Describing Functions

James H. Taylor
Department of Electrical & Computer Engineering
University of New Brunswick
PO Box 4400
Fredericton, NB CANADA E3B 5A3
email: jtaylor@unb.ca

Article prepared for:
Electrical Engineering Encyclopedia,
John Wiley & Sons, Inc., New York, 1999

08 April 1999

Revised: 10 October 1999

Describing Functions

James H. Taylor

In this article we define and overview the basic concept of a describing function, and then we look at a wide variety of usages and applications of this approach for the analysis and design of nonlinear dynamical systems. More specifically, the following is an outline of the contents:

1. concept and definition of describing functions for sinusoidal inputs ("sinusoidal-input describing functions") and for random inputs ("random-input describing functions")
2. traditional application of sinusoidal-input describing functions for limit cycle analysis, for systems with one nonlinearity
3. application of sinusoidal-input describing function techniques for determining the frequency response of a nonlinear system
4. application of random-input describing functions for the performance analysis of nonlinear stochastic systems
5. application of sinusoidal-input describing functions for limit cycle analysis, for systems with multiple nonlinearities
6. application of sinusoidal-input describing function methods for the design of nonlinear controllers for nonlinear plants
7. application of sinusoidal-input describing functions for the implementation of linear self-tuning controllers for linear plants and of nonlinear self-tuning controllers for nonlinear plants

1 Basic Concepts and Definitions

Describing function theory and techniques represent a powerful mathematical approach for understanding (analyzing) and improving (designing) the behavior of nonlinear systems. In order to present describing functions, certain mathematical formalisms must be taken for granted, most especially differential equations and concepts such as step response and sinusoidal input response. In addition to a basic grasp of differential equations as a way to describe the behavior of a system (circuit, electric drive, robot, aircraft, chemical reactor, ecosystem, etc.) certain additional mathematical concepts

are essential for the useful application of describing functions – Laplace transforms, Fourier expansions and the frequency domain being foremost on the list. This level of mathematical background is usually achieved at about the third or fourth year in undergraduate engineering.

The main motivation for describing function (DF) techniques is the need to understand the behavior of nonlinear systems, which in turn is based on the simple fact that every system is nonlinear except in very limited operating regimes. Nonlinear effects can be beneficial (many desirable behaviors can only be achieved by a nonlinear system, e.g., the generation of useful periodic signals or oscillations), or they can be detrimental (e.g., loss of control and accident at a nuclear reactor). Unfortunately, the mathematics required to understand nonlinear behavior is considerably more advanced than that needed for the linear case.

The elegant mathematical theory for linear systems provides a unified framework for understanding all possible linear system behaviors. Such results do not exist for nonlinear systems. In contrast, different types of behavior generally require different mathematical tools, some of which are exact, some approximate. As a generality, exact methods may be available for relatively simple systems (ones that are of low order, or that have just one nonlinearity, or where the nonlinearity is described by simple relations), while more complicated systems may only be amenable to approximate methods. Describing function approaches fit in the latter category: approximate methods for complicated systems.

One way to deal with a nonlinear system is to linearize it. The standard approach, often called small-signal linearization, involves taking the derivative or slope of each nonlinear term and using that slope as the gain of a linear term. As a simple example, an important cause of excessive fuel consumption at higher speeds is drag, which is often modeled with a term $B\,v^2 \cdot \mathrm{sgn}(v)$ (a constant times the square of the velocity times the sign of v). One may choose a nominal velocity, e.g., $v_0 = 100$ km/h, and approximate the incremental effect of drag with the linear term $2\,B\,v_0 \cdot (v - v_0) \triangleq 2\,B\,v_0\,\delta v$. For velocity in the vicinity of 100 km/h, perhaps $|\delta v| \leq 5$ km/h, this may be reasonably accurate, but for larger variations it becomes a poor model – hence the terminology "small-signal linearization".

The strong attraction of small-signal linearization is that the elegant theory for linear systems may be brought to bear on the resulting linear model. However, this approach can only explain the effects of small variations about the linearization point, and, perhaps more importantly, it can only reveal linear system behavior. This approach is thus ill-suited for understanding phenomena such as nonlinear oscillation or for studying the limiting or detrimental effects of nonlinearity.

The basic idea of the describing function (DF) approach for modeling and studying nonlinear system behavior is to replace each nonlinear element with a (quasi)linear descriptor or describing function whose gain is a function of input amplitude. The functional form of such a descriptor is governed by several factors: the type of input signal, which is assumed in advance, and the approximation criterion, e.g., minimization of mean squared error. This technique is dealt with very thoroughly in a number of texts for the case of nonlinear systems with a single nonlinearity (Gelb and Vander Velde, 1968 [8]; Atherton, 1975 [3]); for systems with multiple nonlinearities in arbitrary configurations, the most general extensions may be attributed independently to Kazakov [14] and Gelb & Warren [9] in the case of random-input DFs (RIDFs) and jointly to Taylor [22, 24] and Hannebrink et al [12] for sinusoidal-input DFs (SIDFs). Developments for multiple nonlinearities have been presented in tutorial form in Ramnath, Hedrick and Paynter [21].

Two categories of DFs have been particularly successful: sinusoidal-input describing functions and random-input describing functions, depending, as indicated, upon the class of input signals under consideration. The primary texts cited above [3, 8] and some other sources make a more detailed classification, e.g., SIDF for pure sinusoidal inputs, sine-plus-bias DFs if there is a nonzero DC value, RIDF for pure random inputs, random-plus-bias DFs; however, this seems unnecessary since sine-plus-bias and random-plus-bias can be treated directly in a unified way, so we will use the terms SIDF and RIDF accordingly. Other types of DFs also have been developed and used in studying more complicated phenomena, e.g., two-sinusoidal-input DFs may be used to study effects of limit cycle quenching via the injection of a sinusoidal "dither" signal, but those developments are beyond the scope of this article.

The SIDF approach generally can be used to study periodic phenomena. It is applied for two primary purposes: limit cycle analysis (Sects. 2, 5) and characterizing the input/output behavior of a nonlinear plant in the frequency domain (Sect. 3). This latter application serves as the basis for a variety of control system analysis and design methods, as outlined in Sect. 6. RIDF methods, on the other hand, are used for stochastic nonlinear systems analysis and design (analysis and design of systems with random signals), in analogous ways with the corresponding SIDF approaches, although SIDFs may be said to be more general and versatile, as we shall see.

There is one additional theme underlying all the developments and examples in this article: Describing function approaches allow one to solve a wide variety of problems in nonlinear system analysis and design via the use of direct and simple extensions of linear systems analysis "machinery". In point of fact, the mathematical basis is generally different (not based on linear systems theory); however, the application often results in conditions of the same form which are easily solved. Finally, we note that

the types of nonlinearity that can be studied via the DF approach are very general – nonlinearities that are discontinuous and even multivalued can be considered. The order of the system is also not a serious limitation. Given software such as MATLAB[1] for solving problems that are couched in terms of linear system mathematics, e.g., plotting the polar or Nyquist plot of a linear system transfer function, one can easily apply DF techniques to high-order nonlinear systems. The real power of this technique lies in these factors.

1.1 Introduction to describing functions for sinusoids

The fundamental ideas and use of the SIDF approach can best be introduced by overviewing the most common application, limit cycle analysis for a system with a single nonlinearity. A limit cycle (LC) is a periodic signal, $x_{LC}(t+T) = x_{LC}(t)$ for all t and for some T (the period), such that perturbed solutions either approach x_{LC} (a stable limit cycle) or diverge from it (an unstable one).

The study of LC conditions in nonlinear systems is a problem of considerable interest in engineering. An approach to LC analysis that has gained widespread acceptance is the frequency-domain / SIDF method. This technique, as it was first developed for systems with a single nonlinearity, involved formulating the system in the form shown in Fig. 1, where $G(s)$ is defined in terms of a ratio of polynomials, as follows:

$$\begin{aligned} Y(s) \triangleq \mathcal{L}\left(y(t)\right) &= \frac{p(s)}{q(s)}\, E(s) \\ &\triangleq G(s)\ \mathcal{L}\left(e(t)\right) \\ e(t) &= r(t) - f(y(t)) \end{aligned} \tag{1}$$

where $\mathcal{L}(\cdot)$ denotes the Laplace transform of a variable and $p(s)$, $q(s)$ represent polynomials in the Laplace complex variable s, with $\text{order}(p) < \text{order}(q) \triangleq n$. An alternative formulation of the same system is the state-space description,

$$\begin{aligned} \dot{x} &= A\,x + b\,e(t) \\ y(t) &= c^T\,x \\ e(t) &= r(t) - f(y(t)) \end{aligned} \tag{2}$$

where x is an n-dimensional state vector. In either case, the first two relations describe a linear dynamic subsystem with input e and output y; the subsystem input is then given to be the external input signal $r(t)$ minus a nonlinear function of y. There is thus one single-input/single-output (SISO) nonlinearity, $f(y)$, and linear dynamics of arbitrary order. Note that the state-space description is equivalent to the SISO transfer

[1] MATLAB is a registered trademark of The MathWorks, Inc., Natick, Massachusetts, USA.

function by identifying $G(s) = c^T(sI - A)^{-1}b$. Thus either system description is a formulation of the conventional "linear plant in the forward path with a nonlinearity in the feedback path" depicted in Fig. 1. The single nonlinearity might be an actuator or sensor characteristic, or a plant nonlinearity – in any case, the following LC analysis can be performed using this configuration.

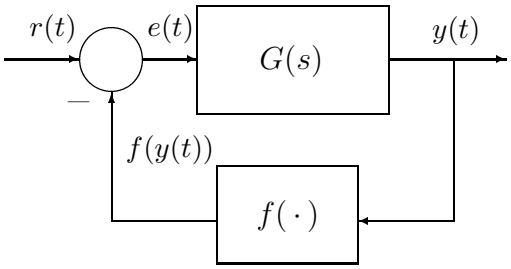

Figure 1: System Configuration with One Dominant Nonlinearity

In order to investigate LC conditions with no excitation, $r(t) = 0$, the nonlinearity is treated as follows: First, we assume that the input y is essentially sinusoidal, i.e., that a periodic input signal may exist, $y \cong a\cos(\omega t)$, and thus the output is also periodic. Expanding in a Fourier series, we have

$$f(a\cos(\omega t)) = \sum_{k=1}^{\infty} \mathrm{Re}\left\{b_k(a)\, e^{j\omega t}\right\} \tag{3}$$

By omitting the constant or DC term from Eqn. 3 we are implicitly assuming that $f(y)$ is an odd function, $f(-y) = -f(y)$ for all y, so that no rectification occurs; cases when $f(y)$ is not odd present no difficulty, but are omitted to simplify this introductory discussion. Then we make the approximation

$$\begin{aligned} f(a\cos(\omega t)) &\cong \mathrm{Re}\left\{b_1(a)\cdot e^{j\omega t}\right\} \\ &\triangleq \mathrm{Re}\left\{N_s(a)\cdot a\, e^{j\omega t}\right\} \end{aligned} \tag{4}$$

This approximate representation for $f(a\cos(\omega t))$ includes only the first term of the Fourier expansion of Eqn. 3; therefore the approximation error (difference between $f(a\cos(\omega t))$ and $\mathrm{Re}\left\{N_s(a)\cdot a\, e^{j\omega t}\right\}$) is minimized in the mean squared error sense [15]. The Fourier coefficient b_1 (and thus the gain $N_s(a)$) is generally complex unless $f(y)$ is single-valued; the real and imaginary parts of b_1 represent the in-phase (cosine) and quadrature (–sine) fundamental components of $f(a\cos(\omega t))$, respectively. The so-called describing function $N_s(a)$ in Eqn. 4 is, as noted, amplitude dependent, thus retaining a basic property of a nonlinear operation.

By the principle of harmonic balance, the assumed oscillation – if it is to exist – must result in a loop gain of unity (including the summing junction), i.e., substituting $f(y) \cong N_s(a)\, y$ in Eqn. 1 yields the requirement $N_s(a)\, G(j\omega) = -1$, or

$$G(j\omega) = -1/N_s(a) \tag{5}$$

For the state-space form of the model, using $X(j\omega)$ to denote the Fourier transform of x, $X(j\omega) = \mathcal{F}(x)$, and thus $j\omega X = \mathcal{F}(\dot{x})$, and again substituting $f(y) \cong N_s(a)\, y$ in Eqn. 2 yields the requirement

$$\left| \, j\omega I - A + N_s(a)\, b\, c^T \, \right| \, X(j\omega) = 0 \tag{6}$$

for some value of ω and $X(j\omega) \neq 0$. This is exactly equivalent to the condition in Eqn. 5.

The condition in Eqn. 5 is easy to verify using the polar or Nyquist plot of $G(j\omega)$; in addition the LC amplitude a_{LC} and frequency ω_{LC} are determined in the process. More of the details of the solution for LC conditions are exposed in Sect. 2. Note that the state-space condition, Eqn. 6, appears to represent a quasilinearized system with pure imaginary eigenvalues; this is merely the first example showing how the describing function approach gives rise to conditions that seem to involve linear systems-theoretic concepts.

It is generally well understood that the classical SIDF analysis as outlined above is only approximate, so caution is always recommended in its use. The standard caveat that $G(j\omega)$ should be "low pass to attenuate higher harmonics" (so that the first harmonic in Eqn. 3 is dominant) indicates that the analyst has to be cautious. Nonetheless, as demonstrated by a more detailed example in Sect. 2, this approach is simple to apply, very informative, and in general quite accurate. The main circumstance in which SIDF limit cycle analysis may yield poor results is in a borderline case, i.e., one where the DF just barely cuts the Nyquist plot, or just barely misses it.

The next step in this brief introductory exposition of the SIDF approach involves showing a few elementary SIDF derivations for representative nonlinearity types. The basis for these evaluations is the well-known fact that a truncated Fourier series expansion of a periodic signal achieves minimum mean squared approximation error [15], so we define the DF $N_s(a)$ as the first Fourier coefficient divided by the input amplitude (we divide by a so that $N_s(a)$ is in the form of a quasilinear gain):

1. **Ideal relay:** $f(y) = D \cdot \mathrm{sgn}(y)$ where we assume no DC level, $y(t) = a\cos(\omega t)$ – we set up and evaluate the integral for the first Fourier coefficient divided by a as follows:

$$N_s(a) \;\; = \;\; \frac{1}{\pi\, a} \int_0^{2\pi} f(a\, \cos(x)) \cdot \cos(x)\, dx$$

$$
\begin{aligned}
&= \frac{2\,D}{\pi\,a}\int_0^{\pi}\cos(x)\,dx \quad \text{(by symmetry)} \\
&= \frac{4\,D}{\pi\,a} \qquad (7)
\end{aligned}
$$

2. **Cubic nonlinearity:** $f(y) = K\,y^3(t)$; again, assuming $y(t) = a\cos(\omega t)$, we can directly write the Fourier expansion using the trigonometric identity for $\cos^3(x)$ as follows:

$$
\begin{aligned}
f(a\cos(\omega t)) &= K[a\cos(\omega t)]^3 \\
&= K\,a^3\,[\frac{3}{4}\cos(\omega t) + \frac{1}{4}\,\cos(3\omega t)] \\
&\simeq \frac{3\,K\,a^2}{4}\cdot a\cos(\omega t) \qquad (8)
\end{aligned}
$$

   so the SIDF is $N_s(a) = 3\,K\,a^2/4$. Note that this derivation uses trigonometric identities as a shortcut to formulating and evaluating Fourier integrals; this approach can be used for any power-law element.

The plots of $N_s(a)$ versus a for these two examples are provided in Fig. 2. Note the sound intuitive logic of these relations: a relay acts as a very high gain for small input amplitude but a low gain for large inputs, while just the opposite is true for the function $f(y) = K\,y^3(t)$. One more example is provided in Sect. 2, to show that a multivalued nonlinearity (relay with hysteresis) in fact leads to a complex-valued SIDF. Other examples and an outline of useful SIDF properties are provided in a companion article [5]. Observe that the SIDFs for many nonlinearities can be looked up in [3, 8] (SIDFs for 33 and 51 cases are provided, respectively).

