International Journal of Automation and Computing 12(1), February 2015, 50-61
DOI: 10.1007/s11633-014-0865-7

Observer Design — A Survey

Chia-Chi Tsui
743 Clove Road, Staten Island, NY 10310, USA

Abstract: This paper surveys the results of observer design for linear time-invariant (L-T-I) deterministic irreducible open-loop systems (OLS), the most basic type of OLS. An observer estimates $K\boldsymbol{x}(t)$ signal where K is a constant and $\boldsymbol{x}(t)$ is the state vector of the OLS. Thus, an observer can be used as a feedback controller that implements state feedback control (SFC) or $K\boldsymbol{x}(t)$-control, and observer design is therefore utterly important in all feedback control designs of state space theory. In this survey, the observer design results are divided into three categories and for three respective main purposes. The first category of observers estimate signal $K\boldsymbol{x}(t)$ only with a given K, and this survey has four conclusions: 1) Function observer that estimates $K\boldsymbol{x}(t)$ directly is more general than state observer that estimates $\boldsymbol{x}(t)$, and may be designed with order lower than that of state observer, and the additional design objective is to minimize observer order; 2) The function observer design problem has already been simplified to the solving of a single set of linear equations only while seeking the lowest possible number of rows of the solution matrix, and an apparently most effective and general algorithm of solving such a problem can guarantee unified upper and lower bounds of the observer order; 3) Because such a single set of linear equations is the simplest possible theoretical formulation of the design problem and such theoretical observer order bounds are the lowest possible, and because the general, simple, and explicit theoretical formula for the function observer order itself do not exist, the theoretical part of this design problem is solved; 4) Because the function observer order is generically near its upper bound, further improvement on the computational design algorithm so that the corresponding observer order can be further reduced, is generically not worthwhile. The second category of observers further realize the loop transfer function and robustness properties of the direct SFC, and the conclusion of this survey is also fourfold: 1) To fully realize the loop transfer function of a practically designed $K\boldsymbol{x}(t)$-control, the observer must be an output feedback controller (OFC) which has zero gain to OLS input; 2) If parameter K is separately designed before the observer design, as in the separation principle which has been followed by almost all people for over half of a century, then OFC that estimates $K\boldsymbol{x}(t)$ does not exist for almost all OLS′s; 3) As a result, a synthesized design principle that designs an OFC first and is valid for almost all OLS′s, is proposed and fully developed, the corresponding K will be designed afterwards and will be constrained by the OFC order as well as the OFC parameters; 4) Although the $K\boldsymbol{x}(t)$-control is constrained in this new design principle and is therefore called the "generalized SFC" (as compared to the existing SFC in which K is unconstrained), it is still strong enough for most OLS′s and this new design principle overcomes many fundamental drawbacks of the existing separation principle. The third category of observers estimate $K\boldsymbol{x}(t)$ signal at special applications such as fault detection and identification and systems with time delay effects. Using directly the result of OFC that estimates $K\boldsymbol{x}(t)$ of the second category, these observers can be generally and satisfactorily designed.

Keywords: Low order function observer, robust output feedback observer, synthesized design principle, faults and time-delays.

1 Why need an observer to fulfill the stated tasks

1.1 System and observer definitions

State space control theory is based on the state space model of the system. For a linear time-invariant (L-T-I) deterministic system, its state space model is

$$\frac{\mathrm{d}\boldsymbol{x}(t)}{\mathrm{d}t} = A\boldsymbol{x}(t) + B\boldsymbol{u}(t) \tag{1a}$$

$$\boldsymbol{y}(t) = C\boldsymbol{x}(t) \tag{1b}$$

where system state $\boldsymbol{x}(t)$, control input $\boldsymbol{u}(t)$ and output measurement $\boldsymbol{y}(t)$ are vectors of dimensions n, p and m respectively, and system parameters (A, B, C) are real constant matrices with appropriate dimensions. For observer design problems, the system is assumed controllable and observable, or is assumed irreducible[1]. Equation (1a) describes the dynamic part of system (1) that determines $\boldsymbol{x}(t)$ from its initial condition and control input $\boldsymbol{u}(t)$, while (1b) describes the static relation between $\boldsymbol{x}(t)$ and output $\boldsymbol{y}(t)$.

System order n is usually much higher than dimensions p and m. For example a challenging circuit system can have over 100 capacitors, but much less controlled voltage/current sources and voltage/current sensors. Furthermore, m is usually no less than p, because adding sensors is usually much easier than adding dynamic control inputs.

State space model (A, B, C) provides much more information about the system, especially its detail on the internal structure that involves with $\boldsymbol{x}(t)$, than the transfer function model $G(s)$ of the same system. For example, only the polynomial matrix fraction description (MFD) form of $G(s)$ can be used effectively in design[1−3], while this form of $G(s)$ has a one-to-one parametric relation with a special canonical form (with state matrix $\underline{A}$) of state space model[1, 4, 5]. The numbers of unknown parameters of state matrices A and $\underline{A}$ are $n \times n$ and only (m or p) $\times n$, respectively.

State space model is usually much more reliable and never less reliable, than the transfer function model (in

Survey paper
Manuscript received August 12, 2014; accepted September 25, 2014
Recommended by Editor-in-Chief Guo-Ping Liu

polynomial MFD form) of the same system, because the transformation from A to $\underline{A}$ implies parameter compression from $n \times n$ to (m or p)$\times n$, and the computation of this parameter compression is usually and intrinsically very unreliable[6, 7].

An observer is also an L-T-I system with the most general state space model

$$\frac{\mathrm{d}\boldsymbol{z}(t)}{\mathrm{d}t} = F\boldsymbol{z}(t) + L\boldsymbol{y}(t) + TB\boldsymbol{u}(t) \tag{2a}$$

$$\boldsymbol{\omega}(t) = K_z\boldsymbol{z}(t) + K_y\boldsymbol{y}(t) \tag{2b}$$

and with order r. All observer parameters are free to design except B of (2a) and (1a). An observer is proper or strictly proper if gain K_y in (2b) is none-zero or zero, respectively[1]. It is required in almost all feedback control applications that observer output $\boldsymbol{\omega}(t)$ converges to $K\boldsymbol{x}(t)$ where K is a constant, and be fed into the control input $\boldsymbol{u}(t)$ of system (1). Observers do not have a gain to control input $\boldsymbol{u}(t)$ in (2b) in all feedback control applications. Observer (2) is the most general feedback controller of state space control theory that implements $K\boldsymbol{x}(t)$-control.

1.2 Why need to estimate $K\boldsymbol{x}(t)$ signal

State feedback control (SFC) is defined as $\boldsymbol{u}(t) = -K\boldsymbol{x}(t)$, where K is a constant. The corresponding feedback system state matrix is $A - BK$. Hence an SFC is based on the best information of current system state ($\boldsymbol{x}(t)$), and on the best information of system model and structure (Subsection 1.1). Notice that these advantages are valid even if the constant gain K is generally constrained by $K = \underline{KC}$ ($\underline{K}$ is free but rank (given $\underline{C}$) $\equiv q \leq n$)). For example, a static output feedback control (SOFC) is defined as $\boldsymbol{u}(t) = -K_y\boldsymbol{y}(t)$, or $K = K_yC$ with rank $(C) = m < n$.

A stronger feedback control must improve better system performance (mainly defined as stability and faster and smoother transient response) and robustness (mainly defined as low sensitivity against system model uncertainty and input disturbance), because both system properties are understandably critical. Among these two system properties, robustness/reliability is even more critical than performance in feedback control because it is well known that such control is aimed mainly at reducing sensitivity against system model uncertainty and input disturbance, and at the price of reduced performance. In other words, the main purpose of feedback control is for improving system robustness but not performance. In addition, because these two system properties are contradictory to each other[8], an effective feedback control must adjust effectively the tradeoff between these two system properties, based on actual design requirements.

If both p and q are higher than one, then an SFC can assign all n eigenvalues and eigenvectors of matrix $A - BK(= A - B\underline{KC})$ if $q = n$[9], and all n eigenvalues and at least $\min\{n - p, n - q\}$ eigenvectors if $n < p + q$[5, 10, 11]. It is well known that system poles can determine most directly and accurately, and far more directly and accurately than bandwidth, the system transient response and performance. It is also well established that eigenvectors determine the sensitivities of their corresponding eigenvalues[6]. To summarize, an SFC can most directly and effectively improve feedback system performance and robustness, and is an ideal and by far the most effective form of feedback control.

SOFC is a special form of SFC with $q = m$. Thus, only systems (1) with $n \geq p + m$ are considered generally nontrivial – only such systems require a more sophisticated dynamic controller such as an observer (SOFC is not effective enough). This is also the reason that if an observer can only estimate $\underline{C}\boldsymbol{x}(t)$ instead of $I\boldsymbol{x}(t)$ (I = identity matrix) (see Section 3), then it is desirable to maximize rank $(\underline{C})(\equiv q)$ as to maximize the effectiveness of the corresponding SFC.

