New computational paradigms in solving fault detection and isolation problems

Andreas Varga

German Aerospace Center, DLR-Oberpfaffenhofen, Institute of System Dynamics and Control, D-82234 Wessling, Germany

Abstract

A representative set of fault detection and isolation problems are formulated for linear time-invariant systems with additive faults. For all problems general existence conditions of their solutions are given. An overview of recently developed computational methods for the synthesis of fault detection filters is presented for all formulated problems. Two remarkable computational paradigms emerged in these developments, which are instrumental in developing generally applicable, numerically reliable and computationally efficient synthesis methods. The first paradigm is the use of integrated synthesis algorithms, where the resulting fault detection filters are determined by successive updating of partial syntheses addressing specific requirements. The second paradigm is the use of the nullspace method as a first synthesis step to reduce all synthesis problems to a simple standard form which allows to easily check solvability conditions and address least order synthesis problems.

Keywords: Fault detection, fault detection and isolation, linear synthesis, nullspace method, numerical methods

1. Introduction

There exists a vast literature on fault detection and related problems, which includes several textbooks (Gertler, 1998; Chen and Patton, 1999; Blanke et al., 2003; Saberi et al., 2007) and thousands of journal and conference papers. Attempts for a systematic categorization of different problems have been made in these books as well as in some survey articles (Saberi et al., 2000; Ding et al., 2000), where both exact and approximate solutions to various problems are described. Although the theoretical developments of the fault detection topics for linear time-invariant systems are essentially completed, still we can often observe differences in the formulations of the same problem. This diversity in problem formulations is partly the consequence of employing particular system theoretical frameworks or focusing on particular classes of solution methods.

To face this situation, we start our presentation by providing a set of intuitive formulations of several exact and approximate fault detection problems in form of requirements on specific transfer-function matrix gains. These problem formulations are independent of any possible solution method and allow the derivation of general solvability conditions in terms of ranks of certain transfer-function matrices. An important aspect to emphasise is that the formulations of approximate problems include the formulations of exact problems.

The development of computational methods for solving the synthesis problems of fault detection filters was a constant activity which complemented most theoretical works. Unfortunately, there are many signs for a general lack of numerical awareness in the fault diagnosis community. For example, many of proposed methods employ highly questionable numer-

ical techniques, as polynomial manipulations, operations involving matrix products and powers, or even the computation of highly sensitive canonical forms. All these methods suffer of the lack of *guaranteed* numerical reliability, and may potentially produce inaccurate results even for well-conditioned computational problems. Therefore, they are generally unsuited for solving large order problems. Another class of computational methods includes synthesis procedures, which are strongly tailored to existing computational tools (e.g., H_∞-synthesis methods available in popular environments as MATLAB). Typically, the applicability of these tools is conditioned by technical assumptions, which are however not necessary for solving specific fault detection problems. Another popular synthesis approach involves the use of so-called unknown input observers as fault detection filters. In spite of lack of generality, the underlying synthesis methods (e.g., based on eigenvector assignment or geometric methods) are applicable in many practical cases.

Despite many algorithmic developments, it is rather surprising that, with a few notable exceptions, the vast literature on fault detection until around 2000 contains almost no results on the development of reliable numerical methods along the well established criteria for satisfactory algorithms in the field of linear numerical algebra (Moler and Van Loan, 1978). Because of the lack of generality and/or the lack of numerical reliability, most of popular synthesis techniques of fault detection filters (e.g., parity space methods, geometric methods, unknown input observer based methods) cannot be considered as satisfactory numerical approaches. To remedy this situation, a new generation of numerically reliable computational algorithms has been developed by the author during the last decade. The new algorithms are able to solve various fault detection filter synthesis problems in the most general setting, without any technical assumptions. In this paper we give an overview of these new al-

gorithmic developments, by describing suitable synthesis procedures using the high level input-output description. Details of concrete computational algorithms are discussed using state-space system based descriptions.

In the development of the new computational techniques two computational paradigms emerged, which are instrumental in developing generally applicable, numerically reliable and computationally efficient synthesis methods. The first paradigm is the use of so-called integrated synthesis algorithms, where the resulting fault detection filters are determined by successive updating of partial syntheses addressing specific requirements. Since each partial synthesis represents a valid fault detection filter, this approach has an increased flexibility in combining different synthesis techniques when compared with the traditionally used one-shot techniques. However, the main strength of the integrated algorithms lies in their ability to exploit at each updating step all available structural information at the previous step, which overall leads to very efficient structure exploiting computations.

The second paradigm is the use of the nullspace method (Patton and Hou, 1998; Frisk and Nyberg, 2001; Varga, 2003) as a first synthesis step to simplify and even solve various synthesis problems. The main appeals of the nullspace based fault detection filter synthesis are: generality, being applicable to both standard and singular (or non-proper) systems; numerical reliability, by relying on numerically sound computational techniques; and flexibility, by leading to simplified problem formulations which allow to easily check solvability conditions and address least order synthesis problems.

Structure of the paper

In Section 2 we formulate several exact and approximate fault detection problems and give existence conditions of their solutions. For convenience, we employ an input-output description based formulation, which leads to simple but general existence conditions involving only rank conditions for rational matrices. In Section 3 we present an overview of the latest developments in the synthesis algorithms to solve the formulated fault detection problems. Conceptual algorithms (not suitable for numerical computations) are presented for each of the formulated problems using the input-output formalism. Two computational paradigms can be clearly observed. The first is the development of algorithms based on detector updating techniques, which led to the concept of integrated synthesis algorithms. The second is the important role of the nullspace based approach, as an universal preprocessing step to reduce the formulated fault detection problems to simpler standard forms. Among other advantages, this approach also leads to easy to check solvability conditions. In Section 4, after the definition of the concept of satisfactory algorithm, we discuss the two paradigms (nullspace based preprocessing and integrated algorithms) on the basis of a detailed synthesis procedure for the solution of the exact fault detection problem. Although this is the simplest to solve of the formulated problems, still the synthesis procedure can be seen as a basic computational step in solving the more involved fault isolation problems. The computational

procedures have been developed for systems described by standard or descriptor state space system models. A comparison of several methods to solve the exact fault detection problem is presented. Short presentations of other integrated algorithms conclude this section. The main conclusions and some of recent extensions of the presented techniques are mentioned in the Section 5. Several appendices provide additional details on the theoretical and computational issues discussed in the paper.

2. Fault detection and isolation problems

Consider additive fault models described by input-output representations of the form

$$\mathbf{y}(\lambda) = G_u(\lambda)\mathbf{u}(\lambda) + G_d(\lambda)\mathbf{d}(\lambda) + G_w(\lambda)\mathbf{w}(\lambda) + G_f(\lambda)\mathbf{f}(\lambda), \quad (1)$$

where $\mathbf{y}(\lambda)$, $\mathbf{u}(\lambda)$, $\mathbf{d}(\lambda)$, $\mathbf{w}(\lambda)$, and $\mathbf{f}(\lambda)$ are Laplace-transformed (in the continuous-time case) or Z-transformed (in the discrete-time case) vectors of the p-dimensional system output vector $y(t)$, m_u-dimensional control input vector $u(t)$, m_d-dimensional disturbance vector $d(t)$, m_w-dimensional noise vector $w(t)$ and m_f-dimensional fault vector $f(t)$, respectively, and where $G_u(\lambda)$, $G_d(\lambda)$, $G_w(\lambda)$ and $G_f(\lambda)$ are the *transfer-function matrices* (TFMs) from the control inputs to outputs, disturbance inputs to outputs, noise inputs to outputs, and fault inputs to outputs, respectively. For complete generality of our problem formulations, we will allow that these TFMs are general rational matrices (proper or improper) for which we will not *a priori* assume any further properties.

A linear residual generator (or fault detection filter, or simply fault detector) processes the measurable system outputs $y(t)$ and known control inputs $u(t)$ and generates the residual signals $r(t)$ which serve for decision making on the presence or absence of faults. The input-output form of this filter is

$$\mathbf{r}(\lambda) = Q(\lambda) \begin{bmatrix} \mathbf{y}(\lambda) \\ \mathbf{u}(\lambda) \end{bmatrix} \quad (2)$$

where $Q(\lambda)$ is the TFM of the filter. For a physically realizable filter, $Q(\lambda)$ must be *proper* (i.e., only with finite poles) and *stable* (i.e., only with poles having negative real parts for a continuous-time system or magnitudes less than one for a discrete-time system). The (dynamic) *order* of $Q(\lambda)$ (also known as *McMillan degree*) is the dimension of the state vector of a minimal state-space realization of $Q(\lambda)$. The dimension q of the residual vector $r(t)$ depends on the fault detection problem to be solved. The form (2) of the fault detection filter is called the *implementation form* (Gertler, 1998).

The residual signal $r(t)$ in (2) generally depends via the system outputs $y(t)$ of all system inputs $u(t)$, $d(t)$, $w(t)$ and $f(t)$. The *internal form* of the filter (Gertler, 1998) is obtained by replacing in (2) $\mathbf{y}(\lambda)$ by its expression in (1), and is given by

$$\mathbf{r}(\lambda) = R_u(\lambda)\mathbf{u}(\lambda) + R_d(\lambda)\mathbf{d}(\lambda) + R_w(\lambda)\mathbf{w}(\lambda) + R_f(\lambda)\mathbf{f}(\lambda) \quad (3)$$

where

$$[\, R_u(\lambda) \mid R_d(\lambda) \mid R_w(\lambda) \mid R_f(\lambda) \,] :=$$
$$Q(\lambda) \begin{bmatrix} G_u(\lambda) & G_d(\lambda) & G_w(\lambda) & G_f(\lambda) \\ I_{m_u} & 0 & 0 & 0 \end{bmatrix} \quad (4)$$

2

For a successfully designed filter $Q(\lambda)$, the corresponding internal representation is a proper and stable system, which achieves additionally specific fault detection requirements.

For the solution of fault detection problems it is always possible to completely decouple the control input $u(t)$ from the residual $r(t)$ by requiring $R_u(\lambda) = 0$. Regarding the disturbance input $d(t)$ and noise input $w(t)$ we aim to impose a similar condition on the disturbance input $d(t)$ by requiring $R_d(\lambda) = 0$, while minimizing simultaneously the effect of noise input $w(t)$ on the residual (e.g., by minimizing the norm of $R_w(\lambda)$). Thus, from a practical synthesis point of view, the distinction between $d(t)$ and $w(t)$ lies solely in the way these signals are treated when solving the residual generator synthesis problem.

More precisely, the disturbance inputs in $d(t)$ are additive effects from which exact decoupling of the residuals is presumably possible and is targeted in the detector synthesis. On the other hand, the noise input vector $w(t)$ contains everything else, including proper random noise or "ordinary" disturbances in excess of those which may be exactly decoupled. It may even contain fictive inputs which model the effect of parametric uncertainties in the process model. This distinction between $d(t)$ and $w(t)$ allows to address the solution of both exact (when $G_w(\lambda) = 0$) and approximate (when $G_w(\lambda) \neq 0$) fault detection problems using a unique computational framework.

In all fault detection problems formulated in what follows, we require that by a suitable choice of a stable fault detection filter $Q(\lambda)$, we achieve that the residual signal $r(t)$ is fully decoupled from the control input $u(t)$ and disturbance input $d(t)$. Thus, the following *decoupling conditions* must be generally fulfilled:

$$
\begin{aligned}
(i) \quad & R_u(\lambda) = 0, \\
(ii) \quad & R_d(\lambda) = 0.
\end{aligned}
\tag{5}
$$

For each fault detection problem specific requirements have to be fulfilled, which are formulated as additional conditions in what follows. For all formulated problems we also give the existence conditions of the solutions of these problems.

2.1. Exact fault detection problem – EFDP

The basic additional requirement is simply to achieve by a suitable choice of a fault detection filter $Q(\lambda)$ that in the absence of noise input (i.e., $G_w(\lambda) = 0$), the residual $r(t)$ is influenced by all fault components $f_j(t)$. Let $R_{f_j}(\lambda)$ denote the j-th column of $R_f(\lambda)$. This requirement can be expressed as the following *detection condition* to be fulfilled for all faults:

$$
(iii) \quad R_{f_j}(\lambda) \neq 0, j = 1, \ldots, m_f \ \text{ with } \ R_f(\lambda) \ \text{stable.}
\tag{6}
$$

The solvability conditions of the EFDP have been established for proper systems, see for example (Ding and Frank, 1991; Nyberg, 2002), and are valid for non-proper systems as well:

Theorem 1. *For the system (1) with $G_w(\lambda) = 0$ the EFDP is solvable if and only if*

$$
\text{rank} \, [\, G_{f_j}(\lambda) \ G_d(\lambda) \,] > \text{rank} \, G_d(\lambda), \quad j = 1, \ldots, m_f
\tag{7}
$$

where $G_{f_j}(\lambda)$ denotes the j-th column of $G_f(\lambda)$.

Generically, the condition (7) is fulfilled if $p > m_d$, which basically says that the system must have a sufficiently large number of measurements. For the case $m_d = 0$, this condition reduces to the simple fault input observability conditions:

$$
G_{f_j}(\lambda) \neq 0, \quad j = 1, \ldots m_f
\tag{8}
$$

2.2. Approximate fault detection problem – AFDP

The effects of the noise input $w(t)$ can usually not be fully decoupled from the residual $r(t)$. In this case, the basic requirements for the choice of $Q(\lambda)$ can be expressed to achieve that the residual $r(t)$ is influenced by all fault components $f_j(t)$ and the influence of the noise signal $w(t)$ is negligible. Thus, the following two additional conditions have to be fulfilled:

$$
\begin{aligned}
(iii) \quad & R_{f_j}(\lambda) \neq 0, j = 1, \ldots, m_f \ \text{ with } \ R_f(\lambda) \ \text{stable,} \\
(iv) \quad & R_w(\lambda) \approx 0, \ \text{ with } \ R_w(\lambda) \ \text{stable.}
\end{aligned}
\tag{9}
$$

Here, (iii) is the detection condition of all faults employed also in the EFDP, while (iv) is the *attenuation condition* for the noise input. The condition $R_w(\lambda) \approx 0$ expresses the requirement that the transfer gain $\|R_w(\lambda)\|$ (measured by any suitable norm) can be made arbitrarily small.

The solvability conditions of the formulated AFDP can be easily established:

Theorem 2. *For the system (1) the AFDP is solvable if and only if the EFDP is solvable.*

Proof. We can always determine a solution $Q(\lambda)$ of the EFDP such that additionally the resulting $R_w(\lambda)$ is stable. Moreover, by rescaling $Q(\lambda)$ with a constant factor γ, the norm of $R_w(\lambda)/\gamma$ can be made arbitrarily small. The necessity is trivial, because any solution of the AFDP is also a solution of the EFDP. $\qquad \square$

2.3. Exact strong fault detection and isolation problem – ESFDIP

Let $M_r(\lambda)$ be a given $q \times m_f$ TFM of a reference model (i.e., stable, proper) specifying the desired input-output transfer from the faults to residuals as $\mathbf{r}(\lambda) = M_r(\lambda)\mathbf{f}(\lambda)$. Thus, we want to achieve by a suitable choice of a stable and proper $Q(\lambda)$ satisfying (i) and (ii) in (5), that we have additionally $R_f(\lambda) = M_r(\lambda)$. For example, a typical choice for $M_r(\lambda)$ is an $m_f \times m_f$ diagonal and invertible TFM, which ensures that each residual $r_i(t)$ is influenced only by the fault $f_i(t)$. Such a choice corresponds to the so-called *directional residuals* used in (Gertler, 1998).

To determine $Q(\lambda)$, we have to solve the linear rational equation (4), with the settings $R_u(\lambda) = 0$, $R_d(\lambda) = 0$, and $R_f(\lambda) = M_r(\lambda)$ ($R_w(\lambda)$ and $G_w(\lambda)$ are empty matrices). The choice of $M_r(\lambda)$ may lead to a solution $Q(\lambda)$ which is not proper or is unstable or has both these undesirable properties. Therefore, besides determining $Q(\lambda)$, we also consider the determination of a suitable updating factor $M(\lambda)$ of $M_r(\lambda)$ to ensure the stability and properness of the solution $Q(\lambda)$ for $R_f(\lambda) = M(\lambda)M_r(\lambda)$. Obviously, $M(\lambda)$ must be chosen a proper, stable and invertible TFM. Additionally, $M(\lambda)$ must be chosen diagonal, in order to preserve the zero/nonzero structure encoded in $M_r(\lambda)$ (see also

Section 2.5). The use of a diagonal $M(\lambda)$ with the above properties has been proposed by Gertler (1998) (see also his earlier papers cited in this book).

