A Necessary and Sufficient Condition for Transform Optimality in Source Coding

Emrah Akyol and Kenneth Rose
{eakyol, rose}@ece.ucsb.edu
Department of Electrical and Computer Engineering
University of California, Santa Barbara, CA-93106

***Abstract*—It is well-known for transform coding of multivariate Gaussian sources, that the Karhunen Loeve transform (KLT) minimizes the mean square error distortion. However, finding the optimal transform for general non-Gaussian sources has been an open problem for decades, despite several important advances that provide some partial answers regarding KLT optimality. In this paper, we present a necessary and sufficient condition for optimality of a transform when high resolution, variable rate quantizers are employed. We present not only a complete characterization of when KLT is optimal, but also a determining condition for optimality of a general (non-KLT) transform. This necessary and sufficient condition is shown to have direct connections to the well studied source separation problem. This observation can impact source separation itself, as illustrated with a new optimality result. Finally, we combine the transform optimality condition with algorithmic tools from source separation, to derive a practical numerical method to search for the optimal transform in source coding.**

***Index Terms*—Transform coding, source coding, source separation, quantization**

I. Introduction

Transform coding is a computationally attractive approach to source coding, and is widely used in audio, image and video compression. In the basic transform coding setting, an input vector is linearly transformed into a vector in the transform domain whose components (also called transform coefficients) are scalar-quantized. The decoder reconstructs the quantized coefficients and performs linear (inverse) transformation to obtain an estimate of the source vector. The design goal is to find the optimal transform pair and bit allocation to scalar quantizers, which minimize distortion. In general, transform coding underperforms optimal vector quantization due to space filling loss in scalar quantizers, even if the transform generates independent coefficients. Nevertheless, due to its low complexity, transform coding is commonly employed in practical multimedia compression systems [1], [2].

Transform coding has been studied extensively. In their seminal paper, Huang and Schulthesis have shown [3] that if the vector source is Gaussian and the bit budget is asymptotically large, then the Karhunen Loeve transform (KLT) and its inverse are an optimal pair of transforms for fixed-rate coding. In a more recent paper Goyal, Zhuang and Vetterli improve that result by showing that KLT is optimal for Gaussian sources without making any high resolution assumptions [4]. Their results require a mild scale invariance assumption and apply to both the fixed and the variable rate quantizers.

The optimality of KLT in transform coding of Gaussian sources is often explained intuitively by the assertion that scalar quantization is better suited to the coding of independent random variables than to the coding of dependent random variables. Thus, the optimality of KLT for transform coding of Gaussian sources is understood to be a consequence of the fact that it yields independent transform coefficients. The application of KLT in transform coding of non-Gaussian sources is then justified using the intuitive argument that KLT's coefficient decorrelation represents, for general sources, a rough approximation to the desired coefficient independence.

In [5], the "popular trust" in the optimality of KLT is challenged and it is demonstrated by examples that KLT can be suboptimal for both fixed and variable rate quantization, at asymptotically high rate (with high resolution approximations). A theoretical result is also obtained, namely, a sufficient condition for optimality of KLT: when KLT generates independent coefficients then it is the optimal transform for variable rate coding.

In [6], a significant positive result is obtained regarding the optimality of KLT: KLT is optimal in conjunction with variable rate high resolution coding, not only for Gaussians but for the broader family of Gaussian vector mixtures, which includes Gaussian mixture models.

The problem is approached from a more practical perspective of numerical design in [7]. The authors proposed a gradient descent iterative algorithm to optimize the optimal orthogonal transform in conjunction with optimization of the quantization scheme. In simulations, they were able to demonstrate performance gains of the optimized transform-quantizer pair over KLT for practical sources.

In this paper, we return to the fundamental theoretical problem of optimal transform coding. The main result is a necessary and sufficient condition for optimality of a transform in conjunction with variable rate coding at high resolution. Specifically, we show that the optimal transform is the one that minimizes the divergence between the joint distribution of the coefficients and the product of their marginals. In other words, it minimizes a quantitative measure of the dependence between the transform coefficients. Note furthermore that this result not only resolves the question of when KLT is optimal (at high

This work is supported in part by the NSF under grants CCF-0728986 and CCF-1016861

resolution), but it also determines the optimal transform when it is not KLT.