Finally, we demonstrate that condition 5 is easy to verify, using standard linear system analytic machinery and software:

Example 1 – The developments so far provide the basis for a simple example of the traditional application of SIDF methods to determine LC conditions for a system with one dominant nonlinearity, defined in Eqn. 1. Assume that the plant depicted in Fig. 1 is modeled by

$$
G(s) = \frac{100\,(s+1)\,(s+70)}{(s+0.4)\,(s+2)\,(s+5+7j)\,(s+5-7j)\,(s+20)} \qquad (9)
$$

This transfer function might represent a servo amplifier and DC motor driving a mechanical load with friction level and spring constant adjusted to give rise to the lightly damped complex conjugate poles indicated, and the question is: will this cause limit cycling?

The Nyquist plot for this fifth-order linear plant is portrayed in Fig. 3, and the upper limit for stability is $K_{max} = 2.07$; in other words, a linear gain k in the range [0, 2.07)

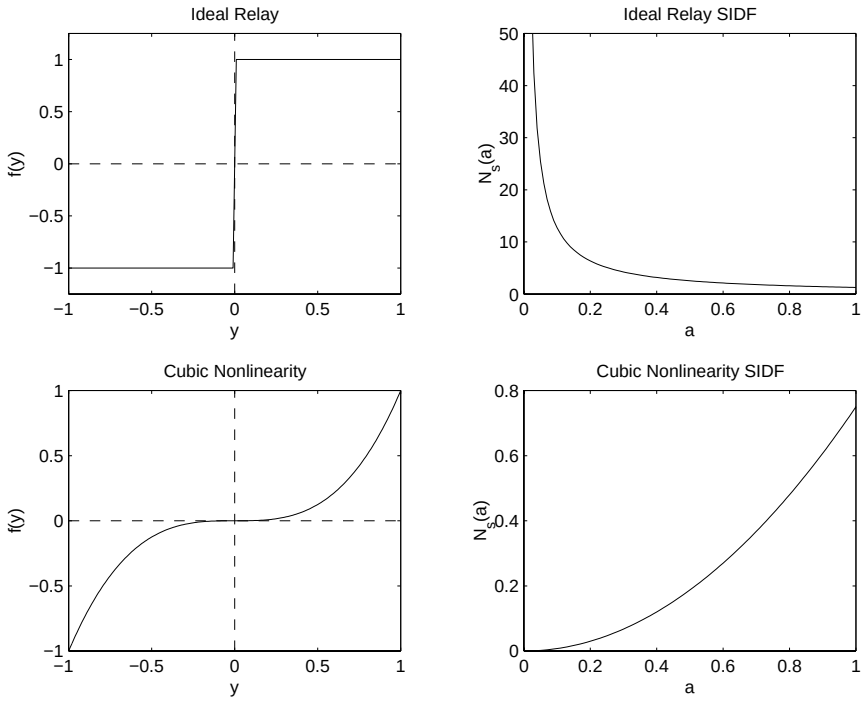

Figure 2: Illustration of Elementary SIDF Evaluations

will stabilize the closed-loop system if $f(y)$ is replaced by $k\,y$. According to Eqn. 5 LCs are predicted if there is a nonlinearity $f(y)$ in the feedback path such that $-1/N_s(a)$ cuts the Nyquist curve, or in this case if $N_s(a)$ takes on the value 2.07. For the two nonlinearities considered so far, the ideal relay and the cubic characteristic, the SIDFs lie in the range $[\ 0, \infty\ )$, so LCs are indeed predicted in both cases. Furthermore, one can immediately determine the corresponding amplitudes of the LCs, namely setting $N_s(a) = 2.07$ in Eqns. 7,8 yields an amplitude of $a_{LC} = 4\,D/(2.07\,\pi)$ for the relay and $a_{LC} = \sqrt{8.28/(3\,K)}$ for the cubic. In both cases $-1/N_s(a)$ cuts the Nyquist curve at the standard cross-over point on the real axis, so the LC frequency is $\omega_{LC} = \omega_{CO} = 8.11$ rad/sec.

The LCs predicted by the SIDF approach in these two cases are fundamentally different, however, in one important respect: *stability* of the nonlinear oscillation. An LC is said to be stable if small perturbations from the periodic solution die out, i.e., the waveform returns to the same periodic solution. While the general analysis for determining the stability of a predicted LC is complicated and beyond the scope of this article (see [3, 8]), the present example is quite simple: If points where $N_s(a)$ slightly exceeds K_{max} correspond to $a > a_{LC}$ then the LC is unstable, and conversely. Therefore, the LC produced by the ideal relay would be stable, while that produced by the cubic characteristic is unstable, as can be seen referring to Fig. 2. Note again that this argument appears to be based on linear systems theory, but in fact the significance is

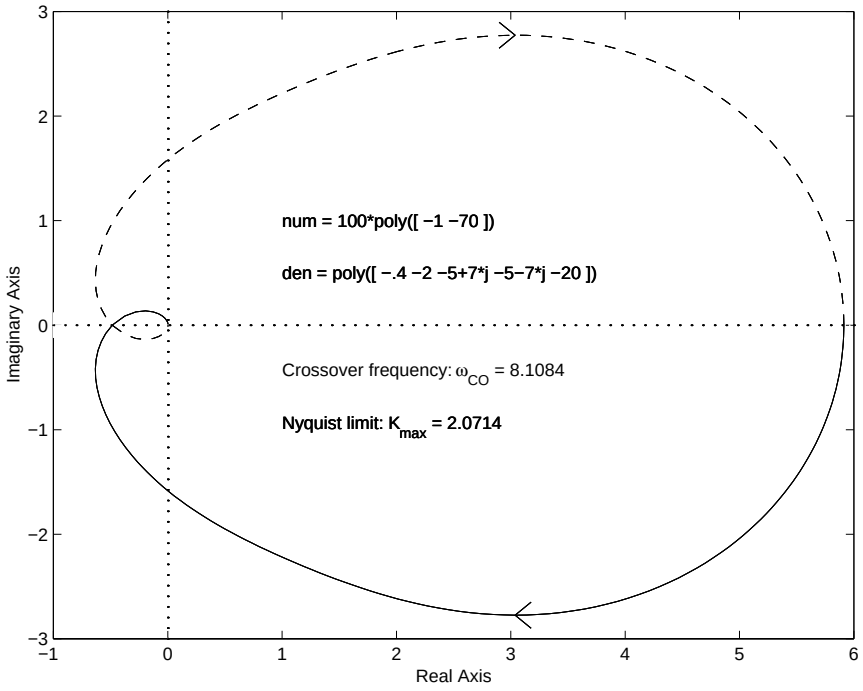

Figure 3: Nyquist Plot for Plant in Example 1

that the loop gain for a periodic signal should be less than unity for $a > a_{LC}$ if a is not to grow.

To summarize this analysis, we first noted the simple approximate condition that allows a periodic signal to perpetuate itself, Eqn. 5, i.e., the loop gain should be unity for the fundamental component. We then illustrated the basis for and calculation of SIDFs for two simple nonlinearities. These elements came together using the well-known Nyquist plot of $G(j\omega)$ to check if the assumption of a periodic solution is warranted and, if so, what the LC amplitude would be. We also briefly investigated the stability question, i.e., for which nonlinearity the predicted LC would be stable.

1.2 Formal definition of describing functions for sinusoidal inputs

The preceding outline of SIDF analysis of LC conditions illustrates the factors mentioned previously, namely the dependence of DFs upon the type of input signal and the approximation criterion. To express the standard definition of the SIDF more completely and formally,

- The nonlinearity under consideration is $f(y(t))$, and is quite unrestricted in form; e.g., $f(y)$ may be discontinuous and/or multivalued;
- The class of input signals is $y(t) = y_0 + a\cos(\omega t)$ and the input amplitudes are

quantified by the parameters (values) y_0, a;

- The SIDFs are denoted $N_s(y_0, a)$ for the sinusoidal component and $F_0(y_0, a)$ for the constant or DC part, i.e., the nonlinearity output is approximated by

$$f(y(t)) \cong F_0(y_0, a) + \mathrm{Re}\,\{N_s(y_0, a) \cdot a\, e^{j\omega t}\} \tag{10}$$

- The approximation criterion in Eqn. 10 is the minimization of mean squared error,

$$E_{ms} \triangleq \int_0^{2\pi/\omega} [f(y_0 + a\cos(\omega t)) - F_0(y_0, a) - \mathrm{Re}\,\{N_s(y_0, a) \cdot a\, e^{j\omega t}]^2 dt \tag{11}$$

Again, under the above conditions it can be shown that $F_0(y_0, a)$ and $a\,N_s(y_0, a)$ are the constant (DC) and first harmonic coefficients of a Fourier expansion of $f(y_0 + a\cos(\omega t))$ [15]. Note that this approximation of a nonlinear characteristic actually retains two important properties: amplitude dependence and the coupling between the DC and first harmonic terms. The latter property is a result of the fact that superposition does not hold for nonlinear systems, so, for example, N_s depends on both y_0 and a.

1.3 Describing functions for general classes of inputs

An elegant unified approach to describing function derivation is given in Gelb & Vander Velde [8], using the concept of *amplitude density functions* to put all DF formulae on one footing. Here we will not dwell on the theoretical background and derivations, but just present the basic ideas and results.

Assume that the input to a nonlinearity comprises a bias component b and a zero mean component $z(t)$, in other words, $y(t) = b + z(t)$. In terms of random variable theory, the expectation of z, denoted $E(z)$, is zero and thus $E(y) = b$. The nonlinearity input y may be characterized by its amplitude density function, $p(\alpha)$, defined in terms of the amplitude distribution function $P(\alpha)$ as follows:

$$p(\alpha) = \frac{d\,P(\alpha)}{d\,\alpha}\,; \;\; P(\alpha) = \mathrm{Prob}\,[\,y(t) < \alpha\,] \tag{12}$$

A well-known density function corresponds to the Gaussian or normal distribution,

$$p_n(\alpha) = \frac{1}{\sqrt{2\,\pi}\sigma_n} \exp\left(-\frac{(\alpha - b)^2}{2\,\sigma_n^2}\right) \tag{13}$$

Other simple amplitude density functions are called *uniform*, where the amplitude of y is assumed to lie between $b - A$ and $b + A$ with equal likelihood $1/2A$,

$$p_u(\alpha) = \begin{cases} 0\,, & \alpha < b - A \\ \frac{1}{2\,A}\,, & (b - A) \le \alpha \le (b + A) \\ 0\,, & \alpha > b + A \end{cases} \tag{14}$$

and *triangular*,

$$p_t(\alpha) = \begin{cases} 0\,, & \alpha < b - A \\ \frac{1}{A}\left(1 - \frac{|\alpha - b|}{A}\right), & (b - A) \le \alpha \le (b + A) \\ 0\,, & \alpha > b + A \end{cases} \tag{15}$$

Note that these three density functions are formulated such that the expected value of the variable is b in each case, and the area under the curve is unity ($\text{Prob}\,[\,y(t) < \infty\,] = \int_{-\infty}^{\infty} p(\alpha)\,d\alpha = 1$). From the standpoint of random variable theory the next most important expectation is $\Sigma = E((y(t) - b)^2)$, or the *variance* or mean squared value. In the normal case this is simply $\Sigma_n = \sigma_n^2$, where σ_n thus represents the *standard deviation*; for the other amplitude density functions we have $\Sigma_u = A^2/3$ and $\Sigma_t = A^2/6$. To express the corresponding DFs in terms of amplitude, the most commonly accepted measure is the standard deviation, $\sigma_u = A/\sqrt{3}$ and $\sigma_t = A/\sqrt{6}$.

Now, in order to unify the derivation of DFs for sinusoids as well as the types of variables mentioned above we need to express the amplitude density function of such signals. A direct derivation, given $y(t) = b + a\cos(\omega t)$, yields

$$p_s(\alpha) = \begin{cases} 0\,, & \alpha < b - a \\ \frac{1}{\pi\sqrt{a^2 - (\alpha - b)^2}}\,, & (b - a) \le \alpha \le (b + a) \\ 0\,, & \alpha > b + a \end{cases} \tag{16}$$

Again, this density function is written so that $E(y) = b$; it is well known that the root mean squared (RMS) value of $a\cos(\omega t)$ is $\sigma_s = a/\sqrt{2}$.

The above notation and terminology has been introduced primarily so that random-input describing functions (RIDFs) can be defined. We have, however, put signals with sinusoidal components into the same framework, so that one definition fits all cases. This leads to the following relations:

$$F_0(\sigma,\, b) \;=\; \int_{-\infty}^{\infty} f(b + z)\, p(z)\, dz \tag{17}$$

$$N_z(\sigma,\, b) \;=\; \frac{1}{\Sigma}\int_{-\infty}^{\infty} z\, f(b + z)\, p(z)\, dz \tag{18}$$

These relations again provide a minimum mean squared approximation error. For separable processes [3], this amounts to assuming that the amplitude density function of the nonlinearity output is of the same class as the input, e.g., the RIDF for the normal case provides the gain that fits a normal amplitude density function to the actual amplitude density function of the output with minimum mean squared error. There is only one restriction compared with the Fourier-series method for deriving SIDFs: multivalued nonlinearities (such as a relay with hysteresis) cannot be treated using the amplitude density function approach. The case of evaluating SIDFs for multivalued nonlinearities is illustrated in Example 2 in Sect. 2; it is evident that the same approach will not work for signals defined only in terms of amplitude density functions.

1.4 Describing functions for normal random inputs

The material presented in Sect. 1.3 provides all the machinery needed for defining the usual class of RIDFs, namely those for Gaussian, or normally distributed, random variables:

$$F_0(\sigma_n,\, b) \;=\; \frac{1}{\sqrt{2\,\pi}\sigma_n}\int_{-\infty}^{\infty} f(b+z)\exp\left(-\frac{(z-b)^2}{2\,\sigma_n^2}\right)dz \tag{19}$$

$$N_n(\sigma_n,\, b) \;=\; \frac{1}{\sqrt{2\,\pi}\sigma_n^3}\int_{-\infty}^{\infty} z\, f(b+z)\,\exp\left(-\frac{(z-b)^2}{2\,\sigma_n^2}\right)dz \tag{20}$$

Considering the same nonlinearities discussed in Sect. 1.1, the following results can be derived:

1. **Ideal relay:** $f(y) = D\cdot\mathrm{sgn}(y)$ where we assume no bias ($b = 0$), $y(t) = z(t)$ a normal random variable – we set up and evaluate the expectation in Eqn. 20 as follows:

$$\begin{aligned} N_n(\sigma_n) &= \frac{1}{\sqrt{2\,\pi}\sigma_n^3}\int_{-\infty}^{\infty} z\, f(b+z)\,\exp\left(-\frac{z^2}{2\,\sigma_n^2}\right)dz \\ &= \frac{2\,D}{\sqrt{2\,\pi}\sigma_n^3}\int_{0}^{\infty} z\,\exp\left(-\frac{z^2}{2\,\sigma_n^2}\right)dz \quad \text{(by symmetry)} \\ &= \frac{\sqrt{2}\,D}{\sqrt{\pi}\sigma_n} \end{aligned} \tag{21}$$

2. **Cubic nonlinearity:** $f(y) = K\,y^3(t)$; again, assuming no bias, the random-component DF is:

$$\begin{aligned} N_n(\sigma_n) &= \frac{1}{\sqrt{2\,\pi}\sigma_n^3}\int_{-\infty}^{\infty} z\, f(b+z)\,\exp\left(-\frac{z^2}{2\,\sigma_n^2}\right)dz \\ &= \frac{2\,K}{\sqrt{2\,\pi}\sigma_n^3}\int_{0}^{\infty} z^4\,\exp\left(-\frac{z^2}{2\,\sigma_n^2}\right)dz \quad \text{(by symmetry)} \\ &= 3\,K\,\sigma_n^2 \end{aligned} \tag{22}$$

1.5 Comparison of describing functions for different input classes

Given the unified framework in Sect. 1.3, it is natural to ask: how much influence does the assumed amplitude density function have on the corresponding DF? To provide some insight, we may investigate the specific DF versus amplitude plots for the four density functions presented above and for the unity limiter or saturation element,

$$f(y) = \begin{cases} y\,, & |\,y\,| \,<\, 1 \\ \mathrm{sgn}(y)\,, & |\,y\,| \,\geq\, 1 \end{cases} \tag{23}$$

and the cubic nonlinearity. These results are obtained by numerical integration in MATLAB and shown in Fig. 4. From an engineering point of view, the effect of varying the assumed input amplitude density function is not dramatic. For the limiter, the spread, (SIDF – RIDF)/2, is less than 10 % of the average, (SIDF + RIDF)/2, which provides good agreement. For the cubic case, the ratio of the RIDF to the SIDF is more substantial, namely a factor of two (taking into account that $a^2 = 2\,\sigma_s^2$ in Eqn. 8).