The formulation of the ESFDIP corresponds to the *model-matching approach* of (Chen and Patton, 1999; Blanke et al., 2003). Accordingly, given a stable and proper $M_r(\lambda)$, it is required to determine a stable and proper filter $Q(\lambda)$ and a diagonal, proper, stable and invertible TFM $M(\lambda)$ such that the following condition is additionally fulfilled:

(iii) $R_f(\lambda) = M(\lambda)M_r(\lambda)$

The solvability condition of the ESFDIP (discussed also in Varga (2004b)) is the standard solvability condition of (under-determined) systems of linear equations:

Theorem 3. *For the system (1) with $G_w(\lambda) = 0$ and a given $M_r(\lambda)$, the ESFDIP is solvable if and only if the following condition is fulfilled*

$$rank\,[\,G_f(\lambda)\ G_d(\lambda)\,] = rank \begin{bmatrix} G_f(\lambda) & G_d(\lambda) \\ M_r(\lambda) & 0 \end{bmatrix} \qquad (10)$$

Proof. A solution $\widetilde{Q}(\lambda)$ which satisfies the linear equations $R_u(\lambda) = 0$, $R_d(\lambda) = 0$ and $R_f(\lambda) = M_r(\lambda)$ exists if the condition (10), representing the solvability condition for this linear system, is fulfilled. A stable solution $Q(\lambda)$ satisfying condition (iii) is obtained by expressing $\widetilde{Q}(\lambda)$ in the left coprime factorized form $\widetilde{Q}(\lambda) = M^{-1}(\lambda)Q(\lambda)$, with $M(\lambda)$ and $Q(\lambda)$ proper and stable TFMs, and $M(\lambda)$ diagonal. $\qquad\square$

When $M_r(\lambda)$ has full column rank m_f, the solvability condition of the ESFDIP has been derived by Frank and Ding (1994). In this case, (10) reduces to

$$rank\,[\,G_f(\lambda)\ G_d(\lambda)\,] = m_f + rank\,G_d(\lambda) \qquad (11)$$

Generically, the condition (11) is fulfilled if $p \geq m_f + m_d$, which implies that the system must have a sufficiently large number of measurements. For the case $m_d = 0$, this condition reduces to the simple left invertibility condition:

$$rank\,G_f(\lambda) = m_f \qquad (12)$$

2.4. Approximate strong fault detection and isolation problem – ASFDIP

The formulation of the ASFDIP extends the model-matching approach of (Chen and Patton, 1999; Blanke et al., 2003), by including the determination/updating of the reference model. Specifically, for a given stable and proper TFM $M_r(\lambda)$, it is required to determine a stable and proper filter $Q(\lambda)$ and a diagonal, proper, stable and invertible TFM $M(\lambda)$ such that the following conditions are additionally fulfilled:

(iii) $R_f(\lambda) \approx M(\lambda)M_r(\lambda)$, with $R_f(\lambda)$ stable;
(iv) $R_w(\lambda) \approx 0$, with $R_w(\lambda)$ stable. $\qquad (13)$

The condition (iii) means that we strive to achieve that $\|R_f(\lambda) - M(\lambda)M_r(\lambda)\| \approx 0$.

A sufficient condition for the solvability of ASFDIP is simply the solvability of the ESFDIP:

Theorem 4. *For the system (1) and a given $M_r(\lambda)$, the ASFDIP is solvable if the ESFDIP is solvable.*

Proof. Let $Q(\lambda)$ and $M(\lambda)$ be a solution of the ESFDIP, which also ensures that $R_w(\lambda)$ is stable. By simultaneously rescaling both $Q(\lambda)$ and $M(\lambda)$ with the same constant factor γ, an arbitrary small norm of $R_w(\lambda)/\gamma$ can be achieved. $\qquad\square$

2.5. Exact weak fault detection and isolation problem – EWFDIP

The ESFDIP is often not solvable, due to the lack of a sufficiently large number of measurements. For fault isolation purposes, we can formulate a weaker alternative to the ESFDIP, by interpreting $M_r(\lambda)$ as a structured TFM aiming to ensure that the residual $r_i(t)$ is influenced by the fault $f_j(t)$, only if the element (i, j) of $M_r(\lambda)$ is non-zero and is decoupled of $f_j(t)$ otherwise. In this case, only the zero-nonzero structure of $M_r(\lambda)$ is relevant and instead $M_r(\lambda)$, the corresponding *structure matrix* $\mathcal{S}(M_r(\lambda))$ with only 0/1 entries can be used to encode these dependencies. The matrix $S = \mathcal{S}(M_r(\lambda))$ is determined as follows:

$$S_{i,j} := \begin{cases} 1, & \text{if } [M_r(\lambda)]_{i,j} \neq 0 \\ 0, & \text{if } [M_r(\lambda)]_{i,j} = 0 \end{cases}$$

According to the nomenclature of Gertler (1998), for a desired structure matrix S, the i-th row of S is called the i-th *specification*, while the j-th column of S is called the j-th *fault signature*. The specifications serve mainly for synthesis purposes (see Section 3.6), while the fault signatures are used mainly for the isolation of specific faults by comparing the signature of fired (nonzero) residuals with those coded in the columns of S. The maximally achievable structure matrix for the system (1) can be determined using a numerically reliable state space based algorithm proposed in (Varga, 2009c).

The formulation of the EWFDIP corresponds to the formulation with *structured residuals* employed in (Gertler, 1998). For a given structure matrix S, it is required to determine a stable and proper filter $Q(\lambda)$ such that the following condition is additionally fulfilled:

(iii) $\mathcal{S}(R_f(\lambda)) = S$, with $R_f(\lambda)$ stable.

The solvability conditions of the EWFDIP can be expressed by formulating for each specification i (i.e., the i-th row of S) the solvability condition corresponding to an EFDP, which arises by redefining each fault signal to be decoupled (i.e., the corresponding element $S_{i,j} = 0$) as a disturbance input. Let $\widehat{G}_d^{(i)}(\lambda)$ be the TFM formed from the columns of $G_f(\lambda)$ for which $S_{i,j} = 0$. We have the following solvability conditions for the EWFDIP:

Theorem 5. *For the system (1) with $G_w(\lambda) = 0$ and a given $q \times m_f$ structure matrix S the EWFDIP is solvable if and only if for $i = 1, \ldots, q$*

$$rank\,[\,G_d(\lambda)\ \widehat{G}_d^{(i)}(\lambda)\ G_{f_j}(\lambda)\,] > rank\,[\,G_d(\lambda)\ \widehat{G}_d^{(i)}(\lambda)\,] \quad (14)$$

for all j such that $S_{i,j} \neq 0$.

2.6. Approximate weak fault detection and isolation problem – AWFDIP

Let S be a given $q \times m_f$ structure matrix and let $\overline{S}$ be the negated S with the (i, j)-th element $\overline{S}_{i,j} := 1 - S_{i,j}$. Let $X \circ Y$ denote the element-wise multiplication of matrices X and Y of same size. The AWFDIP can be formulated as follows. For a given structure matrix S it is required to determine a stable and proper filter $Q(\lambda)$ such that the following conditions are additionally fulfilled:

$$(iii) \quad \mathcal{S}(S \circ R_f(\lambda)) = S , \ \overline{S} \circ R_f(\lambda) \approx 0, \ \text{with } R_f(\lambda) \text{ stable,}$$
$$(iv) \quad R_w(\lambda) \approx 0, \ \text{with } R_w(\lambda) \text{ stable.} \tag{15}$$

It is straightforward to show that a sufficient condition for the solvability of the AWFDIP is the solvability of the EWFDIP.

Theorem 6. *For the system (1) and a given structure matrix S, the AWFDIP is solvable if the EWFDIP is solvable.*

Proof. Let $Q(\lambda)$ be a solution of the EWFDIP, which also ensures that $R_w(\lambda)$ is stable. By rescaling $Q(\lambda)$ with a constant factor γ, arbitrary small norms of $R_w(\lambda)/\gamma$ and $\overline{S} \circ R_f(\lambda)/\gamma$ can be achieved. $\square$

3. Nullspace based synthesis algorithms

In this section we present an overview of the recently developed synthesis procedures of fault detection filters to solve the formulated fault detection problems. In these developments two computational paradigms emerged, which are instrumental in developing generally applicable, numerically reliable and computationally efficient synthesis methods. The first paradigm is the use of so-called integrated synthesis algorithms, where the resulting fault detection filters are determined by successive updating of partial syntheses addressing specific requirements. In all synthesis algorithms described in this section it is possible to express the TFM of the final filter in a factored form as

$$Q(\lambda) = Q_K(\lambda) \cdots Q_2(\lambda) Q_1(\lambda) \tag{16}$$

where $Q_1(\lambda)$, $Q_2(\lambda)Q_1(\lambda)$, ..., can be interpreted as partial syntheses addressing specific requirements. Since each partial synthesis may represent a valid fault detection filter, this approach has a high flexibility in using or combining different synthesis techniques.

The second paradigm is the use of the nullspace method as a first synthesis step to reduce all synthesis problems to a simple standard form which allows to easily check solvability conditions and address least order synthesis problems. In this section, we present solution procedures based on the nullspace method for each of the formulated fault detection problems. The concrete numerical aspects of using nullspace methods are described in the next section, where two examples of integrated algorithms are also presented.

3.1. Nullspace based reduction to standard form

Assume first that $m_u + m_d > 0$ and consider $Q(\lambda)$ in a factored form

$$Q(\lambda) = \overline{Q}_1(\lambda) Q_1(\lambda), \tag{17}$$

where $Q_1(\lambda)$ is a proper left rational nullspace basis satisfying $Q_1(\lambda)G(\lambda) = 0$, where $G(\lambda)$ is defined as

$$G(\lambda) = \begin{bmatrix} G_u(\lambda) & G_d(\lambda) \\ I_{m_u} & 0 \end{bmatrix} \tag{18}$$

and $\overline{Q}_1(\lambda)$ is a factor to be subsequently determined. With this choice it follows that $Q(\lambda)$ automatically fulfills the two conditions (i) and (ii) in (5), namely $R_u(\lambda) = 0$ and $R_d(\lambda) = 0$, intervening in all fault detection problems formulated in the previous section. The resulting $Q_1(\lambda)$ has full row rank $p - r_d$, where $r_d = \text{rank} \, G_d(\lambda)$. Thus, the existence condition of a non-empty rational nullspace basis $Q_1(\lambda)$ is simply $r_d < p$. This condition is guaranteed provided the existence conditions of Theorems 1–6 are fulfilled.

The fault detector in (2) can be rewritten as

$$\mathbf{r}(\lambda) = \overline{Q}_1(\lambda) Q_1(\lambda) \begin{bmatrix} \mathbf{y}(\lambda) \\ \mathbf{u}(\lambda) \end{bmatrix} = \overline{Q}_1(\lambda) \overline{\mathbf{y}}(\lambda) \tag{19}$$

where

$$\overline{\mathbf{y}}(\lambda) := Q_1(\lambda) \begin{bmatrix} \mathbf{y}(\lambda) \\ \mathbf{u}(\lambda) \end{bmatrix} = \overline{G}_f(\lambda)\mathbf{f}(\lambda) + \overline{G}_w(\lambda)\mathbf{w}(\lambda) \tag{20}$$

with

$$[\, \overline{G}_f(\lambda) \ \overline{G}_w(\lambda) \,] := Q_1(\lambda) \begin{bmatrix} G_f(\lambda) & G_w(\lambda) \\ 0 & 0 \end{bmatrix} \tag{21}$$

We can choose $Q_1(\lambda)$ stable and such that both $\overline{G}_f(\lambda)$ and $\overline{G}_w(\lambda)$ defined in (21) are proper and stable TFMs (Varga, 2008).

If $m_u = m_d = 0$, we can determine $Q_1(\lambda)$ simply from a proper and stable left coprime factorization

$$[\, G_f(\lambda) \ G_w(\lambda) \,] = Q_1^{-1}(\lambda)[\, \overline{G}_f(\lambda) \ \overline{G}_w(\lambda) \,] \tag{22}$$

With this first preprocessing step, we reduced the original problems formulated for the system (1) to ones formulated for the reduced system (20) without control and disturbance inputs, and we have to determine the TFM $\overline{Q}_1(\lambda)$ of the simpler detector (19). At this stage we can assume that the reduced system (20) is proper and even stable, provided $Q_1(\lambda)$ has been chosen a stable TFM (always possible).

The main facts established in (Forney, 1975; Kailath, 1980) about nullspace bases (both polynomial and rational) are summarized in Appendix A. The so-called *simple rational bases* are useful when solving fault detection problems with least order detectors. For example, building linear combinations of basis vectors to obtain least order candidate detectors is very straightforward. Moreover, left multiplication with diagonal $M(\lambda)$ is possible without altering the simple structure of

the basis. Unfortunately, the computation of simple bases involves delicate computational steps leading to potential accuracy losses (Varga, 2011c). As an alternative to simple rational bases, we use *minimal proper rational bases*, which can be computed by using a numerically stable reduction method of a linear matrix pencil as described in the Appendix B.

In what follows, we discuss the synthesis of fault detection filters which solve the formulated fault detection problems and we give simpler existence conditions for the solutions of the formulated exact fault detection problems, which can serve for checking the existence of solutions. An important aspect for detector synthesis is the determination of residual generators with least order dynamical orders. Only very few methods proposed in the literature are able to address this aspect.

3.2. Solving the EFDP

The solvability conditions of Theorem 1 lead for the reduced system (20) without control and disturbance inputs to the following simplification:

Corollary 1. *For the system (20) with $\overline{G}_w(\lambda) = 0$ the EFDP is solvable if and only if*

$$\overline{G}_{f_j}(\lambda) \neq 0, \quad j = 1, \dots m_f \tag{23}$$

Any stable and proper rational nullspace basis $Q_1(\lambda)$ already represents a solution of the EFDP, provided the conditions (23) are fulfilled. In what follows, we discuss the computation of a special solution called the *least order* solution of the EFDP, which is a solution having the least possible McMillan degree. The basic synthesis procedure determines $\overline{Q}_1(\lambda)$ in (17) in the factored form

$$\overline{Q}_1(\lambda) = Q_3(\lambda)Q_2(\lambda),$$

where $Q_2(\lambda)$ is determined such that $Q_2(\lambda)Q_1(\lambda)$ has least order and $Q_3(\lambda)$ is determined such that the overall filter $Q(\lambda) = Q_3(\lambda)Q_2(\lambda)Q_1(\lambda)$ and $R_f(\lambda) = Q_3(\lambda)Q_2(\lambda)\overline{G}_f(\lambda)$ are stable. Additionally we must ensure that $Q(\lambda)$ is *admissible*, that is, the conditions (6) are fulfilled.

In what follows we sketch a synthesis procedure for a least order scalar output detector (i.e., with $q = 1$), which serves only to understand the basic principles, but evidently is not suitable as basis of a reliable computational procedure. Assume $Q_1(\lambda)$ is a simple proper rational basis (see Appendix A) formed of $p - r_d$ rational row vectors $v_i(\lambda)/d_i(\lambda)$, where $v_i(\lambda)$ is a polynomial vector of degree n_i and $d_i(\lambda)$ is a stable polynomial of degree n_i. We assume that the basis vectors $v_i(\lambda)$ are ordered such that $n_1 \leq n_2 \leq \dots \leq n_{p-r_d}$ and each $d_i(\lambda)$ divides $d_j(\lambda)$ for $i < j$. It follows immediately, that a linear combination $h^{(i)}Q_1(\lambda)$ of the first i rows with $h^{(i)}$ of the form $h^{(i)} = [h_1, \dots, h_i, 0, \dots, 0]$, $h_i \neq 0$, has a McMillan degree n_i. Thus, choosing the least index i such that $h^{(i)}Q_1(\lambda)$ is admissible, solves the synthesis problem with $Q_2(\lambda) = h^{(i)}$ and $Q_3(\lambda) = 1$. State space computations based algorithms for the least order synthesis will be described in details in Section 4.2.

A short historical note is appropriate in this place. The nullspace method (without using this naming) in a state space

based formulation has been originally employed in (Patton and Hou, 1998) to solve the EWFDIP using structured residuals and extended in (Hou, 2000) to general non-regular descriptor systems. The least order synthesis problem has been apparently addressed for the first time in (Frisk and Nyberg, 2001), where a minimal polynomial basis based solution has been proposed. In this work, a product form parametrization of all possible residual generators to solve the EFDP has been also introduced. The polynomial basis method is applicable to non-proper systems as shown in (Nyberg and Frisk, 2006). A numerically reliable state space approach to the least order synthesis relying on rational nullspace bases has been proposed in (Varga, 2003). Computational details of this approach in a state space based setting are discussed in Section 4.2 and Appendix B.

3.3. Solving the AFDP

We have seen in the proof of Theorem 2, that the solution of the EFDP can be used to solve the formulated AFDP. However, better solutions can be obtained by trying to maximize the gap between the fault detectability and noise attenuation requirements. An optimization based approach can be used to achieve this goal. Let $\gamma > 0$ be an admissible level for the influence of the noise signal $w(t)$ in the residual $r(t)$, which can be measured using suitable norms of $R_w(\lambda) = \overline{Q}_1(\lambda)\overline{G}_w(\lambda)$, as for example the H_2- or H_∞-norms. As a measure of the fault detectability, several definitions of the index $\|R_f(\lambda)\|_-$ have been used in the literature. The definitions used in (Ding et al., 2000; Jaimoukha et al., 2006; Liu and Zhou, 2007, 2008; Zhang and Ding, 2008) are in terms of the least singular values of the frequency-response of $R_f(\lambda) = \overline{Q}_1(\lambda)\overline{G}_f(\lambda)$ and therefore are meaningful only when $m_f \leq p$. An alternative measure $\|R_f(\lambda)\|_{2/\infty-}$ has been used in (Varga, 2009a; Glover and Varga, 2011)

$$\|R_f(\lambda)\|_{2/\infty-} = \min_{1 \leq j \leq m_f} \|R_{f_j}(\lambda)\|_{2/\infty}, \tag{24}$$

The requirement $\|R_f(\lambda)\|_{2/\infty-} > 0$ merely asks for non-zero columns $R_{f_j}(\lambda)$ and thus is equivalent to the fault *detection condition* (6).