In actual practical design, it is only effective to optimize a unified criterion that measures both system performance and robustness. Apparently, a gauge of robust stability called stability margin (defined as the sensitivity or likelyhood to become unstable at system (state matrix) parameter variation) has been such a criterion ever since. Such criterion in classical control theory is gain/phase margin, and this criterion has been critical in all existing practical designs[12]. The most basic stability criterion is that all system poles ($\lambda_i, i = 1, \cdots, n$) have negative real parts ($\mathrm{Re}(\lambda_i) < 0, i = 1, \cdots, n$), and the sensitivities of these poles ($s(\lambda_i), i = 1, \cdots, n$) are determined by their corresponding eigenvectors[6]. As a result, in early 1990s, a new stability margin that is in terms of all $\mathrm{Re}(\lambda_i)'$s and $s(\lambda_i)'$s was proposed as[5, 13]

$$\min\{s(\lambda_i)^{-1}|\mathrm{Re}(\lambda_i)|\}, \quad 1 \leq i \leq n \tag{3}$$

and is convincingly proved to be more generally accurate than other existing stability margins of state space theory. In this stability margin, the distance of variation of λ_i to reach unstable region is $|\mathrm{Re}(\lambda_i)|$, because a $0°$ direction of this variation is assumed. If we know a priori that the direction of this variation is θ_i instead of $0°$, then min $\{s(\lambda_i)^{-1}|\mathrm{Re}(\lambda_i)|\cos^{-1}(\theta_i)\}$ should be even more accurate than (3).

There are two obvious and distinct advantages of stability margin (3) over the gain/phase margins. First, (3) is more generally accurate than gain/phase margin. Stability margins by definition must be determined by state matrix only as so in (3), while gain/phase margins are not only generally inaccurate[14] especially when $p > 1$, but also determined by loop transfer function in which the state matrix is mixed up with un-relevant system parameters such as B and C of (1). Second, stability margin (3) can be much more easily and effectively maximized than gain/phase margin. While existing computational algorithms of [9] can assign eigenvalues/vectors generally and effectively, the gain/phase margin is well known very difficult to be increased generally and effectively.

As a proof of all said in this subsection, the superiority of SFC was demonstrated in reality, when optimal SFC was applied to rocket control in former Soviet Union during the 1950′s[15]. Apparently, this application success promoted state space control theory most effectively. Because most systems have $n > m$ and cannot have all n elements of $\boldsymbol{x}(t)$ measured directly, an observer is therefore needed to estimate $K\boldsymbol{x}(t)$ and to implement SFC (or $K\boldsymbol{x}(t)$-control)[16].

Section 2 surveys the results of observers (2) that estimate $K\boldsymbol{x}(t)$ as its only purpose. The first observer that provides the estimation of $\boldsymbol{x}(t)$ is the well know Kalman filter[17], whose estimation has an additional property of minimum variance. At deterministic cases with zero variance and zero measurement noises, the remaining design freedom is fully used to minimize the observer order. This order reduction is possible if the observer estimates $K\boldsymbol{x}(t)$ directly instead of $\boldsymbol{x}(t)$, and such an observer is named "function (or functional) observer". For thirty years since the last survey of observer design[18], very significant progress was made in function observer design problem and the problem is claimed essentially solved[19].

1.3 Why need to further realize the SFC′s loop transfer function

A state observer estimates system state $\boldsymbol{x}(t)$. Multiply this signal by a previously and separately designed gain (K) and feed it into the control input of OLS (1), the observer feedback system is designed and is designed inherently according to the well known separation principle[20]. Further analysis on this feedback system revealed the well known separation property that the feedback system poles are the same as the eigenvalues of two separate matrices $A-BK$ and F of observer (2)[18].

Because system poles determine corresponding system performance such as stability, separation property guarantees the performance of observer feedback system in par with the performance of the corresponding SFC system and observer system. People also assumed the similar robust properties, but almost all real applications of the 1960s and 1970s showed very different robust properties, between the designed SFC system and the actual corresponding observer feedback system.

In 1978, John Doyle pointed out the reason of this difference in robust properties – the loop transfer functions of these two feedback systems are different[21]. Therefore, it is essential that an observer not only estimates signal $K\boldsymbol{x}(t)$, but also has the loop transfer function $L(s)$ of its corresponding feedback system equal to $-K(sI-A)^{-1}B$, the loop transfer function of $K\boldsymbol{x}(t)$-control.

Section 3 surveys all design results of this line of observers, initiated by[21] and named as "loop transfer recovery" (LTR)[22]. All subsequent works except that of the author of this survey involve with state observers and intrinsically separation principle, and these results are unfortunately invalid to most OLS (1) (such as non-minimum phase systems), require large observer gain L which is forbidden in robust control (high gain always means high sensitivity), and $L(s) \neq -K(sI-A)^{-1}B$ (the name "recovery" implies $L(s) = -K(sI-A)^{-1}B$ only approximately at the best). However, a synthesized design principle (as opposite to the separation principle) developed in the early 1990s, can guarantee $L(s) = -K(sI-A)^{-1}B$ at almost all OLS conditions[5, 23].

The remaining of this subsection shows why designing $L(s)$ directly is far less effective in improving feedback system performance and robustness, than designing SFC (and its associated $-K(sI-A)^{-1}B$), as partly explained already in Subsection 1.2. Because the results of LTR of the 1980s have not been satisfactory as described in the previous paragraph, and because robustness is most critical in feedback control, people turned away from modern/state space control theory and back to the classical control theory, and turned to the direct design of $L(s)$.

The design of $L(s)$ to increase gain/phase margins is known to be ineffective, as described in Subsection 1.2. So people resorted to the design of $L(s)$ aimed at minimizing the sensitivity function $(I-L(s))^{-1}$[12]. The most interested work of this line is named H_∞, or to[24]

$$\min[\max\{\|(I-L(\mathrm{j}\omega))^{-1}\|_\infty\}], \quad 0 \leq \omega \leq \infty. \qquad (4)$$

In fact, reference [24] which initiated H_∞ problem appeared in the same special issue of *IEEE Transactions on Automatic Control* as [12], with [12] (LTR) as its first paper, and the subject that has the most papers in the control systems theory area during the 1980s and 1990s is apparently robust control.

However, the H_∞ design (4) has the following two critical disadvantages. First, unlike SFC design which not only maximizes a far more accurate robust stability margin (3) but also assigns all feedback system poles, design (4) itself does not address at all system performance which is not only critical but also contradictory to robustness. Merely adding a bandwidth requirement as a constraint to (4) in problem formulation does not advance at all the basic theory and basic understanding of classical control, and makes an already very difficult problem (4) even more difficult. In fact, no really solid analytical results have been developed that can significantly simplify the design of (4). Second, unlike the observer feedback design with aims which were rooted on deep and solid analytical understandings of Subsection 1.2 and which were inherited with distinct advantages of Subsection 1.1, the design of (4) by definition aims at the norms (or gains) only but not the phase, and thus "missing a very important aim" (quote of Prof. Sheldon S-L Chang during a 1982 faculty interview). The work on both gain and phase is also virtually impossible in other direct designs of $L(s)$.

Nonetheless, these two critical disadvantages of H_∞ design were apparently missed by many people. And as a result, these people cannot appreciate the significance of this line of observer design. As a matter of fact, reference [23] was rejected recently at SCL with a single-sentenced reason that "robust control design should still be resorted to H_∞"!

Finally, we will use a well known example to further emphasize the importance of designing both gain and phase of $L(\mathrm{j}\omega)$. This example is about the ideal $L(\mathrm{j}\omega)$ of quadratic optimal control which is guaranteed only by SFC with $\text{rank}(\underline{C}) = n$[25]. This ideal $L(\text{j}\omega)$ is proven to satisfy Kalman Inequality[26] and thus must have values only outside of the unit circle that is centered at -1 point. Because large gain $|L(\text{j}\omega)|$ is bad, the real ideal values of this $L(\text{j}\omega)$ must be near the origin, and these values must have phase angles between $0°$ to $\pm 90°$!

