Application of a Control-Law Nonlinearity Measure to the Chemical Reactor Analysis

Alexander J. Stack and Francis J. Doyle III
School of Chemical Engineering, Purdue University, West Lafayette, IN 47907

The optimal control structure has been introduced as a design tool to measure the control-law nonlinearity of a given process design. In this context, control-law nonlinearity is the optimal degree of nonlinear compensation in the controller, a system property distinct from open-loop nonlinearity and determined by a performance objective and the region of operation as well as the nature of the open-loop system. This approach is extended to the analysis of multivariable systems with output feedback through the application of an extended Kalman filter. Coherence estimation is used as a practical method to measure continuous, open-loop multivariable system nonlinearity. The CSTR with van de Vusse kinetics, a system that features output feedback and a control-law nonlinearity that changes with operating points, is analyzed. The optimal control structure approach with coherence estimation correctly indicates changes in the control-law linearity between different operating points and changes as the regions of operation change around a particular operating point.

Introduction

Process design engineers are being called upon to increase process integration in the face of increased economic demands. However, there is little existing art to bridge the gap between process design and control. At best, this engenders an iterative process, where design and control concerns alternate, and at worst leads to no consideration of control concerns by the design engineer. The key difficulty is that there is no practical, standard method to evaluate the operability of competing plant designs.

In Ogunnaike et al. (1993), an approach was proposed to classify systems based on three control-relevant criteria: (1) process nonlinearity; (2) the dynamic character of the process; and (3) the degree of multivariate interaction. These attributes form the axes of a three-dimensional (3-D) space defined as the process characterization cube. Different control strategies may be related to specific regions in the 3-D space, thus forming a mapping between the space of the process systems and appropriate control schemes. While appropriate measures for dynamic character and degree of interaction exist (for example, pole-zero location and relative gain array (RGA), respectively), a measure for the degree of nonlinearity remains an active area of research.

Correspondence concerning this article should be addressed to F. J. Doyle III.

Open-loop nonlinearity measures

Previous attempts to assess process nonlinearity have focused upon the development of open-loop nonlinearity measures, and have assumed, albeit implicitly, that the appropriate control law should fully compensate for, or equal, the open-loop nonlinearity. Some of the earliest work in this area is detailed by Haber (1985), in which he described ten computationally tractable methods for measuring nonlinearity from input–output records. In general, these methods are based on a quantification of the violation of properties that are known to hold for linear systems. An example for such a property is the coherence function. While coherence estimation has been used for some years in other fields (Carter, 1993), its application to chemical engineering studies has been limited to the detection of unmodeled disturbances in discrete time systems (Pearson and Ogunnaike, 1994a,b). Unlike most of the nonlinearity metrics described below, it provides frequency-dependent information. Moreover, as detailed in the third section, the estimation procedure is relatively straightforward. Coherence analysis alone (when applied to the open-loop process operator) can lead to inconsistent ranking of nonlinear systems, hence, more complicated methods have been proposed.

Abstract-operator-based methods have been proposed by

Nikolaou and Hanagandi (1993) and Ogunnaike et al. (1993). In the former, an inner product is defined, and a model's departure from an optimal linear approximation is assessed. Notably, the inner product can be iteratively computed to a desired standard deviation. However, it is difficult to relate the results to process fundamentals, performance objectives, or to changes with the region of operation. In the second method, gain change estimates are quantified over a region of operation. However, this method only measures static nonlinearity, making its use for assessing dynamic control systems unclear. Allgöwer and Gilles suggest two approaches, one relying on the invariance of linear systems under a Hilbert transform (Allgöwer, 1995), and the second on convex optimization (Allgöwer and Gilles, 1992). The latter involves the assessment of system nonlinearity over a given region of operation, and includes practical methods to estimate bounds on the measure. Another contribution to the area of nonlinearity characterization has been advanced by Guay et al. (1995). In this work they recognize two versions of nonlinearity: open-loop and control-law nonlinearity. Elegant scale-independent measures of directional and root-mean-square curvature are proposed, and specific control strategies to compensate for particular nonlinearities are presented. These results are limited to static nonlinearity, more recent extensions to dynamic nonlinearity have been proposed (Guay, 1996).

Closed-loop performance issues

In the context of nonlinearity measures it is of course also of interest to capture the influence that nonlinearities have with respect to achieving certain performance goals in the closed loop. Sometimes this is referred to as *control-relevant nonlinearity*. For instance, one issue that has received considerable attention in the recent literature is the determination of the appropriate application of nonlinear vs. linear control for a particular problem.

From the perspective of stabilization, the relevant question is: Does there exist a nonlinear (or linear) controller that guarantees a particular type of stability (e.g., input–output)? A nice review of work that addresses this question is provided by (Poolla et al., 1990), in which a more difficult problem is also addressed: *robust stabilization*. In an effort to determine whether time-varying nonlinear control is vastly superior, or only marginally superior to linear time-invariant control, they suggest the *plant uncertainty principle*. A very simplified summary of their findings is that parametric uncertainty favors the application of nonlinear control, while dynamic uncertainty favors the application of linear control.

The next level of analysis would address performance of a controlled system with a typical question being: Does a particular plant benefit significantly from the application of nonlinear (vs. linear) control to achieve a specific performance objective? For instance, it is well known that the optimal controller to a quadratic objective function for a linear known plant without constraints and subject to additive Gaussian white noise is a linear controller (Kwakernaak and Sivan, 1972). However, if any of these conditions are not satisfied (e.g., nonlinear plant, nonquadratic objective function, or saturation constraints), experience shows that in many cases a nonlinear controller will yield considerable improvement over a linear controller (Bernstein, 1993).

The issues of stabilization and closed-loop performance are clearly tied to the intrinsic characteristics of the particular process in question. Indeed, as demonstrated by J. Doyle et al. (1993), Morari and Zafiriou (1989), F. J. Doyle et al. (1991), and Shinskey (1988), there exist classes of nonlinear plants that are optimally and robustly controlled by a linear controller. Clearly, a simple analysis of the open-loop system does not include all the information needed to determine the control-relevant nonlinearlity of a system. In particular, *it does not include the effects of a performance objective or the costs of control action*. In this article, we describe the application of a nonlinearity measure that incorporates the region of operation, process description, *and* the closed-loop performance objective. The approach is demonstrated on a complex chemical reactor benchmark problem.

Simple motivating example

As described earlier, the open-loop nonlinearity does not correspond to the nonlinearity of the best control scheme for many systems. Consider the regulatory problem for the heat balance of a perfectly mixed batch reactor (F. J. Doyle et al., 1991; Morari and Zafiriou, 1989):

$$C_p \frac{dT}{dt} = (-\Delta H_r) r - UA(T - T_c)$$

$$r = k_0 e^{-E/RT}, \tag{1}$$

where the nomenclature is defined in Table 1.

The nonlinear reaction term r can be linearized as

$$r = r^0 + k_T (T - T_0), \tag{2}$$

where $k_T(T) = \partial r / \partial T$, which varies nonlinearly with temperature. Note that this is a local linearization, not a global approximation. This results in the following transfer function between T_c and T:

$$T(s) = \frac{UA/C}{s + a} T_c(s), \tag{3}$$

where

$$a = \frac{UA + \Delta H_r k_T}{C_p}. \tag{4}$$

The pole uncertainty in Eq. 4 arises from the term $k_T(T)$. Letting the possible values of a be represented by a norm-bounded perturbation, Δ_E, weighted by w_E gives

Table 1. Notation for the Perfectly Mixed Batch Reactor

T	Reactor temperature	(°C)
T_c	Coolant temperature	(°C)
r	Reaction rate	(mol/s)
ΔH_r	Heat of reaction	(J/mol)
C_p	Heat capacity of fluid in reactor	(J/°C)
UA	Overall heat transfer coefficient	(J/s°C)
E/R	Energy of activation/gas constant	(°C)
k_0	Preexponential term	(mol/s)

$$a = \tilde{a}(1 + r_a \Delta_E), \qquad |\Delta_E| \le 1, \qquad \Delta_E \text{ real}, \tag{5}$$

where $\tilde{a}$ is the nominal pole location and r_a is the relative uncertainty of a. This uncertainty may be expressed as an inverse multiplicative perturbation $(I + w_E \Delta_E)^{-1}$, so that the family of uncertain plants is given by:

$$\frac{1}{s+a} = \frac{1}{s+\tilde{a}} \cdot \frac{1}{1 + w_E \Delta_E}, \tag{6}$$

where

$$w_E(s) = \frac{r_a}{1 + s/\tilde{a}}. \tag{8}$$

Similarly, neglected high-frequency dynamics can be modeled with an output multiplication uncertainty with perturbation Δ_O and weighting w_O. Using these uncertainty models, robust stability is guaranteed if and only if

$$|w_E \tilde{\epsilon}| + |w_O \tilde{\eta}| < 1. \tag{9}$$

Ordinarily, there is a trade-off between robust stability and performance, since $\tilde{\epsilon} + \tilde{\eta} = 1$. The sensitivity function, $\tilde{\epsilon}$, reflects the effect of disturbances on the plant output, with smaller values corresponding to better disturbance rejection. The complementary sensitivity function, $\tilde{\eta}$, relates the setpoint to the output, with $\tilde{\eta} = 1$ representing perfect setpoint tracking. Usually, good setpoint tracking and disturbance rejection ($\tilde{\epsilon} = 0$, $\tilde{\eta} = 1$) are limited by an upper bound on $\tilde{\eta}$ imposed by model uncertainty (Morari and Zafiriou, 1989). In other words, the permissible values of w_E are constrained by the values of w_O at a given frequency. However, since it is assumed in this particular uncertainty formulation that there are no neglected high-frequency dynamics, $w_O = 0$, and there is no effective upper constraint on $\tilde{\eta}$. The robust stability criterion becomes

$$|w_E \tilde{\epsilon}| < 1, \tag{10}$$

and, since the plant is minimum phase, may be satisfied for any fixed value of w_E by making the sensitivity function $\tilde{\epsilon}$ small, equivalent to using linear control with arbitrarily large gain.

Consider the implications of this example. An approach based on analyzing the open-loop plant operator will determine that the system is nonlinear. However, the system is robustly and optimally controlled by a linear controller. This motivates the definition of control-law nonlinearity as a property separate and distinct from open-loop nonlinearity (Stack and Doyle III, 1995). Under this definition, control-law nonlinearity is determined by three factors: the open-loop system, the performance objective, and the scaling or region of operation of the system inputs and outputs. In (Stack and Doyle III 1995), analysis of the optimal control structure (OCS) was introduced for the assessment of control-law nonlinearity. The batch reactor described earlier was analyzed using the OCS approach in (Stack, 1995), with results that are in agreement with the analysis given before. The previously cited application of the OCS approach was limited to single-input, single-output (SISO) state-feedback systems. In this article, it is extended to a richer and more practical case—the continuous stirred-tank reactor (CSTR) with van de Vusse kinetics, with multivariate and output feedback features. The van de Vusse reaction scheme, consisting of competitive series and parallel reactions, carried out in a CSTR has been studied by many researchers (Chen et al., 1995; Engell and Klatt, 1993a,b; Ogunnaike and Ray, 1994; van de Vusse, 1964), and has become an unofficial benchmark (Chen et al., 1995) for nonlinear control studies. A model of the production of cyclopentanol from cyclopentadiene, it exhibits a wide range of nonlinear dynamic behavior, and over the range of operating conditions fits into six of the eight categories described earlier for the process characterization cube (Ogunnaike and Ray, 1994). It is analyzed for both the state- and output-feedback formulations below in the subsection titled "CSTR with van de Vusse case study."

