Set Descriptions of White Noise and Worst Case Induced Norms

Fernando Paganini
Electrical Engineering, M/S 116-81
California Institute of Technology
Pasadena, CA 91125

October 15, 1993

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 described in a worst case setting. This type of modeling of noise and disturbances very much fits the philosophy of both the behavioral and robust control settings. The description is based on properties of finite records of signals, which may be verified directly on experimental data. Bounds of system gain for input signals in these sets are given, and their asymptotic behavior for long data records is analyzed.

1 Introduction

The presence of low-correlated disturbances (noise) in physical systems has usually been modeled by thinking of the disturbance as the realization of a stochastic process, which is white in the sense of having zero autocorrelations in the expected value sense. The basic result for analysis of linear systems in the presence of stochastic noise is that if unit variance white noise is input to a stable linear system, the output variance (expected value of power) is given by the 2-norm of the system function. Moreover, the spectral characteristics of the output signal are given by the filter.

However, if in a real-world situation we want to use results of this type, we will have to convince ourselves that our disturbances can be accurately modeled as a stochastic white noise trajectory. Trying to decide this from experimental data leads to a statistical hypothesis test on a finite record of the signal. In other words, we will accept our signal as white noise if it belongs to a certain set, designed to give us reasonable confidence in the whiteness of the source. This set is typically described in terms of time autocorrelations.

This paper intends to show that these sets themselves contain sufficient information about the signals to predict properties of the system response. In other words, since what we measure are time correlations of the signals, and what we are trying to predict are system gains and spectral characteristics of the outputs, it seems natural to tie these together directly, without need of modeling the (stochastic or deterministic) mechanism that generates the disturbances.

The advantage of this approach, and the main motivation for this work, is that set descriptions and worst case gains are the usual tools in which robust control has been able to integrate both disturbances and other descriptions of uncertainty in system models. For a review of this we refer to [5]. So far, however most work in this field (a notable exception is [8]) has taken the conservative step of allowing larger classes of disturbances (arbitrary bounded power signals, or arbitrary signals in a ball of L_2 or L_∞), ignoring spectral properties of the disturbances.

In many practical situations these larger classes imply a large conservatism; we do not expect a perturbation such as, for example, thermal noise on a resistor to be a persisting sinusoid; we know it will be a "white" signal. The only reason robust control has looked at a larger class is to allow a simple unified framework with further uncertainty in the system model. By using deterministic versions of spectral constraints, we attempt here to integrate these "whiteness" properties with the worst case paradigm.

The deterministic approach to spectral analysis is not new and goes as far back as Wiener [1]; a modern reference is [2]. In those treatments, however, most of the attention is devoted to estimation of the underlying spectrum and stochastics are usually avoided by writing only asymptotic results (as the length of the data record goes to infinity). Here we will focus on worst case gains (we care about disturbance rejection, rather than estimation) and since in practical applications we only have finite records of the signals, we will try to get as much as we can from them and later consider asymptotic results.

We consider finite records of the signals, and define sets of signals in terms of constraints on correlations (or dually, of FFTs), including parameters that characterize the "whiteness" of the signals in the set. The norm induced in the system (worst case gain in energy with signals in the set) can be bounded in terms of these parameters. By appropriate choice of the evolution of these parameters as the length of the data record grows, we can recover asymptotically the 2-norm as the relevant system norm, now understood in a *worst case* setting. Furthermore, this is consistent with defining the sets such that they include a stochastic white noise realization with "large" probability, so nothing is lost with this approach.

Finally, this way of describing connects up very naturally with behavioral descriptions of linear systems, as introduced by Willems [3]. In that framework, systems are described by behaviors, i.e. sets of trajectories compatible with the system laws. To study these systems in the presence of noise in some of the variables amounts to intersecting the behavior with the noise sets in these variables. It is further shown that the constraints which define our sets can be expressed as *uncertain* behavioral equations as introduced in [4]. This would provide a unified framework to analyze systems in the presence of white noise as well as other forms of uncertainty.

2 Assumptions and notation

We will consider discrete time signals and systems. We are interested in describing noise signals and characterizing system response to them in steady state (ignoring transients) in terms of finite records of signals. We will assume that the systems are linear time invariant and stable, and to get sensible answers that the length N of our data record is much larger than the system time constants (in the FIR case, $N \gg T$, duration of system response). Under this assumption, there is no loss in generality 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 of the input signals. Not only do we get more tractable expressions, but this is also the natural thing to look at if we are interested in "steady state" response of the system to persistent disturbances as opposed to the transient response.

Let $x(t)$ be a periodic, real valued signal, of period N, which we will often identify 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

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

We introduce the autocorrelation sequence of x (circular autocorrelations of the vector $x(0)\cdots x(N-1)$):

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

It is well known that the sequences $r_x(\tau), \tau = 0\cdots N-1$ and the "power spectrum" $S(k) = |X(k)|^2$, $k = 0\cdots N-1$ form a DFT pair.

For vector-valued signals $x(t) \in \mathbb{R}^n$, we consider the matrix autocorrelation (prime denotes transpose)

$$R_x(\tau) = \sum_{t=0}^{N-1} x(t+\tau)x'(t)$$

The DFT can be defined for a vector or matrix valued sequence with the same formulas as above, and in this case we find that $S(k) = X(k)X^*(k), k = 0\cdots N-1$ is the DFT of $R_x(\tau)$.

For a vector-valued N-periodic signal x(t), we will use as norm the energy over the period,

$$\|x\|^2 = \sum_{t=0}^{N-1} \|x(t)\|^2 = trace(R_x(0))$$

We consider a stable, discrete time linear time invariant system with in general m inputs and p outputs (q is the shift operator):

$$\mathcal{H}(q) = \sum_{t=-\infty}^{\infty} H(t)q^{-t}$$

The frequency response (Fourier transform of H(t)) is denoted by $\mathcal{H}(e^{j\omega})$. The 2-norm of the system is given by

$$\|\mathcal{H}\|_2^2 = \sum_{t=-\infty}^{\infty} trace(H(t)H'(t)) = \frac{1}{2\pi}\int_0^{2\pi} trace\left(\mathcal{H}(e^{j\omega})^*\mathcal{H}(e^{j\omega})\right) d\omega$$

We define the autocorrelation matrix of H by

$$R_H(\tau) = \sum_{t=-\infty}^{\infty} H(t+\tau)H'(t)$$

assuming convergence for every τ . This implies $\|\mathcal{H}\|_2^2 = trace(R_H(0))$.

