Optimal Parallel 2-D FIR Digital Filter with Separable Terms

Vidya Venkatachalam and Jorge L. Aravena
Department of Electrical and Computer Engineering
Louisiana State University, Baton Rouge, LA 70810.
Phone: (504) 388-5537, fax: (504) 388-5200,
email: aravena@gate.ee.lsu.edu

April 30, 1998

EDICS CLASSIFICATION: SP 4.1.1 m-D Filtering
All correspondence related to this paper should be addressed to the second author

Abstract

This paper completely solves the optimal Weighted Least Mean Square (WLMS) design problem using sums of separable terms. For any fixed number of separable terms (less than or equal to the rank of the unconstrained solution), the problem is solved as a sequence of separable filter approximations. An efficient computational algorithm based on necessary conditions is presented. The procedure allows a high degree of flexibility in the choice of filter orders and the number of separable terms, but it may converge to a local minimum. An improved approximation can be obtained by computing more terms than required and then performing a truncation of the coefficient matrix using a singular value analysis. A significant computational advantage is that the procedure requires neither the solution of the unconstrained WLMS problem nor the singular value analysis of the ideal filter.

1 Introduction

The design of fast-acting 2-D FIR digital filters is a much researched area in Digital Signal Processing. Several authors (see for example [1], [20], [26], [29]) have proposed new algorithms to achieve good quality designs with reduced computational complexity. A *good quality design* is obtained by finding the *optimal* filter coefficients that satisfy a given constraint. *Reduced computational complexity* is obtained by eliminating redundant operations, making acceptable approximations and by putting to use the inherent symmetry properties in the desired filter response. Unfortunately, good quality and reduced complexity are, normally, conflicting in nature, and there is a trade-off between them in any standard design technique.

With the advances in VLSI technology and the advent of high speed processors which allow a high degree of parallelism, there is new interest ([3]-[8]) in digital filter design algorithms which readily lend themselves to a parallel architecture. Such algorithms provide for a fast implementation *without deterioration in filter quality* by allowing for several operations to be performed concurrently, thus reducing the trade-off inherent in the standard design techniques. This paper integrates quality of the filter and parallel implementation by *approximating a desired filter with sums of simpler and faster filters.* The 2-D filtering action is now accomplished by several pairs of 2-D separable filters, all acting *concurrently* on the image. The transfer function of the *k th* such separable FIR filter is then given by

$$H_k(z_1, z_2) = \sum_{n_1=-N_1'}^{N_1'} \sum_{n_2=-N_2'}^{N_2'} a_k(n_1) b_k(n_2) z_1^{-n_1} z_2^{-n_2} \tag{1}$$

[3],[4] give the details of one approach to design such filter pairs. This approach entails the use of the *Singular Value Decomposition* (SVD) of the desired response to find the optimal separable responses. These responses are then approximated by 1-D FIR filters, using standard design algorithms (like for example, the Remez algorithm ([30])). In a second stage, the SVD of the filter coefficient matrix is used to reduce the number of parallel channels. The main drawback of the SVD design procedure is that it involves an *approximation* by FIR filters, once the optimal 1-D frequency responses are determined. This leads to a final design which is suboptimal. There is, therefore, a need to formulate an algorithm to design truly optimal 1-D FIR filters and avoid the need for a double SVD.

This paper develops a parallel decomposition in terms of separable components by setting the filter design problem as a constrained Weighted Least Mean Square (WLMS) problem in the coefficients. Separability of each component is imposed by constraining the coefficients of 2-D filters to be rank one matrices. The technique is shown to be equivalent to an SVD with a different measure of orthogonality.

Notation can become very cumbersome and cloud some developments with unnecessary complexity. For this reason, the next section states the problem in the conventional way and then simplifies the notation by introducing an operator based notation which encompasses both discrete and continuous cases. This general formulation is solved in section three for the case of one separable filter and extended to a sum of separable filters. Section four presents design results. The final section contains conclusions.

2 Problem Statement

Consider a linear shift invariant 2-D filter with frequency response $D(\omega_1, \omega_2)$. This filter must be approximated by an FIR filter of the form

$$H(\omega_1, \omega_2) = \sum_{(k_1,k_2)\in I_z} h(k_1, k_2) e^{-j(k_1\omega_1 + k_2\omega_2)}$$

The merit index for the design is the cost function

$$J = \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} W|D - H|^2 d\omega_1 d\omega_2$$

$$= \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} W(\omega_1, \omega_2)|D(\omega_1, \omega_2) - \sum_{(k_1,k_2)\in I_z} h(k_1, k_2) e^{-j(k_1\omega_1 + k_2\omega_2)}|^2 d\omega_1 d\omega_2$$

The index set I_z is a subset of the integer numbers and normally is a rectangular region. In this case the collection $\{h(k_1, k_2)\}$ can be arranged as a matrix. The function $W(\omega_1, \omega_2)$ is a non negative function that can be used to assign more weight to performance in certain regions of the frequency domain.

The index is clearly a function of the FIR filter coefficients. Its minimization will determine the optimal coefficients. This is the standard WLMS problem. With the obvious modifications, one can set the problem in the discrete frequency domain, or extend it to m-D filters. For simplicity, the presentation is concentrated on the 2-D case, with appropriate extension to the general m-D case. If one wishes to restrict the minimization to the class of separable filters, then the coefficient matrix must be of rank one.

Remark 2.1 *It is well known that the general formulation can be simplified, from a computational point of view, if one makes use of symmetry conditions [25], [26]. However, as long as the filter is linear in the coefficients and the cost function is quadratic, one can always manipulate the design problem to the general formulation discussed below.*

2.1 The General Formulation

We offer here a formulation of the WLMS problem which applies to both discrete and continuous frequency cases and general m-D filters. The main goals are to reduce notational complexity, to highlight the common aspects of the problem, and establish conditions applicable to all cases.

The filter to be designed is an element of a 'filter space', $\mathbf{F}$, which is required to be a Hilbert space. For the continuous m-D case this space is $L^2[-\pi, \pi]^m$ while for the discrete case it will be a Euclidian space with dimension depending on the number of frequency points.

The filter is determined by a set of coefficients which are elements of a 'coefficient space', $\mathbf{C}$, which will also be a Hilbert space. If the coefficients are required to be real, this space is a Euclidian space, otherwise it can be taken as a space of complex numbers. If $c \in \mathbf{C}$ is a set of coefficients then a feasible filter solution can be represented as $H = \mathcal{F}(c)$, where $\mathcal{F} : \mathbf{C} \to \mathbf{F}$ is a given map describing the filter in terms of its coefficients. For FIR filters, this map is linear.

Remark 2.2 *In the unconstrained 2-D case , the parameters are normally arranged in a matrix, C. Since $N_1 \times N_2$ matrices can also be considered elements of a Hilbert space, $E^{N_1 \times N_2}$, with inner product $< A, B >= tr\{A^* B\}, A, B \in E^{N_1 \times N_2}$; one can formulate the problem directly in terms of the matrix. For this, one can use the 'stacking' isometry $\mathcal{S} : E^{N_1 \times N_2} \to E^{N_1 N_2}$. If $C \in E^{N_1 \times N_2}$ is an $N_1 \times N_2$ matrix, the vector $c = \mathcal{S}(C) \in E^{N_1 N_2}$ is obtained by stacking the columns of C following a left to right order. It follows easily that the 'de-stacking operator' is actually the adjoint $\mathcal{S}^* : E^{N_1 N_2} \to E^{N_1 \times N_2}$ and that $\mathcal{S}^* \mathcal{S}$ is the identity transformation. For example, if the filter is of the form*

$$H(z_1, z_2) = \sum_{n_1=-N_1'}^{N_1'} \sum_{n_2=-N_2'}^{N_2'} h(n_1, n_2) z_1^{-n_1} z_2^{-n_2}, \tag{2}$$

(where $N_1 = 2N_1' + 1$, $N_2 = 2N_2' + 1$),
one can define the matrix of coefficients

$$C = [h(k_1, k_2)]; \qquad -N_i' \leq k_i \leq N_i'; \qquad i = 1, 2. \tag{3}$$

The definition of the operator $\mathcal{FS}$ is then

$$\mathcal{FS}(C) = \sum_{n_1=-N_1'}^{N_1'} \sum_{n_2=-N_2'}^{N_2'} h(n_1, n_2) e^{j(n_1 \omega_1 + n_2 \omega_2)}$$

Clearly, the same type of representation can be established for the general $m - D$ case.