2 Results of observers that estimate the $Kx(t)$ signal

2.1 Design requirements and formulation

For a general observer (2), proposed in 1963–1966[16, 27], to generate $K\boldsymbol{x}(t)$ signal with a constant K and constant $[K_z K_y]$ in (2b), it is only obvious that the observer state $\boldsymbol{z}(t)$ must converge to $T\boldsymbol{x}(t)(\boldsymbol{z}(t) \to T\boldsymbol{x}(t))$ for a constant T in (2a) and as $\boldsymbol{z}(t) - T\boldsymbol{x}(t) = \text{e}^{Ft}[\boldsymbol{z}(0) - T\boldsymbol{x}(0)]$. It can be simply proved that the necessary and sufficient condition of this convergence is[27]

$$TA - FT = LC, \quad F \text{ is stable} \tag{5}$$

and the number of rows of T equals the observer order. Once this convergence $\boldsymbol{z}(t) \to T\boldsymbol{x}(t)$ is achieved, it is only obvious from (2b) that $\boldsymbol{\omega}(t)$ converges to $K\boldsymbol{x}(t)$ if and only if

$$K = K_z T + K_y C = [K_z : K_y][T' : C']' \equiv \underline{KC}. \tag{6}$$

For example, a state observer estimates $\boldsymbol{I}\boldsymbol{x}(t)$ (or $K = I$), and the corresponding $\underline{C}$ in (6) must be square with $\text{rank}(\underline{C}) = n$, or the state observer order (number of rows of T) must be $n - m$. For strictly proper ($K_y = 0$) state observers such as Kalman filters, (6) becomes $I = K_z T$ and the observer order (number of rows of T) must be n. In fact, the Kalman filter state space model is $\frac{\text{d}\boldsymbol{x}(t)}{\text{d}t} = A\boldsymbol{x}(t) + B\boldsymbol{u}(t) + L[\boldsymbol{y}(t) - C\boldsymbol{x}(t)]$, simply equivalent to $F = A - LC$ and $T = I$ as in the general observer model (2a), and simply showing (5) is satisfied. As a Kalman filter can be designed for all observable systems, (5) and (6) can be satisfied for all observable systems and for any K.

In addition, because matrix $A - LC$ is the exact dual of $A - BK$[5], arbitrarily given observer poles can be assigned to matrix $A - LC$ for all observable systems (1). Because observer performance such as convergence and convergence rate must also be guaranteed, and because further design objectives such as order reduction[28, 29] and LTR[30, 31] cannot be helped effectively by observer pole selection, the observer poles are arbitrarily given in almost all existing designs with very few exceptions.

As (6) is a single set of linear equations which separately and fully determines parameter $[K_z K_y]$ of the static output part of the observer (2b), (5) is far more complicated and determines fully the far more important dynamic part $[F, T, L]$ of the observer (2a). This may be the reason that (5) was named the "observer equation" or "Luenberger equation" in a lot of literature[32].

These state observer design results have been covered in almost all first courses of state space control theory since the 1960s $[1, 2, \cdots]$. However, no textbooks except [5] and its first edition presented any significant and satisfactory development from these early results, despite of the fact that observer design *is* the feedback controller design in state space control theory.

One of the main reasons of this situation is that few people derived or used the solution of (5) that is general, decoupled, and with remaining freedom expressed in a form that is really usable. For example, in some literature, only the so called "Sylvester equation" $TA - FT = C$ (instead of (5)) is studied and the corresponding solution does not exist if A and F share common eigenvalues. This result is certainly not general in light of the expected arbitrary pole assignment, and this deficiency is caused by the missing of design freedom L of (5)[33]. The rows of solution (F, T, L) of (5) must also be completely decoupled because a decoupled design solution has many technical and computational advantages[5, 34]. For example, a designer can freely design the observer order (or the number of rows of (F, T, L)) only if these rows are decoupled. Isn't it true that in any real good controller design methodology, this most important parameter (order) of the controller must be completely free to be designed? In fact, this special feature of observer order design/adjustment freedom is a key breakthrough that enabled all significant developments of observer design, as described in the rest of this survey.

Fortunately, such a closed-form solution of (5) was derived in the middle of 1980s[34, 35], and is described in the following. For simplicity of presentation, only distinct and real observer poles (eigenvalues of F) $\lambda_i (i = 1, 2, \cdots)$ are used. For general eigenvalue cases, this solution can easily be generalized[5, 34, 35]. Let F be in Jordan form Λ (diagonal in this eigenvalue case). Let the system parameters (A, C) be simply transformed so that $C = [C_1 : 0]$ and so that (5) can be separated into two parts. While the left part of m columns of (5), $(TA - \Lambda T)[I_m : 0]' = LC_1$, can always be satisfied by free parameter L for any T, the solution T that always satisfies the right $n - m$ columns of (5) can be simply expressed as

$$\boldsymbol{t}_i = \boldsymbol{c}_i D_i \tag{7}$$

where $\boldsymbol{t}_i (i = 1, 2, \cdots)$ is the i-th row of T, the m-dimensional row $\boldsymbol{c}_i$ is the complete remaining freedom of $\boldsymbol{t}_i$, and the $m \times n$ dimensional basis vector matrix D_i of $\boldsymbol{t}_i$ can always be computed from the corresponding right $n-m$ columns of (5), $D_i[A - \lambda_i I][0 : I_{n-m}]' = 0$[5, 34, 35].

The dual version of this solution is $\boldsymbol{v}_i = D_i \boldsymbol{c}_i$, where $\boldsymbol{v}_i$ is the i-th column of solution matrix V of the dual of (5), $AV - V\Lambda = BK$. In 1985, general and effective algorithms for SFC robust eigenstructure assignment were developed and were based and enabled by this solution[9].

The next natural significant improvement of the design of observers that estimate $K\boldsymbol{x}(t)$ signal, is observer order reduction. This design task is satisfactorily achieved only by using the above solution (7) of (5), and is surveyed in the next subsection.

2.2 Results of function observer design for reduced order

Function observers estimate the $K\boldsymbol{x}(t)$ signal directly instead of $I\boldsymbol{x}(t)$ as in the state observers, and has minimized order. Therefore, mathematically, this design is to compute the solutions of (5) and (6) with minimized number of rows of solution (F, T, L) of (5) and (6).

Physically speaking, because the number of signals of $K\boldsymbol{x}(t)$ (say p, if $K\boldsymbol{x}(t)$ is an SFC signal) is less than n, the number of signals of $\boldsymbol{x}(t)$, function observer order may be less than the state observer order (n or $n-m$ if strictly proper or proper). Mathematically in (2b) and (6), because rank(K) (say p, if K is an SFC gain) is less than n, the number of rows of T may be less than the number that makes matrix $\underline{C}$ rank n (n or $n-m$ for strictly proper or proper state observers).

Function observer design problem was proposed in 1966[27], and was actively studied and reported until 1986[1, 2, 18, 28, 29, 36−40]. However, none of these results used the above solution (7) of (5), and none derived general design solution except [1, 2, 40], which were based on polynomial MFD only. The generally guaranteed upper bound of observer order of [1, 2] is $p(\nu_1 - 1)$, where ν_1 is the largest observability index of system (1). It can be assumed without loss of generality that the m observability indices of system (1) are in descending order, or $\nu_1 \geq \cdots \geq \nu_m$ and by definition $\nu_1 + \cdots + \nu_m = n$.

In 1985, the above solution (7) of (5) was used and substituted into (6) for the first time[41]. Because $C = [C_1 : 0]$, as for the solution of (5), (6) can be similarly separated into two parts. While the left part of m columns of (6) can always be satisfied by the free parameter K_y for any T, the right $n-m$ columns of (6) can now be expressed as[41]

$$\underline{K} = K_z \mathrm{diag}\{\boldsymbol{c}_1 \cdots \boldsymbol{c}_r\}[\underline{D}_1' : \cdots : \underline{D}_r']' \tag{8}$$

where $\underline{K}$ and $\underline{D}_i (i = 1, \cdots, r)$ are the right $n-m$ columns of K and $D_i (i = 1, \cdots, r)$ of (7), respectively. Our design problem is now simplified to solve (8) only with minimum value of r (or minimum observer order).

Equation (8) is really a single set of linear equations, because the unknown variables of (8) are at the same side of the equation. Although these unknown variables have a special data structure of two parts (K_z and $\{\boldsymbol{c}_1, \cdots, \boldsymbol{c}_r\}$), K_z is the full remaining freedom of (2b) after K_y while $\{\boldsymbol{c}_1, \cdots, \boldsymbol{c}_r\}$ is the full remaining freedom of (2a) and (5). This equation is concluded as the simplest possible theoretical formulation of functional observer design problem[42, 43]. This is also the most significant progress on the theoretical formulation of the design problem, and this progress is enabled only by the use of this new solution (7) of (5).

Mathematically, this basic and simple problem (8) is covered in first linear algebra courses of undergraduate engineering programs such as[44], as a dual problem of $\boldsymbol{k} = D\boldsymbol{c}$, where $\boldsymbol{k}$ and $\boldsymbol{c}$ are column vectors and $\boldsymbol{c}$ is unknown with minimized dimension (r). The apparently most effective algorithm of this mathematical problem is to upper echelonize matrix $[D : \boldsymbol{k}]$ until the linear dependency of $\boldsymbol{k}$ on some (say r) columns of D is revealed, and then the unknown solution $\boldsymbol{c}$ can be computed by simple back-substitution[44].

Likewise, the design algorithm of [41] for problem (8) lower echelonizes repeatedly a working matrix $[\underline{D}_1' : \cdots : \underline{D}_{n-m}' : \underline{K}']'$ until the linear dependency of the rows of $\underline{K}$ on the rows of $\underline{D}_i (i = 1, 2, \cdots)$ is gradually revealed. The unknown solutions $\{\boldsymbol{c}_1, \cdots, \boldsymbol{c}_r\}$ and then the corresponding row of K_z can be computed by back-substitution, every time a row of $\underline{K}$ is revealed to be linearly dependent on the rows of $\underline{D}_i (i = 1, 2, \cdots, r)$. Like the conclusion of [44], this is apparently the most effective algorithm for the problem (8).

