Unitary System - II: Application to $H_\infty$/H_\infty Optimization of Strictly-Proper Systems

Zhongyu Zhao*, Wen-Fang Xie**, Henry Hong***, and Youmin Zhang****

Department of Mechanical & Industrial Engineering, Concordia University, Montreal, QC, Canada, H3G 1M8
*(e-mail: zhao_zho@encs.concordia.ca)
**(Tel: 514-848-2424, ext. 4193; fax: 514-848-3175; e-mail: wfxie@encs.concordia.ca)
*** (e-mail: henhong@vax2.concordia.ca)
**** (e-mail: ymzhang@encs.concordia.ca)

Abstract: In Part II of the paper, the unitary system method is applied to the $H_\infty$/H_\infty optimization of strictly-proper systems by transforming the fault-to-residual transfer matrix to a unitary system. In the area of model-based fault detection, the $H_\infty$/H_\infty optimization is a problem of seeking the balance between the robustness to disturbances and the sensitivity to faults. Although the theoretical optimum cannot be obtained due to the involvement of the inverse of strictly-proper systems, the proposed solution can approximate to the theoretical optimum at any required accuracy because of the special property of a unitary system.

Keywords: $H_\infty$, unitary system, strictly-proper system, fault detection observer, singular values

1. INTRODUCTION

In the studies of model-based fault detection (Chen and Patton 1999, Frank 1990, Isermann 1997a and 1997b, Saberi et al. 2000, Venkatsubramanian et al. 2003, Frank and Ding 1997), it is important to differentiate faults from the signals corrupted by disturbances. The ideal approach is to construct disturbance-decoupled residuals that only respond to the occurring faults as shown in Chen et al. (1996), Duan and Patton (2001), Patton and Chen (1992, 2000), and Xiong and Saif (2000). When such residuals are not available, an alternative approach is to construct residuals based on optimization results so that the residuals are robust or insensitive to disturbances as in Song and Collins (2000), Collins and Song (2000), Douglas and Speyer (1996, 1999), Khosrowjerdi et al. (2005). However, the increase of the robustness to disturbances usually also reduces the sensitivity to faults.

To enhance the accuracy of fault detection, it is desirable to construct the residuals with both the robustness to disturbances and the sensitivity to faults, which leads to the studies of $H_\infty$/H_\infty optimization (Hou and Patton 1996) minimizing the difference between the robustness to disturbances ($H_\infty$) and the sensitivity to faults ($H_\infty$) of the residuals. Different approaches had been proposed for the optimization problem. In Rank and Niemann (1998, 1999), Henry and Zoghadri (2005, 2006), the $H_\infty$/H_\infty optimization is formulated into two $H_\infty$ minimization problems. In Chen et al. (1996), Rambeaux et al. (2000), Chung and Speyer (1997), Liu and Zhou (2008), and Wang et al. (2007), the $H_\infty$/H_\infty optimization is solved with numerical methods such as Genetic Algorithm, Linear Matrix Inequality, or Iterative Linear Matrix Inequality. The exact solutions to the optimization are given only in Jaimoukha et al. (2006) and Ding et al. (2000), where the optimization is tackled through transfer matrix factorization.

These methods mainly address just-proper systems since the full rank of $D_f$ - the direct-feedthrough from faults to residuals - is usually assumed, which makes these methods more suitable to sensor faults than actuators and components faults where it is common to have $D_f = 0$. To relax that requirement, different strategies have been applied. In Ding et al. (2000) multiple observers are suggested for strictly-proper systems so that each observer only addresses one fault. In Rank and Niemann (1998, 1999) and Henry and Zoghadri (2005, 2006), weighted filters are applied. In Wang et al. (2007), artificial elements are added to make systems just-proper.

However, the solution to the $H_\infty$/H_\infty optimization of strictly-proper systems is still absent. Since $H_\infty$ of a strictly-proper system is always zero in the whole frequency range, the optimization is solvable only when the frequency range is considered, which makes the problem more difficult.