The $H_-/H_{2/\infty}$ optimization problem to be solved is the following: Given $\gamma > 0$, determine $\beta > 0$ and a stable and proper fault detection filter $\overline{Q}_1(\lambda)$ such that

$$\beta = \max_{\overline{Q}_1(\lambda)} \left\{ \left\| \overline{Q}_1(\lambda)\overline{G}_f(\lambda) \right\|_{2/\infty-} \ \middle| \ \left\| \overline{Q}_1(\lambda)\overline{G}_w(\lambda) \right\|_{2/\infty} \leq \gamma \right\}$$

Numerically reliable computational methods to solve these optimization problems have been proposed in (Varga, 2009a; Glover and Varga, 2011). Computational procedures relying on alternative definitions of the $\| \cdot \|_-$ index have been proposed in (Ding et al., 2000; Jaimoukha et al., 2006; Liu and Zhou, 2007, 2008; Zhang and Ding, 2008).

The procedure of (Varga, 2009a) determines $\overline{Q}_1(\lambda)$ in the form $\overline{Q}_1(\lambda) = \overline{Q}_2(\lambda)Q_2(\lambda)$, where $Q_2(\lambda)$ ensures that $Q_2(\lambda)\overline{G}_w(\lambda)$ has full row rank and $Q_2(\lambda)Q_1(\lambda)$ has least dynamical order. The *standard case* is when $Q_2(\lambda)\overline{G}_w(\lambda)$ has no unstable zeros on the boundary of the stability domain (i.e., the extended imaginary axis in the continuous-time case, or the unit

circle centered in the origin in the discrete-time case). The *non-standard case* corresponds to the presence of such zeros.

At the next step, using the dual of the algorithm of Oară and Varga (2000) for the continuous-time case and the dual of the algorithm of Oară (2005) for the discrete-time case, we compute the quasi-outer–co-inner factorization

$$Q_2(\lambda)\overline{G}_w(\lambda) = G_{wo}(\lambda)G_{wi}(\lambda),$$

where the quasi-outer factor $G_{wo}(\lambda)$ is an invertible TFM which has only stable zeros, excepting possible zeros on the boundary of the stability domain, and $G_{wi}(\lambda)$ is co-inner (i.e., $G_{wi}(\lambda)G^*_{wi}(\lambda) = I$ with $G^*_{wi}(s) = G^T_{wi}(-s)$ in the continuous-time case, and $G^*_{wi}(z) = G^T_{wi}(1/z)$ in the discrete-time case).

In the standard case, $\overline{Q}_2(\lambda)$ is determined in the form $\overline{Q}_2(\lambda) = Q_4(\lambda)Q_3(\lambda)$, where $Q_3(\lambda) = G^{-1}_{wo}(\lambda)$ and $Q_4(\lambda)$ is the optimal solution (or an approximation of it) of the reduced problem

$$\beta = \max_{Q_4(\lambda)}\{\|Q_4(\lambda)\widetilde{G}_f(\lambda)\|_{2/\infty-}\ |\ \|Q_4(\lambda)\|_{2/\infty} \le \gamma\}$$

where $\widetilde{G}_f(\lambda) = Q_3(\lambda)Q_2(\lambda)\overline{G}_f(\lambda)$.

In the non-standard case, $\overline{Q}_2(\lambda) = Q_3(\lambda)$ is determined by directly solving

$$\beta = \max_{Q_3(\lambda)}\{\|Q_3(\lambda)\widehat{G}_f(\lambda)\|_{2/\infty-}\ |\ \|Q_3(\lambda)G_{wo}(\lambda)\|_{2/\infty} \le \gamma\}$$

where $\widehat{G}_f(\lambda) = Q_2(\lambda)\overline{G}_f(\lambda)$. Suitable algorithms for this purpose have been developed in (Glover and Varga, 2011).

The order of the resulting residual generator in the standard case, is the order of $Q_3(\lambda)$ if we choose $Q_4(\lambda)$ a constant matrix. This order results from the conditions that $Q_2(\lambda)\overline{G}_w(\lambda)$ has full row rank and $Q_2(\lambda)Q_1(\lambda)$ has least order and is admissible (i.e., $\|Q_2(\lambda)\overline{G}_f(\lambda)\|_- > 0$). For each candidate $Q_2(\lambda)$, the corresponding optimal $Q_3(\lambda)$ results automatically, but the different "optimal" detectors for the same level γ of noise attenuation performance can have significantly differing fault detection performance levels (measured via the optimal cost β). Finding the best compromise between achieved order and the achieved performance (measured via the gap β/γ), should take into account that it is to be expected that larger orders and larger number of detector outputs q lead to better performance.

3.4. Solving the ESFDIP

The solvability condition of Theorem 3 leads for the reduced system (20) to the following simplification:

Corollary 2. *For the system (20) with $\overline{G}_w(\lambda) = 0$ and a given $M_r(\lambda)$, the ESFDIP is solvable if and only if the following condition is fulfilled*

$$\mathrm{rank}\,\overline{G}_f(\lambda) = \mathrm{rank}\begin{bmatrix} \overline{G}_f(\lambda) \\ M_r(\lambda) \end{bmatrix} \tag{25}$$

In the case when $M_r(\lambda)$ has full column rank m_f, the solvability condition (11) reduces to the left invertibility condition

$$\mathrm{rank}\,\overline{G}_f(\lambda) = m_f,$$

which is also a sufficient condition for arbitrary $M_r(\lambda)$.

In the general case, we can determine $\overline{Q}_1(\lambda)$ in the factored form

$$\overline{Q}_1(\lambda) = Q_3(\lambda)Q_2(\lambda),$$

where $Q_2(\lambda)$ is determined such that

$$Q_2(\lambda)\overline{G}_f(\lambda) = M_r(\lambda) \tag{26}$$

and has least order, while the diagonal factor $Q_3(\lambda) := M(\lambda)$ is determined such that $Q(\lambda) = Q_3(\lambda)Q_2(\lambda)Q_1(\lambda)$ is stable and proper. A state-space realizations based algorithm for solving the linear rational equation (26) is described in Appendix C. Since the resulting partial detector $Q_2(\lambda)Q_1(\lambda)$ is possibly non-proper, for the determination of $M(\lambda)$, coprime factorization techniques for non-proper systems in generalized state-space (descriptor system) representations have to be used. Suitable algorithms to compute proper and stable coprime factorizations have been proposed in (Varga, 1998). Each diagonal element of $M(\lambda)$ can be independently obtained by applying these algorithms to the corresponding row of $Q_2(\lambda)Q_1(\lambda)$.

An alternative approach for the solution of the ESFDIP is to determine $Q(\lambda)$ as $Q(\lambda) = Q_2(\lambda)Q_1(\lambda)$, where $Q_1(\lambda)$ is the least order solution of the linear system

$$Q_1(\lambda)\begin{bmatrix} G_u(\lambda) & G_d(\lambda) & G_f(\lambda) \\ I_{m_u} & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & M_r(\lambda) \end{bmatrix} \tag{27}$$

computed using the computational algorithm described in Appendix C, while the diagonal TFM $Q_2(\lambda) := M(\lambda)$ is determined to ensure that $Q(\lambda)$ is proper and stable. Further details on this procedure are presented in (Varga, 2004b). The main advantage of this approach is that it allows the determination of detectors of least order.

An important particular case in practical applications is when $M_r(\lambda)$ is diagonal, stable, proper and invertible. In this case, the solution of the ESFDIP allows the detection of several simultaneous faults. Fault reconstruction or fault estimation problems can be addressed by choosing $M_r(\lambda) = I_{m_f}$. The resulting detector for these cases has the explicit form

$$Q(\lambda) = M(\lambda)M_r(\lambda)\overline{G}^+_f(\lambda)Q_1(\lambda), \tag{28}$$

where $\overline{G}^+_f(\lambda)$ is a left inverse of $\overline{G}_f(\lambda)$ and $M(\lambda)$ is determined such that $Q(\lambda)$ is stable and proper. This synthesis method is sometimes called in the literature the *inversion based method*. For numerical computations, an integrated synthesis algorithm is proposed in Section 4.5. Another approach widely studied in the literature is to solve this particular ESFDIP for $M_r(\lambda) = I_{m_f}$ by using the synthesis method to solve EWFDIP (see Section 3.6), with the associated diagonal structure matrix $S = I_{m_f}$. The solution consists of a bank of m_f residual generators with scalar outputs, designed by using synthesis methods for solving the EFDP.

A historical note is appropriate in this place. The *inversion based method* has been pioneered by Gertler and Monajemy (1995) to solve the ESFDIP for discrete-time systems for the particular case when $G_d(z) = 0$ and $G_f(z)$ is invertible. The developed approach can be interpreted as a particular instance

of the general approach above, where the employed nullspace basis is $Q_1(z) = [\, I_p \;\; -G_u(z)\,]$, and thus $\overline{G}_f(z) = G_f(z)$. The computation of the inverse of $\overline{G}_f(z)$ relies on the *explicit* inversion of a first degree polynomial matrix $\Gamma(z)$ of potentially large order, representing the system matrix pencil associated to a state-space representation of $\overline{G}_f(z)$. The inverse is expressed as a quotient of the polynomial adjoint matrix and the determinant of $\Gamma(z)$. $M(z)$ is determined as a product of two diagonal rational matrices, which separately ensure the properness and stability (e.g., dead-beat behaviour) of the overall detector $Q(z) = M(z)M_r(z)\overline{G}_f^{-1}(z)Q_1(z)$. The main difficulty with the underlying computational techniques discussed in (Gertler, 2000) is the need for manipulating polynomials (via their coefficients or roots) and polynomial matrices, which can significantly worsen the sensitivity of the synthesis problem (e.g., in the case of systems with multiple zeros) and thus lead to numerical instability (Van Dooren, 1981).

3.5. Solving the ASFDIP

To fulfill the last two conditions in (13) we can solve for a fixed $M(\lambda)$ a $H_{2/\infty}$-norm minimization problem to determine a stable and proper $\overline{Q}_1(\lambda)$ such that

$$\left\| \left[\, \overline{Q}_1(\lambda)\overline{G}_f(\lambda) - M(\lambda)M_r(\lambda) \;\; \overline{Q}_1(\lambda)\overline{G}_w(\lambda)\,\right] \right\|_{2/\infty} = \min \quad (29)$$

This $H_{2/\infty}$ model matching problem can be easily reformulated as a standard $H_{2/\infty}$-norm minimization based "controller" synthesis problem (Zhou et al., 1996) for which software tools exist, as for example, the functions `h2syn`/`hinfsyn` available in the MATLAB ROBUST CONTROL Toolbox. The main problem when employing standard tools like `h2syn`/`hinfsyn`, is that, although a stable and proper solution of the ASFDIP may exist (according to Theorem 4), this solution cannot be computed because of an inappropriate choice of $M(\lambda)M_r(\lambda)$ or because of the need to fulfill some technical assumptions.

To face the above limitations, general synthesis procedures have been developed for which no such limitations exist. The key parameter to guarantee the stability and properness of the detector is $M(\lambda)M_r(\lambda)$, the desired TFM relating the faults to the residuals. Often, good candidates for $M(\lambda)$ and $M_r(\lambda)$ result from an exact synthesis (for $G_w(\lambda) = 0$) (Varga, 2004b). However, in (Varga, 2005) a synthesis procedure has been proposed, where the choice of a suitable updating factor $M(\lambda)$ of an initial $M_r(\lambda)$ is part of the solution process. This procedure has been refined in (Varga, 2010, 2011a), by developing an integrated approach for the fault detector synthesis. An important feature of these computational approaches is that they rely on repeated updating of an initial fault detector. The final detector can be thus expressed in a factored form as in (16).

In what follows, we present a synthesis procedure based on solving the $H_{2/\infty}$-norm minimization problem (29). For this purpose, we assume first $M(\lambda) = I_q$ and solve the $H_{2/\infty}$-norm minimization problem for $\|\overline{\mathcal{R}}(\lambda)\|_{2/\infty}$ to determine $\overline{Q}_1(\lambda)$, where

$$\overline{\mathcal{R}}(\lambda) := \overline{F}(\lambda) - \overline{Q}_1(\lambda)\overline{G}(\lambda), \quad (30)$$

with $\overline{G}(\lambda) = [\,\overline{G}_f(\lambda)\;\;\overline{G}_w(\lambda)\,]$ and $\overline{F}(\lambda) = [\, M_r(\lambda)\;\; O\,]$. We assume that $M_r(\lambda)$ is the TFM of a given reference model, and therefore is stable and proper.

Let ℓ be the rank of the $(p - r_d) \times (m_f + m_w)$ TFM $\overline{G}(\lambda)$. If $\ell < p - r_d$ (i.e., $\overline{G}(\lambda)$ has no full row rank), we take $\overline{Q}_1(\lambda) = \overline{Q}_2(\lambda)Q_2(\lambda)$, where $Q_2(\lambda)$ is a $\ell \times (p - r_d)$ TFM used to build ℓ linear combinations of basis vectors which ensure that $Q_2(\lambda)\overline{G}(\lambda)$ has full row rank ℓ. A suitable choice of $Q_2(\lambda)$ which also minimizes the McMillan degree of $Q_2(\lambda)Q_1(\lambda)$ is described in (Varga, 2008, 2011c) (see also Section 4.4).

At the next step, using the dual of the algorithm of Oară and Varga (2000) for the continuous-time case and the dual of the algorithm of Oară (2005) for the discrete-time case, we compute the quasi-outer–co-inner factorization

$$Q_2(\lambda)\overline{G}(\lambda) = [\, G_{o,1}(\lambda)\; 0\,] \left[\begin{array}{c} G_{i,1}(\lambda) \\ G_{i,2}(\lambda) \end{array}\right] := G_o(\lambda)G_i(\lambda), \quad (31)$$

where the quasi-outer factor $G_{o,1}(\lambda)$ is an invertible TFM which has only stable zeros, excepting possible zeros on the boundary of the stability domain, and $G_i(\lambda)$ is a square co-inner factor.

$\overline{Q}_2(\lambda)$ is determined in the form $\overline{Q}_2(\lambda) = \overline{Q}_3(\lambda)Q_3(\lambda)$, with $Q_3(\lambda) = G_{o,1}^{-1}(\lambda)$ and $\overline{Q}_3(\lambda)$ as the optimal solution which minimizes $\|\widetilde{\mathcal{R}}(\lambda)\|_{2/\infty}$, where

$$\widetilde{\mathcal{R}}(\lambda) := \overline{\mathcal{R}}(\lambda)G_i^*(\lambda) = \left[\; \widetilde{F}_1(\lambda) - \overline{Q}_3(\lambda) \;\middle|\; \widetilde{F}_2(\lambda)\;\right]$$

with $\widetilde{F}_1(\lambda) := \overline{F}(\lambda)G_{i,1}^*(\lambda)$ and $\widetilde{F}_2(\lambda) := \overline{F}(\lambda)G_{i,2}^*(\lambda)$. Since $G_i(\lambda)$ is a co-inner TFM, we have $\|\widetilde{\mathcal{R}}(\lambda)\|_{2/\infty} = \|\overline{\mathcal{R}}(\lambda)\|_{2/\infty}$.

The overall detection filter $Q(\lambda)$ has the product form

$$Q(\lambda) = \overline{Q}_3(\lambda)Q_3(\lambda)Q_2(\lambda)Q_1(\lambda) := \overline{Q}_3(\lambda)\overline{Q}(\lambda),$$

where $\overline{Q}(\lambda) = Q_3(\lambda)Q_2(\lambda)Q_1(\lambda)$ can be interpreted as a partial detector, to which correspond the TFMs of the partial internal representation

$$[\,\overline{R}_f(\lambda)\;\overline{R}_w(\lambda)\,] := Q_3(\lambda)Q_2(\lambda)[\,\overline{G}_f(\lambda)\;\overline{G}_w(\lambda)\,] \quad (32)$$
$$= [\, I_\ell\;\; 0\,] \left[\begin{array}{c} G_{i,1}(\lambda) \\ G_{i,2}(\lambda) \end{array}\right] = G_{i,1}(\lambda)$$

It is shown in (Varga, 2012a) using state-space realizations based manipulations that pole-zero cancelations occur in forming $\overline{Q}(\lambda)$, such that $\overline{Q}(\lambda)$ has as only poles the zeros of $G_{o,1}(\lambda)$. The *standard case* corresponds to the situation when $G_{o,1}(\lambda)$ has only stable finite zeros and no infinite zeros, in which case $\overline{Q}(\lambda)$ is also stable. The *non-standard case* is when some poles correspond to zeros of $G_{o,1}(\lambda)$ on the boundary of the stability domain (i.e., on the extended imaginary axis in the continuous-time case, or on the unit circle in discrete-time case).

