Minimum sensitivity IIR filter design using principal component approach

Prof. B.-S. Chen, PhD
B.-W. Chioù, MSc
S.-C. Peng, MSc

Indexing terms: Filters and filtering, Design

Abstract: An IIR filter design algorithm in the state space model is presented. The methods of principal component analysis and balanced realisation are employed to solve the design problem so that the designed filter in state space form achieves minimum sensitivity to parameter variation and/or roundoff noise. Two design examples are also presented to indicate the advantages.

1 Introduction

Various methodologies and approximation algorithms such as the Padé approximation [1], the least-squares inverse technique [1] and the Chebyshev approximation [2], have been reported for IIR filter design.

Modern concepts of VLSI design enable the direct implementation of algorithms of linear algebra and thus lead to a feasible realisation of digital signal processors. It is well known that direct implementation of the signal flow graph corresponds to the implementation of the state equations [9]. It is desirable to take into consideration the parameter variation and roundoff noise to achieve a design of optimal state space realisation according to practical requirements. The state space realisation with minimum sensitivity to parameter variation and/or roundoff noise is thus desirable in practical filter design.

Several papers [15, 16, 17] have shown that an IIR filter with internally balanced realisation has minimum sensitivity to the variation of coefficients and the condition for a digital filter which simultaneously minimises sensitivity to parameter variation and roundoff noise has also been obtained. A minimum sensitivity IIR filter design in state space model using principal component approximation and a balanced realisation technique is proposed. Although there exist numerous techniques for approximation [3–7], the resultant designs are not always stable. It is often necessary to find and modify the unstable poles of the filter, which requires more work than the filter design itself [6, 7]. For example, any unstable poles can always be dealt with by inversion with respect to the unit circle. However, the phase response can be affected.

Since the IIR filter is in an internally balanced realisation, it is obvious that the input-to-state coupling and the state-to-output coupling are weighted equally, i.e., the controllability gramian is equal to the observability gramian [8]. The state components which are weakly coupled to both the input and output can thus be discarded [8, 11, 12] for model reduction. The filter design proposed is based on the dominant part of the singular value decomposition of the Hankel matrix formed from the modified impulse response of desired filter, and discards those state components which are weakly coupled to both input and output. The order reduced filter which is still realised in balance form has minimum sensitivity to parameter variation. The resulting realisation can simultaneously attain minimum sensitivity to both parameter variation and roundoff noise by transforming the balanced state space model using a specified matrix [8, 9, 16, 17]. This method is simple because standard matrix software can be employed to solve the computation problem. The resulting designs are always stable, so it is not necessary to find and modify the unstable poles of the filter.

Principal component analysis and problem formulation is first addressed. The balanced realisation based on singular value decomposition is then given. Order reduced filter design is described. Two design examples are then given to illustrate the design algorithm.

2 Principal component analysis and problem formulation

Suppose an ideal amplitude characteristic of a digital filter (for example: an ideal low-pass filter) is given. The first step of digital filter design using an approximation method is to transform this ideal amplitude characteristic of the desired digital filter into a high-order analytic function

$$H(z) = \sum_{i=1}^{\infty} h_i z^{-i}$$  \hspace{1cm} (1)

which has been windowed. The shifted-truncated technique with band-limited window can be applied to this problem. In eqn. 1, the sequence \( \{h_i\} \) denotes the modified impulse response. The next step is to employ an approximation technique to find a rational transfer function \( H(z) \) with the desired order to approximate \( H(z) \).

The internally balanced state space model with several good properties is employed to solve the problem from the viewpoint of principal component analysis. For the analytic function \( H(z) \) in eqn. 1, let the associated Hankel matrix be defined as follows:

$$\Phi(H) = \begin{bmatrix} h_1 & h_2 & \cdots & h_n \\ h_2 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ h_n & \cdots & \cdots & 0 \end{bmatrix}$$  \hspace{1cm} (2)
Let \( \phi(H) \) have the singular value decomposition as follows:
\[
\phi(H) = U \Sigma V^T
\]
where \( \Sigma \) is the diagonal matrix consisting of the nonzero singular values of \( \phi(H) \), i.e., \( \Sigma = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_m) \) with the ordering \( \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_m > 0 \), and \( U = [u_1, u_2, \ldots, u_m] \), \( V = [v_1, v_2, \ldots, v_m] \), with the \( u_i \) and the \( v_i \) being the left singular vector and right singular vector, respectively. \( U \) and \( V \) are unitary matrices, so
\[
U^T U = V^T V = I
\]
It, from eqn. 3, can be shown that
\[
\phi(H) = \sum_{i=1}^{m} \sigma_i u_i v_i^T
\]
where the singular values are ordered as \( \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_m > 0 \), and \( u_i \) and \( v_i \) are the left and right singular vector associated with \( \sigma_i \). They are mutually orthogonal unit vectors.

