BG Pulse Journal USA
ISSN:1947-2536 http://www.bgpulseusa.com/index.php/bg
BG PULSE JOURNAL

A REVIEW OF OBSERVER-BASED FAULT DETECTION AND ISOLATION FOR MECHATRONIC SYSTEMS

Shohjakhon Jonuzokov

University of Birmingham

BEng Mechatronic and Robotic Engineering, Year Three

sxj454@student.bham.ac.uk

Doi: https://doi.org/10.5281/zenodo.22643351

Abstract: Observer-based fault detection and isolation (FDI) exploits the analytical redundancy in a dynamic model of a plant to detect faults without additional hardware, generating a residual that departs from zero only when a fault occurs. This review consolidates the observer-based FDI literature for mechatronic systems, analysing methods within a single state-space fault model to enable direct comparison across notations. The fault detection filter, diagnostic observer, Kalman innovation filter and parity space method are shown to be alternative parameterisations of one residual generator. The central difficulty is the conflict between fault sensitivity and robustness to disturbance and model uncertainty: exact decoupling via unknown input observers is bounded by a rank condition on the sensor set, while approximate decoupling trades that guarantee for generality. An analytical case study on a two-mass servo drive shows that a single rank test determines what any design can achieve. Validation practice remains the field's weakest point.

Keywords: fault detection and isolation, observer-based diagnosis, analytical redundancy, unknown input observer, residual generation, mechatronic systems.

Introduction

Hardware redundancy like duplicate sensors and voting logic, is effective but multiplies mass, cost and the number of components that can themselves fail. Analytical redundancy replaces the second physical measurement with a prediction from a model, and the difference between the two, the *residual*, carries information about the health of the system. Observer-based FDI generates that prediction using a state observer. Its properties can be established analytically before deployment, which is decisive in certified applications where a statistical performance claim on a test set is not an acceptable safety argument.

Mechatronic systems are a natural target. They are modellable from first principles with few states and interpretable parameters, precisely the situation favouring model-based methods over data-driven ones requiring scarce fault data. They are closed-loop, and feedback masks incipient faults just as it rejects disturbances: a controller with integral action silently absorbs a developing actuator degradation until saturation is approached. They are multi-domain, so a shorted winding appears mechanically as torque ripple and bearing wear appears electrically as a current signature — cross-couplings a coupled observer exploits and single-domain monitoring misses.

The literature is large and is fragmented by notation, with each method typically presented in the terms of its originator. The aim of this review is therefore twofold: to place the principal

observer-based methods in a common framework so that their structural relationships become visible, and to establish what the resulting conditions imply for mechatronic plants specifically, which are typically sensor-poor and parameter-uncertain.

Methods

Source selection and analytical approach

This is a theoretical review. No experiments were performed and no experimental results are reported. Sources comprise the foundational control-theoretic literature on analytical redundancy and observer design, together with subsequent work applying that theory to electromechanical plants; monographs and prior surveys were used to establish coverage, and primary papers were consulted for the derivations reproduced here.

Two analytical procedures were applied throughout. Every method was rewritten in the common state-space notation introduced below, so that methods originating in different formalisms could be compared on their assumptions rather than on their presentation. Existence conditions were then evaluated symbolically against a representative mechatronic plant model, which yields a decision on applicability without recourse to simulation or hardware.

Plant and fault model

Faults and disturbances enter the linear plant model as additional input channels:

$$\dot{x}=Ax+Bu+E_d d+E_f f, \qquad y=Cx+Du+F_d d+F_f f$$

with $x\in\mathbb{R}^n$, $u\in\mathbb{R}^p$, $y\in\mathbb{R}^m$. Here f collects the faults, d collects unknown inputs — disturbances, noise, and model uncertainty recast as a fictitious disturbance — and (C,A) is assumed detectable. Discrete-time analogues follow directly and the structural conditions are unchanged.

Three fault locations are distinguished. Actuator faults act on the input, so a loss of effectiveness $\alpha\in[0,1]$ gives $E_f=B$ with $f=-(1-\alpha)u$. Sensor faults act on the output, $F_f=I$. Component faults are parameter changes: $A\to A+\Delta A$ contributes $\Delta A\, x$, written additively with $f=\Delta\theta\cdot x$. The last case is state-dependent, so a stiffness reduction produces no residual at all when the system is at rest, however severe — a structural property of parametric faults that explains why parameter estimation is preferred for component faults while observer-based methods dominate for actuator and sensor faults.

