Ordered and Disordered Source Coding

Lav R. Varshney and Vivek K Goyal
Research Laboratory of Electronics
Massachusetts Institute of Technology

***Abstract*— The separation of the information in a vector into order (a permutation) and value (a multiset) is studied. We formulate rate-distortion problems in which order is irrelevant, for which we give conclusive results in cases where the set size approaches infinity. Conversely, we present a distortion measure on partial orders and source coding techniques based on partial orders. Multiple partial orders can be mutually refining, leading to a form of permutation coding for multiple descriptions. Permutation coding of a harmonic frame expansion gives low-complexity vector quantizers with low-rate performance superior to that of entropy-constrained scalar quantization.**

I. Introduction

Many students struggle at the beginning of a probability class with the fact that sometimes order matters and other times it does not. By rote, they start dividing counts by $n!$ when a problem involves the dealing of n cards. Indeed, in most games in which cards are dealt the order of the cards is immaterial. But an audience's expectation that the order of cards is irrelevant is exploited in various card tricks, such as Fitch's Five Card Trick.[1] In the trick, covert communication is established between the magician and his/her assistant by encoding information in the order of cards.

The distinction between order information and value information is not just useful for card tricks; it is also useful in compression and communication. For some source-destination pairs, order is irrelevant so a *sequence* of source letters can be treated as a *set* of source letters, with commensurate reduction in the required coding rate [3]. Conversely, when value information is irrelevant, a sequence of source letters can be treated as a *permutation* of the integers. When both order and value are relevant, separating the two may be a convenient and effective divide-and-conquer strategy for source coding.

An example in which the source letters are naturally treated as a set is the storage of records in a database. The order of records is generally irrelevant; all that matters is the ability to retrieve them. This property also applies to human memory, as noted in studies of psychological experiments on recognition and recall [4]. One way to exploit the irrelevance of order is to find a reordering that maximimizes compressibility with some fixed compressor [5]. Another example where the order of source letters is irrelevant is the communication of kernel-based signal decompositions and density estimates, where the signal reconstruction is a permutation-invariant function [3], [6]. There are also sources, such as arrival processes, for which the order is predetermined.

Contrary to order-irrelevant sources, there are sources for which the order is the only relevant information. Prime examples are offered by the comparative probability theory [7]. A typical comparative probability statement, when tossing a strange coin is "*heads* is at least as probable as *tails*." Other examples include the rankings in sports leagues. Similarly, for the results of imprecise or data-limited hypothesis testing.

Outline

This paper gathers several new and recent results on coding only value or only order. In the next section, we discuss the partitioning of order and value information. Then, Section III develops source coding results for order-irrelevant scenarios. The fidelity criteria that arise are not single-letter, however they are still sufficiently tractable to allow information theoretic statements. Section IV similarly presents a possible fidelity criterion for order—with no dependence on value—and develops several ramifications. The coding of partial order (only) with the goal of reproducing a vector is the essence of permutation source codes [8]. The bulk of the section provides new variants on permutation codes, one for multiple description coding and the other employing a frame expansion. The final section gives closing comments.

II. Separating Order and Value

Transform coding has been the workhorse of practical source coding [9], with a great bias towards linear, orthogonal transforms. The Karhunen-Loéve Transform (KLT) is often considered as an optimal transform since it yields independent transform coefficients for Gaussian sources. For non-Gaussian sources, however, lifting the orthogonality and linearity restrictions on the transform can often yield better source coding results, e.g. transformation to phase and magnitude yields improvements for Gaussian scale mixture sources [10].

Here we consider sorting as a nonlinear transform that produces order and value as "transform coefficients." We use a collapsed two-line notation for permutations to express the output of sorting.

$$(x_i)_{i=1}^n \xrightarrow{\text{sort}} \begin{pmatrix} i_1 & i_2 & \cdots & i_n \\ x_1 & x_2 & \cdots & x_n \end{pmatrix} = \begin{pmatrix} j_n \\ \{x_i\}_{i=1}^n \end{pmatrix},$$

where the indices $i_1, \ldots, i_n$ are a permutation of the integers $1, \ldots, n$ for sets, but have repetitions for multisets. The ordering is collapsed into a single variable j. Similarly, the values $x_1, \ldots, x_n$ are collapsed into a set $\{x_i\}_{i=1}^n$.