We note minimizing a measure of dependence is closely related, at the high level, to the objective of the well studied problem of source separation. This observation is beneficial in two ways. First, we can leverage a rich reservoir of numerical algorithms, most importantly relating to independent component analysis [8], [9], in order to approximate the optimal transform. Moreover, our necessary and sufficient condition leads to contributions in source separation.

The main objective of source separation is exactly that of finding an orthogonal matrix that will generate coefficients "as independent as possible". Such matrices can be found by maximizing an ad hoc cost function ([10], [11], [12], [13]), called contrast function, that purports to quantify how close to statistically independent the resulting components are. One can choose one of many ways to define the contrast function, and this choice governs the form of the algorithms. The two broadest definitions of independence are based on minimization of mutual information or maximization of "non-Gaussianity". The latter is motivated by the central limit theorem, uses kurtosis and negentropy. The former family of algorithms is obviously closely related measures involving the Kullback-Leibler (KL) divergence.

Our main result yields the precise connection between the problem of finding the optimal transform in high resolution variable rate coding and the source separation problem, when the objective (contrast) function is effectively the divergence. The optimal transform for the former (source coding) problem is shown to minimize the objective of the latter problem. This suggests that advances in transform coding may have an impact directly in source separation. An example of such a result is presented in Section IV, where our necessary and sufficient condition for optimality maps the result of [6] to ensure the optimality of KLT for source separation of Gaussian vector mixtures.

The paper is organized as follows: we present the problem formulation and the review of prior work in Section II. The main result is presented in Section III and further results related to source separation are presented in Section IV. A numerical algorithm to optimize the transform in transform coding is presented with preliminary results in Section V, and conclusions are presented in Section VI.

II. Review of Prior Results

A. Preliminaries and Notation

The entropy of a discrete random variable X taking values in $\mathcal{X}$ is

$$H(X) = -\sum_{x\in\mathcal{X}} P(X=x)\log P(X=x) \tag{1}$$

where logarithm is base 2. The differential entropy of a continuous random variable X with probability density function $f_X(x)$ is

$$h(X) = -\int f_X(x)\log f_x(x)dx \tag{2}$$

The divergence between two continuous distributions f_X and g_X, is given by

$$\mathcal{D}(f_X||g_X) = \int f_X(x)\log\frac{f_X(x)}{g_X(x)}dx \tag{3}$$

The divergence between two continuous distributions f_X and g_X, is given by

$$\mathcal{D}(f_X||g_X) = \int f_X(x)\log\frac{f_X(x)}{g_X(x)}dx \tag{4}$$

Let source $\mathbf{X}$ be an N dimensional random vector, with real components, $X_1, X_2..., X_N$. Without loss of generality, we assume $\mathbb{E}(\mathbf{X}) = \mathbf{0}$, and hence $\mathbf{R_X} = \mathbf{E}(\mathbf{XX^T})$. Let the transform $\mathbf{U}$ be a real $N\times N$ orthogonal matrix ($\mathbf{U}^{-1} = \mathbf{U}^T$) and let

$$\mathbf{Y} = \mathbf{UX} \tag{5}$$

be the transformed random vector with coefficients $Y_1, Y_2, ..Y_N$. A scalar quantizer Q is a mapping $Q:\mathbb{R}\rightarrow\mathbb{R}$. We restrict this paper to variable rate analysis, and the rate needed to describe source X after quantization by quantizer Q is

$$R(Q) = H[Q(X)] \tag{6}$$

A transform coding scheme is a structured vector quantizer where the random vector $\mathbf{X}$ is transformed into $\mathbf{Y}$ by $\mathbf{Y} = \mathbf{UX}$ and then each component Y_i is quantized with scalar quantizers Q_i. The total rate of the transform coder is

$$R_T = \sum_i H(Q_i(Y_i)) \tag{7}$$

At the decoder, inverse transformation by the matrix $\mathbf{U}^{-1} = \mathbf{U}^T$ is used to obtain an estimate of the source vector. The corresponding distortion is measured as mean square error,

$$D_T = \mathbb{E}\{||\mathbf{X} - \mathbf{U}^T\mathbf{Q}(\mathbf{UX}))||_2^2\} \tag{8}$$

where $\mathbf{Q}(\mathbf{X}) = [Q_1(X_1), .., Q_N(X_N)]^T$.