The determination of a stable and proper $\overline{Q}_3(\lambda)$ which minimizes $\|\widetilde{\mathcal{R}}(\lambda)\|_{2/\infty}$ is a least-distance problem (Francis, 1987). We determine $\overline{Q}_3(\lambda)$ in the factored form $\overline{Q}_3(\lambda) = Q_5(\lambda)Q_4(\lambda)$, where depending on the employed norm, different solutions for $Q_4(\lambda)$ have to be computed.

In the standard-case, we take $M(\lambda) := Q_5(\lambda) = I$ and determine the stable and proper $Q_4(\lambda) = \overline{Q}_3(\lambda)$ which minimizes $\|\widetilde{\mathcal{R}}(\lambda)\|_{2/\infty}$. In the non-standard-case, we determine $Q_4(\lambda)$ as in

the standard-case and determine a diagonal, invertible, stable and proper $M(\lambda) := Q_5(\lambda)$ such that

$$Q(\lambda) = Q_5(\lambda)Q_4(\lambda)Q_3(\lambda)Q_2(\lambda)Q_1(\lambda)$$

is stable. For this purpose, coprime factorization techniques based on state-space algorithms as described in (Varga, 1998) can be employed.

The solution of the least distance problem for $\overline{Q}_3(\lambda) = Q_4(\lambda)$ in the case of H_2-norm is straightforward and involves computing the stable projection

$$Q_4(\lambda) = \{\widetilde{F}_1(\lambda)\}_+$$

where $\{\cdot\}_+$ denotes the stable proper part of the underlying TFM.

In the case of H_∞-norm, we determine a stable $Q_4(\lambda)$ as the solution of the suboptimal two-blocks least distance problem

$$\left\| \left[\ \widetilde{F}_1(\lambda) - Q_4(\lambda)\ \ \widetilde{F}_2(\lambda)\ \right] \right\|_\infty < \gamma, \tag{33}$$

where $\gamma_{opt} < \gamma \leq \gamma_{opt} + \varepsilon$, with ε an arbitrary user specified (accuracy) tolerance for the least achievable value γ_{opt} of γ. With the following lower and upper bounds for γ_{opt}

$$\gamma_l = \|\widetilde{F}_2(\lambda)\|_\infty, \quad \gamma_u = \left\| \left[\ \widetilde{F}_1(\lambda)\ \ \ \widetilde{F}_2(\lambda)\ \right] \right\|_\infty \tag{34}$$

such a γ-suboptimal solution $Q_4(\lambda)$ can be computed using the bisection-based γ-iteration approach of Francis (1987). For further details and state-space based computational formulas see (Varga, 2011a, 2012a)

3.6. Solving the EWFDIP

For a given specification i (contained in the i-th row of S), we define a new disturbance input $\hat{d}^{(i)}$, which contains all components f_j of f for which $S_{ij} = 0$, define a new fault input $\hat{f}^{(i)}$, which contains all components f_j for which $S_{ij} \neq 0$ and we build $\widehat{G}_d^{(i)}(\lambda)$ and $\widehat{G}_f^{(i)}(\lambda)$ from the corresponding set of columns of $\overline{G}_f(\lambda)$. We can rewrite the reduced system (20) as

$$\overline{\mathbf{y}}(\lambda) = \widehat{G}_d^{(i)}(\lambda)\hat{\mathbf{d}}^{(i)}(\lambda) + \widehat{G}_f^{(i)}(\lambda)\hat{\mathbf{f}}^{(i)}(\lambda) + \overline{G}_w(\lambda)\mathbf{w}(\lambda) \tag{35}$$

The solvability conditions of Theorem 5 lead for the reduced system (20) to the following simplification:

Corollary 3. *For the system (20) with $\overline{G}_w(\lambda) = 0$ and a given $q \times m_f$ structure matrix S, the EWFDIP is solvable if and only if for $i = 1, 2, \ldots, q$ the following conditions are fulfilled*

$$rank\,[\ \widehat{G}_d^{(i)}(\lambda)\ \overline{G}_{f_j}(\lambda)\ ] > rank\,\widehat{G}_d^{(i)}(\lambda), \tag{36}$$

for all j such that $S_{i,j} \neq 0$.

For the reduced system (20) with $\overline{G}_w(\lambda) = 0$ we can determine, using the standard approach to solve the EFDP, a bank of q scalar output detectors of least orders, each of the form

$$\mathbf{r}_i(\lambda) = \overline{Q}^{(i)}(\lambda)\overline{\mathbf{y}}(\lambda),\ i = 1, \ldots, q \tag{37}$$

such that each detector $\overline{Q}^{(i)}(\lambda)$ solves the EFDP for the rewritten system (35) with $\overline{G}_w(\lambda) = 0$ and thus achieves the i-th specification coded in the i-th row of S. The TFM of the final detector can be assembled from q individual detectors as

$$Q(\lambda) = \begin{bmatrix} Q^{(1)}(\lambda) \\ \vdots \\ Q^{(q)}(\lambda) \end{bmatrix} := \begin{bmatrix} \overline{Q}^{(1)}(\lambda) \\ \vdots \\ \overline{Q}^{(q)}(\lambda) \end{bmatrix} Q_1(\lambda) \tag{38}$$

According to the synthesis procedure for scalar output detectors described in Section 3.2, each individual detector $\overline{Q}^{(i)}(\lambda)$ can be determined in the factored form

$$\overline{Q}^{(i)}(\lambda) = Q_3^{(i)}(\lambda)Q_2^{(i)}(\lambda)Q_1^{(i)}(\lambda)$$

where $Q_1^{(i)}(\lambda)$ is a left proper nullspace basis of $\widehat{G}_d^{(i)}(\lambda)$, $Q_2^{(i)}(\lambda)$ is a rational vector used to build a linear combination of the basis vectors in $Q_1^{(i)}(\lambda)$, and $Q_3^{(i)}(\lambda)$ is a stable transfer function which ensures a desired dynamics for the i-th detector.

3.7. Solving the AWFDIP

For a given specification i (contained in the i-th row of S), we can rewrite the reduced system (20) analogously to Section 3.6 in the form (35). Using the approach to solve the AFDP in Section 3.3, we determine a bank of q scalar output least order detectors of the form (37), such that each detector $\overline{Q}^{(i)}(\lambda)$ achieves approximately the i-th specification coded in the i-th row of S. For this purpose, we can solve the following $H_-/H_{2/\infty}$ optimization problem: Given $\gamma > 0$, determine $\beta^{(i)} > 0$ and a stable and proper fault detection filter $\overline{Q}^{(i)}(\lambda)$ such that

$$\beta^{(i)} = \max_{\overline{Q}^{(i)}(\lambda)} \left\{ \left\|\widehat{R}_f^{(i)}(\lambda)\right\|_{2/\infty-} \,\middle|\, \left\| \left[\widehat{R}_d^{(i)}(\lambda)\ \widehat{R}_w^{(i)}(\lambda)\right] \right\|_{2/\infty} \leq \gamma \right\},$$

where $\widehat{R}_f^{(i)}(\lambda) = \overline{Q}^{(i)}(\lambda)\widehat{G}_f^{(i)}(\lambda)$, $\widehat{R}_d^{(i)}(\lambda) = \overline{Q}^{(i)}(\lambda)\widehat{G}_d^{(i)}(\lambda)$, and $\widehat{R}_w^{(i)}(\lambda) = \overline{Q}^{(i)}(\lambda)\overline{G}_w^{(i)}(\lambda)$. The TFM of the final detector can be assembled from q individual detectors as in (38), where in accordance with the synthesis procedure described in Section 3.3, each individual detector $Q^{(i)}(\lambda)$ is determined in a factored form.

Concluding remark

An important aspect worth to mention regarding the new synthesis algorithms to solve the approximate synthesis problems AFDP, ASFDIP and AWFDIP is that the main focus in developing these algorithms lies not on solving the associated optimization problems but on obtaining "useful" solutions of these synthesis problems, in the most general setting and using reliable numerical techniques. Although the proposed solution approaches in (Varga, 2009a, 2010, 2011a) follow the usual solution processes to determine the optimal solutions, still the resulting filters are usually not optimal in the non-standard cases. The characterization of "useful" solutions must take into account the actual signal bounds on the contribution of noise inputs in the residual signal, which determines the minimum detectable amplitudes of fault signals. For a pertinent discussion of these aspects see (Gertler, 1998).

4. Computational aspects

In this section we formulate first the requirements for satisfactory computational algorithms to solve the synthesis problems of fault detection filters. Then the two computational paradigms already mentioned in Section 3 are discussed. First, the nullspace based preprocessing is presented and the details of a computational algorithm are given in Appendix B. Then, the concept of an integrated algorithm is explained in terms of simultaneous updating of the implementation and internal forms of the fault detection filter. An integrated algorithm to solve the EFDP is presented with details as an example. As it will be apparent, the main strength of the integrated algorithms lies in their ability to exploit at each updating step all available structural information at the previous step, which overall leads to very efficient structure exploiting computations. We also discuss two alternative approaches to solve the EFDP: the *parity space* and the *observer* based methods. As it will be shown, none of them can be considered as a satisfactory numerical approach. The main reasons for this are the lack of generality and/or the lack of numerical reliability. We also describe a second integrated algorithm to solve the ESFDIP and indicate some integrated algorithms to solve the rest of formulated problems.

4.1. Developing satisfactory synthesis algorithms

Before discussing the new computational paradigms for computational algorithms for the synthesis of fault detection filters, we first briefly recall three key requirements for a *satisfactory* numerical algorithm: generality, numerical stability, and efficiency (Moler and Van Loan, 1978). A method which is considered satisfactory can then serve as basis for a robust numerical software implementation.

A *general* synthesis algorithm is one which has no limitations for its applicability of any technical nature. It follows, that it should be applicable without any other limitation than the rank or gain conditions expressing the solvability conditions of various fault detection problems given in Theorems 1–6 in the terms of the TFMs in (1). For example, a synthesis method to solve the EFDP which is applicable only if $G_d(\lambda)$ has no unstable zeros or is only applicable to *strictly* proper systems can not be considered satisfactory. A desirable feature in this context is that the synthesis algorithms should be applicable regardless the underlying system (1) is proper or not in both continuous- and discrete-time settings.

Numerical stability (more precisely, *backward stability*) of an algorithm means that the results computed by that algorithm are exact for slightly perturbed original data. As a consequence, a numerically stable algorithm applied to a well conditioned problem will produce guaranteed accurate results. This is why numerical stability is a key feature for a satisfactory algorithm. A basic ingredient to achieve numerical stability is the exclusive use of orthogonal transformations. The use of these transformations often leads to bounds for perturbations of the initial data which are equivalent to the cumulative effect of round-off errors occurring during the computations. This is an important way to prove the numerical stability of such an algorithm.

A main technique used to solve complex problems is to perform orthogonal transformations on the original problem data to reduce the original problem to an equivalent one which is easier to solve (e.g., reduction to condensed forms). The use of non-orthogonal transformations should generally be avoided, unless there are guarantees for their well-conditioning. Notoriously ill-conditioned transformations arise during reductions to canonical forms (Jordan, Kronecker, etc.) and therefore must be completely avoided. For similar reasons, algorithms involving computational detours, as for example, model conversions to polynomial forms, cannot be considered satisfactory.

Because of the intrinsic complexity of many computational problems, it is not always possible to develop numerically stable algorithms for them. Numerically reliable algorithms are those algorithms, which are either numerically stable or the loss of numerical stability during the computations can be easily detected. Therefore one often imposes only the more general requirement of *numerical reliability* on the individual steps of the algorithm. Although this is not sufficient to guarantee numerical reliability of the global algorithm, one can still expect that it will perform accurately on well-conditioned problems.

The *efficiency* of an algorithm requires a computational complexity of at most $O(n^3)$, where n is the largest dimension in the problem formulation, which typically is the dimension of the underlying state space realization of the input-output model (1). For example, performing repeatedly rank determinations of large matrices may lead to a complexity of $O(n^4)$, which can represent an unacceptable computational burden for large order systems.

The above requirements rule out many popular detector synthesis approaches as candidates for satisfactory numerical methods. In what follows, we shortly mention other aspects which are important in developing satisfactory computational methods to solve fault detection problems.

4.1.1. Choosing adequate system representations

Choosing the right system representation is of primordial importance for reliable numerical computations. This is why, for developing reliable computational algorithms for the synthesis of fault detection filters we employ, instead of the input-output representation (1), an equivalent *descriptor* state-space realization

$$
\begin{aligned}
E\lambda x(t) &= Ax(t) + B_u u(t) + B_d d(t) + B_w w(t) + B_f f(t) \\
y(t) &= Cx(t) + D_u u(t) + D_d d(t) + D_w w(t) + D_f f(t)
\end{aligned}
\tag{39}
$$

with the n-dimensional state vector $x(t)$. Here, $\lambda x(t) = \dot{x}(t)$ or $\lambda x(t) = x(t+1)$ depending on the type of the system, continuous or discrete, respectively. The corresponding TFMs in (1) are

$$
\begin{aligned}
G_u(\lambda) &= C(\lambda E - A)^{-1} B_u + D_u \\
G_d(\lambda) &= C(\lambda E - A)^{-1} B_d + D_d \\
G_w(\lambda) &= C(\lambda E - A)^{-1} B_w + D_w \\
G_f(\lambda) &= C(\lambda E - A)^{-1} B_f + D_f
\end{aligned}
\tag{40}
$$

or in an equivalent notation

$$
\begin{bmatrix} G_u(\lambda) \ G_d(\lambda) \ G_w(\lambda) \ G_f(\lambda) \end{bmatrix} := \left[\begin{array}{c|cccc} A - \lambda E & B_u & B_d & B_w & B_f \\ \hline C & D_u & D_d & D_w & D_f \end{array} \right]
$$

We can always assume that the representation (39) is *minimal*, that is, the descriptor pair $(A - \lambda E, C)$ is *observable* and the pair $(A - \lambda E, [B_u \; B_d \; B_w \; B_f])$ is *controllable*. While this assumption is useful to determine fault detectors of least orders, still it is not a necessary requirement for any of the presented algorithms.

Employing the input-output descriptions via TFMs or matrix polynomial quotients raises serious concerns, because of the intrinsic ill-conditioning of polynomial based representations. In this context, any computational detour involving model conversions to these forms is highly susceptible to numerical instability (Van Dooren, 1981). Even if E is non-singular but potentially ill-conditioned, the reduction of the system (39) to a standard state space form (i.e., with $E = I_n$), should be avoided.

4.1.2. Checking existence conditions

Reliably checking the existence conditions of the solution of various problems is part of any synthesis algorithm. These checks can be performed at the beginning using the original system data from (39) and testing the rank conditions of Theorems 1–6, or can be performed after the first nullspace based preprocessing step by testing the simpler conditions in Corollaries 1–3. In this latter case, it is important that the problem conditioning is not worsened during the preprocessing step. Therefore, only reductions based on orthogonal transformations should be allowed before testing the existence conditions. As we will see, this aim is fully fulfilled using nullspace based computational methods.

4.1.3. Addressing least order synthesis problems

The synthesis of least order fault detection filters is a desirable goal for all synthesis methods and contributes substantially to the reduction of real-time computational demands when implementing these filters. Because of the lack of suitable information on the achievable least orders, only few of the currently employed synthesis techniques are able to address the synthesis of least order fault detection filters. The nullspace based methods have a privileged position among all synthesis methods, because they can benefit from the available full information regarding the possible filter orders. The lack of such knowledge leads to a trial and error process, where typically fault detection filters of increasing orders are tested until they fulfil the design requirements. Other methods, as parity space methods, can address least order synthesis problems, by repeatedly computing nullspace bases of large matrices of increasing sizes. This may lead to unacceptable computational efforts for large order systems.

4.1.4. Optimal tuning of free parameters

Supporting the optimal tuning of free parameters of the fault detection filters is desirable in achieving the best possible fault detection performance. Typical free parameters are the poles of the fault detection filter which impose the speed of the fault detection process, the relative gains used to detect several faults, or simply the relative scaling factors used in the case of a bank of fault detection filters. The optimal choice of these parameters is clearly desirable and the methods described in the previous section usually provide access to the free parameters. Still, the optimal tuning aspect of free parameters has been until now not widely discussed in the literature.

4.2. Nullspace based numerical methods

As apparent from the description of the new generation of synthesis techniques in Section 3, the nullspace method as a first synthesis step serves to reduce all synthesis problems to simple standard forms which allow to easily check solvability conditions and address least order synthesis problems. In this section, we show how this reduction can be performed by employing numerically stable algorithms, applicable to both standard and singular (or non-proper) systems.

