Fault Detection and Diagnosis of Rotating Machinery

Kenneth A. Loparo, *Fellow, IEEE*, M. L. Adams, Wei Lin, *Senior Member, IEEE*, M. Farouk Abdel-Magied, and Nadar Afshari

***Abstract*—A model-based approach to the detection and diagnosis of mechanical faults in rotating machinery is studied in this paper. For certain types of faults, for example, raceway faults in rolling element bearings, an increase in mass unbalance, and changes in stiffness and damping, algorithms suitable for real-time implementation are developed and evaluated using computer simulation.**

***Index Terms*—Fault detection, fault diagnosis, rotating machinery.**

I. Introduction

OTATING machines are essential components in most of today's manufacturing and production industries. Because it is usually not practical or economical to use redundant systems, real-time monitoring and diagnostics for rotating machinery equipment is required. In this paper, model-based techniques are developed for monitoring, fault detection, and diagnosis of faults in rotating machines. For the detection of unbalance, changes in stiffness and damping, etc., a collection of statistical observers or nonlinear filters is designed. In this multiple model framework, each fault to be identified is associated with a certain structure and parameters in the rotating machinery model [1]–[8]. We also present a novel technique that incorporates failure detection filters and sliding-mode detectors for the detection and diagnosis of faults in a rolling element bearing [9].

Rotating machinery diagnostics is an important area. Frequency-domain analysis of vibration data, in particular, the Fourier transform, is a common approach. Here, connections between certain spectral properties and the fundamental nature of various vibration problems and their origins have been perfected and widely used. However, while spectrum analysis has the potential of providing a major improvement over conventional time-based analysis techniques in diagnosing vibration related problems, there are numerous examples where spectrum analysis does not provide all the insights needed to identify the cause of the vibration problem. Other analysis tools such as spectrograms, wavelets, etc., have also been used for the same purpose.

Manuscript received January 15, 1999; revised June 20, 2000. Abstract published on the Internet July 1, 2000. This work was supported in part by the Electric Power Research Institute, Palo Alto, CA, CAMP Inc., First Energy Corporation, Cleveland OH, and the National Science Foundation under Grant ECS-9906218.

K. A. Loparo, M. L. Adams, and W. Lin are with Case Western Reserve University, Cleveland, OH 44106-7070 USA (e-mail: kal4@po.cwru.edu).

M. F. Abdel-Magied is with the Department of Electrical Engineering, Cairo University, Cairo, Egypt.

N. Afshari is with Cutler-Hammer/Eaton Corporation, Pittsburgh, PA 15275 USA.

Publisher Item Identifier S 0278-0046(00)08831-6.

Certain fault modes in rotating machinery are accompanied by an increase in vibration intensity, e.g., rub impact between rotating and nonrotating parts. The vibration signals can exhibit rich dynamics such as periodic, quasi-periodic, and chaotic behavior. Inherent in this trend from regular behavior (e.g., small-amplitude linear orbital motion characteristic of a normal machine) to complicated behavior (e.g., large-amplitude quasi-periodic or chaotic orbital motion associated with a faulted operating condition) is the emergence of nonlinear characteristics in the system's vibration response. Thus, a major goal of this research was to develop signal processing techniques that could extract special temporal features associated with fault modes in rotating machinery from noise corrupted vibration data. This is accomplished by using a statistical model-based approach for the study of rub impact and a novel implementation of failure detection filters for raceway faults in rolling element bearings.

II. Mathematical Model of Rub Impact

"Rub impact" is the phenomenon of intermittent or continuous contact between rotating and stationary parts at close-running clearance locations in a rotating machine. Rotating machinery generally operates most efficiently when the close sealing clearances between rotor and stator are kept to a minimum. Misalignment and rotor mass unbalance can readily lead to an operating condition where the rotor rubs against the stationary parts (journal to bearing, blades to casing, shaft to seal, etc.). These intermittent rubs can cause complicated nonlinear dynamic behavior that is difficult to predict or diagnose using conventional time-based analysis or spectrum analysis techniques. It is important to detect and diagnose this operating condition since continued rubbing can lead to accelerated wearing of close-clearance parts, or machine failure.

The rub impact model consists of a Laval rotor and a nonlinear finite boundary stiffness with Coulomb friction. The Laval rotor is modeled as a massless shaft mounted in two bearings at each end. A thin rigid disc (the rotor) is located symmetrically between the bearings. The disc of mass M rotates with a constant angular velocity Ω with radius R around the center O of the disc. It is not necessary that the center of mass β coincide with the center of rotation O (see Fig. 1).

The transverse bending vibrations of the rotor are modeled with two generalized coordinates. The disc translates in the $\vec{e}_x$ and $\vec{e}_y$ directions and rotates around $\vec{e}_z$. The position of $\boldsymbol{O}$ with respect to $\boldsymbol{O}^\circ$ is given by the vector $\vec{r}_o$ and the position of β with respect to $\boldsymbol{O}$ by $\vec{r}_\beta$, where $\vec{r}_\beta$ and $\vec{r}_o$ are expressed in terms of $\vec{e}_x$ and $\vec{e}_y$ as

$$\vec{r}_o = X\vec{e}_x + Y\vec{e}_y$$
$$\vec{r}_\beta = \varepsilon\cos(\Omega t)\vec{e}_x + \varepsilon\sin(\Omega t)\vec{e}_y.$$

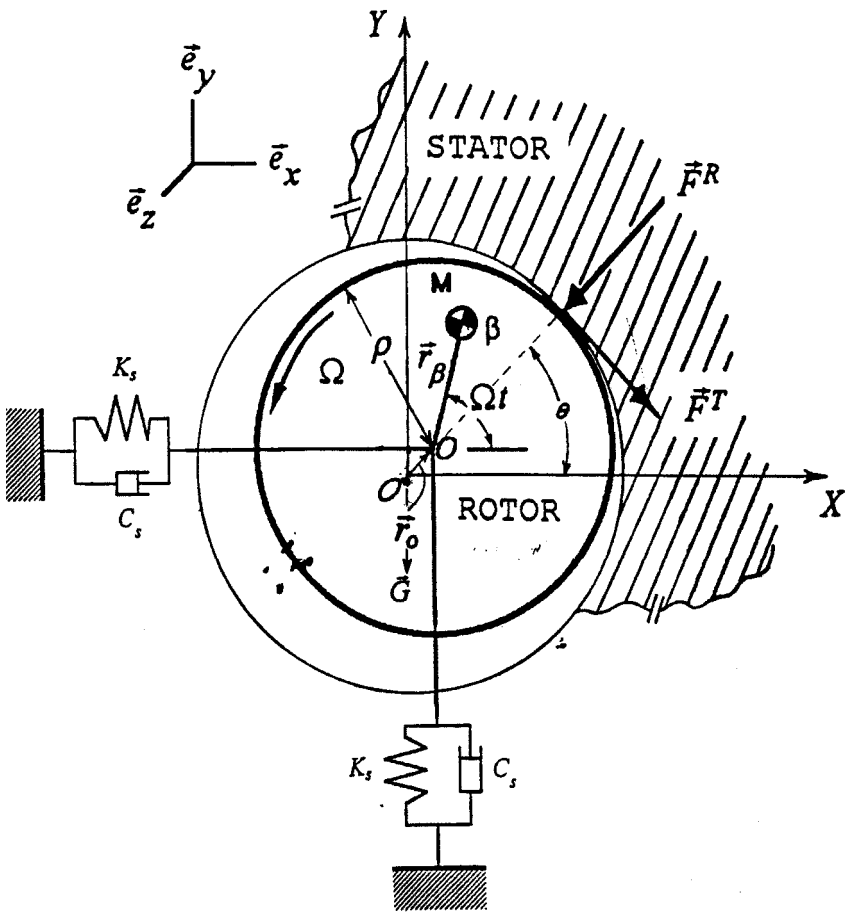


Fig. 1. Two-degrees-of freedom rotor/stator system.