B. High rate approximations

The quantization operation is nonlinear and difficult to analyze mathematically. However, for both fixed and variable rate quantization, high resolution approximations can be made. Specifically, if the density of a scalar random variable is reasonably smooth, then at sufficiently high rate the distribution within a quantization interval is uniform. It is well known that uniform quantizers are asymptotically (at high resolution) optimal for variable rate coding, irrespective of the density of the source to be quantized [14]. Therefore, we use uniform quantizers throughout the paper. Let Δ_i be step size for i^{th}transform coefficient. This assumption results in quantization noise that is uniformly distributed over $(-\Delta_i, \Delta_i)$. Thus, at high resolution the distortion D_i is approximated as:

$$D_i = \frac{\Delta_i^2}{12} \tag{9}$$

The following straightforward auxiliary lemma relates the differential entropy of a continuous random variable with the entropy of its reproduction after uniform quantization at high resolution:

Lemma 1 (e.g., [15])**.** *If density $f_X(x)$ of random variable X is Riemann integrable, and $Q(X)$ is its reproduction after uniform quantization with step size Δ, then the following holds asymptotically, as $\Delta \to 0$:*

$$H(Q(X))) + \log \Delta \to h(X) \tag{10}$$

This lemma will be used in the proof of Theorem 1.

C. On Optimality of KLT

Definition (KLT): An orthogonal $N \times N$ matrix $\mathbf{K}$ is a KLT of N dimensional source vector $\mathbf{X}$ with covariance matrix $\mathbf{R}_X$ if $\mathbf{K}\mathbf{R}_X\mathbf{K}^T = \mathbf{\Lambda}_X$, where $\mathbf{\Lambda}_X$ is diagonal.

In other words, KLT generates uncorrelated coefficients. It is well known that KLT is optimal for "zonal sampling" or "truncated expansion": if the source estimate is approximated by expansion from a pre-determined subset of the transform coefficients, then KLT minimizes the approximation error. Another optimality aspect of KLT is shown in [16] for Gaussian sources: KLT minimizes the expected number of expansion terms (or transform coefficients) if the reconstruction error is required to be below a prescribed threshold. It has more recently been shown that KLT is optimal for Gaussian sources for both variable and fixed rate and at any operating rate regime, i.e., without any high resolution approximations [4]. Note that KLT is not necessarily unique. As example, when $\mathbf{R}_X = \mathbf{I}$, any orthogonal transform $\mathbf{U}$ "diagonalizes" $\mathbf{R}_X$ as $\mathbf{U}\mathbf{I}\mathbf{U}^T = \mathbf{I}$. Then a natural question arises: do all these KLTs perform equally? A sufficient condition for optimality of a KLT (that resolves this question if satisfied by one of the contenders) was given in [5] and is reproduced here.

Effros-Feng-Zeger Theorem (EFZ) [5]: If a KLT produces independent transform coefficients, then it is optimal for variable-rate transform coding at high resolution.

Note that this sufficient condition for optimality is not necessary. Specifically, there is a family of distributions where KLT has been shown to be optimal for transform coding although it does not generate independent coefficients [6].

Definition (Gaussian Vector Scale Mixtures): A random vector $\mathbf{X}$ taking values in $\mathbb{R}^N$ is called Gaussian Vector Scale Mixture (GVSM) if $\mathbf{X} = \mathbf{C}^T(\mathbf{Z} \odot \mathbf{V})$ where $\mathbf{C}$ is a constant orthogonal matrix, random vector $\mathbf{Z} \sim \mathbb{N}(0, \mathbf{I})$, scale vector $\mathbf{V}$ is a random vector independent of $\mathbf{Z}$ and taking values in $\mathbb{R}^+$, and $\odot$ denotes the element-wise product.

Note that conditioned on $\mathbf{V} = \mathbf{v}$, the GVSM vector $\mathbf{X}$ is Gaussian. Note further that this definition characterizes a a fairly broad set of distributions, including Gaussian mixtures.

Jana-Moulin (JM) Theorem [6]: KLT is optimal for a GVSM source for variable rate coding at high resolution.

This theorem clearly identifies a set of source distributions for which KLT is optimal, but leaves open the question of whether KLT is strictly suboptimal outside this set.