The ideal filter is an element $D \in \mathbf{F}$, while the WLMS cost function is a weighted distance in $\mathbf{F}$ and can be written in the form

$$J(c) = \langle \mathcal{W}(D - \mathcal{F}(c)), D - \mathcal{F}(c) \rangle_{\mathbf{F}}$$

For all cases of practical interest, the map, $\mathcal{W}$, describing the weighting function can be assumed to be self-adjoint and positive semi-definite. The optimal WLMS design consists in determining the coefficient $\hat{c} \in \mathbf{C}$ which minimizes the cost function $J(c)$.

Using conventional properties one can write

$$J(c) = \langle \mathcal{W}(D), D \rangle_{\mathbf{F}} - \langle \mathcal{F}^* \mathcal{W}(D), c \rangle_{\mathbf{C}} - \langle c, \mathcal{F}^* \mathcal{W}(D) \rangle_{\mathbf{C}} + \langle c, \mathcal{F}^* \mathcal{W} \mathcal{F}(c) \rangle_{\mathbf{C}}$$

where the notation $(\cdot)^*$ denotes the adjoint of the corresponding operator.

Remark 2.3 *Since one must work with different spaces, one should use different symbols to denote inner products and norms in the various spaces. Thus* $\langle \cdot, \cdot \rangle_{\mathbf{C}}$ *denotes inner product in the coefficient space. For the sake of simplicity, in the rest of the developments, distinctions are not made when the underlying spaces are clear from context.*

One can use standard variational techniques and derive necessary and sufficient conditions for optimality. Specifically one has the standard result

Theorem 2.4 *The parameter* $\hat{c} \in \mathbf{C}$ *is an* unconstrained *solution to the WLMS problem if and only if it satisfies*

$$\mathcal{F}^*\mathcal{W}\mathcal{F}(\hat{c}) = \mathcal{F}^*\mathcal{W}(D)$$

Remark 2.5 *It is also a standard result that the cost function can be written in terms of an optimal solution as*

$$J(c) = \langle \mathcal{W}(D), D\rangle - \langle \hat{c}, \mathcal{F}^*\mathcal{W}\mathcal{F}\hat{c}\rangle + \langle \hat{c} - c, \mathcal{F}^*\mathcal{W}\mathcal{F}(\hat{c} - c)\rangle \tag{4}$$

Hence, minimization of $J(c)$ *is equivalent to the minimization of*

$$J_e(c) = \langle \hat{c} - c, \mathcal{F}^*\mathcal{W}\mathcal{F}(\hat{c} - c)\rangle \tag{5}$$

This expression will be useful in developing a better understanding of the solutions.

Clearly, the unconstrained solution to the WLMS problem will be unique if the operator $\mathcal{F}^*\mathcal{W}\mathcal{F}$ is positive definite. Notice that the map $\mathcal{F}^*\mathcal{W}\mathcal{F}$ is a linear transformation in the parameter space. For the FIR case, even for the general $m - D$ case, this is a finite dimensional space; the cost function becomes a simple quadratic function in the coefficients; and the solution could, in theory, be obtained using matrix inversion techniques. Particularly for m-D filters, this is not a practical approach and researchers have developed many different approaches (see for example [25, 26]).

2.1.1 The separable $2 - D$ case

The constraint that the filter be a single separable term is easily stated in terms of the matrix of coefficients, C. One must have

$$C = ab^T; \qquad a \in E^{N_1}, b \in E^{N_2}$$

It is then clear that the vector of coefficients obtained by stacking the columns of C is nothing more than the Kronecker product $a \otimes b \in E^{N_1 N_2}$ (i.e., $a \otimes b = \mathcal{S}(ab^T)$). The vectors $a \in E^{N_1}, b \in E^{N_2}$ are unconstrained. The cost function can be put in the form

$$J(a, b) = \langle \mathcal{W}(D), D\rangle - \langle \mathcal{F}^*\mathcal{W}(D), a \otimes b\rangle - \langle a \otimes b, \mathcal{F}^*\mathcal{W}(D)\rangle + \langle a \otimes b, \mathcal{F}^*\mathcal{W}\mathcal{F}(a \otimes b)\rangle \tag{6}$$

The optimal design of the one-term separable filter corresponds to the minimization of the cost function in eq.(6) with respect to the unconstrained parameters a, b.

Remark 2.6 *The m-D separable case follows exactly the same model. The filter coefficients will also be arranged in a rank one matrix, which now will be of the form* $C = a_1 \otimes a_2 \otimes \ldots \otimes a_m$. *Each vector* a_k *contains the coefficients of the* $k-th$ *1-D filter. Their dimensions are determined by the required order in that filter.*

In order to gain insight into the minimization, one can reformulate the problem in terms of the unconstrained solution. Following remark 2.5, the cost function in Eq. 6 can be rewritten as

$$J(a,b) = \langle \mathcal{W}(D), D \rangle - \langle \hat{c}, \mathcal{F}^*\mathcal{W}\mathcal{F}\hat{c} \rangle + \langle \hat{c} - a \otimes b, \mathcal{F}^*\mathcal{W}\mathcal{F}(\hat{c} - a \otimes b) \rangle \tag{7}$$

In a similar manner, the constraint that the filter must consist of k parallel, separable terms, requires a matrix of coefficients

$$C_k = \sum_{i=1}^{k} a_i b_i^T; \qquad a \in E^{N_1}, b \in E^{N_2}$$

In terms of the vectors obtained by stacking columns, a $k-$ terms approximation is a vector

$$c_k = \sum_{i=1}^{k} a_i \otimes b_i; \qquad a_i \in E^{N_1};\ b_i \in E^{N_2}$$

where the vectors $a_i \otimes b_i;\ i = 1, 2, \ldots, k$, form a linearly independent set. It should be clear that the number of separable terms must be at most equal to the rank of the unconstrained solution, and must satisfy the constraint $k \leq min\{N_1, N_2\}$. The cost in this case has the form

$$J(c_k) = \langle \mathcal{W}(D), D \rangle - \langle \hat{c}, \mathcal{F}^*\mathcal{W}\mathcal{F}\hat{c} \rangle + \left\langle \hat{c} - \sum_{i=1}^{k} a_i \otimes b_i, \mathcal{F}^*\mathcal{W}\mathcal{F}(\hat{c} - \sum_{i=1}^{k} a_i \otimes b_i) \right\rangle \tag{8}$$

Remark 2.7 *Finding the best k terms separable approximation is then equivalent to the approximation of the* unconstrained *optimal,* $\hat{c}$ *using a weighted inner product. In particular, if* $\mathcal{F}^*\mathcal{W}\mathcal{F}$ *is the identity operator, one must solve the minimization problem*

$$J(a_1, b_1, \ldots, a_k, b_k) = \| \hat{c} - \sum_{i=1}^{k} a_i \otimes b_i \|^2$$

Using the 'de-stacking' operators, the cost can be rewritten in terms of matrices as

$$J(a_1, b_1, \ldots, a_k, b_k) = \| \hat{C} - \sum_{i=1}^{k} a_i b_i^T \|^2$$

The solution of this problem is known and can be expressed in terms of the k largest singular values, and corresponding singular vectors, for $\hat{C}$. *However, for an arbitrary matrix* $\mathcal{F}^*\mathcal{W}\mathcal{F}$, *the conventional SVD will not, in general, yield an optimal k terms representation.*

3 The Optimal Filter with Separable Components

This section contains the main theoretical results, establishing the existence of optimal approximations with a specified number of separable terms. The development examines the case of one term and then uses those results to establish the general result.