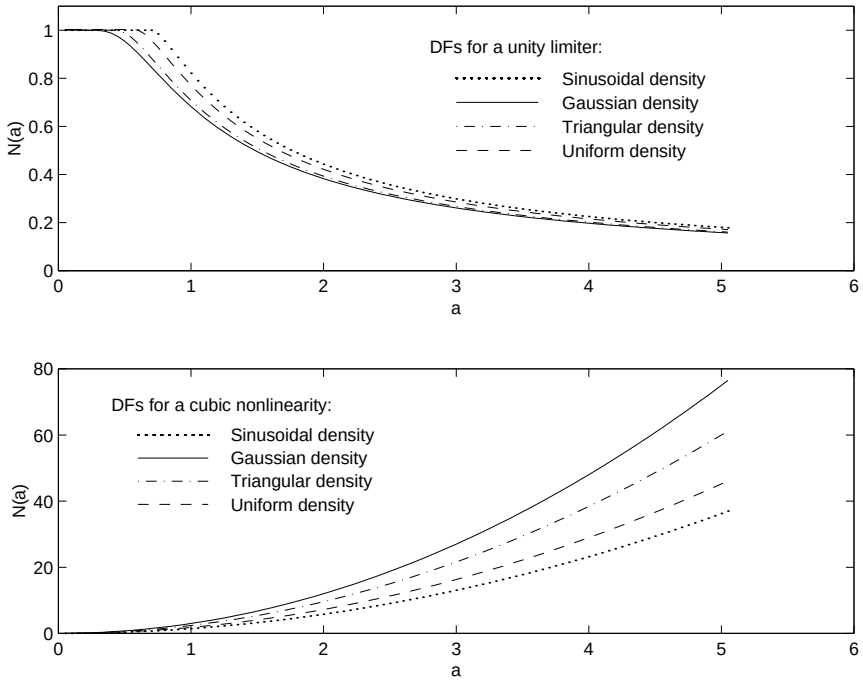

Figure 4: Influence of Amplitude Density Function on DF Evaluation

2 Traditional Limit Cycle Analysis Methods, One Nonlinearity

Much of the process, terminology and derivation for the traditional approach to limit cycle analysis has been presented in Sect. 1.1. We proceed to investigate a more realistic (physically motivated) and complex example.

Example 2 – A more meaningful example – and one that illustrates the use of complex-valued SIDFs to characterize multivalued nonlinearities – is provided by a missile roll control problem from Gibson [10]:

Assume a pair of reaction jets is mounted on the missile, one to produce torque about the roll axis in the clockwise sense and one in the counterclockwise sense. The force

exerted by each jet is $F_0 = 445$ N, and the moment arms are $R_0 = 0.61$ m. The moment of inertia about the roll axis is $J = 4.68$ N·m/sec^2. Let the control jets and associated servo actuator have a hysteresis $h = 22.24$ N and two lags corresponding to time constants of 0.01 sec and 0.05 sec. To control the roll motion, there is roll and roll-rate feedback, with gains of $K_p = 1868$ N/rad and $K_v = 186.8$ N/(rad/sec) respectively. The block diagram for this system is shown in Fig. 5.

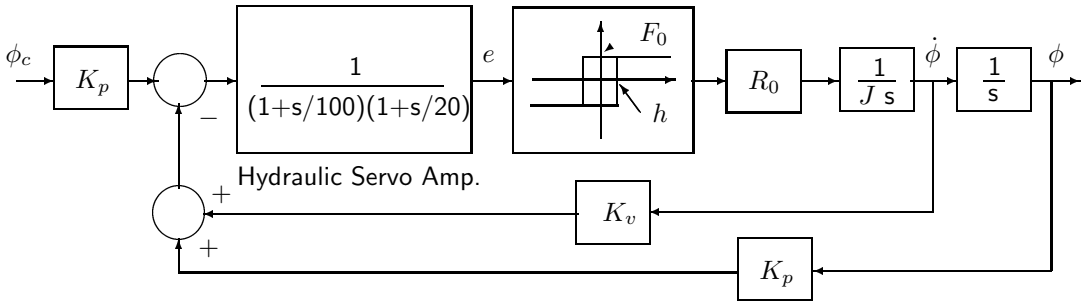

Figure 5: Block Diagram, Missile Roll-Control Problem [Gibson]

Before we can proceed with solving for the LC conditions for this problem, it is necessary to turn our attention to the derivation of complex-valued SIDFs for multivalued characteristics (relay with hysteresis). As in Sect. 1.1 we can evaluate this SIDF quite directly:

Relay with hysteresis: defining the output levels to be $\pm D$ and the hysteresis to be h ($D = F_0$ in Fig. 5), then if we assume no DC level, $y(t) = a\cos(\omega t)$, we can set up and evaluate the integral for the first Fourier coefficient divided by a as follows:

$$\begin{aligned} N_s(a) &= \frac{1}{\pi\, a}\int_0^{2\pi} f(a\,\cos(x))\cdot\exp(-jx)dx \\ &= \frac{2\,D}{\pi\, a}\left\{\int_0^{x_1}\exp(-jx)dx - \int_{x_1}^{\pi}\exp(-jx)dx\right\} \quad \text{(by symmetry)} \\ &= \frac{4\,D}{\pi\, a}\left\{\sqrt{1-(h/a)^2} - j\,h/a\right\} \end{aligned} \tag{24}$$

where the switching point x_1 is $x_1 = \cos^{-1}(-h/a)$. Note that strictly speaking $N_s(a) \equiv 0$ if $a \le h$, because the relay will not switch under that condition; the output will remain at D or $-D$ for all time, so the assumption that the nonlinearity output is periodic is invalid. The real and imaginary parts of this SIDF are shown in Fig. 6.

Given the SIDF for a relay with hysteresis, the solution to the problem of determining LC conditions for the system portrayed in Fig. 5 is depicted in Fig. 7. For a single-valued nonlinearity, and hence a real-valued SIDF, we would be interested in the real-axis crossing of $G(j\omega)$, at $\omega_{CO} = 28.3$ rad/sec, $G_{CO} =$ - 0.5073. In this case, however,

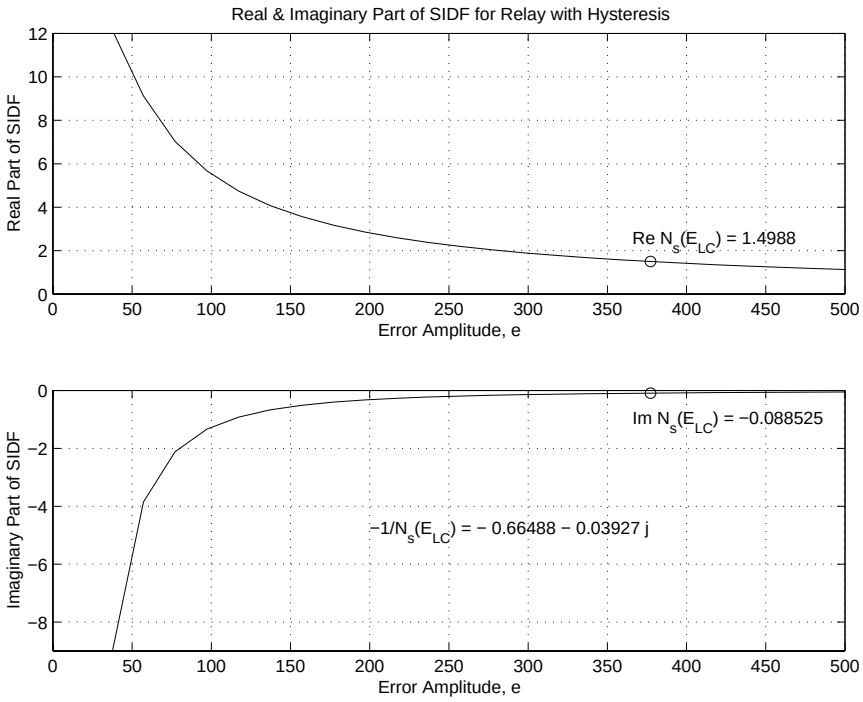

Figure 6: Complex-valued SIDF for Relay with Hysteresis

the intersection of $-1/N_s$ with $G(j\omega)$ no longer lies on the negative real axis, and so $\omega_{LC} = 24.36$ rad/sec $\neq \omega_{CO}$. The amplitude of the variable e is read directly from the plot of $-1/N_s(a)$ as $E_{LC} = 377.2$; to obtain the LC amplitude in roll, one must obtain the loop gain from e to ϕ, which is $G_\phi = R_0\, N_s(E_{LC})/(J\,\omega_{LC}^2) = 1/3033$, giving the LC amplitude in roll to be $A_{LC} = 0.124$ rad (peak). In Gibson, it is said that an analog computer solution yielded $\omega_{LC} = 22.9$ rad/sec and $A_{LC} = 0.135$ rad, which agrees quite well. A highly rigorous digital simulation approach (for which MATLAB-based software is available from the author; [34]) using modes to capture the switching characteristics of the hysteretic relay yielded $A_{LC} = 0.133$, $\omega_{LC} = 23.09$ rad/sec, as shown in Fig. 8, which is in slightly better agreement with the SIDF analysis. As is generally true, the approximation for ω_{LC} is better than that for a – the solution for ω_{LC} is a second-order approximation, while that for a is of first order [3, 8].

Finally, it should be observed that for the particular case of nonlinear systems with relays a complementary approach for LC analysis exists, due to Tsypkin [38]; see also [3, 5]. Rather than assuming that the nonlinearity input is a sinusoid, one assumes that the nonlinearity output is a switching signal, in this case, a signal that switches between F_0 and $-F_0$ at unknown times; one may solve exactly for the switching times and signal waveform by expressing the relay output in terms of a Fourier series expansion and solving for the switching conditions and coefficients. Alternatively, one could extend the SIDF approach by setting up and solving the harmonic balance relations for higher

terms; that approach would converge to the exact solution as the number of harmonics considered increases [17].

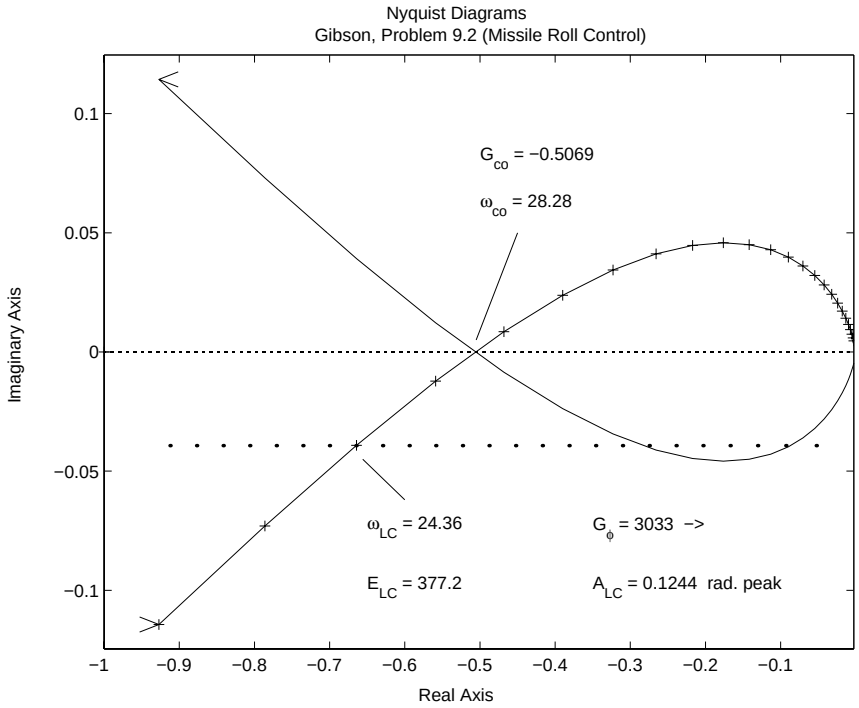

Figure 7: Solution for Missile Roll-Control Problem

3 Frequency Response Modeling

As mentioned in the introductory comments, SIDF approaches have been used for two primary purposes: limit-cycle analysis and characterizing the input/output (I/O) behavior of a nonlinear plant in the frequency domain. In this section we outline and illustrate two methods for determining the amplitude-dependent frequency response of a nonlinear system, hereafter more succinctly called an SIDF I/O model. After that, we discuss some broader but more complicated issues, to establish a context for this process.

3.1 Methods for determining frequency response

As mentioned, SIDF I/O modeling may be accomplished using either of two techniques:

1. **Analytic method using harmonic balance:** Given the general nonlinear dynamics as follows,

$$\dot{x} = f(x,\, u) \tag{25}$$

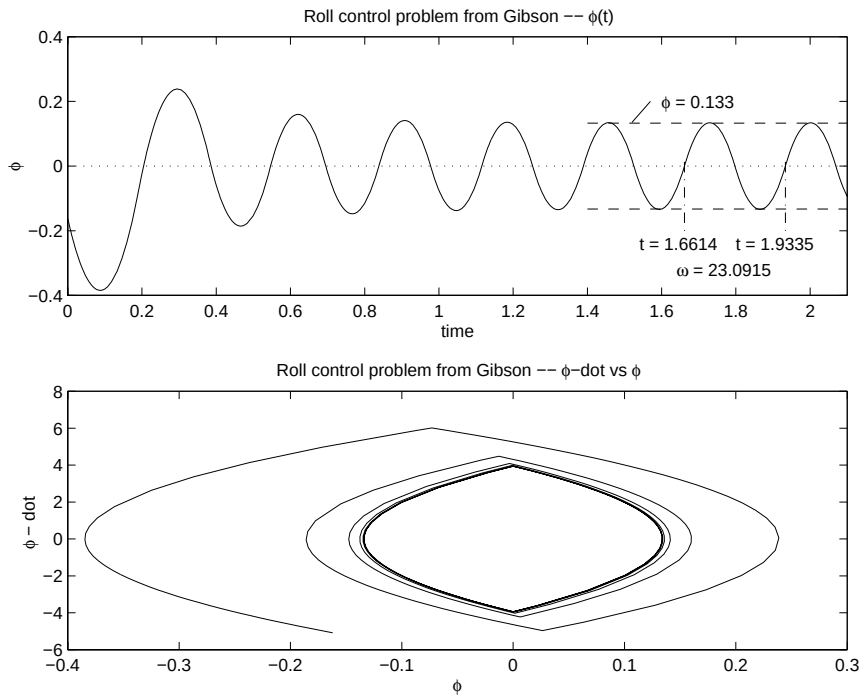

Figure 8: Missile Roll-Control Simulation Result

with a scalar input signal $u(t) = u_0 + b_u \cos(\omega t)$ and the n-dimensional state vector x assumed to be nearly sinusoidal,

$$x \cong x_c + \mathrm{Re}\,[a_x\ \exp(j\omega t)] \tag{26}$$

The variable a_x is a complex amplitude vector (in phasor notation) and x_c is the state-vector center value. We proceed to develop a quasilinear state-space model of the system in which every nonlinear element is replaced analytically by the corresponding scalar SIDF, and formulate the quasilinear equations:

$$\begin{aligned} f(x,\, u) \cong F_{DF}(u_0,\, b_u,\, x_c,\, a_x) \quad &+ \quad \mathrm{Re}\,[A_{DF}(u_0,\, b_u,\, x_c,\, a_x) \cdot a_x\ \exp(j\omega t)] \\ &+ \quad \mathrm{Re}\,[B_{DF}(u_0,\, b_u,\, x_c,\, a_x) \cdot b_u\ \exp(j\omega t)] \end{aligned} \tag{27}$$

Then, formulate the equations of harmonic balance, which for the DC and sinusoidal components are:

$$F_{DF}(u_0,\, b_u,\, x_c,\, a_x) = 0 \tag{28}$$

$$j\,\omega I\, a_x = A_{DF}\, a_x + B_{DF}\, b_u \tag{29}$$

One can, in principle, solve these equations for the unknown amplitudes x_c, a_x given values for u_0, b_u and then evaluate A_{DF} and B_{DF}; however, this is difficult in general and requires special nonlinear equation-solving software. Then, assuming finally that there is a linear output relation $y = C\,x$ for simplicity, the

I/O model may be evaluated as:

$$G(j\omega;\, u_0, b_u) = C\,[\,j\,\omega\,I - A_{DF}\,]^{-1}\,B_{DF} \tag{30}$$

Note that all arrays in the quasilinear model may depend on the input amplitude, u_0, b_u. This approach was used in [33] in developing an automated modeling approach called the *model-order deduction algorithm for nonlinear systems* (MODANS); refer to [27] for further details in the solution of the harmonic balance problem. This approach is subject to argument about the validity of assuming that every nonlinearity input is nearly sinusoidal. It is also more difficult than the following, and not particularly recommended – it is, however, the "pure" SIDF method for solving the problem.