OCS

The approach to control-relevant nonlinearity characterization employed in this work is based on the idea that such a quantity must reflect the degree of nonlinearity in some appropriate control structure. Of course, this requires some *a priori* specification of a particular control structure (e.g., internal model control, proportional integral derivative (PID), model predictive control) that may in fact lead to a biased computation of the control-relevant nonlinearity. In particular, the chosen control structure may lack a sufficiently rich behavioral character that would be required to express appropriate control action to achieve a particular performance objective. Consequently, we have chosen in this article the classical nonlinear optimal control structure (Ray, 1990) as a basis for control design. Such a choice imposes minimal *a priori* assumptions about the feedback architecture other than the existence of a state-space process model and the availability of state measurements (for feedback).

The procedure used is based upon Pontryagin's *maximum principle* (Pontryagin et al., 1962). The ideas behind this procedure as well as the closely related dynamic programming have been used in the chemical engineering community since before the 1960s; for a detailed treatment see, for example Aris (1961), Roberts (1964), and Lapidus and Luus (1967). It is currently used in fields as diverse as process design (Seider et al., 1991), operation (Tieu et al., 1994), and scheduling and planning (Bhalla and Bower, 1993). Within the process-control field, these ideas underly the nonlinear model predictive control (MPC) approach (Li and Biegler, 1988; García et al., 1989; Rawlings et al., 1994), where the optimal control action minimizing an error trajectory to a given time horizon is recalculated and implemented every step.

Analytical derivation of an optimal controller is a two-step process. First, the structure of the controller is derived using Lagrangian optimization, as detailed in the following theorem.

Theorem 1: Weak Maximum Principle (Ray, 1990). Consider the objective function I [Eq. 12, consisting of dynamic ($\boldsymbol{F}(\boldsymbol{x}, \boldsymbol{U})$) and endpoint ($\boldsymbol{G}(\boldsymbol{x}(t_f))$] conditions, subject to the constraints in Eq. 11:

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}), \qquad \boldsymbol{x}(0) = \boldsymbol{x}_0 \tag{11}$$

$$I[u(t)] = G(x(t_f)) + \int_0^{t_f} F(x,u)\, dt. \tag{12}$$

The control $\bar{u}$ maximizes the objective I only if Eq. 13 is true for unconstrained portions of the path, where the Hamiltonian H is defined by Eq. 14 and is maximized along the constrained portions of the path:

$$\frac{\partial H}{\partial u} = 0 \tag{13}$$

$$H = F(x,u) + \lambda^T f(x,u). \tag{14}$$

The time-dependent Lagrange or adjoint multipliers λ are defined by

$$\frac{d\lambda^T}{dt} = -\frac{\partial H}{\partial x}. \tag{15}$$

Equation 16 holds for state variables unspecified at $t = t_f$:

$$\lambda_i(t_f) = \frac{\partial G}{\partial x_i}. \tag{16}$$

It is also necessary that $H(t)$ remain constant along the optimal trajectory, and that $H(t)$ take a constant value of zero if the final time t_f is unspecified. ■

These results provide necessary but not sufficient conditions for optimality. Stronger results can be obtained with the strong maximum principle, which is sufficient under certain convexity conditions (Ray, 1990). This yields a structure and control settings at the final time. However, the initial control law settings must be obtained by solving a two-point boundary-value problem (TPBVP). Many numerical methods exist to approximate these values (Lewis, 1992; Press et al., 1992; Tierno et al., 1995). Hicks and Ray (1971), as well as Allgöwer (1995), avoid this step by synthesizing optimal controllers, and recommend approximate solution methods involving the optimization of parameterizations of the control law.

In a similar spirit, we seek an approximation to the optimal control formulation for the purposes of nonlinearity measurement. While the general optimal controller yields a solution with minimal presupposition about control architecture, it does not admit a convenient causal operator for nonlinearity analysis. In particular, the optimal control solution corresponds to one specific trajectory that requires the TPBVP solution to be evaluated in a closed-loop with the plant operator. Consequently, we extract the mathematical *structure* that underlies the optimal control solution, and examine the response of this operator to a family of input signals. The appropriate definition is as follows.

Definition 1: Optimal Control Structure. The OCS, denoted $\mathfrak{N}$, derived using Lagrangian optimization and a linear quadratic performance objective, is given by the differential equations describing the time-dependent multipliers and the algebraic equation for the partial derivatives of the Hamiltonian with respect to u:

$$\dot{\lambda}^T = -x^T - \lambda^T \left(\frac{\partial f}{\partial x}\right) \tag{17}$$

$$u^T\alpha = \lambda^T \left(\frac{\partial f}{\partial u}\right). \tag{18}$$

In this way, we aim to approximate the nonlinear behavior of the optimal control solution over an operating regime. However, it is important to note that such an analysis will, in general, yield an upper bound on the nonlinearity measure. This is due to the fact that through an analysis of the region of operation many possible trajectories are considered (in addition to the one "optimal" solution), and consequently solutions with higher nonlinear character may influence the calculation.

Analytical determination of the OCS with a quadratic performance objective

The version of the OCS specifically used in this article is based on a quadratic performance objective, and is derived here. Assume the following general state-space model of the system dynamics:

$$\dot{x} = f(x,u) \tag{19}$$

and the following quadratic performance objective (x and u are deviation variables):

$$\min_u \frac{1}{2} \int_0^{t_f} (x^T x + u^T \alpha u)\, dt, \tag{20}$$

where $\alpha = \mathrm{diag}(\alpha_1 \cdots \alpha_m)$ is a matrix of cost factors. Using Theorem 1, the performance index is minimized by u, which solves the following sequence of equations:

$$\dot{\lambda}^T = -x^T - \lambda^T \left(\frac{\partial f}{\partial x}\right) \tag{21}$$

$$u^T\alpha = \lambda^T \left(\frac{\partial f}{\partial u}\right) \tag{22}$$

$$\lambda_i(t_f) = 0. \tag{23}$$

Remark 1. Note that for use of representation, we have chosen the standard quadratic objective function for this derivation. The approach extends in a straightforward manner to a much broader class of problems. A nice theoretical treatment of this operator formulation for more general problems is given in Rouff and Lamnabhi-Lagarrigue (1986).

By Definition 1, Eqs. 21 and 22 comprise the OCS. This formulation involves no assumptions about the control-law structure that would bias the results beyond the assumption of a static-state feedback controller. Analyzing the nonlinearity of the relationship between inputs x and outputs u, by any open-loop method that evaluates the effect of scaling should give an assessment of the control-law nonlinearity of the system and performance objective pairing. Note that the control variable and states for the original state-space model become the output and input for the OCS.

As mentioned earlier, the optimal control design requires the solution of at TPBVP to determine the initial point. Similar to MPC, stability is guaranteed by solving for the initial point from the assumed endpoint solution (Rawlings et al.,

1994). However, the OCS approach is concerned with the nonlinear dynamical structure of the controller, not the particular control action specified by unique end- and initial points. Hence, the OCS (independent of initial and endpoint conditions) facilitates computationally tractable analysis of the control-law nonlinearity.

It is important to note that the resulting nonlinear dynamical system of the OCS is no longer guaranteed to be stable. If, in particular, the open-loop controller is completely unstable (i.e., all of the poles associated with the Jacobian linearization lie in the right-half complex plane), it is possible to use a "reverse-time" stable system approximation. In effect, the substitution $t = -t$ is made into the original nonlinear differential equations. The pole locations are reflected across the imaginary axis to their corresponding position in the opposite half-plane. A similar approach is used in F. J. Doyle et al. (1992) to derive an optimal minimum-phase approximation to a maximum-phase nonlinear plant. As shown in that reference the resulting "time-reversed" approximation retains the essential nonlinear character of the original plant. In the event that the Jacobian of the OCS structure has poles in both the left- and right-half complex plane, then the coherence analysis may require modification for the resulting nonstationary system behavior. This issue is discussed further in the subsection titled "Nonstationary data" below.

Introduction to Coherence Measurements

The magnitude-squared coherence (also referred to as simply "coherence") between stationary random processes $u(t)$ and $x(t)$ is given by (Bendat, 1990; Carter, 1993):

$$\gamma_{ux}^2(f) = \frac{|S_{ux}(f)|^2}{S_{uu}(f)S_{xx}(f)}, \tag{24}$$

where $S(f)$ are the two-sided cross-spectral density functions defined by

$$S_{yw}(f) = \frac{1}{T}E[Y^*(f)W(f)], \qquad -\infty < f < \infty. \tag{25}$$

Consequently the coherence can be interpreted as the fraction of the mean square value of the output, u, that can be accounted for by a linear relationship with the input x, at a given frequency. Conveniently, the coherence, $\gamma_{ux}^2(f)$, is bounded by zero and one, and equal to unity at frequencies where the relationship between $u(t)$ and $x(t)$ is linear. Coherence suppression may actually come from a number of sources: noise in the input and output measurements, unmeasured disturbances, nonlinear relationships, between $u(t)$ and $x(t)$, and, if the coherence is estimated and not derived, errors from the coherence estimation process. In the framework considered in this article, exact values for the input and output and a complete system model will be known. This leaves a nonlinear relationship between the input and output, along with coherence estimation errors, as the causes for coherence suppression.

For purposes of illustration, consider some general trends in the coherences of particular functions and models. As low-order odd nonlinear functions can be more closely approximated by linear functions than all even functions, there is a difference between the coherences of even and odd nonlinearities. Figure 1 shows the coherence for a system of the form $\dot{x} = -u^n - 2x$ for varying values of n. Note that the frequency scale is normalized to the Nyquist frequency. The linear function ($n = 1$) has a coherence of one for all frequencies. As expected, the coherence decreases for odd nonlinearities of higher order ($n = 3, 5$). However, all even nonlinearities ($n = 2, 4$), including the lowest order even nonlinearity ($n = 2$), have a coherence near zero.

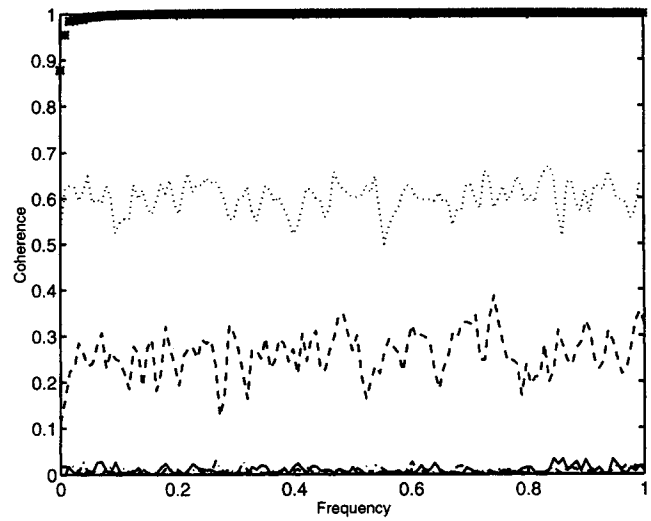

Figure 1. Coherence analysis for $\dot{x} = -u^n - 2x$, $n = 1$ (*'s), $n = 2$ (solid), $n = 3$ (dot), $n = 4$ (dash-dot), and $n = 5$ (dash).