Consider the *descriptor* state space realization (39) of the system (1). The $(p + m_u) \times (m_u + m_d)$ TFM $G(\lambda)$ defined in (18) has the descriptor realization

$$G(\lambda) = \left[\begin{array}{c|cc} A - \lambda E & B_u & B_d \\ \hline C & D_u & D_d \\ 0 & I_{m_u} & 0 \end{array} \right]$$

If $m_u + m_d > 0$, let r_d be the rank of $G_d(\lambda)$. Using the method described in Appendix B, we compute a minimal proper left nullspace basis $Q_1(\lambda)$ of $G(\lambda)$, with minimal order $\underline{n}$. The state space realization of the $(p - r_d) \times (p + m_u)$ TFM $Q_1(\lambda)$ is given by

$$Q_1(\lambda) = \left[\begin{array}{c|c} A_l - \lambda E_l & B_l \\ \hline C_l & D_l \end{array} \right], \tag{41}$$

where the pair $(A_l - \lambda E_l, C_l)$ is observable and E_l is non-singular. Additionally, we can assume that E_l is upper triangular and the pair $(A_l - \lambda E_l, C_l)$ is in a staircase form as in (B.5).

A state space realization of the reduced system (20) can be computed as (Varga, 2011c)

$$[\overline{G}_f(\lambda) \; \overline{G}_w(\lambda)] = \left[\begin{array}{c|cc} A_l - \lambda E_l & \overline{B}_f & \overline{B}_w \\ \hline C_l & \overline{D}_f & \overline{D}_w \end{array} \right], \tag{42}$$

where the row partitioning of

$$Q \left[\begin{array}{cc} B_f & B_w \\ D_f & D_w \\ 0 & 0 \end{array} \right] = \left[\begin{array}{cc} * & * \\ \overline{B}_f & \overline{B}_w \\ \overline{D}_f & \overline{D}_w \end{array} \right]$$

is analogous to the column partitioning in (B.3).

As it can be observed, the state space realizations of the implementation form of the partial detector $Q_1(\lambda)$ and the corresponding internal form characterized by the TFMs $\overline{G}_f(\lambda)$ and $\overline{G}_w(\lambda)$ share the matrix triple (E_l, A_l, C_l), where the $\underline{n} \times \underline{n}$ matrix E_l is invertible (thus all three TFMs are proper) and thus the pair (A_l, E_l) has only finite generalized eigenvalues. If all eigenvalues are stable, then the detector $Q_1(\lambda)$ represents a solution of the EFDP or a (non-optimal) solution of the AFDP, provided the solvability conditions (23) are fulfilled.

If $m_u = m_d = 0$ and $[G_f(\lambda) \; G_w(\lambda)]$ is not proper, then realizations of the forms (41) and (42) can be obtained using the technique proposed in (Varga, 2011c) to determine the

proper factorization (22) by computing a proper nullspace basis $[\, Q_1(\lambda)\ \overline{G}_f(\lambda)\ \overline{G}_w(\lambda)\,]$ satisfying

$$[\, Q_1(\lambda)\ \overline{G}_f(\lambda)\ \overline{G}_w(\lambda)\,] \begin{bmatrix} G_f(\lambda) & G_w(\lambda) \\ -I_{m_f} & 0 \\ 0 & -I_{m_w} \end{bmatrix} = 0$$

The algorithm of Appendix B can be used for this purpose. Once again, the three TFMs $Q_1(\lambda)$, $\overline{G}_f(\lambda)$ and $\overline{G}_w(\lambda)$ will share the matrix triple (E_l, A_l, C_l).

The reduced form (42) allows to easily check the existence conditions of a solution to the EFDP and AFDP. According to Corollary 1, we have to check that $\overline{G}_{f_i}(\lambda) \neq 0$ for $i = 1, 2, \ldots, m_f$. Since the pair $(A_l - \lambda E_l, C_l)$ is observable, checking this condition is equivalent to verify that

$$\begin{bmatrix} \overline{B}_{f_i} \\ \overline{D}_{f_i} \end{bmatrix} \neq 0, \quad i = 1, \ldots, m_f, \tag{43}$$

where $\overline{B}_{f_i}$ and $\overline{D}_{f_i}$ denote the i-th columns of $\overline{B}_f$ and $\overline{D}_f$, respectively.

The existence conditions of Corollary 2 of a solution for the ESFDIP can be checked as part of the solution algorithm presented in Appendix C. To solve the linear rational matrix equation (26), a second orthogonal reduction to a Kronecker-like form is necessary, which allows to check the compatibility condition simply by checking the simple condition (C.8). Since the resulting $Q_2(\lambda)$ may be improper and/or unstable, a third updating step of the detector may be necessary to ensure the stability of $Q(\lambda)$. As already mentioned in Section 3.4, the solution of the ESFDIP can be also addressed by directly solving the system of linear equations (27) (i.e., without performing the explicit computation of a left nullspace basis). This approach is to be preferred if the least synthesis aspect has to be also addressed, for which a suitable algorithm is described in (Varga, 2004b).

The existence conditions of Corollary 3 of a solution for the EWFDIP can be checked by performing q separate checks for the q synthesis problems (e.g., one for each specification coded in the rows of the structure matrix S). For the i-th specification, this involves a left nullspace computation to decouple all the faults f_j for which $S_{i,j} = 0$ and forming a reduced system of the form (20), for which conditions of the form (43) must be checked.

4.3. Integrated synthesis algorithms

Typical algorithmic approaches in control and fault detection can be described as a composition of modularized computational steps, where the problem structure between steps is often not exploited. In contrast, an integrated approach consists of computational steps which are connected at a finer granularity level by exploiting all structural information at the termination of each step. The expected advantages of this approach are: more efficient computations with less overhead and less storage, and also an increased computational reliability, because the structural aspects can be fully exploited. The price for this is a more involved algorithmic development involving careful analysis of structural features and a restricted modularization/reusability of the software implementations.

The synthesis algorithms presented in the previous section determines the TFM of the fault detection filter in a factored form as in (16), where each factor corresponds to a typical computational step. A non-integrated (modular) approach would lead to high order detector models (possibly non-minimal) which is less suited for real-time implementations. In contrast, an integrated synthesis approach would rely on fault detection filter updating techniques, where all cancellations are performed by determining explicit minimal order state space realizations. In such an approach, the nullspace basis $Q_1(\lambda)$ could serve for the initialization of the updating process $Q(\lambda) = Q_1(\lambda)$, while at the successive computational steps $Q(\lambda)$ is replaced by $Q_i(\lambda)Q(\lambda)$, for $i = 2, \ldots, K$. In a similar way, it is possible to perform the updating of the internal form of the detector, where only the non-zero TFMs $R_f(\lambda)$ and $R_w(\lambda)$ must be updated (because the initial choice already guarantees $R_u(\lambda) = 0$ and $R_d(\lambda) = 0$). Thus a typical integrated algorithm for the synthesis of fault detection filters has the following conceptual form:

1. Compute a left proper nullspace basis $Q(\lambda)$ of $G(\lambda)$ in (18) and compute

$$[\, R_f(\lambda)\ R_w(\lambda)\,] := Q(\lambda) \begin{bmatrix} G_f(\lambda) & G_w(\lambda) \\ 0 & 0 \end{bmatrix}$$

2. For $i = 2, \ldots, K$: Determine $Q_i(\lambda)$ in accordance with the goal of the i-th synthesis step and update the design

$$[\, Q(\lambda)\ R_f(\lambda)\ R_w(\lambda)\,] \leftarrow Q_i(\lambda)[\, Q(\lambda)\ R_f(\lambda)\ R_w(\lambda)\,]$$

In the Section 4.2, explicit state space realizations of the initial detector $Q(\lambda) = Q_1(\lambda)$, and of the initial $R_f(\lambda) = \overline{G}_f(\lambda)$ and initial $R_w(\lambda) = \overline{G}_w(\lambda)$ have been derived as

$$[\, Q(\lambda)\ R_f(\lambda)\ R_w(\lambda)\,] = \left[\begin{array}{c|ccc} A_l - \lambda E_l & B_l & \overline{B}_f & \overline{B}_w \\ \hline C_l & D_l & \overline{D}_f & \overline{D}_w \end{array} \right]$$

These realizations have the remarkable property of sharing the observable pair $(A_l - \lambda E_l, C_l)$. This feature is also a very desirable goal to be ensured at the successive computational steps (as long as possible or meaningful). The main advantage when fulfilling this goal is that the stability of the fault detection filter in its implementation form guarantees automatically its stability in the internal form.

Integrated algorithms fulfilling the above goals have been proposed to solve the EFDP (Varga, 2003, 2008, 2011c), the AFDP (Varga, 2009a; Glover and Varga, 2011), the EWFDIP (Varga, 2007b, 2011c) and ASFDIP (Varga, 2010, 2011a). For the solution of the ESFDIP the direct method of (Varga, 2004b) is able to compute the least order solution, while for the AWFDIP the same methods as for the AFDP can be employed. In what follows we describe the development of an integrated algorithm for solving the EFDP, which exhibits all desirable features mentioned previously. A second example addresses the development of a new integrated algorithm to solve a class of ESFDIP by using the inversion based approach.

4.4. Integrated algorithm for solving the EFDP

We discuss the computation of a least order scalar output fault detection filter which solves the EFDP. A synthesis procedure has been already described in the Section 3.2. Here we give some computational details for state space algorithms which can be used to determine the prefilter $Q_2(\lambda)$, to achieves the least order dynamics of the detector, and a second filter $Q_3(\lambda)$, to stabilize the fault detection filter or to impose a desired dynamics of the filter. The determination of each filter exploits the information available from the previous computational step and implements the updating technique described previously.

We choose $Q_2(\lambda)$ of the particular form

$$Q_2(\lambda) = \left[\begin{array}{c|c} A_l + KC_l - \lambda E_l & K \\ \hline HC_l & H \end{array} \right], \tag{44}$$

where H is given and K has to be determined. It is easy to check that $Q_2(\lambda)Q(\lambda)$ has the state space realization

$$Q_2(\lambda)Q(\lambda) = \left[\begin{array}{c|c} A_l + KC_l - \lambda E_l & B_l + KD_l \\ \hline HC_l & HD_l \end{array} \right] \tag{45}$$

We determine the output injection matrix K such that $Q_2(\lambda)Q(\lambda)$ has the least possible McMillan degree. Taking into account Proposition 3 (see Appendix B), this can be only achieved by the cancellation of the maximum number of unobservable poles. For this purpose, minimal dynamic cover techniques can be employed to perform this computation.

Computational procedures of minimal dynamic covers are presented in (Varga, 2004c). The general idea of the cover algorithms is to perform a preliminary orthogonal similarity transformation on the system matrices in (41) by applying a special version of the controllability staircase form algorithm of (Varga, 1990) to the controllable descriptor pair $\left(A_l^T - \lambda E_l^T, \left[C_l^T H^T \ C_l^T \right] \right)$ and then with additional block permutations and non-orthogonal block row/column transformations, to put the transformed system matrices in a *special form* which allows to cancel the maximum number of unobservable eigenvalues. For the so-called Type I dynamic covers (Kimura, 1977), two non-singular transformation matrices U and V result such that

$$U(A_l - \lambda E_l)V = \left[\begin{array}{c|c} \widehat{A}_{11} - \lambda \widehat{E}_{11} & \widehat{A}_{12} - \lambda \widehat{E}_{12} \\ \hline \widehat{A}_{21} & \widehat{A}_{22} - \lambda \widehat{E}_{22} \end{array} \right],$$

$$UB_l = \left[\begin{array}{c} \widehat{B}_1 \\ \hline \widehat{B}_2 \end{array} \right], \quad \left[\begin{array}{c} C_l \\ \hline HC_l \end{array} \right]V = \left[\begin{array}{c|c} \widehat{C}_{11} & \widehat{C}_{12} \\ \hline 0 & \widehat{C}_{22} \end{array} \right],$$

where the pairs $(\widehat{A}_{11} - \lambda \widehat{E}_{11}, \widehat{C}_{11})$ and $(\widehat{A}_{22} - \lambda \widehat{E}_{22}, \widehat{C}_{22})$ are observable, and the submatrices $\widehat{C}_{11}$ and $\widehat{A}_{21}$ have the particular structure

$$\left[\begin{array}{c} \widehat{A}_{21} \\ \widehat{C}_{11} \end{array} \right] = \left[\begin{array}{cc} 0 & A_{21} \\ 0 & C_{11} \end{array} \right]$$

with C_{11} having full column rank. By taking

$$K = U^{-1} \left[\begin{array}{c} 0 \\ K_2 \end{array} \right]$$

with K_2 satisfying $K_2 C_{11} + A_{21} = 0$, we annihilate $\widehat{A}_{21} + K_2 \widehat{C}_{11}$, the (2,1)-block of $U(A_l + KC_l)V$, and thus make all eigenvalues of $\widehat{A}_{11} - \lambda \widehat{E}_{11}$ unobservable. The resulting $Q_2(\lambda)Q(\lambda)$ of least McMillan degree, obtained by deleting the unobservable part, has the state space realization

$$Q_2(\lambda)Q(\lambda) = \left[\begin{array}{c|c} \widehat{A}_{22} + K_2 \widehat{C}_{12} - \lambda \widehat{E}_{22} & \widehat{B}_2 + K_2 D_l \\ \hline \widehat{C}_{22} & HD_l \end{array} \right] \tag{46}$$

Since $Q_2(\lambda)R_f(\lambda)$ has the state space realization

$$Q_2(\lambda)R_f(\lambda) = \left[\begin{array}{c|c} A_l + KC_l - \lambda E_l & \overline{B}_f + K\overline{D}_f \\ \hline HC_l & H\overline{D}_f \end{array} \right],$$

it is straightforward to show that after deleting the unobservable part we obtain

$$Q_2(\lambda)R_f(\lambda) = \left[\begin{array}{c|c} \widehat{A}_{22} + K_2 \widehat{C}_{12} - \lambda \widehat{E}_{22} & \widehat{B}_{f,2} + K_2 \overline{D}_f \\ \hline \widehat{C}_{22} & H\overline{D}_f \end{array} \right],$$

where

$$U\overline{B}_f = \left[\begin{array}{c} \widehat{B}_{f,1} \\ \hline \widehat{B}_{f,2} \end{array} \right]$$

To simplify the notations, after performing the updating operations $Q(\lambda) \leftarrow Q_2(\lambda)Q(\lambda)$ and $R_f(\lambda) \leftarrow Q_2(\lambda)R_f(\lambda)$, we denote the resulting state-space realizations as

$$[Q(\lambda) \ R_f(\lambda)] = \left[\begin{array}{c|cc} \widetilde{A}_l - \lambda \widetilde{E}_l & \widetilde{B}_l & \widetilde{B}_f \\ \hline \widetilde{C}_l & \widetilde{D}_l & \widetilde{D}_f \end{array} \right] \tag{47}$$

Similar formulas apply to update $R_w(\lambda)$ (if necessary).

The crucial issue in determining $Q_2(\lambda)$ is the choice of H. One aspect is that each valid choice of H must guarantee that the resulting $Q(\lambda)$ is admissible. Thus, after performing the minimal dynamic cover based order reduction we need to additionally check

$$\left[\begin{array}{c} \widetilde{B}_{f_i} \\ \widetilde{D}_{f_i} \end{array} \right] \neq 0, \quad i = 1, \dots, m_f \tag{48}$$

If this admissibility test is not fulfilled, the choice of H is rejected.

The second aspect is a suitable choice guaranteeing the least order for an admissible detector. The simplest choice is in the case of a scalar output filter, when we can choose H as a structured row vector of length $p - r_d$

$$H = h^{(i)} := [0, \dots, 0, h_i, \dots, h_1], \tag{49}$$

with $h_i \neq 0$. This choice takes into account the staircase form of the pair $(A_l - \lambda E_l, C_l)$ in (B.5) and, according to (Varga, 2011c, Corollary 1.), leads to an order n_i for the detector $Q(\lambda)$. A systematic search can be performed by generating successive candidates for $h^{(i)}$ with increasing number of non-zero trailing elements and checking for the resulting detector the admissibility conditions (48). These detectors have non-decreasing orders and thus the first admissible detector represents a satisfactory least order design. To speed up the selection, the choice of

13

$i = \mu_0 - \mu_i$ non-zero components of $h^{(i)}$ ensures a tentative order n_i, by building a linear combination of all $\mu_0 - \mu_i$ basis vectors of orders less than or equal to n_i. In this way, repeated checks for the same order are avoided and the search is terminated in at most ℓ steps, where ℓ is the number of diagonal blocks in the staircase form (B.5). Similar ideas can be employed when H is restricted to have a certain row dimension $q > 1$ (e.g., when the updated $R_w(\lambda)$ must have full row rank q).

The algorithm to compute the output injection matrix K is *not* numerically stable, because involves non-orthogonal matrix operations. Still, this algorithm can be categorized as a *numerically reliable* algorithm, because the loss of numerical stability can be easily detected either by detecting large norms of the employed transformation matrices U and V, or a large norm of the resulting K. In both cases, the remedy is usually to generate a new H with the same structure (usually randomly) or to increase the targeted order. For further details, see (Varga, 2004c). For the resulting admissible value of i, it is possible to determine the optimal choice of $h^{(i)}$, which minimizes the norms of the above matrices.