Note that the \( k \)th component is \( u_k v_k^T \) and the \( i \)th component magnitude is \( \sigma_i \). The first \( r \) principal components are defined as
\[
\phi_r(H) = \sum_{i=1}^{r} \sigma_i u_i v_i^T
\]
Using the Frobenius metric defined as follows:
\[
\|X\|_F = \left( \sum_{i=1}^{m} \sum_{j=1}^{n} |x_{ij}|^2 \right)^{1/2}
\]
where \( x_{ij} \) denotes the \( ij \)th entry of the matrix \( X \in \mathbb{R}^{m \times n} \) and defining
\[
d(X, Y) = \left( \sum_{i=1}^{m} \sum_{j=1}^{n} |x_{ij} - y_{ij}|^2 \right)^{1/2}
\]
as the measurement of the distance between any two \( m \times n \) matrices \( X \) and \( Y \), then
\[
\|\phi(H)\|_F = \left( \sum_{i=1}^{m} \sigma_i^2 \right)^{1/2}
\]
It has been shown that the unique matrix \( \phi_r(H) \) of rank \( r \) that lies closest to \( \phi(H) \) in the minimum Frobenius norm sense is given by eqn. 6 [8, 13]. Optimal approximation is achieved when
\[
\|\phi(H) - \phi_r(H)\|_F = \left( \sum_{i=r+1}^{m} \sigma_i^2 \right)^{1/2}
\]
From the above analysis, consider
\[
\|\phi(H)\|_F^2 = \sum_{i=1}^{m} \sigma_i^2
\]
as the total energy of the signal sequence \( \{h_1, h_2, \ldots, h_m\} \), and the component matrix \( u_i v_i^T \) with the magnitude \( \sigma_i \) reflects the spatial distribution of the energy. It is shown that the first \( r \) significant components \( \sum_{i=1}^{r} \sigma_i u_i v_i^T \) are just the optimal approximation of \( r \)-order system to the \( m \)-order system in Frobenius norm sense and that the error (or residual) is equal to
\[
\left( \sum_{i=r+1}^{m} \sigma_i^2 \right)^{1/2}
\]
Let the following state space model be a balanced realisation of eqn. 1:
\[
x(k+1) = Ax(k) + Bu(k) \tag{12a}
\]
\[
y(k) = Cx(k) \tag{12b}
\]
where \( x(k) \in \mathbb{R}^m \) and \( A, B \) and \( C \) satisfy the following constraints [8, 11, 12]
\[
\sum_{k=0}^{\infty} (A^k B)(A^k B)^T = \sum_{k=0}^{\infty} (CA^k)(CA^k)^T = \Sigma \tag{13}
\]
or equivalently
\[
A^T \Sigma A + B B^T = \Sigma \tag{14a}
\]
\[
A^T \Sigma A + C^T C = \Sigma \tag{14b}
\]
One of the advantages of discrete-time balanced realisation is that the output of a digital filter has minimum sensitivity to parameter variation if the filter is realised in a balanced form [15–17].

The first problem is how to find an order reduced filter, which is still balanced
\[
x(k+1) = A x(k) + B u(k) \tag{15a}
\]
\[
y(k) = C x(k) + D u(k) \tag{15b}
\]
to approximate the balanced state space model of eqn. 12, where \( x(k) \in \mathbb{R}^r \) and \( r < m \). A good approximation can be obtained from the viewpoint of principal component analysis. The design strategy is to delete some nonessential 'dynamical elements' of eqn. 12. The approximation considered is primarily a state-space co-ordinate problem. It is necessary to set up the definition of the contribution measurement of each dynamical element of eqn. 12. A main result of recent research in approximation theory is the use of the so-called singular values of Hankel matrix as the contribution measurement of dynamical elements. After a truncation of less significant singular values, the principal components construct an order reduced state space model.

The second problem is how to find an optimal state space model design \( \{A, B, C, D\} \) of a digital filter which simultaneously minimises sensitivity to parameter variation and roundoff noise.

The proposed approximation technique is also an optimal design from the viewpoint of the minimum Frobenius norm [8, 13]. This method is simple because standard matrix software can be applied to solve the computation problem.

### 3 Balanced realisation based on singular value decomposition

For any realisation in eqn. 12, the observability matrix and the controllability matrix are denoted as
\[
\Omega_o = \begin{bmatrix} C \\ CA \\ \vdots \\ C A^{m-1} \end{bmatrix} \tag{16}
\]
\[
\Omega_c = \begin{bmatrix} B & AB & \cdots & A^{m-1} B \end{bmatrix} \tag{17}
\]
It has been shown [14] that
\[
\phi(H) = \Omega_o \Omega_c \tag{18}
\]
\[i.e., \ h_i = C A^{i-1} B, \ i = 1, 2, \ldots, 2m - 1. \]
Note that since eqns. 13, 14a and 14b are all satisfied, \( \{A, B, C\} \) in eqn. 12 is in balanced realisation. If the filter \( \{A, B, C\} \) in eqn. 12 satisfies the following equalities:
\[
\Omega_o = E \Sigma^{1/2} \tag{19}
\]
and
\[
\Omega_c = \Sigma^{1/2} V^T \tag{20}
\]
then
\[
\Omega_{2} \Omega_{1} = U^{1/2} U^{1/2} \Sigma^{1/2} = \Sigma
\]
\[
= \Sigma^{1/2} \Sigma^{1/2} V^{\top} = \Omega_{1} \Omega_{2}^{\top}
\]  
(21)

In system theory, \( W = Q_{2} \Omega_{1}^{\top} \) and \( M = \Omega_{1} Q_{2}^{\top} \) are called the controllability grammian matrix and the observability grammian matrix, respectively. When a filter is realised in a balanced state space form, the controllability grammian matrix and the observability grammian matrix are equal. In other words, if \( (A, B, C) \) is chosen such that eqns. 19 and 20 hold, the filter is in balanced state space realisation.

By comparing eqn. 16 with eqn. 19 and eqn. 17 with eqn. 20, it is seen that \( C \) and \( B \) are given as follows
\[
C = \text{the first row of } U \Sigma^{1/2}
\]  
(22)
\[
B = \text{the first column of } U \Sigma^{1/2} V^{\top}
\]  
(23)

Next, concentrate on the determination of \( A \). Define
\[
\Omega_{1} = \begin{bmatrix} CA & CA^{2} & \cdots & CA^{n-1} \\ CA^{n} & 0 \end{bmatrix}
\]  
(24)
\[
\Omega_{0} = \begin{bmatrix} C & CA & CA^{2} & \cdots & CA^{n-1} \\ CA & 0 \end{bmatrix}
\]
\[
\Omega_{1} \text{ denotes the matrix formed by shifting up one row from } \Omega_{0} \text{ and filling in } 0 \text{ for the last row. Postmultiplying both sides of eqn. 16 by } A
\]
\[
A = \Omega_{0}^{-1} \Omega_{1} = \begin{bmatrix} C & CA & CA^{2} & \cdots & CA^{n-1} \\ CA & 0 \end{bmatrix} = \Omega_{1}^{\top}
\]
\[
\text{since } CA^{n} = 0 \text{ (see Lemma 1 in the Appendix). Thus}
\]
\[
A = \Omega_{0}^{-1} \Omega_{1}^{\top}
\]  
(26)
\[
= \left( U \Sigma^{1/2} V \right)^{-1} \left( U \Sigma^{1/2} V \right)^{\top}
\]  
(27)