For a SISO system we will use lowercase letters, $h(t)$ denoting the system response, $r_h(\tau)$ the autocorrelation function, $r_h(0) = \|\mathcal{H}\|_2^2$.

3 Time domain descriptions

To keep the presentation simple, we will restrict ourselves in this section to scalar noise signals which are inputs to a SISO system. We will try to give a set description of what it means for a noise signal to be white, and characterize the response of the system. Multivariable versions are delayed to a later section. The first result is the following:

Lemma 1 *Let $\mathcal{H}$ be a SISO discrete time system,*

$$\mathcal{H}(q) = \sum_{t=-\infty}^{\infty} h(t)q^{-t}$$

Let $u(t),\quad t = 0\cdots N-1$ be a fragment of an N-periodic input signal to system $\mathcal{H}$, $y(t),\quad t = 0\cdots N-1$ be the corresponding steady state (periodic) output. Then

$$r_y(\tau) = \sum_{t=-\infty}^{\infty} r_h(t)r_u(t-\tau) \tag{1}$$

In particular,

$$\|y\|_2^2 = r_y(0) = r_u(0)r_h(0) + 2\sum_{\tau=1}^{\infty} r_u(\tau)r_h(\tau) \tag{2}$$

Proof:

$$\begin{aligned} r_y(\tau) &= \sum_{n=0}^{N-1} y(n+\tau)y(n) \\ &= \sum_{n=0}^{N-1} \sum_{k=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} h(k)h(m)u(n-k)u(n+\tau-m) \\ &= \sum_{k=-\infty}^{\infty} \sum_{t=-\infty}^{\infty} h(k)h(k+t)r_u(\tau-t) \;=\; \sum_{t=-\infty}^{\infty} r_h(t)r_u(t-\tau) \end{aligned}$$

□

The previous statement shows that the time autocorrelations of u map in a simple way to their counterparts in y. So any information we have on the structure of the r_us can be easily transferred to the r_ys.

We now concentrate in defining what we mean by a "white" signal. Clearly, we are looking for signals such that the autocorrelation vector looks like a delta function. In other words a signal $x(t)$ is white if $r_x(\tau)$ is small for nonzero τ. We will only require this to hold for values of τ smaller than a horizon parameter T. We have the following:

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

$$|r_x(\tau)| \leq \gamma r_x(0) \qquad \tau = 1 \cdots T \tag{3}$$

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

The need for a band as defined by the parameter γ is not surprising, since we do not expect finite length autocorrelations to be exactly zero. A natural description for white signals would involve γ decaying to 0 as N goes to infinity, i.e. as the averaging increases. Determination of an appropriate γ for a given experimental record can be done easily from an autocorrelations plot which is in fact a common tool in time series analysis and system identification. Figure 1 below depicts such a plot for a pseudorandom sequence.

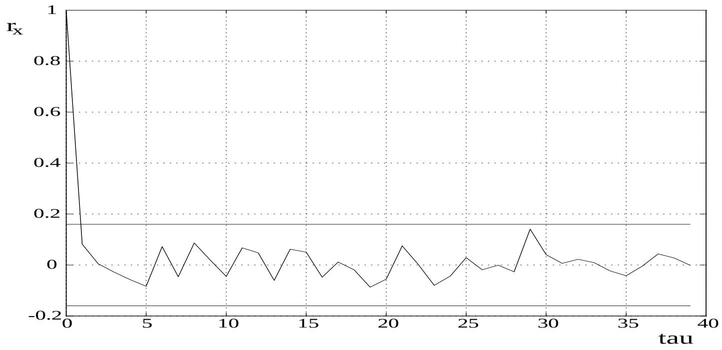

Figure 1: Autocorrelation plot of a pseudorandom sequence

The introduction of a horizon T in which we require our autocorrelations to be small (instead of forcing $T = N-1$) gives us more generality. Typically, we only care about low correlation in time scales where the system responds strongly.

Both γ and T are, in fact, a parameterization of a rectangular weight function which specifies our constraints on the autocorrelation: we could rewrite the constraints in our set as $|r_x(\tau)| \leq v(\tau) r_x(0)$, with $v(\tau) = \gamma$ for $0 < \tau \leq T$, $v(\tau) = 1$ for $\tau > T$.

Other shapes of this weight function v could be considered, we will comment briefly on this at the end of the section.

To gain insight into the response of the system when the input is white noise, we will restrict ourselves for now to the FIR case,

$$\mathcal{H}(q) = \sum_{t=0}^{T} h(t) q^{-t} \tag{4}$$

The next result follows immediately from Lemma 1.

Theorem 1 *Assume $\mathcal{H}$ is FIR as in (4), and the input $u \in W_{N,\gamma,T}$, then*

$$\left| \frac{\|y\|_2^2}{\|u\|_2^2} - \|\mathcal{H}\|_2^2 \right| \leq 2\gamma \sum_{\tau=1}^{T} |r_h(\tau)| \tag{5}$$

Also, if we denote $\|\mathcal{H}\|_{W_{N,\gamma,T}} = \sup\{\frac{\|y\|_2}{\|u\|_2}, u \in W_{N,\gamma,T}\}$ *, the induced system norm (worst case gain) with respect to signals in* $W_{N,\gamma,T}$*, then*

$$\|\mathcal{H}\|_2^2 \leq \|\mathcal{H}\|_{W_{N,\gamma,T}}^2 \leq \|\mathcal{H}\|_2^2(1-\gamma) + \gamma \sum_{\tau=-T}^{T} |r_h(\tau)| \tag{6}$$

If in addition $u \in W_{N,\gamma,2T}$*, then*

$$\left|\frac{r_y(\tau)}{r_u(0)} - r_h(\tau)\right| \leq \gamma \sum_{t=-T}^{T} |r_h(t)| \qquad \tau = -T \cdots T \tag{7}$$

The inequalities (6) are bounds for $\|\mathcal{H}\|_{W_{N,\gamma,T}}$, the measure of disturbance rejection of the system in the presence of disturbances from the set $W_{N,\gamma,T}$. The upper bound is simply expressed as a convex combination, weighted by γ of the 2-norm of the system and the 1-norm of the autocorrelations sequence. If γ is small, we are as expected close to the 2-norm of the system.

