ESD-TR-89-267

2

AD-A221 129

Technical Report
873

Signal Processing on Finite Groups

R.B. Holmes

27 February 1990

Lincoln Laboratory
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
LEXINGTON, MASSACHUSETTS

Prepared for the Department of the Army under Air Force Contract F19628-90-C-0002.

This report is based on studies performed at Lincoln Laboratory, a center for research operated by Massachusetts Institute of Technology. The work was sponsored by the U.S. Army Strategic Defense Command, Department of the Army under Air Force Contract F19628-90-C-0002.

The ESD Public Affairs Office has reviewed this report, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER

Hugh L. Southall

Hugh L. Southall, Lt. Col., USAF
Chief, ESD Lincoln Laboratory Project Office

(2)

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
LINCOLN LABORATORY

SIGNAL PROCESSING ON FINITE GROUPS

R.B. HOLMES
Group 32

TECHNICAL REPORT 873

27 FEBRUARY 1990

LEXINGTON MASSACHUSETTS

ABSTRACT

A unified approach to the design and evaluation of fast algorithms for discrete signal processing is developed. Based on the theory of finite groups, it hence includes the familiar cases of the fast Fourier and Walsh-Hadamard transforms. However, the use of noncommutative groups reveals a large variety of novel methods. Some of these exhibit a superior performance, as measured by both reduced error rates and computational complexity, on nonstationary data.

The recent history of this subject is reviewed first, followed by a detailed examination of the three principal ingredients of the present study: the underlying groups, the signal-processing tasks on which the group-based algorithms are to compete, and the signal models used to define the data environment. Test results and conclusions then follow, the former being based on the use of random correlation matrices.

ACKNOWLEDGMENTS

The author is pleased to acknowledge the invaluable help of Mr. Thomas Loden in coding the program described in Section 5 and in producing the figures displayed in Section 6.

Ms. Laura Parr very capably prepared the manuscript for publication.

We also express appreciation to the Lincoln Laboratory Innovative Research Program Technical Review Committee for its early support of this project, and its ongoing interest and constructive criticism as the work progressed.

TABLE OF CONTENTS

Abstract	iii
Acknowledgments	v
List of Illustrations	ix
List of Tables	ix
1. INTRODUCTION	1
1.1 A General Approach to Signal Processing	1
1.2 The Role of Finite Groups	4
1.3 Recent Work of Other Authors	7
2. THE GROUPS	11
3. GENERIC SIGNAL-PROCESSING TASKS AND THEIR PERFORMANCE MEASURES	15
4. SIGNAL MODELS AND RANDOM CORRELATION MATRICES	19
5. COMPUTATIONAL ASPECTS	21
6. RESULTS AND CONCLUSIONS	25
7. SUMMARY	33
ADDENDUM	35
REFERENCES	37

LIST OF ILLUSTRATIONS

Figure No.		Page
1	Range of signal-processing activity.	1
2	Data compression summary (average geometric mean).	25
3	Data compression summary (average geometric mean).	26
4	Decorrelation data (percent) — stationary.	27
5	Decorrelation data (percent) — nonstationary.	28
6	Decorrelation data (percent) — nonstationary.	28
7	Decorrelation data (percent) — nonstationary.	29
8	Group filter error increase over optimal (percent) — stationary, SNR = 5 dB.	30
9	Group filter error increase over optimal (percent) — nonstationary, SNR = 5 dB.	30
10	Group filter error increase over optimal (percent) — nonstationary, SNR = 5 dB.	31

LIST OF TABLES

Table No.		Page
1	Groups Used in This Study	11
2	Groups of Order 2^n	13

1. INTRODUCTION

This preliminary section is intended to provide a necessary minimal background to motivate the technical results to be presented below. Most of this motivation is set out in great detail in an earlier Technical Report [1] (especially Section III therein), and it is assumed that serious readers are familiar with that material. However, the present section will update [1] by restating the essential role of finite groups in discrete signal processing and by reviewing some pertinent work of other researchers which, in most cases, was not available when [1] was being prepared.

In subsequent sections of this report, we will detail the results of an effort to implement and evaluate the program set forth in Section III of [1], namely the application of the harmonic analysis associated with finite groups to certain standardized signal-processing tasks. The essential purpose is to determine if this somewhat exotic theory can lead to practical signal-processing algorithms that improve on conventional techniques that use the fast Fourier transform. This particular transform is, of course, associated with a particular class of finite groups, namely the cyclic groups. We shall eventually conclude that its use can only be recommended for data derived from a stationary signal source; more general nonstationary data are better processed by transforms associated with other groups.

1.1 A GENERAL APPROACH TO SIGNAL PROCESSING

For several years, I have been evolving both a technical and a philosophical approach to signal processing. Here, I emphasize a particular technical aspect, as a sequel to Section III of [1]. For an associated philosophical viewpoint, refer to Section I of that report where I expressed the view that one of the primary motivations for working on the foundational aspects of signal processing is to provide answers to the basic problem of justifying numerical operations on a data set; that is, why do we do one thing and not another?

Many of the possible answers to such questions are prosaic. A particular algorithm may be employed because it is familiar, convenient (e.g., prepackaged), or fast. Whether or not it is the best thing to do in a given context is often less clear. Indeed, there is always a trade-off among familiarity, speed (e.g., complexity), and error. Even when an optimal procedure is known, its actual implementation may be too computationally intensive for the desired application, and recourse to a faster but suboptimal method must be made.

We can perceive a coarse partition of the broad spectrum of signal-processing activity, as depicted in Figure 1. The spectrum ranges from highly theoretical studies (mathematical foundations, or MFSP), to the design of corresponding reliable and efficient software, to the implementation of this code in suitable hardware, to the final specific field usage of such devices, usually as part of a larger system.

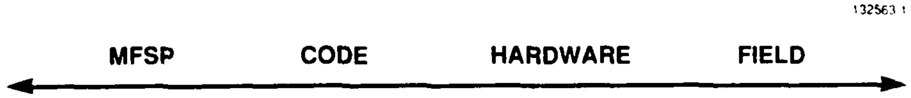


Figure 1. Range of signal-processing activity.

Mindful of the fundamental problem of justifying operations on data, work in MFSP is usually aimed at creating optimal algorithms for some particular data and design structure. But, it may also involve creating new algorithms (optimal or not) by use of some fairly exotic mathematics. As in theoretical physics, this ever-increasing use of ever-deeper and more diverse areas of mathematics is the source of much of the power and the intellectual excitement of this work.

Several quite recent examples of the high level of international interest and activity in MFSP include the following conferences and workshops: the Conference on Mathematics in Signal Processing at University of Bath, September 1985; the SIAM Workshop on Mathematics of Systems and Signal Processing at Stanford University, September 1987; and the summer Program on Signal Processing at the Institute for Mathematics and Its Applications, Minneapolis, mid-1988. Proceedings of the September 1985 conference have been published [2].

Returning momentarily to the context of Figure 1, we may say that the kind of application motivating the work about to be described is a signal-processing device, operating within the confines of a larger system, which has to repeatedly and rapidly process data either to extract information (e.g., estimate on an unobservable signal embedded in noise) or to eliminate redundancy. The ambient system might then use the processor output for fire control or direction change in the first case, or for data transmission or display in the second. In all such applications, the main question concerns the tradeoff between speed (efficiency, complexity, etc.) and error. The methods discussed below will illustrate some possibilities for accepting small increases over optimum in system error rates, in return for increases in computational efficiency.

To facilitate comparisons between competing algorithms, we will henceforth adopt a strictly neutral approach toward applications, in the sense that the discussion will remain free in implementation considerations and specific technologies. Hence, the algorithms will only be compared numerically and statistically on standardized signal-processing tasks, such as data compression and minimum mean-square (Wiener) filtering.

One final comment concerning the coarse structure of MFSP: We see this structure as triangular, with "vertices" consisting of Hilbert space theory, probability and statistics, and abstract harmonic analysis on groups. We must immediately acknowledge that this perceived structure is certainly not all inclusive, but is intended merely to give an approach to discerning some general principles and methodology in a virtually infinite amount of technical detail.

The importance of Hilbert spaces is that these serve as domains of operators, which in turn model various transformations of one signal into another. Operation theory has been intensively developed in the last four decades, as has the more exotic theory of operator algebras, although the most fundamental results (Spectral theorem; Bochner and Stone theorems for positive definite functions and unitary representations of the real line; Peter-Weyl theory for decomposing unitary representatives of compact groups; von Neumann double commutant theorem) date back to the late 1920s. However, most of the operators with direct signal-processing application are of elementary structure (e.g., projections, compact, and unitary operators) and many of the practical difficulties are of a computational nature (e.g., computing singular value and spectral decompositions, adjoints and pseudoinverses, approximation by simpler operators, etc.).

Since real data are always accompanied by noise, "what is measured is not the truth," and, therefore, "careful signal processors must be statisticians" [J. Tukey], so, probability and statistics constitute our second fundamental "vertex." Combined with Hilbert space, this is the basic setup to model second-order random variables and processes, (most of) the theories of Gaussian measures and abstract Wiener spaces, mean-square prediction and filtering (conditional expectations), etc. Much of what is now called linear inverse theory (recovering an unobservable signal passed through some channel or measuring device, and observed in noise) can also be constructed in this context, as can the theory of optimal algorithms and information-based complexity.

This statistical-linear approach to signal processing is by now quite standard, and not anything that requires further commentary here. In [1], I attempted to make the case for harmonic analysis over not necessarily commutative groups as a third foundational "vertex" of signal processing. In general, groups permit us to take advantage of physical or temporal symmetry in a situation, and to produce orthogonal expansions in functions that respect this symmetry. To this end, the vast theory of unitary representations is available, especially the Peter-Weyl theory for compact groups. A familiar example is that of stationary stochastic processes over locally compact commutative groups. In a different direction, I emphasized in [1] the role of finite groups as offering a unified approach to fast unitary transforms, and it is this theme that is developed below.