Part II of the paper addresses the $H_\infty$/H_\infty optimization of strictly-proper systems with an approach of unitary system. As shown in Part I of the paper, the singular values of a unitary system can be assigned as the magnitude frequency response of a first-order transfer function so that for such a system over a given frequency range, $H_\infty$ is the maximum and $H_\infty$ is the minimum of the magnitude of the transfer function. This special property is used to construct an $H_\infty$/H_\infty optimized observer for the purpose of fault detection.

The rest of the paper is organized as follows. Section 2 contains introductions of the $H_\infty$/H_\infty optimization and the unitary system. Section 3 presents the unitary-system-based solution to the $H_\infty$/H_\infty optimization of strictly-proper systems. An example is given in Section 4 to show the application of the presented method in the fault detection observer design. The paper ends in Section 5 with conclusions.
2. PRELIMINARY: $H_\infty/H_1$ OPTIMIZATION AND UNITARY SYSTEM

In this section, preliminary information about the $H_\infty/H_1$ optimization and the unitary system will be presented. For the simplicity, we will only discuss a square system in which the numbers of inputs and outputs are equal.

2.1 $H_\infty/H_1$ optimization

A multi-input multi-output (MIMO) linear time-invariant (LTI) system can be described by a transfer matrix whose elements are transfer functions. The singular values of a transfer matrix are non-negative real functions of frequency. Of all singular values, the largest one $\sigma(s)$ and the smallest one $\sigma(s)$ give the maximal and minimal possible gains of a system for a given frequency $\omega$ ($s = j\omega$). For a transfer matrix $G(s)$ in a given frequency range $[0 \Omega]$, $H_\infty$ (Zhou et al. 1996) is the supreme of the largest singular value $\sigma(s)$ and $H_1$ (Wang et al. 2007) is the infimum of the smallest singular value $\sigma(s)$:

$$H_\infty[G(s)] = \sup_{s \in [0 \Omega]} \sigma[G(s)]$$

$$H_1[G(s)] = \inf_{s \in [0 \Omega]} \sigma[G(s)].$$

In the studies of fault detection, the $H_\infty/H_1$ optimization is a combined optimization to seek the balance of the robustness to disturbances and the sensitivity to faults. Consider a system in the following form:

$$\dot{x} = Ax + Bu + Ed + Ff$$
$$y = Cx$$

where $u$ is the control inputs; $d$ and $f$ are disturbances and faults respectively; $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times m}$, $E \in \mathbb{R}^{n \times p}$, $F \in \mathbb{R}^{n \times q}$, $C \in \mathbb{R}^{m \times n}$. The input-output relationship of system (3) can be expressed in a transfer matrix form as

$$y(s) = G_r(s)u + dG_r(s)d + G_f(s)f,$$

where $G_r(s) = C(sI - A)^{-1}B$ and :

$$G_r(s) = C(sI - A)^{-1}E,$$

$$G_f(s) = C(sI - A)^{-1}F.$$  

An observer for the purpose of fault detection can be built as:

$$\dot{x} = Ax + Bu - L(y - \hat{y})$$
$$\hat{y} = Cx$$

The estimation errors $\tilde{x} = x - \hat{x}$ and $\tilde{y} = y - \hat{y}$ thus have the form of

$$\dot{\tilde{x}} = (A + LC)\tilde{x} + Ed + Ff$$
$$\dot{\tilde{y}} = C\tilde{x}$$

The residuals are defined as weighted estimation errors

$$r = W\tilde{y}$$

where $W \in \mathbb{R}^{m \times m}$ is a weight matrix. In the transfer matrix form

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

where:

$$G_{rd}(s) = WC_\infty(sI - A - LC)^{-1}E_\infty = WG_r(s)$$

with

$$G_r = \left[ I - C(sI - A)^{-1}L \right]^{-1}.$$  

The structure of the fault detection system is shown in Fig. 1.

