find nonlinear effects in the actuator or feedback transducer, or even both, whereas the controlled element and loop compensation might well be linear.

To permit reference to certain classes of separable nonlinear elements, it is appropriate to classify them according to type in some sense. The broadest distinction to be made is between *explicit* and *implicit* nonlinearities. In the former case, the nonlinearity output is explicitly determined in terms of the required input variables, whereas in the latter case the output is defined only in an implicit manner, as through the solution of an algebraic or differential equation. Among the explicit nonlinearities, the next division is between *static* and *dynamic* nonlinearities. In the former case the nonlinearity output depends only on the input function, whereas in the latter, the output also depends on some derivatives of the input function. Among the static nonlinearities, a further distinction is drawn between *single-valued* and *multiple-valued* nonlinearities. In the case of a static single-valued nonlinearity, the output is uniquely given in terms of the current value of the input, whereas more than one output may be possible for any given value of the input in the case of a static multiple-valued nonlinearity. The choice among the multiple values is made on the basis of the previous history of the input; thus such a nonlinearity is said to possess *memory*. One can imagine dynamic multiple-valued nonlinearities as well, but we shall not have occasion to refer to any such in this book. These are the major distinctions among nonlinearities from the point of view of the theory to be developed here. Other characteristics, such as *continuous* vs. *discontinuous*, are of little consequence here, but can be of supreme importance in other contexts.

An example of a static, single-valued, continuous, piecewise-linear nonlinearity is the *deadband gain*, or *threshold* characteristic (Fig. 1.1-2*a*). It could represent the acceleration input–voltage output relationship of a pendulous accelerometer, or the input-output characteristic of an analog angular position transducer. It is described by

$$y = \begin{cases} k(x-\delta) & \text{for } x \geq \delta \\ 0 & \text{for } -\delta \leq x < \delta \\ k(x+\delta) & \text{for } x < -\delta \end{cases} \tag{1.1-1}$$

where x and y denote the nonlinearity input and output, respectively. A static, multiple-valued, discontinuous, piecewise-linear nonlinearity is the *relay with deadband and hysteresis* (Fig. 1.1-2*b*). Arrows denote the direction in which this characteristic must be traversed in the determination of the output for a given input. The *history* of the input determines the value of the output in the multiple-valued regions. This characteristic is representative of the actuator switch in a temperature control system (in which case only the first quadrant portion applies) or the on-off gas jets in a spacecraft angular orientation system.

reason to expect that it will fit the actual solution efficiently. For example, if a system actually has a solution of the form

$$y(t) = A \sin \omega t \tag{1.3-1}$$

the assumed solution form

$$y_a(t) = a + bt + ct^2 + \cdots \tag{1.3-2}$$

cannot generate the solution for an interval of time comparable even with one period of the oscillation with a reasonable number of terms in the series.

More rapidly convergent expansions can be made if one can solve for the approximate response of the system and develop the solution in a series of functions which fit this response efficiently. If such a solution to a nonlinear-system problem is to be achieved, the leading term in the expansion must be the solution to a simpler problem which we are able to solve, and each succeeding term must be derivable from this in some tractable manner. If we confess that the only problems we are really able to solve are linear problems (this statement is intended to be a bit overgeneral), we must expect that the leading term in most useful series solutions will be the solution of a linear problem, and subsequent terms in the expansion will attempt to account for the nonlinear characteristics of the system. Such expansions can then be expected to converge rapidly only if the system is "slightly nonlinear," that is, if the system properties are describable to a good approximation as properties of a linear system. But this is not true of some of the simplest and most commonplace of nonlinear systems, such as a relay-controlled servo. Thus series methods, although they will continue to hold an important place in nonlinear-system theory, are almost certain to be restricted in applicability.

LINEARIZATION

The problem of studying a nonlinear system can be avoided altogether by simply replacing each nonlinear operation by an approximating linear operation and studying the resulting linear system. This allows one to say a great deal about the performance of the approximating system, but the relation of this to the performance of the actual system depends on the validity of the linearizing approximations. Linearization of nonlinear operations ordinarily can be justified only for *small departures* of the variables from nominal operating values. This is pictured in Fig. 1.3-1. Any response which carries variables through a range which exceeds the limits of reasonable linear approximation cannot be described using this technique unless the system is repeatedly relinearized about new operating points, and the resulting solutions patched together. In addition, some commonplace nonlinearities, among them the two-level switch, have a discontinuity at the point which should be system, but it will be a biased asymmetric mode. Or again, use of the single-sinusoid-input describing function may indicate that a system has two stable limit cycles, and one might expect to see either one, depending on initial conditions. In some cases, however, use of the bias-plus-sinusoid-input describing function would show that in the presence of one of the limit cycles the system has an unstable small-signal mode. Thus the system is unable to sustain that limit cycle. We conclude that the analysis is tailored to the evaluation of particular response characteristics. The burden of deciding what characteristics should be inquired into rests with the system designer.

