one can expect to identify p_{01}. For suppose $dn_{it} = p_{oi}\phi_{it}\,dt + dm_{it}$, $i = 1,2$, and let $n_t = n_{1t} + n_{2t}$. Then eventually all the observations of n_t are almost entirely those of n_{2t}, which does not yield much information about p_{o1}. Indeed C now becomes $\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$. Similarly, if $\lim_{t\to\infty}\phi_{1t}/\phi_{2t} = c \in (0,\infty)$, one can only expect to identify $cp_{01} + p_{02}$.

It might be difficult to check assumptions 2 and 3 of Theorem 2. Assumption 1 will in general be easy to verify. A sufficient condition for assumptions 1 and 2 to hold is, for example, $\phi_t \sim t^a$ $(a > -1/2)$. A necessary condition for assumption 3 is that the eigenvalues of $\int_0^t \phi_s\phi_s^T\,ds$ are of the same order as $t \to \infty$. Assumption 3 is similar to the notion of persistence of excitation that appears in identification problems for ARMAX systems.

Condition 3 of the theorem appears as a technical condition, necessary for the proof of Theorem 2. It seems, however, to be related to

$$\lim_{t\to\infty} \frac{1}{p_0^T\Phi_t}\int_0^t \frac{\phi_s\phi_s^T}{p_0^T\phi_s}\,ds > 0 \quad \text{almost surely} \tag{6.1}$$

where $\Phi_t = \int_0^t \phi_s\,ds$. Here (6.1) has an appealing interpretation. To see this, define a normalized version of (3.1) by

$$H_t(p) = \frac{1}{p_0^T\Phi_t}J_t(p). \tag{6.2}$$

Then minimization of $H_t(\cdot)$ is equivalent with minimization of $J_t(\cdot)$. One can easily check that for large t $H_t''(p)|_{p=p_0}$ can be approximated by (6.1). Hence (6.1) says that for $t\to\infty$ p_0 is indeed a minimum point of $H_t(\cdot)$.

We have not discussed the asymptotic distribution of the estimates $\hat{p}_t$ generated by (3.2) and (3.3). This issue will be addressed in another publication.

References

[1] P. Brémaud, *Point Processes and Queues*. New York: Springer, 1981.

[2] C. Dellacherie and P.A. Meyer, *Probabilités et potentiel*. Paris, France: Hermann, 1980, chs. V–VIII.

[3] Yu. Lin'kov, "Estimates of parameters of counting processes," *Prob. Inform. Transmission*, vol. 18, pp. 63–76, 1982.

[4] Y. Ogata, "The asymptotic behavior of maximum likelihood, estimators for stationary point processes," *Ann. Inst. Statist. Math.*, vol. 30, pt. A, pp. 243–261, 1978.

[5] H. Robbins and D. Siegmund, "A convergence theorem for nonnegative almost supermartingales and some applications, " in *Optimizing Methods in Statistics*, J.S. Rustagi, Ed. New York; Academic, 1971, pp. 233–256.

[6] J.H. van Schuppen, "Convergence results for continuous-time adaptive stochastic filtering algorithms," *J. Math. Anal, Appl.*, vol. 96, pp. 209–225, 1983.

[7] V. Solo, "The convergence of AML," *IEEE Trans. Automat. Contr.*, vol. AC-24, pp. 958–962, 1979.

Statistical and Computational Performance of a Class of Generalized Wiener Filters

MARK G. KARPOVSKY, SENIOR MEMBER, IEEE, AND LAZAR A. TRACHTENBERG, MEMBER, IEEE

Abstract—**A class of suboptimal Wiener filters is considered, and their computational and statistical performances (and the trade-off between the two) are studied and compared with those for known classes of suboptimal Wiener filters. A general model of a suboptimal Wiener filter over a group is defined, which includes, as special cases, the known filters based on the discrete Fourier transform (DFT) in the case of a cyclic group and the Walsh–Hadamard transform (WHT) in the case of a dyadic group. Statistical and computational performances of various group filters are investigated. The cyclic and the dyadic group filters are known to be computationally the best ones among all the group filters. However, they are not always the best ones statistically and other (not necessarily Abelian) group filters are studied. Results are compared with those for the cyclic group filters (DFT), and the general problem of selecting the best group filter is posed. That problem is solved numerically for small-size signals (≤ 64) for the first-order Markov process and random sine wave corrupted by white noise. For the first-order Markov process with the covariance matrix $B^{(s,l)} = \rho^{|s-l|}$ as ρ increases, the use of various non-Abelian groups results in improved statistical performance of the filter as compared to the DFT. Similarly, for the random sine wave with covariance matrix $B^{(s,l)} = \cos\lambda(s-l)$ as λ decreases, non-Abelian groups result in a better statistical performance of the filter than the DFT does. However, that is compensated for by the increased number of computations to perform the filtering.**

Manuscript received October 31, 1983; revised August 8, 1985.

This work was supported in part by the National Science Foundation under Grants DCR-8317763 and ECS-8512748. This work was presented in part at the International Workshop on Fault Detection and Spectral Techniques, Boston, MA, October 12–14, 1983.