Reference [45] further improved the algorithm of [41] by adding the adjustment of positions of the $\underline{D}_i$ matrices in that working matrix, and proved that the algorithm can guarantee an observer order upper bound of

$$\min\{n - m, (\upsilon_1 - 1) + \cdots + (\upsilon_p - 1)\} \tag{9}$$

and a lower bound of 0. For strictly proper observers where (8) becomes $K = K_z \mathrm{diag}\{\boldsymbol{c}_1, \cdots, \boldsymbol{c}_r\}[D_1' : \cdots : D_r']'$, the corresponding observer order upper bound becomes $\min\{n, \upsilon_1 + \cdots + \upsilon_p\}$ and the lower bound becomes 1. These bounds are unified and are the lowest ever since 1986. This situation is not surprising because later analysis proved that these bounds are the lowest possible bounds of function observer order[42, 43].

Finally, [19] most formally and rigorously proved the above claims of [42, 43], and then made the following two further clear-cut claims: 1) Just like the minimum value of r is a function of every given data of the problem $\boldsymbol{k} = D\boldsymbol{c}$, there is no analytical formula for the minimal observer order (r) itself from problem (8). Thus, (8) and (9) are the best possible theoretical result of minimal order function observer design problem, and the theoretical part of this problem is solved. 2) Just like the minimal value r of the problem $\boldsymbol{k} = D\boldsymbol{c}$ is generically near its upper bound (= dimension of $\boldsymbol{k}$), the actual observer order r is generically near its upper bound say (9). Thus, further improvement of computational algorithm to seek value of r which is even lower than its upper bound is generically not worthwhile, and thus the whole design problem is essentially solved. This is a significant progress from the early authoritative conclusion that this design problem is “difficult” and “unsolved”[1, 2].

It should be mentioned that the above significant and clear-cut results were largely missed by the research community. The claims of [19, 42, 43] were repeatedly rejected before publication that was already many years after 1986. In addition, about a decade and half after 1986, some new results appeared[46, 47]. According to [19], these results reformulated without simplification the original problem formulation (5) and (6) (such as merely combining (5) and (6) together or merely reformulating the solvability of (6) as rank $[K' : \underline{C}'] = \mathrm{rank}[\underline{C}']$), proposed only exhaustive numerical search to compute the solution of the complicated reformulation, and cannot even achieve the analytical result of $p(\nu_1 - 1)$ of 1980[2]. The claim of [46] that it offered the only design solution to the minimal order functional observer design problem, is based only on the assumption that numerical search can yield the minimal order, and is obviously erroneous and misleading.

The design result of [41, 45] not only is significant and clear-cut theoretically, but also offers significant practical advantage in observer order reduction. This is especially true when $p \ll m$ and the observability indices are similar to each other, because in these cases the guaranteed observer order upper bound (9) is significantly lower than the state observer order $n - m$. For example, if $n = 120$, $m = 20$, $p = 3$, and $\upsilon_1 = \cdots = \upsilon_m = 6$, then $n - m = 100$, while our functional observer order is guaranteed not exceeding 15!

This order reduction of function observers is achieved by fully using the remaining freedom $\{\boldsymbol{c}_1, \cdots, \boldsymbol{c}_r\}$ of (5) (or of (2a)) in (8) (or in (2b)). In Section 3, this remaining freedom will be fully used, instead, for an all important goal – realize the loop transfer function of its $K\boldsymbol{x}(t)$-control.

3 Observers that further realize the loop transfer function of its SFC

3.1 Under separation principle, LTR observer design is unsatisfactory

As analyzed in Subsection 1.3, if an observer generates a $K\boldsymbol{x}(t)$ signal for the purpose of realizing this $K\boldsymbol{x}(t)$-control, then it must also have its observer feedback system loop transfer function $L(s)$ equal $-K(sI - A)^{-1}B$, the loop transfer function of this $K\boldsymbol{x}(t)$-control.

Basic system analysis shows the following for the feedback system of observer (2) satisfying (5) and (6). $L(s)$ at the node $\boldsymbol{u}(t)$ equals $-[I - K_z(sI - F)^{-1}TB]^{-1}[K_y + K_z(sI - F)^{-1}L]G(s)$. At a virtual node $\boldsymbol{\omega}(t)$ which is before the feedback of $\boldsymbol{u}(t)$ into the observer (2a), the loop transfer function $L_{Kx}(s)$ equals $-K_z(sI-F)^{-1}TB - [K_y + K_z(sI - F)^{-1}L]G(s)$ and this loop transfer function equals $-K(sI - A)^{-1}B$[5, 48]. Therefore, our problem is to make $L(s)$ equal to $L_{Kx}(s)$.

The difference between $L(s)$ and $L_{Kx}(s)$ is obviously at the term $K_z(sI - F)^{-1}TB$, which is the loop gain of the feedback of $\boldsymbol{u}(t)$ into the observer. If this feedback, with gain TB of (2a) and loop gain $K_z(sI - F)^{-1}TB$, equals zero, then and only then $L(s) = L_{Kx}(s)$[5, 30].

However in practice, parameter K_z must be designed only for a satisfactory feedback system matrix $A - BK = A - B[K_z : K_y][T' : C']'$ (the existing SFC design does not even allow the constraint $K = [K_z : K_y][T' : C']'$), and certainly cannot allow the additional constraint $K_z(\mathrm{j}\omega I - F)^{-1}TB = 0$ for all ω as well as (5). Therefore, $K_z(\mathrm{j}\omega I - F)^{-1}TB$ should be made 0 for all ω and for all K_z[5]. Also because $\mathrm{j}\omega I - F$ must be non-singular for all ω, $L(s) = L_{Kx}(s)$ if and only if in addition to satisfy (5) and (6)[5, 30]

$$TB = 0. \tag{10}$$

It is also obvious that the feedback of $\boldsymbol{u}(t)$ into the observer is really eliminated, or the term $TB\boldsymbol{u}(t)$ of (2a) is eliminated if and only if $TB = 0$ (but not $K_z(\mathrm{j}\omega I - F)^{-1}TB = 0$ for all ω). In actual computation, the remaining freedom of solution (7) of (5), $\{\boldsymbol{c}_1, \cdots, \boldsymbol{c}_r\}$, can be used to simply achieve $TB = 0$ in the formulation diag $\{\boldsymbol{c}_1, \cdots, \boldsymbol{c}_r\}[D_1' : \cdots : D_r']'B = 0$.

All LTR results since [22] acknowledged condition $K_z(\mathrm{j}\omega I - F)^{-1}TB = 0$ for all ω, but apparently not condition (10) except that of the author of this survey. In fact, condition (10) is theoretically not as necessary as condition $K_z(\mathrm{j}\omega I - F)^{-1}TB = 0$ for all ω (because some special values of (K_z, F, T) can be cooked (not practically designed) so that $K_z(\mathrm{j}\omega I - F)^{-1}TB = 0$ for all ω without $TB = 0$[49]) has been a main reason and sometimes the only reason to reject papers like [23] in the past 20 years.

In the past 20+ years, people also expressed strong doubt on the effect of TB = 0, i.e., the loss of feedback information of $\boldsymbol{u}(t)$ which is needed in Kalman filtering and the simplification of observer structure (2a) to an output feedback controller (OFC) (output feedback only and no more input feedback). But OFC has been the dominant controller structure of the well established classical control designs because they achieved robustness, while Kalman filter or state observer based feedback system has no guarantee of robustness at all in deterministic sense (this is also why $\boldsymbol{u}(t)$ is missing in (2b))!

To summarize, the whole design problem is simplified as to satisfy (5), (6) and (10).

Unfortunately, general exact solution or satisfactory approximate solution of this problem does not exist under separation principle, which implies that (6) must be satisfied for a separately designed and arbitrarily given K. Because all remaining design freedom (7) of (5) will now be used for satisfying (10) instead of for minimizing the observer order, guaranteeing (6) for an arbitrarily given K implies rank $(\underline{C}) = n$, or the observer must be a state observer (capable of estimating $\boldsymbol{x}(t)$). Nonetheless, all LTR results except that of the author of this survey are state observers with rank $(\underline{C}) = n$, and are certainly designed under separation principle.

The exact solution of (5), (10), and rank $(\underline{C}) = n$ is equivalent of an "unknown input observer" (UIO), if B is defined as the only gain to the only unknown input in (1a)[50]. Unknown input observer was proposed in the early 1970s[50] and its necessary and sufficient condition was derived in 1980 as 1) minimum-phase (all transmission zeros are stable); 2) $\mathrm{rank}(CB) = p$; and 3) $m \geq p$[51, 52]. Almost all OLS (1) cannot satisfy these three conditions.