$N = 16{,}384$; $\sigma_x^2 = 6.25$; $t_s = 0.05$; $h = 0.01$; $S = 256$.

Hammerstein and Wiener models are often used to model physical systems (for chemical engineering examples, see Eskinat et al. (1991), Ouarti and Edgar (1993), Zhu and Seborg, (1994)). As shown by Pearson and Ogunnaike (1994a), Hammerstein systems have coherences that are invariant with frequency, while Wiener models have coherence functions that decrease at higher frequencies. Figure 2 shows coherence results for a Hammerstein model ($\dot{x} = -u^3 - 2x$, $y = x$) and a Wiener model ($\dot{x} = -u - 2x$, $y = -3x/(10 + 8x)$). These results suggest that if the OCS of a particular system resembles one of the curves in this figure (i.e., flat-frequency dependence, or characteristic roll-off), then it may be appropriately regulated with a controller that has the corresponding Hammerstein or Wiener structure.

Numerical calculation

Data for use in the coherence estimates in this article were generated using a Euler integrator on MATLAB. The input data used are white-noise signals, with variances set to account for the effects of scaling, as described in the subsection titled "Region of operation (scaling)." The actual signal generators are not perfect white-noise sources, and in particular the power spectral density falls off as the frequency approaches zero. As a result, the coherence estimates in this work also decrease as the frequency approaches zero, regardless of the nonlinearity of the system being tested.

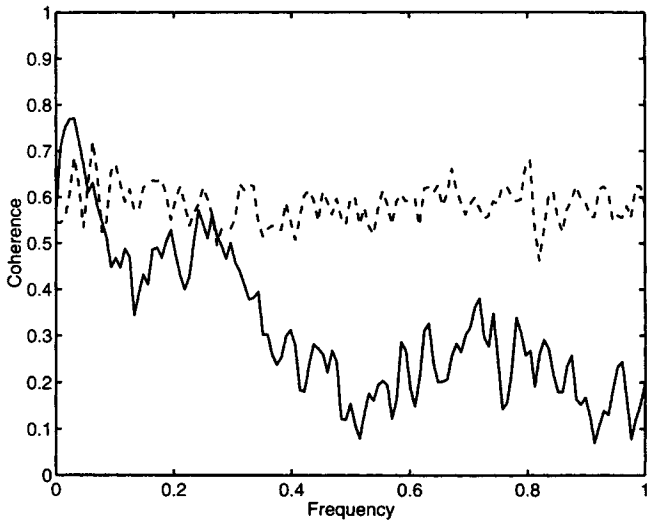

Figure 2. Coherence analysis for Wiener system (solid) and Hammerstein system (dash).

$N = 16{,}384$; $\sigma_x^2 = 6.25$; $t_s = 0.05$; $h = 0.01$, $S = 256$.

Following the weighted overlapped segment averaging (WOSA) method (Carter, 1993), the input and output data are divided into segments, with a 50% overlap between adjacent segments, which are detrended. For example, for 32-point segments, the data would be grouped as points 1–32, 17–48, 33–64, and so forth. Each segment is then multiplied by a Hanning weighting function (Little and Shure, 1992), and a fast Fourier transform (FFT) is performed on each segment. The magnitude-squared values are time averaged over each frequency bin to obtain the power spectral densities of the input and output [$S_{uu}(f)$ and $S_{xx}(f)$], while the products are averaged to form the cross-spectral density $S_{ux}(f)$. The coherence is then calculated as in Eq. 24.

The coherence analysis performed in this work will preferably use segments of 256 points to achieve resolution. However, if the data are nonstationary, as described in the subsection on nonstationary data, a segment of a differing length is preferable. The sampling and integration rates must be chosen judiciously. While it is important to sample fast enough to include all relevant frequencies, sampling at too high a rate can obscure low-frequency information. There should be a number of integrations per sampling period to generate accurate simulations. However, the number of feasible calculations are limited by the available computing power.

Variance, bias, and confidence intervals of the coherence estimate

The variance, bias, and confidence intervals of the magnitude-squared coherence estimate depend on two factors: the number of FFT segments used, and the true coherence value (Carter, 1993; Carter et al., 1983). Not surprisingly, the variance and bias of the estimate decrease as the number of segments is increased. The probability density functions for true coherences on the order of 0.3 to 0.6 are Gaussian in form, and are more peaked outside this region. The maximum variance occurs at a true coherence of 0.3. The maximum bias is the reciprocal of the number of FFT segments averaged, and occurs at $\gamma^2 = 0$.

While multiplication by a Hanning weighting function is necessary to reduce side lobes and increase the peak resolution, it also causes data to be lost at both ends of the data segment. Thus, it is desirable to overlap the data windows so the data are used in an efficient manner. The effect of varying the percent overlap on the variance and bias of the estimate has been investigated in (Carter et al., 1983), and diminishing returns are seen for overlaps of greater than 50%.

The standard deviations of the estimates depend upon the number of segments used as well as the true coherence value. In all cases, the coherence plots is this work use 8,192 points divided as described before into 256- or 32-point windows, resulting in the use of 63 and 511 segments, respectively. While calculating confidence intervals is not a straightforward exercise, graphs of confidence intervals for various estimates may be found in Carter (1993), and using the empirical results from Benignus (1969), the 95% confidence intervals for the graphs employing 32-point windows should be less than ± 0.04, while for the 256-point windows they may be less than ± 0.1.

Nonstationary data

The WOSA method of coherence estimation is strictly limited to stationary time series. However, in general physical, nonlinear systems will generate nonstationary outputs (White and Boashash, 1990). Despite this, nonstationary systems can be analyzed by the preceding methods if it can be assumed that the signals are quasi stationary over the length of the windowed segments (Flandrin and Escudié, 1984; Oppenheim, 1970). If $\{s(t)\}$ is a sequence such that the following limits exist (Ljung, 1987):

$$\text{(i)} \quad E[s(t)] = m_s(t) \qquad |m_s(t)| \le C \qquad \forall t$$

$$\text{(ii)} \quad E[s(t)s(r)] = R_s(t,r) \qquad |R_s(t,r)| \le C, \tag{26}$$

then $\{s(t)\}$ is quasi stationary, where E is the expectation operator. These conditions imply that the mean and covariance function of the sequence should be bounded by C. This is roughly equivalent to stipulating that the system must have a bounded output.

However, this definition of quasi stationarity is very general, and may be met for any bounded output by a sufficiently large choice of C. Thus, it does not suggest a practical measure for stationarity. Instead, run tests may be used to assess the stationarity of three signal attributes: mean value, mean square value, and frequency structure, as measured by the half-period time interval or time between successive baseline crossings (Bendat and Piersol, 1986; Cohen and Sances, 1977). The run test (Walpole and Myers, 1989) is a nonparametric procedure that tests for evidence that a sequence is nonrandom. The observed values are separated into two exclusive categories, and the number of runs or changes in category are determined. This number, along with the actual number of observations that fall into each category, may be compared to a table of P-values to either accept or reject the hypothesis of randomness at a chosen level of significance. Note that when the number of observations in each category is greater than or equal to ten, the sampling distribution approaches a normal distribution, and a Z-test may be used. As an example (Walpole and Myers, 1989), test the hypothesis

that the numbers 3.6, 3.9, 4.1, 3.6, 3.8, 3.7, 3.4, 4.0, 3.8, 4.1, 3.9, 4.0, 3.8, 4.2, 4.1 are random. The mean for the series is 3.9. Letting the numbers be replaced by a "+" if the number is greater than 3.9, "−" if the number is less than the mean, and omitted if it equals the mean, leads to the sequence − + − − − − + + + + − + +, which contains 6 "runs" of 7 plusses and 6 minuses. Two times the corresponding P-value is 0.592, which is greater than 0.05. As a result, the hypothesis is not rejected.

The output data are separated into segments of length S, and run tests are performed on the three aspects for increasing values of S. The largest value of S that does not fail the stationarity test may be used as the window length in the coherence estimation procedure. The run tests were performed at a 95% significance level, and windows of length S were determined to be quasi stationary over short lengths if 80% of the segments passed all three run tests. By comparison, Cohen et al. (1977) accept as stationary window lengths with pass rates for half-periods of 65%, so there is reason to believe that the criteria used in this work are conservative.

Figure 3 shows that a linear unstable system ($\dot{x} = 0.01x - u$) with bounded output can be successfully analyzed with a window length of 256 points. Note that this figure introduces a consistent three-graph pattern that is employed for all of the subsequent coherence plots: a coherence graph is located above its corresponding input and output graphs. The system $\dot{x} = 0.01x - u^3$ has an output that is nonstationary for segment lengths of 256, but is quasi stationary for segments of length 32. Note that using a shorter segment length results in a loss of resolution in the coherence analysis, as shown in Figure 4.

While this approach has been successfully used in practice (Cohen and Sances, 1977; Oppenheim, 1970), it does suffer drawbacks, as outlined in (Kodera et al., 1978). An alternative approach, not followed in this work, would use a true time-frequency representation such as the Wigner-Ville distribution to assess the changes in the spectra with frequency and time (Flandrin and Escudié, 1984; Flandrin, 1986; White and Boashash, 1990). This would remove the requirement of a quasi-stationary signal.

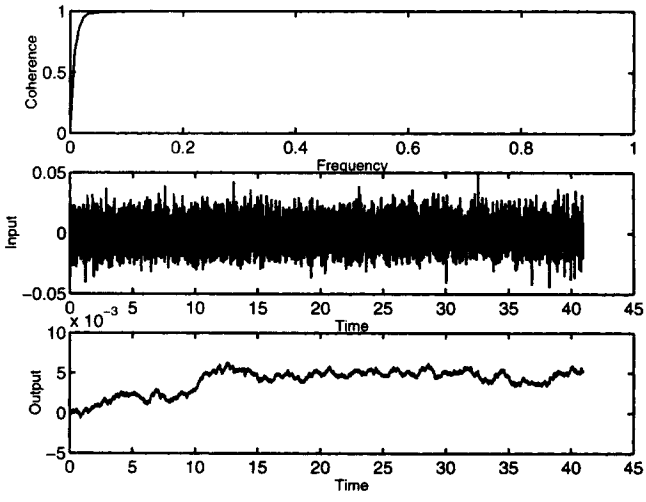

Figure 3. Coherence analysis for system $\dot{x} = 0.01x - u$.

$N = 8{,}192$; $\sigma_x^2 = 1.2e-4$; $t_s = 0.005$; $h = 0.001$; $S = 256$.

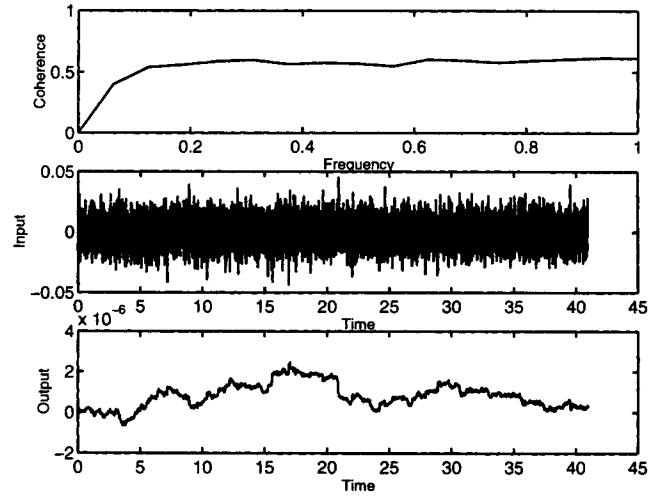

Figure 4. Coherence analysis for system $\dot{x} = 0.01x - u^3$.