M. G. Karpovsky is with the College of Engineering, Boston University, Boston, MA 02215.

L. A. Trachtenberg is with the Department of Electrical and Computer Engineering, Drexel University, Philadelphia, PA 19104.

IEEE Log Number 8406619.

I. Introduction

In recent years interest has grown in utilizing orthogonal transforms in digital signal processing in order to improve statistical or computational performance to permit a trade-off between these two criteria by utilizing a certain chosen orthogonal transform [1], [3], [7], [14].

A common quality shared by many fast transforms which enables their classification (see, e.g., [4], [5]) is that they can be represented as Kronecker products of matrices which may or may not be sparse or structured. By virtue of this Kronecker product representation new transforms can be generated from old ones simply by using the Kronecker product. In a given problem, such as Wiener filtering with given statistical characteristics of a signal and noise, one can *select* a computationally good approximating transform to a statistically optimal transform and the selection can be done out of the family of known fast transforms with a Kronecker product representation. (See [10], where a good reference list can be found, and [1].)

Another approach to the same problem of Wiener filtering would be to *construct* a computationally good approximation to a given statistically optimal transform. A possibility of solving that problem analytically for classes of signals defined by their covariance matrices (e.g., for signals whose covariance matrices are Toeplitz) has been pointed out in [12], [18], [19], [28] and this approach deserves further elaboration. Yet another approach is to *construct experimentally* a computationally good approximating transform to a transform which is known to be good statistically. For example, the discrete cosine transform (DCT) has a nearly optimal statistical performance for highly correlated Markov signals (see [24]), and it has recently been approximated by computationally convenient transforms [8]. Here even for small n (up to 32 vector-components of a signal) the problem is difficult, involves tedious trial and error procedures, and requires artistry rather than clear-cut methods. Another disadvantage is that a success with approximating one transform (as DCT) for some n (say $n = 16$, 32) cannot be generalized to be used to approximate other transforms [8], [26].

A number of researchers [15], [1], [3], [11], [17] have selected a family of fast transforms which are group theoretic by their nature; i.e., they are based on group characters of corresponding Abelian groups: examples are the discrete Fourier transform (DFT) in the case of a cyclic group and the Walsh–Hadamard transform (WHT) (or simply the Walsh transform) in the case of a dyadic group [1], [3], [11], [15]–[17], [27]. The use of non-Abelian groups was discussed in [13], [20].

These transforms exist for any number n, are computed analytically by formulas, and possess Kronecker product representations (which guarantee speed of computation for nonprime

n's). Also, their statistical performances can be computed by formulas, and as will be shown later, in many cases they compare favorably with other transforms. The cyclic and dyadic groups are not always the best to use [3], [7], and attempts have been made to develop a general theory of approximation of systems and signals by those over groups. Elements of this theory as presented in [13], [14] embrace problems in digital signal processing such as filtering, encoding, data compression, etc. (e.g., see [3]).

In the present work we apply methods of this theory to the well studied problem of Wiener filtering. The applications of representations of noncommutative as well as Abelian groups are the subject of our investigation. The results are compared with those in the current literature, and it will be shown (see Sections II and III) that the DFT and the WHT are often not the best group transforms to use for suboptimal Wiener filtering. Using the examples of random sine and first-order Markov processes, we shall show that the use of a noncommutative group may be more advantageous than the use of a commutative group because they yield a better approximation for a given speed (see also [13, sec. 5]).

The correspondence comprises three sections. The suboptimal Wiener filtering problem is treated in Section II as that of constructing the best group Wiener filter approximation to a given classical Wiener filter over a given group. The general solution of this problem is presented and previously studied cases of cyclic and dyadic group filters are deduced. The results of computer experiments are given and analyzed in Section III for stochastic sine wave and first-order Markov processes.

II. Generalized Wiener Filtering over Groups

We consider the set $\mathscr{G} = \{0, 1, \cdots, n-1\}$ of n integers without a group structure being imposed on it. Let u, e be random functions of signal and noise, respectively, defined on $\mathscr{G}$ and taking values in the field of complex numbers C; that is, $u, e: \mathscr{G} \to \mathbb{C}$ with the mean values $E(u(t)) = E(e(t)) = 0$, $t \in \mathscr{G}$. Let $w: \mathscr{G} \times \mathscr{G} \to \mathbb{C}$ be a deterministic impulse response function of a digital device (a filter) which is processing the corrupted signal $f = u + e$. Then the classical Wiener filter problem of separating the signal from noise is to find $w_{\text{opt}}: \mathscr{G} \times \mathscr{G} \to \mathbb{C}$ such that

$$\min_{w} \{ D(\epsilon_w) \} = D_{\text{opt}} \tag{1}$$

where $\epsilon_w(t) = \sum_{\zeta \in \mathscr{G}} w(t, \zeta) f(\zeta) - u(t)$, $t \in \mathscr{G}$; $D(\epsilon_w) = \sum_{t \in \mathscr{G}} D(\epsilon_w(t))$; $D(\epsilon_w(t)) = E(\epsilon_w(t)\epsilon_w(t))/n$ is the dispersion squared which depends upon the choice of impulse response function $w(\cdot, \cdot)$ of the filter; and E represents the expectation operator. Using the notation, $u = (u(0), u(1), \cdots, u(n-1))^T$ and $e = (e(0), e(1), \cdots, e(n-1))^T$ where T stands for the transpose of a row vector, we reformulate the same problem (1) of the optimal filtering of the zero-mean n-vector of noise e from the corrupted, zero-mean signal $f = u + e$ by means of the $(n \times n)$-matrix of the optimal filter $W_{\text{opt}} = (W_{\text{opt}}^{(s,l)})$, $W_{\text{opt}}^{(s,l)} = w_{\text{opt}}(s, l)$, $s, l = 0, 1, \cdots, n-1$ such that