From the above analysis, a balanced realisation filter \( (A, B, C) \) of eqn. 1 can be obtained from eqns. 27 and 22, respectively, using singular value decomposition of Hankel matrix \( \phi(H) \). It is asymptotically stable, controllable, observable and balanced according to Theorem 1 given as follows:

**Theorem 1:** If
\[
A = \left( U \Sigma^{1/2} V \right)^{-1} \left( U \Sigma^{1/2} V \right)^{\top}
\]
\[
B = \text{the first column of } U \Sigma^{1/2} V^{\top}
\]
\[
C = \text{the first row of } U \Sigma^{1/2}
\]
where \( U \) and \( V \) are unitary matrices and \( \Sigma = \text{diag}(\sigma_{1}, \sigma_{2}, \ldots, \sigma_{m}) \in \mathbb{R}^{n \times n} \) in eqn. 3 then the filter \( (A, B, C) \) is asymptotically stable, controllable, observable and balanced.

**Proof:** Since \( A, B \) and \( C \) satisfy eqns. 19 and 20, rank \( \Omega_{0} \) and rank \( \Omega_{1} \) are full rank. The system \( (A, B, C) \) is obviously controllable, observable and balanced. Since the system \( (A, B, C) \) is balanced, eqns. 14a and 14b hold.

Let \( v \) be an eigenvector of \( A^{\top} \) and let \( \lambda \) be the corresponding eigenvalue. Then \( A^{\top} v = \lambda v \). Multiply eqn. 14a from the right by \( v \) and from the left by \( v^{\top} \), then
\[
(\lambda^{2} - 1)v^{\top} \Sigma v = -v^{\top} BB^{\top} v
\]  
(28)

Since the right member is nonpositive and \( v^{\top} \Sigma v \) is positive, it follows that
\[
|\lambda| \leq 1
\]  
(29)
Suppose that \( |\lambda| = 1 \). Then it follows from eqn. 28 that
\[
v^{\top} B = 0
\]  
(30)
The relation
\[
v^{\top} A = \lambda v^{\top}
\]  
(31)
eqn. 30 contradict the controllability on the system \( (A, B, C) \). Therefore, \( |\lambda| \neq 1 \) and it follows that \( |\lambda| < 1 \), i.e. the system \( (A, B, C) \) is asymptotically stable. The proof is complete.

**4 Order reduced filter design**

The minimum sensitivity filter design is introduced with respect to parameter variation and/or roundoff noise.

**4.1 Minimum sensitivity filter design with respect to parameter variation**

From the analysis presented, the design work is reduced to how to obtain the following filter with a lower order, which is still balanced:
\[
x(k + 1) = A_{f} x(k) + B_{f} u(k)
\]
\[
y(k) = C_{f} x(k) + D_{f} u(k)
\]
where
\[
A_{f} \in \mathbb{R}^{r \times r}, \quad B_{f} \in \mathbb{R}^{r \times 1}, \quad C_{f} \in \mathbb{R}^{1 \times r}, \quad D_{f} \in \mathbb{R}
\]
such that \( (A_{f}, B_{f}, C_{f}, D_{f}) \) is a balanced realisation of the first \( r \) significant components
\[
\sum_{i=1}^{r} \sigma_{i} u_{i} v_{i}^{\top}
\]
Partition \( A, B, C \) and \( \Sigma \) compatibly as follows
\[
A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad B = \begin{bmatrix} B_{1} \\ B_{2} \end{bmatrix}, \quad C = \begin{bmatrix} C_{1} & C_{2} \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \Sigma_{1} & 0 \\ 0 & \Sigma_{2} \end{bmatrix}
\]
where
\[
A_{11} \in \mathbb{R}^{r \times r}, \quad A_{21} \in \mathbb{R}^{r \times (n-r)}, \quad A_{22} \in \mathbb{R}^{r \times r}, \quad A_{12} \in \mathbb{R}^{(n-r) \times r}
\]
\[
C_{1} \in \mathbb{R}^{1 \times r}, \quad C_{2} \in \mathbb{R}^{1 \times (n-r)}, \quad B_{1} \in \mathbb{R}^{r \times 1}, \quad B_{2} \in \mathbb{R}^{(n-r) \times 1}
\]
\[
\Sigma_{1} = \text{diag}(\sigma_{1}, \ldots, \sigma_{r}), \quad \Sigma_{2} = \text{diag}(\sigma_{r+1}, \ldots, \sigma_{m})
\]
From eqn. 12
\[
\begin{bmatrix} x_{1}(k+1) \\ x_{2}(k+1) \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} x_{1}(k) \\ x_{2}(k) \end{bmatrix} + \begin{bmatrix} B_{1} \\ B_{2} \end{bmatrix} u(k)
\]
\[
y(k) = C_{1} x_{1}(k) + C_{2} x_{2}(k)
\]
or
\[
x_{1}(k+1) = A_{11} x_{1}(k) + A_{12} x_{2}(k) + B_{1} u(k)
\]
\[
x_{2}(k+1) = A_{21} x_{1}(k) + A_{22} x_{2}(k) + B_{2} u(k)
\]
\[
y(k) = C_{1} x_{1}(k) + C_{2} x_{2}(k)
\]
Since the state space model $(A, B, C)$ using direct truncation is not a balanced realisation, some manipulations are necessary and will be introduced.

