robust performance problems.

The deterministic approach to statistical spectral analysis is not new and goes as far back as Wiener [19]; a modern reference is [10]. However, these treatments rely entirely on asymptotic properties of signals defined on infinite time intervals, and are not focused on the rejection problem and the related robustness issues.

In what follows we will present a very natural formulation compatible with the previous requirements; the starting point is the following question: how does one decide whether a signal can be accurately modeled as a stochastic white noise trajectory? Deciding this from experimental data leads to a statistical hypothesis test on a finite length signal. In other words, one will accept a signal as white noise if it belongs to a certain set. The main idea of our formulation is to take this set as the *definition* of white noise, and carry out the subsequent analysis in a deterministic setting.

The paper is organized as follows: Section 2 contains the notation and some preliminary facts; in Section 3, time domain descriptions are given, and they are analyzed from the point of view of the competing requirements (2) and (3). In Section 4, the same is done with descriptions in the frequency domain. Section 5 provides the multivariable extension of the previous framework. Section 6 summarizes the work and outlines a resulting research direction, which allows for these descriptions to be cast in the mathematical framework of robust control. Some technical proofs are covered in the Appendix.

## 2   Notation and Preliminaries

This paper deals with discrete time signals and linear time invariant systems. In sections 2, 3, and 4 we will consider scalar signals and single input/single output (SISO) systems. The multivariable version is given in Section 5. Notation and elementary properties of spectral analysis are presented in this section, in both finite and infinite time horizon cases.

## 2.1 Finite horizon properties

In the finite horizon case, we wish to characterize white signals among sequences of length $N$, and the *steady state* response of a linear system to such a disturbance. Therefore, to get sensible answers we must assume that $N$ is much larger than the system time constants. Under this assumption, the system gain will not be substantially affected if we consider the signals to be periodic, with period $N$, and we have available the information of one period. In fact, the system will not be sensitive to these "long range" correlations we have introduced in the input signals. As a counterpart, this gives tractable expressions.

Let $x(t)$ be a periodic, real valued signal, of period $N$, which will often be identified with the finite sequence $x(0) \cdots x(N-1)$. The discrete Fourier transform (DFT) $X(k), k = 0 \cdots N - 1$ of the sequence $x(t)$ is defined by the relations

$$X(k) = \sum_{t=0}^{N-1} x(t)e^{-j\frac{2\pi}{N}kt} \quad ; \quad x(t) = \frac{1}{N}\sum_{k=0}^{N-1} X(k)e^{j\frac{2\pi}{N}kt} \tag{1}$$

As a map between $\mathbb{C}^N$ and $\mathbb{C}^N$ the DFT as defined is a unitary operator scaled by $\sqrt{N}$.

The (circular) autocorrelation sequence of $x$ ("correlogram") is given by

$$r_x(\tau) = \sum_{t=0}^{N-1} x(t+\tau)x(t) \quad \tau = 0 \cdots N - 1 \tag{2}$$

and the sequence power spectrum ("periodogram") by $s_x(k) = |X(k)|^2, \; k = 0 \cdots N - 1$. It is easy to show that the sequences $r_x(\tau)$ and $s_x(k)$ form a DFT pair. For an $N$-periodic signal $x(t)$, we will use as norm the energy over the period, $\|x\|^2 = r_x(0) = \frac{1}{N}\sum_{k=0}^{N-1} s_x(k)$.

We consider a stable, LTI, discrete time SISO system $\mathcal{H}(\lambda) = \sum_{t=-\infty}^{\infty} h(t)\lambda^t$, where $\lambda$ is the shift operator. The frequency response (Fourier transform of $h(t)$) is denoted by $\mathcal{H}(e^{j\omega})$.

For a fixed $N$, denote $H(k) \triangleq \mathcal{H}(e^{j\frac{2\pi}{N}k})$ (if $\mathcal{H}$ is FIR of length less than $N$, this corresponds to the DFT of $h(t)$). The autocorrelations of the system $\mathcal{H}$ are defined by

$$r_h(\tau) = \sum_{t=-\infty}^{\infty} h(t+\tau)h(t) \tag{3}$$

assuming convergence for every $\tau$ (this is true if, for example, $\mathcal{H}$ is causal, exponentially stable; this will be assumed in the rest of the paper). The Fourier transform of $r_h(\tau)$ is the power spectrum $s_h(e^{j\omega}) = |\mathcal{H}(e^{j\omega})|^2$. Also, the 2-norm of the system is given by

$$\|\mathcal{H}\|_2^2 = r_h(0) = \sum_{t=-\infty}^{\infty} h(t)^2 = \frac{1}{2\pi} \int_0^{2\pi} s_h(e^{j\omega}) d\omega \tag{4}$$

As an immediate consequence of the previous definitions, the following relationships hold.

**Lemma 1** *Let $\mathcal{H}$ be a SISO discrete time stable system. Let $u(t)$ be an N-periodic input signal to system $\mathcal{H}$, $y(t)$ be the corresponding steady state (periodic) output. Then*

$$(i) \quad r_y(\tau) = \sum_{t=-\infty}^{\infty} r_h(t)r_u(t-\tau) \tag{5}$$

$$(ii) \quad s_y(k) = |H(k)|^2 s_u(k) \tag{6}$$

## 2.2 Infinite horizon properties

Autocorrelations and spectra can be defined in different classes of infinite horizon signals. We will mention here bounded power signals and bounded energy ($l_2$) signals. In the bounded power case, consider the class

$$\mathcal{BP} = \left\{ x(t) : r_x(\tau) = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{t=-N}^{N} x(t+\tau)x(t) \text{ exists for each } \tau \right\} \tag{7}$$

For $l_2$ (square-integrable) sequences, define $r_x(\tau) = \langle x, \lambda^\tau x \rangle$ as in (3). In both cases, it can be shown that there exists a *spectral density* $s_x(e^{j\omega})$ such that

$$r_x(\tau) = \frac{1}{2\pi} \int_0^{2\pi} s_x(e^{j\omega}) e^{j\omega\tau} d\omega \qquad (8)$$

For the bounded power case, this follows from Bochner's theorem (see [5]), and $s_x(e^{j\omega}) d\omega$ is in general a nonnegative measure. In the $l_2$ case, $s_x(e^{j\omega}) = |X(e^{j\omega})|^2$, where $X(e^{j\omega})$ is the Fourier transform of $x(t)$. $l_2$ is equipped with the standard norm $\|x\|_2 = r_x(0)^{\frac{1}{2}}$, $\mathcal{BP}$ with the seminorm $\|x\|_P = r_x(0)^{\frac{1}{2}}$.

Under mild assumptions on the system $\mathcal{H}$, (5) carries through to infinite horizon. Also, the corresponding extension of (6) is $s_y(e^{j\omega}) = |H(e^{j\omega})|^2 s_u(e^{j\omega})$.