Finally, there are two points not made in [1] that should be inserted here. First, as mathematics advances, we are starting to see an influx of novel nonorthogonal bases and expansion devices in (separable) Hilbert space. We have in mind here, *inter alia*, Riesz bases (actually a rather classical notion, defined as a bounded unconditional basis, equivalently, an orthonormal basis transformed by an automorphism) [3]; Dirac bases (a kind of continuous orthogonal basis in the context of "Sobolev triples") [4]; and, inevitably, Dirac-Riesz bases [4], generalized coherent states, and "wavelets." These last two concepts are very much associated with (square-integrable) group representations, appearing as discrete subsets of certain orbits defined by an element of associated Hilbert space known in each case as the "analyzing wavelet." This kind of analysis, developed primarily by French mathematicians and physicists [5, 6, etc.] is rather profound, especially when dealing with non-unimodular groups, and is likely to be of real value in joint time-frequency approaches to signal processing. It extends the classical work of Gabor (wavelets) and Wigner (distributions). It is not germane to give a more precise discussion here, the point simply being yet another instance of group-theoretic methodology in signal processing.

Second, we should bear in mind that the early development of the general or abstract theory of Hilbert space and operators thereon was greatly motivated by attempts to formulate a rigorous theory of quantum mechanics. The chief architect of this fusion was, of course, John von Neumann [7]. For example, the most fundamental of operator structure theorems, the Spectral theorem, states that a not necessarily bounded self-adjoint operator can be synthesized from a family of orthogonal projections, more precisely, from a projection-valued measure (or resolution of the identity) on the real line:

$$T = \int_{-\infty}^{\infty} \lambda \, dE(\lambda)$$

The quantum mechanical interpretation of this formula is that for each system state, as described by a unit vector x in the space of T, the probability measure

$$B \rightarrow \langle E(B)x, x \rangle$$

describes the result of measuring the observable associated with T. The projections $E(B)$ themselves correspond to yes-no "questions" about the observable, so that it at least becomes plausible to expect some sort of determination of such operators by a family of projections.

Now, the only point to be made here is that because of this close connection between the mathematical foundations of quantum mechanics and the basic theory of Hilbert space, and because of our heavy use of this theory in our own work on MFSP (not so much an issue in the present report), we should expect some interesting connections between quantum mechanics and signal processing. In fact, such connections are being noticed in various ways by several researchers, and we intend to discuss some of these in later reports. Hopefully, an eventual unified theory of fundamental quantum mechanics, pattern recognition, information theory, and signal processing will emerge (insofar as this is possible!).

1.2 THE ROLE OF FINITE GROUPS

Let us now move directly to the essence of our subject: the uses of finite groups for certain signal-processing applications. As suggested in [1], a distinction is to be made between the roles of finite and infinite groups: representations of and analysis on the latter serve to define structure and analytic models, such as expansions and transforms of idealized signals, while analysis on the latter leads to direct computational algorithms.

As discussed in [1], finite nonsimple groups permit a unified approach to fast unitary transforms and discrete suboptimal filters. Of course, there are other nongroup-theoretic approaches to fast transforms, and references were provided in [1] to such work (References 10, 15, and 47 of Section III therein). Attempting a moderately reliable analogy, we may say that, at the next level of complexity, the simple groups are to the family of all (finite) groups as the prime numbers are to the set of all positive integers. In fact, each integer admits an essentially *unique prime factorization*, and each finite group admits a composition series (that is, a decreasing chain of subgroups, each of which is a maximal normal subgroup of its predecessor); the associated factor groups are then simple. The Jordan-Holder theorem states that any two composition series for a particular group are equivalent, in the sense that the simple groups associated with the two series can be paired off isomorphically. Thus, each finite group determines a unique list of simple groups. And the converse is, in a certain sense, also valid. That is, if we know all simple groups (which is the case now [8]), then all finite groups can, in principle, be obtained by use of the Schreier extension technique. Hence, all possible groups having a given list of composition factors can be constructed [9], although not very efficiently. In particular, when this theory is applied to the class of cyclic groups, we can recover the prime factorization of integers.

Now, like the integers, the finite groups are elementary mathematical objects and exist independently of any further theories, or of any particular signal-processing application. It was shown in [1] that the complexity of the group transform F_G associated with a particular group G, of order N, is related to the subgroup structure of G, especially to the length of a composition series for G; the greater this length, the greater the possibility of reduction of F_G complexity. The best case has $F_G(x)$ complexity reduced from $O(N^2)$ to $O(N \log_2 N)$, as with the familiar fast Fourier or fast Walsh transforms on the cyclic (resp., dyadic) groups of order $N = 2^n$.

The complexity issue for various group transforms and their inverses, the latter not necessarily being group transforms, has been rather extensively examined in recent literature and will be reviewed in more detail in Subsection 1.3. On the other hand, this is not the only issue of interest in regard to signal-processing applications. There is the further problem of deciding which of the (often many) groups of a given order is most appropriate to a particular signal-processing task. That is, there is the fundamental trade-off between speed, in the sense of reduced complexity, and the error incurred as the result of using a particular group transform or filter in place of the optimum transform or filter.

We must be precise here. Since we are interested in evaluating and comparing the performance of different groups in a signal-processing context, it is necessary to specify four variables, namely:

- Data dimension
- Group
- Task (with performance measure)
- Signal model.

In regard to the first two variables we naturally require that

$$\textit{data dimension} = \textit{order (group)} \quad .$$

The latter three variables are discussed in detail in Sections 2 through 4 below. When all four variables have been fixed (as indicated in the later sections), then a definite question can be asked, and thus group performance in various situations can be contrasted.

In further regard to the relation between the first two variables, we have two alternatives. First, we can fix a data dimension N and then consider some (possibly all) groups of order N. This becomes difficult too rapidly as N increases, especially for the most important cases of highly composite N (such as $N = 16, 32, 48, 64, 96, \ldots$) which are, as already noted, exactly the cases where the greatest reduction in complexity is to be expected. Indeed, for the values of N just mentioned there are, respectively, 14, 51, 52, 267, 230, ... distinct group isomorphism types of that order. Hence, a second approach has proved useful — to study the asymptotic performance of a sequence of groups of orders N_k, as $k \to \infty$. Familiar examples of such sequences are the cyclic groups of order $N_k = k$, the dihedral groups of order $N_k = 2k$, and the dyadic groups of order $N_k = 2^k$. Relevant properties of these groups, and of a fourth sequence, will be discussed in Section 2 below.

We shall look at three canonical signal-processing tasks (data compression, data decorrelation, and Wiener filtering) and their associated performance measures, with which we will challenge the various groups. Other tasks and measures could have been selected (a rate-distortion function, an entropy criterion, resolution ability, etc.), but we had to start, and stop, somewhere and the three tasks that we have elected to work with are familiar and neutral, that is, free of application or hardware specifics.

The choice and analysis of various performance measures and their interrelationships, especially in connection with compression and decorrelation, are a little more subtle than might be expected at first and, in fact, are currently under further investigation by the author. This will be discussed further in Section 3 below. But, we can note here that all the performance criteria utilized in this report have the virtue of being rather simple functions of the data covariance matrix. (Actually, our signal models will be restricted so that all components of each signal used as a test case for any of our algorithms will have the same variance; hence, we may assume that all covariance matrices are of correlation type, that is, have their diagonal entries equal to unity.) From this it follows that a choice of signal model may be reduced to a choice of a *random correlation matrix*. This happens to not be a particularly obvious thing to do and, in fact, leads to a host of interesting problems which are described and, in some cases, solved in two other reports [10, 11]. A summary of the essential points of this work is given in Section 4 below.

Let us now return once more to the general role of finite groups in discrete signal processing. Suppose that we are given a random vector x, of dimension N, with $E(x) = \Theta$. For any particular group G, of order N, we can think about the fundamental idea in two equivalent ways. First is the approach described at length in [1], namely to view x as a function in $L^2(G)$, and to expand x in the special orthonormal basis defined by the irreducible representations of G. The mapping F_G which assigns to x its Fourier coefficients with respect to this basis is exactly the group transform on $L^2(G)$, and by Parseval's theorem is a unitary operator there. It is sometimes more useful to group these coefficients together in sets of size d_i^2, where d_i = dimension of i^{th} irreducible representation of G, so that F_G can be considered to map $L^2(G)$ onto $L^2(G)$, where (G) is the dual object of G consisting of all the irreducible representations of G.

Alternatively, we maintain the natural view of x as an element of the N-dimensional Hilbert space $l^2(N)$, and think of G as a symmetry group of the set $\{1, 2, ..., N\}$, that is, as a subgroup of the symmetric group on N letters (Cayley's theorem). In turn, G may be realized as a group of $N \times N$ permutation matrices. If each point of $\{1, 2, ..., N\}$ is considered to have measure $1/N$, then each $N \times N$ permutation matrix can be considered to be a measure preserving transformation of the trivial measure space $\{1, 2, ..., N\}$. So, we have a unitary representation of G on the space $l^2(N)$. By the usual general theory, this representation can be decomposed into a direct sum of its irreducible components, and so $l^2(N)$ correspondingly splits into a direct sum of invariant subspaces. By selecting an orthonormal basis for each of these subspaces, and expanding x in the resulting coordinates, we obtain a decomposition equivalent to that of the first approach.

In the case of a nonabelian G, there is some unavoidable non-uniqueness in the choice of these bases that goes beyond a mere rearrangement. This is due to the presence of at least one irreducible representation of dimension ≥ 2. Hence, the group transform F_G is only unique up to a unitary equivalence.