Since $x_1(k)$ approaches the steady state more quickly than $x_2(k)$, the effect of $x_2(k)$ to the system is less significant than that of $x_1(k)$. In this situation, with the approximation $x_1(k+1) \approx x_1(k)$ from eqn. 33:

$$ (I - A_{22})x_2(k) = A_{21}x_1(k) + B_2w(k) $$

By Theorem 1, $(A, B, C)$ is asymptotically stable. So

$$ |\lambda_i(A_{22})| < 1 $$

for all $i$ (see Lemma 2 in the appendix) and

$$ (I - A_{22})^{-1} $$

exists

Thus

$$ x_2(k) = (I - A_{22})^{-1}A_{21}x_1(k) + (I - A_{22})^{-1}B_2w(k) $$

Substituting eqn. 37 into eqns. 32 and 34 gives

$$ x_1(k + 1) = [A_{11} + A_{12}(I - A_{22})^{-1}A_{21}]x_1(k) + [B_1 + A_{12}(I - A_{22})^{-1}B_2]w(k) $$

$$ y(k) = [C_1 + C_2(I - A_{22})^{-1}A_{21}]x_1(k) + C_2(I - A_{22})^{-1}B_2w(k) $$

The order reduced filter is defined as

$$ x'(k + 1) = A'x'(k) + B'w(k) $$

$$ y(k) = C'x'(k) + D'w(k) $$

where

$$ A' = A_{11} + A_{12}(I - A_{22})^{-1}A_{21} $$

$$ B' = B_1 + A_{12}(I - A_{22})^{-1}B_2 $$

$$ C' = C_1 + C_2(I - A_{22})^{-1}A_{21} $$

$$ D' = C_2(I - A_{22})^{-1}B_2 $$

The defined order reduced filter is balanced and asymptotically stable according to Theorem 2 given as follows:

**Theorem 2**

(a) The order reduced filter $(A', B', C', D')$ in eqn. 38 satisfies

$$ A'S - S + B'B^T = 0 $$

$$ A'TS - S + C^TC = 0 $$

the order reduced filter in state space form is thus balanced.

(b) If $S_1$ and $S_2$ have no diagonal entries in common, $(A', B', C', D')$ is asymptotically stable.

The proof of this theorem is given in the Appendix.

It is known [15-17] that the balanced state space realisations contain minimum sensitivity to parameter variation. The transfer function associated with $(A', B', C', D')$ is

$$ H'(z) = D' + C'(zI - A')^{-1}B' $$

4.2 Minimum sensitivity filter design with respect to both parameter variation and roundoff noise

The optimal state space model design of a digital filter which simultaneously minimises sensitivity to parameter variation and roundoff noise is considered. The state space realisation $(A', B', C', D')$ corresponding to a transfer function $H'(z)$ is not unique, since for any arbitrary nonsingular matrix $\tilde{T}$

$$ (\tilde{A}, \tilde{B}, C, \tilde{D}) = (T^{-1}A'T, T^{-1}B', C'T, D') $$

will also have transfer function $H'(z)$. In what follows, a realisation $(\tilde{A}, \tilde{B}, C, \tilde{D})$ which simultaneously attains minimum sensitivity to parameter variation and roundoff noise is found. It has been shown that [16, 17] if the controllability grammian matrix and the observability grammian matrix are of the following form:

$$ M = \rho^2W $$

where

$$ \rho = \left[ \frac{\sum_{r=1}^r \sigma_r}{r} \right] $$

$$ W = \begin{bmatrix} 1 & X \\ 1 & X \\ \vdots & \vdots \\ 1 & 1 \end{bmatrix} $$

The filter in state space model will simultaneously achieve minimum sensitivity to parameter variation and roundoff noise. The nonsingular matrix $\tilde{T}$ in eqn. 40 is chosen as

$$ \tilde{T} = \delta P^T $$

with

$$ PP^T = I $$

where

$$ \delta = \left[ \frac{\sum_{r=1}^r \sigma_r}{r} \right]^{1/2} $$

and

$$ P = P_1, P_{r-1}, \ldots, P_1, \ldots, P_2 $$

with

$$ P_i = \begin{bmatrix} I & 0 & \cdots & 0 \\ \vdots & \cos \psi_i & \cdots & \sin \psi_i \\ 0 & I & \cdots & 0 \\ \vdots & -\sin \psi_i & \cdots & \cos \psi_i \\ 0 & 0 & \cdots & I \end{bmatrix} $$

1 = 2, \ldots, r

which can be obtained by Hwang [18], then eqn. 41 is achieved and the realisation will simultaneously achieve minimum sensitivity to parameter variation and roundoff noise.

4.3 Discussions and design algorithm

**Discussion**

(i) The transfer function of the designed filter is given by

$$ H'(z) = D' + C'(zI - A')^{-1}B' $$

$$ = \tilde{D} + \tilde{C}(\tilde{z}I - \tilde{A})^{-1}\tilde{B} $$

(ii) Since the order reduced filter $(A', B', C', D')$ (i.e. eqn. 38) satisfies the balanced realisation conditions eqns. 39a and 39b, it has been proven that this realisation has the advantage of minimum sensitivity to parameter variation [15, 16].

(iii) The error is $\sum_{r=1}^r \sigma_r^{1/2}$ in minimum Frobenius norm sense because the order reduced filter $(A', B', C', D')$ satisfies eqns. 39a and 39b as proved in theorem 2.
(iv) The order reduced filter \((A, B, C, D)\) can simultaneously minimize sensitivity to parameter variation and roundoff noise as proved in Reference 16 and its error is also \(\left(\sum_{i=1}^{m+1} s_i r_i^2\right)^{1/2}\) in minimum Frobenius norm sense because \(CA^{-1}B = CA^{-1}B_k\) for \(k = 1, \ldots, 2m - 1\), i.e., its Hankel matrix remains invariant.

(v) Only singular value decomposition and some matrix manipulations are required. This method is thus computationally simple.

**Design algorithm**

From the above analysis, a design algorithm for the order reduced filter design using a principal component approach is proposed as follows:

**Step 1:** Give the required error bound \(\varepsilon\) to attain \(\|\phi(H) - \phi(H_0)\|_F \leq \varepsilon\).

**Step 2:** Obtain a high-order windowed impulse response \(h\) (see eqn. 1).

**Step 3:** Obtain the associated Hankel matrix \(\phi(H)\) from \(h\) and perform the singular value decomposition \(\phi(H) = U \Sigma V^T\).

**Step 4:** Find filter order \(r\) by \(\sum_{i=1}^{m+1} s_i^2 r_i^2 \leq \varepsilon\) and obtain \(A, B\) and \(C\) from eqns. 27, 23 and 22, respectively.