X and Y are the generalized coordinates, ε is the distance from O and β, while Ωt is the rotation angle. The disc is loaded with a constant gravitational force $\vec{G}$ due to the weight of the disc, a linear restoring force $\vec{R}$, and a nonlinear force $\vec{F}_n$. The linear restoring force $\vec{R}$ acts at O and results from the isotropic stiffness K_s and isotropic damping D_s of the shaft and its bearings. The nonlinear force $\vec{F}_n$ is produced by the contact and rubbing of the rotor against the stationary housing. The contact condition is $\{\delta = ((R/C) - 1)\} \geq 0$. The housing acts on the disc with radial force $\vec{F}^R$ and tangential or friction force $\vec{F}^T$. The radial (Hertzian contact) force is given by

$$\vec{F}^R = \vec{F}^R_X + \vec{F}^R_Y = \begin{cases} \vec{0}, & \delta < 0 \\ -K_\beta \delta^\alpha (\cos\theta \vec{e}_x + \sin\theta \vec{e}_y), & \delta \geq 0. \end{cases}$$

The friction force $\vec{F}^T$ is governed by the Coulomb law with coulomb friction coefficient μ

$$\vec{F}^T = \vec{F}^T_X + \vec{F}^T_Y = \begin{cases} \vec{0}, & \delta < 0 \\ \mu K_\beta \delta^\alpha (\sin\theta \vec{e}_x - \cos\theta \vec{e}_y), & \delta \geq 0. \end{cases}$$

The equations of motion are obtained from Newton's Law

$$M\ddot{\vec{r}}_m = M\left(\ddot{\vec{r}}_o + \ddot{\vec{r}}_\beta\right) = \Sigma \vec{F}.$$

Given in coordinates

$$\vec{e_X}: M\left(\ddot{X} - \varepsilon\Omega^2 \cos(\Omega t)\right) = -K_s X - D_s \dot{X} - F_x$$

$$\vec{e_Y}: M\left(\ddot{Y} - \varepsilon\Omega^2 \sin(\Omega t)\right) = -K_s Y - D_s \dot{Y} - F_y - G - L$$

and

$$F_x = \vec{F}^T_x + \vec{F}^R_x, \quad F_y = \vec{F}^T_y + \vec{F}^R_y.$$

The equations of motion are transformed to nondimensional form by choosing the bearing radial clearance c as the characteristic length and τ as nondimensional time, and $\omega_s = \sqrt{K_s/M}$. The nondimensional substitutions are

$$x = \frac{X}{c}, \quad y = \frac{Y}{c}, \quad e = \frac{\varepsilon}{c}, \quad \overline{\omega} = \frac{\Omega}{\omega_s}, \quad \tau = \omega_s t$$

$$K^*_\beta = \frac{K_\beta}{Mc\omega_s^2}, \quad g = \frac{G}{c\omega_s^2}, \quad k = \frac{K_s}{M\omega_s^2}$$

$$f_x = \frac{F_x}{Mc\omega_s^2}, \quad f_y = \frac{F_y}{Mc\omega_s^2}, \quad \dot{(\,)} = \frac{d(\,)}{d\tau}, \quad \ddot{(\,)} = \frac{d^2(\,)}{d^2\tau}.$$

Due to the presence of modeling errors, input disturbances, and measurement noise, a nonlinear continuous-time stochastic model with additive plant and measurement noises is used. Define the state vector $x(t) = [x_r, \dot{x}_r, y_r, \dot{y}_r]^T$, the input

vector $u(t) = [\cos\overline{\omega}t, \sin\overline{\omega}t, 1]^T$, and the output vector $y(t) = [x_r, y_r]^T$, the state-space model of the system is given by

$$\dot{x}(t) = Ax(t) + Bu(t) + \gamma f(Cx(t)) + Gw(t)$$
$$y(t) = Cx(t) + v(t)$$

where

$$\gamma = \begin{cases} 0 & \delta < 0 \\ 1 & \delta \geq 0 \end{cases} \quad \text{and} \quad \delta = \left(\sqrt{x_r^2 + y_r^2} - 1\right)$$

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -k & -2\zeta & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -k & -2\zeta \end{bmatrix}$$

$$B = \begin{bmatrix} 0 & 0 & 0 \\ e\overline{\omega}^2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & e\overline{\omega}^2 & -g - l \end{bmatrix}$$

$$f(Cx(t)) = K_\beta^* \left(\sqrt{x_r^2 + y_r^2} - 1\right)^\alpha \cdot \begin{bmatrix} 0 \\ \dfrac{x_r}{\sqrt{x_r^2 + y_r^2}} - \mu \dfrac{y_r}{\sqrt{x_r^2 + y_r^2}} \\ 0 \\ \dfrac{y_r}{\sqrt{x_r^2 + y_r^2}} + \mu \dfrac{x_r}{\sqrt{x_r^2 + y_r^2}} \end{bmatrix}$$

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

The initial state $x(0)$, process noise $w(t)$ with gain G, and measurement noise $v(t)$ are assumed to be independent Gaussian white noise processes with known statistics.

III. Rolling Element Bearing Model

Assuming a rotor-bearing system with a stationary outer ring and one accelerometer mounted on the bearing housing as close as possible to the outer race, a one-dimensional vibration model is derived.

Fig. 2 represents a schematic diagram of this vibration model. This model is based on the masses of bodies in motion and various stiffness and damping factors.

1) m_r, k_s, and c_s are the mass of the rotor + inner ring, the stiffness of the shaft and the inner race contact in series, and the damping of the shaft, respectively.
2) m_b, k_c, and c_c are the mass of the balls, the stiffness of the contact, and the damping of the contact, respectively.
3) m_h, k_h, and c_h are the mass of housing + outer ring, the stiffness of the housing with the stiffness of the contact in series, and the damping of the housing, respectively.
4) Ps denote the excitation forces due to eccentricity and faults (inner, outer, and ball).

The values of the masses and the stiffness parameters in the proposed model have a special importance, because they are the components that determine the frequency response of the model.