![Fig. 1 Fault detection observer](image)

The objective of fault detection is, by observing the residuals $r$, to detect the occurring faults $f$ with the interference of disturbances $d$. To increase the accuracy of the fault detection, it is desirable to increase the robustness to disturbances and the sensitivity to faults of the residuals. The measurement of the robustness to disturbance is taken as $H_\infty[G_{rd}(s)]$ since one has

$$\left\| G_{rd}(s)d(s) \right\| \leq H_\infty[G_{rd}(s)]\|d(s)\|.$$  

Similarly, the measurement of the sensitivity to faults $f(s)$ is taken as $H_\infty[G_{rf}(s)]$ since one has

$$\left\| G_{rf}(s)f(s) \right\| \geq H_\infty[G_{rf}(s)]\|f(s)\|.$$  

To increase the robustness to disturbances, it is desirable to reduce $H_\infty[G_{rd}(s)]$. To increase the sensitivity to faults, it is required to increase $H_\infty[G_{rf}(s)]$. To obtain the best balanced robustness and sensitivity, the $H_\infty/H_1$ optimization is defined as to

$$\min_{H_\infty/H_1} H_\infty[G_{rd}(s)]$$

with the selection of constant matrices $W \in \mathbb{R}^{m \times m}$ and $L \in \mathbb{R}^{n \times q}$.

**Remark 2.1:** In Rambeaux et al. (2000), Chung and Speyer (1997), Liu and Zhou (2008), Wang et al. (2007), Jaimoukha et al. (1999) and Ding et al. (2000), $G_f(s)$ has the form of

$$G_f(s) = C(sI - A)^{-1}F + D_f,$$

where $D_f$ is a full rank matrix. This makes $G_f(s)$ a just-proper transfer matrix since

$$\lim_{s \to \infty} G_f(s) = \lim_{s \to \infty} C(sI - A)^{-1}F + D_f = D_f.$$  

The limitation of the full rank assumption on $D_f$ is that, usually, only sensor faults appear as elements in $D_f$, which makes these methods more suitable to sensor faults detection than actuators and components faults detection where it is common to have $D_f = 0$ as shown in (6).

**Remark 2.2:** $G_f(s)$ in (6) is a strictly-proper transfer matrix since, for $s = j\omega$.
\[
\lim_{\omega \to \infty} G_j(s) = \lim_{\omega \to \infty} C(sI - A)^{-1} F = 0_{\text{row}}.
\]

Thus, \( G_j(s) \) is a strictly-proper transfer matrix too. Therefore the optimization in (16) can only be solved for a finite frequency range since, no matter how \( L \) and \( W \) are selected, the singular values of \( G_j(s) \) always approach zero as the frequency increases to infinity, which means the whole frequency range is considered. In this paper, the frequency range is set as \( \omega \in [0, \Omega] \) with \( \Omega \) as any selected upper bound.

### 2.2 Unitary system

The \( H_\infty/H_\infty \) optimization is addressed in this paper with a unitary system approach. A brief introduction of unitary system is given in the Definition and Lemma 1 below, which are taken from Theorem 1 in Part I of this paper.

**Definition:** A stable linear time-invariant system of \( m \)-inputs and \( m \)-outputs is defined as a unitary system if its transfer matrix \( G(s) \) satisfies

\[
\sigma_i(s) = \sigma_i(s) = \cdots = \sigma_m(s),
\]

where \( \sigma_i(s) \) is the \( i \)-th singular value of \( G(s) \), whose singular value decomposition (SVD) has the form of

\[
G(s) = U(s) \Sigma(s) V^*(s),
\]

where \( \Sigma(s) = \text{diag}(\sigma_1, \ldots, \sigma_m) \) is a diagonal matrix of singular values; \( U(s) \) and \( V(s) \) are unitary complex matrices such that \( U(s)V(s) = I_m \) (\( ^* \) denotes the transpose of the conjugate such that \( U^*(s) = U^T(\Sigma) \)).