3.1 Optimal Separable Filter

In the context of the present development, this case is referred to as the *one term separable filter.* In fact it solves the WLMS problem with the additional constraint that the solution must be a separable filter.

For the one term separable filter, the cost function is given by eq. (6). In order to determine an optimal solution, one can use the identity

$$a \otimes b = (a \otimes I_{N_2})b$$

Replacing in the expression for $J(a,b)$ (eq. (6)), and using the property, $< x, Ay > = < A^*x, y >$, one obtains

$$\begin{aligned} J(a,b) = \; & \langle \mathcal{W}(D), D\rangle - \langle (a^* \otimes I_{N_2})\mathcal{F}^*\mathcal{W}(D), b\rangle - \\ & \langle b, (a^* \otimes I_{N_2})\mathcal{F}^*\mathcal{W}(D)\rangle + \langle b, (a^* \otimes I_{N_2})\mathcal{F}^*\mathcal{W}\mathcal{F}(a \otimes I_{N_2})b\rangle \end{aligned} \tag{9}$$

For fixed $a \in E^{N_1}$, the previous equation is a conventional quadratic cost problem in the vector $b \in E^{N_2}$. The problem will have a unique solution $\hat{b}(a)$ if and only if the matrix $(a^* \otimes I_{N_2})\mathcal{F}^*\mathcal{W}\mathcal{F}(a \otimes I_{N_2})$ is positive definite. The following result shows that this is indeed the case whenever the unconstrained problem has a unique solution.

Lemma 3.1 *For any non zero vector $a \in E^{N_1}$, the matrix $(a^* \otimes I_{N_2})\mathcal{F}^*\mathcal{W}\mathcal{F}(a \otimes I_{N_2})$ is positive definite if and only if $\mathcal{F}^*\mathcal{W}\mathcal{F}$ is positive definite.*

The proof is immediate because $\langle b, (a^* \otimes I_{N_2})\mathcal{F}^*\mathcal{W}\mathcal{F}(a \otimes I_{N_2})b\rangle = \langle a \otimes b, \mathcal{F}^*\mathcal{W}\mathcal{F}(a \otimes b)\rangle$.

The unique solution is

$$\hat{b}(a) = [(a^* \otimes I_{N_2})\mathcal{F}^*\mathcal{W}\mathcal{F}(a \otimes I_{N_2}]^{-1}(a^* \otimes I_{N_2})\mathcal{F}^*\mathcal{W}(D) \tag{10}$$

This expression for b can be replaced in the cost function defining

$$\begin{aligned} J_b(a) &= J(a, \hat{b}(a)) \\ &= \langle \mathcal{W}(D), D\rangle - \langle \mathcal{F}^*\mathcal{W}(D), (a \otimes I_{N_2})[(a^* \otimes I_{N_2})\mathcal{F}^*\mathcal{W}\mathcal{F}(a \otimes I_{N_2})]^{-1}(a^* \otimes I_{N_2})\mathcal{F}^*\mathcal{W}(D)\rangle \end{aligned} \tag{11}$$

It is immediately apparent that this cost function is independent of the magnitude of the vector a; hence one can restrict its minimization to the unit ball, $B_a = \{a \in E^{N_1} : \| a \| = 1\}$.

Since the unit ball, B_a, is compact, the existence of a global minimum can be established by showing that $J_b(a)$ is a continuous function on B_a. For this purpose, one can use the following steps

1. If the unconstrained WLMS problem has a unique solution, then $\mathcal{F}^*\mathcal{W}\mathcal{F} > 0$. Hence

   (a)
   $$\mathcal{F}^*\mathcal{W}\mathcal{F} = \Lambda^2$$

   (b)
   $$\lambda_{min}^2 \parallel p \parallel^2 \leq \langle \Lambda p, \Lambda p \rangle \leq \lambda_{max}^2 \parallel p \parallel^2;\ \forall p \in E^{N_1 N_2} \tag{12}$$

2. Define $X(a) = \Lambda a \otimes I_{N_2}$, $Q(a) = X^*(a)X(a)$. Using Eq(12) establish

   (a)
   $$\lambda_{min} \parallel a \parallel \leq \parallel X(a) \parallel \leq \lambda_{max} \parallel a \parallel$$

   (b)
   $$\lambda_{min}^2 \parallel b \parallel^2 \leq \langle b, Q(a)b \rangle \leq \lambda_{max}^2 \parallel b \parallel^2;\ \forall b \in E^{N_2}; a \in B_a$$

   (c)
   $$\lambda_{max}^{-2} \parallel b \parallel^2 \leq \left\langle b, Q^{-1}(a)b \right\rangle \leq \lambda_{min}^{-2} \parallel b \parallel^2;\ \forall b \in E^{N_2}; a \in B_a$$

   (d) If $a_1, a_2 \in B_a$, and $\delta a = a_2 - a_1$ then,
   $$\parallel \delta a \parallel^2 \leq 2 \parallel \delta a \parallel;$$
   $$Q(a_2) - Q(a_1) = X^*(a_1)X(\delta a) + X^*(\delta a)X(a_1) + X^*(\delta a)X(\delta a)$$
   and
   $$\parallel Q(a_2) - Q(a_1) \parallel \leq 4\lambda_{max}^2 \parallel \delta a \parallel$$

3. Since $Q^{-1}(a_1) - Q^{-1}(a_2) = Q^{-1}(a_1)\left(Q(a_2) - Q(a_1)\right)Q^{-1}(a_2)$
   $$\parallel Q^{-1}(a_1) - Q^{-1}(a_2) \parallel \leq 4\lambda_{min}^{-4}\lambda_{max}^2 \parallel \delta a \parallel;\ \forall a_1, a_2 \in B_a$$

4. Since $\mathcal{F}\mathcal{W}(D) = \mathcal{F}^*\mathcal{W}\mathcal{F}(\hat{c})$, the cost function, $J_b(a)$, can be written as
   $$J_b(a) = \langle \mathcal{W}(D), D \rangle - \left\langle X^*(a)\Lambda(\hat{c}), Q^{-1}(a)X^*(a)\Lambda(\hat{c}) \right\rangle$$
   Therefore
   $$\begin{aligned} J_b(a_1) - J_b(a_2) = \; & \langle X^*(a_2)\Lambda(\hat{c}), (Q^{-1}(a_2) - Q^{-1}(a_1))X^*(a_2)\Lambda(\hat{c}) \rangle + \\ & \langle X^*(a_1)\Lambda(\hat{c}), Q^{-1}(a_1)X^*(\delta a)\Lambda(\hat{c}) \rangle + \\ & \langle X^*(\delta a)\Lambda(\hat{c}), Q^{-1}(a_1)X^*(a_2)\Lambda(\hat{c}) \rangle \end{aligned}$$

5. Taking the absolute value, one can see that every inner product in the right hand side can be bounded by $\parallel \delta a \parallel$. Hence the function is continuous.

