Some Properties of Generalized Factorable 2-D FIR Filters

Roberto Manduchi

***Abstract*—Factorable M-dimensional filters are interesting because they can be implemented efficiently: their computational complexity is $\mathcal{O}(Mn)$ instead of $\mathcal{O}(n^M)$ (as in the case of generic non-factorable filters). Unfortunately, the passband support of a factorable filter can assume only very simple shapes (parallelepipeds with edges pairwise parallel to the axes), which are not adequate for most applications. In a recent paper, Chen and Vaidyanathan proposed a new class of non-factorable M-dimensional filters, whose passband support can be any parallelepiped, which can be realized with complexity $\mathcal{O}(Mn)$. In addition, they are designed starting from 1-D prototypes, which makes for a very simple design procedure. In this paper, we show that such filters belong to the class of generalized factorable (GF) filters (whose formal definition we introduce here), and derive some properties of theirs relative to the 2-D case. Our review includes issues such as the relation between minimax frequency response parameters and filter size (which is nontrivial in the multidimensional case), symmetries, 2-D step response, and frequency response constraints.**

I. Introduction

FACTORABLE filters represent a very appealing class of M-D filters— they require $\mathcal{O}(Mn)$ OPS's (operation per input sample), instead of $\mathcal{O}(n^M)$ OPS's as in the case of generic M-D FIR filters. However, factorability is too tight a constraint for most applications: a factorable filter's passband can only be in the shape of a parallelepiped with edges pairwise parallel to the axes.

Recently, Chen and Vaidyanathan [1]–[3] proposed a new class of non-factorable M-D filters which can be implemented with only $\mathcal{O}(Mn)$ OPS's, and whose passband support's shape can be any parallelepiped. The design procedure is quite simple and its computational burden is very light (since the filters are designed starting from M 1-D prototypes). Hence, they are particularly suitable for computer aided design (CAD) systems. The work of Chen and Vaidyanathan extends the previous results by Renfors [4] and by Cortelazzo *et al.* [5].

Chen and Vaidyanathan's filters actually belong to the class of generalized factorable (GF) filters, which we formally define in this paper. A GF filter is such that its polyphase components with respect to a given basis are the tensor product of M 1-D filters oriented along suitable directions. The design procedure by Chen and Vaidyanathan represents the only known algorithm to design GF filters. In the first part of the paper, we review the design procedure by Chen and Vaidyanathan, and derive some important characteristics of the "generalized factorable" implementation. In particular, we show that the number of OPS's required for the generalized factorable implementation of a GF filter defined on a lattice Λ, and of a GF filter with the same frequency response characteristics, defined on a sublattice Γ of Λ, may or may not differ, depending on the passband shape and on the mutual characteristics of the two lattices. Such a result is of extreme importance when GF filters are part of an M-D IFIR structure [6].

Beside possessing an efficient implementation, GF filters enjoy a number of interesting properties, which make them appealing for use in video technology. Hence, even if the efficient "generalized factorable" implementation is not used, GF filters represent a profitable choice for many applications. Some properties (preservation of the zero-phase and of the Nyquist property) have been already described in [3]. In this paper, we derive some further features of interest of GF filters, relative to the 2-D case, as listed in the following.

1) *Frequency Response Characterization:* We show that the minimax parameters of the frequency response of a GF filter are related to the filter order via a simple approximate formula, which is reminiscent of the well–known relation for the 1-D case [7], [8]:

$$N \propto \frac{-10\log(\delta_p\delta_s) - 13}{f_s - f_p} \tag{1}$$

where N is the length of an optimal minimax lowpass 1-D filter, f_p and f_s are its passband and stopband frequencies, and δ_p and δ_s are its passband and stopband ripples. Note that in the literature, similar relations have been described only for a class of circular symmetric filters [9] and for minimax diamond-shaped filters [10]. In the case of GF filters, we derive worst-case relations which hold for any choice of the parallelogram specifying the passband support.

2) *Symmetries:* 2-D GF filters designed starting from zero-phase 1-D filters may or may not satisfy a symmetry property similar to the quadrantal symmetry. We derive a simple condition to verify when such a property is satisfied. Symmetry may be used to reduce the number of multiplications per input sample if the factorable implementation is not employed.

Manuscript received February 6, 1995; revised November 23, 1995. This work was supported by a Fellowship by the Associazione Elettrotecnica Italiano (A.E.I.), Italy. The associate editor coordinating the review of this paper and approving it for publication was Dr. Kamal Premaratne.

The author was with the Department of EECS, University of California, Berkeley, CA. He is now with the Robotics Laboratory, Computer Science Department, Stanford University, Stanford, CA 94305 USA (e-mail: manduchi@cs.stanford.edu).

Publisher Item Identifier S 1053-587X(96)03957-8.

3) *2-D Step Response:* We give a simple characterization of the 2-D step response of GF filters, exploiting the fact that they are designed combining two 1-D filters.
4) *Frequency Response Constraints:* If the filter to be designed is part of a sampling structure converter, it is useful to impose some nulling constraints on its frequency response [11], [12]. We show how such constraints can be translated into simple constraints on the 1-D filters used in the design of the GF filter.

The paper is organized as follows. In Section II we report (together with the adopted nomenclature) a number of results of lattice theory. In particular, we introduce here the notions of the *least dense factorable lattice* containing some given lattice Λ, and of the *densest factorable sublattice* of Λ, which are instrumental in deriving a number of results of this paper. Also in Section II, we give the formal definition of certain useful 2-D filter parameters (passband and stopband curves, passband and stopband ripples), together with some basic properties.

In Section III, we review the procedure of Chen and Vaidyanathan to design GF filters. Some useful remarks to the original algorithm are pointed out, and the formal definition of GF filters is stated. In Section IV, the properties of 2-D GF filters listed above are derived. Section V has the conclusions.

II. Basic Results and Definitions

To develop our theory, it is necessary first to introduce some notions of lattice theory, reported in Section II-A. Section II-B contains the formal definition of some 2-D filter parameters, and derives some basic properties. Although such notions are used extensively by the multidimentional signal processing community, they are dispersed in the literature. In order to make the paper self-contained, we have gathered them here, since new results will be derived from them.

A. Lattice Theory Basics

In this section we report some notions of lattice theory that are used extensively throughout the paper, together with the adopted nomenclature. Section II-A1) contains facts already known in the literature, which we report here in order to make the paper self-contained. For their proofs, as well as for more details, the reader is addressed to [3], [12]-[16]. Section II-A2) reports some novel results.

1) Background and Nomenclature: R denotes the set of real numbers, and Z is the set of integers.

We denote vectors by lowercase boldface letters and matrices by capital boldface letters. Their entries are named after the following example:

$$\mathbf{a} \stackrel{\text{def}}{=} (a_1, a_2)^T \; ; \; \mathbf{A} \stackrel{\text{def}}{=} (\mathbf{a}_1|\mathbf{a}_2) \stackrel{\text{def}}{=} \begin{pmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{pmatrix}. \quad (2)$$

We use the following notation for a generic diagonal matrix $\mathbf{D}$:

$$\mathbf{D} = \text{diag}(\text{D}_1, \text{D}_2, \cdots). \quad (3)$$

Given two sets A and B, we denote their difference (i.e., the set of elements of A that do not belong to B) as $A \setminus B$.

We deal always with square full-rank matrices in this work. For the purpose of this section, we assume that the size of the considered matrices is fixed to M. Matrix $\mathbf{I}$ is the identity matrix.

Given a matrix $\mathbf{V} = (\mathbf{v}_1|\mathbf{v}_2|\cdots|\mathbf{v}_M)$, we define $SPD(\mathbf{V})$ the parallelepiped $\sum_{i=1}^M \alpha_i \mathbf{v}_i$, $-1 \le \alpha_i \le 1$. Given a point $\mathbf{u}$, we define $\mathcal{R}(\mathbf{u})$ the parallelepiped with edges parallel to the axes $\{\mathbf{a} : |a_i| \le |u_i|\}$, and $\underline{\mathcal{R}}(\mathbf{u})$ its boundary.

Given a rational number a, $\text{den}(a)$ denotes the least strictly positive integer such that $a \cdot \text{den}(a)$ is integer. Given a rational matrix $\mathbf{A}$, $\text{den}(\mathbf{A})$ denotes the least strictly positive integer such that $\mathbf{A} \cdot \text{den}(\mathbf{A})$ is an integral matrix. In other words, $\text{den}(\mathbf{A})$ is the least common multiple among $\{\text{den}(A_{i,j})\}$.

Any integral matrix $\mathbf{U}$ such that $\mathbf{U}^{-1}$ is still integral (or equivalently, such that $|\det(\mathbf{U})| = 1$) is called *unimodular*. Two integral matrices $\mathbf{A}_1, \mathbf{A}_2$ such that $\mathbf{A}_2^{-1}\mathbf{A}_1$ is unimodular are called *right-equivalent* or *right-associated*. For each class of right-associates, there is just one *Hermite normal form matrix*, i.e., a matrix $\mathbf{A}$ such that[1]

1) $\mathbf{A}$ is upper triangular,
2) $A_{i,j} \ge 0$,
3) $A_{i,j} < A_{i,i}$ for $1 \le i < j \le M$.

A lattice Λ that admits a basis $\mathbf{A}$ is denoted by $LAT(\mathbf{A})$. In other words, $\Lambda = LAT(\mathbf{A}) = \{\mathbf{A}\mathbf{n} : \mathbf{n} \in Z^M\}$. A unit cell of a lattice Λ is any region $\mathcal{C}$ such that

1) $\mathcal{C} + \mathbf{a}_i \cap \mathcal{C} + \mathbf{a}_j = \emptyset$ for any $\mathbf{a}_i, \mathbf{a}_j \in \Lambda, \mathbf{a}_i \ne \mathbf{a}_j$,
2) $\bigcup_{\mathbf{a}_i \in \Lambda} \mathcal{C} + \mathbf{a}_i = R^M$.

We denote a signal $h(\cdot)$ defined on Λ by $h(\mathbf{a})$ (where $\mathbf{a} \in \Lambda$) or by $h(\mathbf{A}\mathbf{n})$ (where $\mathbf{n} \in Z^M$). The two notations are interchanged liberally. We always use lowercase letters for signals, and the corresponding uppercase letters for their Fourier transforms.

Matrices $\mathbf{A}_1$ and $\mathbf{A}_2$ are bases of the same lattice if (and only if) $\mathbf{A}_2^{-1}\mathbf{A}_1$ is unimodular. When dealing with sampling lattices, we always assume that they are sublattices of Z^M (i.e., they are *integral* lattices, so that they admit only integral bases). Note that any result on integral lattices can be immediately extended to rational lattices (i.e., whose points have rational components). A lattice is said to be *factorable* (or *separable*) if it admits a diagonal basis.

Let $\Gamma = LAT(\mathbf{B})$ be a sublattice of $\Lambda = LAT(\mathbf{A})$. Then $\mathbf{H} = \mathbf{A}^{-1}\mathbf{B}$ is integral. Term $|\det(\mathbf{H})|$ is called the *index* of Γ in Λ (sometimes denoted by $(\Lambda : \Gamma)$ [14]), and it is the ratio between the density of Λ and the density of Γ.

A lattice Λ may be represented in terms of a sublattice Γ and of any set P of coset representatives for the cosets of Γ in Λ [14] (P is also called a Γ*-period of* Λ [17]):

$$\Lambda = \bigcup_{\mathbf{a}_i \in P} \Gamma + \mathbf{a}_i. \quad (4)$$

A Γ–period of Λ is formed by the elements of Λ contained in some unit cell of Γ.

Given a signal defined on Λ, its Γ*–polyphase components* (where Γ is a sublattice of Λ) are

$$h^{\mathbf{r}}(\mathbf{s}) \stackrel{\text{def}}{=} h(\mathbf{s} + \mathbf{r}) \; , \; \mathbf{s} \in \Gamma \; , \; \mathbf{r} \in P \quad (5)$$

[1] In general, the definition of Hermite normal form matrix may be extended to non-full-rank matrices by imposing the further condition: $A_{i,j} = 0$ if $A_{i,i} = 0$ [13].

where P is some Γ-period of Λ. The coset representative $\mathbf{r}$ in formula (5) is termed *polyphase index* of the polyphase component. Note that the Γ-polyphase components are defined on Γ, and that the definition (5) actually depends on the choice of P. Given P, there are $(\Lambda : \Gamma)$ distinct Γ–polyphase components. (In the 1-D case, given a signal $h(x)$, we will say that $h^r(x) = h(Nx + r), 0 \le r < N$, is the r-th N-polyphase component of $h(x)$).