# 3 Time domain descriptions

## 3.1 Finite horizon descriptions

The starting point for a deterministic white noise theory is to characterize white signals among all sequences of length N; when faced with the problem of deciding whether an empirical signal is a sample of white noise, a statistician will perform a hypothesis test in terms of some statistic. A common choice is the sample autocorrelation, which should approximate the expected correlation for white noise (a delta function). In other words a scalar signal is $x(t)$ categorized as white if $r_x(\tau)$ is small compared to $r_x(0)$ for nonzero $\tau$ in a certain range (e.g. for values of $\tau$ smaller than a horizon $T$). Pictorially, the autocorrelation plot, normalized to $r_x(0) = 1$, must fall inside a band around zero, of width $\gamma$, as in Figure 1.

From the classical statistical point of view, the choice of $\gamma$ is associated to a level of significance of the test, which in turn depends on some stochastic model. But regardless of
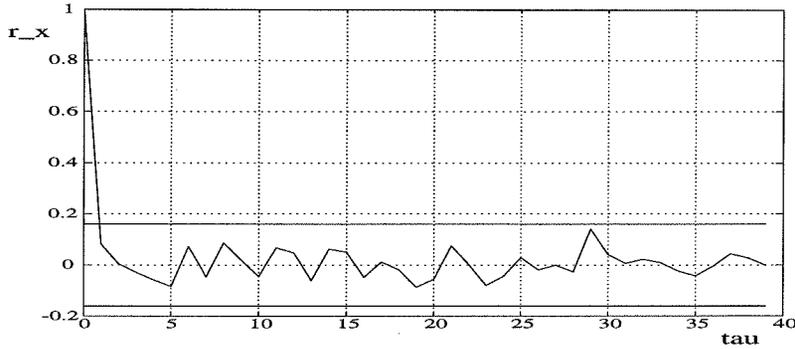
7

Figure 1: Autocorrelation plot of a pseudorandom sequence

the reasoning behind this choice, ultimately the "whiteness" of the signal is decided in terms

of a set, which is parameterized by $\gamma$ (and $T$). This motivates the following:

**Definition 1** *A signal $x(t)$, $t = 0 \cdots N - 1$ is said to be white with accuracy $\gamma$ up to time lag $T$ if it satisfies*

$$|r_x(\tau)| \leq \gamma r_x(0) \qquad \tau = 1 \cdots T \tag{9}$$

*The set of all such signals is denoted $W_{N,\gamma,T}$.*

It is natural to introduce a horizon $T$ in which the autocorrelations are required to be

small; this reduces the number of constraints, and if the response of a system is to be analyzed,

low correlation is only relevant in time scales where the system responds strongly. Both $\gamma$

and $T$ are, in fact, a parameterization of a rectangular weight function which specifies our

constraints on the autocorrelation. Other shapes of this weight function could be considered,

and the following results can be extended with minor modifications.

The response of an LTI system to signals in such sets will now be analyzed. The worst

8

case gain of the system under signals in $W_{N,\gamma,T}$ (a seminorm on systems) will be denoted

$$\|\mathcal{H}\|_{W_{N,\gamma,T}} = \sup\left\{\frac{\|y\|}{\|u\|},\; y = \mathcal{H}u,\; u \in W_{N,\gamma,T}, \|u\| \neq 0\right\} \tag{10}$$

**Theorem 1** *In the conditions of Lemma 1, if $\sum_{-\infty}^{\infty}|r_h(t)| < \infty$,*

*1. If $u \in W_{N,\gamma,T}$, then*

$$\left|\frac{r_y(\tau)}{r_u(0)} - r_h(\tau)\right| \leq \gamma \sum_{\substack{t=\tau-T \\ t\neq\tau}}^{\tau+T} |r_h(t)| + \sum_{|t-\tau|>T} |r_h(t)| \tag{11}$$

*2.*

$$\left|\|\mathcal{H}\|^2_{W_{N,\gamma,T}} - \|\mathcal{H}\|^2_2\right| \leq \gamma \sum_{\substack{t=-T \\ t\neq 0}}^{T} |r_h(t)| + \sum_{|t|>T} |r_h(t)| \tag{12}$$

*3. For the special case $\mathcal{H}(\lambda) = \sum_{t=0}^{T} h(t)\lambda^t$ ($\mathcal{H}$ is FIR),*

$$\|\mathcal{H}\|^2_2 \leq \|\mathcal{H}\|^2_{W_{N,\gamma,T}} \leq \|\mathcal{H}\|^2_2(1-\gamma) + \gamma \sum_{\tau=-T}^{T} |r_h(\tau)| \tag{13}$$

**Proof:** (11) follows immediately from Lemma 1, and the definition of $W_{N,\gamma,T}$. Specializing on $\tau = 0$ gives (12). The upper bound in (13) follows from (12), the lower bound from the fact that the delta function is always a signal in the set $W_{N,\gamma,T}$.

$\square$

**Remarks:**

1. From inequality (11) we conclude that the autocorrelations of the time series $y$ (up to a constant factor $\|u\|^2$) lie in a band centered in the autocorrelations of the filter. Therefore, such a band is a natural set description for *colored* noise, the output of a linear filter under white noise.

9

2. It can be shown (see [14]) that if $\gamma < \frac{1}{T}$, then for large enough $N$ the upper bound in (13) is achieved. This is no longer true for large values of $\gamma$; for example, if $\gamma = 1$, there are no restrictions on the input signal, and the induced norm can be bounded by the $\mathcal{H}_\infty$ norm of the system which in the FIR case is equal to

$$\sup_\omega \left( r_h(0) + 2 \sum_{\tau=1}^{T} r_h(\tau) \cos \omega \tau \right)^{\frac{1}{2}}$$

and is in general strictly less than the bound (13). The role of $\gamma$ in this deterministic approach is therefore to constrain the freedom of the "adversary" (the disturbance), and results in a worst-case gain which varies from the $\mathcal{H}_\infty$ norm for $\gamma = 1$ to the $\mathcal{H}_2$ norm for $\gamma = 0$. Ideally, one could use $\gamma = 0$ but on a finite horizon setting, this would constitute an unrealistically small class of signals. This raises the issue of trading off between the objectives 2 and 3 stated in Section 1: obtaining low worst case gains on sets which accommodate a reasonable class of signals.

To analyze this we turn to asymptotic results, as the length $N$ of the data record goes to infinity and find rates of $\gamma$, $T$ that achieve this compromise. First, we give conditions under which the asymptotic norm is the $\mathcal{H}_2$ norm, which follow obviously from Theorem 1:

**Corollary 1** *1. If $T$ is fixed, $\mathcal{H}$ FIR (T), and $\gamma(N) \stackrel{N\to\infty}{\longrightarrow} 0$, then $\|\mathcal{H}\|_{W_{N,\gamma,T}} \stackrel{N\to\infty}{\longrightarrow} \|\mathcal{H}\|_2$.*

*2. If $T(N) \stackrel{N\to\infty}{\longrightarrow} \infty$, and $\gamma(N) \stackrel{N\to\infty}{\longrightarrow} 0$, $\mathcal{H}$ IIR, then $\|\mathcal{H}\|_{W_{N,\gamma,T}} \stackrel{N\to\infty}{\longrightarrow} \|\mathcal{H}\|_2$ .*

Secondly, a natural requirement for a set description as in definition (1) to be rich enough is that the set have "large" probability when the signal effectively comes from a stochastic white process: this is how these sets are chosen in the standard statistical approach. A reasonably general answer is the following ($\mathcal{P}$ denotes probability):