Remark 3.2 *Notice that*

$$J_b(\hat{a}) \leq J(a, \hat{b}(a)) \leq J(a, b), \forall (a, b)$$

Hence this method indeed computes the globally optimal (one term) separable filter.

It is also clear that if for some collection of nonzero vectors $\{q_1, q_2, \ldots, q_k\}$*, one imposes the additional constraints*

$$\langle a, q_1 \rangle = 0; \; i = 1, 2. \ldots, k$$

the resulting domain is the intersection of the unit ball with a collection of subspaces. This is also a compact subset of the unit ball, and the constrained minimization will have a globally optimal solution. This result will be useful in establishing the existence of an optimal decomposition with a given number of terms.

Now that the existence of the optimal solution has been established, it is possible to develop necessary conditions which will be useful for the development of computationally efficient algorithms. For this, let $\hat{a}, \hat{b}$ be an optimal solution and a, b any other pair of vectors. Using simple algebraic manipulations, one can write

$$\begin{aligned} J(a,b) - J(\hat{a},\hat{b}) = \; & \left\langle a \otimes b - \hat{a} \otimes \hat{b}, \mathcal{F}^* \mathcal{W} \mathcal{F} (a \otimes b - \hat{a} \otimes \hat{b}) \right\rangle + \\ & \left\langle \hat{c} - \hat{a} \otimes \hat{b}, \mathcal{F}^* \mathcal{W} \mathcal{F} (a \otimes b - \hat{a} \otimes \hat{b}) \right\rangle + \left\langle a \otimes b - \hat{a} \otimes \hat{b}, \mathcal{F}^* \mathcal{W} \mathcal{F} (\hat{c} - \hat{a} \otimes \hat{b}) \right\rangle . \end{aligned}$$

By selecting suitable variations one can determine several useful necessary conditions.

Taking $a \otimes b - \hat{a} \otimes \hat{b} = \alpha \hat{a} \otimes \hat{b}$ one has

$$0 \leq \alpha^2 \left\langle \hat{a} \otimes \hat{b}, \mathcal{F}^* \mathcal{W} \mathcal{F} \hat{a} \otimes \hat{b} \right\rangle + \alpha \left\langle \hat{c} - \hat{a} \otimes \hat{b}, \mathcal{F}^* \mathbf{W} \mathcal{F} \hat{a} \otimes \hat{b} \right\rangle + \alpha \left\langle \hat{a} \otimes \hat{b}, \mathcal{F}^* \mathcal{W} \mathcal{F} (\hat{c} - \hat{a} \otimes \hat{b}) \right\rangle .$$

Using the conventional argument, for small values of α, the sign of the right hand side would be determined by the terms linear in α. If they are non zero, one could contradict the condition that $\hat{a}, \hat{b}$ are optimal. Hence one must have

$$\left\langle \hat{c} - \hat{a} \otimes \hat{b}, \mathcal{F}^* \mathcal{W} \mathcal{F} \hat{a} \otimes \hat{b} \right\rangle = 0$$

Taking now $a \otimes b = \hat{a} \otimes b$, and using the identity $a \otimes b = (a \otimes I_{N_2}) b$, one can write

$$\begin{aligned} J(a,b) - J(\hat{a},\hat{b}) = \; & \left\langle \hat{a} \otimes I_{N_2}(b - \hat{b}), \mathcal{F}^* \mathcal{W} \mathcal{F} (\hat{a} \otimes I_{N_2}(b - \hat{b}) \right\rangle + \\ & \left\langle \hat{c} - \hat{a} \otimes \hat{b}, \mathcal{F}^* \mathcal{W} \mathcal{F} (\hat{a} \otimes I_{N_2}(b - \hat{b}) \right\rangle + \left\langle \hat{a} \otimes I_{N_2}(b - \hat{b}), \mathcal{F}^* \mathcal{W} \mathcal{F} (\hat{c} - \hat{a} \otimes \hat{b}) \right\rangle . \end{aligned}$$

The vector $b - \hat{b}$ can be completely arbitrary in E^{N_2}. Repeating again the small variation arguments, one now can establish the condition

$$\hat{a}^* \otimes I_{N_2} \mathcal{F}^* \mathcal{W} \mathcal{F} (\hat{c} - \hat{a} \otimes \hat{b}) = 0$$

In a similar way, taking now $a \otimes b = a \otimes \hat{b}$ and noting that the vectors $a \otimes b$ and $b \otimes a$ are related by a simple permutation; i.e.,

$$a \otimes b = P_{\times} b \otimes a,$$

one can write a new necessary condition. These results are summarized in the following theorem

Theorem 3.3 *The optimal pair $\hat{a}, \hat{b}$ satisfies the necessary conditions*

1.
$$\left\langle \hat{c} - \hat{a} \otimes \hat{b}, \mathcal{F}^* \mathcal{W} \mathcal{F} \hat{a} \otimes \hat{b} \right\rangle = 0$$

2.
$$\hat{a}^* \otimes I_{N_2} \mathcal{F}^* \mathcal{W} \mathcal{F} (\hat{c} - \hat{a} \otimes \hat{b}) = 0$$

3.
$$\hat{b}^* \otimes I_{N_1} P_\times^* \mathcal{F}^* \mathcal{W} \mathcal{F} (\hat{c} - \hat{a} \otimes \hat{b}) = 0$$

Remark 3.4 *It is easy to see that the first necessary condition can be derived from any of the other two. For the sake of clarity, it has been kept separate since its shows the orthogonality characterisitic of all LMS solutions.*

This theorem will be used to establish a numerically simple computational procedure. More details will be presented in section 4, Development of a Computational Algorithm.

3.2 The Optimal Approximation with Several Separable Terms

According to remark 2.7, an optimal approximation in terms of k separable terms is equivalent to the determination of a singular value decomposition by using a weighted inner product to determine orthogonality. The following result makes this statement more clear.

Theorem 3.5 *Assume that the unconstrained optimal $\hat{c}$ can be written in the form*

$$\hat{c} = \sum_{i=1}^{m_0} \sigma_i \hat{a}_i \otimes \hat{b}_i$$

with $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_{m_0} > 0$.

Assume further that the terms are $\mathcal{F}^* \mathcal{W} \mathcal{F}$ conjugate; i.e.,

$$\left\langle \hat{a}_i \otimes \hat{b}_i, \mathcal{F}^* \mathcal{W} \mathcal{F} \hat{a}_j \otimes \hat{b}_j \right\rangle = 0;\ i \neq j \tag{13}$$

and are normalized so that

$$\left\langle \hat{a}_i \otimes \hat{b}_i, \mathcal{F}^* \mathcal{W} \mathcal{F} \hat{a}_j \otimes \hat{b}_j \right\rangle = 1;\ i = j \tag{14}$$

Then, for $1 \leq k \leq m_0$

$$\hat{c}_k = \sum_{i=1}^{k} \sigma_i \hat{a}_i \otimes \hat{b}_i$$

is the best k terms approximation, in the sense that any *other coefficient matrix $\tilde{C}$ of rank less than or equal to k yields a vector $\tilde{c} = \mathcal{S}(\tilde{C})$ such that $J(\tilde{c}) \geq J(\hat{c}_k)$.*

(Note: The normalization condition in Eq (14) is simply a convenience and can be easily removed.)