**Lemma 1:** Given \( G(s) \) with a minimal realization of

\[
G(s) = C(sI - A)^{-1} F,
\]

where \( CF \) is non-singular and \( G(s) \) does not have finite zeros on the imaginary axis, and a unitary system with singular values of \( |s + k + l|^{-1} \) can be constructed as

\[
G_j(s) = (CF)^{-1} C(sI - A - LC)^{-1} F = |s + k + l|^{-1} U(s),
\]

where \( U(s)U^*(s) = I \); and \( L \) is calculated following the procedures of:

1. Calculate:

\[
T = \begin{bmatrix} (CF)^{-1} C & F^{-1} \end{bmatrix}, \quad \hat{A} = TAT^{-1} = \begin{bmatrix} \hat{A}_{11} & \hat{A}_{12} \\ \hat{A}_{21} & \hat{A}_{22} \end{bmatrix};
\] (22)

2. Select \( k \) and calculate

\[
\hat{L}_1 = \begin{bmatrix} -\hat{A}_{11} (CF)^{-1} \\ (kI + \hat{A}_{11}) (CF)^{-1} \end{bmatrix};
\] (23)

3. Calculate

\[
\hat{L}_2 = T^{-1} \hat{L}_1;
\] (24)

4. Solve the positive definite \( Y \) for the Riccati equation

\[
\left[ \hat{A} + \hat{L}_2 C - F(CF)^{-1} C \right] Y + \hat{Y} \left[ \hat{A} + \hat{L}_2 C - F(CF)^{-1} C \right]^T \right] Y = 0
\]

5. Solve the gain

\[
L_2 = \left[ Y (CF)^{-1} Y \right]^{-1} (CF)^{-1} CY = 0
\] (25)

6. Calculate the gain

\[
L = L_1 + L_2.
\] (27)

### 3. A UNITARY SYSTEM SOLUTION TO THE H₂/H₂ OPTIMIZATION

This section addresses the \( H_\infty/H_\infty \) optimization problem in section II:

\[
\min_{H} \| G_m(s) \|_{\infty} = \min_{H} \| W G_h(s) G_m(s) \|_{\infty}
\]

with \( s = j\omega \) and \( \omega \in [0, \Omega] \). In Ding et al. (2000), the solution to the general problem of

\[
\min_{H} \| G_m(s)G_m(s) \|_{\infty},
\]

with an invertible \( G_m(s) \) is given as

\[
G_m(s) = \sigma_0 N(s) G_m^{-1}(s)
\]

and the minimal value is

\[
H = \| G_m^{-1}(s) G_m(s) \|_{\infty}
\]

where, \( \sigma_0 \) is any constant value and \( N(s)N^*(s) = I \). The solution in Ding et al. (2000) is thus to construct a weighted observer so that:

\[
G_m(s) = \sigma_0 N(s) G_m^{-1}(s)
\]

However, this solution is available only for a just-proper \( G_m(s) \), an observer or a filter cannot be constructed to realize \( G_m^{-1}(s) \) which is not proper. Even though, solution (31) still gives the theoretically minimal value to the optimization problem.

In Ding et al. (2000), \( G_j(s) \) is constructed as a transfer matrix of special form that all of its singular values equal to a constant \( \sigma_j \). For a strictly-proper \( G_j(s) \), it is possible to construct such a \( G_j(s) \). However, it is possible to construct a \( G_j(s) \) in the form of a unitary system whose singular values are \( |s + k + l|^{-1} \), or, as the function of frequency, \( (|s| + k + l)^{-2} \). If \( k \) is selected large enough, all singular values of \( G_j(s) \) are equal approximately to a constant value \( |k + l|^{-1} \) in the considered frequency range.