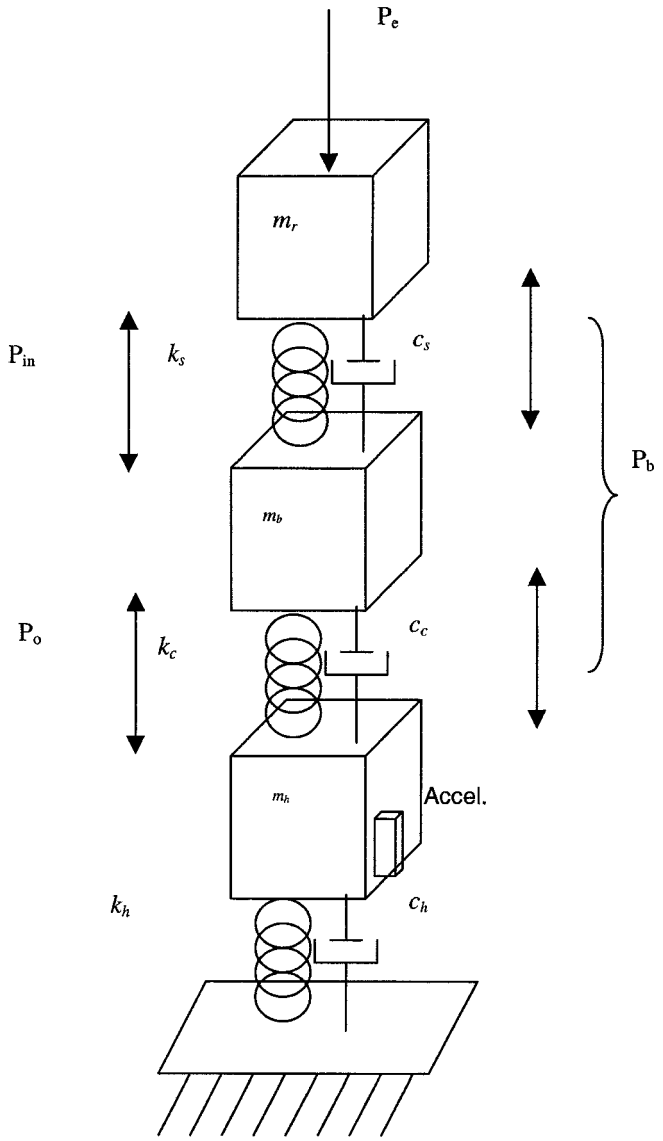


Fig. 2. Schematic diagram of roller bearing model.

The values of these parameters can vary considerably depending on the size, shape, and material used for the bearing and machine components, and the location of the sensor. The manufacturer can normally provide the nominal values for many of these components. However, the stiffness of the shaft and housing vary from one rotor-bearing system to another. Moreover, the radial stiffness of the bearing is a nonlinear function of the surface deformation at the contact points and the radial load, which also varies from case to case and time to time. The contact stiffness is a nonlinear time varying (periodic) function of the relative displacement of the contact surfaces and the temporal location of the rolling elements measured with respect to radial location of the accelerometer. In this study, the contact surfaces of the healthy bearing are assumed to be uniform and a linearized version of the contact stiffness is used in the model. However, for improved numerical simulation performance and a better match with real vibration data, the nonlinear and periodic stiffness was incorporated into the model and the results are given in [9].

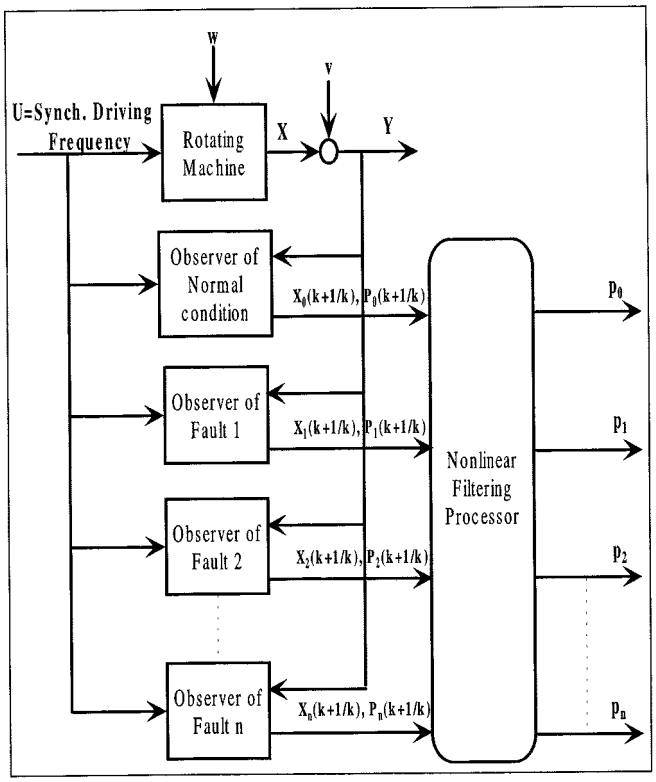


Fig. 3. Multiple-model detection system.

IV. Detection System Design and Implementation

We begin by considering the nonlinear rub impact dynamical system. In this model, certain faults occur due to changes in the model parameters, which, in turn, cause the dynamic behavior of the states x and y to change.

The objective is to monitor the onset and progression of a fault given only the measurement y. Considering the bearing as a local sensor of a nonlinear rotating machine, we can detect faults that occur in the sensor itself (the bearing) or in a remote component of the rotating machine (e.g., another bearing support or station. For rub impact, two types of faults are considered:

1) *external faults* (i.e., external from the bearing being monitored), for example, a change in the unbalance e (for normal operation $e \approx 0$), or a change in the static load l applied to the system;
2) *internal faults* in the sensor (bearing), for example, a change in either the damping ξ (e.g., opening of a seal in a rotating element) or the stiffness K_s (e.g., progression of a crack in a shaft).

A nonlinear filtering approach to fault detection is used. Fault diagnosis is based on statistical testing of the innovations (residuals) of a bank of stochastic nonlinear observers. Fault detection and diagnosis are obtained simultaneously using hypothesis testing techniques. A block diagram of the diagnostic system is given in Fig. 3. The innovations of the various filters are monitored and the conditional probability that each filter model is the process model is computed and in the simplest detection schemes, the filter with the highest probability is declared to match the current operating condition. Each one of the filters is derived from a specific model of the system, which corresponds to a particular operating condition.

Each nonlinear observer during rub–no-rub operation is given by

$$\dot{\widehat{x}}(t) = A\widehat{x}(t) + Bu(t) + \gamma f(C\widehat{x}(t)) + L(y(t) - C\widehat{x}(t))$$
$$\widehat{y}(t) = C\widehat{x}(t)$$

where

$$\gamma = \begin{cases} 0 & \delta < 0 \\ 1 & \delta \geq 0 \end{cases} \quad \text{and} \quad \delta = \left(\sqrt{\widehat{x}_r^2 + \widehat{y}_r^2} - 1 \right).$$

$\widehat{x}_r$ and $\widehat{y}_r$ are the estimated outputs of the nonlinear observer and the clearance has been normalized to the value of one.

Here, $\widehat{x}(t) = E\{x(t)/Y_t\}$, $\widehat{y}(t)$ are the state and the output of the observer, respectively, where $Y_t = \sigma\{y(s), 0 \leq s \leq t\} \equiv$ the sigma-algebra generated by $y(s)$, $0 \leq s \leq t$. The nonlinear observer is a locally exponentially stable unbiased state estimator with uniformly bounded variance. A detailed proof is given in [10].

For the detection and diagnosis of raceway faults in rolling element bearings, we use failure detection filters, a popular model-based technique for actuator and sensor fault detection. A linear time-invariant model of the system given in Fig. 2 can be written in the general form

$$\dot{x} = Ax + Bv + F_i u_i$$
$$y = Cx + Ev + D_i u_i.$$

From this point on, $i = 1, 2$ or subscripts "in" and "out" denote inner and outer race faults, respectively. B and E represent the input maps of known inputs to the state and output spaces, respectively. Also, u_i are excitation inputs due to displacements caused by faults and v represents the known inputs to the system.

In the context of detection filter design, a *fault space* is a subspace of the state space that the inputs are mapped to by the failure maps. The fault spaces are denoted by $\boldsymbol{F_i}$s and the failure maps (matrices) by F_is. The unknown sources of excitation to the rotor-bearing system are the excitations due to the bearing defects. Other sources of excitation such as mass unbalance are treated as known inputs for this discussion. Detection of mass unbalance can be accomplished using the multiple model scheme discussed previously. The full observer is given by

$$\dot{\hat{x}} = (A + LC)\hat{x} + Bv - Ly$$
$$\hat{y} = C\hat{x} + Ev.$$

Note that, because the observer is based on the dynamics of the healthy system, the fault inputs are not included. L represents the output injection part of the observer. The error dynamics are given by

$$\dot{e}_i = (A + LC)e_i - F_i u_i$$
$$y_i - \hat{y}_i = Ce_i + D_i u_i.$$

If (C, A) is observable and L is such that $A + LC$ is stable, then in the absence of disturbances and modeling errors in

steady state, the residual is nonzero only if the failure u_i is nonzero. It follows that in the presence of *a single* fault *any* stable observer is able to detect it, by simply monitoring the residual. When the residual does not converge to zero, a fault has occurred. A more difficult task, diagnosis, is to determine which fault has occurred.