The resulting $\widetilde{A}_l - \lambda\widetilde{E}_l$ may have "bad" eigenvalues, which are either unstable or provide unsatisfactory dynamics for the detector $Q(\lambda)$. For the final design, the resulting dynamics of the detector can be arbitrarily assigned by choosing $Q_3(\lambda)$ in the form

$$Q_3(\lambda) = \left[\begin{array}{c|c} \widetilde{A}_l + \widetilde{K}\widetilde{C}_l - \lambda\widetilde{E}_l & \widetilde{K} \\ \hline \widetilde{C}_l & I_q \end{array} \right]$$

After performing the updating operations

$$[\, Q(\lambda)\ R_f(\lambda)\,] \leftarrow Q_3(\lambda)[\, Q(\lambda)\ R_f(\lambda)\,],$$

the resulting state-space realizations are

$$[\, Q(\lambda)\ R_f(\lambda)\,] = \left[\begin{array}{c|cc} \widetilde{A}_l + \widetilde{K}\widetilde{C}_l - \lambda\widetilde{E}_l & \widetilde{B}_l + \widetilde{K}\widetilde{D}_l & \widetilde{B}_f + \widetilde{K}\widetilde{D}_f \\ \hline \widetilde{C}_l & \widetilde{D}_l & \widetilde{D}_f \end{array} \right]$$

The output injection matrix $\widetilde{K}$ is determined such that $\widetilde{A}_l + \widetilde{K}\widetilde{C}_l - \lambda\widetilde{E}_l$ has only "good" eigenvalues. Note that the least order of the detector $Q(\lambda)$ resulting from the previous step is automatically preserved by the detector updating.

The computation of $\widetilde{K}$ and of the corresponding $Q_3(\lambda)[\, Q(\lambda)\quad R_f(\lambda)\,]$ can be performed using numerically reliable coprime factorization techniques as those proposed in (Varga, 1998). In these algorithms, partial pole assignment is performed in a recursive manner by successively moving the "bad" eigenvalues to desired "good" locations. The state space realization of the resulting numerator TFM $Q_3(\lambda)[\, Q(\lambda)\ R_f(\lambda)\,]$ involves an additional (implicitly performed) coordinate transformation with orthogonal matrices W and Z, such that the resulting $W(\widetilde{A}_l + \widetilde{K}\widetilde{C}_l - \lambda\widetilde{E}_l)Z$ is in a generalized real Schur (or quasi-upper triangular) form. The transformations are automatically performed on all system matrices. See (Varga, 1998) for further computational details.

The choice of a suitable H as in (49) at the previous computational step clearly influences the choice of $\widetilde{K}$ at the final

step. These steps can be easily embedded into an optimization-based tuning procedure to determine an optimal H which ensures a more uniform sensitivity of the detector to individual faults. For this purpose, the *sensitivity condition* introduced in (Gertler, 1998, page 353) can be used as a criterion to be minimized. For a given $R_f(\lambda)$, this criterion is defined as

$$\xi := \max_j \|R_{f_j}(\lambda)\|_\infty / \min_j \|R_{f_j}(\lambda)\|_\infty.$$

Alternatively, a similar *sensitivity condition* can be defined in terms of the gains at a selected frequency λ_s as

$$\xi^s := \max_j \|R_{f_j}(\lambda_s)\|_2 / \min_j \|R_{f_j}(\lambda_s)\|_2.$$

A large value of the sensitivity condition indicates potential difficulties in detecting faults due to a substantial gap between the maximum and minimum gains.

We conclude this section with a comparison of existing computational techniques for solving the EFDP.

4.4.1. Parity space method

The parity relations based approach has been pioneered in (Chow and Willsky, 1984; Lou et al., 1986) for standard discrete-time systems and since then widely discussed in the literature. Extension to standard continuous-time case is done in (Magni and Mouyon, 1994). The parity method primarily targets standard systems (with $E = I$), although an extension to descriptor systems has been proposed in the discrete-time case (Maquin et al., 1993).

The basic approach relies on expressing the input-output dynamics in the form

$$\bar{y}(t) = S_{0,k}x(t) + S_{u,k}\bar{u}(t) + S_{d,k}\bar{d}(t) + S_{f,k}\bar{f}(t)$$

where

$$S_{0,k} = \left[\begin{array}{c} C \\ CA \\ \vdots \\ CA^k \end{array} \right], \quad S_{u,k} := \left[\begin{array}{cccc} D_u & 0 & \cdots & 0 \\ CB_u & D_u & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ CA^{k-1}B_u & \cdots & CB_u & D_u \end{array} \right]$$

and similarly for $S_{d,k}$ and $S_{f,k}$; and

$$\bar{y}(t) := \left[\begin{array}{cccc} y^T(t) & \lambda y^T(t) & \cdots & \lambda^k y^T(t) \end{array} \right]^T$$

and similarly for $\bar{u}(t)$, $\bar{d}(t)$, and $\bar{f}(t)$.

A pseudo-residual is computed as

$$\widetilde{r}(t) = H_k(\bar{y}(t) - S_{u,k}\bar{u}(t)), \tag{50}$$

where H_k is a so-called *parity matrix* with $p(k + 1)$ columns. The main computation is the determination of H_k as a left annihilator satisfying

$$H_k\left[\begin{array}{cc} S_{0,k} & S_{d,k} \end{array} \right] = 0$$

This computation is performed for $k = \nu + 1, \nu + 2, \ldots$, until $H_k S_{f_i,k} \neq 0$, $i = 1, \ldots, m_f$, where ν is the observability index of pair (A, C).

The parity space approach appears to be very simple and also allows to search systematically for a solution of least order. However, the computation of the parity matrix H_k needs to explicitly form the large matrices $S_{0,k}$, $S_{u,k}$, $S_{d,k}$ and $S_{f,k}$, which involves evaluating matrix powers and products (e.g., CA^jB) and leads (almost always) to *numerically unstable* computations. All rank determinations based on these matrices are potentially unreliable (Paige, 1981). In light of these facts, the parity space method can not be considered a satisfactory numerical method.

4.4.2. Observer based methods

Observer based methods proposed for solving the EFDP use so-called diagnostic observers to generate the residual signals. For simplicity, we only discuss standard state space systems for which most of synthesis methods have been developed. The observer based approach has been pioneered by Beard (1971) and Jones (1971) to solve fault isolation problems for standard state space systems using full order Luenberger-type observers. As residual generator, the following Luenberger-type output signal observer is used

$$\lambda \hat{x}(t) = A\hat{x}(t) + B_u u(t) - K(y(t) - C\hat{x}(t) - D_u u(t))$$
$$r(t) = H(y(t) - C\hat{x}(t) - D_u u(t)) \quad , \quad (51)$$

where the matrices K and H are chosen to achieve, via a *stable* observer, the conditions $(ii) - (iii)$ of the EFDP. Note that this form of the observer automatically fulfills the decoupling condition (i) for the control inputs (i.e., $R_u(\lambda) = 0$).

The residual generator in internal form results by introducing $e(t) = x(t) - \hat{x}(t)$ to obtain

$$\lambda e(t) = (A+KC)e(t) + (B_f+KD_f)f(t) + (B_d+KD_d)d(t)$$
$$r(t) = HCe(t) + HD_f f(t) + HD_d d(t)$$

To fulfill requirements (ii) and (iii) of the EFDP, the TFM $R_d(\lambda)$ from d to r must be zero, i.e.

$$R_d(\lambda) = HC(\lambda I - A - KC)^{-1}(B_d+KD_d) + HD_d = 0 \quad (52)$$

and each TFM $R_{f_i}(\lambda)$ from the fault input f_i to r for $i = 1, \ldots, m_f$ must be non-zero

$$R_{f_i}(\lambda) = HC(\lambda I - A - KC)^{-1}(B_{f_i}+KD_{f_i}) + HD_{f_i} \neq 0 \quad (53)$$

The solution of this problem has been addressed by various authors using different methods to design so-called *unknown-input observers* (UIO), as for example, eigenstructure assignment (White and Speyer, 1987; Chen and Patton, 1999) or geometric methods (Massoumnia, 1986) (only for the case $D_d = 0$ and $D_f = 0$). The main difficulty with this approach is that the conditions for the solvability of the EFDP in Theorem 1 and the conditions for the existence of an observer of the form (51) which ensures (52) and (53) (see (Saberi et al., 2007, Theorems 7.6 & 7.9)), are different. Thus there exists cases when although the EFDP is solvable, however no observer of the form (51) can be used as residual generator. Moreover, it is also possible that although both the decoupling and detectability conditions (52) and (53) can be satisfied, however, due to unstable fixed modes (these are the unstable zeros of $G_d(\lambda)$), the resulting detector is

also unstable (Saberi et al., 2007). Therefore, the restriction to use full-order observers is a technical constraint, which unnecessarily limits the class of solvable EFDPs.

The *direct solution approach* of fault detection observers has been considered in several works (Wünnenberg, 1990; Hou and Müller, 1994; Patton and Hou, 1998) to cite a few of them. As was noted in (Hou and Müller, 1994), the main computational problem is the solution of so-called Luenberger equations (see for example (Hou and Müller, 1994)). The approach of Wünnenberg (1990) relies on the Kronecker canonical form of a certain matrix pencil which allows to solve these equations by inspection using the details of this form. In fact, the computations are solely done in terms of the left Kronecker structure, and therefore this method can be seen as a precursor of nullspace methods. The work of Hou and Müller (1994) improves the approach of Wünnenberg (1990) by eliminating the need to compute the Kronecker canonical form. Instead, by using a series of orthogonal transformations (stronly resembling to the structure algorithm of Silverman and Payne (1971)), the disturbance inputs are finally decoupled from the original system and, simultaneosly, the fault detectability conditions can be checked. Finally, the solution is found by designing a standard observer for a reduced order system. The limitation of this approach to consider only the case without direct feedthrough terms from the fault inputs has been removed in (Patton and Hou, 1998) and the method has been latter extended to descriptor systems in (Hou, 2000). The detector results as a full order observer of a certain reduced system. As already mentioned, this method is equivalent to the nullspace method. Note however, that none of these approaches is able to address the least order design aspect.

4.4.3. Comparison of different methods

In our comparison in Table 1 we also included the nullspace method of Frisk and Nyberg (2001), which is based on polynomial matrix manipulations. This comparison indicates that the only satisfactory computational method for solving the EFDP is the rational nullspace based approach. A more thorough comparison of the existing methods to solve the EFDP is done in (Varga, 2009b).

Table 1: Comparison of methods to solve the EFDP.

Approach	Generality	Least order synthesis	Numerical reliability
Nullspace methods			
– polynomial	yes	yes	no
– rational	**yes**	**yes**	**yes**
Parity space method	yes	yes	no
Observer-based design			
– UIO	no	no	yes
– direct solution method	yes	no	yes

4.5. Integrated algorithm for solving the ESFDIP

In what follows, we describe an integrated computational approach for the synthesis of a residual generator for the case

15

when $M_r(\lambda)$ is invertible, by using the inversion based approach presented in Section 3.4. We compute the detector $Q(\lambda)$ in a factored form

$$Q(\lambda) = Q_4(\lambda)M_r(\lambda)Q_3(\lambda)Q_2(\lambda)Q_1(\lambda), \tag{54}$$

which is similar to the product form in (28). To determine $Q(\lambda)$ we use updating techniques typical for integrated algorithms. The initial detector $Q(\lambda) = Q_1(\lambda)$, is just a minimal proper left nullspace basis of $G(\lambda)$ in (18). The state space realization of $Q_1(\lambda)$ is given in (41), while the realization of the reduced system (20) with TFM $R_f(\lambda) = \overline{G}_f(\lambda)$ is given in (42) (recall that $R_w(\lambda) = 0$).

At the second step, we choose a prefilter $Q_2(\lambda)$ such that $Q_2(\lambda)R_f(\lambda)$ is invertible. This choice is always possible if the ESFDIP is solvable, which is equivalent to $R_f(\lambda)$ being left invertible. If $p - r_d = m_f$ we use the trivial choice $Q_2(\lambda) = I_{m_f}$, while in the case when $p - r_d > m_f$ we can always choose $Q_2(\lambda) = H$, where H is a $m_f \times (p - r_d)$ constant matrix with orthonormal rows. We can even determine $Q_2(\lambda)$ as a prefilter of the form (44), where H and K are determined such that $Q_2(\lambda)Q(\lambda)$ has least possible McMillan degree and $Q_2(\lambda)R_f(\lambda)$ is invertible. For a choice as above of H, the corresponding K can be determined using minimal dynamic covers techniques (Varga, 2004c). After performing the updating operations $Q(\lambda) \leftarrow Q_2(\lambda)Q(\lambda)$ and $R_f(\lambda) \leftarrow Q_2(\lambda)R_f(\lambda)$, we obtain the state space realization of $[\, Q(\lambda)\ R_f(\lambda)\,]$ as in (47).

At the next step we choose $Q_3(\lambda) = R_f^{-1}(\lambda)$. A realization of $Q_3(\lambda)Q(\lambda)$ is given by

$$Q_3(\lambda)Q(\lambda) = \left[\begin{array}{cc|c} \widetilde{A}_l - \lambda\widetilde{E}_l & \widetilde{B}_f & \widetilde{B}_l \\ \widetilde{C}_l & \widetilde{D}_f & \widetilde{D}_l \\ \hline 0 & -I_{m_f} & 0 \end{array}\right] \tag{55}$$

After the updating step $Q(\lambda) \leftarrow Q_3(\lambda)Q(\lambda)$, we build the realization of $M_r(\lambda)Q(\lambda)$ using standard series coupling formulas of descriptor systems. Since it is expected that poles-zeros cancellations occur at this step, we apply the algorithm of (Varga, 1990) based on orthogonal similarity transformations to compute an irreducible realization of updated $Q(\lambda) \leftarrow M_r(\lambda)Q(\lambda)$. Observe that at this moment $R_f(\lambda) = M_r(\lambda)$.

At the final step we determine the diagonal TFM $Q_4(\lambda) = M(\lambda)$, such that $M(\lambda)Q(\lambda)$ and $M(\lambda)R_f(\lambda)$ are proper and stable TFMs. For this, we employ the already described coprime factorization techniques for descriptor systems (see Section 3.4). Recall that suitable algorithms to compute proper and stable coprime factorizations have been proposed in (Varga, 1998). Each diagonal element of $M(\lambda)$ can be obtained by applying these algorithms to the corresponding row of $Q(\lambda)$.

All steps in the proposed synthesis algorithm rely on numerically reliable or numerically stable computational algorithms. All structural information available at a particular step have been exploited at the next computational step, by relying on detector updating techniques. This guarantees a lower order of the final detector than the order resulting by summing the orders of factors in (54). In the case when $p - r_d = m_f$, the resulting order is the least possible one.

4.6. Other integrated algorithms

An integrated algorithm for the solution of AFDP has been proposed in (Varga, 2009a) and corresponds to the high-level procedure described in Section 3.3. The determination of the factors $Q_1(\lambda)$ and $Q_2(\lambda)$ is essentially the same as for the solution of the ESFDIP in Section 4.5, with the only difference that the choice of H must ensure that $Q_2(\lambda)R_w(\lambda)$ has full row rank. At the next step we use $Q_3(\lambda) = G_{wo}^{-1}(\lambda)$, where the outer factor $G_{wo}(\lambda)$ in the inner-outer factorization of $Q_2(\lambda)R_w(\lambda)$ is invertible (due to the above choice of $Q_2(\lambda)$) and has a state-space realization of the form (Varga, 2009a)

$$G_{wo}(\lambda) = \left[\begin{array}{c|c} \widetilde{A}_l - \lambda\widetilde{E}_l & \widetilde{B}_{wo} \\ \hline \widetilde{C}_l & \widetilde{D}_{wo} \end{array}\right]$$

Thus, the determination of an explicit state-space realization of $Q_3(\lambda)Q(\lambda)$, with $Q(\lambda) = Q_2(\lambda)Q_1(\lambda)$, is entirely similar to the updating in (55), with $\widetilde{B}_f$ and $\widetilde{D}_f$ replaced by $\widetilde{B}_{wo}$ and $\widetilde{D}_{wo}$, respectively. The details of the final optimization step are described in (Varga, 2009a; Glover and Varga, 2011).

The integrated algorithms for solving the ASFDIP using the $H_{2/\infty}$-norm minimization approach employ the same three steps as above, with the only difference that the choice of H must ensure $Q_2(\lambda)[\, R_f(\lambda)\ R_w(\lambda)\,]$ has full row rank. The details on the final optimization steps are described in (Varga, 2010, 2011a) as well as in the recent overview (Varga, 2012a).

Integrated algorithms for solving the EWFDIP and AWFDIP automatically result when employing the techniques described for EFDP in Section 4.4 and AFDP above, respectively.

5. Conclusions

A new generation of satisfactory numerical algorithms based on detector updating techniques has been developed in the last decade to solve the main classes of fault detection filter synthesis problems. In all these algorithms, the nullspace method has a central role in simplifying the synthesis problems and also in addressing the least order synthesis aspect. The main numerical ingredients of the new methods are numerically stable or numerically reliable algorithms to compute condensed forms of matrices or matrix pairs (Schur, generalized Schur, Kronecker-like forms), factorization methods (proper/stable coprime and inner-outer factorizations), as well as special algorithms developed for the needs to compute least order detectors (minimum dynamic cover, least order solution of linear rational equations). An important aspect is that all basic algorithms use state space system representations in descriptor form, which guarantees full generality and simultaneously leads to better conditioned computational problems by avoiding unnecessary conversions to standard state space representations.