$$\min_{W} \{ 1/nE(\|Wf - u\|) \} = 1/nE(\|W_{\text{opt}} f - u\|), \tag{1'}$$

where

$$1/nE(\|\epsilon_w\|^2) \triangleq 1/nE(\operatorname{tr}(\epsilon_w \epsilon_w^*)) = \sum_{t \in \mathscr{G}} D(\epsilon_w(t)) = D(\epsilon_w).$$

Let B_{uu}, B_{ee}, and B_{ue} be the covariance matrices of the signal and noise which may be either stationary or nonstationary. Then the least square error ϵ between the signal u and its filtered estimate $f \triangleq u + e$ occurs in (1) and (1′) when

$$W_{\text{opt}} B_{ff} = B_{uf} \tag{2}$$

where

$$B_{ff} = B_{uu} + B_{ue} + B_{eu} + B_{ee},$$
$$B_{uf} = B_{uu} + B_{ue}.$$

In that case

$$D_{\text{opt}} = 1/n \operatorname{tr}(B_{uu} - W_{\text{opt}} B_{fu}). \tag{3}$$

The case of uncorrelated signal and noise has been discussed in [15], [3], [17].

The generalized Wiener filter utilizes a unitary transform which is represented by the matrix T in Fig. 1 (see [15], [3]). D_{opt} in (3) is independent of T. The direct Wiener filtering (when T is the identity matrix) is fastest as it requires at most n^2 multiplications and additions. The Karhunen–Loeve transform which uses the T that diagonalizes W_{opt} is the slowest as it requires on the order of $2n^2 + n$ operations to perform the filtering $W_{\text{opt}} f = T^{-1}(TW_{\text{opt}}T^{-1})Tf$, where $TW_{\text{opt}}T^{-1}$ is the diagonal matrix of eigenvalues of W_{opt} and multiplications by vectors are being performed from the right.

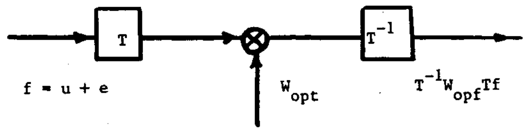

Fig. 1. Generalized Wiener filter.

However, the mathematically optimal D_{opt} in (3) is never achievable in reality (e.g., because of roundoff). We must therefore accept some degradation in performance. The idea in the suboptimal Wiener filter (see [15], [17], [3]) is to allow an acceptable performance degradation in filtering while increasing the computational performance of the filter. That is achieved by using a fast unitary transform for which a procedure exists to simplify the matrix $TW_{\text{opt}}T^{-1}$ by, e.g., making it sparse or structured. The statistical performance $D(\epsilon)$ must be kept within acceptable limits. For example (see [3], [15]–[17]), a family of suboptimal Wiener filters has been obtained by setting all the off-diagonal entries in $TW_{\text{opt}}T^{-1}$ to zero, where T's fast computational algorithm is based on it having a Kronecker product representation (such transforms are classified in [4], [5]). The WHT and DFT which have been used in [1], [3], [15]–[17] are matrices T of characters of the dyadic group and of the cyclic group, respectively. We note that the choice of a dyadic or of a cyclic group G is, generally speaking, not optimal (see [3], [7]), and moreover, replacement of a dyadic or of a cyclic group by some (not necessarily Abelian) group may result in the reduction of $D(\epsilon)$ (see the next section and [13, 26]). Such group transforms exist for any integer n, can be computed analytically by formulas, possess Kronecker product representations [20], [21], [23], and as will be shown later, in many cases compare favorably with other transforms. Therefore, attempts have been made to develop a general approximation theory of systems and signals over groups. Elements of this theory are presented in [13], [14] and embrace approximation problems in digital signal processing such as filtering, encoding, data compression, etc. (see also [1], [3]).

In what follows, we apply ideas and methods of that theory to construct optimal filters over an arbitrary group. Let us introduce a group structure in $\mathscr{G} = \{0, 1, \cdots, n-1\}$. That is, G stands for the group of cardinality n defined on the underlying index set $\mathscr{G}$ with 0 as group identity and with o as the group operation. Let $\mathbb{C}$ be the field of complex numbers. In the space $\{f: G \to \mathbb{C}\}$ the elements of the nonequivalent unitary irreducible representations of G over the field $\mathbb{C}$ will be used as an orthogonal basis [20]–[23]. That is, the representation ω of degree d_ω over $\mathbb{C}$ is a unitary $(d_\omega \times d_\omega)$-matrix. The value of representation ω at the point $t \in G$ will be denoted by $[\omega, t]$, and the functions generated by $[\omega, t]$ when ω and t are fixed will be denoted by $[\omega, \cdot]$ and $[\cdot, t]$, respectively.