Another difficulty which the user of describing function theory must be alert to is the possibility of *multiple solutions*. Formulation of a problem using describing functions results in a simultaneous set of nonlinear algebraic relations to be solved. More than one solution may exist. These solutions represent different possible modes of response, some of which in some cases may be shown to be unstable. But the characteristics of these different solutions may be quite different, and the designer could be badly misled if he did not inquire into the possibility of other solutions. As an illustration of this, the gaussian-plus-sinusoid-input describing function can be used to determine how much random noise must be injected into a system to quench a limit cycle. The equations defining the behavior of the system may have a solution for a zero limit cycle amplitude and a certain rms value of the noise. However, one cannot conclude from this that the calculated rms value of noise will quench the limit cycle until he has assured himself that there is not also another solution for the same rms noise and a nonzero limit cycle amplitude.

A final limitation on the use of describing function theory is the fact that there is no satisfactory evaluation of the *accuracy* of the method. Research into this problem on the part of quite a few workers has resulted in some intuitively based criteria which are rather crude and some analytically based procedures which are impractical to use. All we have, then, is the fact that a great deal of experience with describing function techniques has shown that they work very well in a wide variety of problems. Furthermore, in those cases in which the technique does not work well, it is almost always obvious that the linear part of the system is providing very little low-pass filtering of the nonlinearity output. Finally, since the design of a nonlinear system must be based on the use of approximate analytic techniques, and these techniques will be inadequate to answer all questions regarding system behavior, the design must be checked—preferably by computer simulation—before it is approved. At that point in the design process one need not concern himself with checking the accuracy of the approximate analytic tools he has used in arriving at the design. Rather, his object is to check the design itself, to assure himself of its satisfactory performance in a variety of outputs is

$$
\begin{aligned}
\overline{y(t)y_a(t+\tau)} &= \sum_{i=1}^{n}\int_0^\infty w_{oi}(\tau_1)\overline{y(t)x_i(t+\tau-\tau_1)}\,d\tau_1 \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n}\int_0^\infty d\tau_1\int_0^\infty d\tau_2\, w_{oi}(\tau_1)w_{oj}(\tau_2)\varphi_{ij}(\tau_1-\tau-\tau_2) \\
&= \overline{y_a(t)y_a(t-\tau)} \qquad \tau \le 0 \\
&= \overline{y_a(t)y_a(t+\tau)} \qquad \tau \ge 0
\end{aligned}
\tag{1.5-15}
$$

So the cross correlation between actual and approximate outputs is equal to the autocorrelation of the approximate output over the indicated range of τ. The restrictions on the range of τ for which Eqs. (1.5-14) and (1.5-15) hold are due to the restricted range of τ_1 for which Eq. (1.5-13) is a required condition.

The statistics of the approximation error can now be written

$$
\begin{aligned}
\overline{e(t)} &= \overline{y_a(t)} - \overline{y(t)} \\
&= \sum_{i=1}^{n}\int_0^\infty w_{oi}(\tau_1)\overline{x_i(t-\tau_1)}\,d\tau_1 - \overline{y(t)}
\end{aligned}
\tag{1.5-16}
$$

If the input to the nonlinearity, $x(t) = \sum_{i=1}^{n} x_i(t)$, has a nonzero mean value, there is no prescribed way of associating this constant with the various components $x_i(t)$. The assignment of the mean to the $x_i(t)$ can be made arbitrarily with no loss of generality. The most convenient convention to employ is to assign all of the mean of $x(t)$ to one of the $x_i(t)$ which is just a constant, or bias, function. With this convention, all but one of the $x_i(t)$ are unbiased functions; the remaining $x_i(t)$ is a constant function equal to the mean of $x(t)$. Equation (1.5-16) then becomes

$$
\overline{e(t)} = \overline{x(t)}\int_0^\infty w_{oi}(\tau_1)\,d\tau_1 - \overline{y(t)}
\tag{1.5-17}
$$

where $w_{oi}(\tau_1)$ is the weighting function of the filter which passes the bias input component. This weighting function has yet to be determined on the basis of minimizing the mean-squared approximation error; we shall find that the result also has the desirable property of reducing the mean error as expressed in Eq. (1.5-17) to zero.

The mean-squared approximation error, which is minimized by the set of filters that satisfy the conditions of Eq. (1.5-13), is

$$
\begin{aligned}
\overline{e(t)_o^2} &= \overline{y_a(t)^2} - 2\overline{y_a(t)y(t)} + \overline{y(t)^2} \\
&= \overline{y_a(t)^2} - 2\overline{y_a(t)^2} + \overline{y(t)^2} \\
&= \overline{y(t)^2} - \overline{y_a(t)^2}
\end{aligned}
\tag{1.5-18}
$$

The further change of variable $\theta'' = -\theta'$ was employed in this demonstration of the fact that Eq. (1.5-37), and thus the imaginary part of the describing function for a sinusoidal input to a static single-valued nonlinearity, is zero. For such a nonlinearity, the describing function for a sinusoidal input in the presence of any other uncorrelated inputs is a real static gain, the proportional gain of Eq. (1.5-36).