An observer designed such that it is able to separate faults is called a detection filter. Specifically, a detection filter is an observer with the property that, when a fault input is nonzero, the error $e(t)$ remains in a (C, A)-invariant subspace W_i which contains the reachable subspace of $(A + LC, F_i)$. Hence, the residual remains in the output subspace CW_i. Furthermore, the CW_is need to be independent for fault isolation.

In order to use this approach for the fault detection and isolation of rolling element bearings, we need to redefine the output matrix, C. In order to separate n faults we need n isolated output subspaces CW_i. Because the fault space corresponding to a ball fault is dependent on the inner and outer raceway fault spaces, any subspace that includes the ball fault space will have a nonempty intersection with any subspace that includes the inner and outer fault spaces. As a result of this, we will focus our attention only on inner and outer raceway faults.

Focusing on inner and outer raceway faults, the output matrices (C and D) are shown at the bottom of the page.

This output matrix represents two measurements of the rotor-bearing system: the vibration of the rotor (the first row of C) and the vibration of the casing of the bearing (the second row of C).

The detection problem is to find a set of subspaces W_i $i = 1, 2, \cdots, q$, where q is the number of possible faults such that for some L the following conditions (detectability conditions) are met [5]:

$(A+LC)W_i \subseteq W_i$	subspace invariance
$\boldsymbol{F_i} \subseteq W_i$	fault inclusion
$CW_i \cap (CW_j) = 0,$ for all $j = 1, 2, \cdots, q$ and $j \neq i$	output separability
$Real(\sigma(A+LC)) \in R^-$	stability.

The output separability condition is satisfied for the fault spaces of inner and outer raceway faults. That is, $CF_i \cap (CF_j) = 0$, for $j = 1, 2$ and $j \neq i$.

The detectability conditions can be rewritten with W_i replaced by F_i where the one-dimensional subspace $F_i = \text{span}\{F_i\}$. The detectability conditions can be rewritten as

$(A+LC)\boldsymbol{F_i} \subseteq \boldsymbol{F_i}$	subspace invariance
$C\boldsymbol{F_i} \cap (C\boldsymbol{F_j}) = 0$	output separability, $j = 1, 2, \quad j \neq i$
$Real(\sigma(A+LC)) \in R^-$	stability.

In order to satisfy subspace invariance, select L so that the F_is are eigenvectors of the closed-loop system matrix $(A+LC)$. At the same time, L needs to be chosen such that $(A + LC)$ is a Hurwitz (stable) matrix.

The left eigenstructure assignment method can be used to select L, i.e.,

$$(A + LC)F_i = \lambda_i F_i, \qquad i = 1, 2$$

where λ_is are negative real values corresponding to the eigenvalues of the closed-loop system $(A+LC)$. The λ_is are chosen to be real because they correspond to the real eigenvectors F_is. A solution is obtained using a left inverse of CF_i. In order to accommodate both fault cases in the design, the λ_is are chosen as distinct negative real numbers and a solution for L is obtained for $i = 1, 2$ according to

$$L = -(A[F_1\ F_2] - [\lambda_1 F_1\ \lambda_2 F_2])(C[F_1\ F_2])^{-1}.$$

After solving for L, we need to check the stability of the closed-loop system $(A+LC)$. The steady-state error when the input is constant is given by

$$y_i - \hat{y}_i = \left(-\frac{1}{\lambda_i} CF_i + D_i\right) u_i.$$

In the case of no fault, the error dynamics reduce to

$$\dot{e} = (A + LC)e$$
$$y - \hat{y} = Ce.$$

Because the closed-loop system $(A + LC)$ is stable, the error will converge to zero.

In the presence of noise, a rank condition to determine the linear independence of the failure directions is not practical. Therefore, a robust technique capable of extracting these predetermined directions is required. Inspired by the concepts of sliding-mode observers and controllers [11], [12], the theory of sliding-mode detectors has been developed [9]. Sliding-mode detectors are used in conjunction with detection filters to continuously monitor the behavior of the residuals in the error space of the detection filters. The sliding surfaces are positioned according to the fault subspaces (i.e., fault directions). If the inputs to the sliding-mode detector (the detection filter residuals) are in the same direction as the sliding surface, the trajectory of the sliding-mode detector will remain on the sliding surface. On the other hand, when the input is not on the sliding surface, the trajectory will not necessarily stay on the sliding surface. This property provides a means for determining the existence of a fault (detection) and the monitoring of the fault condition.

$$C = \begin{bmatrix} -k_s/m_r & -c_s/m_r & k_s/m_r & c_s/m_r & 0 & 0 \\ 0 & 0 & k_c/m_h & c_c/m_h & (-k_h - k_c)/m_h & (-c_h - c_c)/m_h \end{bmatrix}$$

$$D_{\text{in}} = \begin{bmatrix} k_s/m_r \\ 0 \end{bmatrix}$$

$$D_{\text{out}} = \begin{bmatrix} 0 \\ k_c/m_h \end{bmatrix}$$

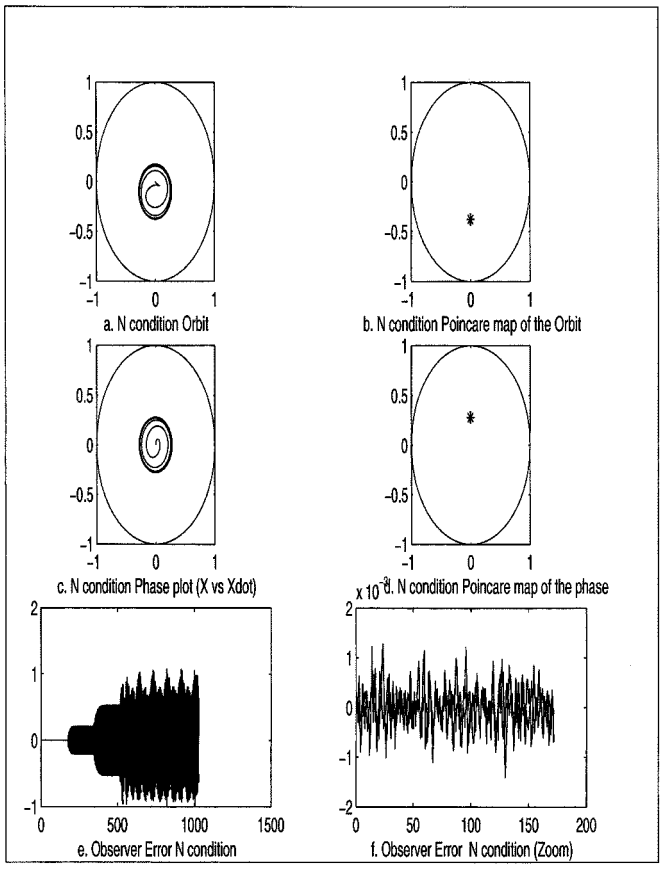


Fig. 4. Stochastic observer for normal operating condition.