Let $\hat{G} = \{\omega\}$ denote the set of all nonequivalent unitary representations of G, indexed so that ω is of degree d_ω. $\hat{G}$ is the dual object for G. In the important case of Abelian groups, G

may be represented as a direct product of its cyclic subgroups:

$$G = H_1 \times \cdots \times H_r,$$

$$t \in G, \quad t = (t_1, \cdots, t_r), \quad t_l \in \{0, 1, \cdots, n_l - 1\},$$

n_l is a power of a prime number, the group operation is componentwise addition mod n_l, $l = 1, 2, \cdots, r$. In this case $d_\omega = 1$ for all $\omega \in \hat{G}$, $\hat{G} = \hat{H}_1 \times \cdots \times \hat{H}_r$, $\hat{G}$ is the multiplicative group of characters which is isomorphic to G and $\hat{H}_l$ is isomorphic to H_l, i.e., $\omega = (\omega_1, \cdots, \omega_r)$, $\omega_l \in \{0, 1, \cdots, n_l - 1\}$ and we have

$$[\omega, t] = \prod_{l=1}^{r} \exp(2\pi i \omega_l t_l / n_l), \qquad \omega, t \in \{0, 1, \cdots, n-1\},$$

$$\omega_l, t_l \in \{0, 1, \cdots, n_l - 1\}, \quad i = \sqrt{-1}, \quad n = \prod_{l=1}^{r} n_l. \quad (4)$$

If $n_1 = n_2 = \cdots = n_r$, then $[\omega, t]$ is known as the Chrestenson function and for $n_1 = n_2 = \cdots = n_r = 2$ it is known as the Walsh function [13], [23], [27].

Let $f: G \to \mathbb{C}$. It follows by the orthogonality relations for the n functions $\{[\omega, \cdot]^{(s,t)}\}$ $(s, t = 1, 2, \cdots, d_\omega)$ (see, e.g., [22], [23], [13]) that the Fourier transform $F_G: f \to \hat{f}$ on the group G may be defined as

$$\hat{f}(\omega) \triangleq d_\omega / n \sum_{t \in G} f(t)[\omega, t^{-1}], \qquad \omega \in \hat{G}. \quad (5)$$

Computation of Fourier F_G and inverse Fourier F_G^{-1} transforms can be done using fast algorithms, and it is based on the following representation of elements of G by the Kronecker product of matrices. Let G be a group, isomorphic to a direct product of some groups H_l, $l = 1, 2, \cdots, r$, $G = \prod_{l=1}^{r} H_l$. Then (see [22])

$$[\omega, t] = \bigotimes_{l=1}^{r} [\omega_l, t_l], \quad (6)$$

where $\omega_l \in \hat{H}_l$, $t_l \in H_l$. It was proved in [20], [21] that the computation of f or $\hat{f}$ requires $n \sum_{l=1}^{r} n_l$ multiplications and additions and n memory locations.

Properties of Fourier transform F_G such as linearity, translation of arguments, group convolution, Wiener–Khinchine, Plancherel, and Poisson theorems are valid (see [13]).

We will now treat the signal and noise u, e as centralized random functions defined on the group G, i.e., $u, e: G \to \mathbb{C}$ and $E(u(t)) = E(e(t)) = 0$, $t \in G$. The action of the group filter can now be described by group convolution:

$$\epsilon_h(t) = \sum_{\zeta \in G} h(\zeta^{-1} o t) f(\zeta) - u(t), \quad (7)$$

where $f = u + e$ and $h: G \to \mathbb{C}$ is the impulse response function of the group filter. Then $h_{\text{opt}}: G \to \mathbb{C}$ of the optimal group filter will be obtained by minimizing the dispersion

$$\min_{h: G \to \mathbb{C}} \{D(\epsilon_h)\} = D_0. \quad (8)$$

We note that the same results can be obtained by computing the optimal Wiener filter (2) and then choosing the group G of the given order n and constructing the best Hilbert–Schmidt approximation to W_{opt}. That is, the optimal group is the unique solution of the following minimization problem

$$\min_{H \in \text{cir}(G)} \|W_{\text{opt}} - H\| = \|W_{\text{opt}} - H_{\text{opt}}\| \quad (9)$$

where cir(G) is the set of all impulse response matrices of group filters defined as follows

$$\text{cir}(G) = \{H | H = (h(\zeta^{-1} o t)), \quad h: G \to \mathbb{C}\}. \quad (10)$$

The solution of that problem is considered in [13, 14]. Such optimal group filters (which are suboptimal models of classical Wiener filter (2)) have been considered for the cases of dyadic and cyclic groups G (see, e.g., [1], [3], [15]–[17]). By using other Abelian and noncommutative groups we maintain about the same speed of computation (see [13], [20], [21], [23]) and the approximation error may be decreased (see the next section).