In summary, several natural follow-up questions remain open: when is KLT optimal for transform coding of general non-Gaussian sources? What is a conclusive condition for optimality of a general (not necessarily KLT) transform? If KLT is suboptimal, how can we numerically find the optimal transform? In this paper, we present a necessary and sufficient condition for optimality of any transform, naturally including KLT. Also, when KLT is suboptimal, we propose an algorithm to find the optimal transform.

III. Main Result

The main result is stated in the following theorem.

Theorem 1. *Orthogonal transform $\mathbf{U}^*$ is optimal if and only if the following is satisfied:*

$$\mathbf{U}^* = \underset{\mathbf{U}}{\operatorname{argmin}}\, \mathcal{D}(f_Y(\mathbf{y}) || \prod_{i=1}^{N} f_{y_i}(y_i)) \tag{11}$$

where $\mathcal{D}$ is divergence.

Note: Theorem 1 subsumes Effros-Feng-Zeger theorem [5] as an extreme special case where KLT yields independent coefficients.

The proof will make use of a trivial auxiliary lemma, which we state without proof:

Lemma 2. *The joint entropy is invariant to orthogonal transformation: Let $\mathbf{X}$ be a random vector and $\mathbf{U}$ be an orthogonal matrix, then*

$$h(\mathbf{U}\mathbf{X}) = h(\mathbf{X}) \tag{12}$$

Proof of Theorem 1: Using high resolution approximation for variable rate quantization, we get the following for total distortion,

$$D_T = \sum_i \frac{\Delta_i^2}{12}, \tag{13}$$

and for total rate,

$$R_T = \sum_i H(Q(y_i)) \tag{14}$$

Since the distortion is independent of the distribution of the transform coefficients, the aim of the transform coder is to minimize the total rate R_T. Using Lemma 1, we can rewrite (13) as,

$$R_T = -\sum_i \int f_{y_i}(y_i) \log f_{y_i}(y_i) dy_i + \log \Delta_i \tag{15}$$

where f_{y_i} is the marginal density of the i^{th} transform coefficient. Since the quantization intervals are fixed, the optimal transform must minimize the first term, hence the cost function:

$$\begin{aligned} J &= -\sum_i \int f_{y_i}(y_i) \log(f_{y_i}(y_i)) dy_i \\ &= -\int f_Y(\mathbf{y}) \left[\sum_i \log(f_{y_i}(y_i)) \right] d\mathbf{y} \end{aligned} \tag{16}$$

Using Lemma 2, we write the differential entropy $h(\mathbf{y})$ as

$$-\int f_Y(\mathbf{y})\log f_Y(\mathbf{y})d\mathbf{y} = -\int f_X(\mathbf{x})\log f_X(\mathbf{x})d\mathbf{x} = C \quad (17)$$

where C is used to emphasize that the joint entropy is determined by the source distribution and is hence constant with respect to the transform. Subtracting the constant C from both sides of (16), and noting that minimizing J is equivalent to minimizing $J - C =$

$$-\int f_Y(\mathbf{y})\left[\sum_i \log(f_{y_i}(y_i))\right]d\mathbf{y} + \int f_Y(\mathbf{y})\log f_Y(\mathbf{y})d\mathbf{y}$$

$$= \mathcal{D}(f_Y(\mathbf{y})||\prod_{i=1}^{N} f_{y_i}(y_i)) \quad (18)$$

which completes the proof. ■

Note that Theorem 1 essentially states that the optimal transform is the one that minimizes the statistical dependence of the transform coefficients. KLT considers second order statistics and decorrelates the transform coefficients, but this is neither necessary nor sufficient to minimize the overall statistical dependence as measured by the above divergence. The theorem also suggests that the optimal transform deviates from KLT whenever second order statistics are not a good representative of the overall dependence. This result also subsumes as a direct corollary the EFZ Theorem [5], and the well known optimality of KLT for jointly Guassian sources at high resolution variable rate coding [2].

IV. Source Separation Problem

Hyvarinen and Oja [8] give the following definition for the noise free linear source separation problem, which is of interest here.