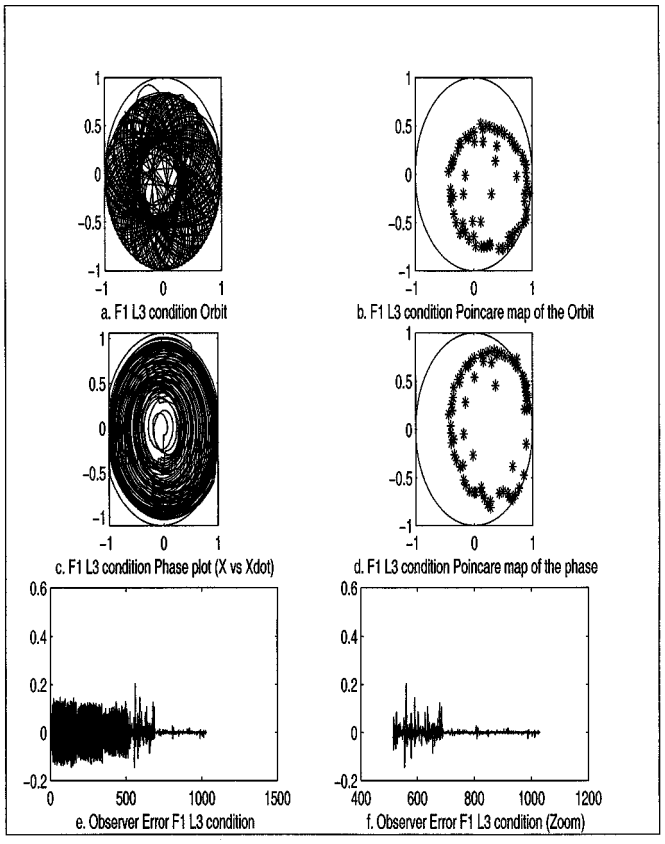


Fig. 5. Stochastic observer for Fault 1 Level 3 operating condition.

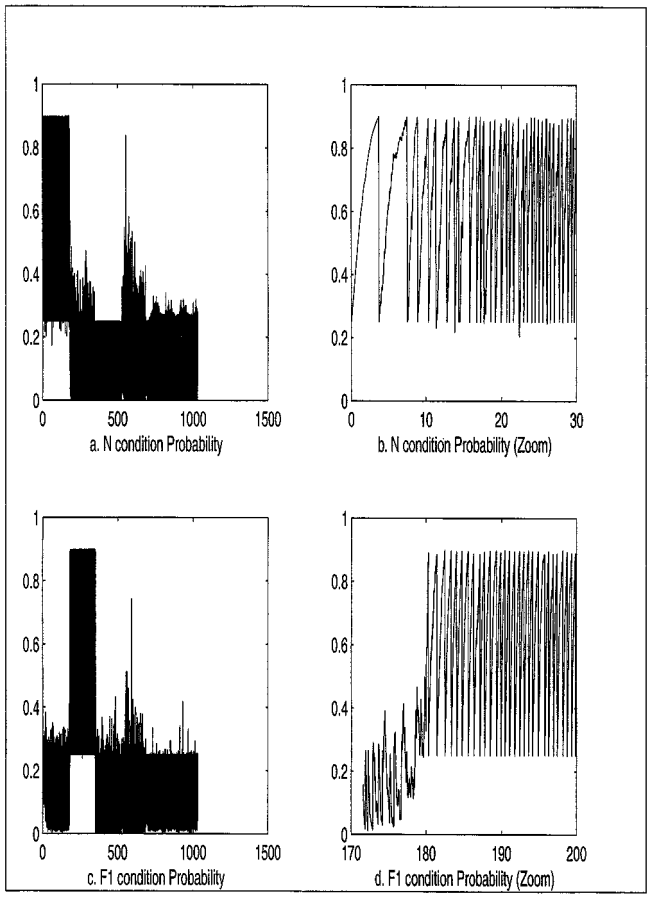


Fig. 6. Probability for Normal and Fault 1 operating conditions.

The sliding-mode detector consists of multiple detection units, each corresponding to particular fault. The free dynamics of a detection unit are given by

$$\dot{x}_1 = x_2$$
$$\dot{x}_2 = -ax_1 - bx_2.$$

The parameters a and b are chosen such that the equilibrium of the model (0, 0) is asymptotically stable.

The sliding surface is chosen to be a one-dimensional subspace passing through the origin. The sliding-mode structure for the detection units is given by

$$\dot{x}_1 = x_2 - k_1 \operatorname{sgn}(a_1x_1 + a_2x_2)$$
$$\dot{x}_2 = -ax_1 - bx_2 - k_2 \operatorname{sgn}(a_1x_1 + a_2x_2)$$

where $\operatorname{sgn}()$ is the signum function and the line $a_1x_1 + a_2x_2 = 0$ is the sliding surface representing one of the fault directions in the output space. The ks are chosen such that when the trajectory is on one side of the switching surface, the corresponding equilibrium is on the opposite side. This makes the sliding surface attractive and with the asymptotically stable free dynamics this guarantees that the trajectory would reach the sliding surface in finite time. The values of the ks determine the rate at which the trajectories approach the sliding surface. Appropriate values for these parameters (ks) are determined to provide a compromise between the rate at which trajectories approach and deflect from the sliding surface.

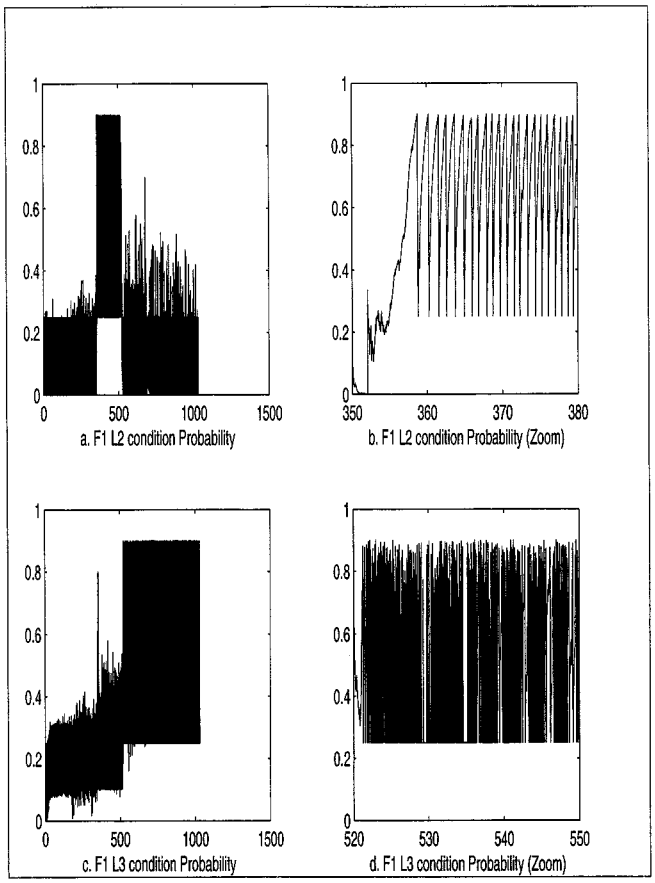


Fig. 7. Probability for Fault 1 Level 2 and Fault 1 Level 3 operating conditions.

The detection units are combined into a sliding-mode detector. The individual detection units for the inner and outer raceway faults incorporate switching (sliding) surfaces positioned at the projected fault directions $CF_i - D_i$. Without loss of generality after scaling, the vectors representing the fault directions, were found to be (1, 6) and (1, 2) for the outer and inner raceway faults, respectively.