**Step 5:** Partition \(A\), \(B\) and \(C\) compatibility into \(A_{11}\), \(A_{12}, A_{22}, B_1, B_2, C_1\) and \(C_2\) (see eqns. 32-34).

**Step 6:** Obtain \(A', B', C', D\) using eqn. 38, i.e., obtain the minimum sensitivity realisation to parameter variation.

**Step 7:** Obtain \(A, B, C, D\) using eqn. 40 with \(\tilde{T} = uP\) (see eqn. 42), i.e., obtain the minimum sensitivity realisation to both parameter variation and roundoff noise.

**5 Examples**

**5.1 Low pass filter design**

Consider the ideal low-pass filter, whose characteristic is specified as follows:

\[
\begin{align*}
|H(e^{j\omega})| &= 1 \quad |\omega| < 1 \\
|H(e^{j\omega})| &= 0 \quad 1 < |\omega| < \pi
\end{align*}
\]

Giving an error bound \(\varepsilon = 0.002\), it is necessary to find \(H(\varepsilon)\) such that \(\|\phi(H) - \phi(H_0)\|_F \leq \varepsilon\). Choose the high-order analytic function

\[
H(\varepsilon) = \sum_{i=1}^{10} h_i \varepsilon^{-1}
\]

where the sequence \(\{h_i\}\) denotes the modified impulse response as shown in Fig. 1. In this case, a Hamming window is employed to modify the impulse response.

Using the proposed design algorithm, two filters \((A', B', C', D')\) and \((A, B, C, D)\) with order 24 are obtained. Their overall and passband frequency responses \(|H(e^{j\omega})|\), overlaid with the desired \(|H(e^{j\omega})|\), are shown in Fig. 2. The error in the Frobenius norm is

\[
(s_2^2 + s_4^2 + \cdots + s_{10}^2)^{1/2} = 0.0019
\]

Two filters of order 8 \((A', B', C', D')\) and \((A, B, C, D)\), which respectively contain eight principal components, are also given as follows:

\[
\begin{align*}
\text{(i) The minimum sensitivity realisation to parameter variation} \\
A' &= \begin{bmatrix} 0.9653 & -0.2469 & 0.0275 & -0.0442 & 0.0062 & -0.0093 & 0.0004 & -0.0010 \\
0.2469 & 0.8768 & -0.3598 & 0.0206 & -0.0749 & 0.0002 & -0.0114 & -0.0010 \\
0.0275 & 0.3598 & 0.7425 & -0.4472 & -0.0040 & -0.0074 & -0.0048 & -0.0089 \\
0.0442 & 0.0206 & 0.4472 & 0.5689 & -0.4960 & -0.0228 & -0.0621 & -0.0078 \\
0.0062 & 0.0749 & -0.0040 & 0.4960 & 0.3956 & -0.5093 & -0.0239 & -0.0498 \\
0.0093 & 0.0002 & 0.0074 & -0.0228 & 0.5093 & 0.2565 & -0.4980 & -0.0386 \\
0.0004 & 0.0114 & -0.0048 & 0.0621 & -0.0239 & 0.4980 & -0.0386 & 0.1679 & -0.5081 \\
0.0010 & -0.0010 & 0.0089 & -0.0078 & 0.0498 & -0.0386 & 0.5081 & -0.1207
\end{bmatrix}
\end{align*}
\]

\[
B' = \begin{bmatrix} 0.0848 \\
-0.2297 \\
0.3640 \\
-0.3834 \\
0.2753 \\
-0.1410 \\
0.0541 \\
-0.0204
\end{bmatrix}
\]

\[
C' = \begin{bmatrix} 0.0848 & 0.2297 & 0.3640 & 0.3834 & 0.2753 & 0.1410 & 0.0541 & 0.0204
\end{bmatrix}
\]

\[
D' = 3.0381 \times 10^{-3}
\]
(ii) The minimum sensitivity realisation to both parameter variation and roundoff noise

\[
\lambda = \begin{bmatrix}
0.3969 & 0.1481 & -0.0062 & -0.0501 & -0.4664 & -0.1600 & 0.1022 & -0.4327 \\
-0.1486 & 0.5057 & 0.0748 & 0.4341 & -0.3682 & -0.0285 & -0.3420 & 0.1748 \\
-0.0062 & -0.0795 & 0.5018 & -0.3947 & 0.1234 & -0.0265 & -0.5956 & 0.0116 \\
-0.0170 & -0.4220 & -0.0835 & 0.4939 & 0.0894 & -0.3784 & -0.0220 & 0.0791 \\
-0.4972 & 0.3513 & -0.1847 & -0.0978 & 0.4335 & -0.2251 & 0.1695 & 0.0720 \\
-0.1608 & 0.0274 & -0.0606 & 0.3664 & 0.2611 & 0.5081 & -0.3331 & -0.4804 \\
-0.3711 & -0.1305 & 0.6201 & 0.0171 & -0.1573 & 0.2506 & 0.5085 & -0.0928 \\
0.1612 & 0.3542 & 0.0198 & -0.0794 & -0.0393 & 0.3385 & 0.0279 & 0.5046 \\
\end{bmatrix}
\]

\[
\bar{B} = \begin{bmatrix}
0.0631 \\
-0.1720 \\
0.2296 \\
-0.6368 \\
0.0825 \\
0.0567 \\
-0.1940 \\
-0.6085 \\
\end{bmatrix}
\]

\[
\mathbf{C} = \begin{bmatrix}
0.0509 & 0.1372 & 0.2484 & -0.0371 & 0.0572 & -0.3378 & -0.0368 & 0.0897 \\
\end{bmatrix}
\]

\[
\bar{D} = 3.0381 \times 10^{-3}
\]