Definition (Source Separation Problem): Let random vector $\mathbf{X}$ of size N be obtained by

$$\mathbf{X} = \mathbf{BS} \quad (19)$$

where $\mathbf{B}$ is a constant $N \times N$ "mixing" matrix, elements S_i in the vector $\mathbf{S} = (S_1, ..., S_N)^T$ are assumed to be mutually independent. $\mathbf{X}$ is observed while both $\mathbf{B}$ and $\mathbf{S}$ are unknown. The aim of the problem is to find $\mathbf{S}$ (or alternatively the matrix $\mathbf{B}$), by maximizing some form of independence among the transform coefficients.

We choose the objective function of divergence between the product of the marginals of the transform coefficients and joint density of the transformed vector, i.e. the cost function

$$J(\mathbf{U}) = \mathcal{D}(f_Y(\mathbf{y})||\prod_{i=1}^{N} f_{y_i}(y_i)) \quad (20)$$

where $\mathbf{Y} = \mathbf{UX}$.

Our main result provides two prospective directions to pursue: i) It allows us to develop an algorithm for the long standing problem of optimal transform coding by leveraging a large bank of algorithms from the source separation literature, and ii) to apply the theoretical optimality (or suboptimality) results of transform coding to source separation problems. An algorithm for finding the optimal transform is presented in the next section. In the remainder of this section we use the JM Theorem to obtain a new optimality result in source separation.

Theorem 2. *The optimal orthogonal transform for source separation of a Gaussian vector scale mixture is KLT, when the contrast function is the divergence-based cost of (20).*

Proof: The proof follows from Theorem 1 and the JM theorem. ■

The theorem establishes the optimality of KLT and hence renders source separation algorithms for this family of sources unnecessary.

V. Algorithm

In this section, we propose a modified version of the algorithm by Pham [17], [18] which seeks to find the orthogonal transform that minimizes the contrast function expressed in (20). The minimization of the cost can be done through a gradient descent algorithm, where the update for transform matrix $\mathbf{U}$ involves a matrix ϵ yielding $\mathbf{U} + \epsilon\mathbf{U}$. We expand $\mathbf{U} + \epsilon\mathbf{U}$ with respect to ϵ up to second order terms and then minimize the resulting cost with respect to ϵ to obtain the optimal ϵ and hence a new estimate. The Taylor expansion of $J(\mathbf{U} + \epsilon\mathbf{U})$ can be expressed as follows:

$$J(\mathbf{U} + \epsilon\mathbf{U}) = J(\mathbf{U}) + \sum_{i,j}\epsilon_{ij}[\mathbb{E}(Y_j\Phi_i(Y_i)) - \mathbb{E}(Y_i\Phi_j(Y_j))]$$
$$+ \frac{1}{2}\sum_{i,j}\epsilon_{ij}^2[\mathbb{E}(\Phi_i^2(Y_i))\mathbb{E}(Y_j^2) - \mathbb{E}(\Phi_j^2(Y_j))\mathbb{E}(Y_i^2) - 2]$$
$$+ O(\epsilon^3) \quad (21)$$

where Φ is the gradient of the entropy function, also known as score function and $O(\epsilon^3)$ accounts for higher order terms which we will neglect. Setting the partial derivative with respect to ϵ to zero, we find ϵ as follows:

$$\epsilon_{ij} = \frac{\mathbb{E}(Y_j\Phi_i(Y_i)) - \mathbb{E}(Y_i\Phi_j(Y_j))}{\mathbb{E}(\Phi_i^2(Y_i))\mathbb{E}(Y_j^2) - \mathbb{E}(\Phi_j^2(Y_j))\mathbb{E}(Y_i^2) - 2} \quad (22)$$

In this expression, the probability density functions being unknown, the score function $\Phi(Y)$ is replaced by an estimate (see [18]) and the expectations are estimated from training samples assuming ergodicity. There is no guarantee that $\mathbf{U} + \epsilon\mathbf{U}$ will be orthogonal. To solve this problem, we replace the resulting matrix $\mathbf{U}$ with its closest (in terms of Frobenius norm) orthogonal approximation which can be obtained by polar decomposition[1].

We obtained some preliminary results using the proposed algorithm. We first generate the samples of $\mathbf{X}$ by $\mathbf{X} = \mathbf{BS}$ where $\mathbf{S}$ consists of four independent and identically distributed random variables, and $\mathbf{B}$ is a random orthogonal mixing matrix. The proposed algorithm finds the correct matrix $\mathbf{U} = \mathbf{B}^{-1}$ precisely. We note that an obvious KLT choice

[1]We employed a fast method as to repeatedly average $\mathbf{U}$ with its transpose inverse until convergence [19].

is the identity $\mathbf{I}$ since the source is already uncorrelated. It follows from the examples in [5], that the gain of the optimal transform over standard KLT (in this case the identity matrix, $\mathbf{I}$) can be unbounded.