$N = 8{,}192$; $\sigma_x^2 = 1.2e-4$; $t_s = 0.005$; $h = 0.001$; $S = 32$.

Multivariate coherences

Coherence analysis is defined for one input signal and one output signal. Obviously, with a multivariable system, a number of coherences can be calculated. If there is no noise added to the output, and the inputs are uncorrelated, the coherences for all inputs may be summed for each output to determine a system nonlinearity (Bendat and Piersol, 1986, 1993), called the cumulative coherence.

Assume that we have a system with two inputs and one output, y. Then $G_{yy} = G_{nn} + G_{vv}$, where G_{yy} is the autospectral density of the output, and G_{vv} and G_{nn} are signal densities arising from linear and nonlinear relationships with the two inputs. This can be rewritten as

$$G_{nn} = G_{yy} - \left|\frac{G_{1y}}{G_{11}}\right|^2 G_{11} - \left(\frac{G_{1y}}{G_{11}}\right)^* \frac{G_{2y}}{G_{22}} G_{12} - \left(\frac{G_{2y}}{G_{22}}\right)^* \frac{G_{1y}}{G_{11}} G_{21} - \left|\frac{G_{2y}}{G_{22}}\right|^2 G_{22}, \quad (27)$$

where 1 and 2 in the subscripts represent the first and second inputs. Since the inputs are uncorrelated, $G_{12} = G_{21} = 0$, and Eq. 27 becomes

$$G_{nn} = G_{yy} - \left|\frac{G_{1y}}{G_{11}}\right|^2 G_{11} - \left|\frac{G_{2y}}{G_{22}}\right|^2 G_{22} = \left[1 - \gamma_{1y}^2 - \gamma_{2y}^2\right] G_{yy}, \quad (28)$$

and $\gamma_{1y}^2 + \gamma_{2y}^2$ is clearly a measure of the output's overall linearity with respect to the two inputs. Note that if the individual variables are scaled with respect to expected fluctuations (Skogestad and Postlethwaite, 1996), then the individual coherences for each input, γ_{iy}, reflect the fraction of energy accounted for in the output by a linear relationship with the input i. Consider an application of multivariate coherence analysis to the OCS problem. A relatively small coherence in one OCS input channel (i.e., process output) in the case where the cumulative coherence is nearly one indicates that that

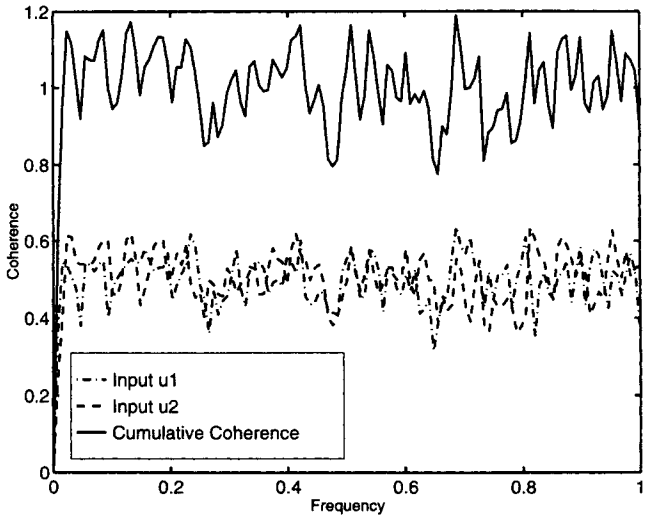

Figure 5. Coherence analysis for the system $\dot{x} = -x + u_1 + u_2$.

$N = 8{,}192$; $\sigma_x^2 = 1$; $t_s = 0.005$; $h = 0.001$; $S = 256$.

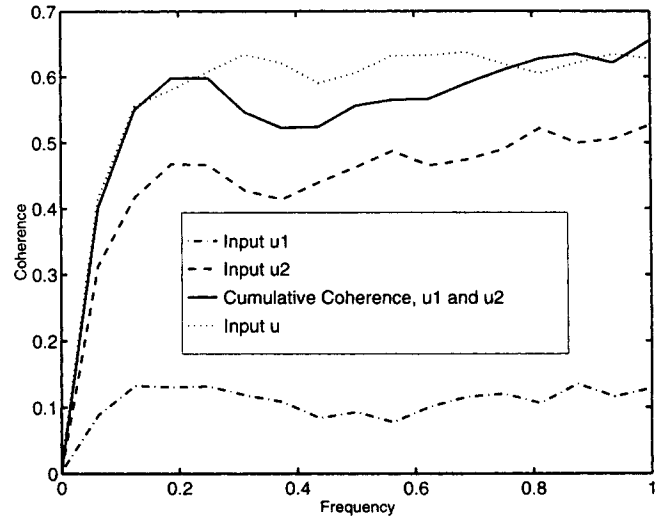

Figure 7. Coherence analysis for systems $\dot{x} = 0.01x - u_1^3 - 2u_2^3$ and $\dot{x} = -0.01x + u^3$.

$N = 8{,}192$; $\sigma_x^2 = 1$; $t_s = 0.005$; $h = 0.001$; $S = 32$.

channel has a relatively low gain with respect to the OCS output (i.e., process input) *for that chosen scaling.*

For example, Figures 5 and 6 show coherences calculated for the system $\dot{x} = -x + u_1 + nu_2$ for $n = 1$, and 2.5, respectively. Note that the frequency scale is normalized to the Nyquist frequency. Although inputs u_1 and u_2 each have a linear relationship to $\dot{x}$ in both figures, the presence of the other signal impedes the analysis of nonlinearity for the individual channels. However, the two signals sum to one (within the bounds of estimation error), indicating that the output is fully explained by linear relationships with the two inputs. Note the effect of varying the relative gains of the two inputs upon the coherences. The coherences between the inputs and the output are suppressed and raised in tandem with the values of n. This allows this cumulative coherence test to account not only for the system nonlinearity, but the relative weighting of two channels with respect to system dynamics.

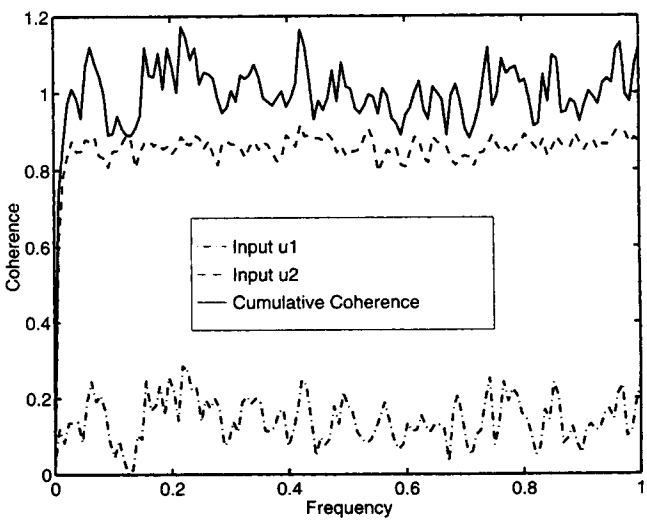

Figure 6. Coherence analysis for the system $\dot{x} = -x + u_1 + 2.5u_2$.

$N = 8{,}192$; $\sigma_x^2 = 1$; $t_s = 0.005$; $h = 0.001$; $S = 256$.

This also holds true for systems with suppressed coherences, as shown in Figure 7. The sum of the two coherences from $\dot{x} = 0.01x - u_1^3 - 2u_2^3$ is equal to the coherence of $\dot{x} = -0.01x + u^3$ within the bounds of estimation error. Note also that the multiple-input single-output (MISO) system is unstable, demonstrating that instability does not defeat coherence analysis if an appropriate segment length (to account for nonstationarity) can be chosen.

Remark 2. Note that the application of coherence analysis to the OCS implies a general analysis of the input–output properties of this operator. As discussed earlier, an optimal solution is the single solution corresponding to the particular TPBVP. Hence, the result proposed in this work involves a "probing" of many possible initial conditions. It is worth noting that although there is a single unique optimal trajectory, under conditions of analyticity of the functions in Eqs. 11 and 12, the optimal solution holds for a neighborhood of initial conditions (Al'Brekht, 1961).

Issues in OCS Formulation

Region of operation (scaling)

The region of operation of the process is one of the three critical factors that determine the control-law nonlinearity of a given system. This consideration is central to the question of controller nonlinearity, as a control scheme that might be feasible in one operating regime may not be feasible in another. Furthermore, the breadth of an operating regime will often correlate strongly with the appropriate controller design. Variable scaling is also of critical importance in a number of other systems tasks, such as process identification and process optimization. For example, Guay et al. (1995) elegantly and rigorously scale the process to an ellipsoidal state-space region.

In the context of the OCS approach, control-input-variable (process-output) scaling can be addressed directly, while control-output-variable (process-input) scaling is handled in an indirect manner. The input to the OCS represents the plant measurements, and in the computation of the coher-

ence for the OCS operator, arbitrary selection of the variance for this signal is possible. Thus control input variable scaling is directly addressed by its two-norm. For control outputs (the manipulated variables for the process), the problem is somewhat more complicated. An iterative procedure was used in this work to determine bounds on the control inputs that admit control outputs that are within suitable bounds (e.g., saturation constraint bounds). There are, of course, alternative approaches. For example, the linear or quadratic approximation to the steady-state map may be examined to approximately identify the input (process-variable) region associated with a given output (process-input) region. An alternative performance objective could be utilized, including a weighting matrix for the states that would reflect the scaling of the variables.

While rigorous analysis of the region of operation is not a central theme to this research, the use of coherence analysis as an open-loop nonlinearity measure provides a simplistic but effective method to examine varying regions of operation.

Output feedback controller

In the preceding analysis, it has been assumed that the system states are available for feedback through the OCS. However, in many systems, some states are not available for measurement. Indeed, none of the states may be measurable. Instead, the unmeasured states may be estimated from available information, and then used by the OCS. The estimator and OCS combination may be evaluated by any open-loop nonlinearity test.

Optimal nonlinear state estimation — The extended Kalman filter

The method of nonlinear state estimation used here is extended Kalman estimation, first introduced by Kalman in Kalman (1960), and described more recently in Ray (1990) and Lewis (1992).

The success of an estimation scheme rests on the observability of the system. Nonlinear observability tests are complex, and often not tractable. Therefore, in general (Ray, 1990), nonlinear systems are analyzed by testing the local weak observability of the linearized system at a number of discrete operating points.

Consider the system:

$$\begin{aligned} \dot{x}(t) &= f(x,u,t) + G\xi(t) \\ x(0) &= x_0 + \xi_0 \\ y(t) &= h(x,t) + \eta(t) \end{aligned} \tag{29}$$

where x, y, and u are the system states, measured outputs, and inputs; ξ and ξ_0 are random process disturbances; and $\eta(t)$ is a measurement error, all assumed to be white-noise processes. The noise-shaping matrix G is used to model process disturbances that are nonwhite. Define the matrices A and C by

$$\begin{aligned} A &= \frac{\partial f}{\partial x} \\ C &= \frac{\partial h}{\partial x}. \end{aligned} \tag{30}$$

System 29 is weakly observable if the standard observability matrix L_0 has rank equal to the state dimension. Ideally, this test should be run at all points within the region of operation, a cumbersome task. However, the observability is often determined by the system structure, and is not dependent upon x_0 in a complex manner. In these cases, simple linearized observability tests are usually adequate for nonlinear problems (Ray, 1990), and the system may be said to be locally weakly observable (Hermann and Krener, 1977).