Supported in part by Texas Instruments Leadership University Consortium.

[1] This card trick and variations of it seem to have been popularized in the information theory community by Elwyn Berlekamp and Thomas Cover [1], among others. For extensions, achievability proofs, and information theoretic converses, see [2] and references therein.

The two transform coefficients, j (order) and $\{x_i\}_{i=1}^n$ (value), are independent for i.i.d. sources. An information theoretic demonstration involves decomposing the entropy into two independent parts. Define $H((X_i)_{i=1}^n)$ as entropy of the original sequence, $H(\{X_i\}_{i=1}^n)$ as multiset entropy, and $H(J_n)$ as the entropy of the chance variable representing order. Suppressing subscripts,

$$\begin{aligned} H((X)) &\overset{(a)}{=} H((X)) + H(\{X\}) - H((X)|J) \\ &= H(\{X\}) + I((X);J) \\ &\overset{(b)}{=} H(\{X\}) + H(J) - H(J|(X)) \\ &= H(\{X\}) + H(J), \end{aligned} \tag{1}$$

where (a) follows from noting that $H(\{X\}) = H((X)|J)$ for i.i.d. sources and (b) from $H(J|(X)) = 0$ in general. Since the transform coefficients are independent, we can code them separately without loss of optimality. Moreover, a rate allocation favoring one or the other should often be used, much as some subbands are favored in linear transform coding.

III. Order Irrelevance

When order is irrelevant, the source coding problem reduces to a problem of coding multisets, detailed in [3]. We present cases when the source is drawn from a finite alphabet and from a continuous alphabet separately. Fidelity criteria where multiset size remains fixed and where it grows with block length are also discussed separately.

A. Growing Multisets from Finite Alphabets

In the finite situation, the multiset of values that we want to code can be represented by the type (empirical distribution) of the multiset. Consequently, there is no requirement for an order on the source alphabet. An alternative view is that an arbitrary space-filling curve through the alphabet can be used to define an order. As a result, the theory applies not just to alphabets of numerical scalars, but to alphabets of numerical vectors, language texts, images, and videos. Since the multiset is equivalent to a type, we can use the type to define the word distortion measure and the fidelity criterion. The word distortion measure of length n is defined to be

$$d_n(x_1^n, y_1^n) = \begin{cases} 0, & \mathrm{type}(x_1^n) = \mathrm{type}(y_1^n) \\ 1, & \mathrm{type}(x_1^n) \neq \mathrm{type}(y_1^n) \end{cases}$$

and the fidelity criterion is

$$F_{\mathrm{val}_1} = \{d_n(x_1^n, y_1^n), n = 1, 2, \ldots\}.$$

This is not a single-letter fidelity criterion on the source, though it does have a frequency of error interpretation on the types. As for other fidelity criteria [11], finding an upper bound to the $R(0)$ point reduces to an enumeration problem, here the enumeration of types. By well-known combinatorics,

$$R(0) \leq \lim_{n\to\infty} \tfrac{1}{n} \log |\mathcal{K}(\mathcal{X}, n)| \leq \lim_{n\to\infty} \tfrac{1}{n} \log (n+1)^{|\mathcal{X}|} = 0,$$

where $|\mathcal{K}(\mathcal{X}, n)|$ is the number of types for source alphabet $\mathcal{X}$ and blocklength n. By the information inequality, $R(D) \geq 0$, so $R(D) = 0$ for any source with fidelity criterion F_{val_1}.

B. Fixed-Size Multisets from Finite Alphabets

For fidelity criterion F_{val_1}, as n increases, larger and larger multisets are used to compute the distortion. This trivializes the problem, in the sense that no rate is needed to achieve zero distortion. An alternate view of order irrelevance is to fix the size of the multiset, at say N, and take larger and larger numbers of multisets. In effect, we now have a block source with a single-letter fidelity criterion. Using the previous word distortion on blocks of length N, the fidelity criterion is

$$F_{\mathrm{val}_2} = \left\{ \frac{N}{n} \sum_{k=1}^{n/N} d_N(x_{kN-N+1}^{kN}, y_{kN-N+1}^{kN}), n = N, 2N, \ldots \right\}.$$

Since the F_{val_2} notion of fidelity casts the problem into a frequency of error framework, assuming that the multisets to be coded are independent and identically distributed, the results of Erokhin apply [12]. If the letters within the multisets are also i.i.d., the probability values needed for the reverse waterfilling characterization are computed using the multinomial distribution. The multiset types used in the representation alphabet are given by the type classes in the typical set

$$T_{p_X}^{\epsilon(\theta)} = \{k : D(p_k \| p_X) \leq \epsilon(\theta)\},$$

where p_X is the source density used to generate the multinomial distribution, p_k is the probability created by normalizing the type k, θ is the rate distortion curve slope parameter, and $D(\cdot\|\cdot)$ is the relative entropy.