A residual generator is a stable dynamic system driven by u and y that yields zero output in the fault-free case. Under the fault model,

$$r(s)=G_{rf}(s)f(s)+G_{rd}(s)d(s)$$

and the design objectives follow directly: $G_{rf}(s)e_i\not\equiv 0$ for detectability of fault i, $G_{rd}(s)$ small for robustness, $G_{rf}(s)$ large for sensitivity, and the columns of G_{rf} distinguishable for isolability.

Detection requires a decision as well as a residual. With $J(t)=\|r\|_{2,[t-T,t]}$ the alarm condition is $J>J_{th}$, and the threshold is properly chosen as $J_{th}=\sup_{d,x_0,f=0}J(t)$, the supremum over all admissible disturbances and initial conditions in the fault-free case. This guarantees zero false alarms over the assumed disturbance class at the cost of missed detections below the threshold. The threshold is therefore not a tuning knob but the point at which robustness is converted into a quantitative limit on the smallest detectable fault.

Case study plant

Existence conditions are evaluated against a motor driving a load through a compliant shaft, with states $x=[\omega_m, \omega_l, \Delta\theta]^\top$:

$$J_m\dot{\omega}_m=T_m-K_s\Delta\theta-B_s(\omega_m-\omega_l)-b_m\omega_m$$

$$J_l\dot{\omega}_l=K_s\Delta\theta+B_s(\omega_m-\omega_l)-b_l\omega_l-T_L$$

$$\dot{\Delta\theta}=\omega_m-\omega_l$$

with input $u=T_m$ and unknown load torque $d=T_L$, so $B=[1/J_m, 0, 0]^\top$ and $E_d=[0, -1/J_l, 0]^\top$. This model is representative of geared and belt-driven axes generally. Three faults are considered: actuator loss of effectiveness ($E_{f_1}=B$), motor speed sensor bias ($F_{f_2}=e_1$), and increased load-side friction ($E_{f_3}=E_d$, $f_3=\Delta b_l\omega_l$). Two sensor configurations are evaluated, one measuring motor speed alone and one measuring both speeds.

Results

Residual generation

Take the Luenberger observer $\dot{\hat{x}}=A\hat{x}+Bu+L(y-\hat{y})$ and the residual $r=V(y-\hat{y})$. The estimation error $e=x-\hat{x}$ obeys

$$\dot{e}=(A-LC)e+(E_d-LF_d)d+(E_f-LF_f)f$$

giving

$$G_{rf}(s)=VC(sI-A+LC)^{-1}(E_f-LF_f)+VF_f$$

and G_{rd} likewise with E_d,F_d. The design freedom is (L,V). The gain L has a dual role: it sets convergence rate and simultaneously shapes both transfer matrices, so fast convergence is not automatically desirable, since a high-gain observer transmits measurement noise straight to the residual. The appearance of L in the numerator terms is what makes decoupling possible and is the algebraic origin of the conditions given below.

The earliest scheme, the fault detection filter (Beard, 1971; Jones, 1973), chooses L so each fault produces a residual confined to a fixed known direction, giving isolation by direction rather than by comparison; Massoumnia (1986) recast this in terms of unobservability subspaces. Its existence conditions are restrictive and the fault directions drift with the model parameters. Relaxing directionality to require only a zero fault-free residual gives the diagnostic observer, the

form above and by far the most widely used, with reduced-order variants exploiting the directly measured states.

If d is modelled as white noise of known covariance, the Kalman innovation is zero-mean with computable covariance S, permitting principled tests: the χ^2 test, the generalised likelihood ratio test (Willsky and Jones, 1976), which also estimates fault onset and magnitude, and CUSUM, which is far more sensitive to small persistent changes. This converts threshold selection into a calculation, but real disturbances are frequently coloured or input-correlated, and when the noise model is wrong the false-alarm rate departs from nominal, often severely.

The parity space approach reaches the same object by a different route. Choosing a parity vector v in the left null space of the extended observability matrix eliminates the state, and robustness follows by requiring $v^{\top} H_d=0$, a generalised eigenvalue problem with a closed-form solution. Parity and observer-based generators are equivalent: the parity form makes the structural design transparent, the observer form permits frequency shaping. They are two views of one object, not competing schools.

Robustness and disturbance decoupling

Faults and modelling errors both appear as discrepancies between prediction and measurement, so a residual sensitive enough to catch a small fault is, unless designed otherwise, equally sensitive to a small parameter error. Robustness is a precondition for usability, not a refinement.