The extended Kalman filter is a nonlinear estimator based on the Kalman filter which, in the case where all noises are Gaussian, as assumed in this article, is the optimal linear estimator in the sense that it minimizes the variance on $\hat{x}$, the estimate of x. These estimates are updated (Eq. 31) by summing a term accounting for system dynamics with a correction factor K, called the Kalman gain, multiplied by the error between the real and estimated measurements.

$$\dot{\hat{x}}(t_f \mid t_f) = f(\hat{x}, t_f) + K[y(t_f) - h(\hat{x}, t_f)]. \tag{31}$$

In Eq. 31, $\hat{x}(t_1 \mid t_2)$ are estimates for x at time t_1 given all the information available up to and including time t_2. In this case, t_f is the actual time of the estimation.

The Kalman gain K is given by

$$K = P(t_f \mid t_f) h_x^T Q(t_f), \tag{32}$$

where the error covariance matrix P is determined by

$$\begin{aligned} \dot{P}(t_f \mid t_f) &= P(t_f \mid t_f)\hat{A}^T + \hat{A}P(t_f \mid t_f) \\ &\quad - P(t_f \mid t_f)\hat{C}^T Q \hat{C} P(t_f \mid t_f) + GR^{-1}G(t_f) \\ P(0 \mid 0) &= P_0. \end{aligned} \tag{33}$$

Diagonal matrices P_0, $R(t)$, and $Q(t)$ are chosen to reflect the error in the initial estimate and the covariances assumed for $\xi(t)$ and $\eta(t)$ in Eq. 29. Equation 33 should be recognized as a Riccati equation. Generally, for the following analyses, errors in the initial estimate and process model are small, and so $P(0)$ and $R(t)$ should be small. In contrast, to ensure that errors between the estimated and real measurements are quickly eliminated, Q is chosen to be relatively large.

The use of the Kalman filter in tandem with the OCS is demonstrated in the following case studies.

Case Studies

CSTR with van de Vusse kinetics case study

The CSTR with van de Vusse kinetics (van de Vusse, 1964) is a model of a real process to produce cyclopentanol from cyclopentadiene, with a number of competing side reactions (Engell and Klatt, 1993a). The reaction scheme is modeled by

$$\begin{aligned} \mathfrak{A} &\rightarrow \mathfrak{B} \rightarrow \mathfrak{C} \\ 2\,\mathfrak{A} &\rightarrow \mathfrak{D} \end{aligned} \tag{34}$$

where $\mathfrak{A}$ is the reactant, cyclopentadiene, $\mathfrak{B}$ is the product, cyclopentanol, and $\mathfrak{C}$ and $\mathfrak{D}$ are unwanted side products cyclopentandiol and dicyclopentadiene, respectively. Assum-

Table 2. Constants for the Nonisothermal van de Vusse Example

Feed concentration of $\mathfrak{A}$ (Constant)	x_{10}	mol/L	5.1
Collision factor for reaction $\mathfrak{A} \rightarrow \mathfrak{B}$	k_{10}	1/h	1.287e12
Collision factor for reaction $\mathfrak{B} \rightarrow \mathfrak{C}$	k_{30}	1/h	1.287e12
Collision factor for reaction $\mathfrak{A} \rightarrow \mathfrak{D}$	k_{30}	1/h	9.043e9
Activation energy for reaction $\mathfrak{A} \rightarrow \mathfrak{B}$	E_1	K	−9,758.3
Activation energy for reaction $\mathfrak{B} \rightarrow \mathfrak{C}$	E_2	K	−9,758.3
Activation energy for reaction $\mathfrak{A} \rightarrow \mathfrak{D}$	E_3	K	−8,560
Volume of reactor	V_R	L	10
Enthalpy of reaction $\mathfrak{A} \rightarrow \mathfrak{B}$	ΔH_1	kJ/mol A	4.2
Enthalpy of reaction $\mathfrak{B} \rightarrow \mathfrak{C}$	ΔH_2	kJ/mol B	−11
Enthalpy of reaction $\mathfrak{A} \rightarrow \mathfrak{D}$	ΔH_3	kJ/mol A	−41.85
Heat removed by heat exchanger	Q	kJ/h	−4,496
Density	ρ	kg/L	0.9342
Heat capacity	C_p	kJ/kg·K	3.01
Temperature of feed stream	T_0	K	403.15

ing a thermal system where the coolant dynamics can be neglected and the cooling rate assumes a constant value Q, the mass balances for $\mathfrak{A}$ and $\mathfrak{B}$ as well as the heat balance for the reactor are (Engell and Klatt, 1993a):

$$\dot{x}_1 = \frac{u}{V_R}(x_{10} - x_1) - k_1 x_1 - k_3 x_1^2$$

$$\dot{x}_2 = -\frac{u}{V_R} x_2 + k_1 x_1 - k_2 x_2$$

$$\dot{x}_3 = \frac{1}{\rho C_p}[k_1 x_1(-\Delta H_1) + k_2 x_2(-\Delta H_2) + k_3 x_1^2(-\Delta H_2)] + \frac{u}{V_R}(T_0 - x_3) + \frac{Q}{\rho C_p V_R}$$

$$k_1 = k_{10} e^{E_1/R x_3}$$

$$k_2 = k_{20} e^{E_2/R x_3}$$

$$k_3 = k_{30} e^{E_3/R x_3}. \tag{35}$$

Table 3. Operating Points for Nonisothermal van de Vusse Reactor

		Point:	A	B
Conc. of $\mathfrak{A}$ in reactor	$x_{1,ss}$	mol/L	2.4946	1.0569
Conc. of $\mathfrak{B}$ in reactor	$x_{2,ss}$	mol/L	1.1003	0.8109
Temp. in reactor	$x_{3,ss}$	K	411.08	399.02
Feed rate	u_{ss}	1/h	80	9.25

The state variables x_1, x_2, and x_3 are the concentrations of $\mathfrak{A}$ and $\mathfrak{B}$ and the reactor temperature; u is the manipulated variable—the inlet flow rate; k_1, k_2, and k_3 are temperature-dependent Arrhenius rate coefficients; and R is the gas constant. The remaining variables are defined in Table 2. For the output feedback case study, it will be assumed that only the concentration of the product and the reactor temperature are available as measured states, denoted

$$y_1 = x_2$$

$$y_2 = x_3 \tag{36}$$

A steady-state plot of the concentration of species $\mathfrak{B}$ (x_2) vs. the feed rate (u) (Figure 8) reveals regions where the gain changes sign, for example, in the region around $u = 9$ (point B), and regions where the relationship is very linear, for example around $u = 80$ (point A). These operating points are defined in Table 3. In regions where the gain changes sign, the gain error exceeds 100%, and no linear controller with integral action can stabilize the system (Morari and Zafiriou, 1989). Although the OCS cannot strictly reproduce this behavior, it may be expected to exhibit greater nonlinearity at operating point B than point A as α approaches zero. The steady-state plots of x_1 vs. u and x_3 vs. u are given in Figures 9 and 10.

If it is assumed that the system nonlinearity can be assessed from the steady-state plots, the system in the region around point A will be relatively linear, with simple dynamics and little interaction, placing it within the bottom, left, front

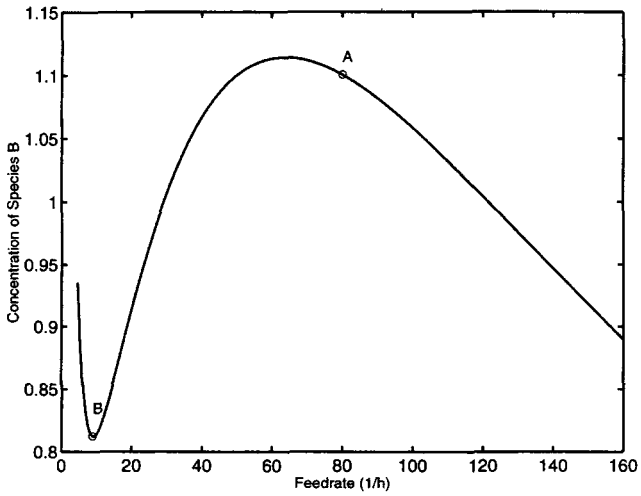

Figure 8. Steady-state concentration of species $\mathfrak{B}$ as a function of reactor feed rate for the nonisothermal van de Vusse reactor.

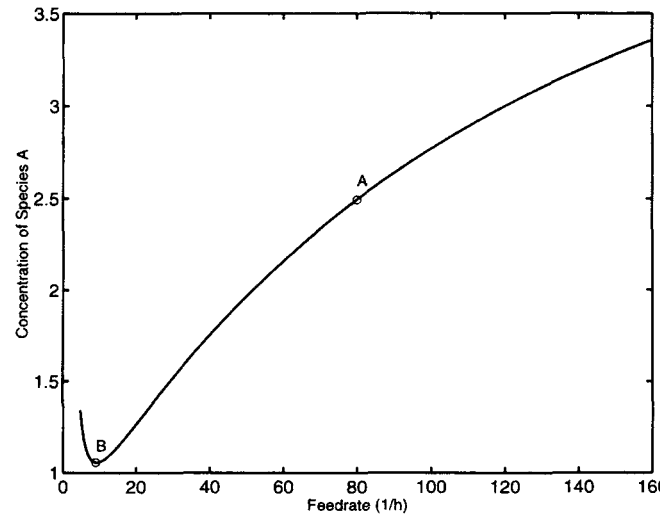

Figure 9. Steady-state concentration of species $\mathfrak{A}$ as a function of reactor feed rate for the nonisothermal van de Vusse reactor.

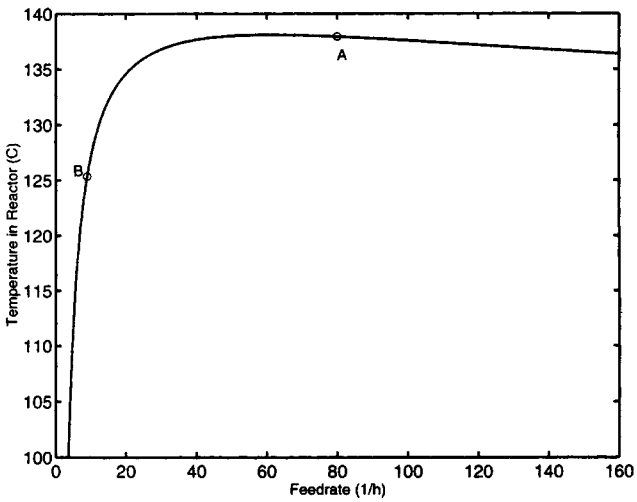

Figure 10. Steady-state reactor temperature as a function of reactor feed rate for the nonisothermal van de Vusse reactor.

subcube of the process characterization cube. Around point B, the system appears relatively nonlinear, has difficult dynamics and has little interaction, which would place it within the bottom, right, back subcube (Ogunnaike and Ray, 1994).

where the OCS operator $\mathfrak{N}$ has inputs x_1, x_2, and x_3 and output u. The "reverse-time" Jacobian for Point B is

$$\begin{bmatrix} -40.0221 & 30.809 & 91.1011 \\ 0 & -30.809 & 120.521 \\ -2.2559 & 0.460509 & 6.91434 \end{bmatrix}, \tag{38}$$

which has eigenvalues -46.557, $-8.67992+3.12143i$, and $-8.67992-3.12143i$. The reverse-time Jacobian for Point A is

$$\begin{bmatrix} -103.969 & 63.1361 & 513.415 \\ 0 & -63.1361 & 246.981 \\ -11.6749 & 5.08348 & 40.5048 \end{bmatrix} \tag{39}$$

which has eigenvalues -97.3357, $-14.6324+38.361i$, and $-14.6324-38.361i$. For both points, the OCS is stable when integrated in reverse time.