C. Growing Sets from Continuous Alphabets

Assume that the source alphabet consists of real scalars, with the usual ordering.[2] When source letters are drawn i.i.d. from a continuous alphabet, the associated multiset is almost surely a set. Rather than using types as natural representatives for sets, we use order statistics. When the sequence of random variables $X_1, \ldots, X_n$ is arranged in ascending order as $X_{(1:n)} \leq \ldots \leq X_{(n:n)}$, $X_{(r:n)}$ is called the rth order statistic. We define the word distortion measure to be

$$\rho_n(x_1^n, y_1^n) = \frac{1}{n} \sum_{i=1}^{n} (x_{(i:n)} - y_{(i:n)})^2,$$

and an associated fidelity criterion to be

$$F_{\mathrm{val}_3} = \{\rho_n(x_1^n, y_1^n), n = 1, 2, \ldots\}.$$

Although not a single-letter fidelity criterion, it is single-letter mean square error on the block of order statistics. For a very large class of sources, the rate distortion function for fidelity criterion F_{val_3} is $R(D) = 0$ [3].

D. Fixed-Size Sets from Continuous Alphabets

Defining a fourth value fidelity criterion, F_{val_4}, in the same manner as F_{val_2}, but substituting $\rho_N(\cdot,\cdot)$ for $d_N(\cdot,\cdot)$, we have an order statistic source coding problem. For the F_{val_4} problem with i.i.d. sets, first look at the high rate quantization theory.

[2] Ordering of real vectors and other continuous sources without natural orders is problematic [13], as is defining multivariate quantile functions [14].

For a high rate uniform quantizer operating on the conditional order statistic distributions with step size Δ, the rate

$$R = \tfrac{1}{N} h(X_1, \ldots, X_N) - \tfrac{1}{N} H(J_N) - \log \Delta,$$

where $H(J_N)$ is as in Section II. The associated distortion is $D = \Delta^2/12$; some further gain is possible if vector quantization is used. When the set elements are drawn from an exchangeable distribution, $h(X_1, \ldots, X_N)/N = h(X_1)$ and $H(J_N) = (\log N!)/N$.

For low rates and coding one set at a time, we can find optimal quantizers through the Lloyd-Max procedure.[3] These quantizers have interesting geometrical relationships to quantizers for more standard fidelity criteria. First note that the joint density of all N order statistics is

$$f_{(1:N)\cdots(N:N)}(x_1, \ldots, x_N) = \begin{cases} N! \prod_{i=1}^{N} f(x_i), & x_1^N \in \mathfrak{R}; \\ 0, & \text{else.} \end{cases}$$

The region of support, $\mathfrak{R} = \{x_1^N : x_1 \leq \cdots \leq x_N\}$, is a convex cone that occupies $(1/N!)$th of $\mathbb{R}^N$.

If the representation points for an MSE-optimal $(k \text{ bit}, n \text{ dimension})$ order statistic quantizer are the intersection of the representation points for an MSE-optimal $(k + \log n!, n)$ quantizer for the unordered variates and $\mathfrak{R}$, then we can interchange sorting and quantization without loss of optimality. This condition can be interpreted as a requirement of permutation polyhedral symmetry on the quantizer of the unordered variates. In fact, the distortion performance of the MSE-optimal (k, n) order statistic quantizer is equal to the distortion performance of the best $(k + \log n!, n)$ unordered quantizer constrained to have the required permutation symmetry.

IV. Coding Order

Dual to Section III, we discuss the problem of coding only order information. We are interested in coding the chance variable J_n, which represents a permutation of a multiset. For ease of exposition, assume no ties occur; the extension is straightforward.