The usefulness of the new synthesis methods would be highly questionable without the availability of robust numerical software to perform all the basic computations. Fortunately, a comprehensive collection of such tools is available in the MATLAB/SLICOT-based Descriptor Systems Toolbox[1] developed by the author in the last decade (Varga, 2000; Huffel et al.,

---

[1]Licensed by SYNOPTIO GmbH http://synmath.synoptio.de/en

2004). Additionally, all described synthesis algorithms are implemented in the Fault Detection Toolbox (Varga, 2006, 2011b) (a proprietary software of DLR, not licensed).

Recent extensions of the synthesis techniques targeted new classes of systems, as *periodic systems* and *linear parameter-varying* (LPV) systems. For periodic systems, the first nullspace based synthesis method has been proposed in (Varga, 2005), and subsequently efforts have been invested to develop the main computational ingredients, as algorithms for minimum periodic dynamic covers (Varga, 2007a) and periodic coprime factorization (Varga, 2009d), to support the periodic counterparts of the EFDP and EWFDIP. For both these fault detection problems, integrated algorithms have been recently proposed (Varga, 2012b). For LPV-systems based synthesis, the first methods have been developed using geometric techniques for systems with affine dependence of scheduling parameters (Bokor and Balas, 2004). A more general LPV-synthesis framework based on symbolic LPV-nullspace computational techniques has been proposed to solve the robust fault detection problem (Varga, 2011d). This approach has been recently turned into a numerical computations based algorithm for the synthesis of fault detection filters in LPV-forms (Varga, 2011e). The proposed computational approach relies on the numerically reliable synthesis techniques for solving the EFDP described in Section 4.2. An interesting feature of the new techniques is that they can address robustness aspects with respect to both measurable uncertain parameters (thus suitable for gain scheduling) as well as unmeasurable uncertain parameters.

Note: This paper is intended in the first line to be a survey of recent developments in the field of numerical methods for the synthesis of residual generators. Therefore, most of the cited literature is primarily relevant to the development of appropriate computational techniques. For a full coverage of the addressed synthesis topics the reader is invited to consult the references listed in the already mentioned textbooks (Gertler, 1998; Chen and Patton, 1999; Blanke et al., 2003; Saberi et al., 2007).

Appendix A. Polynomial and rational nullspace bases

Since polynomial bases represent an important tool in defining the corresponding concepts for the more general *rational bases*, we will recall shortly some of the main results of Forney (1975). Let $G(\lambda)$ be a $p \times m$ rational matrix of rank r and assume that $N_l(\lambda)$ is a $(p - r) \times p$ *polynomial basis* of the left nullspace of $G(\lambda)$, thus

$$N_l(\lambda)G(\lambda) = 0$$

Let denote by n_i, the i-th *index* (or *degree*), representing the greatest degree of the i-th row of $N_l(\lambda)$. Then, the *order* of $N_l(\lambda)$ is defined as $\underline{n} := \sum_{i=1}^{p-r} n_i$, (i.e., the sum of row degrees). A *minimal basis* is one which has least order among all polynomial bases. The indices of a minimal basis are called *minimal indices*.

Some properties of a minimal bases are summarized below (Forney, 1975; Kailath, 1980):

Proposition 1. *Let $N_l(\lambda)$ be a minimal polynomial basis with row indices n_i, $i = 1, \ldots, p - r$. Then the following holds:*

1. *The row indices are* unique *up to permutations (i.e., if $\widetilde{N}_l(\lambda)$ is another minimal basis, then $N_l(\lambda)$ and $\widetilde{N}_l(\lambda)$ have the same minimal indices).*
2. *The minimal indices are the left Kronecker indices of $G(\lambda)$.*
3. *$N_l(\lambda)$ is irreducible, having full row rank for all $\lambda \in \mathbb{C}$ (i.e., $N_l(\lambda)$ has no finite or infinite zeros).*
4. *$N_l(\lambda)$ is row reduced (i.e., the leading row coefficient matrix formed from the coefficients of the highest row degrees has full row rank.)*

If $M(\lambda)$ is a non-singular polynomial or rational matrix, then $\widetilde{N}_l(\lambda) := M(\lambda)N_l(\lambda)$ is also a nullspace basis. Frequently such a matrix $M(\lambda)$ originates from a *left coprime factorizations* of an original basis $N_l(\lambda)$ in the form

$$N_l(\lambda) = M(\lambda)^{-1}\widetilde{N}_l(\lambda), \tag{A.1}$$

where the factors $M(\lambda)$ and $\widetilde{N}_l(\lambda)$ can be chosen to satisfy special requirements (e.g., have only poles in a certain "good" region of the complex plane).

Minimal polynomial bases allow to easily build *proper minimal rational bases*. These are proper rational bases having the least McMillan degree of $\underline{n}$. A proper rational basis with arbitrary poles can be simply constructed by forming $\widetilde{N}_l(\lambda) := M(\lambda)N_l(\lambda)$ with

$$M(\lambda) = \text{diag}\left(\frac{1}{d_1(\lambda)}, \cdots, \frac{1}{d_{p-r}(\lambda)}\right), \tag{A.2}$$

where $d_i(\lambda)$ is a polynomial of degree n_i with arbitrary roots. The resulting basis $\widetilde{N}_l(\lambda)$ has the additional property that the order of any minimal state space realization of $\widetilde{N}_l(\lambda)$ is equal to the sum of orders of the minimal state space realizations of the rows of $\widetilde{N}_l(\lambda)$. Furthermore, $D_l := \lim_{\lambda \to \infty} \widetilde{N}_l(\lambda)$ has full row rank. Such a basis is called *simple minimal proper rational basis* in (Vardulakis and Karcanias, 1984) and is the natural counterpart of the minimal polynomial basis introduced by Forney (1975).

Appendix B. Computation of minimal proper rational nullspace bases

We describe a method for the computation of a minimal proper rational left nullspace basis $N_l(\lambda)$ of a $p \times m$ rational matrix $G(\lambda)$ of rank r using a pencil reduction method based on orthogonal similarity transformations. Consider a state space representation of $G(\lambda)$ in the form

$$G(\lambda) := \left[\begin{array}{c|c} A - \lambda E & B \\ \hline C & D \end{array}\right] \tag{B.1}$$

which satisfies

$$G(\lambda) = C(\lambda E - A)^{-1}B + D$$

The computational method, proposed in (Varga, 2003), exploits the simple fact that $N_l(\lambda)$ is a left nullspace basis of $G(\lambda)$ if and

only if $[\,M_l(\lambda)\ N_l(\lambda)\,]$ is a left nullspace basis of the associated system matrix

$$S(\lambda) = \begin{bmatrix} A - \lambda E & B \\ C & D \end{bmatrix}.$$

Thus, to compute $N_l(\lambda)$ we can determine equivalently a left nullspace basis $Y_l(\lambda)$ of $S(\lambda)$ and then $N_l(\lambda)$ simply results as

$$N_l(\lambda) = Y_l(\lambda) \begin{bmatrix} 0 \\ I_p \end{bmatrix}.$$

The importance of this fact is that $N_l(\lambda)$ can be computed by employing numerically stable linear pencil reduction algorithms based on orthogonal transformations.

Let U and V be orthogonal matrices determined, for instance by using the algorithms of Beelen (1987) or Varga (1996), such that the transformed pencil $\widetilde{S}(\lambda) := US(\lambda)V$ is in the Kronecker-like form

$$\widetilde{S}(\lambda) = \left[\begin{array}{c|c} A_r - \lambda E_r & A_{r,l} - \lambda E_{r,l} \\ \hline 0 & A_l - \lambda E_l \\ \hline 0 & C_l \end{array} \right], \tag{B.2}$$

where the descriptor pair $(A_l - \lambda E_l, C_l)$ is observable, E_l is nonsingular, and $A_r - \lambda E_r$ has full row rank excepting possibly a finite set of values of λ (i.e, the invariant zeros of $S(\lambda)$). It follows that we can choose a proper rational left nullspace basis $\widetilde{Y}_l(\lambda)$ of $\widetilde{S}(\lambda)$ in the form

$$\widetilde{Y}_l(\lambda) = \left[\begin{array}{c|c|c} 0 & C_l(\lambda E_l - A_l)^{-1} & I \end{array} \right]. \tag{B.3}$$

Then, the left nullspace of $G(\lambda)$ is

$$N_l(\lambda) = \widetilde{Y}_l(\lambda) U \begin{bmatrix} 0 \\ I_p \end{bmatrix}$$

and if we partition

$$U \begin{bmatrix} 0 \\ I_p \end{bmatrix} = \begin{bmatrix} B_{r,l} \\ B_l \\ D_l \end{bmatrix}$$

in accordance with the column partition of $\widetilde{Y}_l(\lambda)$, we obtain

$$N_l(\lambda) = C_l(\lambda E_l - A_l)^{-1} B_l + D_l := \left[\begin{array}{c|c} A_l - \lambda E_l & B_l \\ \hline C_l & D_l \end{array} \right] \tag{B.4}$$

which is a descriptor system representation for $N_l(\lambda)$. Note that, to obtain this nullspace basis, we performed exclusively orthogonal transformations on the system matrices. We can prove that all computed matrices are exact for a slightly perturbed original system. It follows that the algorithm to compute the nullspace basis is *numerically backward stable*.

The full column rank subpencil $\begin{bmatrix} A_l - \lambda E_l \\ C_l \end{bmatrix}$ defines the left Kronecker structure of $G(\lambda)$. It is possible to obtain this subpencil in an observability staircase form

$$\left[\begin{array}{c} A_l - \lambda E_l \\ \hline C_l \end{array} \right] = \left[\begin{array}{c|c|c|c} A_{\ell,\ell+1} & A_{\ell,\ell} - \lambda E_{\ell,\ell} & \cdots & A_{\ell,1} - \lambda E_{\ell,1} \\ \hline & A_{\ell-1,\ell} & \ddots & \vdots \\ \hline & & \ddots & A_{1,1} - \lambda E_{1,1} \\ \hline & & & A_{0,1} \end{array} \right], \tag{B.5}$$

where $A_{i,i+1} \in \mathbb{R}^{\mu_i \times \mu_{i+1}}$, with $\mu_{\ell+1} = 0$, are full column rank upper triangular matrices, for $i = 0, \ldots, \ell$. Note that this form is automatically obtained by using the pencil reduction algorithms described in (Beelen, 1987) or (Varga, 1996). The left (or column) Kronecker indices result as follows: there are $\mu_{i-1} - \mu_i$ Kronecker blocks of size $i \times (i-1)$, for $i = 1, \ldots, \ell + 1$. The row dimension of $N_l(\lambda)$ (i.e., the number of linearly independent basis vectors) is given by the total number of Kronecker indices, thus $\sum_{i=1}^{\ell+1}(\mu_{i-1} - \mu_i) = \mu_0$. Applying standard linear algebra results, it follows that $\mu_0 := p - r$.

We give now some properties of the computed rational basis, see (Varga, 2011c) for proofs.

Proposition 2. *If the realization (B.1) of $G(\lambda)$ is minimal, then the rational matrix $N_l(\lambda)$ defined in (B.4) is a minimal proper rational basis of the left nullspace of $G(\lambda)$.*

This result shows that the computed rational basis above has actually the least possible McMillan degree. However, in general, the computed minimal proper basis is not simple. Additionally the following important result holds:

Proposition 3. *If the realization (B.1) of $G(\lambda)$ is minimal, then the realization of $N_l(\lambda)$ defined in (B.4) is maximally controllable.*

This means that for any output injection matrix K, the pair $(A_l + KC_l - \lambda E_l, B_l + KD_l)$ remains controllable. Notice also that due to the particular form of C_l, the resulting $A_l + KC_l$ has only the last block column changed. It follows that $A_l + KC_l$ has the same staircase form structure as A_l.

Appendix C. Solution of linear rational equations

In solving the ESFDIP, the TFM of the detector $Q(\lambda)$ fulfills a linear rational equation having the form

$$Q(\lambda)G(\lambda) = F(\lambda) \tag{C.1}$$

(see for example equations (26) or (27)). Assume that $G(\lambda)$ and $F(\lambda)$ are $p \times m$ and $q \times m$ rational matrices, respectively, and thus the resulting $Q(\lambda)$ is a $q \times p$ rational matrix. The solvability condition for this equation is (see also Corollary 2)

$$\operatorname{rank} G(\lambda) = \operatorname{rank} \begin{bmatrix} G(\lambda) \\ F(\lambda) \end{bmatrix} \tag{C.2}$$

Assume the following descriptor realization of the compound TFM in the right hand side

$$\begin{bmatrix} G(\lambda) \\ F(\lambda) \end{bmatrix} = \left[\begin{array}{c|c} A - \lambda E & B \\ \hline C_G & D_G \\ C_F & D_F \end{array} \right]$$

and let $S_G(\lambda)$ and $S_F(\lambda)$ be the associated system matrix pencils defined as

$$S_G(\lambda) = \begin{bmatrix} A - \lambda E & B \\ C_G & D_G \end{bmatrix}$$

and

$$S_F(\lambda) = \begin{bmatrix} A - \lambda E & B \\ C_F & D_F \end{bmatrix}$$

It is easy to see, that $Q(\lambda)$ is a solution of (C.1) if and only if

$$Y(\lambda) = \begin{bmatrix} Y_{11}(\lambda) & Y_{12}(\lambda) \\ Y_{21}(\lambda) & Q(\lambda) \end{bmatrix}$$

satisfies

$$Y(\lambda)S_G(\lambda) = S_F(\lambda) \qquad (C.3)$$

It follows that instead of solving the linear equation (C.1) with rational coefficients, we can solve the equation (C.3) with polynomial coefficients and take

$$Q(\lambda) = \begin{bmatrix} 0 & I_q \end{bmatrix} Y(\lambda) \begin{bmatrix} 0 \\ I_p \end{bmatrix}$$

In fact, since we are interested in the second block row $Y_2(\lambda)$ of $Y(\lambda)$, we need only to solve

$$Y_2(\lambda) \begin{bmatrix} A - \lambda E & B \\ C_G & D_G \end{bmatrix} = \begin{bmatrix} C_F & D_F \end{bmatrix} \qquad (C.4)$$

and compute $Q(\lambda)$ as

$$Q(\lambda) = Y_2(\lambda) \begin{bmatrix} 0 \\ I_p \end{bmatrix}$$

The condition (C.2) for the existence of a solution becomes for the system (C.4)

$$\mathrm{rank} \begin{bmatrix} A - \lambda E & B \\ C_G & D_G \end{bmatrix} = \mathrm{rank} \begin{bmatrix} A - \lambda E & B \\ C_G & D_G \\ C_F & D_F \end{bmatrix} \qquad (C.5)$$

To solve (C.4), we isolate a full rank part of $S_G(\lambda)$ by reducing it to a particular Kronecker-like form. Let U and V be orthogonal matrices to reduce $S_G(\lambda)$ to the Kronecker-like form $\overline{S}_G(\lambda) := US_G(\lambda)V$ such that

$$\overline{S}_G(\lambda) = \begin{bmatrix} A_r - \lambda E_r & A_{r,reg} - \lambda E_{r,reg} & A_{r,l} - \lambda E_{r,l} \\ 0 & A_{reg} - \lambda E_{reg} & A_{reg,l} - \lambda E_{reg,l} \\ 0 & 0 & A_l - \lambda E_l \\ 0 & 0 & C_l \end{bmatrix}, \quad (C.6)$$

where $A_{reg} - \lambda E_{reg}$ is a regular subpencil which contains the zeros of $G(\lambda)$, the pair $(A_l - \lambda E_l, C_l)$ is observable with E_l nonsingular and the subpencil $A_r - \lambda E_r$ has full row rank. The above reduction can be computed by employing numerically stable algorithms as those proposed in (Beelen, 1987; Varga, 1996).