Inequality (7) lets us conclude that for small γ, (and using a time horizon $2T$) the autocorrelations of the time series y are essentially determined by the autocorrelations of the filter. More precisely, they lie (up to a constant factor $\|u\|^2$) in a band of width $\gamma \sum_{\tau=-T}^{T} |r_h(\tau)|$, centered in the autocorrelations of the filter.

Therefore, this band is a natural set description for colored noise. Note that in this way we are able to reinterpret the usual methods of time series analysis, which essentially consist in writing the observed signal as a filtered version of a white "innovations" signal. The modeling is considered complete when these innovations satisfy a "whiteness" test, usually of the form described by our sets.

Going back to (6), a natural question is whether the upper bound on $\|\mathcal{H}\|_{W_{N,\gamma,T}}$ is tight. It is easy to observe that if we set $\gamma = 1$, there are no restrictions other than periodicity in the input signal, and the induced norm can be bounded by the H_∞ norm of the system which in this 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 obtained. However, for small enough γ the bound is tight:

Proposition 1 *In the conditions of Theorem 1, if* $\gamma < 1/T$*, then for large enough* N *we have* $\|\mathcal{H}\|_{W_{N,\gamma,T}}^2 = \|\mathcal{H}\|_2^2(1-\gamma) + \gamma \sum_{\tau=-T}^{T} |r_h(\tau)|$

Proof: Consider an input $u \in W_{N,\gamma,T}$, with $\|u\| = 1 = r_u(0)$. Then the numbers $r_u(\tau), \tau = 1 \cdots T$ all lie in the interval $[-\gamma, \gamma]$. From Lemma 1,

$$\|y\|_2^2 = r_y(0) = r_h(0) + 2\sum_{\tau=1}^{T} r_u(\tau) r_h(\tau) \tag{8}$$

So if we can find a sequence $u(0) \cdots u(N-1)$ whose autocorrelations take the values $\pm\gamma$ with signs matching those of $r_h(\tau)$, then the upper bound is achieved for this sequence. This leads us to the following question: under what conditions can a set of numbers $r_0 \cdots r_T$ be obtained as the autocorrelations of a sequence? It turns out that if the Toeplitz matrix

$$\mathcal{R}_T = \begin{bmatrix} r_0 & \cdots & \cdots & \cdots & r_T \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & r_0 & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ r_T & \cdots & \cdots & \cdots & r_0 \end{bmatrix}$$

is positive definite it can be shown (see [7], for example) that there is a stochastic process with those autocorrelations, and it is not hard to conclude from here that for sufficiently large N, a finite sequence with those autocorrelations can be found.

It remains to show that if $\gamma < 1/T$, choosing $r_0 = 1$, and $r_\tau = \pm\gamma$ for $\tau = 1 \cdots T$, then the corresponding $\mathcal{R}_T$ matrix is positive definite.

To see this, consider the matrix $I - \mathcal{R}_T$, which has 0 on the diagonal and $\pm\gamma$ elsewhere. Clearly the maximum row sum norm of $I - \mathcal{R}_T$ is $\gamma T < 1$, therefore $\lambda_{max}(I - \mathcal{R}_T) < 1$ and $\mathcal{R}_T$ is positive definite.

$\square$

From the preceding discussion, it is clear that we obtain sharper bounds when the parameter γ is small, so we would want γ as small as possible; however, pushing it too low might leave out the signals we want to describe. To get some insight into this compromise, we turn to asymptotic results. First, an obvious consequence of Theorem 1 is

Proposition 2 *If T is fixed, H as in (4), and $\gamma(N) \stackrel{N\to\infty}{\longrightarrow} 0$, then*

$$\|\mathcal{H}\|_{W_{N,\gamma,T}} \stackrel{N\to\infty}{\longrightarrow} \|\mathcal{H}\|_2$$

The next result is useful in obtaining a consistency check with the stochastic approach, and ensure we are not being too restrictive in our sets. We pose the following question: if the signals effectively come from a stochastic white noise source, what decay rate of $\gamma(N)$ will guarantee that the signals fall in $W_{N,\gamma,T}$ with large probability? A reasonably general answer is the following ($\mathcal{P}$ denotes probability):

Proposition 3 *Let T be fixed, $u(0), \cdots, u(N-1)$ be independent, identically distributed random variables, with 0 mean and finite variance.*

If $\gamma(N)\sqrt{N} \stackrel{N\to\infty}{\longrightarrow} \infty$, then $\mathcal{P}\left(u \in W_{N,\gamma,T}\right) \stackrel{N\to\infty}{\longrightarrow} 1$.

Proof: See the Appendix.

Returning to the general IIR case, we write extensions of Theorem 1. The proof is immediate from Lemma 1 and the definition of $W_{N,\gamma,T}$.

Theorem 2 *Let $\mathcal{H}$ be an IIR filter, with $\sum_{-\infty}^{\infty} |r_h(t)| < \infty$. In the conditions of Lemma 1, if in addition $u \in W_{N,\gamma,T}$, then*

$$|\frac{r_y(\tau)}{r_u(0)} - r_h(\tau)| \leq \gamma \sum_{\substack{t=\tau-T \\ t\neq\tau}}^{\tau+T} |r_h(t)| + \sum_{|t-\tau|>T} |r_h(t)| \tag{9}$$

Focusing on the case $\tau = 0$, and defining $\|\mathcal{H}\|_{W_{N,\gamma,T}}$ as before, we have

$$\left| \|\mathcal{H}\|^2_{W_{N,\gamma,T}} - \|\mathcal{H}\|^2_2 \right| \leq \gamma \sum_{\tau=-T}^{T} |r_h(\tau)| + \sum_{|t|>T} |r_h(t)| \tag{10}$$

The bound on system gain depends explicitly on the parameter T. If T is chosen large with respect to the decay rate of $h(t)$, and γ is chosen small, we expect $\|\mathcal{H}\|_{W_{N,\gamma,T}}$ to be close to the 2-norm of $\mathcal{H}$. As we take longer data records (N large) we can think of also making T grow with N, to cancel the "tail" term asymptotically. Once more, we look at growth rates of these parameters that give us asymptotically the 2-norm, while retaining a rich enough set. The following results extend Propositions 2 and 3, with slightly stronger hypothesis:

Proposition 4 *If $T(N) \stackrel{N\to\infty}{\longrightarrow} \infty$, and $\gamma(N) \stackrel{N\to\infty}{\longrightarrow} 0$, then*

$$\|\mathcal{H}\|_{W_{N,\gamma,T}} \stackrel{N\to\infty}{\longrightarrow} \|\mathcal{H}\|_2$$

Proof: Immediate from (10).