A. Value Irrelevance

Define a binary relation $\bar{\bar{\barwedge}}$ as a *preorder* on the source set if it satisfies the reflexive ($a \bar{\bar{\barwedge}} a$ for all a in $\{X\}$) and transitive ($a \bar{\bar{\barwedge}} b$ and $b \bar{\bar{\barwedge}} c$ implies $a \bar{\bar{\barwedge}} c$) properties. A preorder that also satisfies antisymmetry ($a \bar{\bar{\barwedge}} b$ and $b \bar{\bar{\barwedge}} a$ implies $a = b$) is called a *partial order*. Denote it with the binary operator $\leq$. A total order also satisfies comparability (for any a, b in $\{X\}$, either $a \leq b$ or $b \leq a$).

One can think of a total order as a composition of comparability relations, where knowledge of comparability converts a preorder into a partial order. When all elements can be compared, a total order results. As an example for $n = 4$, assume that the true total order is

$$x_\alpha \overset{(a)}{\leq} x_\beta \overset{(b)}{\leq} x_\gamma \overset{(c)}{\leq} x_\zeta.$$

Then there are $n-1$ possible comparability relations that need to be established explicitly, denoted above as $\{(a), (b), (c)\}$. Other relationships, such as $x_\alpha \leq x_\gamma$ are established implicitly by transitivity. As represented, knowledge of comparability can be treated as a binary vector of length $n-1$. This leads to a distortion measure between a source vector, x_1^n, in terms of its order, $j(x_1^n)$ and a representation order, k, based on how much of the total order j is recreated by the order k. For $n > 1$

$$\delta_n(j(x_1^n), k) = \begin{cases} \frac{1}{n-1} d_H(\tilde{j}_1^{n-1}, \tilde{k}_1^{n-1}), & j \approx k \\ \infty, & \text{else} \end{cases}$$

The binary vectors $\tilde{j}_1^{n-1}$ and $\tilde{k}_1^{n-1}$ represent knowledge of comparability, and $d_H(\cdot, \cdot)$ is the Hamming distance. The relation $\approx$ implies that the orders j and k are consistent with each other. By consistency, we mean that there are no ordering requirements under j and k that are contradictory. For example, the distortion between orderings

$$\begin{aligned} j &: x_\alpha \leq x_\beta \leq (x_\gamma \bar{\bar{\barwedge}} x_\zeta) \\ k &: x_\alpha \leq x_\beta \leq x_\gamma \leq x_\zeta \\ \ell &: x_\alpha \leq x_\beta \leq x_\zeta \leq x_\gamma \end{aligned}$$

satisfy $\delta(j, k) = 1/3$, $\delta(j, \ell) = 1/3$, $\delta(k, \ell) = \infty$. We penalize erasure of comparability knowledge by finite distortion, but error in comparability knowlege by infinite distortion. Without loss of optimality, a source code never uses inconsistent reproductions, since the order with no defined comparability relations is consistent with all total orders and has maximum distortion 1. The fidelity criterion generated by this distortion measure in the manner of F_{val_1} is denoted F_{ord}.

Permutation source codes [8], [16] operating on the original source vector provide a readily-instrumentable source coding method for transmitting partial order relations. If one uses a permutation code with block size n and group sizes $(n_1, \ldots, n_K)$, then the members within each group are incomparable, whereas the groups themselves may be compared to each other. The power set of the possible comparability relations establishes the possible permutation codes (for the $n = 4$ case) in Table I. This may also be drawn as a lattice, as in Fig. 1. One can see that nodes on the same level of the lattice result in the same F_{ord} distortion performance. Since reaching different lattice nodes requires different rates, source coding simply involves choosing the low rate node on the desired distortion level. This is optimal for F_{ord}.

B. Multiple Descriptions of Partial Order

The lattice of partial orders achieved by permutation codes implies that successive refinement and multiple description problems are problems in choosing paths through the lattice, not just for F_{ord}, but also for other fidelity criteria in the operational sense.

[3] Unlike in [15], where each element of the set is coded independently, not taking advantage of the Markovian dependence among elements, we code the entire set together, referring to N as the dimension of the order statistic vector quantizer.

TABLE I

PERMUTATION CODES, COMPARABILITY RELATIONSHIPS DETERMINED BY CORRESPONDING PARTIAL ORDERS, AND LABELS.