Using the definition (5), the convolution theorem, and the linearity of the Fourier transform over G, we have from (7) that

$$\hat{\epsilon}_h(\omega) = n/d_\omega \hat{h}(\omega)\hat{f}(\omega) - \hat{u}(\omega), \qquad \omega \in \hat{G}. \quad (11)$$

Hence, using the Plancherel theorem, for the Fourier transform over G we obtain from (11)

$$D(\epsilon_h) = 1/n E(\|E_h\|^2) = 1/n \sum_{\omega \in \hat{G}} n/d_\omega E(\text{tr}\, \hat{\epsilon}_h(\omega) \hat{\epsilon}_h^*(\omega))$$

$$= \sum_{\omega \in \hat{G}} 1/d_\omega \, \text{tr}\big(n^2/d_\omega^2 \hat{h}(\omega) B_{\hat{f}\hat{f}}(\omega) \hat{h}^*(\omega) + B_{\hat{u}\hat{u}}(\omega)$$

$$- n/d_\omega \big(\hat{h}(\omega) B_{\hat{f}\hat{u}}(\omega) + B_{\hat{u}\hat{f}}(\omega)\hat{h}^*(\omega)\big)\big) \quad (12)$$

where

$$B_{\hat{f}\hat{f}}(\omega) \triangleq E(\hat{f}(\omega)\hat{f}^*(\omega))$$

$$= B_{\hat{u}\hat{u}}(\omega) + B_{\hat{u}\hat{e}}(\omega) + B_{\hat{e}\hat{u}}(\omega) + B_{\hat{e}\hat{e}}(\omega),$$

$$B_{\hat{u}\hat{f}}(\omega) \triangleq E(\hat{u}(\omega)\hat{f}^*(\omega)) = B_{\hat{f}\hat{u}}^*(\omega) = B_{\hat{u}\hat{u}}(\omega) + B_{\hat{u}\hat{e}}(\omega),$$

and

$$B_{\hat{u}\hat{u}}(\omega) \triangleq E(\hat{u}(\omega)\hat{u}^*(\omega)) = d_\omega^2/n^2 \sum_{\zeta, t \in G} B_{uu}^{(\zeta, t)}[\omega, \zeta^{-1} o t],$$

$$\omega \in \hat{G}, \quad (13)$$

and where B_{uu} is the covariance matrix of u.

Analogously, we have

$$B_{\hat{e}\hat{e}}(\omega) = d_\omega^2/n^2 \sum_{\zeta, t \in G} B_{ee}^{(\zeta, t)}[\omega, \zeta^{-1} o t], \qquad \omega \in \hat{G} \quad (14)$$

and

$$B_{\hat{u}\hat{e}}(\omega) = d_\omega^2/n^2 \sum_{\zeta, t \in G} B_{ue}^{(\zeta, t)}[\omega, \zeta^{-1} o t], \qquad \omega \in \hat{G}. \quad (15)$$

It follows that the minimal dispersion in (12) is achieved when

$$\hat{h}_{\text{opt}}(\omega) B_{\hat{f}\hat{f}}(\omega) = d_\omega / n B_{\hat{u}\hat{f}}(\omega), \qquad \omega \in \hat{G}. \quad (16)$$

Assuming $B_{\hat{f}\hat{f}}(\omega)$ to be invertible for every $\omega \in \hat{G}$, we have the following equation for the optimal group filter

$$\hat{h}_{\text{opt}}(\omega) = d_\omega / n B_{\hat{u}\hat{f}}(\omega) B_{\hat{f}\hat{f}}^{-1}(\omega), \qquad \omega \in \hat{G}. \quad (17)$$

It follows from (17), (12) that the dispersion squared that is achieved by utilizing the optimal group filter (17) is computed by

$$D_0 = \sum_{\omega \in \hat{G}} 1/d_\omega \, \text{tr}\big(B_{\hat{u}\hat{u}}(\omega) - B_{\hat{u}\hat{f}}(\omega) B_{\hat{f}\hat{f}}^{-1}(\omega) B_{\hat{f}\hat{u}}(\omega)\big). \quad (18)$$

The dispersion D_0 depends upon the choice of the group G, i.e., $D_0 = D_0^G$. We formulate the problem of selecting the optimal group (among all the groups of a given order n) minimizing D_0^G. This problem is difficult in the general case, and we shall restrict ourselves to special cases of the first-order Markov process and of the random sine wave corrupted by white noise. For these cases, different group filters will be investigated in the next section using computer experiments. All the computations in optimal group filters (17), (18) have been done using the corresponding algorithms of fast Fourier transforms [20], [21], [23].