Measurement-feedback OCS

The output-feedback model for the CSTR with van de Vusse kinetics is given by Eqs. 35 and 36. The observability matrix

$$L_0 = \left[\begin{array}{ccc|c} 0 & 0 & k_{1,ss} & \frac{1}{\rho C_p}[k_{1,ss}\Delta H_1 + 2k_{3,ss}x_{1,ss}\Delta H_2] \\ 1 & 0 & -u-k_{2,ss} & \frac{-1}{\rho C_p}[k_{2,ss}\Delta H_2] \\ 0 & 1 & x_{1,ss}\frac{\partial k_{1,ss}}{\partial x_{3,ss}} - x_{2,ss}\frac{\partial k_{2,ss}}{\partial x_{3,ss}} & \frac{\partial \dot{x}_{3,ss}}{\partial x_{3,ss}} \end{array}\right] \begin{array}{|c} \\ \\ (A^T)^2 C^T \end{array} \tag{40}$$

will have rank three, equal to the state dimension, for all points along the locus of operation as $k_{1,ss} \neq 0$, where the subscript denotes a steady-state value. Note that A and C are defined in Eq. 30. This system should therefore be locally and weakly observable.

The measurement-feedback OCS is

$$\dot{\lambda}_1 = -\hat{x}_1 - \lambda_1(-k_1^* - 2k_3^*\hat{x}_1 - u) - \lambda_2 k_1^* - \lambda_3 \frac{-(k_1^*\Delta H_1 + 2k_3^*\Delta H_3 \hat{x}_1)}{C_p \rho}$$

$$\dot{\lambda}_2 = -y_1 - \lambda_2(-k_2^* - u) - \lambda_3 \frac{-k_2^*\Delta H_2}{C_p \rho}$$

$$\dot{\lambda}_3 = -y_2 - \lambda_1\left(-\hat{x}_1 \frac{\partial k_1^*}{\partial y_2} - \hat{x}_1^2 \frac{\partial k_3^*}{\partial y_2}\right) - \lambda_2\left(\hat{x}_1 \frac{\partial k_1^*}{\partial y_2} - y_1 \frac{\partial k_2^*}{\partial y_2}\right) - \lambda_3 \frac{-\left(\hat{x}_1 \Delta H_1 \frac{\partial k_1^*}{\partial y_2} + y_1 \Delta H_2 \frac{\partial k_2^*}{\partial y_2} + \hat{x}_1^2 \Delta H_3 \frac{\partial k_3^*}{\partial y_2}\right) - uC_p\rho}{C_p \rho}$$

State-feedback OCS

The state-feedback OCS is given by

$$\dot{\lambda}_1 = -x_1 - \lambda_1(-k_1 - 2k_3 x_1 - u) - \lambda_2 k_1 - \lambda_3 \frac{-(k_1 \Delta H_1 + 2k_3 \Delta H_3 x_1)}{C_p \rho}$$

$$\dot{\lambda}_2 = -x_2 - \lambda_2(-k_2 - u) - \lambda_3 \frac{-k_2 \Delta H_2}{C_p \rho}$$

$$\dot{\lambda}_3 = -x_3 - \lambda_1\left(-x_1 \frac{\partial k_1}{\partial x_3} - x_1^2 \frac{\partial k_3}{\partial x_3}\right) - \lambda_2\left(x_1 \frac{\partial k_1}{\partial x_3} - x_2 \frac{\partial k_2}{\partial x_3}\right) - \lambda_3 \frac{-\left(x_1 \Delta H_1 \frac{\partial k_1}{\partial x_3} + x_2 \Delta H_2 \frac{\partial k_2}{\partial x_3} + x_1^2 \Delta H_3 \frac{\partial k_3}{\partial x_3}\right) - uC_p\rho}{C_p \rho}$$

$$u = \frac{-\lambda_1(x_{10} - x_1) - \lambda_2(-x_2) - \lambda_3(T_0 - x_3)}{\alpha}, \tag{37}$$

$$\dot{\hat{x}}_1 = -k_1\hat{x}_1 - k_3\hat{x}_1^2 + u(x_{10} - \hat{x}_1)$$
$$+(Q_{11}P_{12} + Q_{21}P_{13})(y_1 - \hat{x}_2) + (Q_{12}P_{12} + Q_{22}P_{13})(y_2 - \hat{x}_3)$$

$$\dot{\hat{x}}_2 = k_1\hat{x}_1 - k_2\hat{x}_2 - u\hat{x}_2 + (Q_{11}P_{22} + Q_{21}P_{23})(y_1 - \hat{x}_2)$$
$$+(Q_{12}P_{22} + Q_{22}P_{23})(y_2 - \hat{x}_3)$$

$$\dot{\hat{x}}_3 = \frac{1}{\rho C_p}[k_1\hat{x}_1(-\Delta H_1) + k_2\hat{x}_2(-\Delta H_2) + k_3\hat{x}_1^2(-\Delta H_2)]$$
$$+u(T_0 - \hat{x}_3) + \frac{Q}{\rho C_p V_R} + (Q_{11}P_{32} + Q_{21}P_{33})(y_1 - \hat{x}_2)$$
$$+(Q_{12}P_{32} + Q_{22}P_{33})(y_2 - \hat{x}_3)$$

$$\dot{P} = AP + PA^T + GR^{-1}G^T - PC^TQCP$$

$$u = \frac{-\lambda_1(x_{10} - \hat{x}_1) - \lambda_2(-y_1) - \lambda_3(T_0 - y_2)}{\alpha}$$

$$k_1 = k_{10}e^{E_1/\hat{x}_3} \quad k_2 = k_{10}e^{E_1/\hat{x}_3} \quad k_3 = k_{30}e^{E_3/\hat{x}_3}$$
$$k_1^* = k_{10}e^{E_1/y_2} \quad k_2^* = k_{10}e^{E_1/y_2} \quad k_3^* = k_{30}e^{E_3/y_2} \tag{41}$$

where

$$P = \begin{bmatrix} P_{11} & P_{12} & P_{13} \\ P_{21} & P_{22} & P_{23} \\ P_{31} & P_{32} & P_{33} \end{bmatrix} \qquad G = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$C = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$A = \begin{bmatrix} -k_1 - 2\hat{x}_1\dfrac{\partial k_3}{\partial \hat{x}_3} & 0 & -\hat{x}_1\dfrac{\partial k_1}{\partial \hat{x}_3} - \hat{x}_1^2\dfrac{\partial k_3}{\partial \hat{x}_3} \\ k_1 & -k_2 & \hat{x}_1\dfrac{\partial k_1}{\partial \hat{x}_3} - \hat{x}_2\dfrac{\partial k_2}{\partial \hat{x}_3} \\ \dfrac{-\hat{x}_1\Delta H_1\dfrac{\partial k_1}{\partial \hat{x}_3} - \hat{x}_2\Delta H_2\dfrac{\partial k_2}{\partial \hat{x}_3}}{C_p\rho} & \dfrac{-k_2\Delta H_2}{C_p\rho} & \dfrac{-\hat{x}_1\Delta H_1\dfrac{\partial k_1}{\partial \hat{x}_3} - \hat{x}_2\Delta H_2\dfrac{\partial k_2}{\partial \hat{x}_3} - \hat{x}_1^2\Delta H_3\dfrac{\partial k_3}{\partial \hat{x}_3}}{C_p\rho} \end{bmatrix}$$

$$R^{-1} = \begin{bmatrix} R_{11}^{-1} & 0 & 0 \\ 0 & R_{22}^{-1} & 0 \\ 0 & 0 & R_{33}^{-1} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0.0001 & 0 \\ 0 & 0 & 0.0001 \end{bmatrix}$$

$$Q = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} = \begin{bmatrix} 25000 & 0 \\ 0 & 25000 \end{bmatrix}. \tag{42}$$

This OCS has the input–output form $\mathfrak{N}$ given by

$$\dot{\lambda} = \mathfrak{N}_1(\lambda, y_1, y_2)$$
$$u = \mathfrak{N}_2(\lambda, y_1, y_2, \alpha), \tag{43}$$

where the inputs to the operator are y_1 and y_2, and the output is u. In the state-feedback OCS, the values of x_1 are determined by the input signal. In the output-feedback OCS formulation, the values of x_1 are estimated, and are free to vary as determined by the equations given earlier. As a result, the results for the output feedback OCS will not necessarily match the results for the state-feedback OCS.

The Jacobians, which are omitted due to length, were evaluated at points A and B for a number of values of α. The eigenvalues indicate that this OCS has both left and right half-plane poles, so the measurement-feedback OCS is unstable in both the normal and "reverse" time sense. This implies that although coherence measurements are still possible, care must be taken to ensure that the outputs are bounded and a segment length is chosen to ensure quasi stationarity. The measurement-feedback OCS was integrated in reverse time when generating the results used in the next subsection.

Coherence analysis results

Results are shown for $\alpha = 0.1$, as OCSs with smaller values of α could only be integrated with difficulty. The results detailed in this article are representative: the inputs, outputs, and coherences for specific input–output channels for many additional simulations may be found in Stack (1995). The outputs were subjected to the run tests for stationarity, as detailed in the subsection on nonstationary data. Windows of length 32 points were found to be appropriate, and are used in all analyses of the thermal van de Vusse OCS.

The individual channel coherences for one test of the state-feedback OCS at point A are given in Figure 11. Note that these figures use a consistent three-graph pattern: a coherence graph is located above its corresponding input and output graphs. These show that the dynamics for the chosen region of operation are dominated by a relatively linear relationship between input x_3, the reactor temperature, and the output, while inputs x_1 and x_2 have no linear relationship with the output. Similarly, the individual channel coherences for one test of the measurement-feedback OCS at point A are given in Figure 12. Again, the dynamics are dominated by a relatively linear relationship between input y_2, the reactor temperature, and the output, while y_1 has no linear relationship with the output. Assuming that the regions of operation for tests are realistic, these results imply that a control scheme should emphasize control of the reactor temperature for either a state-feedback, or more realistically, an output-feedback case.