V. Simulation Results

To test the proposed algorithms, simulation experiments for the rub-impact model and the rolling element bearing model have been conducted. We begin with a discussion of rub impact given a single fault with different levels. The simulation includes Fault 1, which is a loss of damping with three different levels. The simulation begins with the rotating machine in the normal operating mode with the following set of parameters for the normalized model developed earlier: $w = 1$, $e = 0.1$, $\zeta = 0.18$, $k = 1$, and $g = 0.1$. After operating in the normal mode, Fault 1 is introduced by reducing the damping in the system to $\zeta = 0.1$. Then, Fault 1 Level 2 (F1 L2) is introduced by further reducing the damping to $\zeta = 0.06$. Finally, Fault 1 Level 3 (F1 L3) is introduced by letting the damping drop to $\zeta = 0.02$. The normal operating mode lasts for 170 s (normalized time), i.e., 26 revolutions, F1 and F1 L2 each last for 26 revolutions, and F1 L3 lasts for 80 revolutions. When the damping reduces to 0.02 (F1 L3), rubbing begins and quasi-periodic motion ensues. Multiple-model filters for selected modes of operation, which include the Normal operating condition (N), Fault 1 condition (F1), Fault 1 Level 2 condition (F1 L2), and Fault 1 Level 3 condition (F1 L3) are implemented. The performance of the observers, their sensitivity to different signal-to-noise ratios (SNRs), and their robustness to uncertainty in the modeled dynamics are also investigated. The signal applied to the observers has an SNR between -5 dB and 5 dB, to model the influence of sensor noise. The last observer for F1 L3 has α (the deformation exponent for the Hertzian restoring force) $= 1.0$, and μ (the coefficient of coulomb friction) $= 0.1$, while the simulated process has $\alpha = 1.5$ and $\mu = 0.3$. Hence, there is a mismatch between the fault observers and the simulated process, a situation that will always exist in applications.

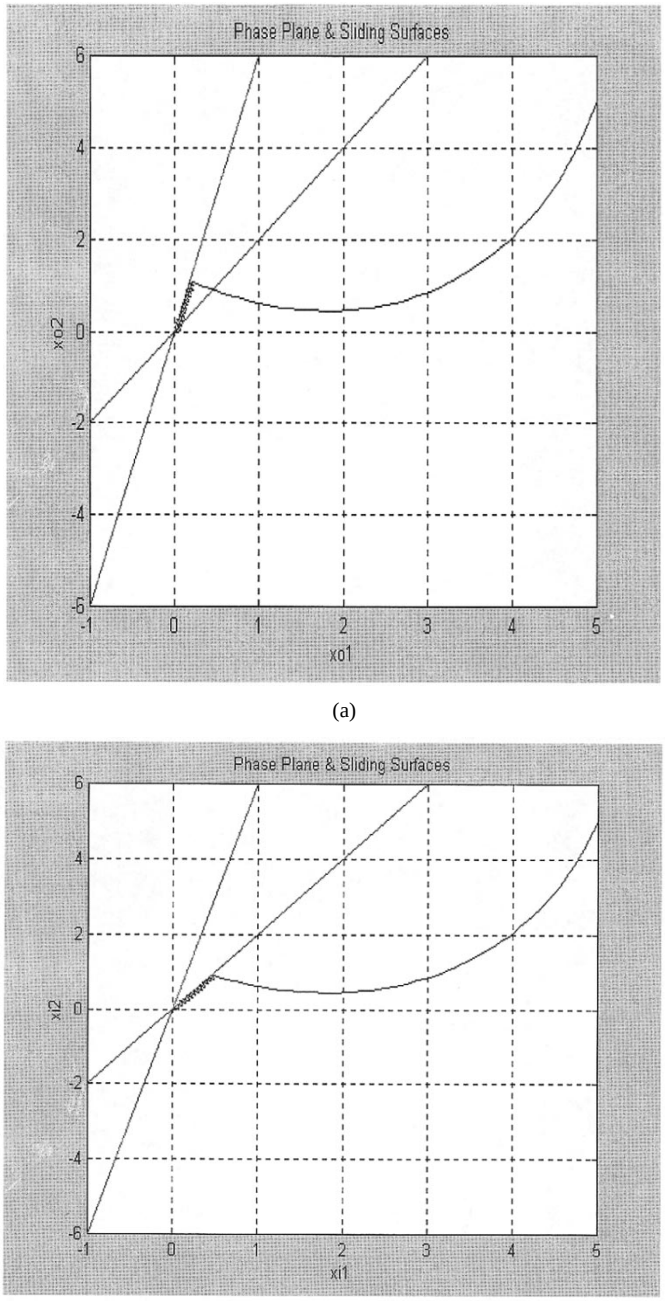
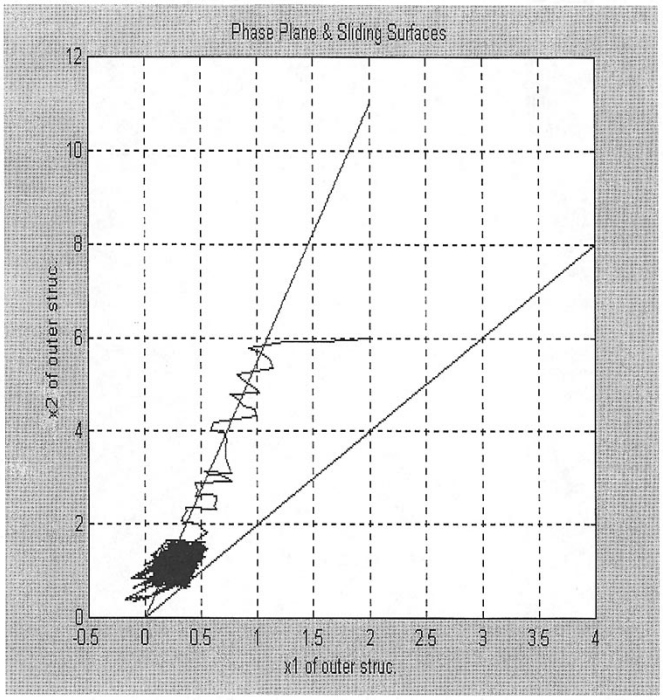
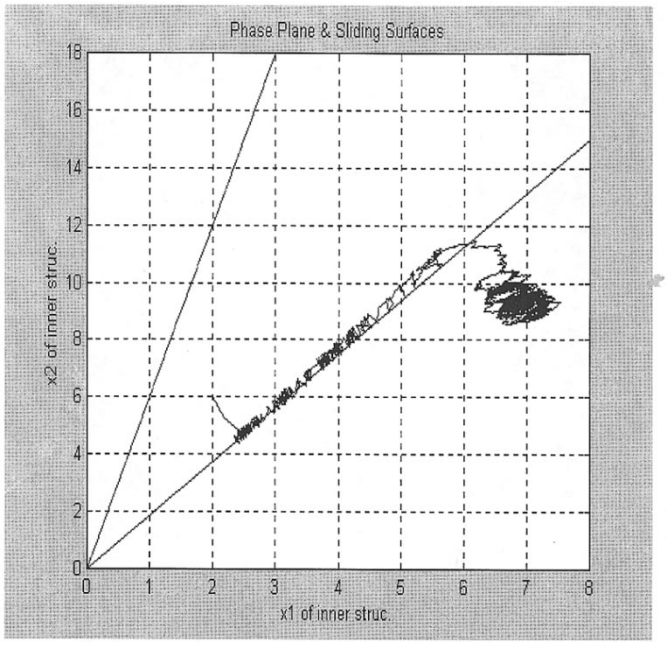


Fig. 8. (a) Outer raceway sliding model (healthy case). (b) Inner raceway sliding model (healthy case).