III. Computer Experiments

In this section, we compare different group filters (17), (18) in the problem of filtering the first-order Markov process and the random sine wave corrupted by white noise. These group filters will be compared with known results [3], [15]–[17], [10], [6], [25] for filters based on the Karhunen–Loeve transform (KLT) which is known to be statistically optimal, that is, its matrix consists of eigenvectors of W_{opt} defined by (2). Other known transforms considered here are the DCT which is asymptotically equivalent to the KLT for the first-order Markov process [24] and the DFT which is represented by the matrix of characters of the cyclic group $G = C_n$ of integers $0, 1, \cdots, n-1$. We shall consider two nonAbelian groups, namely, S_3 (the symmetric group of the third order) and Q_2 (the quaternion group of the eighth order, and C_n (the cyclic group). Their duals $\hat{S}_3, \hat{Q}_2, \hat{C}_n$ are given in [22], [23]. We shall use the direct products $S_3 \times C_n, C_n \times S_3, Q_2 \times C_n, C_n \times Q_2, S_3 \times Q_2, Q_2 \times S_3, S_3 \times S_3, Q_2 \times Q_2$. The corresponding duals are computed by (6) using the Kronecker product property of group representations. The computation of dispersion squared D_0^G in (18) is done using the fast Fourier transform algorithm over the group G (see [20], [21], [23]). The number $n \sum_{l=1}^{r} n_l$ of operations needed to compute the Fourier transform (5) (see Section II) is the upper bound on the computational complexity. The actual number of operations depends upon the number of zeros in the elements of the dual $\hat{G}$ for a given group G. For example, for $G = S_3$ four zeros are among the elements of $\omega = 2$. Therefore, to compute (5) for $S_3 \times C_2$ we need not $12(6+2) = 96$ computer operations but only $12 \cdot 8 - 2 \cdot 4 = 88$ operations. Analogously, 16 zeros are among the elements of $\omega = 4 \in \hat{Q}_2$. Hence, e.g., to compute (5) for $Q_2 \times Q_2$ we need $64 \cdot 16 - 8 \cdot 16 - 8 \cdot 16 = 3 \cdot 2^8 = 768$ computer operations.

It was pointed out in the discussion of Fig. 1 in Section II (see also [15], [1], [3]) that when using a transform T, the filter $T^{-1}(TWT^{-1})Tf$ is performed in the following three steps, which determine the overall amount of computer operations:

1) Tf,
2) $(TWT^{-1})(Tf)$,
3) $T^{-1}((TWT^{-1})(Tf))$.

The numbers of computer operations required to perform generalized filtering in the cases of identity, Karhunen–Loeve, and DFT's are n^2 (step 2 only), $2n^2 + n$ (n^2 operations at steps 1, 3, and n operations to multiply a diagonal matrix TWT^{-1} by a vector Tf), $n^2 + 2n \log_2 n$, $n = 2^r (n \log_2 n$ at steps 1 and 3 and n^2 operations at step 2), respectively.

The suboptimal filtering results in reducing the number of operations at step 2 to the order of n at the expense of setting all the off-diagonal entries in TWT^{-1} to zero (as in the case of DFT, WHT, or any Abelian group based transform T) or by structuring TWT^{-1} to a canonical block diagonal form, uniquely determined by the group G (as in the case of a noncommutative group, see [13]).

Based on these considerations, the results of comparing different transforms (see [3], [9] for DFT and DCT) in suboptimal filtering from the point of view of the number of operations required are given in Table I. Various transforms are compared by their statistical performance in what follows.

TABLE I
NUMBER OF OPERATIONS REQUIRED FOR VARIOUS SUBOPTIMAL FILTERS

	Transform				
n	KLT	DCT	$\hat{C}_n$(DFT)	$\widehat{C_{n/8} \times Q_2}$	$\widehat{Q_2 \times Q_2}$
8	136	92	56	104	
16	528	252	144	272	
32	2080	652	352	736	
64	8256	1612	832	2048	1600

We consider the following signals:

1) the first-order Markov process with the covariance matrix

$$B_{uu} = \left(\rho^{|s-l|}\right), \qquad s, l = 0, 1, \cdots, n-1, \tag{19}$$

where $0 < p < 1$;

2) the random sine wave process $x(t) = \sin(\lambda t + \alpha)$ with phase α distributed uniformly on the segment $[0, 2\pi]$, which has covariance matrix

$$B_{uu} = (\cos \lambda (s - l)), \qquad s, l = 0, 1, \cdots, n-1. \tag{20}$$

The noise e is assumed to be white with identity covariance matrix and the signal and noise are assumed to be uncorrelated.

The KLT is computed for the first-order Markov process in [15]. In the case of the random sine wave, the matrix of eigenvectors of the corresponding W_{opt} in (2) (the KLT) was computed for each n. The DCT is asymptotically equivalent to the KLT of the Markov process (see [24]) and is defined in [6], [25].

The computer experiments for the first-order Markov process are summarized in Tables II and III. Their purpose is to compare statistical performance of suboptimal group filters (see Table II) as well as to compare the known results for the DFT with other (not necessarily Abelian) groups. It follows (see Tables II, III) that as ρ increases, the use of various nonAbelian groups results in improved statistical performance as compared with the DFT. That is, however, compensated for by the increased number of computations.

It follows also that D_0^G increases as ρ decreases, e.g., for $G = S_3 \times C_2$ see Table III. That happens because B_{uu} approaches the identity matrix as ρ decreases. It follows that as ρ increases, various noncommutative groups compare favorably with C_n. We note that the order of the groups G_1 and G_2 in their direct product affects the dispersion without affecting the computational speed; i.e., the number of operations required to compute the dispersion is the same. In the general case it is a difficult problem to select the optimal group among all the groups of a given order [13]. However, this problem can be solved numerically for moderate values of n.