**Theorem 1:** For the system (3) with \( CF \) non-singular, there exists a feedback gain \( L \), which is calculated following the procedures in Lemma 1, and a weight matrix \( W = (CF)^{-1} \) so that

\[
G_j(s) = WC(sI - A - LC)^{-1} F
\]

is a unitary system with singular values \( |s + k + l|^{-1} \); the same \( L \) and \( W \) also solves the \( H_\infty/H_\infty \) optimization problem in (28) so that
that transforms $G_s G_s H G_s$ into a unitary system in the form of $G_j(s) = (CF)^{-1}(CsI - A - LC)^{-1}F$.

Since also from Lemma 1, it is derived

$$G_j(s) = [s + k + 1]^I U(s) G_j^{-1}(s)$$

and thus from (11), one obtains

$$G_j(s) = [s + k + 1]^I U(s) G_j^{-1}(s) G_j(s)$$

which means

$$H_{\omega}[G_j(s)]$$

For $s = j\omega$ and $\omega \in [0, \Omega]$, one has

$$H_{\omega}[G_j(s)]$$

Therefore, the following inequalities hold:

$$H_{\omega}[G_j(s)] \leq \sup_{\omega} j\omega + k + 1 \cdot |U(s) G_j^{-1}(s) G_j(j\omega)|$$

In the frequency range of $\omega \in [0, \Omega]$, inequalities (41) are the same as

$$H_{\omega}[G_j(s)] \geq |j\omega + k + 1| \cdot H_{\omega}[G_j^{-1}(j\omega) G_j(j\omega)]$$

On the other hand, one has

$$H_{\omega}[G_j(s)] = \sup_{\omega} j\omega + k + 1 \cdot |U(s) G_j^{-1}(j\omega)|$$

Thus, the following inequalities are obtained:

$$H_{\omega}[G_j(s)] \geq \sup_{\omega} j\omega + k + 1 \cdot |U(s) G_j^{-1}(j\omega)|$$

Remark 3.1: We have discussed that the theoretical minimum of the $H_{\omega}/H_{\omega}$ optimization is $H_{\omega}[G_j^{-1}(s) G_j(s)]$. For a strictly-proper system $G_j(s)$, this solution is not realizable. However by transforming $G_j(s)$ into a unitary system, the $H_{\omega}/H_{\omega}$ can be minimized closely to its theoretical value by selecting a large parameter $|k+1|$. The solution in this paper is at most $\sqrt{\Omega^2 / (k+1)^2 + 1}$ times larger than the theoretical value and can be further reduced by increasing $|k+1|$.

Remark 3.2: The absolute value of $H_{\omega}[G_j(s)]$ can be increased with a smaller $|k+1|$ as shown in (43). However, the optimal value of $H_{\omega}/H_{\omega}$ might be also increased since its upper bound has increased according to (44).

4. AN EXAMPLE

In this section, the $H_{\omega}/H_{\omega}$ optimization is solved for a system in the form of (1), which is adopted from Khosrowjerdi et al. (2005) for the fault detection of a four-tank system shown in Fig. 2, with:

$$A = \begin{bmatrix} -0.0159 & 0 & 0.0419 & 0 \\ 0 & -0.0111 & 0 & 0.0333 \\ 0 & 0 & 0.0419 & 0 \\ 0 & 0 & 0 & -0.0333 \end{bmatrix}, \quad C = \begin{bmatrix} 0.5 & 0 \\ 0.0833 & 0 \\ 0 & 0.0333 \\ 0 & 0.0312 \end{bmatrix}$$

$$B = F = \begin{bmatrix} 0.0833 & 0 \\ 0 & 0.0718 \\ 0 & 0.0479 \\ 0.0312 & 0 \end{bmatrix}, \quad E = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0.0357 \\ 0 & 0.0313 \end{bmatrix}$$