For a given data dimension N, the various group transforms (defined by the groups of order N) formalize the initiative notion of the passage from physical to spectral coordinate domains. Each such transform is to be thought of as a potential approximant of the discrete Karhunen-Loève transform (DKLT) associated with a particular statistical signal class. We contemplate using such approximants because they are fast to compute (depending on the group, and relative to a general $N \times N$ unitary transform), and because we may not know the exact statistical nature of our data.

Similarly, group filters, defined abstractly as right convolution operators on $L^2(G)$, are to be thought of as potential approximants to some specific operator on $L^2(G)$ which has arisen in a signal-processing context. The example that we emphasize in this report is the Wiener-filter operator which is the optimal solution (in the sense of minimum mean-square error) to the problem of estimating a random signal with known covariance structure that has been observed in additive white noise (uncorrelated with the signal). By passing to the spectral domain via the group transform, we can obtain a fast algorithm for such operators, just as is done in the classical case of computing digital filters by means of multiplying the corresponding transfer functions in the (fast) Fourier domain. Also discussed at length in [1] was the problem of choosing the optimal filter, for a fixed G, to best approximate a particular operator such as a Wiener filter. As usual, we are interested in the trade between the reduced complexity of the optimum group filter (at least for certain groups, as discussed below) and the increase in error over the theoretical optimum.

1.3 RECENT WORK OF OTHER AUTHORS

Before delving into the technical details of our own work, it seems useful to briefly survey some other related approaches to the main problems. This is done primarily for balance and completeness, and we will concentrate on work not already referenced in our preceding report [1].

Let us first consider the problem of approximating the DKLT of a random vector x with $E(x) = \Theta$. This is the essential issue in data compression and decorrelation, since this transform has the desired optimality properties of maximum variance-packing and complete decorrelation, among others relevant to feature selection for pattern recognition [12].

Since the spectrum of a Hermitian operator such as the covariance operator of x is unique only up to order and multiplicity, the DKLT of x is not uniquely specified. At the very least, namely when the eigenvalues of the covariance operator are distinct (i.e., have multiplicity one), any unitary matrix U that diagonalizes a matrix representation of this operator can be replaced by UQ, where Q is a permutation matrix. So, when we talk about approximating the DKLT we shall be referring to a unitary matrix U with the property that U^*AU is "approximately diagonal," where A is a covariance matrix of x. Of course, this phrase needs to be carefully defined and, in fact, is a somewhat subtle problem, as already mentioned. We will touch on this in Section 3.

In general, we can discern at least three approaches to approximating the DKLT, which we will term "direct," "asymptotic," and "fixed suboptimal." Let us say a few words about each of these in turn. The reader should keep in mind that there are always two conflicting aspects of this problem: the accuracy of the approximation, and its complexity.

The direct approach deals with an explicitly known covariance matrix A, which is associated with a vector of samples from a weakly stationary stochastic process. Hence, A is also a Toeplitz matrix. A method developed by A. Solodovnikov [13, in Russian] leads to a fast algorithm for the exact DKLT of such a matrix in the case where its dimension N has the form 2^n. The basic idea is to factor the DKLT matrix as the product of n sparse matrices, so that each element of A can be expressed as a product of just one element of each factor. The proof is carried out by induction on n, beginning with the trivial case where $n = 1$:

$$\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} \sigma & \rho \\ \rho & \sigma \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = 2 \begin{bmatrix} \sigma + \rho & 0 \\ 0 & \sigma - \rho \end{bmatrix} .$$

Notice that in this case the diagonalizing matrix is just the 2×2 Walsh-Hadamard (W-H) transform. In the general induction step the $2^n \times 2^n$ W-H transform is used, along with a permutation, to reduce A to block diagonal form to which, in turn, the induction hypothesis can be applied.

A second direct approach has been suggested in the Russian literature by I. Kaporin [14], under the alternate assumption that the covariance matrix A is tridiagonal. This is of course a significant restriction since, otherwise, to first put A in this form (say by Householder reduction) requires $O(N^3)$ operations. But when A has this special form, recursive linear computations can be defined to yield an $O(N \log^3 N)$ operations count in affecting the DKLT of a vector.

Next we turn to the asymptotic approach. Once again we consider samples $(x_1, \ldots, x_N)$ drawn from a weakly stationary stochastic process, but without a definite bound on N. The corresponding covariance martrices will again be Toeplitz. Early treatments of this topic [15, 16] proceeded by making some assumption about the rate of decrease of the autocorrelation function of the process, and then showing that some particular sequence of unitary transforms of interest, such as the sequence of N-point DFTs, would asymptotically decorrelate the $N \times N$ covariance matrices. Depending on the precise nature of the assumptions, this last phase could mean either that some fixed correlation between two of the transformed components of $(x_1, \ldots, x_N)$ tends to 0, or that some composite measure of all these correlations does so.

More recently, M. Unser [17] has shown that several popular fast unitary transforms (DCT, DFT, DOFT, ...) achieve asymptotically the effect of the DKLT, provided such effects are carefully described by appropriate performance measures, namely as separable convex or concave functions of the variances of the transformed samples. The analysis nicely makes use of a classical theorem on the asymptotic distribution of the eigenvalues of Hermitian Toeplitz matrices, due to Grenander and Szego, as adapted by Gray [18, Thm. 2.3]. To my knowledge, this result of Unser's determination is the most general result presently known on asymptotic approximation of the DKLT for Toeplitz covariance matrices.

Lastly, we consider the fixed suboptimal approach. This is appropriate in all cases not yet covered, that is, the problem size is fixed but the exact signal covariance is not known (or, if known, is not Toeplitz). In this case, we can attempt to measure the performance of some fixed N-dimensional unitary transform over a class of $N \times N$ covariance matrices. By doing this repeatedly, we could hope to rank the transforms relative to the fixed performance measure and covariance class. Of course, they could also be ranked relative to computational complexity, which can be defined independently of any signal-processing task.

Along with M. Karpovsky and E. Trachtenberg [19, 20], I also have adopted this sort of approach. All this activity involves the notion of transforms (and filters) defined by finite groups, the general theory of which was set down at length in [1]. Actually, the most recent work of Trachtenberg [21, 22] utilizes only abelian groups. In [21], for example, such a group G of order N is given, and then a class of so-called Frobenius group matrices is introduced. Such a matrix A is defined by specifying first a scalar function f on G and then defining

$$a_{ij} = f(g_i \oplus g_i) \quad .$$

where $G = \{g_i: i = 0, 1, \ldots, N-1\}$; A turns out to be unitary exactly when $f(g)\,f(g^{-1}) = 1$, for all $g \in G$. The associated matrix transforms can be computed by fast algorithms in many ways, depending on how G is factored into a product of (cyclic) subgroups, and on the support of f. Any remaining degrees of freedom in the choice of f can be utilized to relate the performance of the A-transform to a given signal-processing task, such as DKLT approximation. In this fashion, a whole spectrum of trades between complexity and error can be attained.

Our final comments in this section pertain to the issue of group transform complexity. This issue was discussed at some length in [1], primarily for the case of abelian groups. For such groups a satisfactory theory exists, due to the facts that all irreducible representations are one-dimensional (characters) and that the dual object has a group structure (dual group), so that the inverse transform is again a group transform. For nonabelian groups, the only result I knew of at the time [1] was written (due to Karpovsky [23]) quantified the complexity of the direct and inverse group transform as $O[\mathrm{ord}(G)\ \Sigma\ \mathrm{ord}(G_i)]$, when the group G can be decomposed into a direct product of groups G_i. This same result was established independently by M. Atkinson [24]. More recently, this has been investigated by T. Beth [25] for the much more general case of solvable groups. For such a group G, there is a composition series

$$\{e\} = G_{r+1} \subset G_r \subset \ldots \subset G_1 = G \quad .$$

with each G_k a normal subgroup in G_{k-1}, of prime index p_k. Then, the computational complexity of each transform can be bounded by ord $G(\Sigma\, p_k) + O[\mathrm{ord}(G)^{u/2}]$ steps, where $u \leq 3$ is the exponent for which $N \times N$ matrices can be multiplied in $O(N^u)$ arithmetical steps.

Of course, not all groups of potential interest in statistics and engineering are solvable. Perhaps the foremost example is the symmetric group S_n on n letters. For this group, a special analysis of transform complexity has been made by Diaconis and Rockmore [26], who show the possibility of reducing the usual $O[(n!)^2]$ count to $n(n!)^{u/2}$.

2. THE GROUPS

Recalling the preliminary discussion in Section 1.2, we now indicate precisely the four classes of finite groups used in the present work. These are the cyclic groups, the dyadic groups, the dihedral groups, and, for lack of a more traditional terminology, the "Russian groups." There is a cyclic group C_n of all orders, while the other types exist only for restricted orders as indicated parenthetically in Table 1.

TABLE 1

Groups Used in This Study

Commutative	Noncommutative
Cyclic (n)	Dihedral (2^n)
Dyadic (2^n)	"Russian" (2^n)

Three of these classes of groups have a geometric significance, while the Russian groups are more abstract, being in general only defined in terms of generators and relations. Thus, the cyclic group C_n can be identified with the group of counterclockwise rotations of the plane through an angle of $2\pi/n$ radians. The dyadic group D_n is the set of all n-dimensional binary (0,1) vectors, with componentwise addition mod 2. Thus, equivalently, $D_n = C_2 \times ... \times C_2$, n times, and we can think of D_n as constituted of the vertices of the unit hypercube in R^n. The dihedral group Di_n is the group of symmetries of the regular n-gon. It is generated by a rotation a and a reflection b so that $a^n = b^2$ = identity, and $aba = b$. Thus, Di_n contains C_n as a normal subgroup and is of order $2n$. It is the semidirect product of C_n and C_2, and is hence the simplest example of a noncommutative group.