Let n be an integer. Then the distinct sublattices having index n in $LAT(\mathbf{A})$ are $\{LAT(\mathbf{AH}_i)\}$, where $\{\mathbf{H}_i\}$ are the matrices in Hermite normal form with determinant equal to n [15].

We adopt the following definition for the Fourier transform of a signal $h(\mathbf{a})$ defined on a lattice $\Lambda = LAT(\mathbf{A})$:

$$H(\mathbf{f}) = \sum_{\mathbf{a} \in \Lambda} h(\mathbf{a}) e^{-j2\pi \mathbf{f}^T \mathbf{a}} = \sum_{\mathbf{n} \in Z^M} h(\mathbf{An}) e^{-j2\pi \mathbf{f}^T \mathbf{An}} \quad (6)$$

$H(\mathbf{f})$ is periodic on the *dual lattice* $\Lambda^* = LAT(\mathbf{A}^{-T})$, where $\mathbf{A}^{-T} \stackrel{\text{def}}{=} (\mathbf{A}^{-1})^T$.

The Fourier transform of the signal $h_s(\mathbf{a})$defined on the sublattice Γ of Λ, obtained subsampling the signal $h(\mathbf{a})$ defined on Λ as by

$$h_s(\mathbf{a}) \stackrel{\text{def}}{=} h(\mathbf{a}) \ , \ \mathbf{a} \in \Gamma \quad (7)$$

is

$$H_s(\mathbf{f}) = \frac{1}{(\Lambda : \Gamma)} \sum_{\mathbf{r} \in P} H(\mathbf{f} + \mathbf{r}) \quad (8)$$

where P is any Λ^*-period of Γ^*.

Note that we use the term "filter" meaning both the filter's impulse response (denoted by small letter) and the filter's frequency response (denoted by capital letter).

2) Novel Results: Consider lattice $\Lambda = LAT(\mathbf{A})$ with integral $\mathbf{A}$. As described in Section II-A-1, Λ admits a basis $\mathbf{A}^u$ in Hermite normal form. A geometric interpretation is the following one (see Fig. 1, where $\mathbf{A}^u = \begin{pmatrix} 5 & 3 \\ 0 & 1 \end{pmatrix}$): $(A^u_{1,1}, 0)^T$ is the point of Λ on the horizontal positive half-axis closest to the origin, while $(A^u_{1,2}, A^u_{2,2})^T$ is the point of Λ in the first quadrant to the left of $(A^u_{1,1}, 0)^T$, which has minimum distance to the horizontal axis. From this geometric standpoint, it is straightforward to argue that the same argument should apply interchanging the role of the horizontal and of the vertical axes. For example, from Fig. 1 we can find a "dual" basis $\mathbf{A}^l = \begin{pmatrix} 1 & 0 \\ 2 & 5 \end{pmatrix}$. Generalizing such an idea, we can infer that any integral lattice admits a basis $\mathbf{A}^l$ in *lower Hermite normal form*, i.e., such that

1) $\mathbf{A}^l$ is lower triangular,
2) $A^l_{i,j} \ge 0$,
3) $A^l_{i,j} < A^l_{i,i}$ for $1 \le j < i \le M$.

The Hermite normal form matrices as defined in Section II-A1) thus correspond to the *upper Hermite normal form* matrices. Note that, once one has got an algorithm to find the upper Hermite normal form matrix right-equivalent to a given matrix $\mathbf{A}$ (see for example [13]), it is trivial to modify the algorithm to obtain the lower Hermite normal form matrix right-equivalent to $\mathbf{A}$.

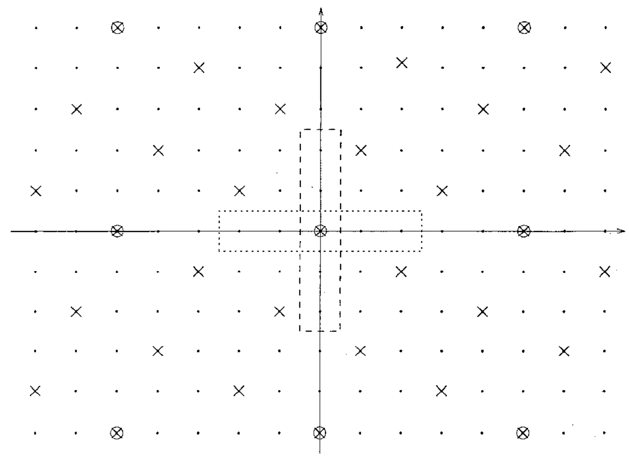


Fig. 1. Lattice Λ (crosses), the LDFL containing Λ (dots), the DFS of Λ (circles), and the two rectangular unit cells of Λ (dotted and dashed rectangles).

We consider only the 2-D case in the following. From our previous geometric arguments, one readily recognizes that, once the upper and the lower Hermite normal form bases $\mathbf{A}^u$ and $\mathbf{A}^l$ of Λ are known, finding a basis of the densest factorable sublattice (DFS) Γ of $\Lambda = LAT(\mathbf{A})$ (see Fig. 1) is straightforward:

$$\Gamma = LAT(\mathbf{S}) \ , \ \mathbf{S} = \text{diag}(S_1 = A^u_{1,1}, S_2 = A^l_{2,2}). \quad (9)$$

Since Γ is a sublattice of Λ, we have that $(\mathbf{A}^u)^{-1}\mathbf{S} = \mathbf{H}$ with integral $\mathbf{H}$. Then, $\mathbf{H} = \begin{pmatrix} 1 & k_1 \\ 0 & k_2 \end{pmatrix}$ with

$$A^l_{2,2} = k_2 A^u_{2,2} \ , \ -k_2 A^u_{1,2}/A^u_{1,1} = k_1 \ , \ \text{integer } k_1, k_2. \quad (10)$$

In particular, in order for Γ to be the densest factorable sublattice, k_2 should be the smallest integer for which (10) holds true, i.e.

$$k_2 = \text{den}(A^u_{1,2}/A^u_{1,1}). \quad (11)$$

Hence, $A^l_{2,2} = A^u_{2,2}\text{den}(A^u_{1,2}/A^u_{1,1})$. Then, since $\det(\mathbf{A}^u) = \det(\mathbf{A}^l)$, we have

$$A^u_{1,1} = A^l_{1,1}\text{den}(A^u_{1,2}/A^u_{1,1}). \quad (12)$$

This way, we are able to write $\mathbf{S}$ in terms of $\mathbf{A}^u$ only: $\mathbf{S} = (\text{diag} A^u_{1,1}, A^u_{2,2}\text{den}(A^u_{1,2}/A^u_{1,1}))$. In particular, the index of $\Gamma = LAT(\mathbf{S})$ in $\Lambda = LAT(\mathbf{A})$ is

$$(\Lambda : \Gamma) = \det(\mathbf{H}) = \det(\mathbf{S})/\det(\mathbf{A}^u) = \text{den}(A^u_{1,2}/A^u_{1,1}). \quad (13)$$

Another important notion is that of the least dense factorable lattice (LDFL) Ψ containing $\Lambda = LAT(\mathbf{A})$. Let $\mathbf{R} = \text{diag}(R_1, R_2)$ be a basis of Ψ. The entry R_i is the greatest common divisor of the entries of the ith row of matrix $\mathbf{A}$. We have that $\mathbf{RK} = \mathbf{A}^u$ for some integral matrix $\mathbf{K}$. It follows that $\mathbf{K} = \begin{pmatrix} n_1 & n_2 \\ 0 & n_3 \end{pmatrix}$, with

$$A^u_{1,1} = n_1 R_1 \ , \ A^u_{1,2} = n_2 R_1 \ , \ A^u_{2,2} = n_3 R_2 \quad (14)$$

for some integer n_1, n_2, n_3. Since Ψ is the least dense factorable lattice among those containing Λ, it must be (using (12))

$$R_1 = A^u_{1,1}/\mathrm{den}(A^u_{1,2}/A^u_{1,1}) = A^l_{1,1}\ ,\ R_2 = A^u_{2,2}. \quad (15)$$

In particular, the index of Λ in Ψ is again

$$(\Psi : \Lambda) = \det(\mathbf{K}) = \det(\mathbf{A}^u)/\det(\mathbf{R}) = \mathrm{den}(A^u_{1,2}/A^u_{1,1}). \quad (16)$$

Combining relations (9), (12), (13), and (15), we maintain that

$$(\Psi : \Lambda) = (\Lambda : \Gamma) = S_1/R_1 = S_2/R_2. \quad (17)$$

Making use of the upper and lower Hermite normal form bases $\mathbf{A}^u$ and $\mathbf{A}^l$, we can also determine the rectangular unit cells of Λ centered at the origin with sides pairwise parallel to the axes (throughout this paper, we refer to them simply by the term "rectangular"). One may prove, starting from our previous geometric observations, that a 2-D integral lattice admits only two rectangular unit cells, namely, $\mathcal{R}(A^u_{1,1}/2, A^u_{2,2}/2)$ and $\mathcal{R}(A^l_{1,1}/2, A^l_{2,2}/2)$ (see Fig. 1). In particular, from (12), we infer that the two rectangular unit cells coincide only if $\mathrm{den}(A^u_{1,2}/A^u_{1,1}) = 1$, i.e., if $A^u_{1,2} = 0$ (Λ is factorable).

B. Some Filter Parameter Definitions and Basic Results

This section contains some novel definitions and basic results that are extensively used throughout the paper. Only zero-phase filters (i.e., having purely real frequency response) are considered in the following. Unitary sampling period is assumed for the 1-D filters.

The typical minimax parameters of a 1-D FIR low-pass filter $h(n)$ are the passband and stopband frequencies f_p and $f_s > f_p$, the passband and stop–band ripples $\delta_p \geq 0$ and $\delta_s \geq 0$, and the filter length N. We consider here only low–pass filters with frequency response approximating a unitary step. Note that quantities f_p and δ_p (as well as f_s and δ_s) are related to one another as

$$\delta_p = \max_{\{0 \leq f \leq f_p\}} |1 - H(f)|\ ,\ \delta_s = \max_{\{f_s \leq f \leq 0.5\}} |H(f)| \quad (18)$$

where $H(f)$ is the frequency response of $h(n)$.

As a matter of fact, f_p, f_s and δ_p/δ_s (or other combinations of the parameters, see [18]) are fixed as *design* parameters. On the other side, when analyzing a given low-pass frequency response (not necessarily optimal), it can be useful to parametrize it in a similar fashion. For this purpose, we introduce here the following definitions of the *analysis* parameters of a given filter $h(n)$:

$$\delta_p = \max_{\{f_i : \frac{dH(f)}{df}|_{f=f_i}=0, |H(f_i)|>0.5\}} |1 - H(f_i)| \quad (19)$$

$$f_p = \min\{f_i : f_i > 0,\ H(f_i) = 1 - \delta_p, \frac{dH(f)}{df}|_{f=f_i} \neq 0\} \quad (20)$$

$$\delta_s = \max_{\{f_i : \frac{dH(f)}{df}|_{f=f_i}=0, |H(f_i)|<0.5\}} |H(f_i)| \quad (21)$$

$$f_s = \min\{f_i : f_i > 0,\ H(f_i) = \delta_s, \frac{dH(f)}{df}|_{f=f_i} \neq 0\}. \quad (22)$$

Note that the couples (f_p, δ_p) and (f_s, δ_s) identified by (20),(19) and (22), (21) respectively, satisfy relations (18), unless $\delta_p > 0.5$ or $\delta_s > 0.5$, in which case the low–pass filter is unsuitable for any practical purpose. We will assume hereinafter that $\delta_p < 0.5$ and $\delta_s < 0.5$.

We now extend the previous notions to 2-D filters. We start with the case of factorable filters. Given two low–pass FIR filters $h_1(n)$ and $h_2(n)$ (with frequency response $H_1(f)$ and $H_2(f)$, respectively), characterized respectively by parameters $(f_{p_i}, f_{s_i}, \delta_{p_i}, \delta_{s_i}, N_i), i = 1, 2$, consider the 2-D filter

$$h(n_1, n_2) = h_1(n_1)h_2(n_2). \quad (23)$$

The frequency response of such a filter (periodic on Z^2) is $H(\mathbf{f}) = H_1(f_1)H_2(f_2)$. In the ideal case

$$\delta_{p_1} = \delta_{p_2} = \delta_{s_1} = \delta_{s_2} = 0 \quad (24)$$

we have that $H(\mathbf{f}) = 1$ for $\mathbf{f} \in \mathcal{R}(f_{p_1}, f_{p_2})$ and $H(\mathbf{f}) = 0$ for $\mathbf{f} \in \mathcal{R}(0.5, 0.5) \setminus \mathcal{R}(f_{s_1}, f_{s_2})$. In such a case, the *transition region* is "naturally" defined as $\mathcal{R}(f_{s_1}, f_{s_2}) \setminus \mathcal{R}(f_{p_1}, f_{p_2})$..