10

**Theorem 2** *Let $x(0), \cdots, x(N-1), \cdots$ be independent, identically distributed random variables, with 0 mean and finite variance. For each $N$, denote $x_N = (x(0), \cdots, x(N-1))$.*

1. *If $T$ is fixed, and $\gamma(N)\sqrt{N} \overset{N\to\infty}{\longrightarrow} \infty$, then $\mathcal{P}\left(x_N \in W_{N,\gamma,T}\right) \overset{N\to\infty}{\longrightarrow} 1$.*

2. *If the $x(t)$ are bounded, and $\gamma(N)\sqrt{\frac{N}{log(N)}} \overset{N\to\infty}{\longrightarrow} \infty$, then $\mathcal{P}\left(x_N \in W_{N,\gamma,N-1}\right) \overset{N\to\infty}{\longrightarrow} 1$.*

3. *If the $x(t)$ are Gaussian, and $\gamma(N)\frac{\sqrt{N}}{log(N)^{\frac{3}{2}}} \overset{N\to\infty}{\longrightarrow} \infty$, then $\mathcal{P}\left(x_N \in W_{N,\gamma,N-1}\right) \overset{N\to\infty}{\longrightarrow} 1$.*

**Proof:** See the Appendix.

For parts 2, 3, $T(N) = N - 1$ was chosen (*all* the entries of $r_x(\tau)$ are constrained; in fact, it is equivalent to constrain up to $T(N) = \lfloor \frac{N}{2} \rfloor$). This type of result is different from those which have typically been considered in the statistical literature, where a small number of $\tau$ values is employed, and substantial "averaging" is available. As will be discussed later, there are also reasons in this setting to reduce the number of constraints. It seems interesting, however, that the autocorrelation plot captures "whiteness" uniformly across all values of $\tau$; this is a strong argument in favor of the use of this statistic. This result also supports the notion of distance employed in the correlogram, in terms of the vector $\infty$ norm. In the next section we will see that the "periodogram" is not as well behaved.

In any event, by showing that the asymptotic probability of $W_{N,\gamma,N-1}$ is 1, it immediately follows that the same holds for $W_{N,\gamma,T}$, with a smaller growth rate of $T$.

A simple way to summarize the preceding results in relation to stochastic white noise, is to say that the *expected* gain of an LTI system to white noise (the $\mathcal{H}_2$ norm), is essentially the same as the *worst case* gain of the system in a set of signals which is "typical" from the point

of view of the probability, when the mechanism which generates the disturbances is assumed to be stochastic.

As remarked before, this assumption cannot be directly verified, and there is evidence that non-stochastic systems (e.g. deterministic chaos, see [1] and references therein) can produce similar spectral properties.

Another situation where disturbances are considered is as "residuals" of some system identification technique, i.e. an error variable needed to explain the experimental data. Though the system identification theory assumes a stochastic model for this disturbance, in practice it always includes other deterministic (e.g. nonlinear) effects.

The previous results show that in terms of rejection, what matters is the *statistical* information (which may be directly tied to experiments), not the generating mechanism. Auto-correlation constraints which characterize a disturbance (and may or may not be consistent with the levels for stochastic noise) can be incorporated into a worst-case rejection measure.

## 3.2 Infinite horizon descriptions

We conclude the section by considering the infinite horizon counterpart of definition 1. For brevity, we will treat $l_2$, $\mathcal{BP}$ signals simultaneously.

**Definition 2** *A signal in $l_2$ (or in $\mathcal{BP}$) is said to be white, with accuracy $\gamma$ up to time lag $T$ if it satisfies*

$$|r_x(\tau)| \leq \gamma r_x(0) \qquad \tau = 1 \cdots T \tag{14}$$

*The set of all such signals is denoted $W_{\gamma,T}$.*

For an exponentially stable $\mathcal{H}$, defining the norm $\|\mathcal{H}\|_{W_{\gamma,T}}$ analogously to (10) it can be concluded in a similar way as before that

$$\|\mathcal{H}\|_2^2 \le \|\mathcal{H}\|_{W_{\gamma,T}}^2 \le \|\mathcal{H}\|_2^2 + \gamma \sum_{\tau=-T}^{T} |r_h(\tau)| + \sum_{|t|>T} |r_h(t)| \qquad (15)$$

It is tempting to consider the set $W_{0,\infty} = \{x(t) \in \mathcal{BP}, r_x(\tau) = 0 \ \forall \tau \ne 0\}$. In this class, (which is the one used in other deterministic spectral analysis treatments, such as [10]) the induced norm is exactly $\|\mathcal{H}\|_{W_{0,\infty}} = \|\mathcal{H}\|_2$, and moreover, for the bounded power case the class contains trajectories of stochastic white noise:

**Theorem 3** *Let $x(0), \cdots, x(t), \cdots$ be independent, identically distributed random variables, with 0 mean and finite variance. Then $\mathcal{P}(x \in W_{0,\infty}) = 1 \ (W_{0,\infty} \subset \mathcal{BP})$.*

**Proof:** For a fixed $\tau \ne 0$, referring to [5] (proposition 6.31), we find that the random process $z(t) = x(t)x(t+\tau)$ is ergodic, so with probability 1,

$$\lim_{N \to \infty} \frac{1}{2N+1} \sum_{t=-N}^{N} x(t+\tau)x(t) = E[x(t+\tau)x(t)] = 0 \qquad (16)$$

Therefore $W_{0,\infty}$ has probability 1 (countable intersection of probability 1 sets).

$\square$

These results on $W_{0,\infty}$ are not, however, particularly useful for the following reasons. In the first place, the constraints on a bounded power signal depend exclusively on its asymptotic behavior: *any* sequence in $\mathbb{R}^N$ is a valid truncation of a white power signal. From a practical perspective, it is impossible to know whether a disturbance is in $W_{0,\infty}(\mathcal{BP})$, just as it is to verify that a signal is generated by a stochastic white process. In this respect, it seems that the $l_2$ version is a better behaved infinite horizon abstraction; if a signal in $l_2$ is truncated with

13

large enough $N$ so as to capture most of its energy, then the autocorrelations are essentially determined by the truncation.

Also, when dealing with uncertain systems, constraints on the signals must be enforced *explicitly* in any analysis or synthesis procedure, as will be discussed below. The definition of $W_{0,\infty}$ requires an infinite number of constraints, which cannot be handled. In the case of a finite number, such as in $W_{\gamma,T}$ ($W_{0,T}$ in particular), these constraints can be naturally introduced into a robustness analysis or synthesis problem, as is shown in [15].