The computer experiments for the random sine wave process are summarized in Table IV. It follows (see Tables I, IV) that as λ decreases, the use of various nonAbelian groups results in improved statistical performance as compared with the DFT. For example, for $n = 64$ and $\lambda = 0.01$ (or $\lambda = 0.05$) the statistical gain in 17.59 percent (or 20.95 percent) for $Q_2 \times Q_2$ as compared with C_{64}. That is, however, compensated for by an increase in speed of nearly 100 percent in the DFT as compared with the group $Q_2 \times Q_2$ (see Table I). It follows also that the dispersion D_0^G increases as λ increases. This happens because B_{uu} approaches the matrix all of whose entries are ones as λ decreases, i.e., correlation between u-components increases and the dispersion decreases. Alternatively, as λ increases, B_{uu} approaches the identity matrix and the dispersion increases. It follows that for small λ various noncommutative groups compare favorably statistically with the cyclic group C_n. We note that as λ increases, the group transforms compare favorably with the DCT; that is, the cyclic group C_n provides a better approximation to the KLT than does the DCT. It can be expected that the more a stochastic process differs from the first-order Markov (in the behavior of its covariance matrix), the worse the results of employing the DCT will be because DCT is equivalent to the KLT of the first-order Markov process [24]. At the same time, a great variety of fast group transforms can always be used to choose the best from using (17) and (18). Therefore, group filters might find their use in practical situations in which we do not know computationally good approximating transforms for the KLT for a given process.

IV. SUMMARY

Different group transforms have been compared among themselves and with the DFT in the problem of constructing a suboptimal Wiener filter. Computational and statistical performance of various filters was considered as well as the problem of trade-off between the two.

The best group model is constructed for a given group which performs the filtering using a fast group transform algorithm. The problem of choosing an optimal group among all the groups of the given order has been discussed. Its solution is very difficult in the general case, but for small n (up to 64) it can be solved numerically for given statistics of the signal and noise.

TABLE II
GROUP WITH THE OPTIMAL STATISTICAL PERFORMANCE FOR THE FIRST-ORDER MARKOV PROCESS

									n							
ρ	6	8	12	16	18	24	30	32	36	40	42	48	54	56	60	64
.9	S_3	Q_2	$C_2 \times S_3$	C_{16}	C_{18}	C_{24}	C_{30}	C_{32}	C_{36}	C_{40}	C_{42}	C_{48}	C_{54}	C_{56}	C_{60}	C_{64}
.99	S_3	Q_2	$C_2 \times S_3$	$C_2 \times Q_2$	$C_3 \times S_3$	$C_3 \times Q_2$	$C_5 \times S_3$	$C_4 \times Q_2$	$S_3 \times S_3$	$Q_2 \times C_5$	$S_3 \times C_7$	$S_3 \times Q_2$	$S_3 \times C_9$	$Q_2 \times C_7$	$S_3 \times C_{10}$	$Q_2 \times Q_2$

TABLE III
DISPERSION FOR THE GROUP FILTER WITH $G = S_3 \times C_2$

ρ	0.5	0.6	0.7	0.8	0.9	0.92	0.94	0.96	0.99
$D^{S_3 \times C_2}$	0.4545	0.4294	0.3944	0.3437	0.2628	0.2398	0.2128	0.1800	0.111

TABLE IV
GROUP WITH THE OPTIMAL STATISTICAL PERFORMANCE FOR THE RANDOM SINE WAVE

									n							
λ	6	8	12	16	18	24	30	32	36	40	42	48	54	56	60	64
.01	C_6 S_3	C_8 Q_2	C_{12} $S_3 \times C_2$	C_{16} $Q_2 \times C_2$	$S_3 \times C_3$	$S_3 \times C_4$	$S_3 \times C_5$	$Q_2 \times C_4$	$S_3 \times C_6$	$Q_2 \times C_5$	$S_3 \times C_7$	$S_3 \times C_8$	$S_3 \times C_9$	$Q_2 \times C_7$	$S_3 \times C_{10}$	$Q_2 \times C_3$ $Q_2 \times Q_2$
.05	C_6 S_3	Q_2	$C_2 \times S_3$	$C_2 \times Q_2$	$S_3 \times C_3$	$S_3 \times C_4$	$S_3 \times C_5$	$Q_2 \times C_4$	$S_3 \times S_3$	$Q_2 \times C_5$	$S_3 \times C_7$	$S_3 \times Q_2$	$S_3 \times C_9$	$Q_2 \times C_7$	$S_3 \times C_{10}$	$Q_2 \times Q_2$
.1	S_3	Q_2	$C_2 \times S_3$	$C_2 \times Q_2$	$S_3 \times C_3$	$S_3 \times C_4$	$S_3 \times C_5$	$C_4 \times Q_2$	$C_6 \times S_3$	$C_5 \times Q_2$	$C_7 \times S_3$	C_{48}	C_{54}	C_{56}	C_{60}	C_{64}

ACKNOWLEDGMENT

The authors wish to thank Mr. O. Zimmerman for conducting the computer experiments of Section III.

REFERENCES

[1] D. F. Elliott and K. R. Rao, *Fast Transforms: Algorithms, Analyses and Applications*. New York: Academic, 1983.

[2] H. C. Andrews, *Picture Processing and Digital Filtering*. New York: Springer-Verlag, 1975.