In any practical situation, condition (24) is never met, and the intuitive notion of "transition region" needs to be defined precisely. For this purpose, we may adopt a procedure which is reminiscent of the one-dimensional case. We define *transition region* of a 2-D filter $H(\mathbf{f})$ the region of the unit cell $\mathcal{R}(0.5, 0.5)$ delimited by the *passband curve* $\underline{\mathcal{P}}$ and the *stopband curve* $\underline{\mathcal{S}}$ (when they are univocally determined), defined as follows:

$$\underline{\mathcal{P}} = \{\mathbf{f} : H(\mathbf{f}) = 1 - \delta_p, \|\nabla H(\mathbf{f})\|^2 \neq 0\} \quad (25)$$

$$\underline{\mathcal{S}} = \{\mathbf{f} : H(\mathbf{f}) = \delta_s\ , \|\nabla H(\mathbf{f})\|^2 \neq 0\} \quad (26)$$

where ∇ indicates the gradient operator and

$$\delta_p = \max_{\{\mathbf{f}_i : \|\nabla H(\mathbf{f}_i)\|^2 = 0, |H(\mathbf{f}_i)|>0.5\}} |1 - H(\mathbf{f}_i)| \quad (27)$$

$$\delta_s = \max_{\{\mathbf{f}_i : \|\nabla H(\mathbf{f}_i)\|^2 = 0, |H(\mathbf{f}_i)|<0.5\}} |H(\mathbf{f}_i)|. \quad (28)$$

The region $\mathcal{P}$ contained within $\underline{\mathcal{P}}$ is called the *passband region* of $H(\mathbf{f})$, while the region of $\mathcal{R}(0.5, 0.5)$ outside $\underline{\mathcal{S}}$ is called the *stopband region* of $H(\mathbf{f})$. Note that our definitions may be easily extended to the case of filters defined on a nonorthogonal lattice. Instead of region $\mathcal{R}(0.5, 0.5)$, some other suitable unit cell of the frequency repetition lattice centered at the origin (e.g., the Voronoi cell [14]) may be chosen.

It may be interesting to check the proposed definitions for the case of a factorable filter $h(n_1, n_2) = h_1(n_1)h_2(n_2)$, where $h_1(n)$ and $h_2(n)$ are low-pass optimal in a minimax sense. Cases of interest for $\nabla H(\mathbf{f}) = \left(\frac{dH_1(f_1)}{df_1}H_2(f_2), H_1(f_1)\frac{dH_2(f_2)}{df_2}\right)^T = \mathbf{0}$ are points such that $\frac{dH_1(f_1)}{df_1} = \frac{dH_2(f_2)}{df_2} = 0$. Extremal interesting points $\mathbf{f}$ thus belong either to region $\mathcal{R}(f_{p_1}, f_{p_2})$ or to region $\mathcal{R}(0.5, 0.5) \setminus \mathcal{R}(f_{s_1}, f_{s_2})$. It is easily seen that

$$\delta_p = \delta_{p_1} + \delta_{p_2} + \delta_{p_1}\delta_{p_2} \simeq \delta_{p_1} + \delta_{p_2} \quad (29)$$

while

$$\delta_s = \max\ \{\delta_{s_1} + \delta_{s_1}\delta_{p_2}, \delta_{s_2} + \delta_{s_2}\delta_{p_1}\} \simeq \max\ \{\delta_{s_1}, \delta_{s_2}\} \quad (30)$$

Consider now the pass–band curve $\underline{\mathcal{P}}$. It is readily seen that $\underline{\mathcal{P}}$ is contained within region $\mathcal{R}(\bar{f}_1, \bar{f}_2) \setminus \mathcal{R}(f_{p_1}, f_{p_2})$, where $\bar{f}_1$ and $\bar{f}_2$ are such that $H_1(\bar{f}_1) = (1 - \delta_{p_1} - \delta_{p_2})/(1 + \delta_{p_2})$ and $H_2(\bar{f}_2) = (1-\delta_{p_1}-\delta_{p_2})/(1+\delta_{p_1})$. Note that $\bar{f}_1$ and $\bar{f}_2$ belong to the transition bands of $H_1(f)$ and $H_2(f)$, respectively, (so long as $1 - \delta_{p_1} - \delta_{p_2} > \delta_{s_1}$ and $1 - \delta_{p_1} - \delta_{p_2} > \delta_{s_2}$).

The bandwidth $(\bar{f}_1 - f_{p_1})$ depends mainly on 1) the passband ripple δ_{p_2} and 2) the behavior of $H_1(f)$ in its transition band. Approximating $H_1(f)$ in its transition band by a linear function, one can show that, for small values of the ripples, $(\bar{f}_1 - f_{p_1})$ can be approximated by $2\delta_{p_2} B_{t_1}$, where $B_{t_1} = (f_{s_1} - f_{p_1})$. Similar considerations apply to the bandwidth $(\bar{f}_2 - f_{p_2})$. In what follows, we will always assume that quantities $2\delta_{p_2} B_{t_1}$ and $2\delta_{p_1} B_{t_2}$ are small enough as to allow us to approximate $\underline{\mathcal{P}}$ with $\underline{\mathcal{R}}(f_{p_1}, f_{p_2})$.

The case of the stopband curve $\underline{\mathcal{S}}$ can be treated in a similar fashion. It can be seen that (assuming that the product $\delta_{p_i}\delta_{s_k}$ is negligible), if $\delta_{s_1} > \delta_{s_2}$, then $\underline{\mathcal{S}} = \underline{\mathcal{R}}(f_{s_1}, \tilde{f}_2)$, where $\tilde{f}_2$ is such that $H_2(\tilde{f}_2) = \delta_{s_1}$. Again, approximating $H_2(f)$ in its transition band by a linear function, we have that the bandwidth $(f_{s_2} - \tilde{f}_2)$ can be approximated by $(\delta_{s_1} - \delta_{s_2}) B_{t_2}$. In addition, in this case, we will assume that such term is small, allowing us to approximate $\underline{\mathcal{S}}$ by $\underline{\mathcal{R}}(f_{s_1}, f_{s_2})$. Similar considerations hold if $\delta_{s_1} < \delta_{s_2}$.

It is useful now to define the *conventional filter length* of an FIR filter as the number of samples of its impulse response not constrained to zero. Such a definition turns out to be profitable when dealing with filters defined on a non–factorable lattice. The conventional length of a filter is proportional to the number of OPS's required in a direct form realization.

Consider now the case of a filter obtained by subsampling an impulse response $\bar{h}(\mathbf{n})$ on a given lattice $\Gamma = LAT(\mathbf{A})$, obtaining $h(\mathbf{An})$. Such a filter may be suitable for processing signals defined on Γ, or to obtain different shapes of the passband and stopband regions via a change of basis (see Section III-A). If $\bar{N}_c$ is the conventional length of $\bar{h}(\mathbf{n})$, the conventional length of $h(\mathbf{n})$ is approximately equal to $\bar{N}_c/|\det(\mathbf{A})|$.

One could analyze $||\nabla H(\mathbf{f})||$ in order to obtain parameters $\underline{\mathcal{P}}, \underline{\mathcal{S}}, \delta_p$ and δ_s, according to the previous definitions (25)–(28). Unfortunately, no general result is to be found, because the position of the zeros of $||\nabla H(\mathbf{f})||$ is *a priori* unknown. However, a simple argument based on the number of replicas generated by subsampling [3] shows that

$$\delta_p \leq \bar{\delta}_p + (|\det(\mathbf{A})| - 1)\bar{\delta}_s \tag{31}$$

and

$$\delta_s \leq |\det(\mathbf{A})|\bar{\delta}_s \tag{32}$$

where $\bar{\delta}_p$ and $\bar{\delta}_s$ are the passband and stopband ripples of $\bar{H}(\mathbf{f})$ as in (27) and (28).

Combining inequalities (31) and (32), an upper bound for $\delta_p\delta_s$ can be found. Note that, in the case of a 1-D filter, one can find a lower bound for $\delta_p\delta_s$ too. Let M be the decimation ratio (corresponding to $|\det(\mathbf{A})|$). Then, recalling from (1) that the lower bound for $\bar{\delta}_p\bar{\delta}_s$ is a function of the product $\bar{N}\bar{B}_t$ (where $\bar{N}$ is the length of $\bar{h}(n)$ and $\bar{B}_t$ its transition band), and observing that

$$NB_t \simeq \bar{N}\bar{B}_t \tag{33}$$

(where $N \simeq \bar{N}/M$ and $B_t \simeq \bar{B}_t M$ are referred to $h(n)$), we have that

$$\delta_p\delta_s \geq \left(\bar{\delta}_p\bar{\delta}_s\right)^{\text{opt}} \tag{34}$$

where $\left(\bar{\delta}_p\bar{\delta}_s\right)^{\text{opt}}$ is the product of the passband and the stopband of the optimal minimax filter with length $\bar{N}$ and transition band $\bar{B}_t$. The determination of similar lower bound relations for the multidimensional case (where relation (1) does not apply) is the object of current research.

The displacement of the passband and stopband curves $\underline{\mathcal{P}}$ and $\underline{\mathcal{S}}$ (of $H(\mathbf{f})$) with respect to $\bar{\underline{\mathcal{P}}}$ and $\bar{\underline{\mathcal{S}}}$ (of $\bar{H}(\mathbf{f})$) depends mainly on values δ_s and $|\det(\mathbf{A})|$, and on the relative position of the ripples in the passband and in the stopband regions of $\bar{H}(\mathbf{f})$. Intuitively, the smaller such values, the "closer" the two curves $\underline{\mathcal{P}}$ and $\underline{\mathcal{S}}$ to $\bar{\underline{\mathcal{P}}}$ and $\bar{\underline{\mathcal{S}}}$, respectively. In our simplified analysis, we will approximate $\underline{\mathcal{P}}$ and $\underline{\mathcal{S}}$ with $\bar{\underline{\mathcal{P}}}$ and $\bar{\underline{\mathcal{S}}}$, respectively.

III. Generalized Factorable Filters

In this section, we review the algorithm proposed by Chen and Vaidyanathan [1]–[3] to design M-D filters with pass–band in the shape of a parallelepiped. In Section III-A we briefly restate the design procedure, adding some remarks to the original algorithm, and in Section III-B we review the "generalized factorable" implementation. An important result of Section III-B is that the number of OPS's required for the generalized factorable implementation of a filter $h(\mathbf{a})$ defined on a lattice Λ, and of a filter with the same frequency response specs of $h(\mathbf{a})$, but defined on a sublattice Γ of Λ, may or may not differ, depending on the joint geometrical characteristics of Λ and Γ. The formal definition of GF filters is then introduced in Section III-C.

A. Chen and Vaidyanathan's Design Technique

Throughout the remainder of this paper, the signal definition lattice will be denoted as $\Lambda = LAT(\mathbf{C})$. The procedure proposed by Chen and Vaidyanathan [1]–[3] enables one to design an M-D FIR filter defined on Λ, with passband region approximating a parallelepiped centered at the origin, starting from M suitable 1-D lowpass filters.

Let

$$\text{SPD}(\mathbf{P}) = \left\{ \sum_{i=1}^{M} \alpha_i \mathbf{p}_i \ , \ -1 \leq \alpha_i \leq 1 \right\} \tag{35}$$

be a parallelepiped (representing the desired pass–band region), characterized by matrix $\mathbf{P} = (\mathbf{p}_1|\mathbf{p}_2|\cdots|\mathbf{p}_M)$. We assume that $\mathbf{P}$ has only rational entries. Consider the parallelepiped SPD($\mathbf{P}_t = \mathbf{C}^T\mathbf{P}$) (matrix $\mathbf{C}$ may actually be any basis of Λ). SPD($\mathbf{P}_t$) represents a "transformed" version of the passband region SPD($\mathbf{P}$).