## 4  Frequency domain descriptions

### 4.1  Finite horizon descriptions

In the frequency domain, the natural object of study is the power spectrum; as the name implies, a "white" signal is characterized by a flat power spectrum. Referring to finite length signals, presumably a set characterization of whiteness can be obtained by specifying the periodogram $s_x(k)$ to be close to a constant across $k$; it is difficult, however, to find a notion of distance in which this holds for typical white noise signals. Figure 2 shows a typical periodogram of a signal obtained from a pseudorandom number generator. As we can see, the periodogram is very erratic, and is not close to its average value in a pointwise sense. Various authors (see [6] and references therein) have studied the stochastic properties of the periodogram. In the case of Gaussian noise, for example, the following can be shown:

**Proposition 1** *Let $x(t), t = 0 \cdots N - 1$, be independent, complex Gaussian random variables, of mean zero and variance 1. Then $\frac{s_x(k)}{N}, k = 0 \cdots N - 1$ are independent random variables of*
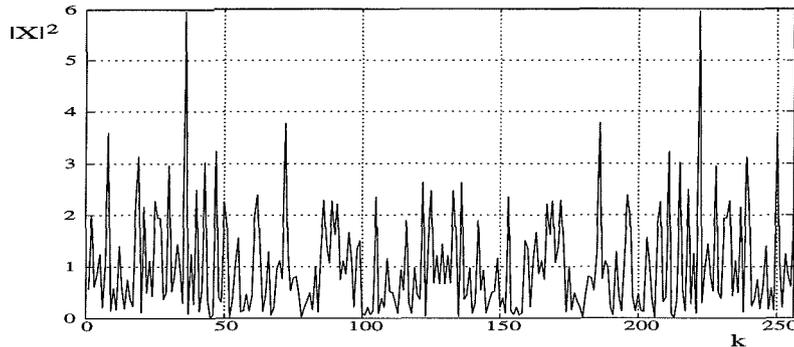
Figure 2: Periodogram of a pseudorandom sequence

*exponential distribution, with expected value 1, and the expected value of $max_k \frac{s_x(k)}{N}$ grows as $Log(N)$ when $N \to \infty$.*

This means that if we attempt to characterize white noise by a band of power spectra around a constant value, this band would have to grow arbitrarily to be able to capture stochastic noise. It is not hard to show that the worst case gain under signals in such a band would approach the $\mathcal{H}_\infty$ norm of the system; therefore, this description would not be tight enough for our purposes. Sets of spectra defined in terms of other simple vector norms can also be shown to be not satisfactory for similar reasons.

The fact that a "raw" periodogram of a noise signal is not a very well behaved statistic has long been recognized in the statistical spectral analysis community (see [3]). Peaks in the periodogram do not necessarily correspond to underlying periodicities in the time-series, and from this point of view the autocorrelation plot is more significant. Another way to say this is that the frequency domain is not a natural set of "coordinates" to uncover trends in noisy data. Rotating the data back to the time domain (autocorrelation plot) gives a tight

15

description for "whiteness": small distance to the delta function in the vector $\infty$ norm.

There is, however, a way around this difficulty that has been used extensively in statistical spectral analysis, in terms of *smoothing* of the periodogram (see [18, 6]): adequate local averaging in the periodogram reveals the process spectral information. In this section, we will pursue the same smoothing approach to provide set descriptions of white noise. Instead of smoothing by convolution (as in [14]) in this paper we will adopt an approach of averaging in a set of bands: given a uniform filter bank $V_m(k), m = 0 \cdots M - 1$, a white signal will be characterized by having approximately equal energy on these bands. Various designs for the filter bank could be considered; in this paper we will assume for simplicity that the length of the signal is a integer multiple of the number of bands $N = MK$, and that the filter bank is made of ideal bandpass filters of the form

$$V_m(k) = \begin{cases} \frac{1}{K} & mK \leq k < (m+1)K \\ 0 & otherwise \end{cases} \tag{17}$$

**Definition 3** *Let $N = MK$. Define the filter bank $V = \{V_m(k), m = 0 \cdots M - 1\}$ as in (17)*

*A signal $x(t)$ of length $N$ is said to be white with accuracy $\alpha$, with respect to the filter bank $V$, if its periodogram $s_x(k)$ verifies*

$$\max_m \frac{\langle s_x, V_m \rangle}{\|x\|^2} \leq 1 + \alpha \tag{18}$$

*We denote the set of all such signals as $\hat{W}_{N,\alpha,V}$.*

In the definition, the inner product $\langle s_x, V_m \rangle$ averages the periodogram in the band. The requirement is that the (normalized) band averages be close to the global average of 1 (since these are nonnegative quantities, constraining from above suffices). With these definitions,

16

results which parallel those in the time domain can be written. The worst-case induced norm of a system $\mathcal{H}$ under signals in the set $\hat{W}_{N,\alpha,V}$ will be denoted $\|\mathcal{H}\|_{\hat{W}_{N,\alpha,V}}$.

**Theorem 4** *Consider a stable LTI system $\mathcal{H}$, such that $B \triangleq \max_\omega |\frac{d}{d\omega}|\mathcal{H}(e^{j\omega})|^2|$ is finite. Then*

$$\|\mathcal{H}\|_2^2 - \frac{B\pi}{M} \leq \|\mathcal{H}\|_{\hat{W}_{N,\alpha,V}}^2 \leq \|\mathcal{H}\|_2^2(1+\alpha) + \frac{B\pi}{M} \tag{19}$$

**Proof:** Denote $h_m = \frac{M}{2\pi}\int_{\frac{2\pi m}{M}}^{\frac{2\pi(m+1)}{M}}|\mathcal{H}(e^{j\omega})|^2 d\omega$. Note that $\frac{1}{M}\sum_{m=0}^{M-1} h_m = \|\mathcal{H}\|_2^2$.

If $mK \leq k < (m+1)K$, then

$$\left| h_m - |H(k)|^2 \right| \leq \frac{M}{2\pi}\int_{\frac{2\pi m}{M}}^{\frac{2\pi(m+1)}{M}}\left| |\mathcal{H}(e^{j\omega})|^2 - |\mathcal{H}(e^{j\frac{2\pi k}{N}})|^2 \right| d\omega \leq \frac{BM}{2\pi}\int_{\frac{2\pi m}{M}}^{\frac{2\pi(m+1)}{M}}\left|\omega - \frac{2\pi k}{N}\right| d\omega \leq \frac{B\pi}{M} \tag{20}$$

Let $u(t)$ be a signal in $\hat{W}_{N,\alpha,V}$, $\|u\|^2 = \frac{1}{N}\sum_{k=0}^{N-1} s_u(k) = 1$, $y$ the corresponding output.

$$\begin{aligned}
\|y\|^2 &= \frac{1}{N}\sum_{k=0}^{N-1} s_u(k)|H(k)|^2 = \frac{1}{M}\sum_{m=0}^{M-1}\sum_{k=0}^{N-1} s_u(k)|H(k)|^2 V_m(k) = \\
&= \frac{1}{M}\sum_{m=0}^{M-1} h_m \sum_{k=0}^{N-1} s_u(k) V_m(k) + \frac{1}{M}\sum_{m=0}^{M-1}\sum_{k=0}^{N-1} s_u(k)(|H(k)|^2 - h_m) V_m(k) \quad (21)
\end{aligned}$$

For the second term in (21), the bound (20) gives

$$\left| \frac{1}{M}\sum_{m=0}^{M-1}\sum_{k=0}^{N-1} s_u(k)(|H(k)|^2 - h_m) V_m(k) \right| \leq \frac{B\pi}{M}\left( \frac{1}{M}\sum_{m=0}^{M-1}\sum_{k=0}^{N-1} s_u(k) V_m(k) \right) = \frac{B\pi}{M} \tag{22}$$

Since $u \in \hat{W}_{N,\alpha,V}$,

$$\langle s_u, V_m \rangle = \sum_{k=0}^{N-1} s_u(k) V_m(k) \leq 1 + \alpha \tag{23}$$

Also, for signals with $s_u(k) \equiv 1$ the left hand side of (23) achieves the value 1. Therefore the supremum of the first term in (21) is bounded between $\|\mathcal{H}\|_2^2$ and $(1+\alpha)\|\mathcal{H}\|_2^2$. Incorporating (22), the supremum of $\|y\|^2$ is between the bounds in (19).