Proof: If the operator $\mathcal{F}^*\mathcal{W}\mathcal{F}$ is positive definite, the operation

$$\langle p_1, p_2 \rangle_{\mathcal{F}^*\mathcal{W}\mathcal{F}} = \langle p_1, \mathcal{F}^*\mathcal{W}\mathcal{F}p_2 \rangle, p_1, p_2 \in \mathbf{C}$$

defines a new inner product in the space $\mathbf{C}$ and consequently induces a new definition of orthogonality.

Any collection of nonzero vectors $p_1, p_2, \ldots, p_k$ such that $\langle p_i, \mathcal{F}^*\mathcal{W}\mathcal{F}p_j \rangle = 0;\ i \neq j$ are necessarily linearly independent, since they are orthogonal in the new inner product. In particular if $p_i = a_i \otimes b_i; \forall i$, then the matrix

$$C_k = \sum_{i=1}^{k} a_i b_i^T$$

must be exactly of rank k.

If the vectors $a_i \otimes b_i$ satisfy the normalization condition in equation (14), they form an orthonormal basis (in the new inner product) for the subspace

$$V_k = span\{a_i \otimes b_i;\ 1 \leq i \leq k\}$$

Moreover, the vector $\hat{c} - \sigma_i \hat{a}_i \otimes \hat{b}_i$ is clearly orthogonal (in the new inner product) to the vector $\sigma_i \hat{a}_i \otimes \hat{b}_i$. Hence, the vector $\hat{c}_k = \sum_{i=1}^{k} \sigma_i \hat{a}_i \otimes \hat{b}_i$ will be the orthogonal projection of $\hat{c}$ onto this subspace V_k.

Let now $\tilde{C}$ be a coefficient matrix, and $\tilde{c} = \mathcal{S}(\tilde{C})$ be the corresponding vector of coefficients. Assume that

$$\langle \hat{c} - \tilde{c}, \mathcal{F}^*\mathcal{W}\mathcal{F}(\hat{c} - \tilde{c}) \rangle < \langle \hat{c} - \hat{c}_k, \mathcal{F}^*\mathcal{W}\mathcal{F}(\hat{c} - \hat{c}_k) \rangle$$

Since $\hat{c}_k$ is the orthogonal projection of $\hat{c}$ onto V_k, the previous inequality implies that the vector $d = \tilde{c} - \hat{c}$ cannot belong to subspace V_k.

In terms of the coefficient matrices, one has

$$\tilde{C} = \sum_{i=1}^{k} \hat{a}_i \hat{b}_i^T + D$$

where the matrix D cannot be expressed as a linear combination of the rank one matrices $\hat{a}_i \hat{b}_i^T$. Hence, the rank of the matrix $\tilde{C}$ must be strictly larger than k. Therefore, $\hat{C}_k$ defines the best approximation with rank k. The theorem is established.

Since the cost function for the WLMS design (Eq. 4)

$$\begin{aligned} J(c) = &\quad \langle \mathcal{W}(D - \mathcal{F}c), D - \mathcal{F}c \rangle \\ = &\quad \langle \mathcal{W}(D), D \rangle - \langle \hat{c}, \mathcal{F}^*\mathcal{W}\mathcal{F}\hat{c} \rangle + \langle \hat{c} - c, \mathcal{F}^*\mathcal{W}\mathcal{F}(\hat{c} - c) \rangle \end{aligned}$$

one can easily derive

Corollary 3.6 *The first term in the decomposition, $\sigma_1 \hat{a}_1 \otimes \hat{b}_1$, is the optimal separable filter (determined in the previous section).*

The theorem is constructive and provides sufficient conditions for an optimal decomposition of the unconstrained solution of the WLMS problem. The following argument shows that one can always construct an optimal sequence. Hence, it is possible to establish a constructive technique to compute the optimal k terms approximation as a sequence of one term optimizations. Consider an approximation of the form

$$\begin{aligned} c_{k+1} &= \sum_{i=1}^{k+1} a_i \otimes b_i \\ &= \sum_{i=1}^{k} a_i \otimes b_i + a_{k+1} \otimes b_{k+1} \\ &= c_k + a_{k+1} \otimes b_{k+1} \end{aligned}$$

Assume that, $\hat{c}_k$, the optimal approximation with k terms is known (this is the case for $k = 1$). Assume further that the minimization problem

$$\min_{a_{k+1}, b_{k+1}} J(\hat{c}_k + a_{k+1} \otimes b_{k+1}) \tag{15}$$

with the constraints

$$\left\langle a_{k+1} \otimes b_{k+1}, \mathcal{F}^* \mathcal{W} \mathcal{F} \hat{a}_i \otimes \hat{b}_i \right\rangle = 0;\ 1 \leq i \leq k$$

admits a nonzero solution (see remark 3.2).

It is apparent that if the method is continued until the one term minimization does not permit any improvment, then one has actually constructed a sequence of terms that satisfies the conditions of the theorem (3.5) and has, therefore, computed the optimal solution. Moreover, the optimal solution can be determined sequentially with the one term constrained minimizations.

A severe limitation of many filter design tools is their computational complexity. The decomposition into a sequence of smaller problems has definite advantages. However, the one term minimization is a nonlinear programming problem which could still be considered computationally challenging. On the other hand, the results in the previous section show that the unconstrained one term solution *always* exists. Such a solution must yield a cost which cannot be larger than the constrained case. Hence, it is clear that the optimal solution coincides with an unconstrained one. Moreover, it must coincide with the global optimal.

This argument is attractive because it suggests that the optimization with k terms could be solved with a sequence of unconstrained one term minimizations. Its limitation lies in the fact that when using unconstrained minimizations, one does not insure the orthogonality conditions and may end up with suboptimal results. The next section explores this issue and develops an efficient algorithm for the one term minimization.

4 Development of a Computational Algorithm

The previous section establishes sufficient conditions for the existence of an optimal approximation with a specified number of separable terms; the solution can be obtained as a sequence of one term optimizations. This is a significant result; however, from a computational point of view, the one term minimization is still a complicated procedure. This section will develop an efficient algorithm for its solution, based on the necessary conditions in Theorem 3.3.

The equations of interest here are

$$\hat{a}^* \otimes I_{N_2} \mathcal{F}^* \mathcal{W} \mathcal{F}(\hat{c} - \hat{a} \otimes \hat{b}) = 0$$

$$\hat{b}^* \otimes I_{N_1} P_\times^* \mathcal{F}^* \mathcal{W} \mathcal{F}(\hat{c} - \hat{a} \otimes \hat{b}) = 0$$

The dependence on the unconstraineed solution $\hat{c}$ is eliminated by the use of the identity (see Theorem 2.4)

$$\mathcal{F}^* \mathcal{W} \mathcal{F}(\hat{c}) = \mathcal{F}^* \mathcal{W}(D)$$

Using also the identities $a \otimes b = (a \otimes I_{N_2})b, a \otimes b = P_\times b \otimes a$, one can write

$$(\hat{a}^* \otimes I_{N_2} \mathcal{F}^* \mathcal{W} \mathcal{F} \hat{a} \otimes I_{N_2})\, \hat{b} = \hat{a}^* \otimes I_{N_2} \mathcal{F}^* \mathcal{W}(D) \tag{16}$$

$$\left(\hat{b}^* \otimes I_{N_1} P_\times^* \mathcal{F}^* \mathcal{W} \mathcal{F} P_\times \hat{b} \otimes I_{N_1}\right) a = \hat{b}^* \otimes I_{N_1} P_\times^* \mathcal{F}^* \mathcal{W}(D) \tag{17}$$

The proposed algorithm uses the following steps

1. Select an arbitrary unit vector $a_0 \in E^{N_1}$
2. Given the unitary vector $a_n \in E^{N_1}$
   (a) Compute $\hat{b}(a_n)$ as the solution.of Eq (16), which is a linear equation of size N_2

   Remark 4.1 *An equivalent procedure is to minimize the cost function with respect to b for a fixed a_n. Since this is a quadratic cost function, a conjugate gradient guarantees convergence in at most N_2 steps.*

   (b) Given the vector $b_n = \hat{b}(a_n)$, compute $\hat{a}(b_n)$ by solving equation (17) (or by using a minimization procedure).