Finally, the "Russian groups" form a sequence $\{BG_n\}$ of groups of order 2^n defined as follows. If n is even, $BG_n = BG_{n-1} \times C_2$. If $n = 2k + 1$, then BG_n has generators $\{a, b_1, ..., b_k\}$ obeying the relations

$$ab_i = b_i a \quad ,$$

$$b_{2i-1} b_{2i} = b_{2i} b_{2i-1} a \quad , \quad i = 1, \ldots, k$$

$$b_i b_j = b_j b_i, \text{ otherwise} \quad .$$

It happens that $BG_3 = Di_4$, but otherwise the Russian groups are not dihedral.

The Russian groups were introduced into the signal-processing literature by Berman and Grushko [27]. The major thrust of this article was to prove that the computational complexity of the associated group transform is $0.75\ N \log_2 N + o(N)$, as $\mathrm{ord}(BG_n) \to \infty$. Thus, the effort to compute the BG_n transform is about 1/2 (resp. 3/4) that of the familiar FFT (resp. W-HT). The proof is based on a careful analysis of the unitary dual $\hat{BG}_n$; this set contains just characters and either one or two irreducible representations of dimension 2^k according as $\mathrm{ord}(BG_n) = 2^{2k+1}$ or 2^{2k+2}. The commutator subgroup of BG_n is the group C_2, and its abelianization BG_n/C_2 (the largest commutative quotient group of BG_n) is the dyadic group D_{n-1}.

We note that the authors of [27] did not attempt to evaluate the role that their groups might play in signal-processing applications beyond that of reduced complexity of the associated transform. Looking ahead slightly, we will include such an evaluation below and eventually deduce that, in terms of standard performance measures for the SP tasks that we consider (the Russian groups), performance is comparable to that of the dyadic groups.

As a matter of specificity we record next the general form of the unitary matrix of the group transform of BG_n, $n = 2^{2k+1}$, for a particular choice and ordering of the basis vectors in the spaces of the various irreducible representations. Let

$$H_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \text{and } H_m = H_{m-1} \otimes H_2 \text{ , for } m \geq 3 \quad .$$

These are the well-known Hadamard matrices. Then, the group transform matrix of BG_n is, up to a normalizing constant,

$$\begin{bmatrix} & H_{2^{2k}} & & & & H_{2^{2k}} & \\ cH_{2^k} & & & -cH_{2^k} & & & \\ & cH_{2^k} & & & -cH_{2^k} & & \\ & & \cdots & & & \cdots & \\ & & cH_{2^k} & & & & -cH_{2^k} \end{bmatrix}$$

where $c = 2^{k/2}$.

Let us use the phrase "superfast group" for any group whose associated group transform can be computed "significantly faster" than the usual FFT. Without attempting to be overly precise about the meaning of this phrase, let us agree that the groups $\{BG_n\}$ are at least asymptotically superfast. Open questions now abound: Are there other superfast groups? How can they be recognized or constructed? And what is the limit of complexity reduction that is possible? There are also the important ancillary questions of such groups' value, if they exist, for particular SP tasks.

We have already remarked on the rapid growth of the number N_k of distinct groups of order k when k is a highly composite integer of the form $k = 2^n$ or $k = 3 \times 2^n$ (cf. Subsection 1.2). Table 2 illustrates this remark even more dramatically for the prime power orders $k = 2^n$.

TABLE 2

Groups of Order 2^n

k	N_k
16	14
32	51
64	267
128	2,328
256	56,092
512	>8,400,000

Since groups of order $k = 2^n$ are those of order k for which maximum *complexity reduction is to* be expected, we must concentrate on these. Obviously, the numbers in Table 2 preclude a case-by-case analysis, so that special constructions, as were done for the Russian groups case, will be required. On the other hand, the sheer size of the numbers N_k in Table 2 suggests a degree of *a priori* probability for the existence of other classes of superfast groups. (We might term this argument one of "ample opportunity," faintly analogous to that for the existence of extraterrestrial life based on the large number of stars in our galaxy.) There is a field known as computational group theory whose techniques will likely be of use in this matter.

3. GENERIC SIGNAL-PROCESSING TASKS AND THEIR PERFORMANCE MEASURES

We now want to begin consideration of how the various group transforms and optimal group filters can be compared. The most direct approach is to simply let them operate in controlled statistical environments and monitor the results from various SP tasks. Specification of the latter is our present concern. In order not to appeal to any possible bias on the reader's part, we will limit attention to generic SP tasks, without regard to hardware implementation or specific technology application.

Proceeding from this intent, we next list several such SP tasks; others will, of course, occur to knowledgeable readers:

Compression

Decorrelation

Filtering (estimation)

Detection

Discrimination

Pattern recognition

Resolution

Simulation.

Clearly some of these tasks, such as filtering and pattern recognition, are very general, and considerable further specification is necessary to arrive at a well-defined problem to which our particular group-theoretic techniques can be applied.

For present purposes, we have decided to restrict attention to the first three applications listed above. Each of these is discussed in detail in the earlier report [1]; relevant portions of that discussion are reviewed next. The key point is that in each case a natural performance measure is given as a function of the signal covariance. This fact permits us to neatly generate test signals, as we will see in the next section.

The compression and decorrelation tasks can be treated from a common viewpoint, which is called "transform coding" in the engineering literature. A block x of N samples is formed from a signal of constant variance (perhaps, but not necessarily, stationary), and transformed by a unitary matrix U: $y = Ux$. The covariance matrices of x and y are related by

$$\Sigma_y = U \Sigma_x U^* \quad . \tag{3.1}$$

In a complete SP system, the transformed samples $\{y_j\}$ might now be individually quantized prior to transmission over some channel and eventual reconstruction at the receiver. However, here we are primarily interested in the statistical behavior of the original or "physical" coordinates $\{x_j\}$ vis-a-vis that of the transformed or "spectral" coordinates $\{y_j\}$.

In general we may say that compression is determined by some measure of the size of the diagonal of Σ_y, and decorrelation by some measure of the off-diagonal entries. A more elegant approach would in fact use the *same* measure for decorrelation as for compression, but this is a matter still under investigation. So, we will stay with the more ad hoc approach for now.

For notational simplicity, let us set $P = \Sigma_x$ and $Q = \Sigma_y$. We will assume that P is a correlation matrix, so that $P_{ii} = \text{var}(x_i) = 1$, $i = 1, \ldots, N$. It then follows that

$$tr(Q) = tr(P) = N \quad .$$

In this study, we choose to measure compression either as a plot of the fraction of total variance in the largest m spectral coordinates vs m, $1 \leq m \leq N$, or by the single number γ_Q which is the geometric mean of diag (Q). This measure γ_Q is interesting for a couple of reasons. First, its logarithm

$$\log \gamma_Q = \frac{1}{N} \Sigma \log (q_{ii})$$

is a simple example of a concave symmetric function of the $\{q_{ii}\}$. As such, it is part of a more general theory of optimizing compression by minimizing concave (resp., maximizing convex) symmetric functions of the spectral variances. Second, its use is motivated by work in information and rate distortion theory; in particular, the result that the average distortion from an optimal bit assignment in an independent quantization of $\{y_i\}$, subject to the constraint of a given average bit rate, is proportional to γ_Q [28].

As general facts about γ_Q, we note that $\gamma_Q \leq 1$ as long as P is a correlation matrix, and that min γ_Q is achieved when the transformed matrix U is a discrete Karhunen-Loève matrix for P. It can also be shown that γ_Q can be arbitrarily near to 0 for some correlation matrix P.

To measure the amount of decorrelation achieved by the transform U, we will adopt the ad hoc quantity

$$100\left(1 - \frac{|||Q|||}{|||P|||}\right) \quad , \tag{3.2a}$$

where

$$|||A||| = \sum_{i>j} |a_{ij}| \quad . \tag{3.2b}$$

This quantity normally ranges between 0 and 100, with 100 meaning that the spectral components have become decorrelated, and 0 meaning no change in overall correlation structure. However, as we will see, it can happen that $|||Q||| > |||P|||$ for some ill-chosen transform U. In such a case, the above performance measure will become negative, indicating that the transform operation has increased the overall correlation between data coordinates.

We turn now to a precise formulation of the filtering task. The simplest situation is classical mean square (Wiener) filtering of a signal observed in uncorrelated white noise. That is, we are given N-dimensional data

$$x = s + w \quad ,$$

with $E(s\,w^*) = 0$, and noise covariance matrix $\Sigma_w = \lambda I$ ($\lambda > 0$, I = N-dimensional identity matrix). The problem is to form a linear estimate

$$\hat{s} = L(x)$$

to minimize the error

$$E\left(\| s - \hat{s} \|^2\right) = e(L) = tr\left[(I - L)\Sigma_s\right] \quad . \tag{3.3}$$

The solution W = arg min $e(\cdot)$ is the standard (discrete) Wiener filter

$$W = \Sigma_s\left(\Sigma_s + \lambda I\right)^{-1} \quad . \tag{3.4}$$

The computation $s = Wx$ has no particularly efficient algorithm and, further, exact knowledge of the second-order statistics of the data is required before W can even be written down. For these reasons we proposed, and discussed at some length in [1], the replacement of W by fast suboptimal group filters. These latter are special kinds of linear transformations of N-space, having a certain internal symmetry defined by some group. Specifically, they are right convolution operators on $L^2(G)$, ord$(G) = N$ and, as such, they are unitarily equivalent, via the group transform associated with G, to a multiplication operator on $L^2(G)$, where G is the unitary dual. The basic point is that there will be a fast algorithm for their evaluation whenever there is one for the associated group transform.