Let

$$\bar{\mathbf{A}} = \mathbf{P}_t^T \text{den}(\mathbf{P}_t). \tag{36}$$

Note that $\bar{\mathbf{A}}$ is integral. Now consider filter $\bar{h}(\mathbf{n}) = \prod_{i=1}^{M} q(n_i)$, where $q(n)$ is an ideal 1-D filter such that

$$Q(f) = \begin{cases} 1, & |f| \le f_p \\ 0, & f_p < |f| \le 0.5 \end{cases} \tag{37}$$

and $f_p = 1/\mathrm{den}(\mathbf{P}_t)$. Filter $\hat{h}(\mathbf{n}) = |\det(\bar{\mathbf{A}})|\bar{h}(\bar{\mathbf{A}}\mathbf{n})$ is such that, within a suitable unit cell,

$$\hat{H}(\mathbf{f}) = \begin{cases} 1, & \mathbf{f} \in \mathrm{SPD}(\mathbf{P}_t) \\ 0, & \text{otherwise.} \end{cases} \tag{38}$$

The just described algorithm suffers from a "design overhead" [3], in the sense that $|\det(\bar{\mathbf{A}})|$ (the "decimation ratio") is typically higher than necessary, and filter $Q(f)$ may have a very narrow transition band. It is important that $|\det(\mathbf{A})|$ be small for the control of the filter characteristics, as will be stressed later in Section IV-A. It is possible (in general) to lower such a value by allowing for different 1-D filters along the axes [3]. Let

$$\bar{\mathbf{A}} = \mathbf{D}\mathbf{A}, \quad \mathbf{D} = \mathrm{diag}(D_1, D_2, \cdots, D_M) \tag{39}$$

where D_i is the greatest common divisor of the entries of the ith row of $\bar{\mathbf{A}}$. Note that $\mathbf{A}$ is integral. Consider M 1-D ideal filters $q_i(n)$, such that

$$Q_i(f) = \begin{cases} 1, & |f| \le f_{p_i} \\ 0, & f_{p_i} < |f| \le 0.5 \end{cases} \tag{40}$$

where $f_{p_i} = D_i/\mathrm{den}(\mathbf{P}_t)$. Their tensor product

$$\bar{h}(\mathbf{n}) = \prod_{i=1}^{M} q_i(n_i) \tag{41}$$

has spectral support in SPD$(\mathrm{diag}(f_{p_1}, f_{p_2}, \cdots, f_{p_M}))$, while $\hat{h}(\mathbf{n}) = |\det(\mathbf{A})|\bar{h}(\mathbf{A}\mathbf{n})$ has spectral support in SPD$(\mathbf{P}_t)$. Finally, the sought for filter $h(\mathbf{C}\mathbf{n})$ which satisfies

$$H(\mathbf{f}) = \begin{cases} 1, & \mathbf{f} \in \mathrm{SPD}(\mathbf{P}) \\ 0, & \text{otherwise} \end{cases} \tag{42}$$

within a suitable unit cell of the dual lattice Λ^*, is

$$h(\mathbf{C}\mathbf{n}) = \hat{h}(\mathbf{n}) = |\det(\mathbf{A})|\bar{h}(\mathbf{A}\mathbf{n}). \tag{43}$$

Remarks:

1) In the original algorithm by Chen and Vaidyanathan [3], the authors used term $|\det(\mathbf{M})|$ (where $\mathbf{L}^{-1}\mathbf{M}$ is any left coprime factorization of $\mathbf{P}_t^{-T}$) instead of term $\mathrm{den}(\mathbf{P}_t)$ in (36). Their choice yields higher values of $|\det(\bar{\mathbf{A}})|$ than necessary, as $|\det(\mathbf{M})|$ can be larger than $\mathrm{den}(\mathbf{P}_t)$ (for a proof, see the Appendix of [12]). However, if we allow the 1-D filters to differ from one another (like in the second part of the algorithm), any common factor in the entries of $\bar{\mathbf{A}}$ will be "absorbed" by matrix $\mathbf{D}$.
2) Note that, from (43), any different choice $\tilde{\mathbf{C}} = \mathbf{C}\mathbf{U}$, unimodular $\mathbf{U}$, of the basis of Λ, induces a different sampling matrix $\tilde{\mathbf{A}} = \mathbf{A}\mathbf{U}$. Hence, we can always pick a basis $\mathbf{C}$ of Λ such that $\mathbf{A}$ is in upper (lower) Hermite normal form.
3) In practical cases, filters $q_i(n)$ will be characterized by the minimax parameters $(f_{p_i}, f_{s_i}, \delta_{p_i}, \delta_{s_i}, N_i)$ (as defined in Section II-B). Assume that the desired frequency response of the filter $H(\mathbf{f})$ is specified by the passband and stopband surfaces SPD$(\mathbf{P}_p)$ and SPD$(\mathbf{P}_s)$ in the shape of parallelepipeds having pairwise parallel faces, i.e.,
$$\mathbf{P}_s = \mathbf{P}_p\mathbf{T}\ ,\ \mathbf{T} = \mathrm{diag}(T_1, T_2, \cdots T_M)\ ,\ T_i > 1. \tag{44}$$
The corresponding decimation matrices $\mathbf{A}_p$ and $\mathbf{A}_s$, obtained via the previous algorithm, may differ from each other. One way to circumvent such a problem, is to compute the matrix $\mathbf{A}$ and the stop–band frequencies $\{f_{s_i}\}$ following the previous algorithm for $\mathbf{P} = \mathbf{P}_s$, and set the passband frequencies $\{f_{p_1} = f_{s_i}/t_i\}$. Then, within the approximation discussed in Section II-B, the passband and stopband regions of the filter will coincide with SPD$(\mathbf{P}_p)$ and SPD$(\mathbf{P}_s)$, respectively.
4) It can be shown that, if P is a Z^M-period of $LAT(\mathbf{A}^T)$, then
$$H(\mathbf{f}) = \sum_{\mathbf{r} \in P} \bar{H}\big(\mathbf{A}^{-T}(\mathbf{C}^T\mathbf{f} + \mathbf{r})\big) \tag{45}$$
and that the support of the impulse response of $h(\mathbf{C}\mathbf{n})$ is
$$\mathrm{SPD}\left(\mathbf{C}\mathbf{A}^{-1}\mathrm{diag}\left(\left\{\frac{N_i+1}{2}\right\}\right)\right) \cap LAT(\mathbf{C}) \tag{46}$$
where N_i is the length of the ith 1-D filter in (40). In the remainder of this paper, we will approximate the conventional length of $h(\mathbf{C}\mathbf{n})$ with $\prod_{i=1}^{M} N_i/|\det(\mathbf{A})|$.
5) Matrix $\mathbf{D}$ in decomposition (39) has the following geometrical interpretation: $LAT(\mathbf{D})$ is the least dense factorable lattice (LDFL, see Section II-A-2)) containing $LAT(\bar{\mathbf{A}})$. Then, the LDFL containing $LAT(\mathbf{A})$ is Z^M. This in turn implies that, if $\mathbf{A}$ is diagonal, then $\mathbf{A} = \mathbf{I}$.
6) Suppose one is given a GF filter $h(\mathbf{a})$ defined on $\Lambda = LAT(\mathbf{C})$ with spectral support approximating SPD$(\mathbf{P})$. Let $\Gamma = LAT(\mathbf{C}\mathbf{K})$ be a sublattice of Λ, and assume that the spectral support of $h(\mathbf{a})$ is contained within some unit cell of Γ^*. In order to design a filter $g(\mathbf{a})$ defined on Γ with the same spectral support of $h(\mathbf{a})$, one can choose between two procedures. The simpler one is derived immediately from equation (43), letting $g(\mathbf{C}\mathbf{K}\mathbf{n}) = |\det(\mathbf{K})|h(\mathbf{C}\mathbf{K}\mathbf{n})$, i.e., subsampling $h(\mathbf{a})$ on Γ and adjusting the gain. The conventional length N_g of $g(\mathbf{a})$ is approximately equal to the conventional length N_h of $h(\mathbf{a})$ divided by $|\det(\mathbf{K})|$. If the LDFL containing $LAT(\mathbf{A}\mathbf{K})$ is less dense than $LAT(\mathbf{D})$ (see Remark 5 above), one can use the second part of the algorithm instead, and construct the new sampling matrix and the new passband and stopband frequencies for the 1-D filters. It is easily seen (making use of relation (1)) that, if the 1-D filters of (40) are forced to exhibit the same passband and stopband ripples as in the design of $h(\mathbf{a})$, the conventional length of $g(\mathbf{a})$ will be again approximately equal to $N_h/|\det(\mathbf{K})|$.
7) Starting from 1-D optimal minimax filters, one does not necessarily obtain the best minimax approximation (in

the class of GF filters) to the ideal frequency response of (42). The determination of the constraints on the 1-D filters to get the optimal minimax solution, remains an open issue.

B. Generalized Factorable Implementation

Following Chen and Vaidyanathan's algorithm, one designes a GF filter by sampling an appropriate factorable impulse response. Hence, it should not surprise that techniques for the filter's efficient implementation may be devised. Let N_1 and N_2 be the lengths of the 1-D filters $q_1(n)$ and $q_2(n)$ of Section III-A (for the sake of notation's simplicity, in this section we refer only to the 2-D case). If the filter $h(\mathbf{a})$ is implemented in direct form, approximately $N_1N_2/|\det(\mathbf{A})|$ OPS's are required (which corresponds to the conventional length of the filter). Chen and Vaidyanathan cleverly proved in [3] that it is possible to implement the filter with only approximately $N_1 + N_2$ OPS's. Their proof is based on a machinery of formal identities; our proof is (hopefully) more intuitive.

In the effort to be clear, we divide the procedure for the generalized factorable implementation into several steps.

Let $\mathbf{C}$ and $\mathbf{A}$ be bases of the signal definition lattice and of the decimated lattice respectively, as in Section III-A. Assume, without loss of generality, that $\mathbf{A}$ is in upper Hermite normal form (see Remark 2 in Section III-A).

Step 1: Consider a change of basis on the input signal $x(\mathbf{Cn})$ and on the filter $h(\mathbf{Cn})$:

$$\bar{x}(\mathbf{An}) = x(\mathbf{Cn}) \ , \ \bar{h}(\mathbf{An}) = h(\mathbf{Cn}). \tag{47}$$

Step 2.: Filter $\bar{h}(\mathbf{An})$ is not factorable (unless $\mathbf{A}$ is unimodular), but it is made up of factorable polyphase components. To prove this, let $\mathbf{AH} = \mathrm{diag}(S_1, S_2), S_1, S_2 > 0$ be a basis of the densest factorable sublattice (DFS, see Section II-A-2)) of $LAT(\mathbf{A})$. Then, there are $(\mathrm{den}A_{1,2}/A_{1,1})$ $LAT(\mathbf{AH})$-polyphase components in $\bar{h}(\mathbf{An})$:

$$\bar{h}^{\mathbf{r}}(\mathbf{AHn}) \stackrel{\text{def}}{=} \bar{h}(\mathbf{AHn} + \mathbf{r}) \ , \ \mathbf{r} \in P. \tag{48}$$

where P is some $LAT(\mathbf{AH})$-period of $LAT(\mathbf{A})$. Each component $\bar{h}^{\mathbf{r}}(\mathbf{a}), \mathbf{a} \in LAT(\mathbf{AH})$, is factorable. In fact, let $\mathbf{a} = (k_1S_1, k_2S_2^T)$ and $\mathbf{r} = (r_1, r_2^T)$; then

$$\bar{h}^{\mathbf{r}}(\mathbf{a}) = q_1^{r_1}(k_1)q_2^{r_2}(k_2). \tag{49}$$

where $q_i^{r_i}$ is the r_ith S_i-polyphase component of filter $q_i(n)$:

$$q_i^{r_i}(n) \stackrel{\text{def}}{=} q_i(nS_i + r_i) \ , \ 0 \le r_i < S_i. \tag{50}$$

Step 3: At this point, it should be clear that it is possible to filter $\bar{x}(\mathbf{a})$ with $\bar{h}(\mathbf{a})$ in a factorable fashion. It just needs to write down the $LAT(\mathbf{AH})$-polyphase decomposition of $\bar{h}(\mathbf{a})$ to get

$$\begin{aligned}\bar{X}(\mathbf{f})\bar{H}(\mathbf{f}) &= \bar{X}(\mathbf{f}) \sum_{\mathbf{r}\in P} e^{-j2\pi\mathbf{f}^T\mathbf{r}} \bar{H}^{\mathbf{r}}(\mathbf{f}) \\ &= \sum_{\mathbf{r}\in P} \left(\bar{X}(\mathbf{f})e^{-j2\pi\mathbf{f}^T\mathbf{r}}\right) \bar{H}^{\mathbf{r}}(\mathbf{f}).\end{aligned} \tag{51}$$

Each term $\left(\bar{X}(\mathbf{f})e^{-j2\pi\mathbf{f}^T\mathbf{r}}\right)\bar{H}^{\mathbf{r}}(\mathbf{f})$ represents the (factorable) filtering, by the $LAT(\mathbf{AH})$-polyphase component of $\bar{h}(\mathbf{a})$ of index $\mathbf{r}$, of a version of $\bar{x}(\mathbf{a})$ displaced by $\mathbf{r}$.