Proposition 5 *Let $u(0), \cdots, u(N-1)$ be independent, identically distributed random variables,with 0 mean and finite 4th order moment.*

If for some constant $\nu > 1$, we have $\gamma(N)\sqrt{\frac{N}{log(N)}} > \nu$, and $T(N)\sqrt{\frac{log(N)}{N}} \stackrel{N\to\infty}{\longrightarrow} 0$, then $\mathcal{P}\left(u \in W_{N,\gamma,T}\right) \stackrel{N\to\infty}{\longrightarrow} 1$.

Proof: See the Appendix.

In the previous result we had to pose restrictions on the growth of $T(N)$ to guarantee that our set of signals is asymptotically at least as rich as stochastic white noise. This motivates the inclusion of this horizon parameter in a general noise description.

As a comment, even though these asymptotic results are comforting, we will have no guarantees as to the accuracy of any stochastic model for our disturbances; so other choices of evolution of γ and T that make the sets richer than stochastic white noise might be of interest. From a practical point of view, we will always be dealing with finite data records, so it is the finite horizon results which are most useful to characterize disturbance rejection properties in the system.

To elaborate on this issue a little further, let us assume that using some standard system identification technique (as can be found, for example, in [6]) we have obtained a model for some part of a system which includes some "residuals", i.e. a disturbance variable needed to explain the experimental data. Typically the model we have fit is a great simplification from the physical reality; so the disturbance will account for the more complex phenomena which we have neglected (probably nonlinear and not close at all to a stochastic white process).

From the empirical values of this disturbance, however, we may directly obtain autocorrelation constraints which we have confidence the disturbance will satisfy, whatever its nature. The previous results will give us hard bounds on system rejection of this disturbance. If we get tight values of γ, we will be close to the 2-norm of the system; $\gamma = 1$ gives the $\mathcal{H}_\infty$ norm of the system. For intermediate values we can use the bounds obtained above.

The only (inevitable) leap of faith when we predict future behavior of the system is that the experiment was representative of this behavior (so that the disturbance will not be "worse" in the future) and that the system response is negligible for time scales longer than the length of the experiment.

To summarize, we have described white noise signals in the time domain by constraining the autocorrelation function by a rectangular weight function. We have given hard bounds on the system gain in the sets described by these constraints, and conditions on the evolution of the weight function parameters with the data record to ensure desirable asymptotic properties. The shape chosen for the weight function could of course be changed, and the results in theorems 1 and 2 could be extended with little effort, the bounds involving a weighted 1 norm of the function $r_h(t)$.

4 Frequency domain descriptions

We consider dual descriptions of white noise signals in terms of DFTs.

Lemma 2 *In the conditions of Lemma 1, let U(k),Y(k), be the N-point DFTs of the signals u(t), y(t). Let* $H(k) = \mathcal{H}(e^{j\frac{2\pi}{N}k})$ *where* $\mathcal{H}(e^{j\omega})$ *is the frequency response (in the FIR case,* $T < N$*, H(k) coincides with the N-point DFT of h(t)). Then*

$$Y(k) = H(k)U(k) \tag{11}$$

and therefore the power spectra verify

$$|Y(k)|^2 = |H(k)|^2|U(k)|^2 \tag{12}$$

Proof:

$$Y(k) = \sum_{n=0}^{N-1}\left(\sum_{t=-\infty}^{\infty} h(t)u(n-t)\right)e^{-j\frac{2\pi}{N}kn} =$$

$$= \sum_{t=-\infty}^{\infty} h(t)e^{-j\frac{2\pi}{N}kt}\sum_{n=0}^{N-1} u(n-t)e^{-j\frac{2\pi}{N}k(n-t)} = H(k)U(k)$$

□

Now we want to translate what we mean by "white" signals in terms of the DFTs: we want the power spectrum to be close to a constant. If X(k) is the DFT of a noise signal, we might think of requiring $\frac{|X(k)|^2}{\|x\|^2} - 1$ to be small for all k. However, observing a typical DFT of a noise signal we find that the $|X(k)|^2$ is very erratic and only close to constant in an average

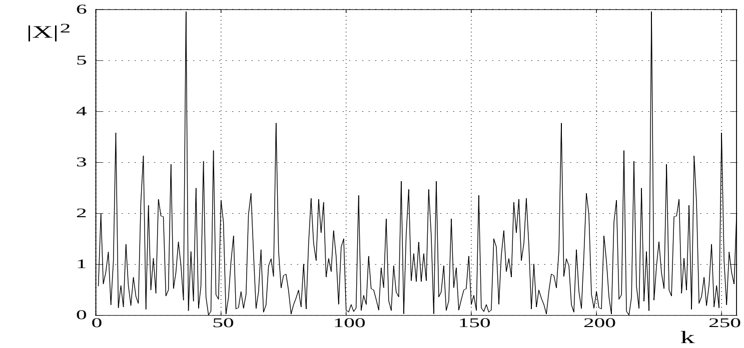

Figure 2: Power spectrum of a pseudorandom sequence

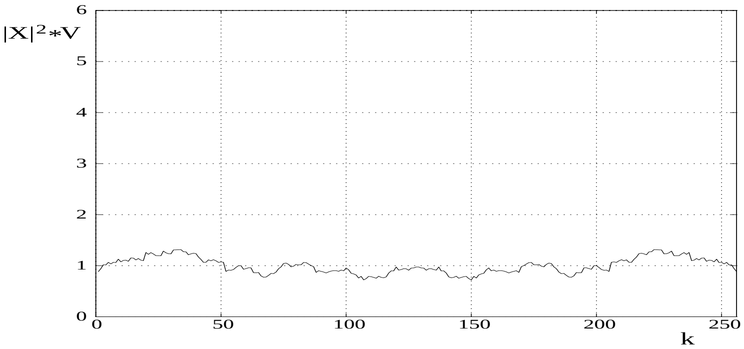

Figure 3: Smoothed power spectrum

sense. This is illustrated in figure 2, which depicts a power spectrum of a signal generated by a pseudorandom number generator.