If $\Phi(G)$ denotes the N-dimensional operator algebra of all group filters on G, then the optimal group filter for our filtering problem is

$$T_w = \arg\min E\left[\| s - T(x) \|^2\right] \quad , \tag{3.5}$$

taken over all $T \in \Phi(G)$. Of course, this depends on the choice of the group G. Varying G subject to ord$(G) = N$ will yield optimal filters with a variety of mean-square errors and computational complexities. In particular, we have

$$e(T_w) = tr\left[(I - T_w)\Sigma_s\right] \quad , \tag{3.6}$$

which can be compared with the benchmark formula for $e(W)$.

Note that all these formulas implicitly involve the positive parameter λ, which is related to a natural signal-to-noise ratio (SNR) expression by

$$\text{SNR} = \frac{tr(\Sigma_s)}{\lambda N} = \frac{I}{\lambda} \quad , \tag{3.7}$$

if we assumed that Σ_s is a correlation matrix, so that $tr(\Sigma_s) = N$. In this study, we have considered SNR values between ± 10 dB.

All the preceding formulas can, of course, be generalized to situations where the signal s has the form $s = Ah$, where h is a random element of some Hilbert space H, A is a linear operator on H of finite rank, and the noise w is colored. The operator A represents the effect of some kind of measuring device, or communications channel perhaps, on the unknown element h. The formation of optimal linear estimates $h = L(x)$ in this context is the subject of linear inverse theory, but this added generality exceeds our present need of a simple linear filtering task on which to test some new algorithms.

As noted previously in Subsection 1.2, all performance measures defined above have the virtue of being explicit functions of the data covariance matrix Σ_x or the signal covariance matrix Σ_s. We will put this feature to good use in the next section, by first defining several signal classes in terms of their covariance structure, and then indicating how representatives of these various classes can be picked at random, for purposes of a simulation to test the efficacy of our new algorithms.

5. COMPUTATIONAL ASPECTS

As a first attempt to assess the relative merits of our various algorithms, we decided to simply let them compete with one another in the two particular statistical environments just discussed. This is in keeping with the general approach of using the computer as a kind of "laboratory," either for the testing of conjectures or for the generation of data, from which we might attempt to evolve explanatory theories.

A FORTRAN program was prepared to carry out the necessary calculations. First, a data dimension N is selected. In this work, N was always taken as 2^n, $n \leq 6$. Next, a signal model class is selected from $AR(p)$, nonparametric stationary, nonstationary, as discussed in the preceding section. Then, 50 $N \times N$ random correlation matrices are generated from this signal class; these provide the test cases on which our evaluations will be based. Next, one of the four groups of Section 2 is specified, and a corresponding group transform matrix U is calculated, representing the group transform for a particular choice of basis. The minimal information for this is the set of values of a complete set of irreducible representations of the group taken on a set of generators.

At this point, we must decide on the type of signal-processing task on which to test our group transform. If the task is to be compression or decorrelation, we simply transform each of our correlation matrices according to Equation (3.1) and then compute the relevant performance measure. This is, respectively, the geometric mean γ_Q of the diagonal of the transformed correlation matrix, or the percentage decrease in overall correlation, defined by Equation (3.2). On the other hand, if the task is to be Wiener filtering, then a little more work is required. We first supply as an added input a signal-to-*noise figure, SNR*, defined by Equation (3.7). Then, we must compute the Wiener filter error from Equations (3.3) and (3.4), and the optimal group filter error from Equations (3.5) and (3.6). Carrying out the minimization in Equation (3.5) is not an entirely trivial task, and is described in Section 3.5 of [1]. The result is that a system $Ac = b$ of linear equations must be solved to yield coefficients $\{c_1, \ldots, c_N\}$ in the expression

$$T_w = \sum_i c_i \, R(h_i) \quad ,$$

where h_i runs through all the group elements, and $R(\cdot)$ is the regular representation of the group. The matrix A here has entries of the form

$$a_{ij} = \left\langle P + \mathrm{SNR}^{-1} I, R\left(h_j^{-1} h_i \right) \right\rangle \quad ,$$

where P is the input correlation matrix, and from this form it follows that A is positive definite, so the corresponding linear system is not troublesome numerically.

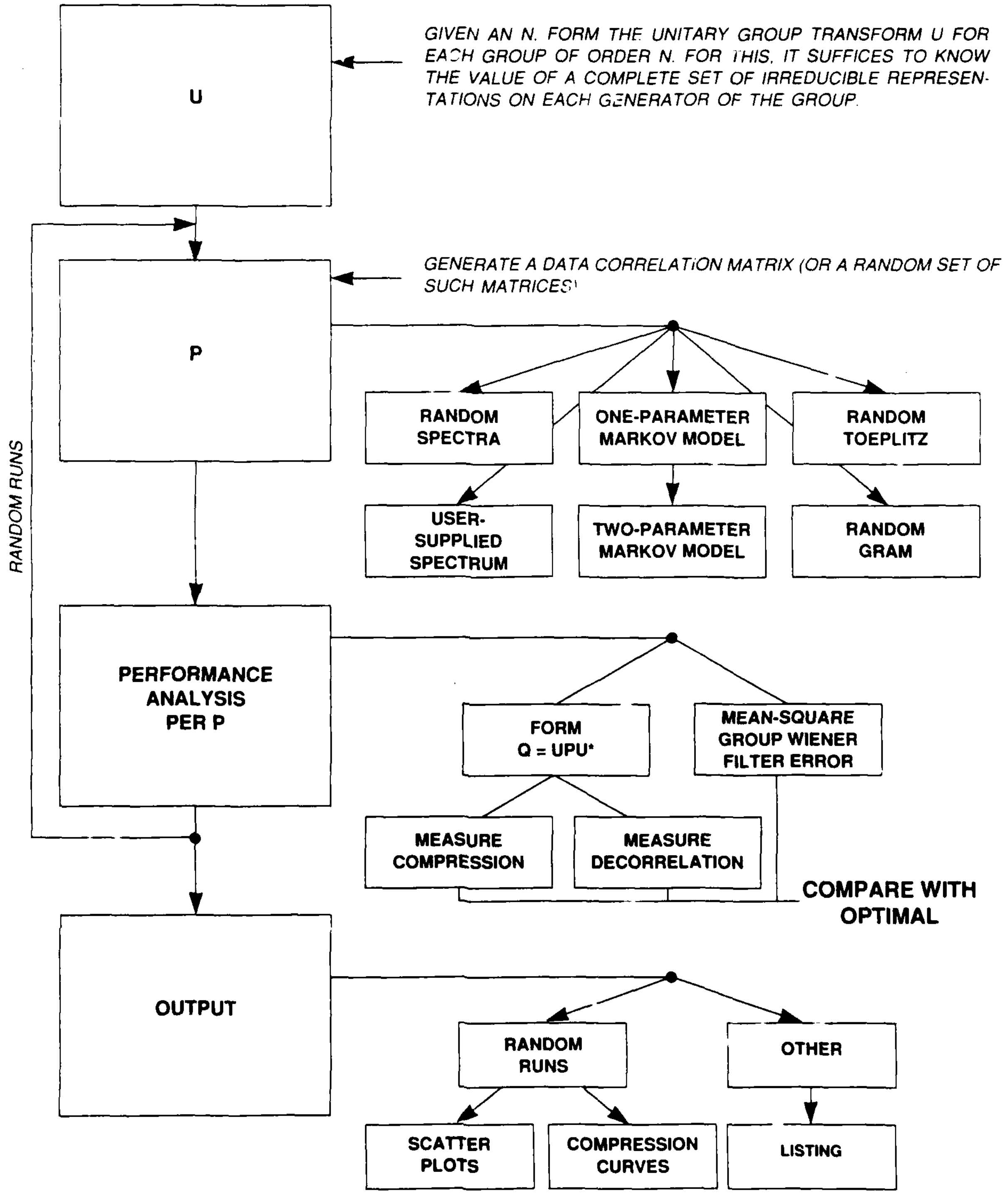
Finally, our computer program must display the results so obtained in a visually compelling format, and with useful summary statistics. For the tasks of decorrelation and filtering, we have elected to display the results in scatterplot form, for one pair of groups at a time, so that the two groups can be directly compared with one another. Fixing the pair of groups, say (G_1, G_2), a two-dimensional point

(p_1, p_2) is plotted for each sample correlation matrix. In the decorrelation case, p_i is the performance measure of Equation (3.2); while in the filtering case, p_i is the percentage increase of the optimal G_i-filter error over the Wiener-filter error. So, in the decorrelation case, $p_1 > p_2$ means that G_1 gave a *better* result than group G_2 (i.e., achieved a higher degree of decorrelation), while in the filtering case, $p_1 > p_2$ means that group G_1 gave a *worse* result than group G_2 (since its optimal group filter had a larger error than that of the other group).

For the task of data compression, we simply plot the average value of γ_Q for each group, along with two standard deviation error bars. Here, the smaller this average and the narrower the error bars, the better the group has performed. Note that we have also included the corresponding result for the discrete cosine transform (DCT), even though it is not a group filter, because of its popularity in engineering applications.

A flowchart of the program operation is presented on the following page.