17

□

**Corollary 2** *In the conditions of Theorem 4, if* $\alpha(N) \overset{N \to \infty}{\longrightarrow} 0$, *and* $M(N) \overset{N \to \infty}{\longrightarrow} \infty$, *then*

$$\|\mathcal{H}\|_{\hat{W}_{N,\alpha,V}} \overset{N \to \infty}{\longrightarrow} \|\mathcal{H}\|_2 \tag{24}$$

**Theorem 5** *Let* $x(0), \cdots, x(N-1), \cdots$ *be independent, Gaussian, zero mean random variables. For each* $N$, *denote* $x_N = (x(0), \cdots, x(N-1))$. *Assume* $\alpha(N) \overset{N \to \infty}{\longrightarrow} 0$, $M(N) \overset{N \to \infty}{\longrightarrow} \infty$, *and* $\frac{N\alpha^2}{MLog(M)} \overset{N \to \infty}{\longrightarrow} \infty$. *Then*

$$\mathcal{P}\left(x_N \in \hat{W}_{N,\alpha,V}\right) \overset{N \to \infty}{\longrightarrow} 1 \tag{25}$$

**Proof:** See the Appendix.

The previous statements provide the frequency domain counterpart of the time domain asymptotic results. Provided $\alpha(N) \overset{N \to \infty}{\longrightarrow} 0$, $M(N) \overset{N \to \infty}{\longrightarrow} \infty$ with appropriate rates (e.g. $M(N) = \sqrt{N}$, $\alpha(N) = \frac{Log(N)}{N^{1/4}}$), the worst case disturbance rejection measure approaches the $\mathcal{H}_2$-norm of the system, while the class of signals contains asymptotically all typical instances of stochastic white noise, at least in the Gaussian case.

We have introduced essentially dual descriptions in time and frequency domains, of parameterized sets which describe white noise. With the chosen definitions, however, these sets are different: the time and frequency domain constraints do not correspond to each other through DFT. Such a correspondence can be achieved if the *Euclidean* norm is used to measure distance in both domains (the DFT is unitary), and smoothing in the frequency domain

18

is performed by convolution. The resulting signal constraints are more complicated, however, since they involve fourth powers of the signals.

As mentioned in Section 1, an important objective in the choice for these descriptions is their simplicity and compatibility with representations for uncertainty. From this point of view, constraints which are *quadratic* in the signals involved are especially adequate. This fact has been recognized in the work of Yakubovich [20] and Megretski [13], where uncertainty is described in terms of integral quadratic constraints (IQCs). Alternatively, in [15] it shown how these set characterizations for white noise can be easily fit into the standard linear fractional transformation (LFT) framework for uncertainty, expressing these sets as *kernels* of LFTs on uncertain operators. This is the underlying motivation for our particular choices for the sets.

In any event, for large $N$ and these choices of parameters, both time and frequency definitions capture sets of signals which are "likely" if we adhere to a probabilistic model, and which produce approximately the same worst-case disturbance rejection measure; from the point of view of the objectives of this paper, they are equivalently adequate descriptions for white noise.

## 4.2   Infinite horizon descriptions

An infinite horizon (in $l_2$, or $\mathcal{BP}$) counterpart of the set $\hat{W}_{N,\alpha,V}$ is now given. A filter bank $V = \{V_m(e^{j\omega}), m = 0 \cdots M - 1\}$ is used, where the integral across frequency of each $V_m$ is 1.

**Definition 4** *Given a filter bank $V = \{V_m(e^{j\omega}), m = 0 \cdots M - 1\}$, an $l_2$ (or $\mathcal{BP}$) signal $x(t)$ is said to be white with accuracy $\alpha$, with respect to the filter bank $V$, if its spectral density*

$s_x(e^{j\omega})$ *verifies*

$$\max_m \frac{1}{2\pi} \int_0^{2\pi} s_x(e^{j\omega}) V_m(e^{j\omega}) d\omega \leq (1+\alpha) \frac{1}{2\pi} \int_0^{2\pi} s_x(e^{j\omega}) d\omega \tag{26}$$

*We denote the set of all such signals as* $\hat{W}_{\alpha,V}$.

For the case of an filter bank with ideal bandpass filters,

$$V_m(e^{j\omega}) = \begin{cases} M & \frac{2\pi m}{M} \leq \omega \leq \frac{2\pi(m+1)}{M} \\ 0 & otherwise \end{cases} \tag{27}$$

we arrive in a similar manner to the bound

$$\|\mathcal{H}\|_2^2 \leq \|\mathcal{H}\|_{\hat{W}_{\alpha,V}}^2 \leq \|\mathcal{H}\|_2^2 (1+\alpha) + \frac{\pi B}{M} \tag{28}$$

# 5   The multivariable case

This section outlines how the previous methodology can be extended to deal with vector valued white noise signals. For reasons of brevity and simplicity we will only consider the case of infinite horizon $l_2$ signals, which demonstrates all the necessary extensions.

For vector-valued signals $x(t) \in l_2(\mathbb{R}^n)$, the matrix autocorrelation (prime denotes transpose, * denotes conjugate transpose) is given by

$$R_x(\tau) = \sum_{t=-\infty}^{\infty} x(t+\tau) x'(t) \tag{29}$$

and the matrix spectrum by $S_x(e^{j\omega}) = X(e^{j\omega}) X^*(e^{j\omega})$, where $X(e^{j\omega})$ is the Fourier transform of $x(t)$. They are related as in (8). The 2-norm of the signal verifies

$$\|x\|_2^2 = trace(R_x(0)) = \frac{1}{2\pi} \int_0^{2\pi} trace(S_x(e^{j\omega})) d\omega \tag{30}$$

Consider a stable, discrete time linear time invariant system with in general $n$ inputs and $p$ outputs, $\mathcal{H}(\lambda) = \sum_{t=-\infty}^{\infty} H(t) \lambda^t$, with frequency response $\mathcal{H}(e^{j\omega})$. We define $R_H(\tau) =$

$\sum_{t=-\infty}^{\infty} H(t+\tau)H'(t)$ and $S_H(e^{j\omega}) = \mathcal{H}(e^{j\omega})\mathcal{H}^*(e^{j\omega})$. The 2 norm of $\mathcal{H}$ satisfies the relation (30). If $u(t) \in l_2(\mathbb{R}^n)$, $y(t) \in l_2(\mathbb{R}^p)$ are respectively, the input and output to $\mathcal{H}$, then the following relations hold:

$$R_y(\tau) = \sum_{t=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} H(s)R_u(\tau+t-s)H'(t) \tag{31}$$

$$S_y(e^{j\omega}) = \mathcal{H}(e^{j\omega})S_x(e^{j\omega})\mathcal{H}(e^{j\omega})^* \tag{32}$$

Now we give set descriptions of vector valued white noise. In the time domain, $R_x(\tau)$ should be small for $\tau \neq 0$, and that $R_x(0)$ must be approximately a constant times the identity matrix. This implies that in addition to the components of $x(t)$ being scalar white noise signals, they must be "spatially" uncorrelated.