So to uncover the underlying "whiteness" we need to perform some local averaging in the signals; this idea is very well known in statistical spectral analysis (smoothing of "periodograms" by windowing, as appears for example in [2]). The idea is to convolve the DFT with a window function $V(k)$ which performs the averaging, and we can require that the smoothed signal be close to constant. Figure 3 shows the result of smoothing on the previous power spectrum, using a rectangular frequency window of width (see definition below) $K_V \approx \sqrt{N}$.

We give the following definitions:

Definition 2 *A frequency window is an N-periodic sequence V(k), such that* $\sum_{n=0}^{N-1} V(k) = 1$. *The circular convolution of an N-periodic sequence X(k) with V(k) is defined as*

$$(X * V)(k) = \sum_{l=0}^{N-1} X(k-l)V(l) = \sum_{-\frac{N}{2}<l\leq\frac{N}{2}} X(k-l)V(l)$$

The width parameter of a window is defined as

$$K_V = 2 \sum_{-\frac{N}{2} < k \leq \frac{N}{2}} |kV(k)|$$

Motivation for the choice of the width parameter can be found in the rectangular window defined as $V(k) = \frac{1}{2B}$ for $-B < k \leq B$, and as 0 in the rest of the period, which has width B.

Definition 3 *A DFT $X(k)$ of a signal x(t) of length N is said to be white with level α, with respect to a frequency window V(k), if*

$$\max_k \left| \frac{(|X|^2 * V)(k)}{\|x\|^2} - 1 \right| \leq \alpha \tag{13}$$

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

This smoothing by convolution in the frequency domain corresponds in the time domain to multiplying the autocorrelation sequence by a weight. A variety of shapes of windows could be chosen, leading to slightly different sets of noise signals. The main parameter, however, is the width of the window. We explore its influence on the system norm:

Theorem 3 *Assume $M = \max_\omega \frac{d}{d\omega} |\mathcal{H}(e^{j\omega})|^2$ is finite. Then*

$$\|\mathcal{H}\|_2^2 - \frac{2\pi M}{N} \leq \|\mathcal{H}\|_{\hat{W}_{N,\alpha,V}}^2 \leq \left(\|\mathcal{H}\|_2^2 + \frac{2\pi M}{N} \right)(1+\alpha) + \frac{\pi K_V M}{N} \tag{14}$$

Proof:

First observe that, setting once more $H(k) = \mathcal{H}(e^{j\frac{2\pi}{N}k})$, we have

$$\left| \|\mathcal{H}\|_2^2 - \frac{1}{N} \sum_{k=0}^{N-1} |H(k)|^2 \right| \leq \frac{2\pi M}{N}$$

This can be seen by approximating the integral $\int_0^{2\pi} |\mathcal{H}(e^{j\omega})|^2 d\omega$ and bounding the difference using the derivative bound (in fact, in the FIR case the difference is 0). So it suffices to show that

$$\frac{1}{N} \sum_{k=0}^{N-1} |H(k)|^2 \leq \|\mathcal{H}\|_{\hat{W}_{N,\alpha,V}}^2 \leq (1+\alpha) \frac{1}{N} \sum_{k=0}^{N-1} |H(k)|^2 + \frac{\pi K_V M}{N}$$

The left inequality is immediate by considering the gain on the signal $U(k) \equiv 1$ which is in the set. For the right hand side, take any $U(k) \in \hat{W}_{N,\alpha,V}$, $\frac{1}{N} \sum_{k=0}^{N-1} |U(k)|^2 = 1 = \|u\|^2$. Then

$$\begin{aligned}
& \left| \frac{1}{N} \sum_{k=0}^{N-1} |Y(k)|^2 - \frac{1}{N} \sum_{k=0}^{N-1} |H(k)|^2 (|U|^2 * V)(k) \right| \\
= \ & \frac{1}{N} \left| \sum_{k=0}^{N-1} \sum_{-\frac{N}{2} < l \leq \frac{N}{2}} |Y(k-l)|^2 V(l) - \sum_{k=0}^{N-1} |H(k)|^2 \sum_{-\frac{N}{2} < l \leq \frac{N}{2}} |U(k-l)|^2 V(l) \right| \\
\leq \ & \frac{1}{N} \sum_{k=0}^{N-1} \sum_{-\frac{N}{2} < l \leq \frac{N}{2}} \left| |H(k-l)|^2 - |H(k)|^2 \right| |U(k-l)|^2 |V(l)| \\
\leq \ & \frac{\pi M}{N} \sum_{-\frac{N}{2} < l \leq \frac{N}{2}} 2|lV(l)| \frac{1}{N} \sum_{k=0}^{N-1} |U(k-l)|^2 = \frac{\pi K_V M}{N}
\end{aligned}$$

In the last inequality, $|H(k-l)|^2 - |H(k)|^2$ is the increment of the function $|\mathcal{H}(e^{j\omega})|^2$ on an interval of length $|l|2\pi/N$, so it can be bounded by $M|l|2\pi/N$. The previous inequalities give

$$\|y\|_2^2 = \frac{1}{N}\sum_{k=0}^{N-1}|Y(k)|^2 \leq \frac{1}{N}\sum_{k=0}^{N-1}|H(k)|^2(|U|^2 * V)(k) + \frac{\pi K_V M}{N}$$

Using the fact that $U(k) \in \hat{W}_{N,\alpha,V}$ we have

$$\|y\|_2^2 = \frac{1}{N}\sum_{k=0}^{N-1}|Y(k)|^2 \leq (1+\alpha)\frac{1}{N}\sum_{k=0}^{N-1}|H(k)|^2 + \frac{\pi K_V M}{N}$$

□

The following is an immediate consequence:

Proposition 6 *In the same hypothesis as above, if in addition* $\alpha(N) \stackrel{N\to\infty}{\longrightarrow} 0$ *and* $\frac{K_V(N)}{N} \stackrel{N\to\infty}{\longrightarrow} 0$*, then*

$$\|\mathcal{H}\|^2_{\hat{W}_{N,\alpha,V}} \stackrel{N\to\infty}{\longrightarrow} \|\mathcal{H}\|_2^2$$

As to the requirements on the growth of K_V that would allow usual white signals to fall in the set $\hat{W}_{N,\alpha,V}$, we haven't pursued analytic proofs with stochastic white noise (which are not as clean), but extensive simulations with pseudorandom number generators show that a choice of $K_V \approx \sqrt{N}$ is compatible with $\alpha(N) \stackrel{N\to\infty}{\longrightarrow} 0$.