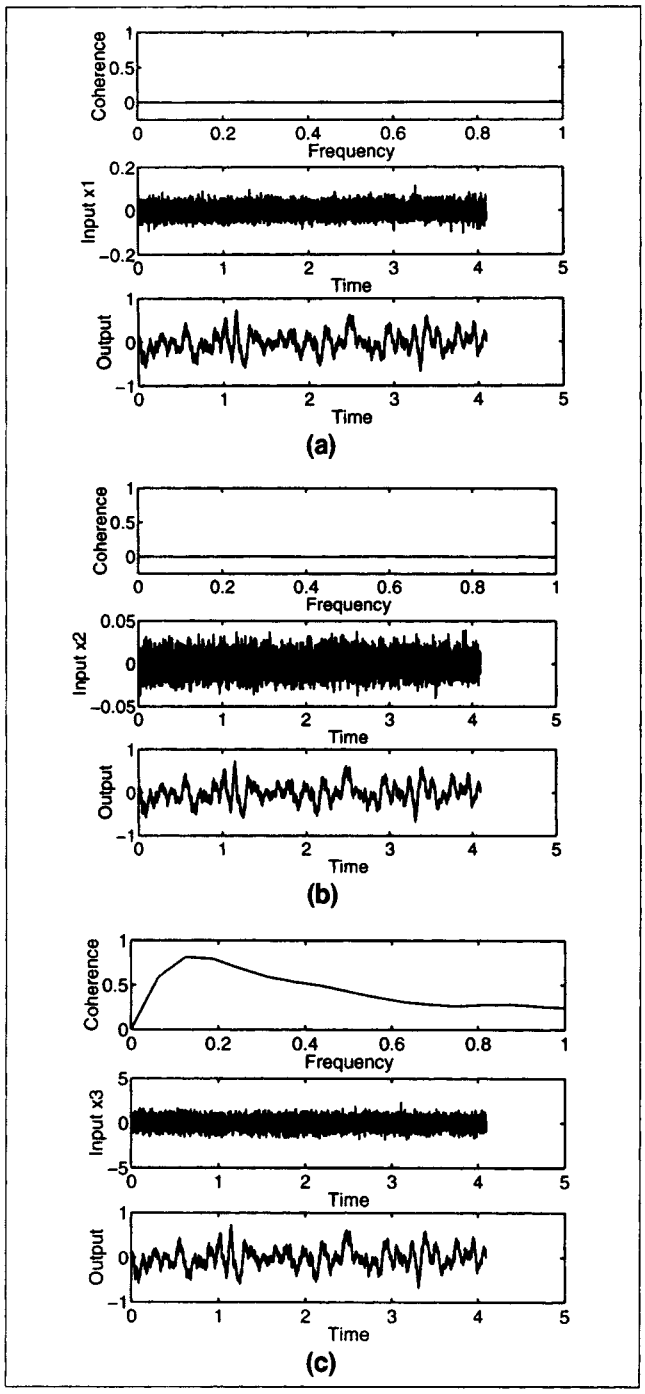

Figure 11. Coherence analysis for the state-feedback nonisothermal van de Vusse reactor for point *A* with $\alpha = 0.1$.

$N = 8{,}192$; $\sigma_{x_1}^2$: $(x_{1,ss}/100)^2$; $\sigma_{x_2}^2$: $(x_{2,ss}/100)^2$; $\sigma_{x_3}^2$: $(x_{3,ss}/250)^2$; $t_s = 0.0005$; $h = 0.0001$; $S = 32$. (a) Input x_1. (b) Input x_2. (c) Input x_3.

Cumulative coherence estimation results are presented for two sets of scaled input conditions (labeled scalings 1 and 2 on the figures) for the state-feedback OCS in Figure 13, and for the measurement-feedback OCS in Figure 14. For both problem formulations, point B was determined to be more control-law nonlinear than point A for both sets of scaling.

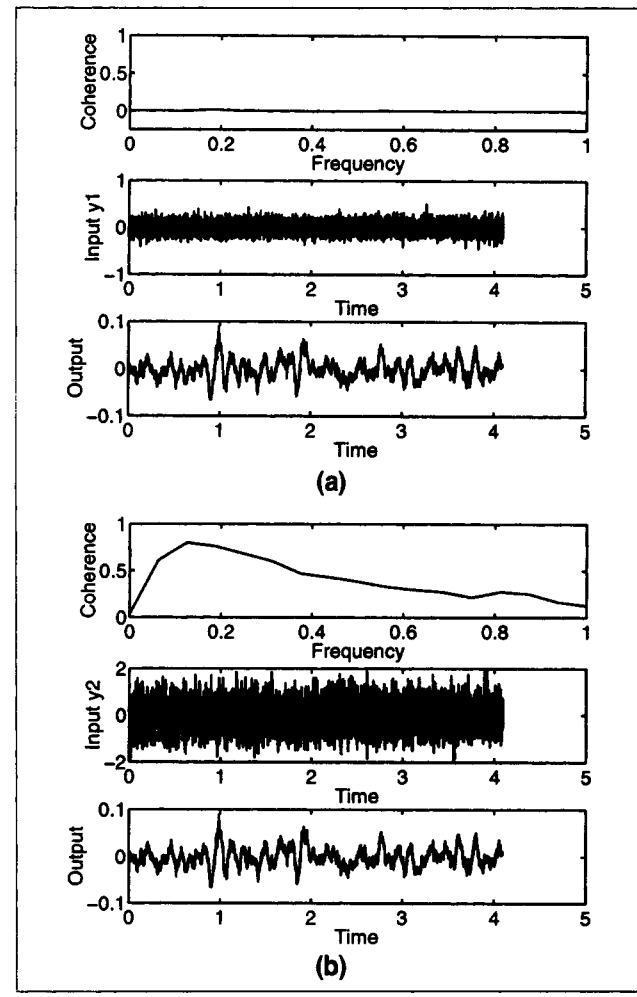

Figure 12. Coherence analysis for the measurement feedback nonisothermal van de Vusse reactor for point *A* with $\alpha = 0.1$.

$N = 8{,}192$; y_1: $(1.1004/10)^2$; y_2: $(137.9329/250)^2$; $t_s = 0.0005$; $h = 0.0001$; $S = 32$. (a) Input y_1. (b) Input y_2.

This is consistent with the expected results described earlier: since no integrative linear controller can control the system at point B in the regulatory case, it was expected that point A would be more control-law linear than point B.

The OCS approach also detects changes in the control-law nonlinearity as the region of operation changes for a particular operating point. For both the state-feedback and measurement-feedback analyses, the coherence estimates for point A fall at higher frequencies as the region of operation for the reactor temperature is widened, or as we move from scalings 1 and 2 on the figures. As a result, the frequency range over which there is a significant difference in the control-law nonlinearity of points A and B shrinks as the assumed variance in the reactor temperature increases.

The same tests were repeated using equal (or unscaled) inputs for both points on the same order of magnitude as the inputs used in Figures 13 and 14. The results were virtually the same as for the scaled case.

The control-law nonlinearity of the points could also be examined for output, rather than input, regions of operation by imposing constraints on the energy of the output. Unlike

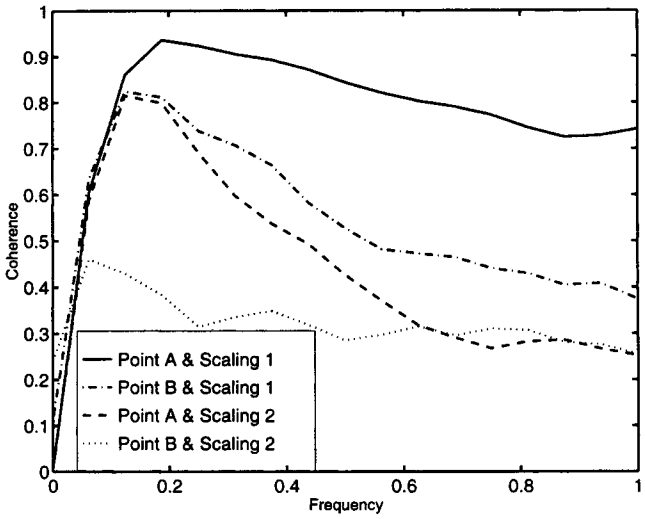

Figure 13. Cumulative coherence analyses for the state-feedback thermal van de Vusse reactor with $\alpha = 0.1$.

$N = 8{,}192$; $\sigma_{x_1}^2$: $(x_{1,ss}/100)^2$; $\sigma_{x_2}^2$: $(x_{2,ss}/100)^2$; $\sigma_{x_3}^2$: $(x_{3,ss}/1{,}000)^2$ (scaling 1), $\sigma_{x_3}^2$: $(x_{3,ss}/250)^2$ (scaling 2); $t_s = 0.0005$; $h = 0.0001$; $S = 32$.

the previous results, point A was found to be more control-law nonlinear than point B over the examined frequency range. However, to excite comparable outputs, it is necessary to use input variances for the reactor temperature for point A approximately an order to magnitude larger than the inputs used for point B. For example, to generate an output bounded by $\pm 0.1(1/h)$ for the output feedback case, the input variances for point A were $\sigma_{y_1}^2 = 0.0066$ and $\sigma_{y_2}^2 = 0.2535$ compared to $\sigma_{y_1}^2 = 0.000065983$ and $\sigma_{y_2}^2 = 0.0158$ for point B. Therefore, simple comparisons based on a common output region of operation are questionable, and need to be interpreted with care. Note that this is the opposite conclusion from results for the batch reactor in Stack (1995), where comparison based on the input was uninformative, and the output was the appropriate basis for comparison.

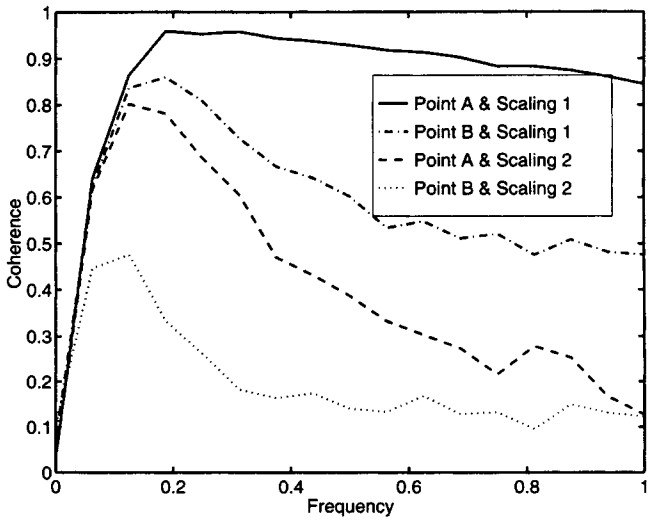

Figure 14. Cumulative coherence analyses for the measurement feedback thermal van de Vusse reactor with $\alpha = 0.1$.

$N = 8{,}192$; $\sigma_{y_1}^2$: $(y_{1,ss}/100)^2$; $\sigma_{u_2}^2$: $(y_{2,ss}/1{,}000)^2$ (scaling 1); $\sigma_{y_1}^2$: $(y_{1,ss}/10)^2$, $\sigma_{y_2}^2$: $(y_{2,ss}/250)^2$ (Scaling 2); $t_s = 0.0005$; $h = 0.0001$; $S = 32$.

Conclusions

Control-law nonlinearity, defined for this article as the appropriate degree of nonlinearity in the controller for a system, has been assessed using the OCS approach. Previously, the OCS approach had only been developed for SISO state-feedback systems. In this article, the approach is extended to systems with output feedback through the introduction of an appropriate Kalman filter. It is also demonstrated that coherence estimation can be used to measure the nonlinearity of an output relative to multiple inputs. With these two advances, the OCS can now be used to analyze multivariate systems with measurement feedback, showing the potential to be used as a design tool in practical situations. It is demonstrated that coherence estimation can be performed upon nonstationary data as long as a segment length can be found that is quasi stationary. These capabilities, taken together, suggest that a wide range of nonlinear systems should be amenable to assessment by the combination of the OCS and coherence analysis.