2. **Simulation method:** Apply a sinusoidal signal to the nonlinear system model, perform direct Fourier integration of the system output in parallel with simulating the model's response to the sinusoidal input, and simulate until steady state is achieved to obtain the dynamic or frequency-domain SIDF $G(j\omega;\, u_0,\, b_u)$ [31].

To elaborate on the second method and illustrate its use, we assume for simplicity that $u_0 = 0$ and focus on determining $G(j\omega,\, b)$ for a range of input amplitudes $[b_{min},\, b_{max}]$ to cover the expected operating range of the system and frequencies $[\omega_{min},\, \omega_{max}]$ to span a frequency range of interest. Then specific sets of values $\{b_i\} \in [b_{min},\, b_{max}]$ and $\{\omega_j\} \in [\omega_{min},\, \omega_{max}]$ are selected for generating $G(j\omega_j,\, b_i)$. The nonlinear system model is augmented by adding new states corresponding to the Fourier integrals

$$\begin{aligned}
\mathrm{Re}\,(G_K(j\omega_j, b_i)) &= \frac{\omega_j}{\pi b_i}\int_{KT}^{(K+1)T} y(t)\cos(\omega t)dt \\
\mathrm{Im}\,(G_K(j\omega_j, b_i)) &= \frac{-\omega_j}{\pi b_i}\int_{KT}^{(K+1)T} y(t)\sin(\omega t)dt
\end{aligned}$$

where $\mathrm{Re}\,(\cdot)$ and $\mathrm{Im}\,(\cdot)$ are the real and imaginary parts of the SIDF I/O model $G(j\omega_j,\, b_i)$, $T = 2\pi/\omega_j$, and $y(t)$ is the output of the nonlinear system. In other words, the derivatives of the augmented states are proportional to $y(t)\cos(\omega t)$ and $y(t)\sin(\omega t)$. Achieving steady state for a given b_i and ω_j is guaranteed by setting tolerances and convergence criteria on the magnitude and phase of G_K where K corresponds to the number of cycles simulated; the integration is interrupted at the end of each cycle and the convergence criteria checked to see if the results are within tolerance (G_K is acceptably close to G_{K-1}), so that the simulation could be stopped and $G(j\omega_j, b_i)$ reported. For further detail, refer to [31, 32]. It should be mentioned that convergence can be slow if the simulation initial conditions are chosen without thought, especially if lightly-damped modes are present. We have found that a converged solution point from the simulation for ω_j can serve as a good initial condition for the simulation for ω_{j+1}, especially if the frequencies are closely spaced. The MATLAB-based software for performing this task is available from the author.

Example 3 – First, a brief demonstration of setting up and solving harmonic balance relations: Given a simple closed-loop system composed of an ideal relay and a linear dynamic block $W(j\,\omega)$, as shown in Fig. 9. If the input is $u(t) = b\cos(\omega t)$ then $y(t) \cong \mathrm{Re}\,[\,c(b)e^{j\omega t}\,]$ and, similarly, $e(t) \cong \mathrm{Re}\,[\,e(b)e^{j\omega t}\,]$, where, in general, c and e are complex phasors. These three phasors are related by

$$\begin{aligned} c &= N_s(\,|\,e\,|\,)W(j\,\omega)\,e \\ &= N_s(\,|\,e\,|\,)W(j\,\omega)\,(b-c) \end{aligned} \tag{31}$$

The SIDF for the ideal relay is $N_s = 4\,D/(\pi\,|\,e\,|\,)$, so the overall I/O relation is

$$G(j\omega,\,b) = \frac{c}{b} = \frac{4\,D}{\pi\,|\,b-c\,|} \cdot \frac{(b-c)}{b} \tag{32}$$

Taking the magnitude of this relation factor by factor yields

$$|\,G(j\omega,\,b)\,| = \frac{4\,D}{\pi\,b}\,|\,W(j\,\omega)\,| \tag{33}$$

It is interesting to note that the feedback does not change the frequency dependence of $|\,W(j\,\omega)\,|$, just the phase – this is not surprising, since the output of the relay always has the same amplitude which is then modified only by $|\,W(j\,\omega)\,|$. The relationship between the phases of G and W is not so easy to determine, even in this special case (ideal relay).

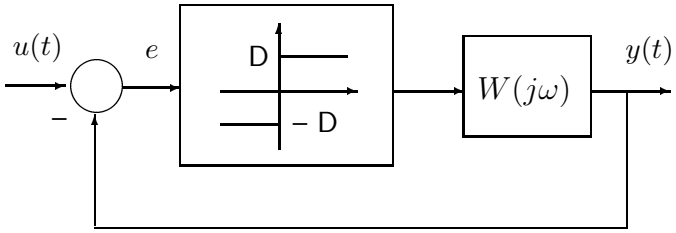

Figure 9: System with Relay and Linear Dynamics

Example 4 – To illustrate the simulation approach to generating SIDF I/O models, consider the simple nonlinear model of a motor and load depicted in Fig. 10, where the primary nonlinear effects are torque saturation and stiction (nonlinear friction characterized by "sticking" whenever the load velocity passes through zero). This model has been used as a challenging exemplary problem in a series of projects studying various SIDF-based approaches for designing nonlinear controllers, as discussed in Sect. 6. The mathematical model for stiction is given by the torque relation

$$T_m = \begin{cases} T_e - f_v\dot{\theta} - f_c\,\mathrm{sgn}(\dot{\theta})\,, & \dot{\theta} \neq 0 \text{ or } |T_e| > f_c \\ 0\,, & \text{otherwise} \end{cases} \tag{34}$$

where T_e and T_m denote electrical and mechanical torque, respectively, and, of necessity, we include a viscous friction term $f_v\dot{\theta}$ along with a Coulomb component of value f_c.

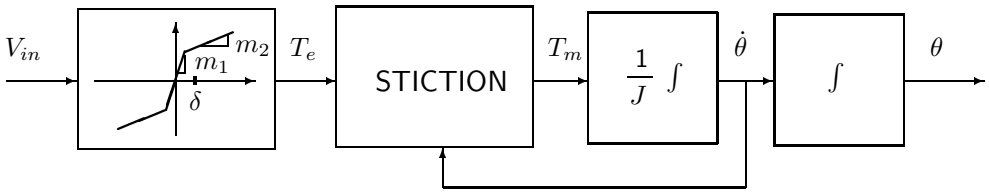

Figure 10: Motor Plus Load: Model Schematic

To generate the amplitude-dependent SIDF models we selected twelve frequencies, from 5 rad/sec to 150 rad/sec, and eight amplitudes, from quite small ($b_1 = 0.25$ volts) to quite large ($b_8 = 12.8$ volts). The results are shown in Fig. 11. The magnitude of $G(j\omega, b)$ varies by nearly a factor of 8, and the phase varies by up to 45 deg, showing that this system is substantially nonlinear over this operating range.

3.2 Methods for accounting for nonlinearity

Various ways exist to account for amplitude sensitivity in nonlinear dynamic systems. This is a very important consideration, both generally and particularly in the context of models for control system design. Approaches for dealing with static nonlinear characteristics in such systems include replacing nonlinearities with linear elements having gains that lie in ranges based on:

- nonlinearity sector bounds,
- nonlinearity slope bounds,
- random-input describing functions (RIDFs), or
- sinusoidal-input describing functions (SIDFs).

In brief, frequency-domain plant input/output models based on SIDFs provide an excellent trade-off between conservatism and robustness in this context. In particular, it can be shown by example that sector and slope bounds may be excessively conservative, while RIDFs are generally not robust, in the sense that a nonlinear control system design predicted to be stable based on RIDF plant models may limit cycle or be unstable. Another important attribute of SIDF-based frequency-domain models is that they account for the fact that the effect of most nonlinear elements depends on frequency

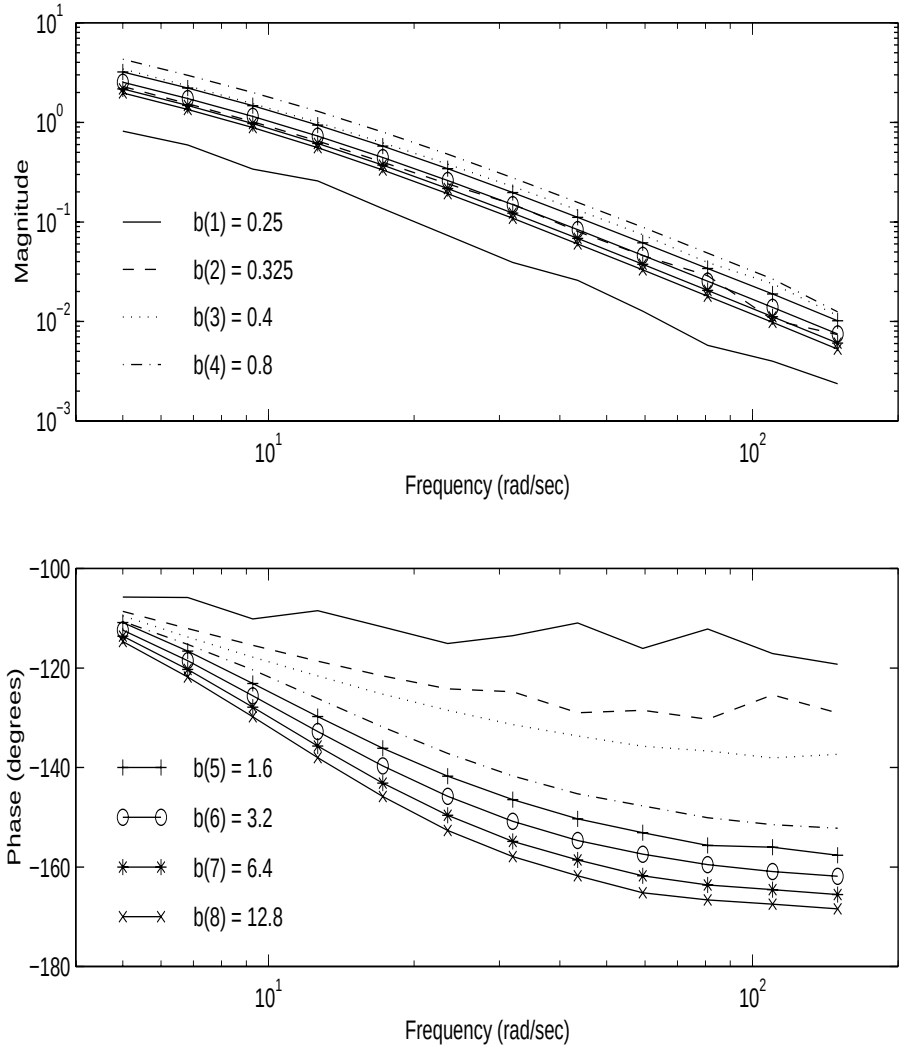

Figure 11: Motor Plus Load: SIDF I/O Models

as well as amplitude; none of the other techniques capture both of these traits. These points are discussed in more detail below; note that it is assumed that no biases exist in the nonlinear dynamic system, for the sake of simplicity; extending the arguments to systems with biases is straightforward.

Linear model families ($\dot{x} = Ax + Bu$) can be obtained by replacing each plant nonlinearity with a linear element having a gain that lies in a range based on its sector bound or slope bound. We will hereafter call these model families *sector* I/O *models* and *slope* I/O *models,* respectively. (Robustness cannot be achieved using *one* linear model based on the slope of each nonlinearity at the operating point for design, so that alternative is not considered.)

From the standpoint of robustness in the sense of maintaining stability in the presence of plant I/O variation due to amplitude sensitivity, it has been established that none of the model families defined above provide an adequate basis for a *guarantee.* The idea that sector I/O models would suffice is called the Aizerman conjecture, and the premise that slope I/O models are useful in this context is the conjecture of Kalman; both have been disproven even in the case of a single nonlinearity (for discussion, see [18]). Both SIDF and RIDF models similarly can be shown to be inadequate for a robustness guarantee in this sense (see also [18]).

Despite the fact that these model families cannot be used to guarantee stability robustness, it is also true that in many circumstances they are conservative. For example, a particular nonlinearity may pass well outside the sector for which the Aizerman conjecture would suggest stability and yet the system may still be stable. On the other hand, only very conservative conditions such as those imposed by the Popov criterion [20] (MATLAB-based software for applying the Popov criterion is available from the author [36]) and the off-axis circle criterion [6] serve this purpose rigorously – however, the very stringent conditions these criteria impose and the difficulty of extension to systems with multiple nonlinearities generally inhibit their use. Thus many control system designs are based on one of the model families under consideration as a (hopeful) basis for robustness. *It can be argued that designs based on* SIDF I/O *models that predict that* LC*s will not exist by a substantial margin is the best one can achieve in terms of robustness* (see also Atherton [4]). In SIDF-based synthesis the frequency-domain design objective (see Sect. 6.1) must ensure this.

Returning to conservatism, considering a static nonlinearity and assuming that it is single-valued and its derivative exists everywhere, it can be stated that slope I/O models are always more conservative than sector I/O models, which in turn are always more conservative than SIDF models. This is because the range of an SIDF cannot exceed the sector range, and the sector range cannot exceed the slope range. An additional argument that sector and slope model families may be substantially more conservative than SIDF I/O models is based on the fact that only SIDF models account for the frequency dependence of each nonlinear effect. This is especially important in the case of multiple nonlinearities, as illustrated by the simple example depicted in Fig. 12: Denoting the minimum and maximum slopes of the gain-changing nonlinearities f_k by $\underline{m}_k$ and $\overline{m}_k$ respectively, we see that the sector and slope I/O models correspond to all linear systems with gains lying in the indicated rectangle, while SIDF I/O models only correspond to a gain trajectory as shown (the exact details of which depend on the linear dynamics that precede each nonlinearity). In many cases, the SIDF model will clearly prove to be a less restrictive basis for control synthesis.