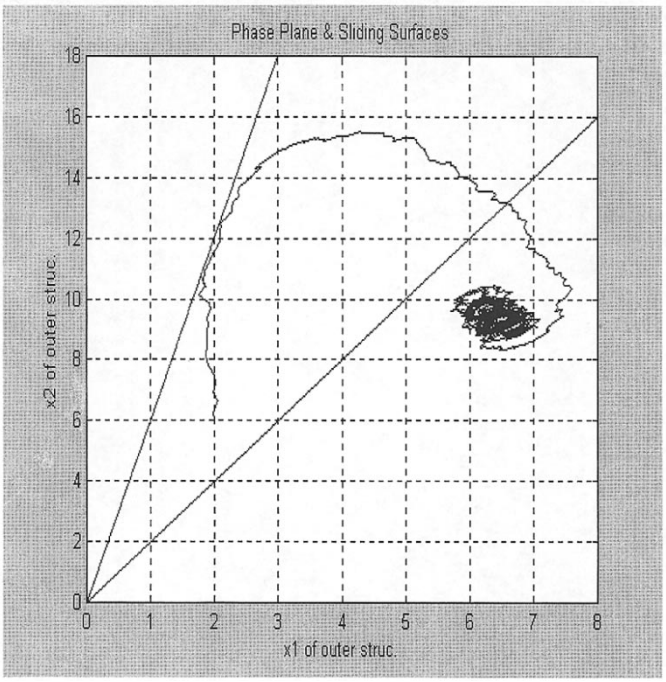


(a)

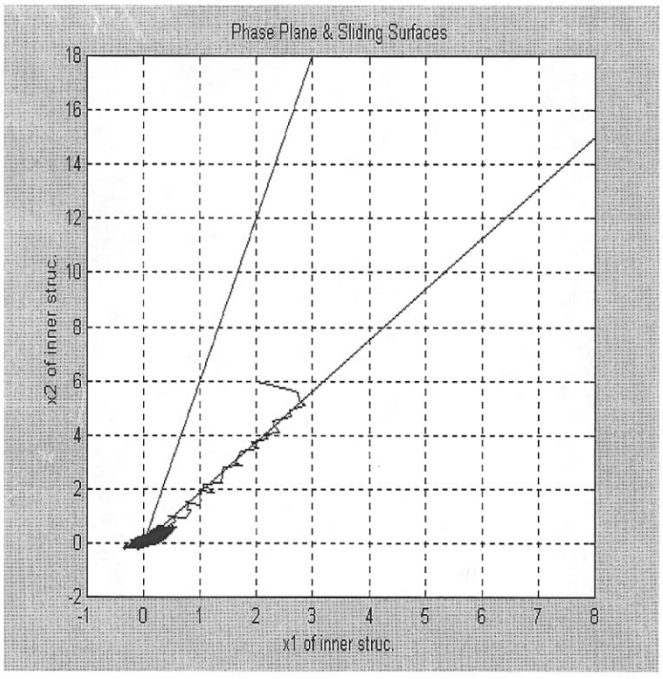


(b)

Fig. 9. (a) Outer raceway sliding model excited by outer raceway fault. (b) Inner raceway sliding model excited by outer raceway fault.



(a)



(b)

Fig. 10. (a) Outer raceway sliding model excited by inner raceway fault. (b) Inner raceway sliding model excited by inner raceway fault.

Fig. 4 shows the orbit plot, the Poincaré map of the orbit plot, the phase plot, the Poincaré map of the phase plot, and the error dynamics for the normal operating condition (N) as an example for the performance of the linear observers. The performance of the Kalman filters in the linear (normal operating) regime is evident. It is also clear from Fig. 5 that the stochastic nonlinear observers, in the presence of sensor noise and unmodeled dynamics, are capable of accurately reconstructing the nonlinear orbit, the Poincaré map, and the phase plot. The nonlinear observer error is given in Fig. 5(e) and (f). The F1 L3 observer error begins to converge at $t = 520$ s, indicating that this observer is tracking the true operating condition of the process. The outputs of the fault observers are used as inputs to a nonlinear post processor, where the conditional probability for each operating mode is computed. Figs. 6(a)–(d) and 7(a)–(d) illustrate the operation of the detection system for the N, F1, F1 L2, and F1 L3 operating modes, respectively. The conditional probability of each operating mode takes values in the interval [0, 1], and the conditional probabilities are constrained to sum to

one. An operating mode is detected when a conditional probability exceeds a given threshold, 0.9 in the examples given in this paper. As is evident from the simulations, once a conditional probability exceeds the threshold and the operating mode of the process is determined, all of the fault filters are reinitialized and the mode probabilities are reset to their initial values (0.25 for this set of simulations). An advantage of the nonlinear (multiple-model) filtering approach to fault detection discussed in this paper is that mean time between failure (MTBF) statistics can be used to reinitialize the mode probabilities after resetting. This can help the detection system decide between two "similar" faults based on the *a priori* probability of their occurrence. The resetting also prevents "windup" in the fault filters and enhances their ability to track abrupt changes (faults) in the operating mode of the system.

Next, we discuss the detection and diagnosis of inner and outer raceway faults in rolling element (ball) bearings. The detection units include switching (sliding) surfaces that are positioned at the projected fault directions, $CF_i - D_i$. After scaling, the vectors representing the fault directions were found to be (1, 6) and (1, 2) for the outer and inner raceway faults, respectively.

Fig. 8(a) and (b) illustrates the free response of the sliding-mode detection units for outer and inner raceway faults, respectively. The trajectories of both detection units chatter to the origin (without deflection) once the sliding surface is reached. Fig . 8(a) and (b) is typical of a healthy rotor-bearing system where fault excitations are absent. In this case, it is expected that the trajectories of each sliding mode detector will converge to the origin along their respective sliding directions.

Figs. 9(a) and (b) and 10(a) and (b) show the behavior of the sliding-mode detectors when driven by the scaled outer and inner raceway fault residuals from the detection filter, respectively. The trajectory of the matching fault model stays along the sliding surface of that fault model, where the trajectory of the other model tends to diverge from the surface, eventually. This model-based scheme is capable of detecting and diagnosing inner and outer raceway faults in rolling element bearings. Because a ball fault can be represented as a linear combination of an inner and an outer raceway fault, the proposed scheme is also capable of detecting this fault condition. However, it will not be possible to distinguish a ball fault from a combined inner and outer raceway fault condition. Finally, although not thoroughly tested, because the sliding-mode detectors are invariant to disturbances along their respective sliding surfaces, this model-based approach to roller bearing fault detection should be robust to modeling errors, sensor noise, and unknown disturbances.

VI. Conclusions

This paper has presented model-based techniques for the detection and diagnosis of rotating machinery faults. A nonlinear filtering approach was developed for a rub-impact model of a rotating machine where unbalance, changes in the stiffness, and damping of the rotor and bearing system, etc., can lead to a complex nonlinear vibration response. The statistical filter incorporates a nonlinear post processor that generates conditional probabilities that track the evolution of the fault modes in the system. Simulations were used to illustrate the performance of the statistical filters in a low signal-to-noise measurement environment. Also, a fault detection filter in combination with a sliding-mode detector is used to detect and isolate inner and outer raceway faults in rolling element bearings. Simulation studies were used to demonstrate the performance of the proposed detection and diagnosis schemes. Results from laboratory and field testing of the techniques developed in this paper are currently underway and will be reported in a subsequent publication.

References

[1] E. Chow and A. S. Wilsky, "Analytical redundancy and the design of robust failure detection systems," *IEEE Trans. Automat. Contr.*, vol. AC-29, pp. 603–614, July 1984.