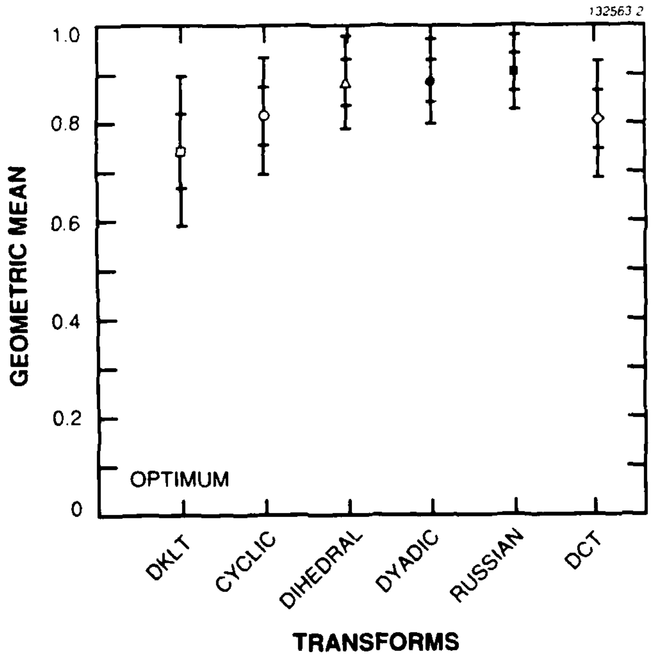
"GROUPS IFTRAN" OVERVIEW



6. RESULTS AND CONCLUSIONS

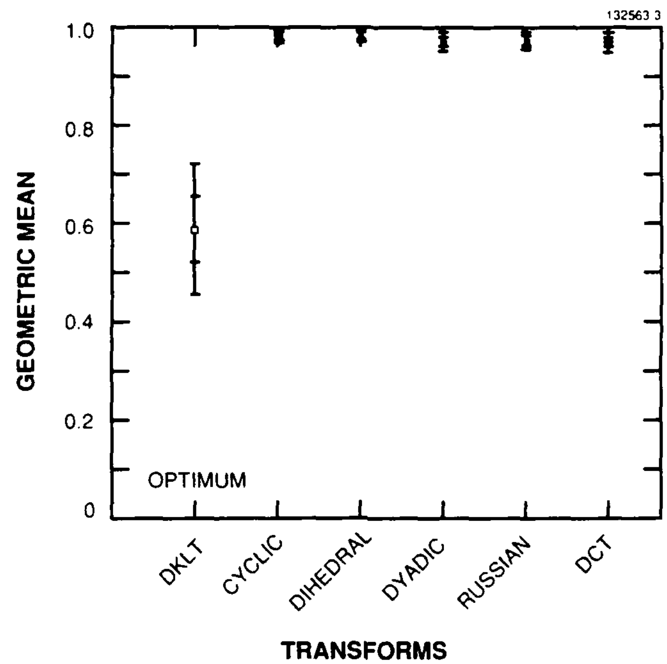
We will now illustrate the operation of the program just described by presenting several examples of output. Most of this output will be for the case of data dimension = 32, as being typical of the dimensions we considered ($8 \leq N \leq 64$). After output pertaining to each of the three signal tasks under consideration, we will offer some conclusions, based on not only the displayed results but all results relevant to that particular task.

Figures 2 and 3 show the ability of the various transforms to compress data vectors, with the geometric mean of the variances of the transformed vector as a xperformance criterion. We see that for samples from stationary data, the DFT (cyclic group transform) and the DCT are about equally effective, and close to the theoretical optimum of the DKLT. The other transforms are marginally less effective. However, for the nonstationary data samples, very little compression is evident by any of the transforms, even though the DKLT achieves a significant degree of compression.



STATIONARY: DATA POINTS = 50, N = 32 (Error Bars Are in Increments of 1 Standard Deviation)

Figure 2. Data compression summary (average geometric mean).



NONSTATIONARY: DATA POINTS = 50, N = 32 (Error Bars Are in Increments of 1 Standard Deviation)

Figure 3. Data compression summary (average geometric mean).

The empirical evidence presented here, that stationary data are highly compressed (relative to the optimum) by both the DFT and DCT, will not come as any surprise to readers familiar with the transform coding literature. Recall, for example, the work of Unser [17], wherein it is shown that the DFT and DCT (among others) are asymptotically equivalent to the DKLT, for data derived from (weakly) stationary processes.

Perhaps the result for nonstationary data will not be surprising either, given the total lack of structure in the correlation matrices. In the case of RCMs with random spectra, it can be proved that, asymptotically, the average geometric mean associated with the DKLT is $e^{-1/2} \approx 0.61 \ldots$, while that value associated with any of the transforms seems to tend to 1. Empirically, the discrete W-HT does best, but not well enough to be of any practical interest. The situation for Gram RCMs is, in a sense, even more striking, as the geometric mean figure associated with the DKLT seems to be asymptotic to a value $\approx 0.36 \ldots$.

These results tend to suggest that, given a correlation matrix P, the functional γ_Q defined on the unitary group by

$$\gamma_Q(U) = GM\left[\mathrm{diag}(UPU^*)\right]$$

which has a maximum value of 1 at $U = I$, has a steep and sharp minimum at U_p = DKLT (= matrix of normalized eigenvectors of P; the fact that the minimum occurs at U_p is not at question, but rather the behavior of γ_Q near this minimum). From this perspective, it appears that other methods of approximating the DKLT, for general non-Toeplitz correlation matrices, along the lines already mentioned in Section 1.3 are needed.

Although similiar conclusions might be expected when we turn to the second task, decorrelation, there is now an interesting twist. Recall that we are using the ad hoc measure of decorrelation defined by Equation (3.2). This is just the average magnitude of the off-diagonal entries of the covariance matrix $Q = UPU^*$. The larger this value, the more decorrelation has taken place under the action of the transform U. Figure 4 shows a typical result for stationary data: both the W-HT and the DFT achieve a significant level of decorrelation, with the latter slightly better on average (64 vs 58 percent). For such data the DCT and DFT are essentially comparable, with a slight advantage to the former (73 vs 64 percent); the W-HT more decisively outperforms the Russian groups transform (58 vs 40 percent).

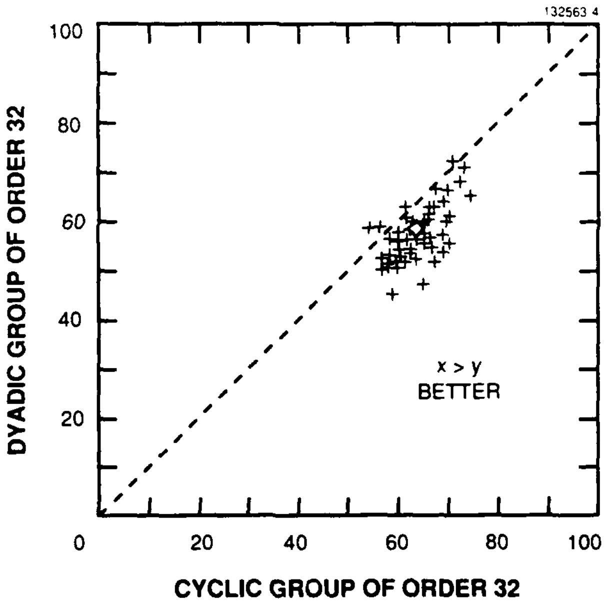


Figure 4. Decorrelation data (percent) — stationary.

Now let us consider nonstationary data. We first report that the DCT, W-HT, and Russian groups transform all perform essentially identically. However, the DFT actually produces, on average, a *negative* degree of decorrelation, indicating an *increase* in overall correlation. This is shown in Figures 5 through 7 for the cases of data dimension = 16, 32, and 64, respectively.

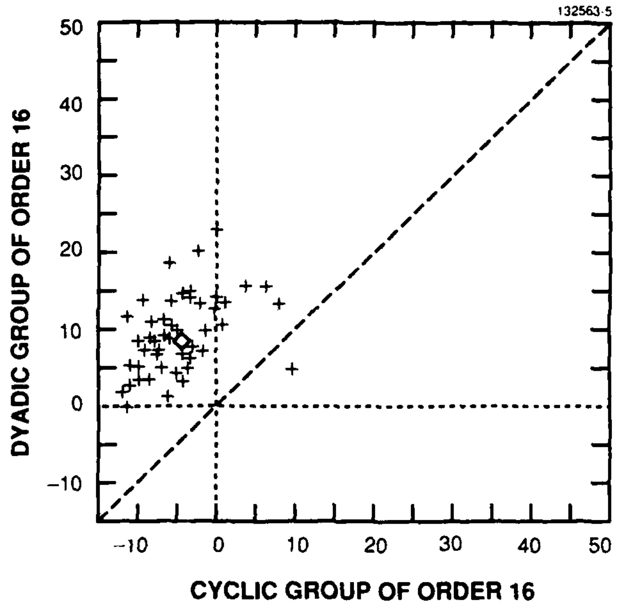


Figure 5. Decorrelation data (percent) — nonstationary.

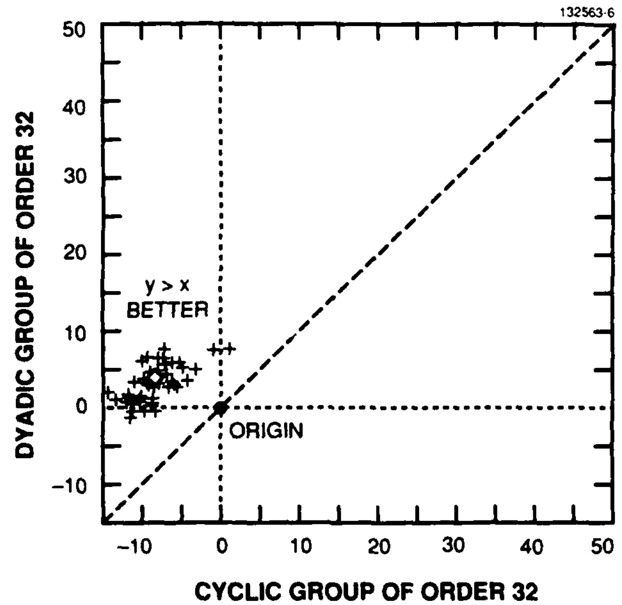


Figure 6. Decorrelation data (percent) — nonstationary.