For example, the probability of a transmission zero is stable/unstable can reasonably be assumed as $\frac{1}{2}$, because the stable and unstable regions are about half and half. So the probability of minimum phase with $n{-}m$ transmission zeros is only $\left(\frac{1}{2}\right)^{n-m} = 0.125, 0.0625, \cdots$ as $n - m = 3, 4, \cdots$[5]. We recall systems with $\mathrm{rank}(CB) = p$ and $m = p$ always have $n - m$ transmission zeros[53], and that observers are needed only for non-trivial systems which have $n \gg m$

(Subsection 1.1). Condition $\mathrm{rank}(CB) = p$ is also unsatisfied by many important systems such as airborne systems even if $m > p$.

The main approximate solution of this problem of (5), (10) and $\mathrm{rank}(\underline{C}) = n$ is called "asymptotic LTR". Asymptotic LTR is achieved by asymptotically increasing the input noise level while designing a Kalman filter[22], or by asymptotically increasing the time scale while designing a state observer[54, 55]. Through this way, the corresponding observer gain L is asymptotically increased so that the effect of gain TB is overwhelmed in (2a). This technique was also known in Kalman filter design[56]. But large gain L is always prohibitive in practice especially in robust control[57−59]. Large gain L will cause instability if OLS is non-minimum phase, so this LTR result is not general (see the previous paragraph). Even at a very large gain $L, L(s)$ is still very different from $L_{Kx}(s)$ at low frequency, and even for very simple and typical example of [22, 60].

Another approximate LTR solution computes directly $L(s)$ so that $\|L(\mathrm{j}\omega) - L_{Kx}(\mathrm{j}\omega)\|_\infty$ is bounded[61]. However, there is apparently no limit on this bound itself–it can be very large[62], and there is no consideration on the phase of $L(\mathrm{j}\omega) - L_{Kx}(\mathrm{j}\omega)$ at all. Clearly, the reason for this unguaranteed result is that $L_{Kx}(\mathrm{j}\omega)$ is arbitrarily given (or is separation principle itself)[5].

There is another approximate solution to the same problem, which is aimed directly at minimizing TB by selecting the $n - m$ observer poles clustered around the existing stable transmission zeros while designing a function observer[30, 48]. However, this pattern of observer poles will result in a near singular matrix $\underline{C}$ which in turn will result in a large observer output gain $\underline{K}$ while trying to satisfy (6).

3.2 All drawbacks are overcome under a new and synthesized design principle

In 1990 and in an 86°F room one summer afternoon, while stuck by the dilemma that making $TB = 0$ would make $\mathrm{rank}(\underline{C}) < n$, it suddenly occurred to this author that $\mathrm{rank}(\underline{C})$ needs not to be n (or observer order needs not be fixed at $n - m$), that SFC gain K needs not to be pre-designed (it can be designed afterwards and indirectly by designing $\underline{K}$ in the form of (6) ($K = \underline{KC}$) using the existing SOFC methods), and separation principle need not to be adhered after all!

With only (5) and (10) need to be satisfied without requiring $n - m$ linearly independent rows of T, the exact solution becomes valid for almost all OLS′s (1). Let z_i be the i-th eigenvalue of a diagonal matrix F, and let $\boldsymbol{t}_i$ and $\boldsymbol{l}_i$ be the corresponding rows of matrices T and $L\,(i = 1, \cdots, r)$. Then (5) and (10) can be expressed together as

$$[\boldsymbol{t}_i : \boldsymbol{l}_i]\begin{bmatrix} A - z_iI & : B \\ -C & : 0 \end{bmatrix} = 0. \tag{11}$$

Obviously, exact solution $[\boldsymbol{t}_i : \boldsymbol{l}_i]$ exists if and only if either $m > p$, or z_i is a (stable) transmission zero of OLS (1)[1]. Because the number of rows of T is free to design, exact solution of (5) and (10) exists if and only if either $m > p$ or system (1) has at least one stable transmission zero[63−69].

This condition is satisfied by almost all OLS (1). The paragraph following (1) explained that $m \geq p$ is true for almost all systems (1). At $m = p$, the system generically have $n - m$ transmission zeros[53]. Based on the same argument on the probability of minimum-phase for systems with $m = p$, the probability of at least one stable transmission zero among $n - m$ transmission zeros is $1 - (\frac{1}{2})^{n-m}$ and is almost 100 % if $n - m > 3$!

Now, the first general OFC that can estimate the $K\boldsymbol{x}(t)$ signal is claimed in [66]. This controller has the key advantage of full realization of the critical robust property of this all-powerful $K\boldsymbol{x}(t)$-control ($K = \underline{KC}$).

To make this $\underline{KC}\boldsymbol{x}(t)$-control as effective as possible, in addition to (5) and (10), $\mathrm{rank}(\underline{C})(= r + m)$ or r should be maximized (see the 4th paragraph of Subsection 1.2), and this maximization can be guaranteed by the algorithms of [9] because (5) and (10), or $\mathrm{diag}\{\boldsymbol{c}_1, \cdots, \boldsymbol{c}_r\}[D_1' : \cdots : D_r']'B = 0$ is very simple and is the same form on which the algorithms of [9] are based[63−69]. This means that if the OLS (1) satisfies the three conditions of UIO[51, 52], then $\mathrm{rank}(\underline{C}) = n$ (or $r = n - m$) is guaranteed. It is also proven that separation property holds as long as (5) is satisfied[60]. Thus, this synthesized design also guarantees the performance of its results (see the first two paragraphs of Subsection 1.3).

The second main step of this synthesized design is the design of $\underline{K}$ based on the already designed parameter $\underline{C}\,(= [T' : C']'$, rank $(\underline{C}) \equiv q = r + m)$ in the form of (6) ($K = \underline{KC}$). This design is the same as the existing SFC design if $\mathrm{rank}(\underline{C}) = n$, and is the same as the existing SOFC if $\mathrm{rank}(\underline{C}) < n$ (see the first four paragraphs of Subsection 1.2). This $\underline{KC}\boldsymbol{x}(t)$-control is therefore named the "generalized state feedback control"[5, 63−69].

The observer order r, or the number of linearly independent rows of T, is free to design based on the actual OLS conditions and actual design requirements, and contrary to many people, this is not a disadvantage at all! For example, if the actual design requirement is only partial eigenvector assignment or only $p + q > n$ (or $q > n - p$)[5, 10, 11], then a lower observer order $r < n - m$ but $r > n - p - m$ can be eventually chosen even if $\mathrm{rank}(\underline{C}) = n$ or $r = n - m$ is already achieved. A lower order r means a simpler controller as well as easier making of $TB = 0$ because less equations to satisfy (easier robust realization). On the other hand, if another actual design requirement requires a high value of r (say $r = n - m$) which is not achievable (say the conditions for UIO are not met), then our design still can set r equal that desired high value while satisfying $TB = 0$ only approximately in least-square sense (or to realize $L_{Kx}(s)$ only approximately)[69, 70]. To summarize, this synthesized design can adjust the levels of feedback system performance and robustness effectively by adjusting order r[5, 69].

Although the $\underline{KC}\boldsymbol{x}(t)$-control can be constrained, there are good reasons to believe that this control is effective enough for most OLS conditions and most design requirements, because rank($\underline{C}$) $= r + m > m$. The following example may show some light on this conclusion[23].

This example includes all the 5th order systems with two inputs and two outputs ($p = m = 2$). Many important practical systems such as aircrafts and engines have a 5th order model. We will assume $3\,(= n - m)$ transmission zeros in this example[53] and the probability of minimum-phase and exact LTR is therefore only 12.5 %. Because $m \times p < n$, the SOFC of this OLS is too weak to guarantee acceptable properties (such as stability and arbitrary pole assignment)[71], and this example is definitely non-trivial.

Because $m = p$, our synthesized design guarantees the full realization of a $\underline{KC}\boldsymbol{x}(t)$-control with rank($\underline{C}$) $= q = m + r$, where r is the number of OLS stable transmission zeros[5, 23, 63−69]. The probability that this control is very effective (able to have partial eigenvector assignment or better, or $q + p > n$ and $r > n - p - m = 1$) is 50 %, while the probability that this control is effective (able to assign generically all poles or better, or $q \times p > n$[71] and $r > \frac{n}{p} - m = 0$) is 87.5 %[23]. Thus, this $\underline{KC}\boldsymbol{x}(t)$-control is effective enough for most systems of this example.

Finally, the only remaining and the most severe criticism of this synthesized design principle, that the corresponding generalized $\underline{KC}\boldsymbol{x}(t)$-control is allowed to be constrained and weaker than the existing ideal $K\boldsymbol{x}(t)$-control (corresponding to rank($\underline{C}$) $= n$ and under separation principle) when rank($\underline{C}$) $= n$ cannot be achieved along with (5) and (10) (such as the 87.5 % systems of the above example), is answered in the following with a more philosophical point of view.