The unknown input observer (UIO) achieves $G_{rd}\equiv 0$ exactly. With the structure $\dot{z}=Fz+TBu+Ky$, $\hat{x}=z+Hy$, requiring the error to be independent of d gives $(HC-I)E_d=0$, $T=I-HC$ and $F=TA-K_1C$ Hurwitz. Existence requires

$$\operatorname{rank}(CE_d)=\operatorname{rank}(E_d)$$

together with detectability of a reduced pair (Darouach et al., 1994; Chen and Patton, 1999). The rank condition is the binding one and has a clear reading: **the number of disturbances that can be decoupled cannot exceed the number of independent measurements.** This is a hard limit and it is why UIO methods often fail on mechatronic plants, which are sensor-poor — a servo axis with a single encoder can decouple at most one disturbance direction. Eigenstructure assignment (Patton and Chen, 1991) achieves the same result by placing left eigenvectors orthogonal to the disturbance directions, and meets the same obstruction.

When exact decoupling is impossible the design becomes an explicit trade-off between $\| G_{rd} \|_{\infty}$ and the sensitivity index

$$\| G_{rf} \|_{-}=\inf_{\omega\in\Omega}\sigma_{\min}\left(G_{rf}(j\omega)\right)$$

which is a minimum over frequency, since a residual generator is only as good as its response to the fault direction it detects worst. Maximising $\| G_{rf} \|_{-}/\| G_{rd} \|_{\infty}$ is solvable by linear matrix inequalities. The family is general and accommodates frequency weighting naturally, but the resulting gain has no physical interpretation and the design is only as good as an uncertainty weighting for which no principled selection procedure exists.

A third option leaves the residual generator alone and makes the threshold state-dependent. Since the uncertainty contribution scales with the operating point, a fixed threshold must be set for the worst case and is needlessly insensitive elsewhere. A threshold of the form $J_{th}(t)=J_0+\|W(s)u\|+\|V(s)y\|$ expands during transients and contracts at steady state, recovering sensitivity in exactly the quiescent conditions where incipient faults are most visible.

Fault isolation

Detection needs one residual; isolation needs a set. In the structured-residual approach a bank $\{r_1,\ldots,r_N\}$ is designed so each residual is sensitive to a specified subset of faults, summarised by a binary incidence matrix Φ whose columns are the fault signatures. Weak isolability requires the columns to be non-zero and distinct. Strong isolability additionally forbids one column being contained in another, since containment allows a single missed detection to convert one fault's signature into another's and produce confident misdiagnosis.

Insensitivity to a fault is decoupling with the fault treated as an unknown input, so isolability obeys the same rank limits as disturbance decoupling and is bounded by the sensor set. In many applications certain fault pairs are simply not distinguishable with the available instrumentation, and reporting this is the correct response.

For sensor faults two bank structures are canonical. In the dedicated observer scheme each observer is driven by a single output, giving an identity incidence matrix and immediate isolation, but requiring $\left(c_i^{\top},A\right)$ detectable for every i — frequently violated. In the generalised observer scheme each observer is driven by all outputs except one, so a fault in sensor j alarms every residual but r_j. Detectability is required only from $m-1$ outputs, so this applies far more often, at the cost of being unable to isolate simultaneous faults. A third possibility supersedes isolation altogether: fault estimation, reconstructing $f(t)$ itself, which gives magnitude and time profile as well as location.

Nonlinear observers

Linearisation error is exactly the kind of model uncertainty that masks small faults, so nonlinear observers are frequently a necessity for mechatronic diagnosis rather than a refinement.

For $\dot{x}=Ax+Bu+\phi(x,u)$ with ϕ Lipschitz of constant γ, the natural observer converges if $\min_{\omega}\sigma_{\min}(A-LC-j\omega I)>\gamma$ (Rajamani, 1998), an LMI in L. This covers smooth friction and manipulator kinematics, but the Lipschitz constant of Coriolis terms grows with velocity, so a globally valid γ may admit no feasible L.

Sliding mode observers add a discontinuous injection term that establishes a sliding motion on the output error surface in finite time and maintains it despite the fault. The output error is driven to zero, which would appear to destroy the residual, but the equivalent output injection required to sustain the sliding motion is a direct reconstruction of the fault (Edwards et al., 2000). This is qualitatively different from detection: magnitude and time profile are recovered, exactly in the ideal sliding regime rather than asymptotically. The costs are chattering, which interacts badly with sampled-data implementation on embedded hardware, and a switching gain that must upper-bound the fault magnitude — awkward when that magnitude is the unknown of interest.