where $x$ is the level of water; $u = [u_1, u_2]^T$ is the voltage to Pump 1 and 2; $f = [f_1, f_2]^T$ is the fault to the pumps as gain loss; $d = [d_1, d_2]^T$ is the leaking of Tank 3 and 4. The objective is to minimize the $H_{\omega}/H_{\omega}$ as shown in (16) in the frequency range of $\omega \in [0, 1]$ rad/s. By following the procedures Lemma 1 and selecting $k=9$, the observer gain is calculated as

$$L = \begin{bmatrix} -20.1471 & 0.0031 & 42.9487 & 0.9327 \\ -0.0180 & -19.9838 & -4.3145 & -1.7765 \end{bmatrix}$$

With $W = (CF)^{-1}$ and $L$ calculated above, the $H_{\omega}/H_{\omega}$ optimization is solved according to Theorem 1 so that:

$$G_{rd} = (CF)^{-1}(sI - A - LC)^{-1}F$$

The singular values plots of $G_j(s)$, $G_j(s)$, $G_j(s)$ and $G_{rd}(s)$ are shown in Fig. 2. The values of $H_{\omega}[G_j(s)]$,....
From the results, one can see that:
\[ H_{\infty}[G_r(s)] / H_{\infty}[G_f(s)] = 31.89 \text{dB} \]
\[ H_{\infty}[G_{rd}(s)] / H_{\infty}[G_{rf}(s)] = -7.63 \text{dB} \]
which means that the \( H_{\infty}/H_{\infty} \) value is reduced 39.53 dB after the optimization. The theoretical optimum is -7.68 dB. The solution in this paper is only 0.05 dB from the optimum and the difference can be further reduced by increasing \(|k+1|\).

Moreover, if a larger frequency range needs to be considered for the optimization, the same accuracy can be obtained by increasing \(|k+1|\).

To show the use of the proposed unitary system method, simulations are carried out to detect pump faults of 30% gain loss or \( f = -0.3u \), where \( u \) is calculated with an output feedback controller. In the first simulation, the disturbance is assumed to be zero to show the effect of unitary system in the fault detection observer. In the second simulation, constantly leaking of \( \mathbf{d} = [1 \ 1]^T \) is assumed to show the effect of the \( H_{\infty}/H_{\infty} \) optimization. In both cases, \( y \) is controlled to track sine signals of frequency 1 rad/s and the faults occur at 10 second.

If there is no disturbance as in Simulation 1, the singular values of the resulted unitary system is \([\omega^2 + (k+1)^2]^{-1/2} = 1/(k+1)\) for the considered frequency range, which means that the magnitude (2-norm) of \( r \) has the form of \( ||r|| = ||f||/(k+1) \). This is consistent with the result in Fig. 4, where it can be seen that \( ||f|| \) is approximately \( k+1 \) (in this example 10) times of \( ||f|| \). The magnitude of the occurring faults thus can be determined by just observing the magnitude of \( r \). In Simulation 2 with disturbance, it can be seen from Fig. 5 that the occurring faults can be easily detected since \( ||r|| \) deviates from its normal value instantaneous at the occurring of faults.

5. CONCLUSIONS

In Part II of the paper, we presented a solution to the \( H_{\infty}/H_{\infty} \) optimization of strictly-proper systems by transforming the fault-to-residual transfer matrix into a unitary system. The theoretical minimal value of the \( H_{\infty}/H_{\infty} \) optimization cannot be obtained for strictly-proper systems - which usually represent faults in actuators and components in model-based fault detection - since the inverse of such systems are unavailable. However, in the solution presented in this paper, it can be approached as close as possible for any given frequency range starting from zero by increasing the value of \(|k+1|\).
Furthermore, the maximal inaccuracy of the optimization result can be calculated with the given upper bound of frequency and the selected parameter $(k+1)$. An example shows the improvement over the $H_\infty$ value due to the optimization.

REFERENCES