It is interesting to derive the number of OPS's required for the generalized factorable implementation. Let the length of the two 1-D filters $q_1(n)$ and $q_2(n)$ be N_1 and N_2 respectively. Then the lengths of their S_1– and S_2–polyphase components as in (50) are approximately N_1/S_1 and N_2/S_2, respectively. Now, from Section II-A2) we have that

$$\#\ \text{LAT}(\mathbf{AH})\text{–polyphase components} = S_1/R_1 = S_2/R_2 \tag{52}$$

where $\mathrm{diag}(R_1, R_2), R_1, R_2 > 0$, is a basis of the LDFL containing $LAT(\mathbf{A})$, and symbol "#" stands for "number of." But, according to Remark 5 of Section III-A, $R_1 = R_2 = 1$ (and consequentely $S_1 = S_2$). Hence, the number of OPS's required for the realization is approximately

$$\begin{aligned}\#\ \text{OPS's} &= \#\ \text{LAT}(\mathbf{AH})\text{–polyphase components} \\ N_1/S_1 + N_2/S_2 &= N_1 + N_2.\end{aligned} \tag{53}$$

Consider now the case of a filter defined on the sublattice $\Gamma = LAT(\mathbf{CK})$, designed by subsampling $h(\mathbf{a})$ (as in the first procedure of Remark 6 above). Let $\mathrm{diag}(R_1', R_2'), R_1', R_2' > 0$ be a basis of the LDFL containing $LAT(\mathbf{AK})$, and let $\mathbf{AH}' = \mathrm{diag}(S_1', S_2'), S_1', S_2' > 0$ be a basis of the DFS of $LAT(\mathbf{AK})$ (note that $S_1' \ge S_1, S_2' \ge S_2$). Now we have that, in general, $R_1' \ne R_2'$, so that $S_1' \ne S_2'$. The number of OPS's required for the realization in this case is

$$\#\ \text{OPS's} = S_1'/R_1' \cdot (N_1/S_1' + N_2/S_2') = N_1/R_1' + N_2/R_2'. \tag{54}$$

Let us summarize this last result: the number of OPS's required for the generalized factorable implementation of a GF filter defined on $\Lambda = LAT(\mathbf{C})$, and of its version subsampled on a sublattice $\Gamma = LAT(\mathbf{CK})$, may or may not differ, according to (53) and (54). In particular, let $\mathbf{A}$ be the basis of the subsampling lattice like in Section III-A. Then, if the LDFL containing $LAT(\mathbf{A})$ and the LDFL containing $LAT(\mathbf{AK})$ coincide, the number of required OPS's will be the same the two cases. Such a result is in contrast with the case of non-factorable implementation, where subsampling the impulse response reduces the number of OPS's by a factor approximately equal to the subsampling ratio. Similar arguments hold adopting the second procedure of Remark 6.

As an example, consider the case of $\Lambda = Z^2$ and $\mathbf{A} = \begin{pmatrix} 2 & 1 \\ 0 & 1 \end{pmatrix}$. Lattice $LAT(\mathbf{A})$ is represented by dots in Fig. 2(a) and (b). A basis of the LDFL containing $LAT(\mathbf{A})$ is $\mathrm{diag}(R_1 = 1, R_2 = 1)$, while a basis of the DFS of $LAT(\mathbf{A})$ is $\mathrm{diag}(S_1 = 2, S_2 = 2) = \mathbf{AH}$. Hence, there are $|\det(\mathbf{H})| = 2$ $LAT(\mathbf{AH})$-polyphase components. If the lengths of the 1-D filters $q_1(n), q_2(n)$ are N_1 and N_2 respectively, we have that the length of the S_1- and S_2-polyphase components is approximately $N_1/2$ and $N_2/2$, respectively, so that approximately $N_1 + N_2$ OPS's are required. Consider now the sublattice of $LAT(\mathbf{A})$ with basis $\mathbf{AK} = \begin{matrix} 4 & 1 \\ 0 & 1 \end{matrix}$ (represented by small circles in

Fig. 2(a)). A basis of the LDFL containing $LAT(\mathbf{AK})$ is diag $(R_1' = R_1 = 1, R_2' = R_2 = 1)$, whereas a basis of the DFS of $LAT(\mathbf{AK})$ is $\mathrm{diag}(S_1' = 4, S_2' = 4) = \mathbf{AH}'$. The number of $LAT(\mathbf{AH}')$-polyphase components is $|\det(\mathbf{H}')| = 4$, whereas the lengths of the S_1' and S_2'–polyphase components of the two 1-D filters are approximately $N_1/4$ and $N_2/4$. Hence, $N_1 + N_2$ OPS's are also required in this case since the LDFL containing $LAT(\mathbf{A})$ and the LDFL containing $LAT(\mathbf{AK})$ coincide. One can easily verify that in the case of the factorable sublattice with basis $\mathbf{AK}' = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$ (which is represented by small circles in Fig. 2(b)), only $(N_1 + N_2)/2$ OPS's are required for its realization. In this case, the LDFL containing $LAT(\mathbf{AK}')$ (i.e., $LAT(\mathbf{AK}')$ itself), is different from the LDFL containing $LAT(\mathbf{A})$ (i.e., Z^2).

The number of multiplications in the implementation of the zero-phase 1-D filters can be reduced by exploiting the symmetry of the coefficients. Note however that the r–th S-polyphase component of a zero-phase 1-D filter $q(n)$ satisfies the symmetry property only if $2r/S$ is integer (as it can be easily proved). Since it is $0 \le r < S$, we have that only the zeroth S–polyphase component (if S is odd) or only the zeroth and the $S/2$th polyphase component (if S is even) satisfy the symmetry property.

Step 4: We should now get back to the definition lattice $LAT(\mathbf{C})$ via the inverse of transformation (47). Each coset $LAT(\mathbf{AH}) + \mathbf{r}$ is mapped into $LAT(\mathbf{CH}) + \mathbf{CA}^{-1}\mathbf{r}$ (note that $\mathbf{A}^{-1}\mathbf{r} \in Z^2$ as $\mathbf{r} \in LAT(\mathbf{A})$). In particular, points $\{kS_i\}$ on the i–th axis are mapped into points $\{k\mathbf{T}_i\}$, where $\mathbf{T}_i$ is the ith column of $\mathbf{CH}$. Hence, the factorable filtering of Step 3 corresponding to the polyphase index $\mathbf{r}$, becomes the cascade of two generalized 1-D filterings (the ith of which by the r_ith S_i-polyphase component of $q_i(n)$ along the direction $\mathbf{T}_i$) of signal $x(\mathbf{a})$, displaced by vector $\mathbf{CA}^{-1}\mathbf{r}$.

Before concluding this section, we would like to point out that in certain applications, computational complexity is not the main issue. For example, in the case of spatio-temporal video filters, computational complexity trades off for memory requirement.

C. Generalized Factorable Filters: A Formal Definition

Let $\Lambda = LAT(\mathbf{C})$ be an M-D lattice. Consider the set of points of Λ aligned along a given direction:

$$\mathcal{L}_{\mathbf{v}} = \{k\mathbf{v}, \text{ integer } k\} \tag{55}$$

where $\mathbf{v} = \mathbf{Cn}$ with $n_1, n_2, \cdots, n_M$ coprime (not necessarily pairwise coprime). $\mathcal{L}_{\mathbf{v}}$ spans all the points of Λ along the line $\alpha\mathbf{v}, \alpha \in R$. A filter $h(\mathbf{a})$ such that $h(\mathbf{a}) \neq 0$ only for $\mathbf{a} \in \mathcal{L}_{\mathbf{v}}$ will be called a *generalized 1-D filter on* $\mathbf{v}$. Let N be the number of non-null elements of $h(\mathbf{a})$ (assume for simplicity's sake that between two non-null samples of $h(\mathbf{a})$ there is always a non-null sample). N will be called the *generalized length* of $h(\mathbf{a})$. In general, N OPS's are required to implement $h(\mathbf{a})$ (more precisely, in case no symmetry is present, in the direct form realization N multiplications and $N - 1$ sums per input sample are required).

Suppose now Λ is factorable (without loss of generality, we can assume $\Lambda = Z^M$). Then $h(\mathbf{n})$ is said to be factorable

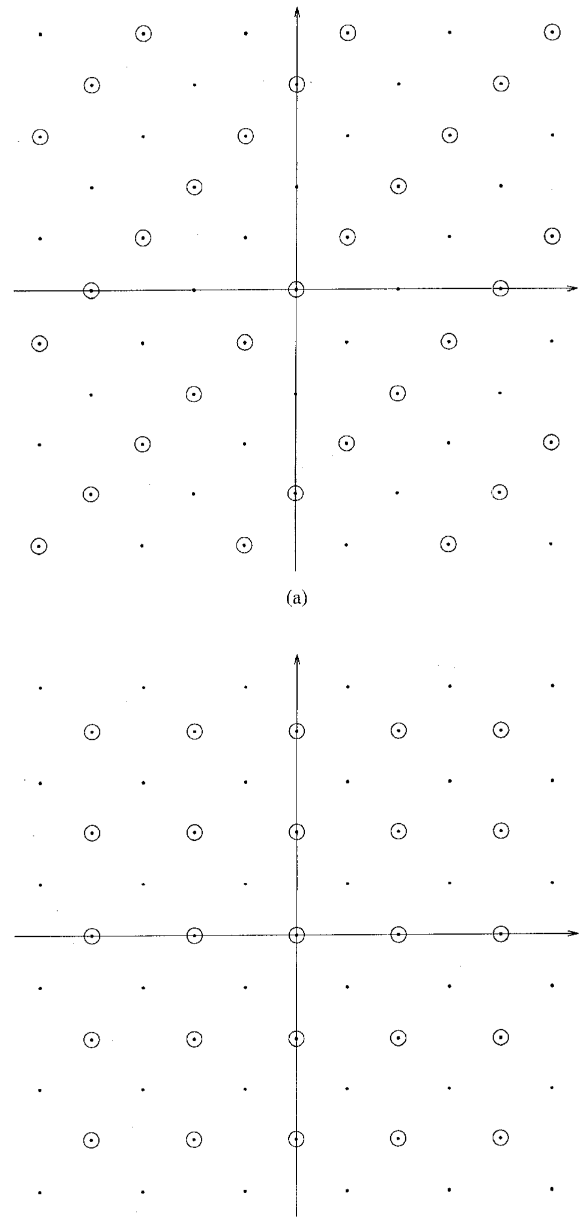


Fig. 2. (a) Lattice $LAT(\mathbf{A})$ (dots) and its sublattice $LAT(\mathbf{AK})$ (circles) (see text). (b) Lattice $LAT(\mathbf{A})$ (dots) and its sublattice $LAT(\mathbf{AK}')$ (see text).

[9] if it is the tensor product of M 1-D filters: $h(\mathbf{n}) = \prod_{i=1}^{M} q_i(n_i)$. We can extend such a notion exploiting the idea of generalized 1-D filters: Given a lattice Λ and a basis of its $\mathbf{C} = \mathbf{c}_1|\mathbf{c}_2|\cdots|\mathbf{c}_M$ of Λ, we will say that $h(\mathbf{a})$ is *generalized factorable* (GF) *on* $\mathbf{C}$ if $h(\mathbf{Cn}) = \prod_{i=1}^{M} q_i(\mathbf{c}_i n_i)$, where each filter $q_i(\mathbf{a})$ is generalized 1-D on $\mathbf{c}_i$. Note that such a definition applies to filters defined on any lattice, and not only on factorable ones.

The idea of generalized factorability allows us to classify filters that, although not factorable, can be transformed, via a change of basis of the definition lattice, into factorable filters. Hence, they can be implemented with a number of OPS's that grows linearly with the sum of the lengths of their impulse response's edges (instead of growing with the product of such lengths, like in the "direct" implementation). We extend further our definition to include the class of filters designed using the algorithm by Chen and Vaidyanathan.