[2] R. Patton, P. M. Frank, and R. Clark, *Fault Diagnosis in Dynamic Systems*. Englewood Cliffs, NJ: Prentice-Hall, 1989.

[3] P. M. Frank, "Evaluation of analytical redundancy for fault diagnosis in dynamic systems," in *Proc. AIPAC*, vol. 1, 1989, pp. 7–21.

[4] J. Gertler, "Survey of model-based failure detection and isolation in complex plants," *IEEE Contr. Syst. Mag.*, vol. 13, pp. 3–11, Dec. 1993.

[5] V. Krishnaswami and G. Rizzoni, "Model based health monitoring of vehicle steering system using sliding mode observer," in *Proc. American Control Conf.*, Seattle, WA, June 1995, pp. 1652–1656.

[6] V. Garg and J. K. Hedrik, "Fault detection filters for a class of nonlinear systems," in *Proc. American Control Conf.*, Seattle, WA, June 1995, pp. 1647–1651.

[7] S. Nowakowski, M. Boutayeb, and M. Darouach, "A new failure detection and correction method for nonlinear systems," in *Proc. IEEE Int. Conf. Systems, Man, and Cybernetics*, vol. 3, 1993, pp. 424–428.

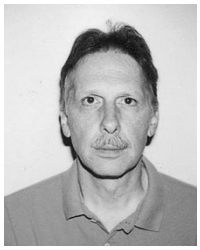
[8] R. Iserman, "Process fault detection based on modeling and estimation methods—A survey," *Automatica*, vol. 20, pp. 378–404, 1984.

[9] N. Afshari, "Model based techniques for real-time fault detection of rolling element bearings," Ph.D. dissertation, Dep. Elect. Eng. Comput. Sci., Case Western Reserve University, 1998.

[10] M. F. Abdel-Magied, "Fault detection of rotating machinery using model-based techniques," Ph.D. dissertation, Dep. Elect. Eng. Comput. Sci., Case Western Reserve Univ., Cleveland, OH, 1997.

[11] J.-J. Slotine, "On sliding observers for nonlinear systems," *J. Dynam. Syst., Meas., Contr.*, vol. 109, pp. 247–252, Sept. 1987.

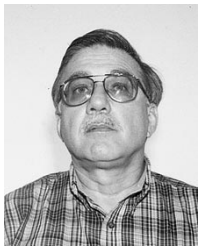
[12] R. A. DeCarlo, "Variable structure sliding mode controller design," in *CRC Control Systems Handbook*. Boca Raton, FL: CRC Press, 1995.



Kenneth A. Loparo (S'75–M'77–SM'89–F'99) received the Ph.D. degree in systems and control engineering from Case Western Reserve University, Cleveland, OH, in 1977.

He was an Assistant Professor in the Mechanical Engineering Department, Cleveland State University, Cleveland, OH, from 1977 to 1979. He has been a member of the faculty of The Case School of Engineering, Case Western Reserve University, since 1979. He is a Professor of Electrical Engineering and Computer Science and holds academic appointments in the Departments of Mechanical and Aerospace Engineering and Mathematics. He was Associate Dean of Engineering from 1994 to 1997 and Chair of the Department of Systems Engineering from 1990 to 1994. His research interests include stability and control of nonlinear and stochastic systems with applications to large-scale electric power systems, nonlinear filtering with applications to monitoring, fault detection, diagnosis, and reconfigurable control, and information theory aspects of stochastic and quantized systems with applications to adaptive and dual control and the design of digital control systems.

Dr. Loparo has held numerous positions in the IEEE Control Systems Society, including Chair of the Conference Audit and Finance Committees, Member of the Board of Governors, Member of the Conference Editorial Board and Technical Activities Board, and Associate Editor of the IEEE TRANSACTIONS ON AUTOMATIC CONTROL and *IEEE Control Systems Magazine*. He has received numerous awards, including the Sigma Xi Research Award for contributions to stochastic control, the John S. Diekoff Award for Distinguished Graduate Teaching, the Tau Beta Pi Outstanding Engineering and Science Professor Award, the Undergraduate Teaching Excellence Award, and the Carl F. Wittke Award for Distinguished Undergraduate Teaching.



M. L. Adams received the B.S.M.E. degree from Lehigh University, Bethlehem, PA, the M.S. degree from The Pennsylvania State University, University Park, and the Ph.D. degree from the University of Pittsburgh, Pittsburgh, PA, in 1963, 1970, and 1977, respectively.

He is a Professor of Mechanical and Aerospace Engineering, Case Western Reserve University, Cleveland, OH. He has spent more than 35 years working in rotating machinery engineering, the first 15 years in industry with Allis–Chalmers, Worthington, Franklin Institute Research Laboratories, and Westinghouse R&D Center.

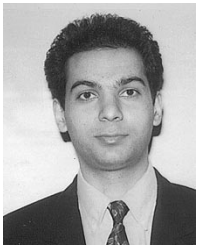
Wei Lin (S'91–M'94–SM'99) received the B.S. degree from Dalian Institute of Technology, Dalian, China, and the M.S. degree from Huazhong University of Science and Technology, Wuhan, China, both in electrical engineering. He also received the M.S. and D.Sc. degrees in systems science and mathematics from Washington University, St. Louis, MO, in 1991 and 1993, respectively.

From 1986 to 1989, he was a Lecturer in the Department of Mathematics, Fudan University, Shanghai, China. During 1994–1995, he was a Postdoctoral Research Associate in the School of Engineering and Applied Science, Washington University. In 1996, he joined Case Western Reserve University, Cleveland, OH, as an Assistant Professor in the Department of Systems, Control and Industrial Engineering, currently the Department of Electrical Engineering and Computer Science. On October 1, 1998, he was appointed the Warren E. Rupp Assistant Professor of Science and Engineering. His current research interests include nonlinear geometric control theory, dynamic systems, H_∞ and robust control, adaptive control, control of mechanical systems with nonholonomic constraint, discrete-time systems, fault detection, and their applications to robotics and induction motors.

Dr. Lin received the CAREER Award from the U.S. National Science Foundation in 1999. He currently serves as an Associate Editor of the IEEE TRANSACTIONS ON AUTOMATIC CONTROL.

M. Farouk Abdel-Magied received the B.Sc. and M.Sc. degrees from the Faculty of Engineering, Cairo University, Cairo, Egypt, and Ph.D. degree from Case Western Reserve University, Cleveland, OH, in 1991, 1993, and 1997, respectively.

Following receipt of the Ph.D. degree, he held a post-doctoral position in the Systems Engineering Department, Case Western Reserve University, where he was involved in research projects in the areas of fault diagnostics with the Electric Power Research Institute and Reliance Electric. In 1998, he joined the faculty of Cairo University, where he is currently an Assistant Professor in the Electrical Engineering Department within the Automatic Control Group. He has also worked as a Consultant for various companies, including Foxboro. His core technology expertise is in the areas of fault detection, computer control, signal processing, and artificial intelligence and neuro-fuzzy control. He has authored and coauthored several publications in these sareas.



Nadar Afshari received the B.S.E.E. degree from Tehran Azad University, Tehran, Iran, the M.S.E.E. degree from California State Polytechnic University, Pomona, and the Ph.D. degree from Case Western Reserve University, Cleveland, OH, in 1986, 1995, and 1998, respectively.

Since 1998, he has been a member of the Electronics Research and Development Group, Cutler–Hammer/Eaton Corporation, Pittsburgh, PA, where he develops monitoring systems for power distribution systems. His research interests include fault detection and predictive monitoring techniques for rotating machinery and power systems.