Adaptive observers estimate states and parameters jointly, addressing the fault class the additive framework handles worst, since the estimated parameter is the diagnostic quantity and carries a physical unit. The critical requirement is persistency of excitation: a system at a constant set point is not persistently exciting, so the estimate drifts precisely in the steady-state condition where the plant spends most of its time.

Takagi–Sugeno and LPV observers represent the plant as a convex combination of local linear models blended by measurable scheduling variables, which lets the entire linear apparatus — UIO conditions, $\mathscr{H}_\infty$ bounds, fault estimation — extend as LMIs in the local gains. Where the nonlinearity is dominated by a measurable variable such as manipulator configuration or rotor speed this is effective and implementable; the main conservatism is the common Lyapunov function.

Case study outcome

With a single sensor, $y=\omega_m$, the product $CE_d=0$ while E_d has rank one, so the rank condition fails and no unknown input observer exists. The reason is physical: load torque acts on the load inertia and reaches the motor only through the compliant shaft, so its effect on ω_m is filtered but not eliminated, and no combination of the single measurement annihilates it. Detection under varying load then requires an adaptive threshold, or a frequency-selective residual exploiting the fact that load disturbances are low-frequency while the faults of interest excite the torsional mode. Note also that $E_{f_3}=E_d$ exactly: the friction fault and the load disturbance share a channel and are structurally indistinguishable. No observer design separates them and no quantity of data would help. The only escapes are added instrumentation or exploiting the state dependence $f_3=\Delta b_l\omega_l$, which correlates with the measured speed while an exogenous T_L does not.

With two sensors, $y=[\omega_m, \omega_l]^\top$, the product $CE_d=[0, -1/J_l]^\top \neq 0$, so $\operatorname{rank}(CE_d)=\operatorname{rank}(E_d)=1$ and the condition is satisfied. The torsional state remains observable from the two speeds, so a UIO exists and a residual completely decoupled from load torque can be built with $H=E_d(CE_d)^+$. Faults f_1 and f_2 remain detectable since neither direction lies in the span of E_d; f_3 is annihilated with the disturbance, and the diagnosis must be reported as a load-side anomaly rather than resolved further.

Comparison of methods

Method	Plant class	Robustness	Isolation	Fault estimate	Key limitation
Luenberger residual	LTI	None inherent	Via bank	No	No disturbance rejection
Fault detection filter	LTI	Geometric	Directional	No	Restrictive existence conditions
Unknown input observer	LTI	Exact decoupling	Via bank	No	Requires $n_d \leq m$
Parity space	LTI	Null-space	Structured	No	Deadbeat; no dynamic shaping
Kalman / GLR	LTI stochastic	Statistical	Via bank	Magnitude	Needs valid noise model

Method	Plant class	Robustness	Isolation	Fault estimate	Key limitation
$\mathcal{H}_-/\mathcal{H}_\infty$	Uncertain LTI	Optimised	Via bank	No	Conservative; opaque tuning
Lipschitz observer	Nonlinear	Via LMI	Via bank	No	Fails for large γ
Sliding mode observer	Nonlinear, matched	Insensitivity	Direct	Yes, exact	Chattering; needs fault bound
Adaptive observer	Parametric faults	Adaptation	Direct	Yes, physical	Requires excitation
T–S / LPV observer	Nonlinear, scheduled	LMI	Via bank	Yes	Needs scheduling variable

Discussion

No method dominates

Selection is determined by the plant rather than by preference. If the uncertainty is low-rank and the number of disturbances does not exceed the number of measurements, the UIO gives an exact guarantee that no optimisation-based method can match; if it does, the UIO is simply unavailable and $\mathcal{H}_-/\mathcal{H}_\infty$ becomes the coherent choice. Where the fault is parametric and the system is persistently excited, an adaptive observer gives a physically interpretable estimate that additive methods cannot. The tendency to present each method as general-purpose obscures this.

Structural limits precede design

The rank conditions governing decoupling and isolability are properties of the plant and its sensors, not of the designer's skill. Effort spent on a problem that fails the rank test is wasted, and the correct output of the analysis is a specification for additional instrumentation. The case study demonstrates this concretely: adding a second speed measurement converted an impossible decoupling problem into a solvable one, while one fault remained indistinguishable from the disturbance in both configurations. This reframes part of the diagnostic design problem as a sensor placement problem, and it can be settled analytically before any hardware is committed.