Let $\mathbf{V}$ be an integral matrix. A filter $h(\mathbf{a})$ defined on $\Lambda = LAT(\mathbf{C})$ is said to be *generalized factorable* (GF) *on* $\mathbf{CV}$ if its $LAT(\mathbf{CV})$-polyphase components $h^{\mathbf{r}}(\mathbf{s})$ are GF on $\mathbf{CV}$, where

$$h^{\mathbf{r}}(\mathbf{s}) \stackrel{\text{def}}{=} h(\mathbf{s}+\mathbf{r}) \;,\; \mathbf{s} \in LAT(\mathbf{CV}) \;,\; \mathbf{r} \in P \tag{56}$$

and P is any $LAT(\mathbf{CV})$-period of $LAT(\mathbf{C})$. A filter defined on $LAT(\mathbf{C})$, which is GF on $\mathbf{CV}$, is also GF on $\mathbf{CVD}$ for any integral diagonal matrix $\mathbf{D}$. Moreover, for any FIR filter $h(\mathbf{a})$ defined on $LAT(\mathbf{C})$, it is always possible to find a matrix $\mathbf{V}$ such that $h(\mathbf{a})$ is GF on $\mathbf{CV}$. To prove this, one just needs to choose a matrix $\mathbf{V}$ such that the support of $h(\mathbf{a})$ is contained within some unit cell of $LAT(\mathbf{CV})$.

Note that the procedure of Chen and Vaidyanathan represents the only known algorithm to design filters GF on $\mathbf{CV}$, where $\mathbf{AV}$ is a diagonal basis of the DFS of $LAT(\mathbf{A})$, and $\mathbf{A}$ is the decimation matrix, according to the notation of Section III-A.

IV. Some Properties of 2-D GF Filters

A. Minimax Parameters

We derive here some relations among conventional filter length (see Section II–A2)), passband and stopband ripples, and some measure of the "size" of the transition region, that hold for GF filters designed from optimal minimax 1-D filters.

Adopting the notation of Section III-A, from Remark 4 in the same section it is seen that

$$\mathbf{F}_p \stackrel{\text{def}}{=} \operatorname{diag}(f_{p_1}, f_{p_2}) = \mathbf{A}^{-T}\mathbf{C}^T\mathbf{P}_p \tag{57}$$

$$\mathbf{F}_s \stackrel{\text{def}}{=} \operatorname{diag}(f_{s_1}, f_{s_2}) = \mathbf{A}^{-T}\mathbf{C}^T\mathbf{P}_s \tag{58}$$

Noting that $|\det(\mathbf{F}_s - \mathbf{F}_p)| = |\det(\mathbf{A}^{-T}\mathbf{C}^T(\mathbf{P}_s - \mathbf{P}_p))| = B_{t_1}B_{t_2}$, where $B_{t_i} = f_{s_i} - f_{p_i}$, we have that, for given passband and stopband parallelograms specifying the support of $H(\mathbf{f})$, the product of the transition bandwidths of filters $q_1(n)$ and $q_2(n)$ (and, therefore, term $1/N_1N_2$, for given passband and stopband ripples δ_{p_i} and δ_{s_i} of the two 1-D filters, see (1)) is proportional to $|\det(\mathbf{C})|/|\det(\mathbf{A})|$. The conventional length N_c of filter $h(\mathbf{a})$, which corresponds to the number of OPS's required in the direct form realization, is approximately equal to $N_1N_2/|\det(\mathbf{A})|$; we thus maintain that, for given passband and stopband ripples of $Q_1(f)$ and $Q_2(f)$, N_c can be considered approximately inversely proportional to $|\det(\mathbf{C})|$ and independent of $\mathbf{A}$. Similarly, for fixed value of $|\det(\mathbf{C})|/|\det(\mathbf{A})|$, the conventional length N_c is approximately inversely proportional to $\det(\mathbf{P}_s - \mathbf{P}_p)$ (which represents the area $\mathcal{A}$ of the "corner region" depicted in Fig. 3):

$$\mathcal{A} \cdot N_c \simeq \text{constant}. \tag{59}$$

The resemblance of such a result with the relationship between transition bandwidth, ripples and length of minimax 1-D filters (1), is apparent. A more careful exam, though, shows that the situation is actually quite tricky. As a matter of fact, relation (59) holds for fixed ripples $\delta_{p_i}, \delta_{s_i}$ of the two 1-D filters. The relation between $\{\delta_{p_i}, \delta_{s_i}\}$ and $\{\delta_p, \delta_s\}$ (the ripples of the resulting 2-D filter), is obtained combining together inequalities (29)–(32):

$$\delta_p \le \delta_{p_1} + \delta_{p_2} + (|\det(\mathbf{A})| - 1)\max\{\delta_{s_1}, \delta_{s_2}\} \tag{60}$$

$$\delta_s \le |\det(\mathbf{A})| \max\{\delta_{s_1}, \delta_{s_2}\}. \tag{61}$$

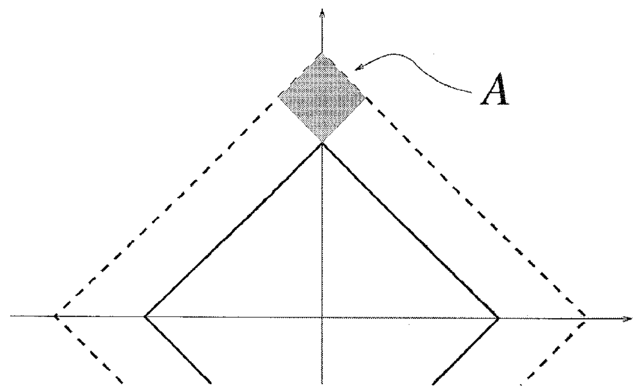


Fig. 3. Example of "corner region" for a diamond-shaped spectral mask. Solid line: passband curve. Dashed line: stopband curve.

We can obtain a simple worst-case relationship between $\delta_p\delta_s$ and N_c (for fixed area $\mathcal{A}$) if $\delta_{p_1} = \delta_{p_2} = \delta_{s_1} = \delta_{s_2} \stackrel{\text{def}}{=} \delta$. In such a case, we have that

$$\delta_p\delta_s \le |\det(\mathbf{A})|(|\det(\mathbf{A})| + 1)\delta^2. \tag{62}$$

As a matter of fact, the derived upper bound is quite pessimistic, and experimental tests show that one can actually reach much smaller values for $\delta_p\delta_s$. On the other side, the lack of a theoretical expression for the lower bound for $\delta_p\delta_s$ in the 2-D case (see Section II-B), makes it difficult to predict the actual characteristics of the resulting filter.

As an operative rule of thumb, we can accept the following simple approximation in the case of GF filters: term $\delta_p\delta_s$ decreases as the product $\mathcal{A} \cdot N_c$ increases. Due to the unpredictable effects of the contributions of the frequency response oscillations of the two 1-D filters, our statement may not hold true in some instances. A general theory capable of predicting such behaviors is beyond the scope of the present work. Nonetheless, experimental tests show that these situations are not frequent in practice. Clearly, our simple rule is more likely to hold true for small values of $|\det(\mathbf{A})|$. The higher the number of overlapping spectral repetitions, the more unpredictable the behavior of the frequency response. This is one of the reasons why, in the GF filter design algorithm, one seeks for decimation matrices $\mathbf{A}$ with the smallest value of $|\det(\mathbf{A})|$ (another reason being that—as already noted—large values of $|\det(\mathbf{A})|$ typically induce the passband and transition band of filters $q_i(n)$ to be narrow, with consequent increase of the design burden).

The just described relations among filter parameters have been used as a guideline criterion for the design of 2–D IFIR structures proposed in [6].

B. Symmetries

The impulse responses of M-D filters defined on lattices often enjoy symmetries, which may be exploited in order to reduce the computational weight (in terms of number of multiplications per input sample) if the generalized factorable implementation is not adopted. The type of symmetry we consider here is the one deriving from a *spatially complete congruent* mapping (i.e., $\Lambda \to \Lambda$ bijective) [19]. In other words, given filter $h(\mathbf{Cn})$, we are looking for an integral unimodular matrix $\mathbf{Q}$ such that

$$h(\mathbf{Cn}) = h(\mathbf{CQn})\ , \mathbf{n} \in Z^2. \quad (63)$$

It is useful, for the arguments of this section, to define the following integral unimodular matrices:

$$\mathbf{Q}_1 = \mathbf{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},\ \mathbf{Q}_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$
$$\mathbf{Q}_3(b) = \begin{pmatrix} 1 & b \\ 0 & -1 \end{pmatrix},\ \mathbf{Q}_4(c) = \begin{pmatrix} 1 & 0 \\ c & -1 \end{pmatrix} \quad (64)$$

where b and c are integer. The matrices of (64) characterize the set $\mathcal{Q}$ of 2×2 integral matrices that coincide with their own inverse:

$$\mathcal{Q} = \{\pm\mathbf{Q}_1, \pm\mathbf{Q}_2, \pm\mathbf{Q}_3(b), \pm\mathbf{Q}_4(c), \ \text{integer } b, c\}. \quad (65)$$

If the 1-D filters $q_1(n)$ and $q_2(n)$ of Section III-A are zero–phase (as we have assumed so far), then filter $\bar{h}(\mathbf{n})$ as in (41) satisfies property

$$\bar{h}(\mathbf{n}) = \bar{h}(\mathbf{Qn}),\ \mathbf{Q} \in \{\pm\mathbf{Q}_1, \pm\mathbf{Q}_3(0)\}. \quad (66)$$

Its sampled version $\bar{h}(\mathbf{An})$ (see (43)) satisfies property (from (66))

$$\bar{h}(\mathbf{An}) = \bar{h}(\mathbf{QAn}),\ \mathbf{Q} \in \{\pm\mathbf{Q}_1, \pm\mathbf{Q}_3(0)\} \quad \mathbf{QAn} \in LAT(\mathbf{A}). \quad (67)$$

In order for the mapping specified by $\mathbf{Q}$ to be complete, the last condition of (67) should hold true for each $\mathbf{n} \in Z^2$. This is equivalent to (remember that $\mathbf{Q}$ is unimodular)

$$LAT(\mathbf{QA}) = LAT(\mathbf{A}) \quad (68)$$

i.e., to

$$\mathbf{A}^{-1}\mathbf{QA} \text{ is integral.} \quad (69)$$

This condition is trivially satisfied if $\mathbf{A}$ is unimodular. Assume now $|\det(\mathbf{A})| > 1$. Note that, since $\det \mathbf{A}^{-1}\mathbf{QA} = \pm 1$, condition (69) is equivalent to

$$(\mathbf{A}^{-1}\mathbf{QA})^{-1} = \mathbf{A}^{-1}\mathbf{Q}^{-1}\mathbf{A} = \mathbf{A}^{-1}\mathbf{QA} \text{ is integral} \quad (70)$$

and, therefore, $\mathbf{A}^{-1}\mathbf{QA} = \mathbf{Q}_j$ for some $\mathbf{Q}_j \in \mathcal{Q}$. We can rewrite this last identity as

$$\mathbf{QA} = \mathbf{AQ}_j\ ,\ \mathbf{Q}_j \in \mathcal{Q}. \quad (71)$$

Assume $\mathbf{A}$ is in upper Hermite normal form (see Remark 2 of Section 3.1). We should now determine when (71) is satisfied for $\mathbf{Q} \in \{\pm\mathbf{Q}_1, \pm\mathbf{Q}_3(0)\}$. Cases $\mathbf{Q} = \pm\mathbf{Q}_1$ are trivial. In particular, $-\mathbf{Q}_1\mathbf{A} = -\mathbf{AQ}_1$, i.e., employing zero-phase 1-D filters, the resulting 2-D filter is zero-phase. This property was already derived in [3].

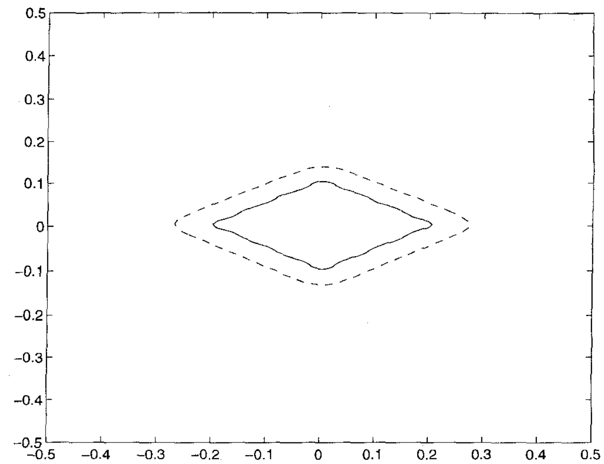


Fig. 4. Passband support (solid line) and stopband support (dashed line) of a GF filter within a unit cell of Z^2.