[3] N. Ahmed and K. R. Rao, *Orthogonal Transform for Digital Signal Processing*. New York: Springer-Verlag, 1975.

[4] B. J. Fino and V. R. Algazi, "A unified treatment of discrete fast unitary transforms," *SIAM J. Computing*, vol. 6, pp. 700–717, Dec. 1977.

[5] H. C. Andrews and K. L. Caspari, "Degrees of freedom and modular structure in matrix multiplication," *IEEE Trans. Comput.*, vol. C-19, pp. 16–25, 1970.

[6] N. Ahmed, T. Natarajan, and K. R. Rao, "Discrete cosine transform," *IEEE Trans. Comput.*, vol. C-23, pp. 90–93, 1974.

[7] J. Pearl, "Basis restricted transformations and performance measures for spectral representations," *IEEE Trans. Inform. Theory*, vol. IT-17, pp. 751–752, 1971.

[8] H. S. Kwak, R. Srinivasan, and K. R. Rao, "C-matrix transforms," submitted to *IEEE Trans. Acoust. Speech, Signal Processing*.

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

[10] K. R. Rao, J. G. K. Kuo, and M. A. Narasimhan, "Slant-Haar transform," *Int. J. Computer Math.*, sec. B, 7, pp. 73–83, 1979.

[11] E. A. Trachtenberg, "Fast Wiener filtering computation technique," in *Proc. 1979 Int. Symp. on the Mathematical Theory of Networks and Systems*, vol. 3, Delft, The Netherlands pp. 174–177.

[12] E. A. Trachtenberg, "Construction of fast unitary transforms which are equivalent to Karhunen-Loeve spectral representations, in *Proc. 1980 IEEE Int. Symp. on EMC*, Baltimore, MD, USA, pp. 376–379.

[13] M. G. Karpovsky and E. A. Trachtenberg, "Some optimization problems for convolution systems over finite groups," *Inform. Contr.*, vol. 34, pp. 227–247, 1977.

[14] J. Pearl, "Optimal dyadic models of time-invariant systems," *IEEE Trans. Comput.*, vol. C-24, 1975, pp. 598–602.

[15] W. K. Pratt, "Generalized Wiener filtering computation technique," *IEEE Trans. Comput.*, vol. C-21, pp. 636–641, 1972.

[16] J. Pearl, "Walsh processing of random signals," *IEEE Trans. Electromagn. Compat.*, vol. EMC-13, pp. 137–141, 1971.

[17] ——, "On coding and filtering stationery signals by discrete Fourier transform," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 229–232, 1973.

[18] ——, "Asymptotic equivalence of spectral representations," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-23, pp. 547–551, 1975.

[19] Y. Yemini and J. Pearl, "Asymptotic properties of discrete unitary transforms," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-1, pp. 366–371, 1979.

[20] M. G. Karpovsky, "Fast Fourier transforms on finite non-Abelian groups," *IEEE Trans. Comput.*, vol. C-26, pp. 1028–1030, 1977.

[21] P. J. Nicholson, "Algebraic theory of finite Fourier transforms," *J. Comput. Syst. Sci.*, pp. 524–547, 1971.

[22] L. Dornhoff, *Group Representation Theory*. New York: Marcel Dekker, 1971.

[23] M. G. Karpovsky and E. A. Trachtenberg, "Fourier transform over finite groups for error detection and error correction in computation channels," *Inform. Contr.*, vol. 40, pp. 335–358, 1979.

[24] M. D. Flickner and N. Ahmed, "A derivation for the discrete cosine transform," *Proc. IEEE*, vol. 70, pp. 1132–1134, 1982.

[25] M. Hamidi and J. Pearl, "Comparison of cosine and Fourier transforms of Markov-1 signals," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-24, pp. 428–429, 1976.

[26] E. A. Trachtenberg, "Systems over finite groups as suboptimal Wiener filters: A comparative study," presented at *1983 Int. Symp. on Mathematical Theory of Systems and Networks*, Beer-Sheva, Israel, 1983.

[27] M. G. Karpovsky, *Finite Orthogonal Series in the Design of Digital Devices*. New York: Wiley, 1976.

[28] E. A. Trachtenberg, "Construction of Group Transforms Subject to Several Performance Criteria," *IEEE Trans. Acoust. Speech, Signal Processing*, vol. ASSP-33, pp. 1521–1531, 1985.

Maximum Likelihood Estimation for Multivariate Mixture Observations of Markov Chains

B. H. JUANG, STEPHEN E. LEVINSON, SENIOR MEMBER, IEEE, AND M. M. SONDHI

***Abstract*—To use probabilistic functions of a Markov chain to model certain parameterizations of the speech signal, we extend an estimation technique of Liporace to the cases of multivariate mixtures, such as Gaussian sums, and products of mixtures. We also show how these problems relate to Liporace's original framework.**

Manuscript received September 10, 1984; revised July 12, 1985. This work was presented at the 1985 IEEE International Symposium on Information Theory, Brighton, England, June 24-28.

The authors are with Acoustics Research Department, Bell Laboratories, Murray Hill, NJ 07974.

IEEE Log Number 8406633.

0018-9448/86/0300-0307$01.00