with

\[
\mathbf{T} = \begin{bmatrix}
0.4793 & 0.0000 & 0.0000 & 0.0000 & -0.4479 & -0.1441 & -0.1243 & -0.1243 \\
0.0000 & 0.4792 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & -0.3552 & 0.3552 \\
0.0000 & 0.0000 & 0.4932 & -0.4725 & 0.0000 & 0.0000 & -0.0960 & 0.0557 & 0.0557 \\
0.0000 & 0.0000 & 0.0000 & 0.1765 & 0.0000 & 0.0000 & -0.5192 & 0.3011 & 0.3011 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.3141 & 0.0000 & -0.3926 & -0.3385 & -0.3385 \\
0.0000 & 0.0000 & 0.4886 & 0.4770 & 0.0000 & 0.0000 & -0.0969 & -0.0562 & -0.0562 \\
0.0000 & 0.0000 & 0.5023 & 0.0000 & 0.0000 & 0.0000 & 0.3388 & -0.3388 \\
0.5023 & 0.0000 & 0.0000 & 0.0000 & 0.4274 & 0.1375 & 0.1186 & 0.1186 \\
\end{bmatrix}
\]

and the transfer function is

\[
H(z) = \frac{b(z)}{a(z)}
\]

with

\[
a(z) = \sum_{i=0}^{n} a(i)z^{-i}
\]

\[
b(z) = \sum_{i=0}^{n} b(i)z^{-i}
\]

and the coefficients of \(a(z)\) and \(b(z)\) given by

\[
a(0) = 1.0000 \quad b(0) = 0.0000
\]

\[
a(1) = -3.8528 \quad b(1) = -0.0017
\]

\[
a(2) = 7.4202 \quad b(2) = 0.0089
\]

\[
a(3) = -9.1197 \quad b(3) = -0.0169
\]

\[
a(4) = 7.7303 \quad b(4) = 0.0252
\]

\[
a(5) = -4.5920 \quad b(5) = -0.0472
\]

\[
a(6) = 1.8569 \quad b(6) = 0.0550
\]

\[
a(7) = -0.4656 \quad b(7) = -0.0370
\]

\[
a(8) = 0.0553 \quad b(8) = 0.0461
\]
Its overall and passband frequency responses \( |H(e^{j\omega})| \), overlaid with the desired \( |H'(e^{j\omega})| \), are shown in Fig. 3. In this case, \( m \) is taken to be 21 and the error in the Frobenius norm is \( (\sigma_5^2 + \sigma_{10}^2 + \cdots + \sigma_{21}^2)^{1/2} = 0.0048 \).

![Frequency responses for lowpass filter](image)

**Fig. 3** Frequency responses for lowpass filter

- \( r = 8 \)
- \( m = 21 \)
- \( \overline{\text{H}(\omega)} \)
- \( \text{H}(\omega) \)

a) Overall response
b) Passband response

5.2 Differentiator design

Consider the ideal digital differentiator whose characteristic is:

\[
H(e^{j\omega}) = j\omega \quad |\omega| \leq \pi
\]

Giving an error bound \( \varepsilon = 0.002 \), it is necessary to find \( H_z(t) \) such that \( |\phi(H) - \phi(H_z)| \leq \varepsilon \). Choose the high-order analytic function

\[
H(z) = \sum_{l=1}^{m} h_l z^{-l}
\]

where the sequence \( \{h_l\} \) denotes the modified impulse response as shown in Fig. 4. Using the proposed design algorithm, two differentiators of order 32 \((A', B', C', D')\) and \((A, B, C, D)\) are obtained whose frequency response \( |H(e^{j\omega})| \), overlaid with the desired \( |H'(e^{j\omega})| \), is shown in Fig. 5. The error in the Frobenius norm is \( (\sigma_3^2 + \sigma_{14}^2 + \cdots + \sigma_{41}^2)^{1/2} = 8.922 \times 10^{-4} \).

![Windowed shifted-truncated impulse response of differentiator](image)

**Fig. 4** Windowed shifted-truncated impulse response of differentiator

**Fig. 5** Amplitude response of designed differentiator

\( r = 32 \)
\( m = 61 \)

6 Conclusions

A new filter design which has a state space realisation using a principal component approximation was proposed. The advantages of the proposed IIR filter design algorithm are summarised as follows:

(i) The method is computationally simple
(ii) The designed digital filter has minimum sensitivity to parameter variation and/or roundoff noise
(iii) The resulting design is always stable. Hence it is not necessary to find and modify the unstable poles of the filter
(iv) It is possible to predict the error between the desired filter and the designed filter.

7 References


IEE PROCEEDINGS, Vol. 138, No. 4, AUGUST 1991
14 CHEN, C.T.; 'Linear system theory and design' (CBS College, 1984)

8 Appendix

Lemma 1: Given
\[ A = (U\Sigma U^T)^{1/2} \]
where \( U \) is unitary matrix and \( \Sigma = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_n) \) then
\[ CA^m = 0 \]

Proof: By comparing eqn. 16 with eqn. 19
\[ CA^{m-1} \text{ is the last row of } U^T \]
Define the left singular vector
\[ u_i = \begin{bmatrix} u_i^1 \\ u_i^2 \\ \vdots \\ u_i^n \end{bmatrix} \]
then
\[ U = \begin{bmatrix} u_1^T \\ u_2^T \\ \vdots \\ u_n^T \end{bmatrix} \Sigma^{1/2} \]
\[ CA^{m-1} = \begin{bmatrix} u_1^T \\ u_2^T \\ \vdots \\ u_n^T \end{bmatrix} \Sigma^{1/2} \]
Since \( U \) is unitary matrix
\[ U^T = U \]
and the rows of \( U \) form an orthonormal set.
\[ CA^m = (CA^{m-1})A = (CA^{m-1})U\Sigma^{1/2} = \begin{bmatrix} u_1^T \\ u_2^T \\ \vdots \\ u_n^T \end{bmatrix} \]
\[ \times \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n \end{bmatrix} \begin{bmatrix} u_1^T \\ u_2^T \\ \vdots \\ u_n^T \end{bmatrix} = 0 \]

Lemma 2:
\[ |\lambda(A_22)| < 1 \quad \text{for all } i \]