It remains to check case $\mathbf{Q} = \pm\mathbf{Q}_3(0)$, which can be verified by direct inspection. Condition (71) is never satisfied for $\mathbf{Q} = \pm\mathbf{Q}_3(0), \mathbf{Q}_j = \mathbf{Q}_2$. Case $\mathbf{Q}_j = \mathbf{Q}_3(0) = \mathbf{Q}_4(0)$ satisfies (69) only if $\mathbf{A}$ is diagonal, but, according to Remark 5 of Section III-A, this implies $\mathbf{A} = \mathbf{I}$, and we are back to the case of unimodular $\mathbf{A}$. The only other case for which condition (69) is satisfied when $\mathbf{Q} = \pm\mathbf{Q}_3(0)$, is for $\mathbf{Q}_j = \pm\mathbf{Q}_3(1)$ if $A_{1,1} = 2A_{1,2}$.

Let us summarize the results of this section. Assume the 1-D filters $q_1(n)$ and $q_2(n)$ are zero-phase. Then $h(\mathbf{Cn}) = h(-\mathbf{Cn})$. Moreover, if the decimation matrix $\mathbf{A}$ is unimodular, then

$$h(\mathbf{Cn}) = h(\pm\mathbf{CQ}_3(0)\mathbf{n}). \quad (72)$$

Such a relation can be regarded to as a quadrantal-like symmetry. The number of multiplications per input sample in this case is one half that of the case of simple zero-phase symmetry.

If $\mathbf{A}$ is not unimodular, let $\mathbf{A}^u$ be the upper Hermite normal form matrix right-equivalent to $\mathbf{A}$. Then relation (72) is verified if (and only if) $A^u_{1,1} = 2A^u_{1,2}$.

As an example, consider the case of a GF filter (defined on Z^2), whose passband support is represented by a solid line in Fig. 4. A basis of the decimated lattice is $\mathbf{A} = \begin{pmatrix} 2 & 1 \\ -2 & 1 \end{pmatrix}$, which is right-equivalent to the upper Hermite normal form matrix $\mathbf{A}^u = \begin{pmatrix} 4 & 1 \\ 0 & 1 \end{pmatrix}$. Since $A^u_{1,1} \neq 2A^u_{1,2}$, the quadrantal-like symmetry is not verified (in spite of the inherent symmetry of the passband support).

One may also look for a weaker symmetry condition, i.e. when relation (72) holds for points belonging to a sublattice $LAT(\mathbf{CH})$ of $LAT(\mathbf{C})$. In such a case, the number of multiplications per input sample required in the realization can be reduced by a factor $|\det(\mathbf{H})|/(|\det(\mathbf{H})| - 0.5)$ with respect to the case of simple zero–phase symmetry.

In our former example, if $\mathbf{H} = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$, then the upper Hermite normal form matrix right-equivalent to $\mathbf{AH}$ is $(\mathbf{AH})^u = \begin{pmatrix} 4 & 2 \\ 0 & 2 \end{pmatrix}$, and therefore, the quadrantal-like symmetry holds on sublattice $LAT(\mathbf{CH})$. The number of multiplications per input sample can be reduced by a factor $4/3$ exploiting such a symmetry.

C. 2-D Step Response

An important characteristic of filters to be implemented in video applications, is the behavior in case of sharp brightness transitions. The "ringing" corresponding to the oscillatory response of the filter in such situations is visually quite noticeable [20].

In the case of 1-D filters, the response to a unitary step characterizes the filter behavior in such critical cases. For 2-D filters, the response to 2-D unitary steps along two or more directions, as well as to other "transition" functions, is usually considered. For example, in [21] a technique to put linear constraints in the filter design algorithm, in order to minimize the maximum amount of ripples in the vertical, horizontal and diagonal step response, as well as in the "quadrantal step" response, is described.

GF filters are completely characterized by the two 1-D filters $q_1(n)$ and $q_2(n)$ of (40), together with the sampling matrix $\mathbf{A}$. Hence, it seems of interest to examine the relationships between the step response characteristics of $q_1(n)$ and $q_2(n)$ and that of the resulting 2-D filter.

We show in the following that for two suitable 2-D steps, the output of filter $h(\mathbf{a})$ is characterized by the step responses of the decimated versions of $q_1(n)$ and $q_2(n)$. To this purpose, let $\mathbf{C} = (\mathbf{c}_1|\mathbf{c}_2)$ be a basis of the filter definition lattice, and consider function

$$S_{\mathbf{c}_2}(k) = \sum_{n_1=-\infty}^{k} \sum_{n_2=-\infty}^{\infty} h(n_1\mathbf{c}_1 + n_2\mathbf{c}_2) \tag{73}$$

where $S_{\mathbf{c}_2}(k)$ represents the response of the filter to a 2-D step oriented along $\mathbf{c}_2$, "read" along $\mathbf{c}_1$. This is a good characterization of the step response, because $S_{\mathbf{c}_2}(k)$ actually "spans" all the values of the output of the filter relative to the 2-D step. In particular, we are interested in the position of the ripples, and in the amount of the difference between each overshoot and the next undershoot. Such parameters, determined by $S_{\mathbf{c}_2}(k)$, are expected to represent a good measure of the "annoyance" of the ringing. We will show in the following how to choose orientations for the 2-D step so that $S_{\mathbf{c}_2}(k)$ can be related to the step responses of two decimated versions of the two 1-D filters.

We can get some insight into $S_{\mathbf{c}_2}(k)$ by considering together (41), (43) and (73):

$$\begin{aligned} S_{c_2}(k) &= |\det(\mathbf{A})| \sum_{n_1=-\infty}^{k} \sum_{n_2=-\infty}^{\infty} \bar{h}(n_1\mathbf{a}_1 + n_2\mathbf{a}_2) \\ &= |\det(\mathbf{A})| \sum_{n_1=-\infty}^{k} \sum_{n_2=-\infty}^{\infty} q_1(n_1A_{1,1} + n_2A_{1,2}) \\ &\quad \cdot q_2(n_1A_{2,1} + n_2A_{2,2}) \end{aligned} \tag{74}$$

where $(\mathbf{a}_1|\mathbf{a}_2) = \mathbf{A}$. Now, suppose that $\mathbf{C}$ has been chosen so that $\mathbf{A}$ is in lower Hermite normal form (see Remark 2 in Section III-A); then $A_{1,2} = 0$ and

$$S_{c_2}(k) = |\det(\mathbf{A})| \sum_{n_1=-\infty}^{k} \left(q_1(n_1A_{1,1}) \sum_{n_2=-\infty}^{\infty} q_2(n_1A_{2,1} + n_2A_{2,2}) \right) \tag{75}$$

Consider an $A_{2,2}$–polyphase decomposition of $q_2(n)$ (with "$\lfloor x \rfloor$" we denote the largest integer less than x):

$$q_2(n) = q_2^{s(n)}(\lfloor n/A_{2,2} \rfloor) \tag{76}$$

where $s(n) = n \bmod A_{2,2}$ and $q_2^s(n) = q_2(A_{2,2}n + s)$. Then, calling

$$Q_2^r = \sum_{n=-\infty}^{\infty} q_2^r(n) \tag{77}$$

and

$$r(n) = (nA_{2,1}) \bmod A_{2,2} \tag{78}$$

we can rewrite (75) as

$$S_{c_2}(k) = |\det(\mathbf{A})| \sum_{n=-\infty}^{k} q_1(nA_{1,1}) Q_2^{r(n)}. \tag{79}$$

Equation (79) shows that the output of filter $h(\mathbf{a})$ to a 2-D step oriented along $\mathbf{c}_2$, "read" along points $k\mathbf{c}_1$, coincides with the step response of the 1-D filter with impulse response $q_1(nA_{1,1})Q_2^{r(n)}$. If function Q_2^r is constant with respect to r, we have the important result that the 2-D step response is the step response of a decimated version of $q_1(n)$, times a multiplicative constant depending on $q_2(n)$. The following equality holds:

$$\begin{aligned} A_{2,2}Q_2^r = Q_2(0) + 2 \sum_{k=1}^{\lfloor (A_{2,2}-1)/2 \rfloor} Q_2\left(\frac{k}{A_{2,2}}\right) \cos\left(2\pi \frac{kr}{A_{2,2}}\right) \\ + \begin{cases} Q_2(0.5), & \text{even } A_{2,2} \\ 0, & \text{odd } A_{2,2} \end{cases} \end{aligned} \tag{80}$$

where $Q_2(f)$ is the Fourier transform of $q_2(n)$. Hence, a sufficient condition to have constant Q_2^r is

$$Q_2(k/A_{2,2}) = 0 \text{ for } k > 0. \tag{81}$$

If such a case is verified, the behavior of the considered 2-D step response is determined only by $q_1(n)$. Note that as long as the stopband frequency of $Q_1(f)$ is smaller than $0.5/A_{1,1}$, it is reasonable to assume that the peaks of the step response of filter $q_1(n)$ do not differ "too much" from the peaks of the step response of its decimated version $A_{1,1}q_1(A_{1,1}n)$ [6].

Completely similar considerations hold for the filter response to a 2-D step oriented along $\hat{\mathbf{c}}_1$, where $(\hat{\mathbf{c}}_1|\hat{\mathbf{c}}_2) \stackrel{\text{def}}{=} \hat{\mathbf{C}}$ and $\hat{\mathbf{C}} = \mathbf{CU}$ is a basis of Λ such that the corresponding subsampling matrix $\hat{\mathbf{A}} = \mathbf{AU}$ is in upper Hermite normal form.

Finally, let us recall that, for an ideal 1-D FIR filter, the risetime is inversely proportional to its passband frequency f_p, while the overshoot goes from 6.8% to 9% as f_p goes from 0.25 to 0.

D. Frequency Response Constraints

If the filter to be designed is part of a sampling structure converter, it is important to impose some constraints on its frequency response. For example, if the filter $H(\mathbf{f})$ is expected to cancel the undesired spectral repetitions occurring when up-sampling from lattice $\Gamma = LAT(\mathbf{CH})$ to lattice $\Lambda = LAT(\mathbf{C})$ ($\Gamma \subset \Lambda$, integral $\mathbf{H}$), $H(\mathbf{f})$ should be vanishing for $\mathbf{f} \in \Gamma^* \setminus \Lambda^*$ [11], [12], so that the aliasing due to flat brightness areas is removed. It is shown in the following how such a constraint on the 2-D frequency response can be converted into constraints on the two 1-D filters $q_1(n)$ and $q_2(n)$ used in the design of the GF filter, as described in Section III-A.