Some matrix norms will be used in the following:

$$\|A\|_\infty = \max_{i,j} |a_{ij}| \qquad \|A\|_1 = \sum_{i,j} |a_{ij}|; \qquad \|A\|_F = (trace(A'A))^{\frac{1}{2}} = \left(\sum_{i,j} |a_{ij}|^2\right)^{\frac{1}{2}}$$

Note that

$$|trace(AB)| \leq \|A\|_\infty \|B\|_1 \qquad |trace(AB)| \leq \|A\|_F \|B\|_F \tag{33}$$

**Definition 5** *A signal $x(t) \in l_2(\mathbb{R}^n)$ is said to be white with accuracy $\gamma$ up to time lag $T$ if it satisfies*

$$\left\| \frac{1}{\|x\|^2} R_x(\tau) - \delta(\tau)\frac{1}{n}I \right\|_\nu \leq \gamma \tag{34}$$

*We will denote the set of all such signals by $W_{\gamma,T}^n$.*

In the definition, $\delta(\tau)$ is the usual delta function, and the norm $\|\cdot\|_\nu$ referred to in the definition can be in principle any matrix norm. The norm $\|\cdot\|_\infty$ has the advantage of giving quadratic constraints on the signal. Defining $\|\mathcal{H}\|_{W_{\gamma,T}^n}$ as in (10), we have:

**Theorem 6** *Defining $W_{\gamma,T}^n$ in terms of the matrix norm $\|\cdot\|_\infty$*

$$\left| \|\mathcal{H}\|_{W_{\gamma,T}^n}^2 - \frac{1}{n}\|\mathcal{H}\|_2^2 \right| \leq \gamma \sum_{\tau=-T}^{T} \|R_{H'}(\tau)\|_1 + \sum_{|t|>T} \|R_{H'}(\tau)\|_F \tag{35}$$

**Proof:** Let $\|u\| = 1, u \in W_{\gamma,T}^n$. Starting from (31),

$$trace(R_y(0)) = \sum_{t=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} trace(R_u(t-s)H'(t)H(s)) = \sum_{\tau=-\infty}^{\infty} trace(R_u(\tau)R_{H'}(\tau))$$

$$trace(R_y(0)) - \frac{1}{n}\|\mathcal{H}\|_2^2 = \sum_{\tau=-\infty}^{\infty} trace\left( \left(R_u(\tau) - \frac{1}{n}\delta(\tau)I\right)R_{H'}(\tau)\right)$$

Using (33), and the fact that $\|R_u(\tau)\|_F \leq \|u\|^2$ for $\tau \neq 0$ we have

$$\left| \|\mathcal{H}\|_{W_{N,\gamma,T}}^2 - \frac{1}{n}\|\mathcal{H}\|_2^2 \right| \leq \sum_{\tau=-T}^{T} \left\| R_u(\tau) - \frac{1}{n}\delta(\tau)I\right\|_\infty \|R_{H'}(\tau)\|_1 + \sum_{|\tau|>T} \|R_u(\tau)\|_F \|R_{H'}(\tau)\|_F \leq$$

$$\leq \gamma \sum_{\tau=-T}^{T} \|R_{H'}(\tau)\|_1 + \sum_{|t|>T} \|R_{H'}(\tau)\|_F$$

$\square$

Under similar growth conditions as in Corollary (1), the asymptotic norm is $\frac{1}{\sqrt{n}}\|\mathcal{H}\|_2$.

For a frequency domain characterization, consider

**Definition 6** *Given a filter bank $V = \{V_m(e^{j\omega}), m = 0\cdots M-1\}$, an $l_2(\mathbb{R}^n)$ signal $x(t)$ is said to be white with accuracy $\alpha$, with respect to the filter bank $V$, if its spectral density $S_x(e^{j\omega})$ verifies*

$$\max_m \left\| \frac{1}{2\pi \|x\|_2^2} \int_0^{2\pi} S_x(e^{j\omega})V_m(e^{j\omega})d\omega - \frac{1}{n}I \right\|_\nu \leq \alpha \tag{36}$$

*The set of all such signals is denoted $\hat{W}_{\alpha,V}^n$.*

Will not pursue the subsequent analysis, which follows the same lines as before.

# 6 Conclusion

In this paper, set characterizations of white noise in terms of constraints in signal space were presented. It was shown how these sets can be "tailored" to adequately capture stochastic noise, retaining its properties in terms of system gain, now understood in a worst case setting. The parameterization allows, however, a greater flexibility in signal characterization, and the finite horizon version allows these descriptions to be tied directly to experimental data.

The bounds obtained for worst-case gain on these sets of signals are useful in showing that this procedure is sound and consistent with the alternative stochastic approach, but they are not exact and too complicated to provide a basis for *robust* performance analysis when the system $\mathcal{H}$ is subject to uncertainty. The major argument given in Section 1 in favor of adopting these deterministic descriptions was, after all, to unify white noise rejection with robustness analysis. Fortunately there is an elegant framework (developed fully in [15]) which encompasses our deterministic descriptions of white noise with other forms of uncertainty in the system. This framework relates to the IQC [20, 13] formulation, and to recent developments in uncertain behavioral systems [7]. The resulting methods for "Robust $\mathcal{H}_2$" analysis and synthesis have been pursued in [15, 8].

This framework can be extended, to some degree, to continuous time systems. While "pure" white noise is a difficult object to define, it is clear that useful approximations can be obtained, for example, from sets of signals defined in terms of frequency domain characterizations, in a similar manner to the discrete time case.

The finite horizon version of this framework can also be applied to the area of worst-case

system identification. Recent work [9, 16] has shown that if noise disturbances are allowed to be arbitrary norm bounded signals, the identification problem has high computational complexity. It is to expect that constraining the disturbance in the style of this paper (constraining the freedom of the "adversary" in the identification problem) will bring some reduction in complexity.

# Appendix

The stochastic results will be proved here. In the sequel, $x(0) \cdots x(N-1) \cdots$ are independent, identically distributed random variables, of 0 mean and finite variance. Since the sets $W_{N,\gamma,T}$ and $\hat{W}_{N,\alpha,V}$ are closed under scalar multiplication, the variance can be normalized to 1.

**Proof of Theorem 2:**

**Part 1:** For the case of a fixed time lag $\tau$, the distribution of the autocorrelation $r_x(\tau)$ has been extensively studied in the statistical literature [3, 2, 12]; exact expressions for the distribution of $r_x(\tau)/r_x(0)$ when $x(t)$ is Gaussian are obtained in [2], and asymptotic normality holds. We outline a proof for completeness.