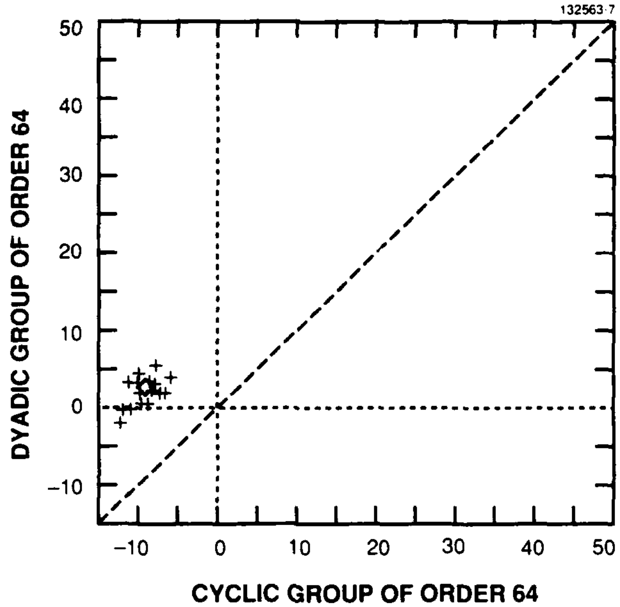


Figure 7. Decorrelation data (percent) — nonstationary.

From such results we conclude that, as with the case of data compression, the DFT/DCT should be used with stationary data. However, for nonstationary data the DFT should definitely not be used. The other transforms all achieve at least a slight decorrelation although, as before, not enough to be of practical interest. It appears that the DFT, so well adapted to stationary data, is a singularly bad transform otherwise. We even ran some experiments with *random* unitary matrices vs the DFT and, on average, the DFT underperformed the random matrices in decorrelating nonstationary data.

However, there is a fringe benefit here. The empirical observation that the DFT is so good (resp., poor) a decorrelator of stationary (resp., nonstationary) data suggests its usage as a test for stationarity. That is, given a stretch of data we compute its DFT and then the measure of decorrelation relative to the original data. A significant level of this figure, say $\geq$20 percent, would be taken to imply stationarity — otherwise, nonstationarity. Our simulation suggests such a test would be 100-percent accurate for data dimension ≥ 8, but a more vigorous justification should still be given. The problem lies in passing from the data stretch, or its transform, to a covariance matrix. The associated sampling errors will act to reduce the test accuracy.

Finally, we turn to the matter of Wiener filtering. Hence, we display results only in dimension 32 and for SNR = +5 dB. Recalling from the preceding section that larger values of the filtering performance measure are worse than smaller values, we see from Figure 8 that, once again, the cyclic group is better for stationary data than the dyadic group, which in turn is slightly better than the Russian groups. The average increase in mean-square error for these optimal group filters over the Wiener-filter error is 3.5, 7.0, and 8.3 percent, respectively.

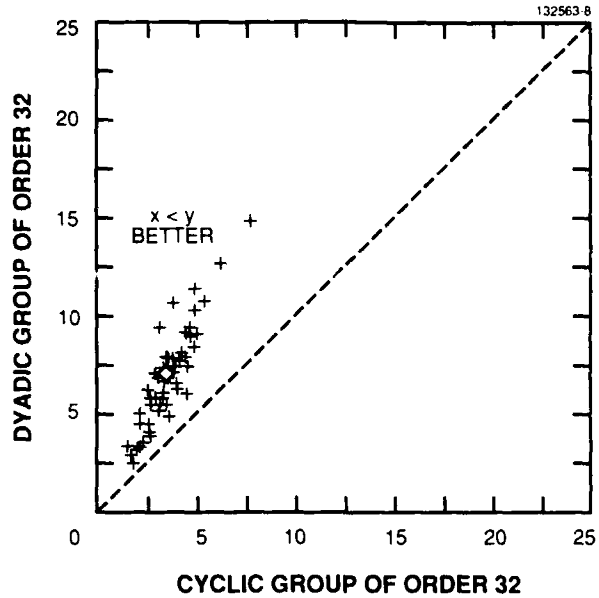


Figure 8. Group filter error increase over optimal (percent) — stationary, SNR = 5 dB.

For nonstationary data, we observe from Figures 9 and 10 a definite, although small, advantage to the dyadic group optimal transform, and a virtual equivalence of performance between this transform and the Russian groups transform.

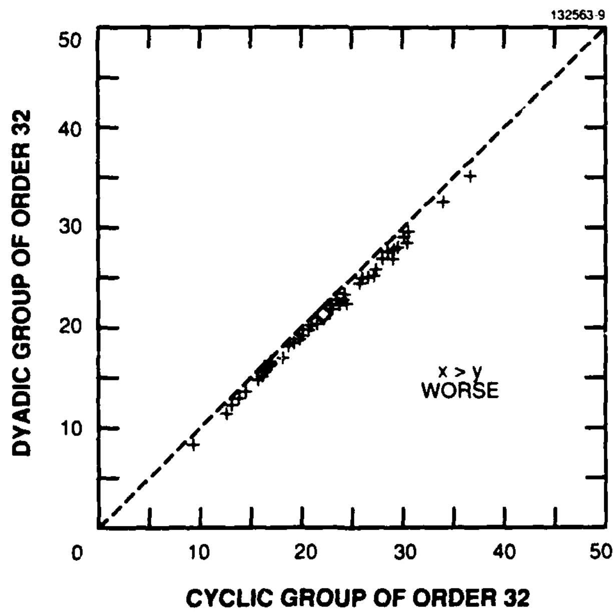


Figure 9. Group filter error increase over optimal (percent) — nonstationary, SNR = 5 dB.

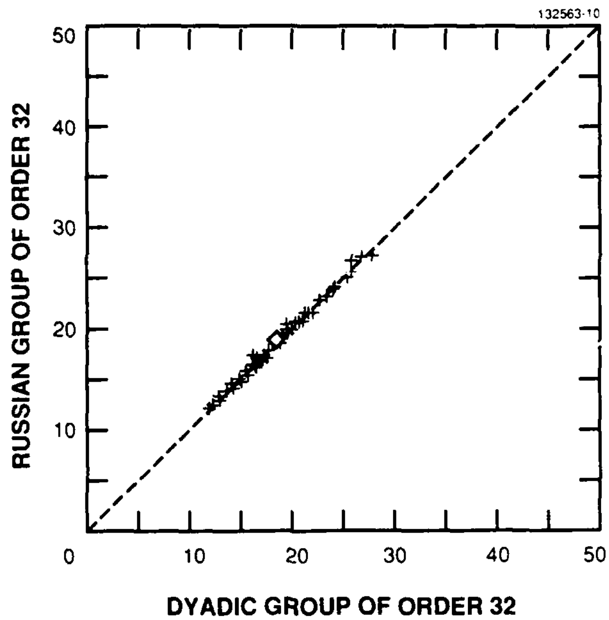


Figure 10. Group filter error increase over optimal (percent) — nonstationary, SNR = 5 dB.

Based on these and related figures for other dimensions and SNRs, we can draw the following conclusion. For filtering stationary data, the use of ordinary digital filters (with computation perhaps accelerated by means of the associated FFT) is recommended. For nonstationary data, the use of the group filters based on the Russian groups is recommended because of the slightly lower error rate and the decrease in complexity of the associated group transform.

At this time, two other remarks are appropriate. The first concerns the dependence of the foregoing conclusions on the SNR level. As SNR increases, all the group filtering errors tend to collapse back to the optimal, which in turn is decreasing to zero [as it must, according to Equations (3.4) and (3.7)]. As SNR decreases, the Wiener filter W tends to 0 from Equation (3.4) (more generally, the resolvent operator vanishes at ∞), and the mean-square filtering error tends to $N = tr(\Sigma_s)$, according to Equation (3.3). In this situation, while the absolute error increases to N, as just noted, the relative group-filter error appears to decrease to 0. It should be checked here whether the mathematics of optimal group filtering necessarily forces the optimal group filter to the zero-operator. We have also noticed a slightly larger degree of scatter in the scatter plots as SNR decreases, so as to favor the use of the dyadic/Russian groups filters.

The second remark pertains to the dependence of our conclusions on the data dimension N. The basic observation here is that, with increasing N, all the point clouds in the various scatter plots tend to decrease in size, both visually and as measured by the magnitude of the determinant of the 2×2 sample covariance matrix of the point cloud (sometimes called the generalized sample variance). This indicates that the effect of the various transforms/filters is becoming constant across the various signal classes. In the filtering studies we also noted that the contracting point cloud tended to cluster along the 45° line as N increased, suggesting an asymptotic equality of performance. Insofar as this is truly valid, it suggests that the main criterion for choice of group, at least for filtering applications, is the complexity of its group transform. This observation thereby tends to validate the emphasis given in the recent mathematical literature to transform complexity rather than error analysis. In particular, for the small collection of groups considered in this work, it suggests the use of the Russian groups filters for Wiener filtering of large dimensional nonstationary data.

The alert reader may have noted the absence of any mention of results pertaining to the dihedral groups, even though this was one of the four classes of groups originally discussed in Section 2. Because it became apparent rather early in our work that the error rates for these groups agreed very closely with those of the cyclic groups, there was really not much point in our suggesting their use in place of the so much more familiar cyclic group transforms (DFT) and digital filters.

7. SUMMARY

The progress of this research project to date can perhaps be summarized best along the lines of

- Where we were
- Where we are
- Where we may be.