$[1,1,1,1]$	$\{a,b,c\}$	U
$[2,1,1]$	$\{b,c\}$	F
$[1,2,1]$	$\{a,c\}$	E
$[1,1,2]$	$\{a,b\}$	D
$[2,2]$	$\{b\}$	B
$[3,1]$	$\{c\}$	C
$[1,3]$	$\{a\}$	A
$[4]$	$\emptyset$	O

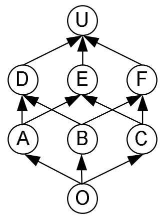

Fig. 1. Lattice of partial orders, where $O = \emptyset$, $A = \{a\}$, $B = \{b\}$, $C = \{c\}$, $D = \{a,b\}$, $E = \{a,c\}$, $F = \{b,c\}$, and $U = \{a,b,c\}$. Note that this lattice actually defines a partial order on the partial orders based on the distortion performance.

For successive refinement of order, sub-permutation codes may be used to define ordering among group elements. The sub-permutation code for the members of the first group would be of blocklength n_1 and group sizes $(m_1,\ldots,m_L)$. When this refinement information is used to supplement the original permutation code information, the distortion performance is exactly equivalent to a blocklength n, group size $(m_1,\ldots,m_L,n_2,\ldots,n_K)$ permutation code. Furthermore, assuming an exchangeable source, the total rate that is required for the two step procedure is

$$\begin{aligned} R_{sr} &= \log \frac{n!}{\prod_{i=1}^{K} n_i!} + \log \frac{n_1!}{\prod_{j=1}^{L} m_i!} \\ &= \log \frac{n! n_1!}{\prod_{j=1}^{L} m_i! \prod_{i=1}^{K} n_i!} = \log \frac{n!}{\prod_{j=1}^{L} m_i! \prod_{i=2}^{K} n_i!}. \end{aligned} \qquad (2)$$

The total rate for a single description permutation code of the same performance would be identical to (2). By repeated application of this property, we see there is no operational rate loss for any exchangeable source in successively refining a permutation code with a sub-permutation code for any distortion measure for which order is relevant.

In a similar fashion, one sees that for the two-channel multiple description problem using permutation codes, we are to choose two nodes in the lattice. For each of these combinations of two nodes, we can compute the multiple description quintuple, $(R_1,R_2,\Delta_0,\Delta_1,\Delta_2)$. The values for R_1, R_2, Δ_1, and Δ_2 are readily apparent, since they are simply the operational rate and distortion of the corresponding single description permutation codes. To determine the central distortion, we make use of the fact that receiving two order informations is equivalent to receiving the union of the comparability relations that are determined by them collectively. The lattice is well-suited to determining the union, and the partial order of partial orders that it defines yields the central distortion. For example, from Fig. 1, the distortion when sending A and B, $D(A,B) = D(A \cup B) = D(D)$ and $D(B,E) = D(B \cup E) = D(U)$.

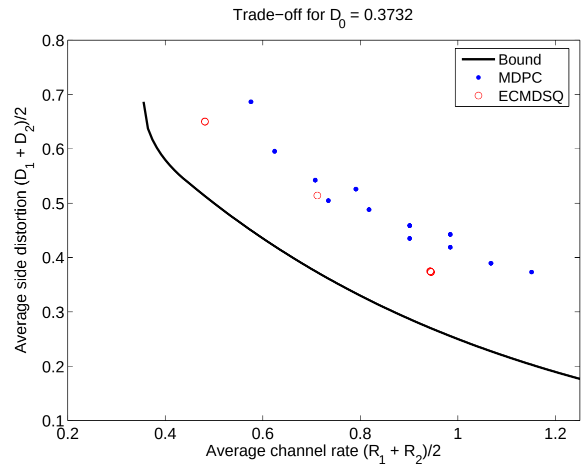

Fig. 2. The performance of some multiple description permutation codes of blocklength 6 on a memoryless $\mathcal{N}(0,1)$ source. Rate in bits; distortion is squared error . Performance of entropy constrained multiple description scalar quantizers (ECMDSQ) [17] and rate distortion bound [18] shown for comparsion. Performance relative to ECMDSQ similar for other values of $D_0 \geq 0.36$ and worse for smaller values of D_0.