A central limit theorem on the $\tau$-dependent (see [4]) random variables $z(n) = x(n)x(n+\tau)$ shows that $\frac{1}{\sqrt{N-\tau}} \sum_{n=0}^{N-\tau-1} x(n)x(n+\tau)$ is asymptotically normal $\mathcal{N}(0,1)$. Since $\tau$ is fixed, the "circular" terms are vanishingly small, and the same holds for $\frac{1}{\sqrt{N}} r_x(\tau)$. Also $\frac{1}{N} r_x(0)$ converges almost surely to 1, so $\sqrt{N} \frac{r_x(\tau)}{r_x(0)}$ is asymptotically normal. Since $\gamma\sqrt{N} \to \infty$, and $T$ is fixed,

$$\mathcal{P}\left(x \notin W_{N,\gamma,T}\right) \leq \sum_{\tau=1}^{T} \mathcal{P}\left(\left|\sqrt{N}\frac{r_x(\tau)}{r_x(0)}\right| > \gamma\sqrt{N}\right) \xrightarrow{N \to \infty} 0 \tag{37}$$

□

In parts 2, 3 of the theorem, the number of correlation constraints grows with the sample size, and the argument with the normal approximation cannot be used: even though each $r_x(\tau)$ for fixed $\tau$ is asymptotically normal, the joint distribution of $(r_x(1), \cdots r_x(N-1))$ is defined on a space of increasing dimension, where no global "averaging" occurs. Our proof relies on a Hoeffding inequality for sums of bounded random variables, [11]:

**Theorem 7 (Hoeffding)** *Let $z_0, \cdots z_{N-1}$ be independent random variables, of mean $\mu$ and bounded ($a \leq z_n \leq b$), $\bar{z} = \frac{1}{N} \sum_{n=0}^{N-1} z_n$. Then for $\epsilon > 0$,*

$$\mathcal{P}\left(\bar{z} - \mu > \epsilon\right) \leq e^{\frac{-2N\epsilon^2}{(b-a)^2}} \tag{38}$$

To apply this inequality to the sum $r_x(\tau) = \sum_{n=0}^{N-1} z(t)$, with $z(t) = x(t)x((t+\tau)\mathrm{mod}N)$, the following lemma takes care of of the slight dependence between the terms, by dividing the sum in three sums, ensuring $z(t)$, $z((t+\tau)\mathrm{mod}N)$, $z((t-\tau)\mathrm{mod}N)$ fall in different groups for each $t$. We omit the elementary proof, which requires a discussion on the values of $N$ and $\tau$.

**Lemma 2** *Let $N \geq 3$, and $x(0), x(1) \cdots x(N-1)$ be independent identically distributed random variables. For fixed $1 \leq \tau \leq \frac{N}{2}$, there exists a partition (depending only on $N, \tau$) of the terms in the sum $r_x(\tau)$ into three groups, giving $r_x(\tau) = S_0 + S_1 + S_2$, where each $S_i$ is the sum of $n_i$ independent random variables, and $n_i > \frac{N}{6}$.*

**Part 2:** Assume $x(0) \cdots x(N-1)$ are *bounded* random variables, $|x(t)| \leq K$. Pick $1 \leq \tau \leq \frac{N}{2}$. From the lemma $r_x(\tau) = S_0 + S_1 + S_2$, where each $S_i$ is the sum of $n_i$ independent,

25

identically distributed random variables, with zero mean and bounded in $[-K^2, K^2]$. Invoking Hoeffding's inequality and $n_i \geq \frac{N}{6}$, we have

$$\mathcal{P}\left(\frac{r_x(\tau)}{N} > \epsilon\right) \leq \sum_{i=0}^{2} \mathcal{P}\left(\frac{S_i}{n_i} > \epsilon\right) \leq \sum_{i=0}^{2} e^{\frac{-n_i \epsilon^2}{2K^4}} \leq 3e^{\frac{-N\epsilon^2}{12K^4}} \tag{39}$$

The same argument can be employed to bound $\mathcal{P}\left(-\frac{r_x(\tau)}{N} > \epsilon\right)$, for each value of $\tau$. This implies

$$\mathcal{P}\left(\max_{1 \leq \tau \leq \frac{N}{2}} \frac{|r_x(\tau)|}{N} > \epsilon\right) \leq 3N e^{\frac{-N\epsilon^2}{12K^4}} = 3e^{Log(N)\left(1 - \frac{N\epsilon^2}{12K^4 Log(N)}\right)} \tag{40}$$

By symmetry of $r_x(\tau)$, $W_{N,\gamma,N-1} = W_{N,\gamma,\frac{N}{2}}$. Now choose $0 < \rho < 1$. The complement of $W_{N,\gamma,N-1}$ can be written as

$$W_{N,\gamma,N-1}^C = \left\{\max_{1 \leq \tau \leq \frac{N}{2}} \frac{|r_x(\tau)|}{r_x(0)} > \gamma\right\} \subset \left\{\max_{1 \leq \tau \leq \frac{N}{2}} \frac{|r_x(\tau)|}{N} > \gamma\rho\right\} \cup \left\{\frac{1}{N}\sum_{t=0}^{N-1} x(t)^2 < \rho\right\} \tag{41}$$

The probability of the first set is bounded by (40), setting $\epsilon = \gamma\rho$. The probability of the second set can be bounded by another use of the Hoeffding inequality, applied to the bounded IID random variables $x(t)^2$. Putting everything together,

$$\mathcal{P}\left(W_{N,\gamma,N-1}^C\right) \leq 3e^{Log(N)\left(1 - \frac{N\gamma^2\rho^2}{12K^4 Log(N)}\right)} + e^{\frac{-2N(1-\rho)^2}{K^4}} \tag{42}$$

The second term clearly goes to to zero as $N \to \infty$, and the same happens with the first term since by hypothesis $\gamma(N)\sqrt{\frac{N}{log(N)}} \overset{N \to \infty}{\longrightarrow} \infty$.

**Part 3:** Assume $x(0) \cdots x(N-1)$ are *Gaussian* random variables, $x(t) \sim \mathcal{N}(0,1)$. Choosing $K(N) = \sqrt{2Log(N)}$, define the random variables $v(t), t = 0 \cdots N-1$ by truncation:

$$v(t) = \begin{cases} x(t) & if |x(t)| \leq K(N) \\ 0 & otherwise \end{cases} \tag{43}$$

26

$$\mathcal{P}\left(x \neq v\right) \leq N\mathcal{P}\left(x(t) \neq v(t)\right) = N\mathcal{P}\left(|x(t)| > K(N)\right) \leq \frac{NC}{K(N)}e^{\frac{-K(N)^2}{2}} = \frac{C}{\sqrt{2Log(N)}} \quad (44)$$

In (44) $x = (x(0)\cdots x(N-1))$, $v = (v(0)\cdots v(N-1))$, and the second inequality follows from a standard bound to the tail of the normal distribution (C is a constant). Observing that

$$\mathcal{P}\left(x \notin W_{N,\gamma,N-1}\right) \leq \mathcal{P}\left(v \notin W_{N,\gamma,N-1}\right) + \mathcal{P}\left(x \neq v\right) \quad (45)$$

it remains to show that $\mathcal{P}\left(v \notin W_{N,\gamma,N-1}\right)$ also vanishes as $N \to \infty$. Since the variables $v(t)$ are bounded by $K(N)$, (42) gives

$$\mathcal{P}\left(v \notin W_{N,\gamma,N-1}\right) \leq 3e^{Log(N)\left(1 - \frac{N\gamma^2\rho^2}{12K^4Log(N)}\right)} + e^{\frac{-2N(1-\rho)^2}{K^4}} \quad (46)$$

The second term clearly has limit 0 as $N \to \infty$. The first term also goes to 0, since by hypothesis $\frac{N\gamma^2\rho^2}{K^4Log(N)} = \frac{\rho^2}{4}\frac{N\gamma^2}{Log(N)^3}$ goes to infinity.