If $\overline{Y}_2(\lambda)$ is a solution of the reduced equation

$$\overline{Y}_2(\lambda)\overline{S}_G(\lambda) = \begin{bmatrix} C_F & D_F \end{bmatrix} V, \qquad (C.7)$$

then $Y_2(\lambda) = \overline{Y}_2(\lambda)U$ and thus

$$Q(\lambda) = \overline{Y}_2(\lambda)U \begin{bmatrix} O \\ I_p \end{bmatrix}$$

is a solution of the equation $Q(\lambda)G(\lambda) = F(\lambda)$. Partition

$$\begin{bmatrix} -C_F & -D_F \end{bmatrix} V = \begin{bmatrix} \overline{C}_1 & \overline{C}_2 & \overline{C}_3 \end{bmatrix}$$

in accordance with the column structure of $\overline{S}_G(\lambda)$. Since $A_r - \lambda E_r$ has full row rank, it follows that (C.5) is equivalent to

$$\overline{C}_1 = 0 \qquad (C.8)$$

Thus, if a solution exists, $\overline{Y}_2(\lambda)$ must have the form

$$\overline{Y}_2(\lambda) = \begin{bmatrix} 0 & \overline{Y}_{22}(\lambda) & \overline{Y}_{23}(\lambda) & \overline{Y}_{24}(\lambda) \end{bmatrix},$$

where the column partitioning of $\overline{Y}_2(\lambda)$ corresponds to the row partitioning of $\overline{S}_G(\lambda)$. Choosing $\overline{Y}_{24}(\lambda) = 0$, we obtain

$$\begin{bmatrix} \overline{Y}_{22}(\lambda) \overline{Y}_{23}(\lambda) \end{bmatrix} = \begin{bmatrix} \overline{C}_2 \overline{C}_3 \end{bmatrix} \begin{bmatrix} \lambda E_{reg} - A_{reg} & \lambda E_{reg,l} - A_{reg,l} \\ 0 & \lambda E_l - A_l \end{bmatrix}^{-1}.$$

Let partition $U \begin{bmatrix} 0 \\ I_p \end{bmatrix}$ in accordance with the row structure of $\overline{S}_G(\lambda)$ as

$$U \begin{bmatrix} 0 \\ I_p \end{bmatrix} = \begin{bmatrix} B_r \\ B_{reg} \\ B_l \\ D_l \end{bmatrix} \qquad (C.9)$$

and denote

$$\overline{A} - \lambda \overline{E} = \begin{bmatrix} A_{reg} - \lambda E_{reg} & A_{reg,l} - \lambda E_{reg,l} \\ 0 & A_l - \lambda E_l \end{bmatrix},$$

$$\overline{B} = \begin{bmatrix} B_{reg} \\ B_l \end{bmatrix}, \quad \overline{C} = \begin{bmatrix} \overline{C}_2 & \overline{C}_3 \end{bmatrix}$$

Then, a *particular* solution $Q_0(\lambda)$ of the equation $G(\lambda) = Q(\lambda)F(\lambda)$ can be expressed in form of a descriptor realization

$$Q_0(\lambda) = \overline{C}(\lambda\overline{E} - \overline{A})^{-1}\overline{B} := \left[\begin{array}{c|c} \overline{A} - \lambda\overline{E} & \overline{B} \\ \hline \overline{C} & 0 \end{array} \right]$$

When solving the ESFDIP, $G(\lambda)$ and $F(\lambda)$ may share common zeros and the associated zero structure (typically unstable and infinite zeros), in order to obtain a stable solution of the linear rational equation (C.1) by enforcing the cancellation of these zeros (Kailath, 1980). Since these zeros are among the generalized eigenvalues of the pencil $A_{reg} - \lambda E_{reg}$, they are the unobservable eigenvalues of the descriptor pair $(A_{reg} - \lambda E_{reg}, \overline{C}_2)$ and can be removed using orthogonal similarity transformations to reduce this pair to an observability staircase form. Suitable algorithms for this purpose have been proposed in (Varga, 1990). An observable realization of $Q_0(\lambda)$ can be determined after applying the resulting orthogonal transformations to the rest of submatrices of $\overline{A} - \lambda\overline{E}$, $\overline{B}$, and $\overline{C}$ by removing their unobservable parts.

To compute a minimal descriptor realization of $Q_0(\lambda)$ we employed exclusively orthogonal similarity transformations. Therefore, this computation is numerically stable, because we can easily show that the computed system matrices in the presence of roundoff errors are exact for an original problem with slightly perturbed data.

19

The solution $Q_0(\lambda)$ is unique if there is no left structure in the Kronecker-like form of $S_G(\lambda)$ (i.e., A_l, E_l and C_l are empty matrices). In the general case, when the left structure is present, the solution is not unique and can be expressed in the form

$$Q(\lambda) = Q_0(\lambda) + X(\lambda)N_l(\lambda)$$

where $N_l(\lambda)$ is a left nullspace basis of the form (B.4) and $X(\lambda)$ is an arbitrary rational matrix. A suitable choice of $X(\lambda)$ allows to determine a least order solution $Q(\lambda)$ (Varga, 2004a).

References

Beard, R. V., 1971. Failure accomodation in linear systems through self-reorganization. Ph. D. Thesis, Dept. Aeronautics and Astronautics, Mass. Inst. Technology, Cambridge, MA.

Beelen, T., 1987. New algorithms for computing the Kronecker structure of a pencil with applications to systems and control theory. Ph. D. Thesis, Eindhoven University of Technology.

Blanke, M., Kinnaert, M., Lunze, J., Staroswiecki, M., 2003. Diagnosis and Fault-Tolerant Control. Springer-Verlag, Berlin.

Bokor, J., Balas, G., 2004. Detection filter design for LPV systems – a geometric approach. Automatica 40, 511–518.

Chen, J., Patton, R. J., 1999. Robust Model-Based Fault Diagnosis for Dynamic Systems. Kluwer Academic Publishers, London.

Chow, E. Y., Willsky, A. S., 1984. Analytical redundancy and the design of robust failure detection systems. IEEE Trans. Automat. Control 29, 603 – 614.

Ding, S. X., Frank, P. M., Ding, E. L., Jeinsch, T., 2000. A unified approach to the optimization of fault detection systems. Int. J. of Adaptive Contr. and Signal Processing 14, 725–745.

Ding, X., Frank, P. M., 1991. Frequency domain approach and threshold selector for robust model-based fault detection and isolation. In: Proc. of IFAC Symposium SAFEPROCESS'1991, Baden-Baden, Germany.

Forney, G. D., 1975. Minimal bases of rational vector spaces with applications to multivariable linear systems. SIAM J. Control 13, 493–520.

Francis, B. A., 1987. A Course in H^∞ Theory. Springer-Verlag, New York.

Frank, P. M., Ding, X., 1994. Frequency domain approach to optimally robust residual generation and evaluation for model-based fault diagnosis. Automatica 30, 789–804.

Frisk, E., Nyberg, M., 2001. A minimal polynomial basis solution to residual generation for fault diagnosis in linear systems. Automatica 37, 1417–1424.

Gertler, J., 1998. Fault Detection and Diagnosis in Engineering Systems. Marcel Dekker, New York.

Gertler, J., 2000. Designing dynamic consistency relation for fault detection and isolation. Int. J. Control 73, 720–732.

Gertler, J. J., Monajemy, R., 1995. Generating directional residuals with dynamic parity relations. Automatica 31, 627–635.

Glover, K., Varga, A., 2011. On solving non-standard $\mathcal{H}_-/\mathcal{H}_{2/\infty}$ fault detection problems. In: Proc. of CDC'2011, Orlando, FL. pp. 891–896.

Hou, M., 2000. Fault detection and isolation for descriptor systems. In: Patton, R. J., Frank, P. M., Clark, R. N. (Eds.), Issues of Fault Dyagnosis for Dynamic Systems. Springer Verlag, London, pp. 115–144.

Hou, M., Müller, P. C., 1994. Fault detection and isolation observers. Int. J. Control 60, 827–846.

Huffel, S. V., Sima, V., Varga, A., Hammarling, S., Delebecque, F., 2004. High-performance numerical software for control. Control Systems Magazine 24, 60–76.

Jaimoukha, I. M., Li, Z., Papakos, V., 2006. A matrix factorization solution to the $\mathcal{H}_-/\mathcal{H}_\infty$ fault detection problem. Automatica 42, 1907–1912.

Jones, H. L., 1971. Failure detection in linear systems. Ph. D. Thesis, Dept. Aeronautics and Astronautics, Mass. Inst. Technology, Cambridge, MA.

Kailath, T., 1980. Linear Systems. Prentice Hall, Englewood Cliffs, N.J.

Kimura, G., 1977. Geometric structure of observers for linear feedback control laws. IEEE Trans. Automat. Control 22, 846–855.

Liu, N., Zhou, K., 2007. Optimal solutions to multiobjective robust fault detection problems. In: Proc. of CDC'07, New Orleans, USA. pp. 981–988.

Liu, N., Zhou, K., 2008. Optimal robust fault detection for linear discrete time systems. Journal of Control Science and Engineering vol. 2008, Article ID 829459, 16 pages, doi:10.1155/2008/829459.

Lou, X.-C., Willsky, A. S., Verghese, G. C., 1986. Optimally robust redundancy relations for failure detection in uncertain systems. Automatica 22, 333–344.

Magni, J.-F., Mouyon, P., 1994. On residual generation by observer and parity space approaches. IEEE Trans. Automat. Control 39, 441 – 447.

Maquin, D., Gaddouna, B., Ragot, J., 1993. Generation of parity equations for singular systems. application to diagnosis. In: Proc. of Int. Conference on Systems, Man and Cybernetics, Le Touquet, France. Vol. 3. pp. 400–405.

Massoumnia, M. A., 1986. A geometric approach to failure detection and identification in linear systems. Ph. D. Thesis, Dept. Aeronautics and Astronautics, Mass. Inst. Technology, Cambridge, MA.

Moler, C. B., Van Loan, C. F., 1978. Nineteen dubious ways to compute the exponential of a matrix. SIAM Review 20, 801–836.

Nyberg, M., 2002. Criterions for detectability and strong detectability of faults in linear systems. Int. J. Control 75, 490–501.

Nyberg, M., Frisk, E., 2006. Residual generation for fault diagnosis of systems described by linear differential-algebraic equations. IEEE Trans. Automat. Control 51, 1995–2000.

Oară, C., 2005. Constructive solutions to spectral and inner-outer factorizations with respect to the disk. Automatica 41, 1855 – 1866.

Oară, C., Varga, A., 2000. Computation of general inner-outer and spectral factorizations. IEEE Trans. Automat. Control 45, 2307–2325.

Paige, C. C., 1981. Properties of numerical algorithms related to computing controllability. IEEE Trans. Automat. Control 26, 130 – 138.

Patton, R. J., Hou, M., 1998. Design of fault detection and isolation observers: a matrix pencil approach. Automatica 34 (9), 1135–1140.

Saberi, A., Stoorvogel, A. A., Sannuti, P., 2007. Filtering Theory – With Applications to Fault Detection, Isolation, and Estimation. Systems & Control: Foundations & Applications. Birkhäuser, Basel.

Saberi, A., Stoorvogel, A. A., Sannuti, P., Niemann, H., 2000. Fundamental problems in fault detection and identification. Int. J. Robust Nonlinear Control 10, 1209–1236.

Silverman, L. M., Payne, H. J., 1971. Input–output structure of linear systems with application to the decoupling problem. SIAM Journal on Control 9 (2), 199–233.

Van Dooren, P., 1981. The generalized eigenstructure problem in linear systems theory. IEEE Trans. Automat. Control 26, 111–129.

Vardulakis, A. I. G., Karcanias, N., 1984. Proper and stable, minimal MacMillan degrees bases of rational vector spaces. IEEE Trans. Automat. Control 29, 1118 – 1120.

Varga, A., 1990. Computation of irreducible generalized state-space realizations. Kybernetika 26, 89–106.

Varga, A., 1996. Computation of Kronecker-like forms of a system pencil: Applications, algorithms and software. In: Proc. CACSD'96 Symposium, Dearborn, MI. pp. 77–82.

Varga, A., 1998. Computation of coprime factorizations of rational matrices. Lin. Alg. & Appl. 271, 83–115.

Varga, A., 2000. A Descriptor Systems toolbox for Matlab. In: Proc. CACSD'2000 Symposium, Anchorage, Alaska. pp. 150–155.

Varga, A., 2003. On computing least order fault detectors using rational nullspace bases. In: Proc. of IFAC Symp. SAFEPROCESS'2003, Washington D.C.

Varga, A., 2004a. Computation of least order solutions of linear rational equations. In: Proc. of MTNS'04, Leuven, Belgium.

Varga, A., 2004b. New computational approach for the design of fault detection and isolation filters. In: Voicu, M. (Ed.), Advances in Automatic Control. Vol. 754 of The Kluwer International Series in Engineering and Computer Science. Kluwer Academic Publishers, pp. 367–381.

Varga, A., 2004c. Reliable algorithms for computing minimal dynamic covers for descriptor systems. In: Proc. of MTNS'04, Leuven, Belgium.

Varga, A., 2005. Numerically reliable methods for optimal design of fault detection filters. In: Proc. of CDC'05, Seville, Spain. pp. 2391–2396.

Varga, A., 2006. A fault detection toolbox for Matlab. In: Proc. of CACSD'06, Munich, Germany. pp. 3013–3018.

Varga, A., 2007a. Computation of minimal order dynamic covers for periodic systems. In: Proc. of ECC'07, Kos, Greece. pp. 1088–1093.

Varga, A., 2007b. On designing least order residual generators for fault detec-

tion and isolation. In: Proc. 16th Internat. Conf. on Control Systems and Computer Science, Bucharest, Romania. pp. 323–330.

Varga, A., 2008. On computing nullspace bases – a fault detection perspective. In: Proc. IFAC 2008 World Congress, Seoul, Korea. pp. 6295–6300.

Varga, A., 2009a. General computational approach for optimal fault detection. In: Proc. SAFEPROCESS'2009, Barcelona, Spain. pp. 107–112.

Varga, A., 2009b. The nullspace method – a unifying paradigm to fault detection. In: Proc. CDC'2009, Shanghai, China. pp. 6964–6969.

Varga, A., 2009c. On computing achievable fault signatures. In: Proc. SAFE-PROCESS'2009, Barcelona, Spain. pp. 935–940.

Varga, A., 2009d. On computing periodic coprime factorizations. In: Proc. ECC'2009, Budapest, Hungary. pp. 1359–1364.

Varga, A., 2010. Integrated algorithm for solving $\mathcal{H}_2$-optimal fault detection and isolation problems. In: Proc. of SYSTOL'2010, Nice, France. pp. 353–358.

Varga, A., 2011a. Integrated computational algorithm for solving $\mathcal{H}_\infty$-optimal FDI problems. In: Proc. IFAC 2011 World Congress, Milano, Italy. pp. 10187–10192.

Varga, A., 2011b. Linear FDI-Techniques and Software Tools. Fault Detection Toolbox V1.0 – Technical Documentation, German Aerospace Center (DLR), Institute of Robotics and Mechatronics.

Varga, A., 2011c. On computing minimal proper nullspace bases with applications in fault detection. In: Van Dooren, P., Bhattacharyya, S. P., Chan, R. H., Olshevsky, V., Routray, A. (Eds.), Numerical Linear Algebra in Signals, Systems and Control. Vol. 80 of Lecture Notes in Electrical Engineering. Springer Verlag, Berlin, pp. 433–465.

Varga, A., 2011d. On parametric solution of fault detection problems. In: Proc. IFAC 2011 World Congress, Milano, Italy. pp. 6697–6702.

Varga, A., 2011e. Synthesis of robust gain scheduling based fault detection filters for a class of parameter uncertain nonlinear systems. In: Proc. of 19th Mediterranean Conference on Control and Automation, Corfu, Greece. pp. 467–472.

Varga, A., 2012a. Descriptor system techniques in solving $\mathcal{H}_{2/\infty}$-optimal fault detection and isolation problems. In: Biegler, L. T., Campbell, S. L., Mehrmann, V. (Eds.), Control and Optimization with Differential-Algebraic Constraints. Vol. 23 of Advances in Design and Control. SIAM, pp. 105–125.

Varga, A., 2012b. A numerically reliable approach for the synthesis of periodic FDI filters. In: Proc. of SAFEPROCESS'12, Mexico City, Mexico. pp. 498–503.

White, J. E., Speyer, J. L., 1987. Detection filter design: Spectral theory and algorithms. IEEE Trans. Automat. Control 32, 593–603.

Wünnenberg, J., 1990. Observer-based fault detection in dynamic systems. Ph.D. Thesis, University of Duisburg, Germany.

Zhang, P., Ding, S. X., 2008. An integrated trade-off design of observer based fault detection systems. Automatica 44, 1886–1894.

Zhou, K., Doyle, J. C., Glover, K., 1996. Robust and Optimal Control. Prentice Hall.

Andreas Varga received the diploma in control engineering and the Ph.D. degree in electrical engineering from the University Polytechnica Bucharest, Romania, in 1974 and 1981, respectively. From 1974 to 1993, he has held various research positions at the Institute of Informatics Bucharest and at the Ruhr-University of Bochum, Germany. From 1990 to 1992, he worked at the Ruhr-University of Bochum as Visiting Research Fellow in the framework of a fellowship award of the Alexander von Humboldt Foundation. Since 1993, he has been at the German Aerospace Center (DLR), Oberpfaffenhofen, where he is currently a Senior Scientist. His main research interests include the numerical methods for linear systems analysis and design, descriptor systems, periodic systems, and robust numerical software for computer aided control system design (CACSD). He coauthored three books, coedited two books, published over 60 papers in refereed journals or book chapters, and has over 145 conference publications. Dr. Varga is Fellow of IEEE and served as Associate Editor for the IEEE Transactions on Automatic Control between 1997-1999. He was the Program Chairman of the 1999 Symposium on CACSD (Hawaii), the General Chair for the 2000 Symposium on CACSD (Anchorage, Alaska), the General Chair of the 2006 IEEE Conference on Control Applications, 2006 IEEE Symposium on CACSD and 2006 IEEE International Symposium on Intelligent Control (Munich, Germany). For 2000-2004 he was the Chairman of the Technical Committee on CACSD within the IEEE Control Systems Society and for 2002-2003 he was a nominated member of the Board of Governors of the IEEE Control Systems Society.