The sets of rate and distortion values $\{R_O,R_A,R_B,\ldots,R_F,R_U\}$, $\{D_O,D_A,D_B,\ldots,D_F,D_U\}$, and the lattice structure which defines the union operation are sufficient to compute the (R_1,R_2,D_0,D_1,D_2) multiple descriptions quintuple for all possible combinations of permutation codes on the two channels. Moreover, if the two vertices used for the two descriptions do not lie on a path from O to U, then there is no operational rate loss for any of the three decoders. An example of the performance of a multiple description permutation code for a standard distortion measure is given in Fig. 2.

C. *Coding Order in a Transform Domain*

Contrary to Sections III and IV-A, here we use a fidelity criterion that takes both value and information into account, but consider a coding scheme that only codes for order: the usual application of permutation codes [8], [16]. In a sense, one gets a rough description of the set for free, and by the asymptotic equipartion principle type concentration of the source output onto a constant composition, this rough description is quite good. Operating on the original source output, permutation codes are as good as and sometimes better than scalar quantizers [16].

Rather than applying permutation coding in the original source domain, what if we first transformed the source into a frequency domain? For sources that we have experimented with, it turns out that order in a frequency domain is more

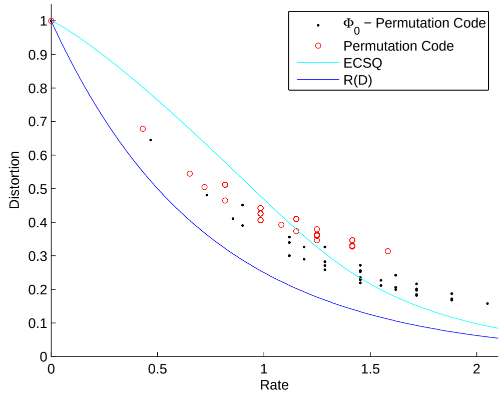

Fig. 3. Operational rate distortion performance for coding memoryless $\mathcal{N}(0,1)$ source. Rate is $\log_2$ of codebook size per source symbol; distortion is squared error per source symbol. Block length 6 codes used. Achievable points shown for directly permutation coding in source domain and for permutation coding in frame expansion domain. The harmonic frame with input block length 6 and output block length 7, Φ_0, and its pseudoinverse used for frame expansion and reconstruction. The performance of entropy constrained scalar quantizers and the rate distortion bound are shown for comparison.

useful (in the rate distortion sense) than order in the original source domain, quite similar to the fact that phase is often much more important than magnitude in the Fourier domain [19], [20]. To transform into a frequency domain, we use overcomplete frame expansions, in particular the the redundancy $(N+1)/N$, overcomplete, harmonic, uniform, tight frame expansion, which is quite similar to a Fourier transform [21]. Since the frames are uniform tight, there is a Parseval like relationship that gives mean square error in one domain equal to the distortion incurred in the other domain. Since the redundancy of the frame is $(N+1)/N$ and the vector sum of the frame vectors is zero, the frame coefficients can be shown to lie in an N dimensional subspace of $\mathbb{R}^{N+1}$. By the property that permutation codewords have the same composition, the permutation codewords will also lie in this subspace, so no out-of-subspace quantization error is incurred [22]. Thus the usual Moore-Penrose pseudoinverse is a consistent (in the sense of [23]) frame reconstruction method. Using the harmonic frame expansion–permutation code–pseudoinverse frame reconstruction system,[4] we generated the operational rate distortion performance shown in Fig. 3. Evidently, there may be improvement.

V. Comments

Our goal here has been to present some theoretical implications of order-irrelevance and value-irrelevance along with practical techniques based on representing only order.

We may interpret the preservation or neglect of order as adopting fidelity criteria of Jerome (Eusebius Hieronymus), the translator of the bible into Latin [24, p. 150]:

> Jerome advises the utmost textual fidelity for sacred works, whose divine qualities come from the very order of the words as well as their meaning. For secular writings, however, such fidelity is a mistake; it "conceals the sense, even as an overgrown field chokes the seeds."

The results of Section III show that coding long secular writings is trivial, whereas coding sacred writing is non-trivial.

Notwithstanding interpretations of Jerome, several non-trivial engineering problems remain. We have no results for lossy coding of sets of vectors, as might be the model for compression of certain numerical databases. Furthermore, we have yet to show how the separation of information into order and values may aid in situations in which sender or receiver has uncertain knowledge of the source distribution.