This gain is subject to an interesting interpretation.

$$N_A = \frac{1}{\pi A}\int_0^{2\pi}\left\{\int_{-\infty}^{\infty} dp_1 \cdots \int_{-\infty}^{\infty} dp_n f_n(p_1, \ldots, p_n) y[A \sin\theta + x_r(0)]\right\} \sin\theta\, d\theta \tag{1.5-40}$$

The definitions of all variables and functions used here are identical with those of Eq. (1.5-37). The quantity in the braces—the integral with respect to $p_1, \ldots, p_n$—is the expectation of the output of the nonlinearity with θ fixed. This expectation is itself a nonlinear function of θ which appears only in the form of $A \sin\theta$. Call this new function $y'(A \sin\theta)$. In terms of this, the describing function for a sinusoidal input component in the presence of any other independent input components is written

$$N_A = \frac{1}{\pi A}\int_0^{2\pi} y'(A \sin\theta) \sin\theta\, d\theta \tag{1.5-41}$$

But this is just an ordinary harmonic analysis of the modified nonlinearity $y'(x)$. This interpretation is also true of N_A for more general nonlinearities than static single-valued. In that case both the sine and cosine components of the output of the modified nonlinearity must be calculated, and this calculation is considerably more difficult. If the remainder includes a random process, the determination of the expectation of the nonlinearity output with θ fixed is somewhat obscure when the present output depends not only on the present input, but also on the history of that input. However, if the input consists only of a bias and any number of sinusoids, the calculation of the modified nonlinearity is straightforward. In summary, then, *the describing function for a sinusoidal input component may be viewed as the amplitude and phase relationship between an input sinusoid and the fundamental harmonic component of the expectation of the output of the nonlinearity taken with respect to all statistical parameters except* θ. This definition of the gain of a nonlinearity to a sinusoid in the presence of other input components was employed by Vander Velde (Ref. 14), based on an intuitive argument. This property has since been utilized for computational purposes by a number of writers, among them Atherton (Ref. 1), Gusev (Ref. 5), Popov (Ref. 10), and Somerville and Atherton (Ref. 13).

Random signal The last of the signal forms which we are considering is the gaussian process.

$$x(t) = r(t) + x_r(t) \tag{1.5-42}$$

$r(t)$ is a member of a stationary ensemble, and the remainder, $x_r(t)$, is uncorrelated with $r(t)$. For $x_i(t) = r(t)$, the autocorrelation function is simply written as

$$\varphi_{ii}(\tau) = \varphi_{rr}(\tau) \tag{1.5-43}$$

and the corresponding form of Eq. (1.5-21) as

$$\begin{aligned} \int_0^\infty w_R(\tau_2)\varphi_{rr}(\tau_1 - \tau_2)\, d\tau_2 &= \overline{y(0)r(-\tau_1)} \\ &= \varphi_{ry}(\tau_1) \qquad \tau_1 \geq 0 \end{aligned} \tag{1.5-44}$$

The right-hand member of this equation, defining the weighting function for the filter which passes a gaussian input component, is the cross correlation between the gaussian input component and the output of the nonlinearity. For a general nonlinearity, the output at time zero, $y(0)$, may depend not only on the current value of the input, but on certain properties of the past history of that input, or on certain derivatives of the input at time zero. The cross-correlation function in Eq. (1.5-44) is an average over all inputs to, and corresponding outputs from, the nonlinearity. The evaluation of this expectation requires the joint probability density function for all the random variables needed to define $y(0)$. Needless to say, this constitutes a formidable task even for dynamic nonlinearities of simple-appearing form.

Even if one is able to evaluate $\varphi_{ry}(\tau_1)$ in some cases, a substantial chore remains. The solution to the integral equation will not be obvious; the equation must be solved in the more general sense. This solution is not difficult if the transform of $\varphi_{ry}(\tau_1)$ can be taken and if the result is a rational function of the transform variable, or can be well approximated by a rational function. If so, the solution to the integral equation can be written down explicitly, since the equation is of the form of the Wiener-Hopf equation. The solution is derived in a number of books, including Refs. 4 and 7 to 9. If transform techniques cannot usefully be employed, the only practical alternative is likely to be numerical solution with computer help. The solution will be some general function for $w_R(\tau_2)$. Thus the optimum linear filter to approximate the effect of the general nonlinearity in passing a gaussian input component is not a static gain, but is indeed some dynamic linear filter, as one would surely expect.

Fortunately, this situation is simplified considerably in the very important case of a static single-valued nonlinearity. In this case the output of the nonlinearity depends only on the current value of the input, and the