Proof: Since the system \((A, B, C)\) is balanced, eqns. 14a and 14b hold. From eqn. 14a
\[ \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} \Sigma_1 \\ 0 \end{bmatrix} = \begin{bmatrix} A_{11}^T & A_{12}^T \\ A_{21}^T & A_{22}^T \end{bmatrix} \begin{bmatrix} \Sigma_1 \\ 0 \end{bmatrix} \]
\[ = \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} + \begin{bmatrix} B_1 & B_1^T \\ B_2 & B_2^T \end{bmatrix} = 0 \]
The equation for the lower right corner of above equation gives
\[ A_{21}\Sigma_1 A_{12} + A_{22}\Sigma_2 A_{12} = 0 \]
Let \( \hat{\theta} \) be an eigenvector of \( A_{12}^T \) and let \( \hat{\lambda} \) be the corresponding eigenvalue. Then \( A_{12}^{1/2} \hat{\theta} = \lambda \hat{\theta} \). Multiply eqn. 43 from the right by \( \hat{\theta}^* \) and from the left by \( \hat{\theta}^* \), then
\[ (|\hat{\lambda}|^2 - 1)\hat{\theta}^* \Sigma \hat{\theta} = -\hat{\theta}^* A_{12} \Sigma_1 A_{12}^T \hat{\theta} + \hat{\theta}^* B_1 B_2^T \hat{\theta} \]
Since the right member is nonpositive and \( \hat{\theta}^* \Sigma \hat{\theta} \) is positive, it follows that
\[ |\hat{\lambda}| \leq 1 \]

Firstly, suppose that \( |\hat{\lambda}| = 1 \). Then it follows from eqn. 44 that
\[ \hat{\theta}^* A_{21} = 0 \]
\[ \hat{\theta}^* B_2 = 0 \]
since \( \Sigma_1 \) is positive definite. Therefore
\[ \begin{bmatrix} 0 & \hat{\theta}^* \\ A_{21} & A_{22} \end{bmatrix} = \hat{\lambda} \begin{bmatrix} 0 & \hat{\theta}^* \\ A_{21} & A_{22} \end{bmatrix} \]
\[ \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} = 0 \]
Eqs. 46 and 47 contradict the controllability on the system \((A, B, C)\). Therefore, \( |\hat{\lambda}| \neq 1 \) and it follows that \( |\hat{\lambda}| < 1 \). The proof is complete.

Proof of theorem 2
(a) Since the system \((A, B, C)\) is balanced, eqns. 14a and 14b hold. From eqn. 14a
\[ \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} \Sigma_1 \\ 0 \end{bmatrix} = \begin{bmatrix} A_{11}^T & A_{12}^T \\ A_{21}^T & A_{22}^T \end{bmatrix} \begin{bmatrix} \Sigma_1 \\ 0 \end{bmatrix} \]
\[ = \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} + \begin{bmatrix} B_1 & B_1^T \\ B_2 & B_2^T \end{bmatrix} = 0 \]
\[ A_{11}\Sigma_1 A_{12} + A_{12}\Sigma_2 A_{12} = 0 \]
\[ A_{11}\Sigma_1 A_{12} + A_{12}\Sigma_2 A_{12} + B_1 B_2^T = 0 \]
\[ A_{21}\Sigma_1 A_{12} + A_{12}\Sigma_2 A_{12} + B_1 B_2^T = 0 \]
\[ A_{21}\Sigma_1 A_{12} + A_{12}\Sigma_2 A_{12} + B_1 B_2^T = 0 \]
\[ A_{21}\Sigma_1 A_{12} + A_{12}\Sigma_2 A_{12} + B_1 B_2^T = 0 \]

Define
\[ \Psi = (I - A_{22})^{-1} \]
Using eqns. 48, 49, 50 and 51
\[ A^T \Sigma + A^T = \Sigma + B^T B^T \]
\[ = (A_1 + A_2 \Sigma A_1) \Sigma (A_1^T + A_2 \Sigma A_1^T) - \Sigma + (B_1 + A_2 \Sigma B_2)^T (B_1^T + A_2 \Sigma B_2^T) \]
\[ = A_{11} \Sigma A_{11} + A_{12} \Sigma A_{12} + \Sigma A_{12}^T A_{12} + A_1 \Sigma A_1^T - \Sigma + B_1 B_1^T \]
\[ + B_2 B_2^T A_{12}^T + A_2 \Sigma B_2 B_2^T + A_2 \Sigma B_2 B_2^T A_{12}^T + A_2 \Sigma B_2 B_2^T \]
\[ = -A_{12} \Sigma A_{12}^T - A_{12} \Sigma A_{12}^T A_{12}^T \]
\[ + A_2 \Sigma A_2 \Sigma A_{22} - A_2 \Sigma A_2 \Sigma A_{22}^T \]
\[ + A_2 \Sigma \Sigma_{22} - A_2 \Sigma A_2 \Sigma A_{22}^T \]
\[ = A_{12} (-\Sigma - \Sigma A_{22}^T \Sigma - \Sigma A_{22} \Sigma + \Sigma A_{22} \Sigma) \]
\[ + (\Sigma \Sigma_{22} - \Sigma A_2 \Sigma A_{22}^T \Sigma A_{22}^T) \]
\[ = A_{12} [(-\Sigma - \Sigma A_{22}^T \Sigma - \Sigma A_{22} \Sigma + \Sigma A_{22} \Sigma)] \]
\[ + (\Sigma \Sigma_{22} - \Sigma A_2 \Sigma A_{22}^T \Sigma A_{22}^T) \]
\[ = 0 \]

Similarly
\[ A^T \Sigma A^T = \Sigma + C^T C = 0 \]

Thus the order reduced filter is balanced.