If (10) cannot be achieved and the critical loop transfer function/robustness property of the ideal $K\boldsymbol{x}(t)$-control cannot be actually realized, then the optimality of this ideal $K\boldsymbol{x}(t)$-control is, as demonstrated in reality ever since the 1960s, all but lost, and the ultimate purpose of observers of realizing the $K\boldsymbol{x}(t)$-control is all but failed! Under separation principle, the $K\boldsymbol{x}(t)$-control is designed not only assuming the ideal information of full measurement of $\boldsymbol{x}(t)$ which is actually unavailable, but also disregard important OLS conditions (such as the conditions to have UIO) and totally disregard the parameters of the observer which actually implements this control, and then force the realization of this unrealistically designed "ideal" $K\boldsymbol{x}(t)$-control for all OLS′s. This is the fundamental reason of the failed realization of the $K\boldsymbol{x}(t)$-control for almost all OLS′s[5, 23].

In human history, if a social policy or a social system (like $K\boldsymbol{x}(t)$-control) that does not fit the actual (say backward) condition of the country (like OLS), disasters almost always happen when such a policy/system is implemented, no matter how ideal and lofty that policy/system is.

Although the new synthesized design principle abandons the ideal $K\boldsymbol{x}(t)$-control in the case that control cannot be truly realized anyway, a flexible/general $\underline{KC}\boldsymbol{x}(t)$-control (rank($\underline{C}$) is all flexible/general from m to n) is fully realized. This control can fully adjust its strength (or constraint) according to different system conditions and different design requirements. This control fully and wisely uses the design freedom of observer (2) including its order and the information of OLS (1) (both are substantially better than their counter parts in classical control theory, see Subsection 1.1), and is therefore far more effective than other forms of control, and is effective enough for most OLS conditions and most design requirements. This development makes the control of modern/state space control theory far more effective and far more useful in applications, and far more effective than that of the classical control theory.

4 Observers that generate $\boldsymbol{Kx(t)}$ signal in other special applications

This section surveys the observers that generate $K\boldsymbol{x}(t)$ signal in two special applications, i.e., fault detection/ identification and systems with time-delays. In both applications, $K\boldsymbol{x}(t)$ is estimated while eliminating the unknown and undesirable effects such as fault signal $\boldsymbol{f}(t)$ and time-delay signal $\boldsymbol{f}(t - \tau)$. Because both undesirable signals are modeled as additional input disturbances in systems model (1a) with gain say B_d, the new and general solution of the matrix equation pair (5) and (10) (with B replaced by B_d) of Subsection 3.2 can be used directly to derive satisfactory and far more satisfactory design solutions to these two applications.

4.1 Observers for input fault detection and identification

The normal observers or normal feedback controllers of Sections 2 and 3 are designed to achieve robustness against system model uncertainties and system input disturbance that are frequent but minor. For example, both sensitivity function $[I - L(s)]^{-1}$[12] and sensitivity of eigenvalues $s(\lambda)$[6] were defined under the assumption that the corresponding parameter change is minor. On the contrary, faults happen only accidentally but are severe and major. Therefore, fault detectors additional to the normal observer/controller are needed to detect and identify the fault occurrence.

In [72, 73], fault can be commonly categorized as input (actuator) fault and output (sensor) fault, with the input fault far more complicated since it affects directly the dynamic of the system (1a). This paper surveys the results dealing with input fault only.

The effect of the fault is modeled as an additional input term $B_d\boldsymbol{f}(t)$ in system model (1a), where the unknown signal $\boldsymbol{f}(t)$ is called "fault signal" with dimension say p, and is non-zero or zero if the fault occurs or not, while B_d is a known and given gain to $\boldsymbol{f}(t)$. For example, if there are two possible groups of p_1 and p_2 faults represented by fault signals $\boldsymbol{f}_1(t)$ and $\boldsymbol{f}_2(t)$ (of dimensions p_1 and p_2), then $B_d\boldsymbol{f}(t) = [B_{d1} : B_{d2}][\boldsymbol{f}_1(t)' : \boldsymbol{f}_2(t)']'$ with $p = p_1 + p_2$. In reality, the number of fault groups is often more than two. For example, we can consider the same system having p groups of single faults.

The task of fault detection is very simple – all we need is a single detector output $\omega(t)$ that is normally (fault free) zero and will become non-zero when fault occurs. For this task, the fault detector (2a) needs only to achieve a convergence $\boldsymbol{z}(t) \rightarrow T\boldsymbol{x}(t)$ or to satisfy (5), because (6) can always be satisfied if $K = 0$ (the rows of T are allowed to be linearly dependent). A threshold on the zero/nonzero determination of this $\omega(t)$ can be established against model uncertainty and measurement noise, and the level of $\boldsymbol{f}(t)$ for which the corresponding $\omega(t)$ can surpass that threshold and be detected, is also derived[5, 74−76].

Like the situation that further realizing the loop transfer function is much more difficult than signal estimation, the further task of fault identification is many degrees more difficult than the mere task of fault detection. Fault identification requires identify among all possible fault groups, which group of faults has occurred (in the above example of only two fault groups, identify actually which fault signal, $\boldsymbol{f}_1(t)$ or $\boldsymbol{f}_2(t)$, is non-zero, not just a lump $\boldsymbol{f}(t)$).

Intrinsically, fault detection and identification is very much like illness detection and diagnosis – becoming ill (like fault occurrence) can be detected by patient (just feel abnormal), while it takes the expertise of a doctor to diagnose from all possible diseases which disease actually occurred and then prescribe the corresponding remedy (fault control in our case). Going back to more general situation, the effect of germs (like noise and system variations) is everywhere and ever present but minor, a robust person (or robust system) can handle it normally and most of time. Only a severe illness (or a fault that can cause $\omega(t)$ to surpass its threshold) requires a special prescription (or fault control) to cure.

A group of fault detectors is required for fault identification, with each fault detector able to identify the occurrence of only one group of faults, or have its output become nonzero if and only if that group of faults occurs. We call such a fault detector "robust fault detector" since it is supposed to be robust (output remains zero) toward the occurrence of all other groups of faults. For example to identify if only $\boldsymbol{f}_1(t)$ or only $\boldsymbol{f}_2(t)$ is nonzero, two robust fault detectors are needed (outputs are $\omega_1(t)$ and $\omega_2(t)$, respectively). Output $\omega_i(t)$ becomes nonzero if and only if $\boldsymbol{f}_i(t)$ is nonzero ($i = 1, 2$).

It is obvious that to achieve this stated task, in addition to satisfy the convergence or (5), each robust fault detector ($F_i, T_i, L_i, i = 1, \cdots$, the number of fault groups) must satisfy $T_i\underline{B}_{di} = 0$ or (10), where $\underline{B}_{di}$ is the given gain to all the faults for which this i-th fault detector is supposed to be robust. In the two-fault example, T_1B_{d2} must be zero and T_2B_{d1} must be zero.

The solution of this design requirement is already described in (11) of Subsection 3.2. Among the two sufficient conditions for the existence of exact solution, the simpler condition $m > \underline{p}$ ($\underline{p}$ is the column rank of $\underline{B}_{di}$) was used to derive a far more satisfactory design solution[77] to a 4th order and four-state-fault water tank fault detection and identification system, first proposed in [78].

Once the fault identification is achieved, the subsequent fault control against the identified fault(s) is relatively easy (just like the prescription of remedy is relatively easy if the diagnosis is specific). One advantage of this fault detector result of [77] is that it provides estimation information (like $T_i\boldsymbol{x}(t), i = 1, \cdots$, the number of fault detectors) in addition to the fault detection and identification. Once a fault situation is identified and based on the priori knowledge of that specific fault situation, one can select and gather the still (relatively) reliable signals among these $T_i\boldsymbol{x}(t)$ signals to form a signal say $T\boldsymbol{x}(t)$ and apply a corresponding and effective fault control $\underline{KC}\boldsymbol{x}(t)$-control for that specific fault situation[5, 75, 76].

4.2 Observers for systems with time-delay effects

The time-delay effects can be modeled as an additional input term $B_d\boldsymbol{f}(t-\tau)$ in system model (1a), where $\boldsymbol{f}(t-\tau)$ is unknown with time delay τ while B_d is given with column rank say p, just like the model of unknown-input systems and the model of fault systems of Subsection 4.1. For example, [79] modeled this $B_d\boldsymbol{f}(t-\tau)$ signal as $A_d\boldsymbol{x}(t-\tau)$.

Obviously, to estimate a required signal $K\boldsymbol{x}(t)$ for this system, observer (2) must satisfy (5), (6) and (10) ($TB_d = 0$)[79]. Thus, the result of Section 3 can be used directly to form a far more satisfactory solution to this problem[80], while [79] did not advance beyond the formulation (5), (6) and (10) at all and did not provide a real solution (just like [47]). Reference [79] also has basic errors such as the claim that the necessary condition of having an exact solution is $m > p$ (should be $m \geq p$), as pointed out by [80]. Reference [80] was previously rejected.

5 Conclusions

Observer is a $L-T-I$ system that generates a $K\boldsymbol{x}(t)$ signal where K is a constant. Therefore, state space feedback controller design is observer design, if $K\boldsymbol{x}(t)$-control is implemented. Section 1 shows that this form of control is by far the most effective, among all basic forms of feedback control.