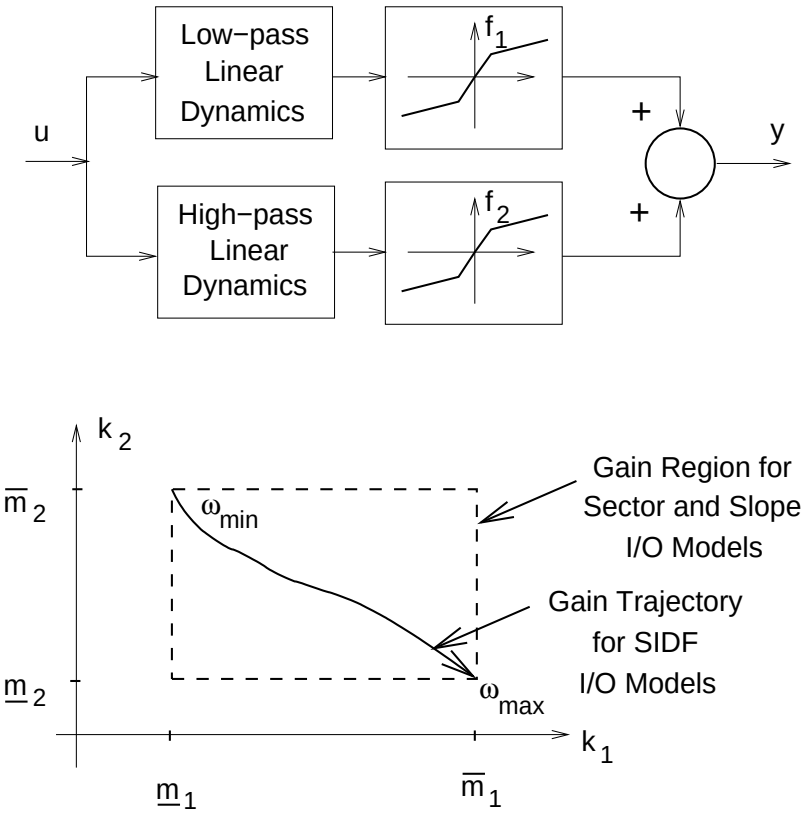

Figure 12: Frequency Dependence of Multiple Limiters

Returning to DF models, there are two basic differences between SIDF and RIDF models for a static nonlinearity, as mentioned above, namely, the assumed input amplitude distribution is different, and SIDFs can characterize the effective phase shift caused by multivalued nonlinearities such as those commonly used to represent hysteresis and backlash, while RIDFs cannot. In Sect. 1.5 we see that the input amplitude distribution issue is generally not a major consideration.

However, there is a third difference [27] that impacts the I/O model of a nonlinear plant in a fundamental way. This difference is related to how the DF is used in determining the I/O model; the result is that RIDF plant models (as usually defined) also fail to capture the frequency dependence of the system nonlinear effects.

This difference arises from the fact that the standard RIDF model is the result of *one* quasilinearization procedure carried out over a wide band of frequencies, while the SIDF model is obtained by multiple quasilinearizations at a *number* of frequencies, as we have seen. This behavior is best understood via a simple example [27] involving a low-pass linear system followed by a saturation (unity limiter), defined in Eqn. 23:

- Considering sinusoidal inputs of amplitude substantially greater than unity, the following behavior is exhibited: Low-frequency inputs are only slightly atten-

uated by the linear dynamics, resulting in heavy saturation and reduced SIDF gain; however, as frequency and thus attenuation increases, saturation decreases correspondingly and eventually disappears, giving a frequency response $G(j\omega, b)$ that approaches the response of the low-pass linear dynamics alone.

- A random input with rms value greater than one, on the other hand, results in saturation at all frequencies, so $G(j\omega, b)$ is identical to the linear dynamics followed by a gain less than unity.

In other words, the SIDF approach captures both a gain change and an effective increase in the transfer function magnitude corner frequency for larger input amplitudes, while the RIDF model shows only a gain reduction.

4 Analyzing the Performance of Nonlinear Stochastic Systems

This application of RIDFs represents the most powerful use of statistical linearization. It also represents a major departure from the class of problems considered so far. To set the stage, we must outline a class of nonlinear stochastic problems that will be tackled and establish some relations and formalism [28].

The dynamics of a nonlinear continuous-time stochastic system can be represented by a first-order vector differential equation in which $x(t)$ is the system state vector and $w(t)$ is a forcing function vector,

$$\dot{x}(t) = f(x, t) + G(t)\, w(t) \tag{35}$$

The state vector is composed of any set of variables sufficient to describe the behavior of the system completely. The forcing function vector $w(t)$ represents disturbances as well as control inputs that may act upon the system. In what follows, $w(t)$ is assumed to be composed of a mean or deterministic part $b(t)$ and a random part $u(t)$, the latter being composed of elements that are uncorrelated in time; that is, $u(t)$ is a white noise process having the spectral density matrix $Q(t)$. Similarly, the state vector has a deterministic part $m(t) = E[x(t)]$ and a random part $r(t) = x(t) - m(t)$; for simplicity, $m(t)$ is usually called the mean vector. Thus the state vector $x(t)$ is described statistically by its mean vector $m(t)$ and covariance matrix, $S(t) = E[r(t)\, r^T(t)]$. Henceforth, the time dependence of these variables will usually not be denoted explicitly by (t).

Note that the input to Eqn. 35 enters in a linear manner. This is for technical reasons, related to the existence of solutions. (In Itô calculus the stochastic differential equation $dx_t = f(x, t)\, dt + g(x, t)\, d\beta_t$ has well-defined solutions, where β_t is a Brownian motion

process. Equation 35 is an informal representation of such systems.) This is not a serious limitation; a stochastic system may have random inputs that enter nonlinearly if they are, for example, band-limited first-order Markov processes, characterized by $\dot{z} = A(t)z + B(t)w$, where again w contains white noise components – in this case, one may append the Markov process states z to the physical system states x so $f(x, z)$ models the nonlinear dependence of $\dot{x}$ on the random input z.

The differential equations that govern the propagation of the mean vector and covariance matrix for the system described by Eqn. 35 can be derived directly, as demonstrated in [11], to be

$$\begin{aligned} \dot{m} &= E[\, f(x,t)\,] + G(t)\, b \\ &\triangleq \hat{f} + G(t)\, b \\ \dot{S} &= E[\, f\, r^T\,] + E[\, r\, f^T\,] + G(t)\, Q\, G^T(t) \end{aligned} \tag{36}$$

The equation for S can be put into a form analogous to the covariance equations corresponding to $f(x,t)$ being linear, by defining the auxiliary matrix N_R through the relationship

$$N_R\, S \triangleq E[\, f(x,\, t)\, r^T\,] \tag{37}$$

Note that the RIDF matrix N_R is the direct vector / matrix extension of the scalar describing function definition, Eqn. 18. Then Eqn. 36 may be written as

$$\begin{aligned} \dot{m} &= \hat{f} + G(t)\, b \\ \dot{S} &= N_R\, S + S\, N_R^T + G(t)\, Q\, G^T(t) \end{aligned} \tag{38}$$

The quantities $\hat{f}$ and N_R defined in Eqns. 36 and 37 must be determined before one can proceed to solve Eqn. 38. Evaluating the indicated expected values requires knowledge of the joint probability density function (pdf) of the state variables. While it is possible, in principle, to evolve the n-dimensional joint pdf $p(x,t)$ for a nonlinear system with random inputs by solving a set of partial differential equations known as the Fokker-Planck equation or the forward equation of Kolmogorov [11], this has only been done for simple, low-order systems – so, this procedure is generally not feasible from a practical point of view. In cases where the pdf is not available the exact solution of Eqn. 38 is precluded.

One procedure for obtaining an approximate solution to Eqn. 38 is to assume the form of the joint pdf of the state variables in order to evaluate $\hat{f}$ and N_R according to Eqns. 36 and 37. Although it is possible to use any joint pdf, most development has been based on the assumption that the state variables are jointly normal; the choice was made because it is both reasonable and convenient.

While the above assumption is strictly true only for linear systems driven by Gaussian inputs, it is often approximately valid in nonlinear systems with nongaussian inputs.

Although the output of a nonlinearity with a Gaussian input is generally nongaussian, it is known from the central limit theorem [19] that random processes tend to become Gaussian when passed through low-pass linear dynamics (filtered). Thus, in many instances, one may rely on the linear part of the system to ensure that nongaussian nonlinearity outputs result in nearly Gaussian system variables as signals propagate through the system. By the same token, if there are nongaussian system inputs that are passed through low-pass linear dynamics, the central limit theorem can again be invoked to justify the assumption that the state variables are approximately jointly normal. The validity of the Gaussian assumption for nonlinear systems with Gaussian inputs has been studied and verified for a wide variety of systems.

From a pragmatic viewpoint, the Gaussian hypothesis serves to simplify the mechanization of Eqn. 38 significantly, by permitting each scalar nonlinear relation in $f(x,t)$ to be treated in isolation [9], with $\hat{f}$ and N_R formed from the individual RIDFs for each nonlinearity. Since RIDFs have been catalogued for numerous single-input nonlinearities [3, 8], the implementation of this technique is a straightforward procedure for the analysis of many nonlinear systems.

As a consequence of the Gaussian assumption, the RIDFs for a given nonlinearity are dependent only upon the mean and the covariance of the system state vector. Thus, $\hat{f}$ and N_R may be written explicitly as

$$\begin{aligned} \hat{f} &= \hat{f}(m,\, S,\, t) \\ N_R &= N_R(m,\, S,\, t) \end{aligned} \tag{39}$$

Relations of the form indicated in Eqn. 39 permit the direct evaluation of $\hat{f}$ and N_R at each integration step in the propagation of m and S. Note that the dependence of $\hat{f}$ and N_R on the statistics of the state vector is due to the existence of nonlinearities in the system.

A comparison of quasilinearization with the classical Taylor series or small-signal linearization technique provides a great deal of insight into the success of the RIDF in capturing the essence of nonlinear effects. Figure 4 illustrates this comparison with concrete examples. If a saturation or limiter is present in a system and its input v is zero-mean, the Taylor series approach leads to replacing $f(v)$ with a unity gain regardless of the input amplitude, while quasilinearization gives rise to a gain that decreases as the RMS value of v, σ_v, increases. The latter approximate representation of $f(v)$ much more accurately reflects the nonlinear effect; in fact, the saturation is completely neglected in the small-signal linear model, so it would not be possible to determine its impact. The fact that RIDFs retain this essential characteristic of system nonlinearities – input-amplitude dependence – provides the basis for the proven accuracy of this technique.

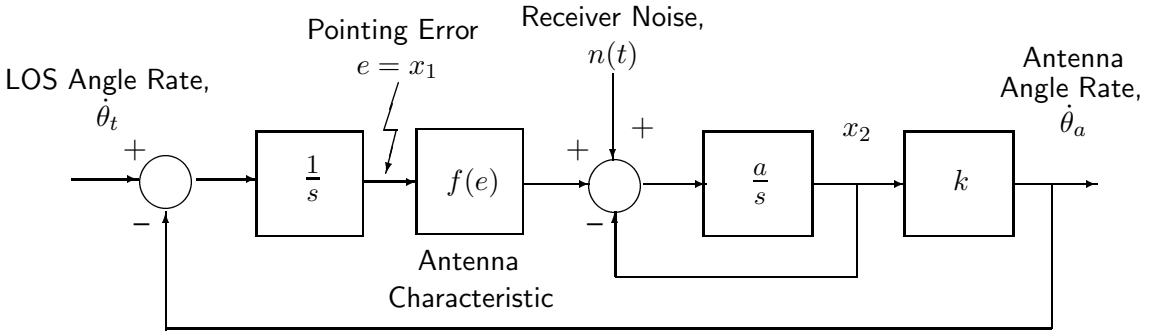

Figure 13: Antenna Pointing and Tracking Model

Example 5 – An antenna pointing and tracking study is treated in some detail, to illustrate this methodology. This problem formulation is taken directly from [16]; a discussion of the approach and results in [16] vis-à-vis the current treatment is given in [23].

The function of the antenna pointing and tracking system modeled in Fig. 13 is to follow a target line-of-sight (LOS) angle, θ_t. Assume that θ_t is a deterministic ramp,

$$\dot{\theta}_t = \Omega \, u_{-1}(t) \tag{40}$$

where Ω is the slope of the ramp and u_{-1} denotes the unit step function. The pointing error, $e = \theta_t - \theta_a$ where θ_a is the antenna centerline angle, is the input to a nonlinearity $f(e)$ which represents the limited beamwidth of the antenna; for the present discussion,

$$f(e) = e(\, 1 - k_a \, e^2 \,) \tag{41}$$

where k_a is suitably chosen to represent the antenna characteristic. The noise $n(t)$ injected by the receiver is a white noise process having zero mean and spectral density q.

In a state space formulation, Fig. 13 is equivalent to

$$\dot{x} = f(x) + w \tag{42}$$

where x_1 is the pointing error, e, x_2 models the slewing of the antenna via a first-order lag, as defined in Fig. 13, and

$$f(x) = \begin{bmatrix} -k \, x_2 \\ a \, (f(x_1) - x_2) \end{bmatrix} \; ; \; w = \begin{bmatrix} \dot{\theta}_t \\ a \, n(t) \end{bmatrix} \tag{43}$$

The statistics of the input vector w are given by

$$E[\, w \,] \stackrel{\Delta}{=} b = \begin{bmatrix} \Omega \\ 0 \end{bmatrix} \tag{44}$$

$$E[\,(w(t)-b)\,(w(\tau)-b)^T\,] \triangleq Q\,\delta(t-\tau) = \begin{bmatrix} 0 & 0 \\ 0 & a^2\,q \end{bmatrix} \delta(t-\tau) \tag{45}$$

The initial state variable statistics, assuming $x_2(0) = 0$, are

$$E[\,x(0)\,] \triangleq m_0 = \begin{bmatrix} m_{e0} \\ 0 \end{bmatrix} \;;\; E[\,(x(0)-m_0)\,(x(0)-m_0)^T\,] \triangleq \begin{bmatrix} \sigma_{e0}^2 & 0 \\ 0 & 0 \end{bmatrix} \tag{46}$$

where m_{e0} and σ_{e0} are the initial mean and standard deviation of the pointing error, respectively.

The above problem statement is in a form suitable for the application of the RIDF-based covariance analysis technique. The quasilinear RIDF representation for $f(e)$ in Eqn. 41 is of the form

$$\begin{aligned} f(x_1) &\cong \hat{f} + N_n\,(x_1 - m_1) \\ &= \left(m_1 - k_a\,(m_1^2 + 3\,\sigma_1^2)\,m_1\right) + \left(1 - 3\,k_a\,(m_1^2 + \sigma_1^2)\right)\,(x_1 - m_1) \end{aligned} \tag{47}$$

where m_1 and σ_1^2 are elements of m and S, respectively. The solution is then obtained by solving Eqn. 38, which specializes to

$$\dot{m} = \begin{bmatrix} -k\,m_2 \\ a\,(\hat{f} - m_2) \end{bmatrix} + b \tag{48}$$

$$\dot{S} = N_R\,S + S\,N_R^T + Q \;;\; N_R = \begin{bmatrix} 0 & -k \\ a\,N_n & -a \end{bmatrix} \tag{49}$$

subject to the initial conditions in Eqn. 46. As noted previously, Eqn. 49 is exact if x is a vector of jointly Gaussian random variables; if the initial conditions and noise are Gaussian and the effect of the nonlinearity is not too severe, the RIDF solution will provide a good approximation.

The goal of this study is to determine the system's tracking capability for various values of Ω; for brevity, only the results for $\Omega = 5$ deg/sec are shown. The system parameters are: $a = 50$ sec^{-1}, $k = 10$ sec^{-1}, $k_a = 0.4$ deg^{-2}; the pointing error initial condition statistics are $m_{e0} = 0.4$ deg, $\sigma_{e0} = 0.1$ deg; and the noise spectral density is $q = 0.004$ deg^2. The RIDF solution depicted in Fig. 14 is based on the Gaussian assumption.