(b) This part will be proven by a contradiction method. Let \( \Sigma_1 \) and \( \Sigma_2 \) have no diagonal elements in common. Since eqn. 39a holds, all the eigenvalues of \( A^T \) are on or inside the unit circle. Suppose \( A^T \) is not asymptotically stable, then it follows that \( A^T \) has an eigenvalue on the unit circle, say at \( z = e^{\text{th}} \). Let \( R \) be a basis matrix for the right null space of \( (A^T - e^{\text{th}}) \).

\[ A^T \Sigma A^T = \Sigma + C^T C = 0 \]

Multiplying eqn. 39b from the right by \( R \) and from the left by \( R^T \) and using eqns. 52 and 53 gives
\[ C^T R = 0 \]

Multiplying eqn. 39b from the right by \( R \) and using eqns. 52 and 54 gives
\[ A^T \Sigma_1 R = e^{-\text{th}} \Sigma_1 R \]

\[ (R^T \Sigma_1) A^T = e^{\text{th}} (R^T \Sigma_1) \]

Analogously multiplying eqn. 39a from the right by \( \Sigma_1 R \) and from the left by \( R^T \Sigma_1 \) and using eqns. 55 and 56 gives
\[ B^T \Sigma_1 R = 0 \]

Then multiplying eqn. 39a from the right by \( \Sigma_1 R \) and using eqns. 57 and 55 to get
\[ (A^T - e^{\text{th}}) \Sigma_1 R = 0 \]

i.e., columns of \( \Sigma_1 R \) are in the right null space of \( (A^T - e^{\text{th}})D \).

There therefore exists a matrix \( \Sigma_1 \) such that
\[ \Sigma_1 R = \Sigma_1 R \]

and it is possible to choose \( R \) such that \( \Sigma_1 R \) is diagonal with diagonal elements that are subsets of the diagonal elements of \( \Sigma_1 R \).

\[ \text{From eqns. 54 and 57} \]
\[ C_1 R = -C_2 \Sigma A_2 R \]
\[ B_2^T R = -B_2^T \Sigma A_2^T \Sigma_1 R \]

\[ \text{From eqn. 146} \]
\[ A_{12}^T \Sigma_1 A_{12} + A_{12}^T \Sigma_1 A_{12} - A_{12}^T \Sigma_1 A_{12} = 0 \]
\[ C_1 R + C_2^T C_2 = 0 \]

Taking the complex conjugate transpose of eqn. 63, and multiplying from the right by \( R \)
\[ A_{12}^T \Sigma_1 A_{12}^T A_{12} + A_{12}^T \Sigma_1 A_{12}^T A_{12} - A_{12}^T \Sigma_1 A_{12}^T A_{12} = 0 \]
\[ C_1^T C_1 R = 0 \]

Substituting eqn. 60 into eqn. 64 gives
\[ A_{12}^T \Sigma_1 A_{12}^T A_{12} + A_{12}^T \Sigma_1 A_{12}^T A_{12} - C_2^T C_2 \Sigma A_2 R = 0 \]

Substituting eqn. 62 into eqn. 65 gives
\[ A_{12}^T \Sigma_1 A_{12}^T A_{12} + A_{12}^T \Sigma_1 A_{12}^T A_{12} = 0 \]

This implies
\[ A_{12}^T \Sigma_1 (A_1 R + A_{12} \Sigma A_2 R) \]
\[ + (A_{22}^T \Sigma_2 - A_2 \Sigma A_2^T) \Sigma A_2 R = 0 \]

\[ \text{From eqn. 52} \]
\[ A_{12} R + A_{12} \Sigma A_2 \Sigma A_2 R = e^{\text{th}} R \]

\[ \text{Substituting eqn. 68 into eqn. 67 gives} \]
\[ e^{\text{th}} A_{12}^T \Sigma_1 R + (A_{22}^T \Sigma_2 - A_2 \Sigma A_2^T) \Sigma A_2 R = 0 \]

This implies
\[ e^{\text{th}} A_{12}^T \Sigma_1 R - \Sigma A_2^T \Sigma A_2 R = 0 \]

Similarly from eqn. 14a
\[ A_{22}^T \Sigma_2 A_{22} + A_{22}^T \Sigma_2 A_{22} - A_{22}^T \Sigma_2 A_{22} = 0 \]

\[ A_{22}^T \Sigma_2 A_{22} + A_{22}^T \Sigma_2 A_{22} + B_2^T B_2 = 0 \]

Taking the complex conjugate transpose of eqn. 71, multiplying from the right by \( \Sigma_1 R \), using eqns. 69 and 70, simplifying and using eqn. 55 gives
\[ e^{-\text{th}} A_{12}^T \Sigma_1 R - \Sigma A_2^T \Sigma A_2^T R = 0 \]

\[ \text{From eqns. 69 and 72} \]
\[ e^{-\text{th}} A_{12}^T \Sigma_1 R = \Sigma A_2^T \Sigma A_2^T R \]

\[ \text{Therefore} \]
\[ \Sigma A_2 R = 0 \]

\[ \text{Substituting eqn. 59 into eqn. 73 gives} \]
\[ (R^T \Sigma_1 R)^2 = \Sigma_1 (R^T \Sigma_1 R) \]

\[ \text{If} \Sigma_1 \text{ and} \Sigma_2 \text{ have no diagonal elements in common, then} \Sigma_1 \text{ and} \Sigma_2 \text{ have no eigenvalues in common and eqn. 74 has a unique solution. Since it was assumed that} \Sigma_1 \text{ and} \Sigma_2 \text{ have no diagonal elements in common, then it follows that the unique solution of eqn. 74 is given by} \]
\[ \Sigma A_2 R = 0 \]

\[ \text{Since} \Sigma \text{ is nonsingular, it follows that} \]
\[ A_{12}^T \Sigma A_2 R = 0 \]

\[ \text{Substituting the above equation into eqn. 54 gives} \]
\[ e^{\text{th}} R = (A_{11} + A_{12} \Sigma A_2) R = A_{11} R \]

\[ \text{Therefore} \]
\[ \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} R \\ 0 \end{bmatrix} = \begin{bmatrix} A_{11} R \\ 0 \end{bmatrix} = e^{\text{th}} R \]

Thus the matrix \( A \) has an eigenvalue at \( e^{\text{th}} \) which contradicts that the system \( (A, B, C) \) is asymptotically stable as proved in Theorem 1. The reduced order filter \( (A', B', C', D) \) is therefore asymptotically stable.