To estimate $K\boldsymbol{x}(t)$ for a separately designed and given K (under separation principle), only function observer may have order lower than state observer order ($n-m$ or n for proper and strictly proper observers). Section 2 indicates that based on a general and decoupled solution (7) to the most important observer design equation (5), the minimal order function observer design problem was essentially solved in 1986[19].

Subsection 1.3 proves that in addition to the generation of feedback control signal $K\boldsymbol{x}(t)$, the observer must also realize the loop transfer function or the robust property of that $K\boldsymbol{x}(t)$-control. Subsection 3.1 proves that to achieve this goal in practice, one must satisfy (10), or design the observer gain (TB) to system input to be 0, or to make the observer an output feedback controller which has been dominant in classical control theory. Subsection 3.1 also shows that under separation principle of more than half of a century, satisfactory solution to this problem does not generally exist.

Subsection 3.2 shows a synthesized design principle in which K is now designed based on the key observer parameter T and the system parameter C. With a now fully adjustable observer order and based on the solution of (5) as described in Section 2, exact solution to (5) and (10) exists for almost all systems[81]. The clear and fundamental reasons for the superiority[23] of this synthesized principle over the existing separation principle are discussed. Because robustness is the most critical property of feedback control, this development improves the state space control theory so decisively that the theory becomes really and generally useful now, and becomes far more effective than classical control theory.

Section 4 describes the direct application of the solution of (5) and (10) in two more special observer design problems, i.e., fault detection/identification and systems with time-delay effects[81]. The dual version of this solution of (5) and (10) also made the solution of the critical SOFC (or generalized state feedback control) design procedure from numerical to analytical, and this analytical solution made the remaining design freedom as clear as that of [9] and thus made the eigenvector assignment possible[5, 10, 11, 81]. We expect more applications of this all important matrix equation pair[81].

Observer design has been covered in almost all first courses of state space control theory and often at undergraduate level. Many research results were published on this basic and all important topic of state space control theory for more than half of a century, but very few of them (except the extension to MIMO systems[1−3] of the early 1980s) were significant and satisfactory enough to be added to the general textbooks. This survey concentrates on the observer design results that are not merely problem formulations, but general, simple, and effective actual design solutions that can improve significantly the whole feedback control theory. It is expected that these results will be added to the new and general textbooks as attempted in [5], be extended to more general OLS′s such as [82], and find many more successful practical applications.

References

[1] C. T. Chen. *Linear System Theory and Design*, 2nd ed., New York, USA: Holt, Rinehart and Winston, 1984.

[2] T. Kaileth. *Linear Systems*, Englewood Cliffs, USA: Prentice-Hall, 1980.

[3] W. A. Wolovich. *Linear Multivariable Systems*, New York, USA: Springer-Verlag, 1974.

[4] C. C. Tsui, C. T. Chen. An algorithm for companion form realization. *International Journal of Control*, vol. 38, no. 4, pp. 769–779, 1983.

[5] C. C. Tsui. *Robust Control System Design – Advanced State Space Techniques*, 2nd ed., Taylor & Francis, 2004.

[6] J. H. Wilkinson. *The Algebraic Eigenvalue Problem*, London, UK: Oxford University Press, 1965.

[7] A. J. Laub. Numerical linear algebra aspects of control design computation. *IEEE Transactions on Automatic Control*, vol. 30, no. 2, pp. 97–108, 1985.

[8] Y. C. Ho, Q. C. Zhao, D. L. Pepyne. The no free lunch theorems: Complexity and security. *IEEE Transactions on Automatic Control*, vol. 48, no. 5, pp. 783–793, 2003.

[9] J. Kautsky, N. Nichols, P. Van Dooren. Robust pole assignment in linear state feedback. *International Journal of Control*, vol. 41, no. 5, pp. 1129–1155, 1985.

[10] C. C. Tsui. A design algorithm of static output feedback control for eigenstructure assignment. In *Proceedings of IFAC World Congress*, Beijing, China, vol. Q, pp. 405–410,1999.

[11] C. C. Tsui. Six-dimensional expansion of output feedback design for eigen-structure assignment. *Journal of the Franklin Institute*, vol. 342, no. 7, pp. 892–901, 2005.

[12] J. Doyle, G. Stein. Multivariable feedback design: Concepts for a classical/modern synthesis. *IEEE Transactions on Automatic Control*, vol. 26, no. 1, pp. 4–16, 1981.

[13] C. C. Tsui. A new robust stability measure for state feedback systems. *Systems & Control Letters*, vol. 23, no. 5, pp. 365–369, 1994.

[14] M. Vidyasagar. The graph metric for unstable plants and robustness estimates for feedback stability. *IEEE Transactions on Automatic Control*, vol. 29, no. 5, pp. 403–418, 1984.

[15] B. Friedland. *Control System Design: An Introduction to State Space Methods*, New York, USA: McGraw-Hill, 1986.

[16] D. G. Luenberger. Observing the state of a linear system. *IEEE Transactions on Military Electronics*, vol. 8, no. 2, pp. 74–80, 1964.

[17] R. E. Kalman, R. S. Bucy. New results in linear filtering and prediction theory. *Transactions on ASME – Journal of Basic Engineering*, vol. 83, no. 3, pp. 95–108, 1961.

[18] J. O′Reilly. *Observers for Linear Systems*, London, UK: Academic Press, 1983.

[19] C. C. Tsui. The theoretical part of linear functional observer design problem is solved. In *Proceedings of the 33rd Chinese Control Conference*, Nanjing, China, pp. 3456–3461, 2014.

[20] J. C. Willems. Mathematical systems theory: The influence of R. E. Kalman. *IEEE Transactions on Automatic Control*, vol. 40, pp. 978–979, 1995.

[21] J. C. Doyle. Guaranteed margins for LQG regulators. *IEEE Transactions on Automatic Control*, vol. 23, no. 4, pp. 756–757, 1978.

[22] J. Doyle, G. Stein. Robustness with observers. *IEEE Transactions on Automatic Control*, vol. 24, no. 4, pp. 607–611, 1979.

[23] C. C. Tsui. Overcoming eight drawbacks of the basic separation principle of state space control design. In *Proceedings of the 31st Chinese Control Conference*, Hefei, China, pp. 213–218, 2012.

[24] G. Zames. Feedback and optimal sensitivity: Model reference transformations, multiplicative seminorms, and approximate inverse. *IEEE Transactions on Automatic Control*, vol. 26, no. 2, pp. 301–320, 1981.

[25] D. Z. Zheng. Some new results on optimal and suboptimal regulators of LQ problem with output feedback. *IEEE Transactions on Automatic Control*, vol. 34, no. 5, pp. 557–560, 1989.

[26] R. E. Kalman. Contributions to the theory of optimal control. *Boletin de la Sociedad Matematica Mexicana*, vol. 5, pp. 102–119, 1960.

[27] D. G. Luenberger. Observers for multivariable systems. *IEEE Transactions on Automatic Control*, vol. 11, no. 2, pp. 190–199, 1966.

[28] T. E. Fortman, D. Williamson. Design of low-order observers for linear feedback control laws. *IEEE Transactions on Automatic Control*, vol. 17, no. 3, pp. 301–308, 1972.

[29] P. Whistle. Low-order Observer Pole Selection Algorithm, Master dissertation, Department of Electrical and Computer Engineering, Northeastern University, Boston, USA, 1985.

[30] C. C. Tsui. On preserving the robustness of an optimal control system with observers. *IEEE Transactions on Automatic Control*, vol. 32, no. 9, pp. 823–826, 1987.

[31] Soggard-Anderson. Comments on "On the loop transfer recovery". *International Journal of Control*, vol. 45, pp. 369–374, 1987.

[32] D. G. Luenberger. An introduction to observers. *IEEE Transactions on Automatic Control*, vol. 16, no. 6, pp. 596–602, 1971.

[33] C. C. Tsui. Comments on "New technique for the design of observers". *IEEE Transactions on Automatic Control*, vol. 31, no. 6, pp. 592, 1986.

[34] C. C. Tsui. A complete analytical solution to the equation TA−FT=LC and its applications. *IEEE Transactions on Automatic Control*, vol. 32, no. 8, pp. 742–744, 1987.

[35] C. C. Tsui. On the solution to matrix equation TA−FT=LC and its applications. *SIAM Journal on Matrix Analysis*, vol. 14, no. 1, pp. 33–44, 1993.

[36] B. Gopinath. On the control of linear multiple input-output systems. *Bell System Technical Journal*, vol. 50, no. 3, pp. 1063–1081, 1971.

[37] R. D. Gupta, F. W. Fairman, T. Hinamoto. A direct procedure for the design of single functional observers. *IEEE Transactions on Circuits and Systems*, vol. 28, no. 4, pp. 294–300, 1981.

[38] P. Van Dooren. Reduced order observers: A new algorithm and proof. *Systems & Control Letters*, vol. 4, no. 5, pp. 243–251, 1984.