[4]This system is quite computationally simple since FFT-like algorithms exist for harmonic frame expansion and pseudoinverse reconstruction and since permutation coding is a sorting procedure.

References

[1] T. M. Cover, "Poker hands and channels with known state," in *IEEE Commun. Theory Workshop*, Aptos, CA, 1999.

[2] N. Do, "Mathellaneous: A mathemagical card trick," *Australian Mathematical Society Gazette*, vol. 32, pp. 9–15, Mar. 2005.

[3] L. R. Varshney and V. K. Goyal, "Toward a source coding theory for sets," in *Data Compression Conference*, Mar. 2006.

[4] B. R. Judd and N. S. Sutherland, "The information content of nonsequential messages," *Inf. Control*, vol. 2, pp. 315–332, Dec. 1959.

[5] A. L. Buchsbaum, G. S. Fowler, and R. Giancarlo, "Improving table compression with combinatorial optimization," *J. ACM*, vol. 50, pp. 825–851, Nov. 2003.

[6] A. T. Ihler, J. W. Fisher, III, and A. S. Willsky, "Using sample-based representations under communications constraints," M.I.T. Laboratory for Information and Decision Systems, Tech. Rep. 2601, Dec. 2004.

[7] T. L. Fine, *Theories of Probability: An Examination of Foundations*. New York: Academic Press, 1973.

[8] T. Berger, F. Jelinek, and J. K. Wolf, "Permutation codes for sources," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 160–169, Jan. 1972.

[9] V. K. Goyal, "Theoretical foundations of transform coding," *IEEE Signal Processing Mag.*, vol. 18, pp. 9–21, Sept. 2001.

[10] D. E. Ba and V. K. Goyal, "Nonlinear transform coding: Polar coordinates revisited," in *Data Compression Conference*, Mar. 2006.

[11] T. Berger and W. C. Yu, "Rate-distortion theory for context-dependent fidelity criteria," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 378–384, May 1972.

[12] V. Erokhin, "ε-entropy of a discrete random variable," *Theory Probab. Appl.*, vol. 3, pp. 97–100, 1958.

[13] V. Barnett, "The ordering of multivariate data," *J. R. Stat. Soc. Ser. A. Gen.*, vol. 139, pp. 318–355, 1976.

[14] R. Serfling, "Quantile functions for multivariate analysis: Approaches and applications," *Stat. Neerl.*, vol. 56, pp. 214–232, May 2002.

[15] P. P. Gandhi, "Optimum quantization of order statistics," *IEEE Trans. Signal Processing*, vol. 45, pp. 2153–2159, Sept. 1997.

[16] V. K. Goyal, S. A. Savari, and W. Wang, "On optimal permutation codes," *IEEE Trans. Inform. Theory*, vol. 47, pp. 2961–2971, Nov. 2001.

[17] V. A. Vaishampayan, "Design of multiple description scalar quantizers," *IEEE Trans. Inform. Theory*, vol. 39, pp. 821–834, May 1993.

[18] L. Ozarow, "On a source-coding problem with two channels and three receivers," *Bell Syst. Tech. J.*, vol. 59, pp. 1909–1921, Dec. 1980.

[19] A. V. Oppenheim and J. S. Lim, "The importance of phase in signals," *Proc. IEEE*, vol. 69, pp. 529–541, May 1981.

[20] W. A. Pearlman and R. M. Gray, "Source coding of the discrete Fourier transform," *IEEE Trans. Inform. Theory*, vol. IT-24, pp. 683–692, Nov. 1978.

[21] V. K. Goyal, J. Kovačević, and J. A. Kelner, "Quantized frame expansions with erasures," *Appl. Comput. Harmon. Anal.*, vol. 10, pp. 203–233, May 2001.

[22] L. R. Varshney, "Permutations and combinations," M.I.T. 6.961 Report, Dec. 2004.

[23] V. K. Goyal, M. Vetterli, and N. T. Thao, "Quantized overcomplete expansions in $\mathbb{R}^N$: Analysis, synthesis, and algorithms," *IEEE Trans. Inform. Theory*, vol. 44, pp. 16–31, Jan. 1998.

[24] S. L. Montgomery, *Science in Translation: Movements of Knowledge through Cultures and Time*. The University of Chicago Press, 2000.