Prior to the inception of this work, a highly developed theory existed of (weakly) stationary time series, indexed by either the group of integers or the group of all real numbers, with most of the basic constructs carrying over the case where the index set is a general locally compact abelian group. Here, "basic constructs" include Fourier transform, power spectrum, autocorrelation function, convolution operators (linear invariant filters), etc. For applications, there was the discrete Fourier transform (DFT) on the cyclic group of arbitrary order, and various fast algorithms (FFT) for its computation, along with some asymptotic results describing its effect of long segments on a stationary series. Given this theory and the efficient computational procedures, there would have seemed to be no particular reason to expect to discover any significantly better methods with other groups and, indeed, our simulations have supported this belief. As classical transforms well suited to the treatment of stationary data, we are including the discrete cosine transform (DCT) here along with the DFT.

Based on the work and computer experiments reported here, we can announce a small step beyond the state of knowledge just described. Namely, when dealing with general nonstationary data, other group transforms and filters exhibit both better error rates and lower complexity. These are based on both abelian (e.g., dyadic) and nonabelian (e.g., Russian) groups. Three further remarks are pertinent here. First, while the decrease in error rates is not high, we should bear in mind that the results are based on truly random and unstructured data, so there possibly might be a more significant improvement for data derived from a particular nonstationary model. Second, the alternate transforms are just as easy to use as the DFT. Third, there is a general methodology, based in part on the use of random correlation matrices, to permit comparison of arbitrary transforms and filters. The development of both the theory and the computer generation of these matrices has been an important by-product of this project.

Finally, several open questions of general nature remain which we want to record once more. First, there is a case to be made that of the two familiar abelian group transforms studied here, namely the DFT and W-HT, the latter is somehow the one of choice — if a choice must be made. It has by far the simpler arithmetic, exhibits almost comparable performance on stationary data, and better performance across the broader range of nonstationary data. Is there a theoretical basis for this situation? Second, there is the circle of questions raised near the end of Section 2 concerning the existence of other "superfast" (and presumably nonabelian) groups of order 2^n, besides the Russian groups, and their relevance for signal-processing applications. What are the limits of such group-based performance and complexity? And, as a final speculation, we can wonder whether the theory of group transforms and filters is in any reasonable sense sufficient for the needs of signal processing. That is, we have made much of the fact that it offers a unified approach to fast transforms and filters, and we conclude by wondering: Are these the only such transforms and filters that need be considered?

ADDENDUM

During the preparation of this report, two other articles appeared that bear on the "Russian groups" defined in Section 2. These articles, by Clausen [29] and Rockmore [30], both exhibit a large class of nonabelian groups G for which the group transform F_G and its inverse can both be computed in $O(|G| \log |G|)$ operations, where $|G| := \text{ord}(G)$. The groups to which this result applies are the so-called metabelian groups. Recall that G is metabelian if and only if G has normal abelian subgroup H for which the quotient group G/H is also abelian. Equivalently, the commutator subgroup of G is abelian.

The result of Clausen that is particularly relevant here is that any metabelian 2-group has a group (Fourier) transform F such that both F and its inverse can be computed in at most $1.5\ |G| \log |G|$ operations. The term "operations" means, in the present context, additions/subtractions and (scalar) multiplications. We use the more neutral symbol F rather than F_G to emphasize the *non-uniqueness* of the group transform for nonabelian groups, due to the presence of at least one irreducible representation of degree ≥ 2. Now, by the remarks made about the Russian groups sequence $\{BG_n\}$ in Section 2, they are metabelian 2-groups so this general theory applies. However, for those groups the constant 1.5 can be halved, as $n \to \infty$; that was the whole point of [27]. So we now have a refinement of our earlier question about superfast groups; namely, it would seem sensible to restrict our attention to the subclass of metabelian groups, and to attempt a complete understanding of its potentialities before considering any further generality.

REFERENCES

1. R.B. Holmes, "Mathematical foundations of signal processing II. The role of group theory," MIT Lincoln Laboratory, Lexington, Mass., Technical Rep. 781 (13 October 1987). DTIC AD-A188482.

2. T.S. Durrani *et al.* (eds.), *Mathematics in Signal Processing*, New York: Oxford University Press (1987).

3. R.M. Young, *An Introduction to Nonharmonic Fourier Series*, New York: Academic Press (1980).

4. S. van Eijndhoven and J. de Graaf, *A Mathematical Introduction to Dirac's Formalism*, New York: North-Holland, Elsevier (1986).

5. I. Daubechies, A. Grossmann, and Y. Meyer, "Painless nonorthogonal expansions," *J. Math Phys.* 27, 1271 (1986).

6. A. Grossmann, J. Morlet, and T. Paul, "Transforms associated to square integrable group representations I. General results," *J. Math. Phys.* 256, 2473 (1985).

7. J. von Neumann, *Mathematical Foundations of Quantum Mechanics*, Princeton, NJ: Princeton University Press (1955).

8. D. Gorenstein, "The enormous theorem," *Sci. Am.* 253, 104 (1985).

9. J.J. Rotman, *The Theory of Groups*, Boston, Mass.: Allyn and Bacon, Inc. (1965).

10. R.B. Holmes, "On random correlation matrices," MIT Lincoln Laboratory, Lexington, Mass., Technical Rep. 803 (28 October 1988). DTIC AD-A202786. To appear in *SIAM J. Matrix Anal. Appl.*

11. ___________, "On random correlation matrices II. The Toeplitz case," MIT Lincoln Laboratory, Lexington, Mass., Technical Rep. 816 (23 March 1989), DTIC AD-A208229. To appear in *Commun. Statist.*

12. P. Divijver and J. Kittler, *Pattern Recognition*, Englewood Cliffs, NJ: Prentice Hall (1982).

13. A.I. Solodovnikov, "Formation of optimal Karhunen-Loève bases by means of a fast transform algorithm," *Voprosy Teor. Sistem Avtomat. Upravlenija* 5, 10 (1980).

14. I.E. Kaporin, "A fast algorithm for a class of unitary transformations," *Sov. Math.* Engl Transl 28, 316 (1983).

15. M. Hamidi and J. Pearl, "On the residual correlation of finite-dimensional discrete Fourier transforms of stationary signals," *IEEE Trans. Inf. Theory* IT-21, 480 (1975).

16. J. Pearl, "On coding and filtering stationary signals by discrete Fourier transforms," *IEEE Trans. Inf. Theory* IT-19, 229 (1973).

17. M. Unser, "On the approximation of the discrete Karhunen-Loève transform for stationary processes," *Signal Process.* 7, 231 (1984).

18. R.M. Gray, "On the asymptotic eigenvalue distribution of Toeplitz matrices," *IEEE Trans Inf. Theory* IT-18, 725 (1972).

19. M. Karpovsky and E. Trachtenberg, "Some optimization problems for convolution systems over finite groups," *Inf. Control* 34, 227 (1977).

20. ______________________________, "Filtering in a communication channel by Fourier transforms over finite groups," in *Spectral Techniques and Fault Detection*, M. Karpovsky (ed.), New York: Academic Press (1985), pp. 179-216.

21. E.A. Trachtenberg, "Construction of group transforms subject to several performance criteria," *IEEE Trans. Acoust. Speech Signal Process.* ASSP-33, 1521 (1985).

22. _______________, "A remarkable discrete unitary transform," pp. 637-650 in Reference 2.

23. M.G. Karpovsky, "Fast Fourier transforms on finite non-abelian groups," *IEEE Trans. Comput.* C-26, 1028 (1977).

24. M.D. Atkinson, "The complexity of group algebra computations," *Theo. Comput. Sci.* 5, 205 (1977).

25. T. Beth, "On the computational complexity of the general discrete Fourier transform," *Theo. Comput. Sci.* 51, 331 (1987).

26. P. Diaconis and D. Rockmore, "Efficient computation of the Fourier transform on finite groups," Stanford University Dept. of Statistics, Stanford, Calif., Technical Rep. 292 (1988).

27. S. Berman and I. Grushko, "The theory of discrete signal processing," *Probl. Inf. Transm.* 19, 284 (1984).

28. J. Huang and P. Schultheiss, "Block quantization of correlated Gaussian random variables," *IEEE Trans. Circuits Syst.* CS-11, 289 (1963).

29. M. Clausen, "Fast Fourier transforms for metabelian groups," *SIAM J. Comp.* 18, 584 (1989).

30. D. Rockmore, "Fast Fourier analysis for abelian group extensions," to appear in *Adv. Appl. Math.*

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. 27 February 1990	3. REPORT TYPE AND DATES COVERED *Technical Report*

4. TITLE AND SUBTITLE

Signal Processing on Finite Groups

5. FUNDING NUMBERS

C — F19628-90-C-0002
PE — 63304A, 65301A, 63308A
PR — 34

6. AUTHOR(S)

Richard B. Holmes

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Lincoln Laboratory, MIT
P.O. Box 73
Lexington, MA 02173-9108

8. PERFORMING ORGANIZATION REPORT NUMBER

Technical Report 873

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Strategic Defense Command — Huntsville
Sensor Directorate
P.O Box 1500
Huntsville, AL 35807-3801

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

ESD-TR-89-267

11. SUPPLEMENTARY NOTES

None

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

A unified approach to the design and evaluation of fast algorithms for discrete signal processing is developed. Based on the theory of finite groups, it hence includes the familiar cases of the fast Fourier and Walsh-Hadamard transforms. However, the use of noncommutative groups reveals a large variety of novel methods. Some of these exhibit a superior performance, as measured by both reduced error rates and computational complexity, on nonstationary data.

The recent history of this subject is reviewed first, followed by a detailed examination of the three principal ingredients of the present study: the underlying groups, the signal-processing tasks on which the group-based algorithms are to compete, and the signal models used to define the data environment. Test results and conclusions then follow, the former being based on the use of random correlation matrices.

14. SUBJECT TERMS

signal processing	group filter	Walsh-Hadamard transform
finite groups	fast Fourier transform	random correlation matrix
group transform		

15. NUMBER OF PAGES 50

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by AMSI Std. 239-18
298-102