$\square$

## Proof of Theorem 5:

For simplicity we consider the complex Gaussian case. From Proposition 1, $\frac{s_x(k)}{N}$, $0 \leq k \leq N-1$ are independent, exponential random variables with mean 1. (for the real case, the same holds for $0 < k < \frac{N}{2}$, but $s_x(0)$, $s_x(\frac{N}{2})$ have a different distribution, and the spectrum is symmetric around $\frac{N}{2}$; the proof can be extended to this case). If $\Lambda_K$ is the sum of $K$ independent, exponential, mean 1 variables, the bound

$$\mathcal{P}\left(\Lambda_K > u\right) \leq \left(\frac{K+p}{u}\right)^p \quad (47)$$

27

holds for any positive integer $p$. Let $\epsilon = \frac{\alpha}{3}$, $K = \frac{N}{M}$. We can write

$$\hat{W}_{N,\alpha,V}^C \subset \left\{ \frac{1}{N} \|x\|^2 < (1-\epsilon) \right\} \cup \bigcup_{m=1}^{M} \left\{ \langle \frac{s_x}{N}, V_m \rangle > (1+\alpha)(1-\epsilon) \right\} \qquad (48)$$

Focusing on each set in the second union, $K \langle \frac{s_x}{N}, V_m \rangle$ is the sum of $K$ exponential, mean 1 random variables. Choose $p = \lfloor \frac{K\alpha}{3} \rfloor$. Using (47),

$$\mathcal{P} \left( \bigcup_{m=1}^{M} \left\{ \langle \frac{s_x}{N}, V_m \rangle > (1+\alpha)(1-\epsilon) \right\} \right) \leq M \left( \frac{K+p}{K(1+\alpha)(1-\epsilon)} \right)^p = e^{LogM \left[ 1 - \frac{p}{LogM} Log \left( \frac{(1+\alpha)(1-\epsilon)}{1+\frac{p}{K}} \right) \right]}$$

$$(49)$$

As $N \to \infty$, $\alpha \to 0$, and $\frac{p}{K} \sim \frac{\alpha}{3}$, $\epsilon = \frac{\alpha}{3}$. This gives

$$\frac{p}{LogM} Log \left( \frac{(1+\alpha)(1-\epsilon)}{1+\frac{p}{K}} \right) \sim \frac{p}{LogM} \frac{\alpha}{3} \sim \frac{K\alpha^2}{9LogM} = \frac{N\alpha^2}{9MLogM} \xrightarrow{N \to \infty} \infty. \qquad (50)$$

Therefore the probability in (49) goes to zero as $N \to \infty$. As for the first set in (48),

$$\mathcal{P} \left( \frac{1}{N} \|x\|^2 < (1-\epsilon) \right) = \mathcal{P} \left( \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} (x(t)^2 - 1) < -\sqrt{N}\epsilon \right) \xrightarrow{N \to \infty} 0 \qquad (51)$$

since $\sqrt{N}\epsilon \geq \frac{\sqrt{K}\alpha^2}{3} \xrightarrow{N \to \infty} \infty$, and using asymptotic normality.

$\square$

# Acknowledgements

# References

[1] Abarbanel H., *The Analysis of Observed Chaotic Data in Physical Systems*, Reviews of Modern Physics, Vol 65(4), pp.1331-1392, 1993.

[2] Anderson R.L., Anderson T.W., *Distribution of the circular serial coefficient for residuals from a fitted Fourier series*, Ann. Math. Statist., 21, pp. 59-81, 1956.

[3] Bartlett M.S., *On the theoretical Specification and Sampling Properties of Autocorrelated Time-Series*, J.Royal Statist. Society, 8, pp 27-41, 1946.

[4] Billingsley P., *Convergence of Probability Measures*, J.Wiley & Sons, New York, 1968.

[5] Breiman L., 1968, *Probability*, Addison-Wesley, 1968.

[6] Brillinger D., *Time Series Data Analysis and Theory*, Mc Graw-Hill, 1981.

[7] D'Andrea R., Paganini F., *Interconnection of Uncertain Behavioral Systems for Robust Control*, Proceedings 1993 CDC, San Antonio, Texas, pp. 3642-3647.

[8] D'Andrea R., Paganini F., *Robust Control Synthesis with "Behavioral" Constraints* Submitted to 1994 CDC, Orlando, Florida

[9] Dahleh M.A, Theodosopoulos T.V., Tsitsiklis J., *The sample complexity of worst-case identification of FIR linear systems*, Systems and Control Letters, 20, pp 157-166, 1993.

[10] Gardner, W.A., *Statistical Spectral Analysis - A Non-Probabilistic Theory*, Prentice-Hall, Englewood Cliffs, New Jersey, 1988.

[11] Hoeffding W., *Probability Inequalities for Sums of Bounded Random Variables*, American Statistical Association Journal, 59, pp. 13-30, March 1963.

[12] Kendall M.G., Stuart A., *The Advanced Theory of Statistics*, Hafner Publishing, 1966.

[13] Megretski A., *S-Procedure in Optimal Non-Stochastic Filtering,* Tech. Rep. Code Trita/Mat-92-0015, Royal Inst. of Tech., Sweden.

[14] Paganini F., *Set Descriptions of White Noise and Worst Case Induced Norms,* Proceedings 1993 CDC, San Antonio, Texas, pp. 3658-3663.

[15] Paganini F., D'Andrea R., and Doyle J., *Behavioral Approach to Robustness Analysis* Proceedings 1994 ACC, Baltimore, MD, pp. 2782-2786.

[16] Poolla K., Tikku A., *On the Time Complexity of Worst-Case System Identification* IEEE Trans AC., Vol 39 (5), pp. 944-950, 1994.

[17] Stoorvogel A.A., *The Robust $\mathcal{H}_2$ Control Problem: A Worst-Case Design*, IEEE Trans. A.C., Vol 38, 9,pp. 1358-1370, 1993.

[18] Whittle P., *Curve and Periodogram Smoothing*, J. Roy. Stat. Soc., B, 19, pp 38-47, 1957.

[19] Wiener N., 1930, *Generalized Harmonic Analysis*, Acta Mathematika 55, pp. 117-258.

[20] Yakubovich V.A., *Dichotomy and absolute stability of nonlinear systems with periodically nonstationary linear part.* Systems and Control Letters, 11(3), pp. 221-228, 1988.