   (c) Define

$$a_{n+1} = \frac{\hat{a}(b_n)}{\| \hat{a}(b_n) \|}$$

   If $\quad \| a_n - a_{n+1} \| > tolerance$
   set $a_n := a_{n+1}$ and repeat iteration.
3. else
   stop
   end

It is clear that at every step, one is reducing the cost function. Moreover, the sequence of vectors a_n lies on the unit ball in E^{N_1}, which is a compact set, and consequently it must, at least, have a convergent subsequence.

In practice, numerous examples with discrete frequency response cases show that the algorithm converges very rapidly to a solution. Moreover, the algorithm appears to be insensitive to the selection of the starting point. However, as is common in nonlinear programming problems, there is no guarantee that it converges to the global optimum.

Remark 4.2 *The application to the discrete frequency case was analyzed in detail in [31]. It turns out that it is possible to characterize the cases where only real valued separable filters are necessary. For the cases where complex valued vectors are generated, one can constrain the formulation and force only real valued vectors. However, the experimental results showed that constraining the optimization produced slow convergence of the algorithm. It was also established that if a term $a \otimes b$ was a solution to the necessary conditions for a half plane symmetric filter, then the conjugate vector $a^c \otimes b^c$ was also a solution. For these filters, in cases where complex valued vectors were generated, the method forms a component filter with real coefficients using $\mathcal{F}(a \otimes b + a^c \otimes b^c)$.*

4.1 Numerical Examples

The operators $\mathcal{F}, \mathcal{W}$ have been explicitly evaluated in [31] for the case of discrete frequencies

$$\omega_{i,k_i} = \pi \frac{2k_i + 1}{2M_i}; \; k_i = -M_i, \ldots, M_i - 1; \; i = 1, 2$$

The ideal filter D is represented by an $(2M_1 + 1) \times (2M_2 + 1)$ matrix. The approximating FIR filter has the form

$$H = \Omega_1 a b^T \Omega_2^T$$

with

$$\Omega_i(k_1, k_2) = e^{-\jmath\pi \frac{(2k_1+1)}{2M_i} k_2}; \tag{18}$$

$$k_1 = -M_i, \ldots, M_i - 1, k_2 = -N_i', \ldots, N_i', i = 1, 2.$$

Hence

$$\mathcal{F}(a \otimes b) = \Omega_1 \otimes \Omega_2 a \otimes b = \Omega_1 a b^T \Omega_2^T$$

The weighting function is defined by an $(2M_1 + 1) \times (2M_2 + 1)$ matrix , W, with nonnegative entries and the operator $\mathcal{W}$ is defined as a Hadamard, or entry-by-entry matrix product, and denoted here by $\bullet$; i..e,

$$\mathcal{W}(D) = W \bullet D$$

The numerical examples included here divide the frequency range in 128 points (i.e., $M_1 = M_2 = 64$) and specify filter matrix coefficients of size $N_1' = N_2' = 22$. Thus the nonseparable case requires $2N_1'N_2' + N_1' + N_2' + 1 = 1013$ coefficients while each separable filter requires only $2N_1' + 2N_2' + 2 = 90$ coefficients. The cases shown below are:

1. A one quadrant fan filter (Figures 1, 2). This is a good example of an 'almost separable' ideal filter. In this case, the one term approximation yields a very good approximation. In fact, the figure shows one can obtain a very good quality response, comparable to the optimal non-separable response, using only very few (9%) coefficients. The maximum error between the exact filter and the one term approximation is less than 2%. In this case the algorithm required 180 iterations which is much smaller than the unconstrained number of parameters.

2. A filter whose support is a rotated ellipse. The ideal response and a computed approximation are shown in Figures 3 and 4 respectively. This filter has axes of $.7\pi$ and $.3\pi$ and an external transition band of width of $.1\pi$. It is rotated 30^o counterclockwise about the ω_1 axis. The filter is highly non-separable, but [31] established that the solutions to the necessary conditions are always real. One can obtain an approximation with a maximum error of 17% with a relatively small number of terms (11 in the case shown in Fig 4). The evolution of the cost function with the number of terms is also examined and shows a steady decrease in the error as the number of terms increases (see Figure 5). This last figure also shows the number of iterations required for convergence for each of the separable filters.

   Remark 4.3 *In order to interpret properly the* 17% *error, one must consider the fact that the unconstrained solution with the same weighting function also has a very high error. In fact, the corresponding 11 terms approximation derived using SVD analysis of the unconstrained case has a maximal error of* 21% *(see [31]).*

3. A filter with triangular support having axes of $.65\pi$ and $.55\pi$, with an internal transition band of width $.1\pi$. This is a half-plane symmetric filter, similar to the one quadrant fan filter, but is also highly non-separable similar to the rotated elliptical. Figure 6 displays the ideal filter and Figure 7 the approximation with 14 separable terms yielding a maximum error of 6.5%. The cost function and the number of iterations as functions of the number of terms are displayed in Figure 8.

   Remark 4.4 *Each of the cases shows that the cost function varies rapidly for the first few terms and then shows only marginal improvement for each additional term. This property suggests the concept of* critical number of terms *which appears to be related to the singular value structure of the weighted ideal filter.*

   Remark 4.5 *An examination of the data on convergence, shown in Figures 5 and 8, shows that, on the average, the algorithm converges in a number of iterations equal to the order of the separable filter. This speed is comparable to that of the best quadratic algorithms.*

4. The optimality of the approximations is also examined for the last two filters. The procedure is the following:
   First one computes an approximation with a number of terms greater than the *critical number*, n_c. The resulting matrix of coefficients is analyzed for its singular values and a new coefficient matrix is determined using the largest n_c singular values.
   The above procedure produces a remarkable reduction in the number of terms required for the filter with triangular support. As can be seen from the Figure 9, one can get a very good approximation with just the first 4 terms taken from the SVD decomposition of the coefficient matrix computed from the 14 terms originally used. An analysis of the singular values of the matrix shows that the remaining singular values are less than 10% of the highest singular value, and hence do not contribute much to the filter response. For the rotated elliptical filter however, almost no reduction is achievable using this procedure, even though the singular values of its coefficient matrix (computed from the original 11 terms), after 5 terms, are less than 10% of the maximum value. Reduction in the number of terms is not achieved in this case since the originally computed filter has a high error (21%).

   Remark 4.6 *The experimental results show that the algorithm does not compute an optimal approximation. In both cases, the approximation obtained by truncating with the SVD analysis presents superior filter characteristics and smaller error (see for example Figures 9 and 10). However,* the cost function for the one-term obtained using the algorithm is always lower than that for the one-term filter obtained from the SVD-based reduction. *This has been observed regardless of the number of terms computed prior to the SVD analysis. The conclusion is that the algorithm based on necessary conditions does converge to an optimal solution, but the algorithm builds up numerical errors as the number of terms is increased.*

   It is also interesting to point out that the separable filters obtained by SVD reduction yield a large cost but look smoother and have better appearance than the optimal one-term solution. This fact is a reflection of the acknowledged limitations of the mean square criterion.