We have introduced essentially dual descriptions in time and frequency domains, of parameterized sets which describe white noise. A natural question is whether, with the adequate choice of weight/window in each domain, the sets are the same. The difficulty here is that the vector norms used to measure distances (to the delta function in the time domain, to a constant in the frequency domain) do not correspond to each other through DFT. One could think of using the Euclidean norm, which has this property, at the expense of complicating the description of the sets (these would involve Euclidean norms of autocorrelation sequences, implying expressions of fourth order in the original signal). In any case, it is clear that for appropriate growth rates in the parameters, both families capture for large N a rich set of white signals and the disturbance rejection measure turns out to be the 2-norm of the system. Therefore it seems natural, for practical purposes, to use the simplest possible description.

5 Multivariable extensions

We shall show briefly in this section (in the time domain only) how previous results extend to the case of vector valued signals, which are inputs to a MIMO system. Let us consider the time domain case with the definitions for autocorrelation matrices introduced in section 2. We get a new version of Lemma 1, which can be proved in a similar way (in this case, however, we cannot group terms into $R_H(\tau)$ due to non-commutativity of matrix product).

Lemma 3 *Let $\mathcal{H}$ be a MIMO (m inputs, p outputs) discrete time system. Let $u(t) \in \mathbb{R}^m, t = 0, \ldots N-1$ be a fragment of an N-periodic input signal to system $\mathcal{H}$, $y(t) \in \mathbb{R}^p, t = 0 \ldots N-1$ be the corresponding output. Then*

$$R_y(\tau) = \sum_{t=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} H(s) R_u(\tau + t - s) H'(t) \tag{15}$$

Now we turn to defining what we mean by vector-valued white noise. What changes is that now we must deal with spatial correlations (cross-correlations between the variables) as well as time correlations. For a signal x(t) to be white we want the matrix $R_x(0)$ to be close to a constant times the identity, and the matrices $R_x(\tau), \tau \neq 0$, to be small.

Definition 4 *A signal $x(t) \in \mathbb{R}^m, t = 0 \cdots N-1$ is said to be white with level γ up to time lag T if it satisfies*

$$\left\| \frac{1}{\|x\|^2} R_x(\tau) - \delta(\tau) \frac{1}{m} I \right\| \leq \gamma \tag{16}$$

Here $\delta(\tau)$ is the usual delta function, and the norm referred to in the definition can be any matrix norm. A choice which gives simple expressions below is the Frobenius norm, $\|A\|_F = (trace(A'A))^{\frac{1}{2}}$. We will denote the set of all such signals by $W_{N,\gamma,T}$.

Hard bounds similar to (9) on the difference $R_y(\tau) - R_H(\tau)$ may be given assuming $u \in W_{N,\gamma,T}$, though they are not as clean. We will only state the corresponding theorem for induced norms, which extends (10).

Theorem 4 *Under the conditions in Lemma 3, using $\|\cdot\|_F$ to define $W_{N,\gamma,T}$,*

$$\left| \|\mathcal{H}\|^2_{W_{N,\gamma,T}} - \frac{1}{m} \|\mathcal{H}\|_2^2 \right| \leq \gamma \sum_{\tau=-T}^{T} \|R_{H'}(\tau)\|_F + \sum_{|t|>T} \|R_{H'}(\tau)\|_F \tag{17}$$

Proof: Let $\|u\| = 1, u \in W_{N,\gamma,T}$. Starting from (15), standard manipulations yield

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

Using the fact that $|trace(AB)| \leq \|A\|_F \|B\|_F$ for real matrices, and that $\|R_u(\tau)\|_F \leq \|u\|^2$ for $\tau \neq 0$ we have

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

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

$\square$

We immediately obtain that under the same conditions as in Proposition 4, the asymptotic norm is $\frac{1}{\sqrt{m}} \|\mathcal{H}\|_2$.

6 Connections to behavioral descriptions of systems

In the behavioral approach to system theory as described in [3], a linear system is characterized by an equation $\mathcal{R}(q)w = 0$, where w is the vector of manifest variables which we wish to describe. The behavior $\mathcal{B}$ is the set of all signals $w(t)$ compatible with these equations, and we need not distinguish a priori which of these variables are inputs or outputs. For precise definitions and statements we refer to [3], a summary is given in [5].

Disturbance variables will appear in the vector w, and we could think of imposing whiteness constraints to some of them. In the behavioral framework, this amounts to intersecting the behavior $\mathcal{B}$ with sets such as those described in this paper. In this sense, these descriptions appear to be natural and fit more easily into the behavioral picture than stochastic models.

One could argue, however, that a white noise variable is always an input: we choose to model a complex phenomenon as a simplified part plus noise sources, which only exist in our model, and by construction they are inputs. So apparently not much is gained by adopting the behavioral setup as far as noise descriptions are concerned.

However, the behavioral point of view comes in when we consider the following question: is there a way to write our whiteness constraints on signals to make them compatible with standard robust control models, which are usually written in terms of linear fractional transformations (LFTs) on norm bounded Δ blocks? It turns out that this can be done *if* we adopt the behavioral setup for robust control, as proposed in [4].