Using the notation of Section III-A, from (45) a sufficient condition for $H(\mathbf{f})$ to be null for some $\mathbf{f} = \bar{\mathbf{f}}$ is

$$\bar{H}\big(\mathbf{A}^{-T}(\mathbf{C}^T\bar{\mathbf{f}} + \mathbf{r})\big) = 0\ ,\ \mathbf{r} \in Z^2. \tag{82}$$

In our case of interest, $H(\mathbf{f})$ should vanish for

$$\mathbf{f} \in \Gamma^* \setminus \Lambda^* \tag{83}$$

i.e., for

$$\mathbf{f} \in \big\{(\mathbf{CH})^{-T}\mathbf{n}\ ,\ \mathbf{n} \in Z^2 \setminus LAT(\mathbf{H}^T)\big\} \tag{84}$$

Substituting (84) for $\bar{\mathbf{f}}$ in (82), one gets the corresponding constraint on $\bar{H}(\mathbf{f})$:

$$\bar{H}\big(\mathbf{A}^{-T}(\mathbf{H}^{-T}\mathbf{n} + \mathbf{r})\big) = 0\ , \mathbf{n} \in Z^2 \setminus LAT(\mathbf{H}^T)\ ,\ \mathbf{r} \in Z^2 \tag{85}$$

Hence, $\bar{H}(\mathbf{f})$ should vanish for $\mathbf{f}$ belonging to the set

$$\begin{aligned}&(LAT((\mathbf{AH}^{-T})) + LAT(\mathbf{A}^{-T}))\setminus\\ &(LAT(\mathbf{A}^{-T}) + LAT(\mathbf{A}^{-T}))\\ &= LAT((\mathbf{AH})^{-T}) \setminus LAT(\mathbf{A}^{-T}).\end{aligned} \tag{86}$$

Since $\bar{H}(\mathbf{f})$ is periodic on Z^2, condition (85) is equivalent to

$$\bar{H}(\mathbf{f}) = 0\ ,\ \mathbf{f} \in \mathcal{Z} \stackrel{\text{def}}{=} P \setminus LAT(\mathbf{A}^{-T}) \tag{87}$$

where P is a Z^2–period of $LAT(\mathbf{AH})^{-T}$. To have $\bar{H}(\mathbf{f})$ vanishing in the $|\det(\mathbf{A})|(|\det(\mathbf{H})| - 1)$ points of $\mathcal{Z}$, one should set $Q_1(f_1) = 0$ and/or $Q_2(f_2) = 0$ for $(f_1, f_2)^T \in \mathcal{Z}$. We will call such points the *nulling frequencies* for the 1-D filters. It is easily seen that, given set $\mathcal{Z}$, there exist several combinations of nulling frequencies for $Q_1(f)$ and $Q_2(f)$, such that (87) is satisfied. For example, for each $\mathbf{f}^i = (f_1^i, f_2^i)^T \in \mathcal{Z}$, one could set

$$Q_j(f_j^i) = 0 \quad \text{if}\ \ f_j^i > f_{p_j} \quad j = 1, 2. \tag{88}$$

Note that we do not want to set to zero $H(\mathbf{f})$ in "useful" frequencies; such an occurrence is avoided by the condition in (88). Nevertheless, following the simple criterion expressed by (88), $H(\mathbf{f})$ may be forced to zero also for some frequencies in $LAT(\mathbf{A}^{-T})$, which is not required by (87). As a matter of fact, $\mathcal{Z}_1$ and $\mathcal{Z}_2$ (the sets of nulling frequencies for $Q_1(f)$ and $Q_2(f)$) obtained from (88), are not the ones with minimum cardinality, among those satisfying the nulling constraint (87) (i.e., such that $\mathcal{Z} \subset \mathcal{Z}_1 \times \mathcal{Z}_2$). Such a characteristic is important, because, as is well known, adding constraints to the frequency response of 1-D filters typically increases the minimum filter length required to achieve desired specifics. Thus we are led to seek for sets $\mathcal{Z}_1$ and $\mathcal{Z}_2$ with smaller cardinality. To this purpose, we can exploit the following observation. If $\mathcal{Z}$ contains two or more points $\{\mathbf{f}^1, \mathbf{f}^2, \cdots\}$ with the same component $f_j^1 = f_j^2 = \cdots \stackrel{\text{def}}{=} \bar{f}_j$, then the nulling condition is satisfied for all of them simply by setting $Q_j(\bar{f}_j) = 0$.

The previous argument suggests the following procedure to find "good" sets $\mathcal{Z}_1$ and $\mathcal{Z}_2$ for a given set $\mathcal{Z}$.

1) For each point $(f_1^i, f_2^i)^T$ of $\mathcal{Z}$ such that $f_1^i < f_{p_1}$ or $f_2^i < f_{p_2}$, set $Q_2(f_2^i) = 0$ or $Q_1(f_1^i) = 0$ respectively. Take these points out of $\mathcal{Z}$, and call $\tilde{\mathcal{Z}}$ the set of the remaining points.
2) Determine all the clusters $\{C_1^i\}$ and $\{C_2^l\}$ of points of $\tilde{\mathcal{Z}}$ having common component f_1^i or f_2^l, respectively. In order to find such clusters, one can start from the LDFL containing $LAT((\mathbf{AH})^{-T})$, determine (trivially) the points of it within a rectangular unit cell of $LAT(\mathbf{A}^{-T})$, and for each row (or each column) of the resulting set determine which points belong to $LAT((\mathbf{AH})^{-T})$.
3) Find a "minimum cost covering" of $\tilde{\mathcal{Z}}$ by elements of $\{C_1^i\}$ and $\{C_2^l\}$, i.e., a couple of sets $\tilde{\mathcal{Z}}_1 = C_1^{i_1} \cup C_1^{i_2} \cup \cdots \cup C_1^{i_I}$ and $\tilde{\mathcal{Z}}_2 = C_2^{l_1} \cup C_2^{l_2} \cup \cdots \cup C_2^{l_L}$ with minimum cardinality, such that $\tilde{\mathcal{Z}} \subset \tilde{\mathcal{Z}}_1 \times \tilde{\mathcal{Z}}_2$.
4) Set $Q_1(f) = 0$ for $f \in \tilde{\mathcal{Z}}_1$ and $Q_2(f) = 0$ for $f \in \tilde{\mathcal{Z}}_2$.

Note, that in general, there exist more than one minimum cost covering of $\tilde{\mathcal{Z}}$ by clusters of $\{C_1^i\}$ and $\{C_2^l\}$. In order to design 1-D filters with frequency response constrained to zero in the chosen frequencies, one may use linear programming [22] or the Parks–McClellan algorithm [18] for minimax filters.

As an example, consider the following case: $\mathbf{A}^{-T} = \mathbf{I}$, $(\mathbf{AH})^{-T} = \begin{pmatrix} 1/4 & 1/12 \\ 0 & 1/6 \end{pmatrix}$, $f_{p_1} = 1/24, f_{p_2} = 1/8$. Fig. 5 represents lattice $LAT((\mathbf{AH})^{-T})$, together with the unit cell $\mathcal{R}(1/2, 1/2)$ of $LAT(\mathbf{A}^{-T})$ (continuous line) and the pass-band region of $\bar{H}(\mathbf{f})$ (dashed line). The set $\mathcal{Z}$ corresponds to the set of points of $LAT((\mathbf{AH})^{-T})$ within $\mathcal{R}(1/2, 1/2)$, excluding the origin. A minimum cost covering of $\mathcal{Z}$ is represented by the clusters of points within the regions contoured by dotted lines. Filter $Q_1(f)$ (corresponding to the horizontal axis) is forced to zero for $f \in \{1/4, 1/2\}$, while filter $Q_2(f)$ is forced to zero for $f \in \{1/6, 1/3, 1/2\}$.

V. Conclusions

In this paper, we have presented an analysis of GF filters, which pushes forward the work of Chen and Vaidyanathan [1]–[3]. GF filters are appealing because they can be implemented efficiently, and because they are easy to design and to analyze, as they are built starting from 1-D prototypes. We have described here the relations among minimax filter parameters, the symmetries in the impulse response which

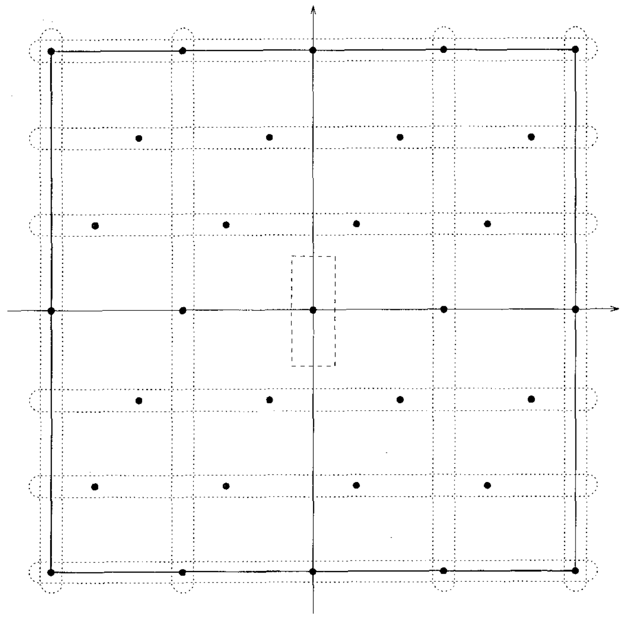


Fig. 5. Example of nulling constraints, relative to the up-sampling from lattice $LAT(\mathbf{AH})$ to lattice $LAT(\mathbf{A}) = Z^2$ (see text). The dots denote the points of lattice $LAT((\mathbf{AH})^{-T})$, the region contoured by solid line is the unit cell $\mathcal{R}(0.5, 0.5)$ of Z^2, the region contoured by dashed line is the passband region of $\bar{H}(\mathbf{f})$, and the regions contoured by dotted lines denote the clusters of points corresponding to the nulling frequencies of the two 1-D filters.

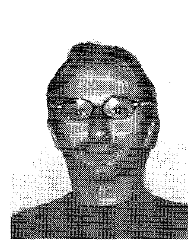
derive from zero-phase symmetry in the 1-D prototypes, the characteristics of the 2-D step response, and how to translate constraints on the 2-D frequency response into constraints on the 1-D prototypes.

ACKNOWLEDGMENT

The author wishes to thank the anonymous referees for their patient reading and valuable comments. T. Chen of Bell Labs is kindly acknowledged for his useful comments, and for correcting the original draft.

REFERENCES

[1] T. Chen and P. P. Vaidyanathan, "Multidimensional multirate filters derived from one-dimensional filters," *Electron. Lett.*, vol. 27, no. 3, pp. 225–228, Jan. 1991
[2] ——, "Recent developments in multidimensional multirate systems," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 3, pp. 116–137, Apr. 1993.
[3] ——, "Multidimensional multirate filters and filter banks derived from one–dimensional filters," *IEEE Trans. Signal Processing*, vol. 41, pp. 1749–1765, May 1993.
[4] M. Renfors, "Multi-dimensional sampling structure conversion with one-dimensional Nth-band filters," in *Proc. IEEE ISCAS'89*, Portland, OR, pp. 1502–1507.
[5] G. M. Cortelazzo, G. A. Mian, and C. Monti, "An effective technique to design hexagonal lattice FIR filters," in *Digital Signal Processing–91*, Firenze, Italy, pp. 25–29.
[6] R. Manduchi, "Two–dimensional interpolated IFIR filters using generalized factorable filters," Univ. of California at Berkeley, 1993, Rep. no. UCB/CSD/93/785.
[7] J. F. Kaiser, "Nonrecursive digital filter design using IO–sinh window function,"in *Proc. IEEE ISCAS'74*, Apr. 1974, pp. 20–23.
[8] O. Herrmann, L. R. Rabiner, and D. S. K. Chan, "Practical design rules for optimum finite impulse response low-pass digital filters," *Bell Syst. Tech. J.*, vol. 52, no. 6, pp. 769–799, July 1973.
[9] D. E. Dudgeon and R. M. Mersereau, *Multidimensional Digital Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1984
[10] P. Carrai, P. and G. M. Cortelazzo, and G. A. Mian, "Characteristics of minimax FIR filters for video interpolation/decimation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 4, pp. 453–467, Oct. 1994.
[11] Knoll, A., "Filter design for the interpolation of highly subsampled pictures," *Signal Processing: Image Commun.*, vol. 3, pp. 239–248, 1991.
[12] R. Manduchi, G. M. Cortelazzo, and G. A. Mian, "Multistage sampling structure conversion of video signals," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 3, pp. 325–340, Oct. 1993.
[13] M. Newman, *Integral Matrices*. New York: Academic, 1972.
[14] E. Dubois, "The sampling and reconstruction of time-varying imagery with applications in video systems," *Proc. IEEE*, vol. 73, pp. 502–522, Apr. 1985.
[15] G. M. Cortelazzo and R. Manduchi, "On the determination of all the sublattices of preassigned index and its application to multidimensional subsampling," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 3, pp. 318–320, Aug. 1993.
[16] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
[17] R. Bernardini and R. Manduchi, "On the reduction of multidimensional DFT to separable DFT by Smith normal form theorem," *Euro. Trans. Telecommun. Related Technol.*, vol. 5, no. 3, pp. 377–380, May–June 1994.
[18] A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1975.
[19] J. H. Lodge and M. M. Fahmy, "K-cyclic symmetries in multidimensional sampled signals," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-31, pp. 847–860, Aug. 1983.
[20] G. A. Reitmeir, "The effect of analog filtering on the picture quality of component digital television systems," *SMPTE J.*, pp. 949–955, Oct. 1981.
[21] V. Ouvrard and P. Siohan, "Design of two-dimensional video filters with spatial constraints," in *Proc. EUSIPCO 92*, Brussels, Belgium, Aug. 1992.
[22] L. Rabiner and B. Gold, *Theory and Application of Digital Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1975.



Roberto Manduchi received the Ph.D. degree in electrical engineering from the University of Padova, Italy.

He worked as a Post-Doctoral Fellow at the University of California at Berkeley, at the California Institute of Technology, Pasadena, and at Stanford University, Stanford, CA, doing research on computer vision and signal processing.