5 Conclusions

The paper completely characterizes the optimal solution of the WLMS problem using separable terms. The characterization is supplemented with a fast numerical algorithm based on necessary conditions. For any fixed number of separable terms (less than or equal to the rank of the unconstrained solution), the problem is solvable as a sequence of separable filter approximations. Extensive numerical results indicate that the algorithm builds up errors as the number of terms increases. However, the technique permits a clear estimation of the number of terms required

for a good approximation to a given filter. An improved approximation can be obtained by computing a few more terms than required and then performing a truncation of the coefficient matrix using a singular value analysis. A significant computational advantage is that the procedure requires neither the solution of the unconstrained WLMS problem nor the singular value analysis of the ideal filter.
Some of the experimental results yield filters with poor characteristics. This is attributable to the known limitations of the LMS criterion. Better designs may be obtained by varying the weighting function, for example using Lawson's type updates ([25, 29]).

References

[1] M.O. Ahmad, and J.D. Wang, "An analytic least mean square solution to the design problem of two-dimensional FIR filters with quadrantally symmetric or antisymmetric frequency response," *IEEE Trans. Circuits and Systems*, Vol CAS-36, pp 968-979, July 1989.

[2] Wu-Sheng Lu, and Andreas Antoniou, "Two-Dimensional Digital Filters," Marcel Dekker,Inc., New York, 1992.

[3] A. Antoniou, and W.S. Lu, "Design of two-dimensional Digital Filters by using the Singular-Value Decomposition," *IEEE Trans. Circuits Syst.*, Vol. CAS-34, pp 1191-1198, Oct. 1987.

[4] W.S. Lu, H.P. Wang, and A. Antoniou, "Design of two-dimensional FIR Digital Filters by using the Singular-Value Decomposition," *IEEE Trans. Circuits and Systems*, Vol. 37, No. 1, pp 35-46, Jan. 1990.

[5] W.S. Lu, H.P. Wang, and A. Antoniou, "Design of two-dimensional Digital Filters using Singular-Value Decomposition and Balanced Approximation Method," *IEEE Trans. on Signal Processing*, Vol. 39, No. 10, pp 2253-2262, Oct. 1991.

[6] S. Treitel, and J.L. Shanks, "The design of multistage separable planar filters," *IEEE Trans. Geosci. Electron.*, Vol. GE-9, pp 10-27, Jan. 1971.

[7] P. Karivaratharajan, and M.N.S. Swamy, "Realization of a 2-dimensional FIR digital filter using separable filters," *Electron. Lett.*, Vol. 14, No. 8, pp 244-251, April 1978.

[8] B.R. Suresh, and B.A. Shenoi, "Exact realization of 2-dimensional digital filters by separable filters," *Electron. Lett.*, Vol. 12, No. 10, pp 242-244, 1976.

[9] W.S. Lu, and A. Antoniou, "Synthesis of 2-D state-space fixed point digital filter structures with minimum roundoff noise," *IEEE Trans. Circuits Syst.*, Vol. CAS-33, pp 965-973, Oct. 1986.

[10] T.S. Huang, J.W. Burnett, and A.G. Deczky, "The importance of phase in image processing filters, *IEEE Trans. Acoust., Speech, Signal Process.*, Vol. ASSP-23, pp 529-542, Dec. 1975.

[11] T.S. Huang, "Two-dimensional windows," *IEEE Trans. Audio Electroacoust.*, Vol. AU-20, pp 88-89, March 1972.

[12] A. Antoniou, and W.S. Lu, "Design of 2-D nonrecursive filters using window method," *IEE Proc.*, Vol. 137, pt. G, pp 247-250, Aug. 1990.

[13] T.C. Speake, and R.M. Mersereau, "A note on the use of windows for two-dimensional FIR filter design," *IEEE Trans. Acoust., Speech, Signal Process.*, Vol. ASSP-29, pp 125-127, Feb. 1981.

[14] J.F. Kaiser, "Nonrecursive digital filter design using the $I_0 - sinh$ window function," *Proc. 1974 IEEE Int. Symp. Circuit Theory*, pp 20-23.

[15] W. Chen, C.H. Smith, and S.C. Fralick, "A fast computational algorithm for the discrete cosine transform," *IEEE Trans. Commun.*, Vol. COM-25, pp 1004-1009, Sep. 1977.

[16] J. Makhoul, "A fast cosine transform in one and two dimensions," *IEEE Trans. Acoust., Speech, and Signal Process.*, Vol. ASSP-28, pp 27-34, Feb. 1980.

[17] A. Gupta, and K.R. Rao, "A fast Recursive Algorithm for the Discrete Sine Transform," *IEEE Trans. Acoust., Speech, Signal Process.*, Vol. ASSP-38, No. 3, pp 553, March 1990.

[18] A. Gupta, and K.R. Rao, "An Efficient Algorithm Based on the Discrete Sine Transform," *IEEE Trans. on Signal Processing*, Vol. 39, No. 2, pp 486, February 1991.

[19] J.H. McClellan, "The design of two-dimensional filters by transformations," *Proc. 7 th Annual Princeton Conf. Information Sciences and Systems*, pp 247-251, 1973.

[20] R.M. Mersereau, W.F.G. Mecklenbräuker, and T.F. Quatieri, Jr., "McClellan Transformations for two-dimensional digital filtering: I-Design," *IEEE Trans. Circuits Syst.*, Vol. CAS-23, pp 405-413, July 1976.

[21] Y. Kamp, and J.P. Thiran, "Chebyshev approximations for two-dimensional nonrecursive digital filters," *IEEE Trans. Circuits Syst.*, Vol. CAS-22, pp 208-218, March 1975.

[22] T.W. Parks, and J.H. McClellan, "Chebyshev approximation for nonrecursive digital filters with linear phase," *IEEE Trans. Circuit Theory*, Vol. CT-19, pp 189-194, March 1972.

[23] J.H. McClellan, and T.W. Parks, "Equiripple approximation of fan filters," *Geophysics*, Vol. 7, pp 573-583, 1972.

[24] L.R. Rabiner, J.H. McClellan, and T.W. Parks, "FIR digital filter design techniques using weighted Chebyshev approximation," *Proc. IEEE*, Vol. 63, pp 595-610, April 1975.

[25] G. Gu, J.L. Aravena, "Weighted Least Mean Square Design of Two-dimensional FIR Digital Filters," IEEE Trans. Sig. Proc., Vol SP-42, No 11, pp 3178-3187, Nov. 1994.

[26] J.L. Aravena and G. Gu," Weighted Least Mean Square Design of 2-D FIR Digital Filters: General Weighting Function," in Proc 27th Asilomar Conference on Signals, Systems and Computers, Asilomar, CA, November 1993, pp 711-715.

[27] C. Charalambous, "A unified review of optimization," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-22, pp 289-300, March 1974.

[28] C. Charalambous, "Acceleration of the least *p th* algorithm for minimax optimization with engineering applications," *Mathematical Programming*, Vol. 17, pp 270-297, 1979.

[29] C. Charalambous, "The performance of an algorithm for minimax design of two-dimensional linear phase FIR digital filters," *IEEE Trans. Circuits Syst*, Vol. CAS-32, pp 1016-1028, Oct. 1985.

[30] D.J. Shpak and A. Antoniou, "A generalized Remez method for the design of FIR digital filters," *IEEE Trans. Circuits Syst.*, Vol. CAS-37, pp 161-174, Feb. 1990.