The CSTR with van de Vusse kinetics, an example of a real process featuring output feedback, was analyzed for both the state-feedback and output-feedback cases. The OCS approach correctly discerned the difference in control-law nonlinearity between an operating point that contains a change in the sign of the gain, indicating that the system cannot be controlled by any linear controller with integrative action at that point, and an operating point where the sign does not change. Furthermore, changes in the control-law nonlinearity with the variance in the reactor temperature were observed at high frequencies, emphasizing that the control-law nonlinearity strongly depends upon the region of operation chosen, and is frequency dependent.

Acknowledgments

The authors would like to acknowledge funding from the National Science Foundation through a NYI award (CTS 9257059) and funding from the Shell Foundation. In addition, the authors would like to acknowledge the helpful suggestions of the anonymous reviewers.

Notation

A = system Jacobian
C = output Jacobian
h = integration time
$m(t)$ = mean function
N = number of points used for estimation
$R(t,r)$ = covariance function
t_s = sampling time
$*$ = complex conjugate
σ^2 = input variance
τ = time constant

Literature Cited

Al'Brekht, E., "On the Optimal Stabilization of Nonlinear Systems," *J. Appl. Math. Mech.*, **25**, 1254 (1961).

Allgöwer, F. "Definition and Computation of a Nonlinearity Measure and Application to Approximate I/O-Linearization," Tech. Rep. 95-1, Stuggart Univ., Stuttgart, Germany (1995).

Allgöwer, F., A., and E. D. Gilles, "Approximate Input/Output Linearization of Nonlinear Systems," AIChE Meeting, Miami (1992).

Aris, R., *The Optimal Design of Chemical Reactors*, Academic Press, New York (1961).
Bendat, J. S., *Nonlinear System Analysis and Identification from Random Data*, Wiley, New York (1990).
Bendat, J. S., and A. G. Piersol, *Random Data: Analysis and Measurement Procedures*, Wiley, New York (1986).
Bendat, J. S., and A. G. Piersol, *Engineering Applications of Correlation and Spectral Analysis*, Wiley, New York (1993).
Benignus, V., "Estimation of the Coherence Spectrum and Its Confidence Interval Using the Fast Fourier Transform," *IEEE Trans. Audio Electroacoust.*, **AU-17**(2), 145 (1969).
Bernstein, D., "Nonquadratic Cost and Nonlinear Feedback Control," *Int. J. Rob. Nonlinear Contr.*, **3**, 211 (1993).
Bhalla, U. S., and J. M. Bower, "Exploring Parameter Space in Detailed Single Neuron Models: Simulations of the Mitral and Granule Cells of the Olfactory Bulb," *J. Neurophysiol.*, **69**, 1948 (1993).
Carter, G. C., ed., *Coherence and Time Delay Estimation*, IEEE Press, New York (1993).
Carter, G. C., C. H. Knapp, and A. H. Nuttal, "Estimation of the Magnitude-Squared Coherence Function Via Overlapped Fast Fourier Transform Processing," *IEEE Trans. Audio Electroacoust.*, **AU-21**, 337 (1983).
Chen, H., A. Kremling, and F. Allgöwer, "Nonlinear Predictive Control of a Benchmark CSTR," *Proc. Euro. Control Conf.*, Rome (1995).
Cohen, B. A., and A. Sances, "Stationarity of the Human Electroencephalogram," *Med. Biol. Eng. Comput.*, **15**, 513 (1977).
Doyle, J., T. Georgiou, and M. Smith, "The Parallel Projection Operators of a Nonlinear Feedback System," *Syst. Contr. Lett.*, **20**, 79 (1993).
Doyle, F. J., III, F. Allgöwer, S. Oliverira, E. Gilles, and M. Morari, "On Nonlinear Systems with Poorly Behaved Zero Dynamics," *Proc. Amer. Control Conf.*, Chicago (1992).
Doyle, F. J., III, M. Morari, and J. C. Doyle, "Some Practical Considerations in the Selection of 'Linearizing' Control over Linear Control," AIChE Meeting, Los Angeles (1991).
Engell, S., and K. Klatt, "Nonlinear Control of a Non-Minimum-Phase CSTR," *Proc. Amer. Control Conf.* San Francisco, p. 2941 (1993a).
Engell, S., and K.-U. Klatt, "Gain Scheduling Control of a Non-Minimum-Phase CSTR," *Proc. Euro. Control. Conf.*, Groningen (1993b).
Eskinat, E., S. H. Johnson, and W. L. Luyben, "Use of Hammerstein Models in Identification of Nonlinear Systems," *AIChE J.*, **37**(2), 255 (1991).
Flandrin, P., "On the Positivity of the Wigner-Ville Spectrum," *Signal Process.*, **11**, 187 (1986).
Flandrin, P., and B. Escudié, "An Interpretation of the Pseudo-Wigner-Ville Distribution," *Signal Process.*, **6**, 27 (1984).
García, C. E., D. M. Prett, and M. Morari, "Model Predictive Control: Theory and Practice—A Survey," *Automatica*, **25**(3), 335 (1989).
Guay, M., *Measurement of Nonlinearity in Chemical Process Control*, PhD Thesis, Queen's Univ. at Kingston, Kingston, Ont., Canada (1996).
Guay, M., P. McLellan, and D. Bacon, "Measurement of Nonlinearity in Chemical Process Control Systems: The Steady State Map," *Can. J. Chem. Eng.*, **73**, 868 (1995).
Haber, R., "Nonlinearity Test for Dynamic Processes," *Proc. IFAC Identification and System Parameter Estimation Symp.*, York (1985).
Hermann, R., and A. J. Krener, "Nonlinear Controllability and Observability," *IEEE Trans. Automat. Contr.*, **AC-22**, 728 (1977).
Hicks, G. A., and W. H. Ray, "Approximation Methods for Optimal Control Synthesis," *Can. J. Chem. Eng.*, **49**, 522 (1971).
Kalman, R. E., "A New Approach to Linear Filtering and Prediction Theory," *J. Basic Eng.*, **82**, 34 (1960).
Kodera, K., R. Gendrin, and C. de Villedary, "Analysis of Time-Varying Signals with Small BT Values," *IEEE Acoust., Speech, Signal Processing*, **ASSP-26**, 64 (1978).
Kwakernaak, H., and R. Sivan, *Linear Optimal Control Systems*, Wiley, New York (1972).
Lapidus, L., and R. Luus, *Optimal Control of Engineering Processes*, Blaisdell, Boston (1967).
Lewis, F. L., *Applied Optimal Control and Estimation*, Prentice Hall, Englewood Cliffs, NJ (1992).
Li, W. C., and L. T. Biegler, "Process Control Strategies for Constrained Nonlinear Systems," *Ind. Eng. Chem. Res.*, **27**, 1421 (1988).
Little, J. N., and L. Shure, *Signal Processing Toolbox for use with MATLAB*, The MathWorks, Inc., Natick, MA (1992).
Ljung, L., *System Identification: Theory for the User*, Prentice Hall, Englewood Cliffs, NJ (1987).
Morari, M., E. Zafiriou, *Robust Process Control*, Prentice Hall, Englewood Cliffs, NJ (1989).
Nikolaou, M., and V. Hanagandi, "When is Nonlinear Dynamical Modeling Necessary?" *Proc. Amer. Control Conf.*, San Francisco, p. 910 (1993).
Ogunnaike, B. A., R. K. Pearson, and F. J. Doyle III, "Chemical Process Characterization: With Applications in the Rational Selection of Control Strategies," *Proc. Euro. Control Conf.*, Groningen (1993).
Ogunnaike, B. A., and W. Ray, *Process Dynamics, Modeling, and Control*, Oxford Univ. Press, New York (1994).
Oppenheim, A. V., "Speech Spectrograms using the Fast Fourier Transform," *IEEE Spectrum*, **7**, 57 (1970).
Ouarti, H., and T. Edgar, "The Use of Approximate Models and Exact Linearization for Control of Nonlinear Processes," *Proc. Amer. Control Conf.*, p. 2268, San Francisco (1993).
Pearson, R. K., and B. A. Ogunnaike, "Application of Coherence Analysis in Linear and Nonlinear Input/Output Model Identification," *Proc. AIChE Meeting* (1994a).
Pearson, R. K., and B. A. Ogunnaike, "Detection of Unmodelled Disturbance Effects by Coherence Analysis," *ADCHEM, IFAC*, Kyoto, Japan (1994b).
Pontryagin, L., V. Boltyanskii, R. Gamkrelidze, and E. Mischenko, *The Mathematical Theory of Optimal Processes*, Interscience, New York (1962).
Poolla, K., J. Shamma, and K. Wise, "Linear and Nonlinear Controllers for Robust Stabilization Problems: A Survey," *Proc. IFAC World Cong.*, IFAC, Tallinn, p. 176 (1990).
Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C The Art of Scientific Computing*, 2nd ed., Cambridge Univ. Press, Cambridge, England (1992).
Rawlings, J., E. Meadows, and K. Muske, "Nonlinear Model Predictive Control: A Tutorial and Survey," *ADCHEM*, IFAC, Kyoto, Japan (1994).
Ray, W., *Advanced Process Control*, McGraw-Hill, New York (1990).
Roberts, S., *Dynamic Programming in Chemical Engineering and Process Control*, Academic Press, New York (1964).
Rouff, M., and F. Lamnabhi-Lagarrigue, "A New Approach to Nonlinear Optimal Feedback Law," *Syst. Contr. Lett.*, **7**, 411 (1986).
Seider, W., D. Brengel, and S. Widagdo, "Nonlinear Analysis in Process Design," *AIChE J.*, **37**(1), 1 (1991).
Shinskey, F., *Process-Control Systems: Application, Design and Adjustment*, McGraw-Hill, New York (1988).
Skogestad, S., and I. Postlethwaite, *Multivariable Feedback Control*, Wiley, New York (1996).
Stack, A., and F. Doyle III, "The Optimal Control Structure Approach to Measuring Control-Relevant Nonlinearity," *Comput. Chem. Eng.* (1995).
Stack, A. J., "The Optimal Control Structure: A Measure for Control-Relevant Nonlinearity," MSChE Thesis, Purdue Univ., West Lafayette, IN, in press (1997).
Tierno, J. E., R. M. Murray, and J. C. Doyle, "An Efficient Algorithm for Performance Analysis of Nonlinear Control Systems," *Proc. American Control Conf.*, Seattle, WA (1995).
Tieu, D., W. Cluett, and A. Penlidis, "Optimization of Polymerization Reactor Operation: Review and Case Studies with the End-Point Collocation Method," *Poly. React. Eng.*, **2**(3), 275 (1994).
van de Vusse, J., "Plug-flow Type Reactor Versus Tank Reactor," *Chem. Eng. Sci.*, **19**, 994 (1964).
Walpole, R. E., and R. H. Myers, *Probability and Statistics for Engineers and Scientists*, Macmillan, New York (1989).
White, L. B., and B. Boashash, "Cross Spectral Analysis of Nonstationary Processes," *IEEE Trans. Inform. Theory*, **IT-36**, 830 (1990).
Zhu, X., and D. E. Seborg, "Nonlinear Model Predictive Control Based on Hammerstein Models," *Proc. Int. Symp. on Process Systems Eng.*, Vol. II, E. S. Yoon, ed., Korean Institute of Chemical Engineers, p. 995 (1994).

Manuscript received Oct. 24, 1995, and revision received July 9, 1996.