Three solutions are presented in Fig. 14, to provide a basis for assessing the accuracy of RIDF-based covariance analysis. In addition to the RIDF results, ensemble statistics from a 500-trial Monte Carlo simulation are plotted, along with the corresponding 95 % confidence error bars, calculated on the basis of estimated higher-order statistics [26]. Also, the results of a linearized covariance analysis are shown, based on assuming that the antenna characteristic is linear ($k_a = 0$). The linearized analysis indicates that the pointing error statistics reach steady-state values at about $t = 0.2$ sec. However, it is

evident from the two nonlinear analyses that this not the case: in fact, the tracking error can become so large that the antenna characteristic effectively becomes a negative gain, producing unstable solutions (the antenna loses track). The same is true for higher slewing rates; for example, $\Omega = 6$ deg/sec was investigated in [16] and [23]; in fact, the second-order Volterra analysis in [16] also missed the instability (loss of track). Returning to the RIDF-based covariance analysis solution, observe that the time histories of $m_1(t)$ and $\sigma_1(t)$ are well within the Monte Carlo error bars, thus providing a excellent fit to the Monte Carlo data. The fourth central moment was also captured, to permit an assessment of deviation from the Gaussian assumption; the parameter λ (*kurtosis*, the ratio of the fourth central moment to the variance squared) grew to $\lambda = 8.74$ at $t = 0.3$ sec, which indicates a substantial departure from the Gaussian case ($\lambda = 3$); this is the reason the error bars are so much wider at the end of the study ($t = 0.3$ sec) compared with those near the beginning when in fact $\lambda \cong 3$ – higher kurtosis leads directly to larger 95 % confidence bands [26].

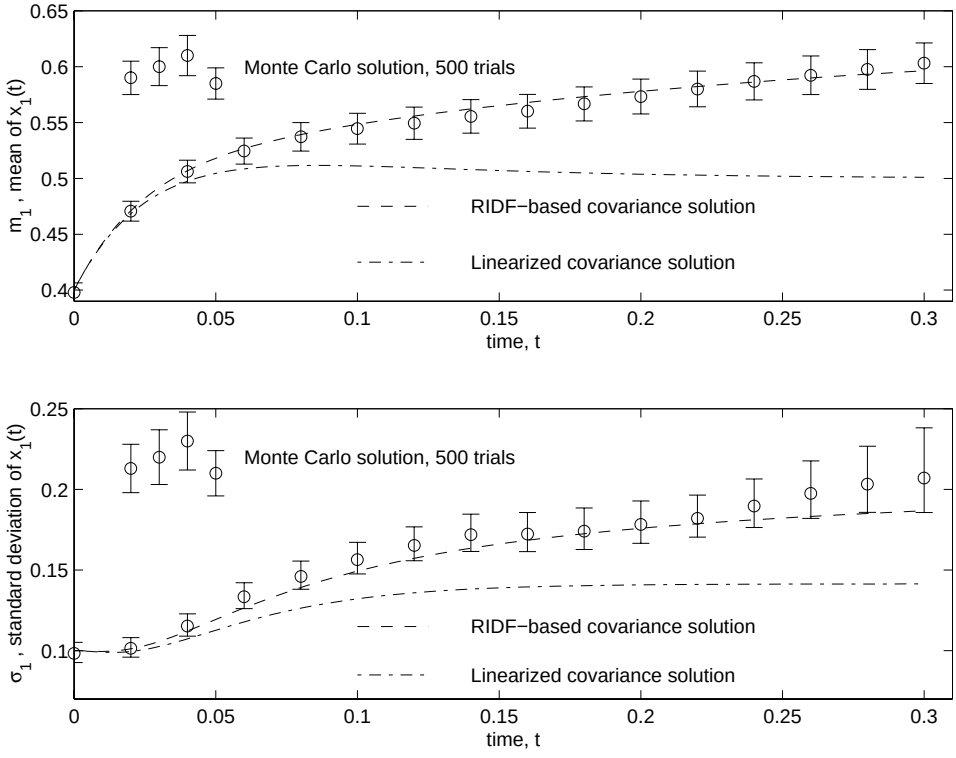

Figure 14: Antenna Pointing Error Statistics

Many other applications of RIDF-based covariance analysis have been performed (for several examples, see [28]). In every case, its ability to capture the significant impact of nonlinearities on system performance has been excellent, until system variables become highly nongaussian (roughly, until kurtosis exceeds about 10 to 15). It is recommended that some cases be spot-checked by Monte Carlo simulation; however, one should

recognize that one will have to perform many trials if kurtosis is high, and that knowing how many trials to perform is itself problematical (for details, see [26]).

5 Limit Cycle Analysis, Systems with Multiple Nonlinearities

Using SIDFs, as developed in Sects. 1 and 2, is a well-known approach for studying LCs in nonlinear systems with one dominant nonlinearity. Once that problem was successfully solved, many attempts were made to extend this method to permit the analysis of systems containing more than one nonlinearity. At first, the nonlinear system models that could be treated by such extensions were quite restrictive (limited to a few nonlinearities, or to certain specific configurations; cf. [3]). Furthermore, some results involved only conservative conditions for LC avoidance, rather than actual LC conditions. The technique described in this section (see [25]) removes all constraints: Systems described by a general state-vector differential equation, with any number of nonlinearities, may be analyzed. In addition, the nonlinearities may be multi-input, and bias effects can be treated. This general SIDF approach was first fully developed and applied to a study of wing-rock phenomena in aircraft at high angle of attack [22]. It was also applied to determine LC conditions for rail vehicles [12]. Its power and use are illustrated here by application to a second-order differential equation derived from a two-mode panel flutter model [13]. While the mathematical analysis is more protracted than in the single nonlinearity case, it is very informative and reveals the rich complexity of the problem.

The most general system model considered here is again as given in Eqn. 25. Assuming that u is a vector of constants, denoted u_0, it is desired to determine if Eqn. 25 may exhibit LC behavior. As before, we assume that the state variables are nearly sinusoidal, Eqn. 26, where a_x is a complex amplitude vector and x_c is the state-vector center value (which is not an equilibrium, or solution to $f(x_0, u_0) = 0$, unless the nonlinearities satisfy certain stringent symmetry conditions with respect to x_0). Then we again neglect higher harmonics, to make the approximation

$$f(x,\, u_0) \cong F_{DF}(u_0,\, x_c,\, a_x) + \mathrm{Re}\,[A_{DF}(u_0,\, x_c,\, a_x) \cdot a_x\, \exp(j\omega t)] \tag{50}$$

The real vector F_{DF} and the matrix A_{DF} are obtained by taking the Fourier expansions of the elements of $f(x_c + \mathrm{Re}\,[a_x\, \exp(j\omega t)\,],\, u_0)$ as illustrated below, and provide the quasilinear or describing function representation of the nonlinear dynamic relation. The assumed LC exists for $u = u_0$ if x_c and a_x can be found so that

$$\begin{aligned} F_{DF}(u_0,\, x_c,\, a_x) &= 0 \\ [j\omega I - A_{DF}(u_0,\, x_c,\, a_x)]\; a_x &= 0\,,\;\; a_x \neq 0 \end{aligned} \tag{51}$$

These nonlinear algebraic equations (51) are often difficult to solve. A second-order system with two nonlinearities (from a two-mode panel flutter model) can be treated by direct analysis, as shown below [24]. Even for this case the analysis is by no means trivial; it is included here for completeness and as guidance for the serious pursuit of LC conditions for multivariable nonlinear systems. It may be mentioned that iterative solution methods (e.g., based on successive approximation) have been used successfully to solve the DC and first-harmonic balance equations above for substantially more complicated problems such as the aircraft wing-rock problem (eight state variables, five multivariable nonlinear relations [22, 25]). More recently, a computer-aided design package for LC analysis of free structured multivariable nonlinear systems was developed [17] using both SIDF methods as well as extended harmonic balance (including the solution of higher-harmonic balance relations); it is noteworthy that the extended harmonic balance approach can, in principle, provide solutions with excellent accuracy, as long as enough higher harmonics are considered – one interesting example included balancing up to the 19^{th} harmonic.

Example 6 – The following second-order differential equation has been derived to describe the local behavior of solutions to a two-mode panel flutter model [13]

$$\ddot{\chi} + (\alpha + \chi^2)\,\dot{\chi} + (\beta + \chi^2)\,\chi = 0 \tag{52}$$

Heuristically, it is reasonable to predict that LCs may occur for negative α (so the second term provides damping that is negative for small values of χ but positive for large values). Observe also that there are three singularities if β is negative: $\chi_0 = 0, \pm\sqrt{-\beta}$. Making the usual choice of state vector, $x = [\,\chi \;\; \dot{\chi}\,]^T$, the corresponding state-vector differential equation is

$$\dot{x} = \begin{bmatrix} \dot{\chi} \\ \ddot{\chi} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\beta & -\alpha \end{bmatrix} x - \begin{bmatrix} 0 \\ x_1^2\,(x_1 + x_2) \end{bmatrix} \tag{53}$$

The SIDF assumption corresponding to Eqn. 26 for this system of equations is that

$$\begin{aligned} x_1 = \chi &\;\cong\; \chi_c + a_1 \cos(\omega t) \\ x_2 = \dot{\chi} &\;\cong\; -a_1\,\omega \sin(\omega t) \end{aligned}$$

(From the relation $x_2 = \dot{\chi}$ it is clear that x_2 has a center value of 0 and that $a_2 = -j\,\omega\,a_1$ – recognizing this at the outset greatly simplifies the analysis.) Therefore, the combined nonlinearity in Eqn. 53 may be quasilinearized to obtain

$$\begin{aligned} x_1^2\,(x_1 + x_2) &= (\chi_c + a_1\cos(\omega t))^2(\chi_c + a_1\cos(\omega t) - a_1\,\omega\sin(\omega t)\,) \\ &\cong (\chi_c^3 + \frac{3}{2}\chi_c\,a_1^2) + (3\,\chi_c^2 + \frac{3}{4}a_1^2)\,a_1\cos(\omega t) \\ &\qquad +(\chi_c^2 + \frac{1}{4}a_1^2)\,(-a_1\,\omega\sin(\omega t)) \\ &\triangleq \Phi_0 + N_{s1}\cdot a_1\cos(\omega t) + N_{s2}\cdot(-a_1\,\omega\sin(\omega t)) \end{aligned} \tag{54}$$

This result is obtained by expanding the first expression using trigonometric identities to reduce $\cos^2(v)$, $\cos^3(v)$ and $\sin(v)\cos^2(v)$ into terms involving $\cos(k\,v)$, $\sin(k\,v)$, $k = 0, 1, 2, 3$ and discarding all terms except the fundamental ones ($k = 0, 1$).

Therefore, the conditions of Eqn. 51 require that

$$F_{DF} = \begin{bmatrix} 0 \\ -\chi_c\,(\beta + \chi_c^2 + \frac{3}{2}a_1^2) \end{bmatrix} = 0 \tag{55}$$

$$A_{DF} = \begin{bmatrix} 0 & 1 \\ -(\beta + 3\chi_c^2 + \frac{3}{4}a_1^2) & -(\alpha + \chi_c^2 + \frac{1}{4}a_1^2) \end{bmatrix} \triangleq \begin{bmatrix} 0 & 1 \\ -\omega_{LC}^2 & 0 \end{bmatrix} \tag{56}$$

where we have taken advantage of the canonical second-order form of an A matrix with imaginary eigenvalues $\pm j\,\omega_{LC}$; again, the canonical form of A_{DF} ensures harmonic balance, not "pure imaginary eigenvalues". The relation in Eqn. 55 shows two possibilities:

- **Case 1:** $\chi_c = 0$, in which case Eqn. 56 yields

$$a_1 = 2\sqrt{-\alpha}\,, \quad \omega_{LC} = \sqrt{\beta - 3\,\alpha} \tag{57}$$

  The amplitude a_1 and frequency ω_{LC} must be real for LCs to exist. Thus, as conjectured, $\alpha < 0$ is required for a LC to exist centered about the origin. The second parameter must satisfy $\beta > 3\,\alpha$, so β can take on any positive value but cannot be more negative than $3\,\alpha$.

- **Case 2:** $\chi_c = \pm\sqrt{(\beta - 6\,\alpha)/5}$, yielding

$$a_1 = 2\sqrt{(\alpha - \beta)/5}\,, \quad \omega_{LC} = \sqrt{\beta - 3\,\alpha} \tag{58}$$

  For the two LCs in Case 2 to exist, centered at $\chi_c = \pm\sqrt{(\beta - 6\,\alpha)/5}$, it is necessary that $3\,\alpha < \beta < \alpha$, so again LCs cannot exist unless $\alpha < 0$. One additional constraint must be imposed: $|\,\chi_c\,| > a_1$ must hold, or the two LCs will overlap; this condition reduces the permitted range of β to $2\,\alpha < \beta < \alpha$.

The stability of the Case 1 LC can be determined as follows: Take any $\epsilon > 0$ and perturb the LC amplitude to a slightly larger value, e.g., $a_1^2 = -4\,\alpha + \epsilon$. Substituting into Eqn. 56 yields

$$A_{DF} = \begin{bmatrix} 0 & 1 \\ -(\beta - 3\alpha + \frac{3}{4}\epsilon) & -\frac{1}{4}\epsilon \end{bmatrix} \tag{59}$$

which for $\epsilon > 0$ has "slightly stable eigenvalues" (leads to loop gain less than unity). Thus a trajectory perturbed just outside the LC will decay, indicating that the Case 1 LC is stable. A similar analysis of the Case 2 LC is more complicated (since a perturbation in a_1 produces a shift in χ_c that must be considered), and thus is omitted.

Based on the SIDF-based LC analysis outlined above, the behavior of the original system Eqn. 53 is portrayed for $\alpha = -1$ and seven values of β in Fig. 15. The analysis has revealed the rather rich set of possibilities that may occur, depending on the values of α and β. One may use traditional singularity analysis to verify the detail of the solutions near each center [25], as shown in Fig. 15, but that is beyond the scope of this article. Also, one may use center manifold techniques to analyze the LC behavior shown here [13], but that effort would require additional higher-level mathematics and substantially more analysis to obtain the same qualitative results. We provide one simulation example in Fig. 16, for $\alpha = -1, \beta = -1.1$, which should produce the behavior portrayed in case E in Fig. 15. The results for two initial conditions, $x_1 = 0.4, 0.5$, $x_2 = 0$ (marked $\circ$), bracket the unstable LC with predicted center at (0.99, 0), while for two larger starting values, $x_1 = 1.6, 2.0$, $x_2 = 0$ (marked $\times$), we observe clear convergence to the stable LC with predicted center at (0 , 0). While the resultant stable oscillation is highly nonsinusoidal (and thus the Case 1 LC amplitude prediction is quite inaccurate), the SIDF prediction of panel flutter behavior is remarkably close.

Finally, it is worth mentioning that the "eigenvectors" of the matrix A_{DF} (state-vector amplitude in phasor form, Eqn. 51) may be very useful in some cases. For the wing-rock problem mentioned previously we encountered obscuring modes, i.e., slow unstable modes that made it essentially impossible to use simulation to verify the SIDF-based LC predictions. We were able to circumvent this difficulty by picking simulation initial conditions based on a_x corresponding to the predicted LC and thereby minimizing the excitation of the unstable mode and giving the LC time to develop before it was obscured by the instability.

6 Designing Nonlinear Controllers

Describing function methods – especially, using SIDF frequency-domain models as illustrated in Fig. 11 – for the design of linear and nonlinear controllers for nonlinear plants has a long history, and many approaches may be found in the literature. In general, the approach for *linear* controllers involves a direct use of frequency-domain design techniques applied to the family or a single (generally "worst-case" in some sense) SIDF model. The more interesting and powerful SIDF controller design approaches are those directed towards *nonlinear* compensation; that is the primary emphasis of the following discussions and examples.

A major issue in designing robust controllers for nonlinear systems is the amplitude sensitivity of the nonlinear plant and final control system. Failure to recognize and accommodate this factor may give rise to nonlinear control systems that behave differently for small *versus* large input excitation, or perhaps exhibit LCs or instability.