[31] Vidya Venkatachalam, "Parallel and Separable 2-D FIR Digital Filter Design," *M.S. Thesis*, Louisiana State University, May 1995.

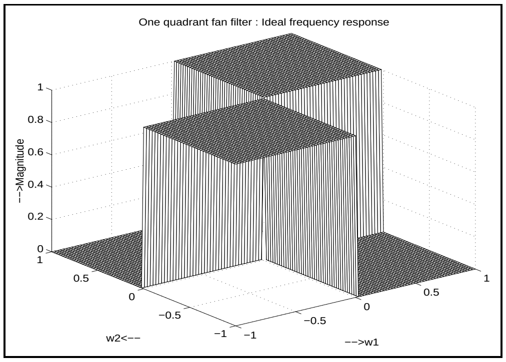

Figure 1: Ideal magnitude frequency response of 2-D quadrant fan filter

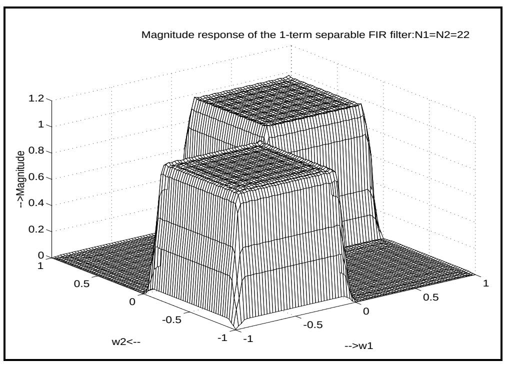

Figure 2: Magnitude frequency response of the optimal 1-term separable FIR quadrant fan filter

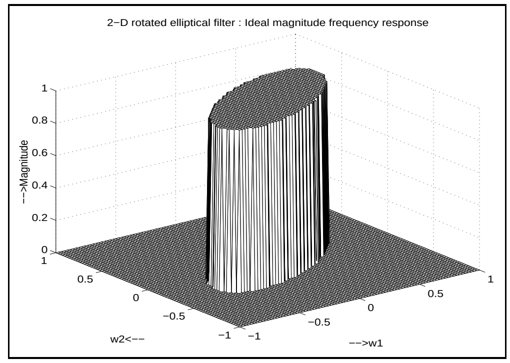

Figure 3: Ideal magnitude frequency response of 2-D rotated elliptical filter

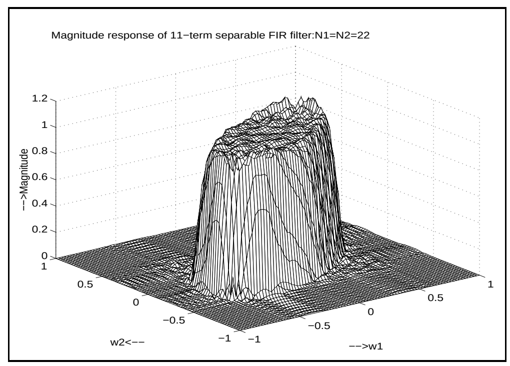

Figure 4: Magnitude frequency response of the optimal 11-term separable FIR rotated elliptical filter

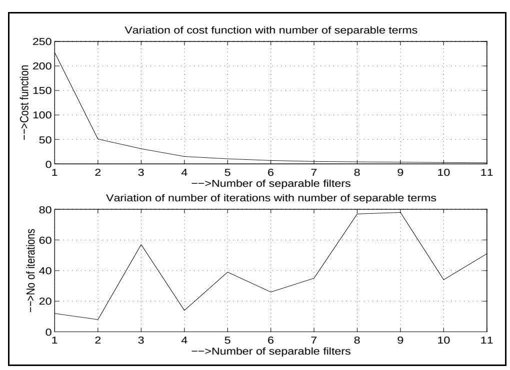

Figure 5: Performance variation with number of terms for the 11-term separable FIR rotated elliptical filter

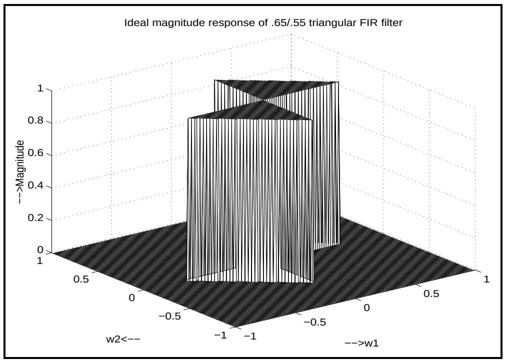

Figure 6: Ideal magnitude frequency response of 2-D triangular filter

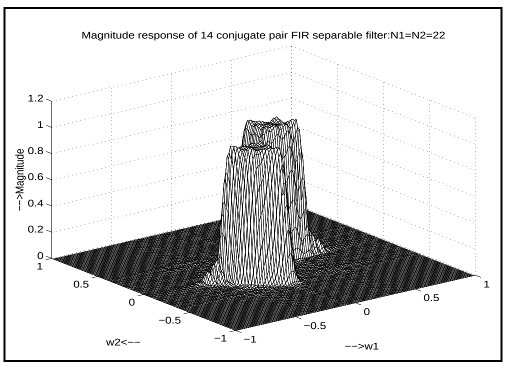

Figure 7: Magnitude frequency response of the optimal 14-term separable FIR triangular filter

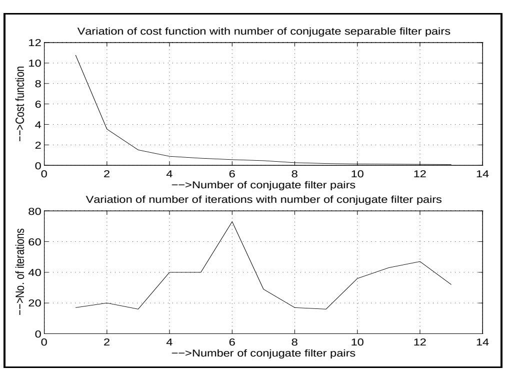

Figure 8: Performance variation with number of terms for the 14-term separable FIR triangular filter

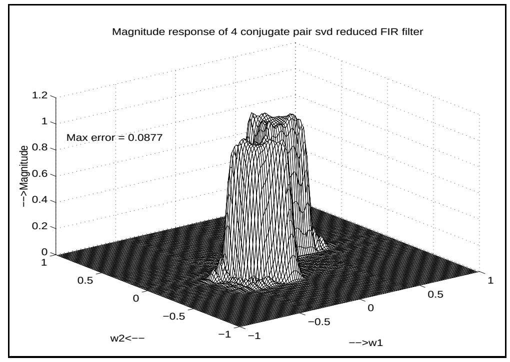

Figure 9: Magnitude frequency response of the 4-term svd reduced FIR triangular filter

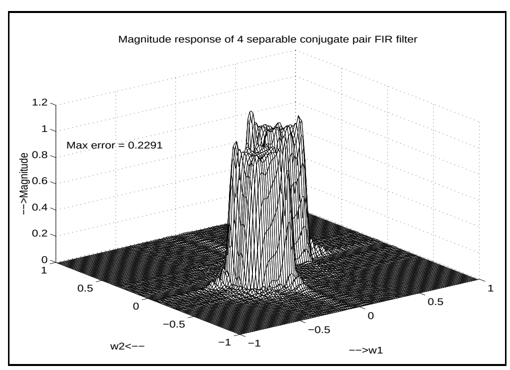

Figure 10: Magnitude frequency response of the optimal 4-term separable FIR triangular filter