In what follows we will show how we can write whiteness constraints in autocorrelations as *uncertain behavioral equations* , which can afterwards be interconnected to a behavioral description of our system (which might also include other sources of uncertainty) to give a unified description in which we can analyze stability and performance of the system in the presence of noise. We consider a simple case which captures the essence of the construction. Assume N, γ, T are fixed, and consider the autocorrelation constraints

$$|r_w(\tau)| \leq \gamma \; r_w(0), \qquad \tau = 1 \cdots T$$

which were introduced in Definition 1 to describe whiteness of a certain variable w. Focusing on, say, $\tau = 1$, we can write the equivalent pair of constraints

$$\pm \; r_w(1) \leq \gamma \; r_w(0) \tag{18}$$

Considering the plus sign, we can write the inequality as

$$\langle w, q^{-1}w \rangle \leq \gamma \, \|w\|^2$$
$$\iff 2(1-\gamma) \, \|w\|^2 \leq \left\| w - q^{-1}w \right\|^2$$

where all norms and the inner product $\langle , \rangle$ are over a period of length N. Choosing $\alpha = \sqrt{2(1-\gamma)}$, we can replace the previous constraint by the existence of an operator δ, $\|\delta\| \leq 1$, and a signal ξ such that

$$\begin{aligned} \xi &= \delta(w - q^{-1}w) \\ 0 &= \xi + \alpha w \end{aligned} \tag{19}$$

This in turn can be written as the uncertain behavioral equation $\mathcal{S}(\delta, M(q))w = 0$ where $\mathcal{S}$ denotes the LFT $M_{22} + M_{21}\delta(I - M_{11}\delta)^{-1}M_{12}$, depicted in the figure below, $M(q)$ being the transfer function in the box.

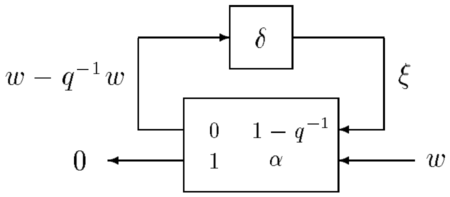

Figure 4: Uncertain behavioral equation for autocorrelation constraint

This equation can also be rewritten as $\mathcal{S}(diag[q^{-1}, \delta], M)w = 0$ which is the discrete time version of the form described in [4].

We can write a similar constraint for the minus sign in (18), and also for other values of τ (here we could also accommodate a different weighting of the autocorrelation). The final result is that the set $W_{N,\gamma,T}$ can be described as the set of N-periodic sequences such that $\mathcal{S}(\Delta, M(q))w = 0$, where Δ is a structured operator, $\|\Delta\| \leq 1$. By interconnecting this equation with a behavioral description of a system we have recast the white noise analysis in the standard LFT paradigm for robust control. There is no clear way to do this if we do not adopt the behavioral framework.

A final comment is that an alternative paradigm for robust control has been recently proposed by Megretsky [8], based on integral quadratic constraints (IQCs). It is easy to see (and is mentioned in [8]) that correlation constraints such as those describing $W_{N,\gamma,T}$ fit the IQC framework without further transformation. In fact, these more general IQCs can also be recast as uncertain behavioral equations.

7 Conclusion

This paper provides a framework in which to analyze spectral properties of disturbances and link them to rejection properties in a linear system. The connection is direct in the sense that information available in experimental data is directly tied to the system response, avoiding the intermediate step of introducing a probabilistic model.

These descriptions are attractive because they match up with existing methods for the analysis of systems containing other forms of uncertainty which are usually written in the worst case paradigm.

In this respect, the uncertain behavioral framework comes in as a global setup where all these descriptions can coexist. In particular, robust performance analysis of a linear system with some structured uncertainty and subject to the restriction that some disturbance signals should be white, can be recast as the question of finding non-trivial solutions to an uncertain behavioral equation of the type $\mathcal{S}(\Delta, M(q))w = 0$. This stimulates a direction of future research, also related to the work in [4], regarding the extension of robustness analysis and synthesis techniques to the behavioral framework. We are currently working on an extension of μ-analysis to this setup.

8 Appendix

We will outline here the proof of Propositions 3 and 5, which relate to the stochastic setup. In the sequel, $u(0)\cdots u(N-1)\cdots$ is a sequence of independent, identically distributed random variables, of 0 mean and finite variance σ^2. To analyze the behavior of $r_u(\tau), \tau > 0$, it is useful to consider the non-circular autocorrelation

$$\hat{r}_u(\tau) = \sum_{n=0}^{N-\tau-1} u(n)u(n+\tau).$$

Let $\Delta r_u(\tau) = r_u(\tau) - \hat{r}_u(\tau)$. We first write the complement of $W_{N,\gamma,T}$ as

$$W_{N,\gamma,T}^C = \bigcup_{\tau=1}^{T} A_{N,\gamma}^{\tau} \quad , \quad A_{N,\gamma}^{\tau} = \{u : |r_u(\tau)| > \gamma r_u(0)\} \tag{20}$$

For any constants $0 < \eta < \rho < 1$, we can write

$$\begin{aligned} A_{N,\gamma}^{\tau} &\subset \left\{u : \frac{|r_u(\tau)|}{\sigma^2\sqrt{N}} > \gamma\sqrt{N}\rho\right\} \cup \left\{u : \frac{r_u(0)}{\sigma^2 N} < \rho\right\} \subset \\ &\subset \left\{\frac{|\hat{r}_u(\tau)|}{\sigma^2\sqrt{N-\tau}} > \gamma\sqrt{N}\eta\right\} \cup \left\{\frac{|\Delta r_u(\tau)|}{\sigma^2 N} > \gamma(\rho-\eta)\right\} \cup \left\{\frac{r_u(0)}{\sigma^2 N} < \rho\right\} \end{aligned} \tag{21}$$

Proof of Proposition 3:

Here we assume T is fixed, and $\gamma\sqrt{N} \to \infty$.

For fixed $\tau > 0$, the random variables $z(n) = u(n)u(n+\tau), n \geq 0$ form a strictly stationary random process, with 0 mean and finite variance $\sigma_z^2 = \sigma^4$, which is τ-dependent (i.e., $(z(0)\cdots z(m))$ and $(z(m+k), z(m+k+1)\cdots)$ are independent for $k > \tau$).