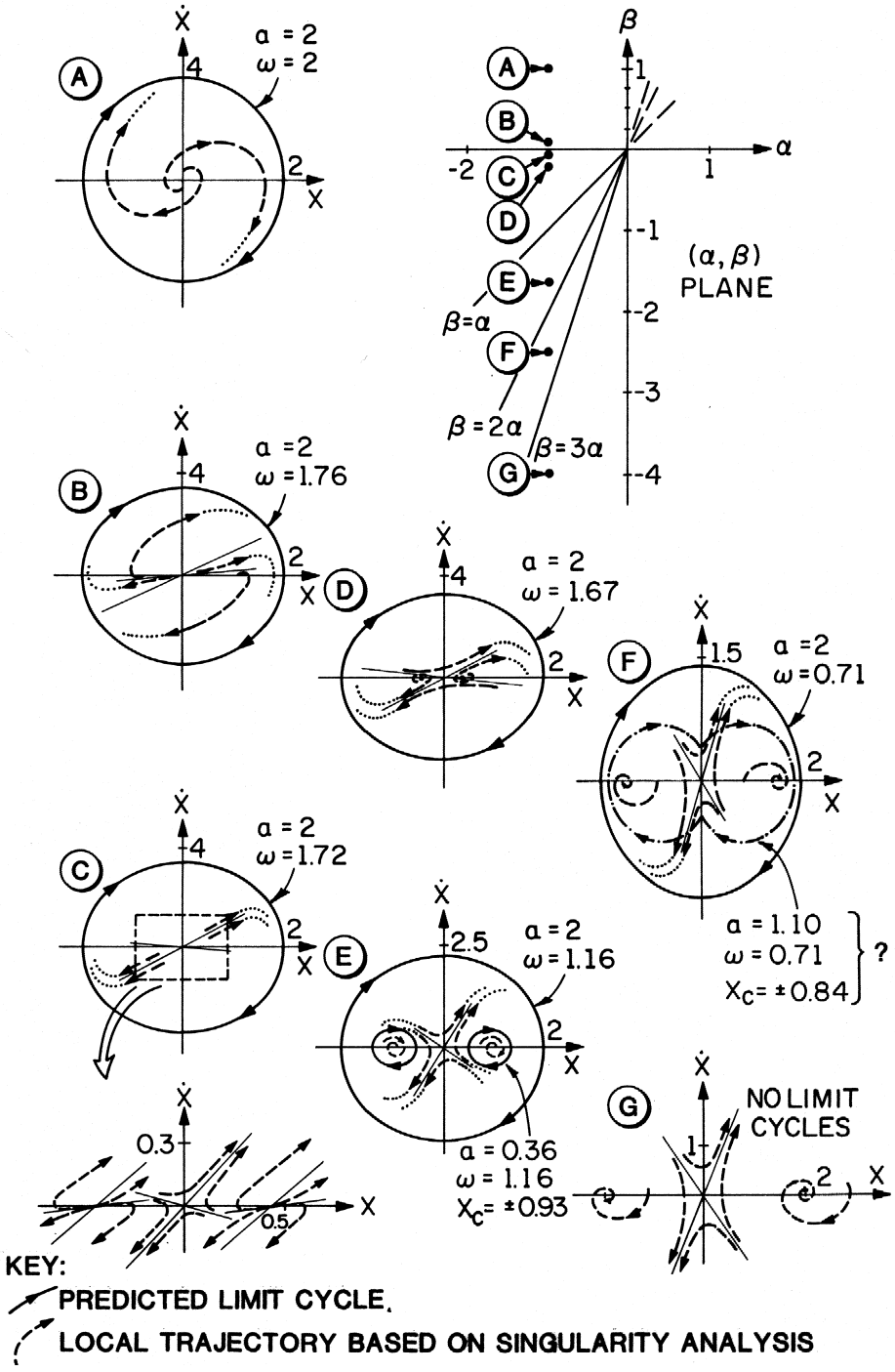

Figure 15: Limit Cycle Conditions for the Panel Flutter Problem (from Ref. [25])

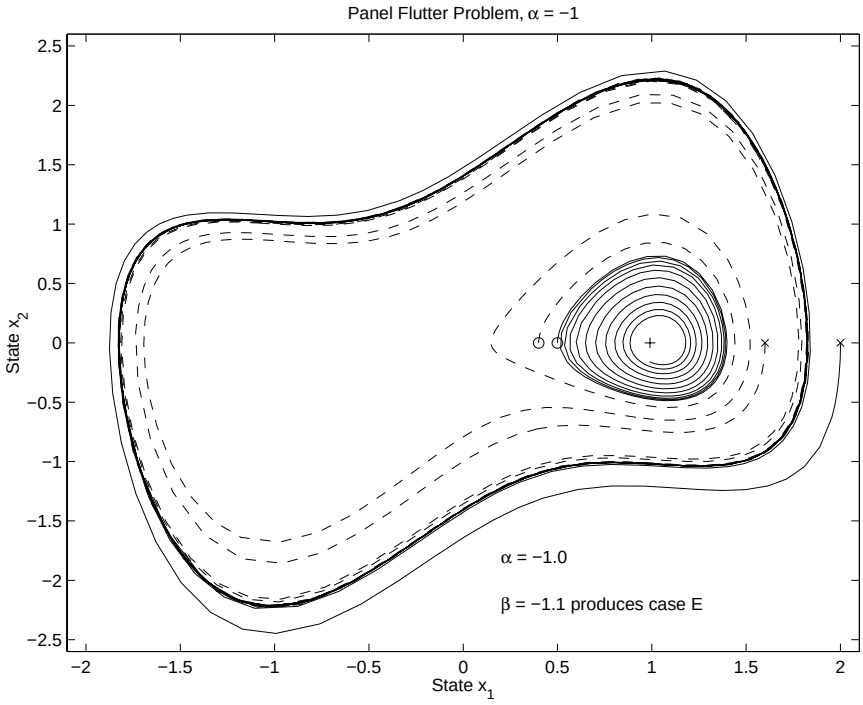

Figure 16: Panel Flutter Problem Simulation Result

Sinusoidal-input describing functions (SIDFs) have been shown to be effective in dealing with amplitude sensitivity in two areas: *modeling* (providing plant models that achieve an excellent trade-off between conservatism and robustness, Sect. 3) and *nonlinear control synthesis.* In addition, SIDF-based modeling and synthesis approaches are broadly applicable, in that there are very few and mild restrictions on the class of systems that can be handled. Several practical SIDF-based nonlinear compensator synthesis approaches are presented here and illustrated via application to a position control problem.

Before delving into specific approaches and results, the question of control system stability must be addressed. As mentioned in Sect. 1.1, the use of SIDFs to determine LC conditions is not exact. Specifically, if one were to try to determine the critical value of a parameter that would just cause a control system to begin to exhibit LCs, the conventional SIDF result might not be very accurate (this is not to say that more detailed analysis such as inclusion of higher harmonics in the harmonic balance method could not eventually give accurate results [17]). Therefore, if one desired to design a system to operate just at the margin of stability (onset of LCs) the SIDF method would not provide any guarantee.

To shun the use of SIDFs for nonlinear controller design for this reason would seem unwarranted, however. Generally one designs to safe specifications (good gain and

phase margin, for example) that are far from the margin of stability or LC conditions. The use of SIDF models in these circumstances is clearly so superior to the use of conventional linearized models, or even a set of linearized models generated to try to cover for uncertainty and nonlinearity, as discussed in Sect. 3.2, that we have no compunctions about recommending this practice. The added information (amplitude dependence) and intuitive support they provide is extremely valuable, as we hope the following examples demonstrate.

6.1 Design of conventional nonlinear controllers

Design approaches based on SIDF models are all frequency-domain in orientation. The basic idea of a family of techniques developed by the author and students [27, 29, 31, 37] is to define a *frequency-domain objective* for the open-loop compensated system and synthesize a nonlinear controller to meet that objective as closely as possible for a variety of error signal amplitudes (e.g., for small, medium and large input signals, where the numerical values associated with the terms "small", "medium" and "large" are based on the desired operating regimes of the final system). The designs are then at least validated in the time domain (e.g., step-response studies [29, 31]); recent approaches have added time-domain optimization to further reduce the amplitude sensitivity [35, 37]. The methods presented below all follow this outline.

Modeling and synthesis approaches based on these principles are broadly applicable. Plants may have any number of nonlinearities, of arbitrary type (even discontinuous or hysteretic), in any configuration. These methods are robust in several senses: In addition to dealing effectively with amplitude sensitivity, the exact form of each plant nonlinearity does not have to be known as long as the SIDF plant model captures the amplitude sensitivity with decent accuracy, and the final controller design is not specifically based on precise knowledge of the plant nonlinearities. The resulting controllers are simple in structure and thus readily implemented, with either piece-wise-linear characteristics [29, 31] or fuzzy logic [35, 37].

Before proceeding, it is important to consider the premises of the SIDF design approaches that we have been developing:

1. The nonlinear system design problem being addressed is the synthesis of controllers that are effective for plants having frequency-domain I/O models that are *sensitive* to input amplitude (e.g., for plants that behave very differently for small, medium and large input signals).
2. Our primary objective in nonlinear compensator design is to arrive at a closed-loop system that is as *insensitive* to input amplitude as possible.

This encompasses a limited but important set of problems, for which gain-scheduled compensators cannot be used (gain-scheduled compensators can handle plants whose behavior differs at different operating points but not amplitude-dependent plants; while a gain scheduled controller is often implemented with piece-wise-linear relations or other curve fits to produce a controller that smoothly changes its behavior as the operating point changes, these curve fits are usually completely unrelated to the differing behavior of the plant for various input amplitudes *at a given operating point*) and for which other approaches (e.g., variable structure systems, model-reference adaptive control, global linearization) do not apply because their objectives are different (e.g., their objectives deal with *asymptotic* solution properties rather than *transient* behavior, or they deal with the behavior of *transformed* variables rather than physical variables).

A number of controller configurations have been investigated as these approaches were developed, ranging from one nonlinearity followed by a linear compensator (which has quite limited capability to compensate for amplitude dependence) to a two-loop configuration with nonlinear rate feedback and a nonlinear proportional-integral (PI) controller in the forward path. Since the latter is most effective we will focus on that option [31].

An outline of the synthesis algorithm for the nonlinear PI plus rate feedback (PI+RF) controller is as follows:

1. Select sets of input amplitudes and frequencies that characterize the operating regimes of interest.
2. Generate SIDF I/O models of the plant corresponding to the input amplitudes and frequencies of interest (Sect. 3).
3. Design amplitude-dependent rate-feedback (RF) gains using an extension of the D'Azzo and Houpis algorithm [7], devised by Taylor and O'Donnell [31].
4. Convert these linear designs into a piece-wise-linear characteristic (RF nonlinearity) by sinusoidal-input describing function inversion.
5. Find SIDF I/O models for the nonlinear plant plus nonlinear RF compensation.
6. Design PI compensator gains using the frequency-domain sensitivity minimization technique described in Taylor and O'Donnell (1990).
7. Convert these linear designs into a piece-wise-linear PI controller, also by sinusoidal-input describing function inversion.
8. Develop a simulation model of the plant with nonlinear PI+RF control.
9. Validate the design through step-response simulation.

Steps 1, 2 and 5 are already described in detail in Sect. 3. In fact, the example and SIDF I/O model presented there (Fig. 11) were used in demonstrating the PI+RF controller

design method. Steps 3 and 4 proceed as follows:

a. The general objective when designing the inner-loop RF controller is to give the same benefits expected in the linear case, namely, stabilizing and damping the system, if necessary, and reducing the sensitivity of the system to disturbances and plant nonlinearities. At the same time, we wish to design a nonlinearity to be used with the controller that will desensitize the inner loop as much as possible to different input amplitudes.

b. As shown in D'Azzo and Houpis, it is convenient to work with inverse Nyquist plots of the plant I/O model, i.e., invert the SIDF frequency-response information in complex-gain form and plot the result in the complex plane. In the linear case, this allows us to study the closed-inner-loop (CIL) frequency response $G_{CIL}(j\omega)$ in the inverse form

$$\frac{1}{G_{CIL}(j\omega)} = \frac{1 + G(j\omega)H(j\omega)}{G(j\omega)} = \frac{1}{G(j\omega)} + H(j\omega)$$

where the effect of $H(j\omega)$ on $1/G_{CIL}(j\omega)$ is easily determined.

c. The inner-loop tachometer feedback design algorithm given by D'Azzo and Houpis and referred to as *Case 2* uses a construction amenable to extension to nonlinear systems. For linear systems, this algorithm is based on evaluating a tachometer gain and external gain in order to adjust the inverse Nyquist plot to be tangent to a given M-circle at a selected frequency. The algorithm is extended to the nonlinear case by applying it to each SIDF model $G(j\omega, b_i)$. Then for each input amplitude b_i a tachometer gain, $K_{T,i}$, and an external (to the inner loop) gain $A_{2,i}$ is found. The gains $A_{2,i}$ are discarded, since the external gain will be subsumed in the cascade portion of the controller that is synthesized in step 6.

d. The set of desired tachometer gains $K_{T,i}(b_i)$ is then used to synthesize the tachometer nonlinearity f_T. As first described in [27], these gain/amplitude data points are interpreted as SIDF information for an unknown static nonlinearity. A least-squares routine is used to adjust the parameters of a general piece-wise-linear nonlinearity so that the SIDF of that nonlinearity fits these gain/amplitude data with minimum mean squared error; this generates the desired RF controller nonlinearity, completing this step. This process of SIDF inversion is illustrated below (Fig. 17).

The final step in the complete controller design procedure is generating the nonlinear cascade PI compensator. The general idea is to first generate SIDF I/O models for the nonlinear plant (which, in this approach, is actually the nonlinear plant with nonlinear rate feedback) over the range of input amplitudes and frequencies of interest. This information forms a frequency-response map as a function of both input amplitude and

frequency. A single nominal input amplitude is selected, b^*, and a linear compensator is found that best compensates the plant *at that amplitude*. This compensator, in series with the nonlinear plant, is used to calculate the corresponding desired open-loop I/O model $CG^*(j\omega; b^*)$, the *frequency-domain objective function*. Then, at each input amplitude b_i a least-squares algorithm is used to adjust the parameters of a linear PI compensator, $K_{P,i}(b_i)$ and $K_{I,i}(b_i)$, to minimize the difference between the resulting frequency response found using the linear compensator and interpolating on the SIDF frequency-response map, and $CG^*(j\omega; b^*)$, as described in [29]. The nonlinear PI compensator is then obtained by synthesizing the nonlinearities f_P and f_I by SIDF inversion.

An important part of this procedure is the process called SIDF inversion, or adjusting the parameters of a general piece-wise-linear nonlinearity so that the SIDF of that nonlinearity fits the gain/amplitude data with minimum mean squared error. This step is illustrated in Fig. 17, where the piece-wise-linear characteristic had two breakpoints (δ_1, δ_2) and three slopes (m_1, m_2, m_3) that were adjusted to fit the gain/amplitude data (small circles) with good accuracy.