Validation practice is the weakest link

Much published work demonstrates detection of one artificially injected fault of generous magnitude on a clean rig, with a threshold selected after the fact, which establishes very little. Laboratory rigs have new bearings and a benign disturbance environment, so the uncertainty against which robustness is demonstrated is unrepresentatively small, and an injected fault has a sharper onset than the corresponding physical degradation. Credible evaluation requires faults swept to the detection limit, thresholds fixed a priori from a stated uncertainty bound, false-alarm rates over realistic profiles, and sensitivity to parameter error quantified rather than assumed small.

Relation to data-driven diagnosis

The relevant axis of comparison is not accuracy but the availability of prior knowledge: observer-based methods need a model and no fault data, data-driven methods the reverse. Where a model is derivable and fault data scarce — the typical mechatronic situation — the observer-based approach is preferable almost by default, and the common comparison against a classifier trained on abundant simulated faults tests a scenario that does not obtain in practice.

Open problems

The threshold must bound the disturbance-induced residual, so the minimum detectable fault is set by model uncertainty, and faults whose early signature falls below that envelope cannot be caught at useful lead time. Change detection over long horizons exploits the persistence of a fault against the transience of a disturbance and is the most promising direction, but its integration with observer-based residual generation remains underdeveloped.

Parametric fault detection requires excitation while regulation requires its absence. Designing minimal probing signals that achieve identifiability with bounded degradation of the controlled variable is open and industrially relevant.

Most published model-and-data hybrids are shallow — a residual fed to a classifier, or a network trained to imitate an observer. Deeper integration, such as using observer structure as an architectural prior or learning bounded corrections to a physical model, is where new work is most likely to be valuable.

Conclusion

Observer-based FDI provides a rigorous route to diagnosis in mechatronic systems, with residual properties establishable analytically before deployment. Two conclusions stand out. Structural conditions govern what is achievable: rank conditions determine whether disturbances can be decoupled and which faults can be distinguished, and both are properties of the plant and its instrumentation rather than of the designer's skill. And the field's theoretical maturity is not matched by its validation practice; sharper evidential standards would do more for industrial uptake than further proliferation of methods. The most promising direction lies at the boundary with data-driven diagnosis, where the complementarity of the two paradigms is widely acknowledged and largely unexploited.

References

1. Beard, R.V. (1971) *Failure Accommodation in Linear Systems through Self-Reorganization*. PhD thesis, MIT.
2. Blanke, M., Kinnaert, M., Lunze, J. and Staroswiecki, M. (2016) *Diagnosis and Fault-Tolerant Control*. 3rd edn. Springer.
3. Chen, J. and Patton, R.J. (1999) *Robust Model-Based Fault Diagnosis for Dynamic Systems*. Kluwer.

4. Chow, E.Y. and Willsky, A.S. (1984) 'Analytical redundancy and the design of robust failure detection systems', *IEEE Transactions on Automatic Control*, 29(7).
5. Darouach, M., Zasadzinski, M. and Xu, S.J. (1994) 'Full-order observers for linear systems with unknown inputs', *IEEE Transactions on Automatic Control*, 39(3).
6. Ding, S.X. (2013) *Model-Based Fault Diagnosis Techniques*. 2nd edn. Springer.
7. Edwards, C., Spurgeon, S.K. and Patton, R.J. (2000) 'Sliding mode observers for fault detection and isolation', *Automatica*, 36(4).
8. Frank, P.M. (1990) 'Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy', *Automatica*, 26(3).
9. Gao, Z., Cecati, C. and Ding, S.X. (2015) 'A survey of fault diagnosis and fault-tolerant techniques', *IEEE Transactions on Industrial Electronics*, 62(6).
10. Gertler, J. (1998) *Fault Detection and Diagnosis in Engineering Systems*. Marcel Dekker.
11. Isermann, R. (2006) *Fault-Diagnosis Systems*. Springer.
12. Jones, H.L. (1973) *Failure Detection in Linear Systems*. PhD thesis, MIT.
13. Massoumnia, M.A. (1986) 'A geometric approach to the synthesis of failure detection filters', *IEEE Transactions on Automatic Control*, 31(9).
14. Patton, R.J. and Chen, J. (1991) 'Robust fault detection using eigenstructure assignment', *Proc. IFAC/IMACS Symposium SAFEPROCESS*.
15. Rajamani, R. (1998) 'Observers for Lipschitz nonlinear systems', *IEEE Transactions on Automatic Control*, 43(3).
16. Willsky, A.S. and Jones, H.L. (1976) 'A generalized likelihood ratio approach to the detection and estimation of jumps in linear systems', *IEEE Transactions on Automatic Control*, 21(1).