[39] R. A. Fowell, D. J. Bender, F. A. Assal. Estimating the plant state from the compensator state. *IEEE Transactions on Automatic Control*, vol. 31, no. 10, pp. 964–967, 1986.

[40] S. Y. Zhang. Generalized functional observer. *IEEE Transactions on Automatic Control*, vol. 35, no. 6, pp. 743–745, 1990.

[41] C. C. Tsui. A new algorithm for the design of multifunctional observers. *IEEE Transactions on Automatic Control*, vol. 30, no. 1, pp. 89–93, 1985.

[42] C. C. Tsui. What is the minimum function observer order? *Journal of the Franklin Institute*, vol. 335, no. 4, pp. 623–628, 1998.

[43] C. C. Tsui. What is the minimum function observer order? In *Proceedings of the European Control Conference*, Cambridge, UK, 2003.

[44] D. C. Lay. *Linear Algebra and Its Applications*, 3rd ed., Boston, USA: Pearson Addison-Wesley, 2006.

[45] C. C. Tsui. On the order reduction of linear function observers. *IEEE Transactions on Automatic Control*, vol. 31, no. 5, pp. 447–449, 1986.

[46] M. Darouach. Existence and design of functional observers for linear systems. *IEEE Transactions on Automatic Control*, vol. 45, no. 5, pp. 940–943, 2000.

[47] T. L. Fernando, H. M. Trinh, L. Jennings. Functional observability and the design of minimum order linear functional observers. *IEEE Transactions on Automatic Control*, vol. 55, no. 5, pp. 1269–1273, 2010.

[48] C. C. Tsui. A new approach of robust observer design. *International Journal of Control*, vol. 47, pp. 745–751, 1988.

[49] B. M. Chen, A. Saberi, P. Sannuti. A new stable compensator design for exact and approximate loop transfer recovery. *Automatica*, vol. 27, no. 2, pp. 257–280, 1991.

[50] S. H. Wang, E. J. Davison, P. Dorato. Observing the states of systems with unmeasurable disturbances. *IEEE Transactions on Automatic Control*, vol. 20, no. 5, pp. 716–717, 1975.

[51] P. Kudva, N. Viswanadham, A. Ramakrishna. Observers for linear systems with unknown inputs. *IEEE Transactions on Automatic Control*, vol. AC-25, pp. 113–115, 1980.

[52] M. Hou, P. C. Muller. Design of observers for linear systems with unknown inputs. *IEEE Transactions Automatic Control*, vol. 37, no. 6,pp. 871–875, 1992.

[53] E. J. Davison, S. H. Wang. Properties and calculation of transmission zeros of linear multivariable systems. *Automatica*, vol. 10, no. 6, pp. 643–658, 1974.

[54] A. Saberi, P. Sannuti. Observer design for loop transfer recovery and for uncertain dynamic systems. *IEEE Transactions Automatic Control*, vol. 35, no. 8, pp. 878–897, 1990.

[55] A. Saberi, B. N. Chen, P. Sannuti. *Loop Transfer Recovery: Analysis and Design*, New York, USA: Springer-Verlag, 1993.

[56] H. Kwakernaak, R. Sivan. *Linea Optimal Control Systems*, New York, USA: Wiley-Intersciences, 1972.

[57] U. Shaked, E. Soroka. On the stability robustness of the continuous-time LQG regulators. *IEEE Transactions on Automatic Control*, vol. 30, no. 10, pp. 1039–1043, 1985.

[58] M. Tahk, J. L. Speyer. Modeling of parameter variations and asymptotic LQG synthesis. *IEEE Transactions on Automatic Control*, vol. 32, no. 9, pp. 793–801, 1987.

[59] M. Y. Fu, Exact, optimal, and partial loop transfer recovery. In *Proceedings of the 29th IEEE Conference on Decision and Control*, Honolulu, HI, USA, pp. 1841–1846, 1990.

[60] C. C. Tsui. On robust observer compensator design. *Automatica*, vol. 24, no. 5, pp. 687–692, 1988.

[61] J. B. Moore, T. T. Tay. Loop recovery via H_∞/H_2 sensitivity recovery. *International Journal of Control*, vol. 49, no. 4, pp. 1249–1271, 1989.

[62] Z. X. Weng, S. J. Shi. H_∞ loop transfer recovery synthesis of discrete-time systems. *International Journal of Robust and Nonlinear Control*, vol. 8, no. 8, pp. 687–697, 1998.

[63] C. C. Tsui. Unified output feedback design and loop transfer recovery. In *Proceedings of the 11th American Control Conference*, Chicago, IL, USA, pp. 3113–3118, 1992.

[64] C. C. Tsui. Unifying state feedback/LTR observer and constant output feedback design by dynamic output feedback. In *Proceedings of the IFAC World Congress*, IFAC, Sydney, Australia, vol. 2, pp. 231–238, 1993.

[65] C. C. Tsui. A new design approach of unknown input observers. *IEEE Transactions on Automatic Control*, vol. 41, no. 3, pp. 464–468, 1996.

[66] C. C. Tsui. The first general output feedback compensator that can implement state feedback control. *International Journal of Systems Science*, vol. 29, no. 1, pp. 49–55, 1998.

[67] C. C. Tsui. A fundamentally novel design approach that defies separation principle. In *Proceedings of the IFAC World Congress*, IFAC, Beijing, China, pp. 283–288, 1999.

[68] C. C. Tsui. Eight irrationalities of basic state space control system design. In *Proceedings of the 6th World Congress on Intelligent Control and Automation*, Dalian, China, pp. 2304–2307, 2006.

[69] C. C. Tsui. High-performance state feedback, robust, and output feedback stabilizing control – A systematic design algorithm. *IEEE Transactions on Automatic Control*, vol. 44, no. 3, pp. 560–563, 1999.

[70] C. C. Tsui. A design example with eigenstructure assignment control whose loop transfer function is fully realized. *Journal of the Franklin Institute*, vol. 336, no. 7, pp. 1049–1053, 1999.

[71] X. A. Wang. Grassmannian, central projection, and output feedback pole assignment of linear systems. *IEEE Transactions on Automatic Control*, vol. 41, no. 6, pp. 786–794, 1996.

[72] P. M. Frank. Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy – A survey and some new results. *Automatica*, vol. 26, no. 3, pp. 459–474, 1990.

[73] J. Gertler. Analysis redundancy methods in fault detection and isolation, survey and synthesis. In *Proceedings of the IFAC Safe-process Symposium*, Boston, Massachusetts, USA, 1991.

[74] A. Emami-Naeini, M. Akhter, S. Rock. Effect of model uncertainty on failure detection, the threshold selector. *IEEE Transactions on Automatic Control*, vol. 33, no. 12, pp. 1106–1115, 1988.

[75] C. C. Tsui, A general failure detection, isolation, and accommodation system with model uncertainty and measurement noise. *IEEE Transactions on Automatic Control*, vol. 39, no. 11, pp. 2318–2321, 1994.

[76] C. C. Tsui. Design generalization and adjustment of a failure isolation and accommodation system. *International Journal of Systems Science*, vol. 28, pp. 91–107, 1997.

[77] C. C. Tsui. On the solution to the state failure detection problem. *IEEE Transactions on Automatic Control*, vol. 34, no. 9, pp. 1017–1018, 1989.

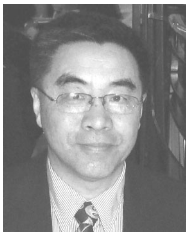
[78] W. Ge, C. F. Fang. Detection of faulty components via robust observation. *International Journal of Control*, vol. 47, no. 2, pp. 581–600, 1988.

[79] H. M. Trinh, P. S. Teh, T. L. Fernando. Time-delay systems design of delay-free and low order observers. *IEEE Transactions on Automatic Control*, vol. 55, no. 10, pp. 2434–2438, 2010.

[80] C. C. Tsui. Observer design for systems with time-delayed states. *International Journal of Automation and Computing*, vol. 9, no. 1, pp. 105–107, 2012.

[81] C. C. Tsui. An overview of the applications and solutions of a fundamental matrix equation pair. *Journal of the Franklin Institute*, vol. 341, no. 6, pp. 465–475, 2004.

[82] G. R. Duan. On the solution to Sylvester matrix equation AV+BW=EVF. *IEEE Transactions on Automatic Control*, vol. 41, no. 4, pp. 612–614, 1996.



Chia-Chi Tsui worked at a state farm in northeast China between 1969 and 1975. He received the B. Sc. degree in computer science from Concordia University, Montreal, Canada in 1979, and received the M. Sc. and Ph. D. degrees from Electrical Engineering Department, State University of New York at Stony Brook in 1980 and 1983, respectively. He has held teaching positions at Northeastern University, City University of New York Staten Island College, and DeVry University New York. Among his numerous publications in control systems theory, he authored two books: *Robust Control System Design – Advanced State Space Techniques* and *Modern Control System Design Theory's New Development* (in Chinese).

His research interests include linear feedback control system design and robust control design.

E-mail: ctsui743@hotmail.com

ORCID iD: 0000-0001-6351-6344