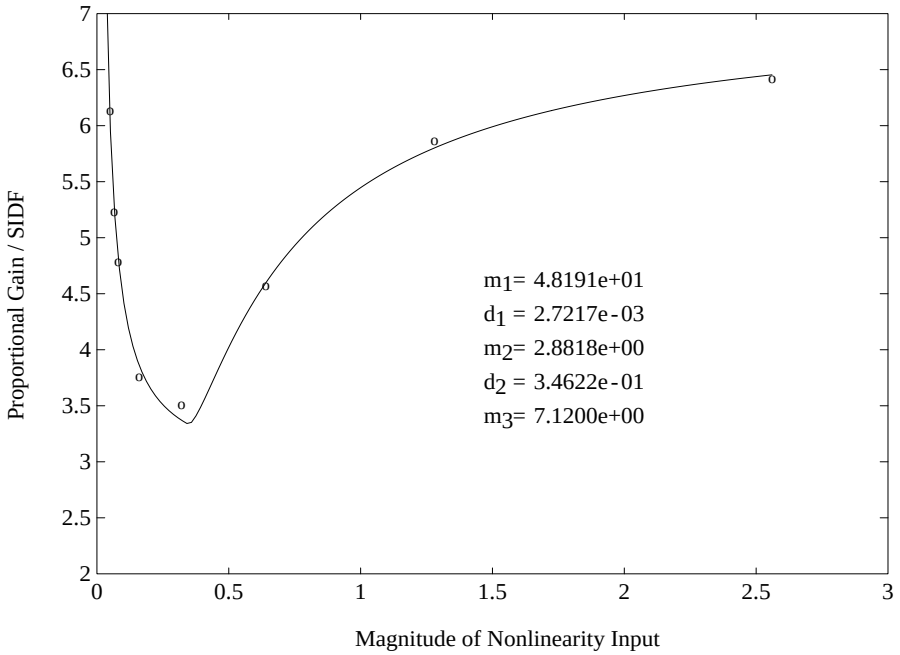

Figure 17: Proportional Nonlinearity Synthesis via DF Inversion

The final validation of the design is to simulate a family of step responses for the nonlinear control system. The results for the controller from [31] (which are identical to the results obtained with a later fuzzy-logic implementation that is also based on the direct application of this SIDF approach [35]) are depicted in Fig. 18, along with similar step responses generated using the linear PI+RF controller corresponding to the

frequency-domain objective function. In both response sets input amplitudes ranging from $b_1 = 0.20$ to $b_8 = 10.2$ were used, and the output normalized by dividing by b_i. Comparing the responses of the two controllers, it is evident that the SIDF design

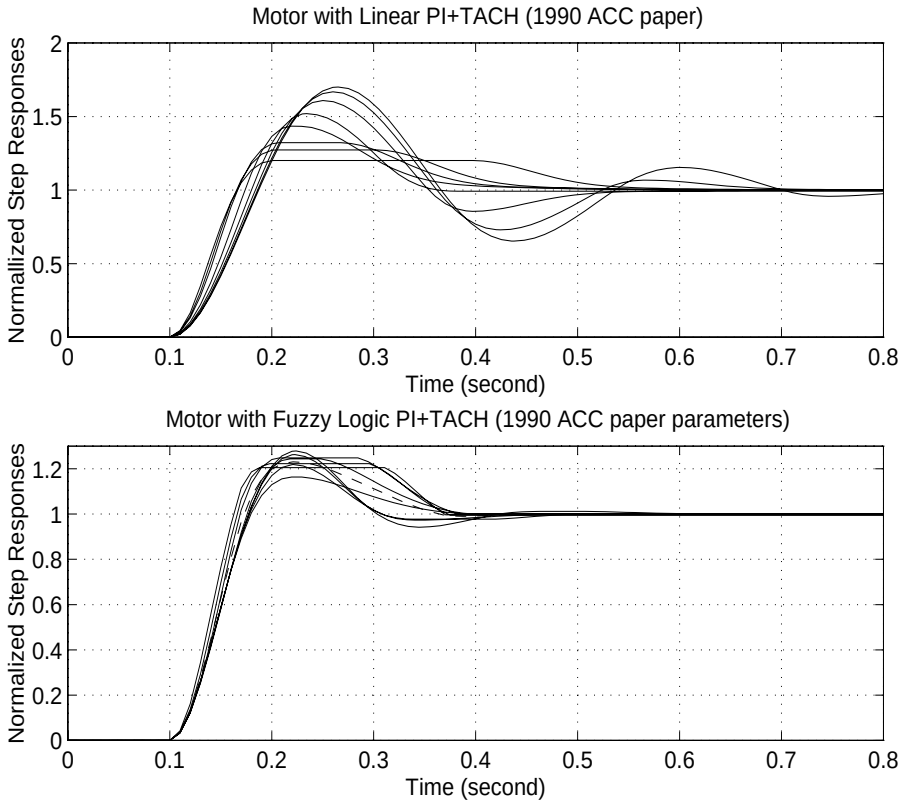

Figure 18: Linear and Nonlinear Controller Step Responses

achieves significantly better performance, both in the sense of lower overshoot and less settling time and in the sense of very low sensitivity of the response over the range of input amplitudes considered. The high overshoot in the linear case is caused by integral windup for large step commands, and the long settling for small step inputs is due to stiction. The nonlinear controller greatly alleviates these problems.

6.2 Design of autotuning linear and nonlinear controllers

Linear autotuning controllers – A clever procedure for the automatic tuning of proportional-integral-derivative (PID) regulators for linear or nearly linear plants has been introduced and used commercially [1]. It incorporates a very simple SIDF-based method to identify key parameters in the frequency response of a plant, to serve as the basis for automatically determining the parameters of a PID controller (a process called autotuning). It is based on performing system identification via relay-induced oscillations. The system is connected in a feedback loop with a known relay to produce a

limit cycle; frequency-domain information about the system dynamics is derived from the LC's amplitude and frequency. With an ideal relay, the oscillation will give the critical point where the Nyquist curve intersects the negative real axis. Other points on the Nyquist curve can be explored by adding hysteresis to the relay characteristic. Linear design methods based on knowledge of part of the Nyquist curve are called tuning rules; the Ziegler-Nichols rules [2] are the most familiar.

To appreciate how these key parameters are identified, refer to Fig. 7: clearly, if the process is seen to be in an LC, one can readily observe the amplitude and frequency of the oscillation. Since the relay height and hysteresis are known, along with the amplitude, $N_s(a)$ can be calculated (Eqn. 24), and the point $-1/N_s(a)$ establishes a point on the Nyquist plot of the process. On changing the hysteresis, the locus of $-1/N_s(a)$ will shift vertically, the LC amplitude and frequency will also change, and one can calculate another point on the Nyquist curve.

Basing identification on relay-induced oscillations has several advantages. An input signal that is nearly optimal for identification is generated automatically, and the experiment is safe in the sense that it is easy to control the amplitude of the oscillation by choosing the relay height accordingly.

The most basic tuning rules require only one point on $G(j\omega)$, so this procedure is fast and easy to implement. More elaborate tuning rules may require more points on $G(j\omega)$, but that presents no difficulty.

Nonlinear autotuning controllers – This same procedure can be extended to the case of nonlinear plants quite directly. The amplitude of the periodic signal forcing the plant is determined by the relay height D. Therefore, by selecting a number of values, D_i, one may identify points on the family of frequency response curves, $G(j\omega, b_i)$. Nonlinear controllers can then be synthesized from this information, using the methods outlined and illustrated in Sect. 6.1. This technique, and a sample application, are presented in [30].

7 Describing Function Methods: Concluding Remarks

Describing function methods all follow the formula "assume a signal form, choose an approximation criterion, evaluate the DF $N(a)$, use this as a quasilinear gain to replace the nonlinearity with a linear term, and solve the problem using linear systems theoretic machinery". We should keep in mind that in so doing we are deciding what type of phenomena we may investigate and thereby avoid the temptation to reach erroneous conclusions; for example, if for some values of (m, S) the RIDF matrix N_R (Eqn. 39)

has "eigenvalues" that are pure imaginary we must not jump to the conclusion that LCs may exist. We should also take care not to slip into linear-system thinking, and read too much into a DF result – for example, $X(j\omega)$ in Eqn. 6 is not, strictly speaking, the Fourier transform of an eigenvector.

The DF approach has proven to be immensely powerful and successful over the 50 years since its conception, especially in engineering applications. The primary reasons for this are:

1. Engineering applications usually are too large and/or too complicated to be amenable to exact solution methods.
2. The ability to apply linear systems theoretic machinery (e.g., the use of Nyquist plots to solve for LC conditions) alleviates much of the analytic burden associated with the analysis and design of nonlinear systems.
3. The behavior of DFs (the form of amplitude sensitivity, Fig. 2) is simple to grasp intuitively, so one can even use DFs in a qualitative manner without analysis.

The techniques and examples presented in this article are intended to demonstrate these points. This material represents a very limited exposure to a vast body of work. The reference books by Gelb & Vander Velde [8] and Atherton [3] detail the first half of this corpus (in chronological terms); subsequent work by colleagues and students of these pioneers plus that of others inspired by those contributions, has produced a body of literature that is massive and of great value to the engineering profession.

8 REFERENCES

[1] K. J. Åström and T. Hägglund, "Automatic Tuning of Simple Regulators for Phase and Amplitude Margin Specifications", *Proc.* IFAC *Workshop on Adaptive Systems for Control and Signal Processing,* San Francisco, CA, 1983.

[2] K. J. Åström and B. Wittenmark, *Adaptive Control,* Second Edition, Addison-Wesley Publishing Co., Reading MA, 1995.

[3] D. P. Atherton, *Nonlinear Control Engineering*, Van Nostrand Reinhold Co., London & New York; full edition 1975, student edition 1982.

[4] D. P. Atherton, *Stability of Nonlinear Systems.* Research Studies Press (Wiley), Chichester, 1981.

[5] D. P. Atherton and S. Spurgeon, "Nonlinear Control Systems, Analytical Methods", Electrical Engineering Encyclopedia, John Wiley & Sons, Inc., New York, 1999.

[6] Y. S. Cho and K. S. Narendra, "An Off-Axis Circle Criterion for the Stability of Feedback Systems with a Monotonic Nonlinearity" *IEEE Trans. on Automatic Control, AC-13,* No. 4, pp. 413-416, 1968.

[7] J. J. D'Azzo and C. H. Houpis, *Feedback Control System Analysis and Synthesis*, McGraw-Hill Book Co., New York, NY, 1960.

[8] A. Gelb and W. E. Vander Velde, *Multiple-Input Describing Functions and Nonlinear System Design*, McGraw-Hill Book Co., New York, NY, 1968.

[9] A. Gelb and R. S. Warren, "Direct Statistical Analysis of Nonlinear Systems: CADET", *AIAA Journal, Vol. 11,* No. 5, May 1973, pp. 689-694.

[10] J. E. Gibson, *Nonlinear Automatic Control*, McGraw-Hill Book Co., New York, NY, 1963.

[11] A. H. Jazwinski, *Stochastic Processes and Filtering Theory,* Academic Press, New York, 1970.

[12] D. N. Hannebrink, H. S. Lee, H. Weinstock, and J. K. Hedrick, "Influence of axle load, track gauge, and wheel profile on rail vehicle hunting", *Trans. ASME - J. Eng. Ind.* pp. 186-195, 1977.

[13] P. J. Holmes and J. E. Marsden, "Bifurcations to Divergence and Flutter in Flow Induced Oscillations - An Infinite Dimensional Analysis", *Automatica, Vol. 14,* pp. 367-384, 1978.

[14] I. E. Kazakov, "Generalization of the method of statistical linearization to multi-dimensional systems", *Avtomatika i Telemekhanika, 26,* pp. 1210-1215, 1965.

[15] T. W. Körner, *Fourier Analysis,* Cambridge University Press, Cambridge, 1988.

[16] M. Landau and C. T. Leondes, "Volterra Series Synthesis of Nonlinear Stochastic Tracking Systems", *Trans. on Aerospace and Electronics Systems, Vol. AES-11,* No. 2, pp. 245-265, March 1975.

[17] O. P. McNamara and D. P. Atherton, "Limit Cycle Prediction in Free Structured Nonlinear Systems", *Proc.* IFAC *World Congress*, Munich, Germany, 1987.

[18] K. S. Narendra and J. H. Taylor, *Frequency Domain Criteria for Absolute Stability*, Academic Press, New York, 1973.

[19] A. Papoulis, *Probability, Random Variables, and Stochastic Processes,* McGraw-Hill Book Co., New York, 1965.

[20] Popov, V. M., "Nouveaux Criteriums de Stabilité pour les Systemès Automatiques Non-Linéaries", *Revue d'Electrotechnique et d'Energetique, Acad. de la Populaire Romaine, 5*, No. 1, 1960.

[21] R. V. Ramnath, J. K. Hedrick, and H. M. Paynter (editors), *Nonlinear System Analysis and Synthesis: Vol 2 - Techniques and Applications*, ASME Press, Book G00178, New York (see chapters 7, 9, 13, 16), 1980.

[22] J. H. Taylor, "An Algorithmic State-Space/Describing Function Technique for Limit Cycle Analysis", TIM-612-1 for the Office of Naval Research, The Analytic Sciences Corporation (TASC), Reading, MA, October 1975; presented as "A New Algorithmic Limit Cycle Analysis Method for Multivariable Systems", IFAC Symposium on Multivariable Technological Systems, Fredericton, NB, Canada, July 1977.

[23] J. H. Taylor, "Comments on 'Volterra Series Synthesis of Nonlinear Stochastic Tracking Systems' ", IEEE *Trans. on Aerospace and Electronics Systems, Vol. AES-14,* No. 2, pp. 390-393, March 1978.

[24] J. H. Taylor, "A General Limit Cycle Analysis Method for Multivariable Systems", Engineering Foundation Conference in New Approaches to Nonlinear Problems in Dynamics, Asilomar, CA, December 1979; published as a chapter in *New Approaches to Nonlinear Problems in Dynamics,* Ed. by P. J. Holmes, SIAM (Society of Industrial and Applied Mathematics), pp. 521-529, 1980.

[25] J. H. Taylor, "Applications of a General Limit Cycle Analysis Method for Multi-Variable Systems", Chapter 9 in [21], pp. 143-159.

[26] J. H. Taylor, "Statistical Performance Analysis of Nonlinear Stochastic Systems by the Monte Carlo Method" (invited paper), *Trans. on Mathematics and Computers in Simulation, Vol. XXIII,* pp. 21-33, April 1981.

[27] J. H. Taylor, "A Systematic Nonlinear Controller Design Approach Based on Quasilinear System Models", *Proc. American Control Conference,* San Francisco, CA, pp. 141-145, June 1983.

[28] J. H. Taylor, C. F. Price, J. Siegel and A. Gelb, "Covariance Analysis of Nonlinear Stochastic Systems Via Statistical Linearization", Chapter 13 in [21], pp. 211-226.

[29] J. H. Taylor and K. L. Strobel, "Nonlinear Compensator Synthesis via Sinusoidal-Input Describing Functions," *Proc. American Control Conference*, Boston MA, pp. 1242-1247, June 1985.

[30] J. H. Taylor and K. J. Åström, "A Nonlinear PID Autotuning Algorithm", *Proc. American Control Conference,* pp. 2118-2123, Seattle, WA, 18-20 June 1986.

[31] J. H. Taylor and J. R. O'Donnell, "Synthesis of Nonlinear Controllers with Rate Feedback via SIDF Methods", *Proc. American Control Conference,* San Diego, CA, pp. 2217-2222, May 1990.

[32] J. H. Taylor and J. Lu, "Computer-Aided Control Engineering Environment for the Synthesis of Nonlinear Control Systems," *Proc. American Control Conference*, San Francisco, CA, pp. 2557-2561, June 1993.

[33] J. H. Taylor and B. H. Wilson, "A Frequency Domain Model-Order-Deduction Algorithm for Nonlinear Systems", *Proc. IEEE Conference on Control Applications,* Albany, NY, pp. 1053-1058, September 1995.

[34] J. H. Taylor and D. Kebede, "Modeling and Simulation of Hybrid Systems", *Proc. IEEE Conference on Decision and Control,* New Orleans, LA, pp. 2685-2687, December 1995. MATLAB-based software is available on the author's web site, www.ee.unb.ca/jtaylor (including documentation).

[35] J. H. Taylor and L. Sheng, "Fuzzy-Logic Controller Synthesis for Electro-mechanical Systems with Nonlinear Friction", *Proc. IEEE Conference on Control Applications,* Dearborn, Michigan, pp. 820-826, September 1996.

[36] J. H. Taylor and C. Chan, "MATLAB Tools for Linear and Nonlinear System Stability Theorem Implementation", *Proc. Sixth IEEE Conference on Control Applications,* Hartford CT, pp. 42-47, October 1997. MATLAB-based software is available on the author's web site, www.ee.unb.ca/jtaylor (including documentation).

[37] J. H. Taylor and L. Sheng, "Recursive Optimization Procedure for Fuzzy-Logic Controller Synthesis", *Proc. American Control Conference,* Philadelphia, PA, pp. 2286-2291, June 1998.

[38] Ya. Z. Tsypkin, "On the Determination of Steady-State Oscillations of On-Off Feedback Systems", *IRE Transactions on Circuit Theory, Vol. CT-9*, No. 3, September 1962; original citation: "Ob Ustoichivosti Periodicheckikh Rezhimov v Relejnykh Systemakh Avtomaticheskogo Regulirovanija", *Avtomatika i Telemekhanika, Vol. 14*, No. 5, 1953.

Contents

List of Figures