Under these conditions of weak dependence we have a version of the central limit theorem (see for example, [9]). This says that

$$\frac{1}{\sigma_z\sqrt{N-\tau}} \sum_{n=0}^{N-\tau-1} z(n) = \frac{\hat{r}_u(\tau)}{\sigma^2\sqrt{N-\tau}}$$

converges in distribution to a normal $\mathcal{N}(0,1)$ law. Since $\gamma\sqrt{N}\eta \to \infty$, the probability of the first set in (21) converges to 0 as $N \to \infty$.

Since τ is fixed, the distribution of $\Delta r_u(\tau)$ is fixed and therefore the probability of the second set in (21) also vanishes asymptotically.

As for the third set, we use the fact that $\rho < 1$ and invoke the law of large numbers to get

$$\mathcal{P}\left(\frac{1}{\sigma^2 N}\sum_{n=0}^{N-1} u(n)^2 < \rho\right) \overset{N\to\infty}{\longrightarrow} 0$$

Adding these terms, $\mathcal{P}\left(A_{N,\gamma}^{\tau}\right) \overset{N\to\infty}{\longrightarrow} 0$, and from (20) we get $\mathcal{P}\left(u \notin W_{N,\gamma,T}\right) \overset{N\to\infty}{\longrightarrow} 0.$

$\square$

Proof of Proposition 5:

Here we assume $\gamma(N)\sqrt{\frac{N}{log(N)}} > \nu > 1$, and $T(N)\sqrt{\frac{log(N)}{N}} \stackrel{N\to\infty}{\longrightarrow} 0$ and that the variables $u(n)$ have finite 4th moment. This implies the variables $z(n) = u(n)u(n+\tau)$ also have this property for $\tau > 0$.

Since the number of sets $A^{\tau}_{N,\gamma}$ will now be a function of N, we will need a bound on $\mathcal{P}\left(A^{\tau}_{N,\gamma}\right)$. Going back to (21), let us choose $\eta = 1/\nu < 1$, and $\eta < \rho < 1$; we will bound separately the probability of each set.

For the first set, we will need a bound on the convergence rate to the normal distribution of $Z_{N-\tau} = \frac{\hat{r}_u(\tau)}{\sigma^2\sqrt{N-\tau}}$. Once more, under the τ-dependence conditions we have here, such bounds are known; we quote one from [10], which requires finite 3rd moment:

$$|\mathcal{P}(|Z_m| > \beta) - \mathcal{P}(|Z| > \beta)| \leq \frac{C_1}{\sqrt{m}} + C_2\tau\frac{log(m)}{m}$$

where Z is a $\mathcal{N}(0,1)$ random variable, (C_i will denote constants from now on). A bound for the tail of the normal distribution is

$$\mathcal{P}(|Z| > \beta) \leq \frac{C_3}{\beta}e^{-\frac{\beta^2}{2}}$$

Setting $\beta = \gamma\sqrt{N}\eta > \sqrt{log(N)}$, and $m = N - \tau$, we get after some manipulations, for sufficiently large N,

$$\mathcal{P}\left(\frac{|\hat{r}_u(\tau)|}{\sigma^2\sqrt{N-\tau}} > \gamma\sqrt{N}\eta\right) \leq \frac{C_4}{\sqrt{N}} + C_5\tau\frac{log(N)}{N}$$

For the second set in (21) we can use a Chebishev inequality (this uses the 4th moment assumption) to get

$$\mathcal{P}\left(\frac{|\Delta r_u(\tau)|}{\sigma^2 N} > \gamma(\rho-\eta)\right) \leq \frac{E(\Delta r_u(\tau))^2}{\sigma^4 N^2\gamma^2(\rho-\eta)^2} \leq \frac{C_6\tau}{(N\gamma)^2} \leq \frac{C_7\tau}{N}$$

For the third set, the finite 4th moment gives us a similar Chebishev bound $\mathcal{P}\left(\frac{r_u(0)}{\sigma^2 N} < \rho\right) \leq \frac{C_8}{N}$. Putting all the terms together, we have

$$\mathcal{P}\left(A^{\tau}_{N,\gamma}\right) \leq \frac{C_4}{\sqrt{N}} + C_5\tau\frac{log(N)}{N} + \frac{C_7\tau}{N} + \frac{C_8}{N}$$

Now we bound τ by T, and use (20) to get the bound

$$\mathcal{P}\left(u \notin W_{N,\gamma,T}\right) \leq \frac{C_4 T}{\sqrt{N}} + C_5 T^2\frac{log(N)}{N} + \frac{C_7 T^2}{N} + \frac{C_8 T}{N} \stackrel{N\to\infty}{\longrightarrow} 0$$

using the fact that $T(N)\sqrt{\frac{log(N)}{N}} \stackrel{N\to\infty}{\longrightarrow} 0$.

□

Acknowledgements

The author would like to thank John Doyle for motivation and helpful discussions at Caltech, where we also had the useful input of Munther Dahleh and Karl Astrom. I also wish to acknowledge P.P.Vaidyanathan and Gonzalo Perera for providing references, and Raffaello D'Andrea for help in the review of this paper. This work was supported by AFOSR, NSF and by the Universidad de la Republica, Uruguay.

References

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

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

[3] Willems, J.C., 1991, **Paradigms and Puzzles in the Theory of Dynamical Systems**, IEEE Transactions on Automatic Control, Vol. 36, pp. 259-294.

[4] D'Andrea, R. and Paganini, F., 1993, **Interconnections of Uncertain Behavioral Systems for Robust Control**, Proceedings 1993 CDC, San Antonio, Texas.

[5] D'Andrea, R., Paganini, F.,and Doyle,J.C., 1993, **Uncertain Behavior**, Proceedings 1993 CDC, San Antonio, Texas.

[6] Ljung, L., 1987, **System Identification Theory for the User**, Prentice-Hall, Englewood Cliffs, New Jersey.

[7] Vaidyanathan, P.P. **Robust Structures for Digital Signal Processing**, Prentice-Hall, in preparation.

[8] Megretsky, A., 1992, **S-Procedure in Optimal Non-Stochastic Filtering** Technical Report, Code Trita/Mat-92-0015, Royal Institute of Technology, Sweden.

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

[10] Tikhomirov, A.N., 1980, **On the Convergence Rate in the Central Limit Theorem for Weakly Dependent Random Variables**, Theory of Probability and its Applications, Vol 25-4, pp. 790-809.