VI. Conclusion

In this paper, we presented a necessary and sufficient condition for transform optimality at high resolution, variable rate coding. Note that this result not only resolves the question of when KLT is optimal (at high resolution), but also determines the optimal transform when it is not KLT. This condition also points to direct connections between the transform coding problem and an important subset of the well studied source separation problems. We used this observation to obtain new results in two directions: developing a numerical algorithm for transform optimization in transform coding by leveraging tools from source separation; and mapping known theoretical optimality results in transform coding to the source separation problem. Preliminary results for transform optimization show the algorithm converging to the optimal transform, although global optimality is not guaranteed in general. In source separation the analogy enables the identification of a fairly broad family of distributions for which the optimality of KLT is guaranteed and numerical optimization algorithms are not needed. The basic ideas in this paper can be (nontrivially) extended to fixed rate coding, to non-orthoognal transforms, to distributed [20] and to multiple descriptions coding [21] scenarios, all of which are the subjects of ongoing investigation.

References

[1] V.K. Goyal, "Theoretical foundations of transform coding," *Signal Processing Magazine, IEEE*, vol. 18, no. 5, pp. 9–21, 2002.

[2] A. Gersho and R.M. Gray, *Vector Quantization and Signal Compression*, Springer, 1992.

[3] J. Huang and P. Schultheiss, "Block quantization of correlated Gaussian random variables," *IEEE Transactions on Communications*, vol. 11, no. 3, pp. 289–296, 1963.

[4] V. Goyal, J. Zhuang, and M. Vetterli, "Transform coding with backward adaptive updates," *IEEE Transactions on Information Theory*, vol. 46, no. 4, pp. 1623–1633, 2000.

[5] M. Effros, H. Feng, and K. Zeger, "Suboptimality of the Karhunen-Loeve transform for transform coding," *IEEE Transactions on Information Theory,*, vol. 50, no. 8, pp. 1605–1619, 2004.

[6] S. Jana and P. Moulin, "Optimality of KLT for High-Rate Transform Coding of Gaussian Vector-Scale Mixtures: Application to Reconstruction, Estimation, and Classification," *IEEE Transactions on Information Theory*, vol. 52, no. 9, pp. 4049–4067, 2006.

[7] C. Archer and TK Leen, "A generalized Lloyd-type algorithm for adaptive transform coder design," *IEEE Transactions on Signal Processing,*, vol. 52, no. 1, pp. 255–264, 2004.

[8] A. Hyvarinen and E. Oja, *Independent component analysis: algorithms and applications*, vol. 13, Elsevier, 2000.

[9] P. Comon, "Independent component analysis, a new concept?," *Signal processing*, vol. 36, no. 3, pp. 287–314, 1994.

[10] A. Hyvarinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Transactions on Neural Networks,*, vol. 10, no. 3, pp. 626–634, 2002.

[11] F.R. Bach and M.I. Jordan, "Kernel independent component analysis," *The Journal of Machine Learning Research*, vol. 3, pp. 1–48, 2003.

[12] E.G. Learned-Miller and W.F. John III, "ICA using spacings estimates of entropy," *The Journal of Machine Learning Research*, vol. 4, pp. 1271–1295, 2003.

[13] T.W. Lee, M. Girolami, and T.J. Sejnowski, "Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources," *Neural Computation*, vol. 11, no. 2, pp. 417–441, 1999.

[14] H. Gish and J. Pierce, "Asymptotically efficient quantizing," *IEEE Transactions on Information Theory,*, vol. 14, no. 5, pp. 676–683, 1968.

[15] T.M. Cover and J.A. Thomas, *Elements of Information Theory*, J.Wiley New York, 1991.

[16] V. Algazi and D. Sakrison, "On the optimality of the Karhunen-Loève expansion," *IEEE Transactions on Information Theory,*, vol. 15, no. 2, pp. 319–321, 1969.

[17] D.T. Pham, "Mutual information approach to blind separation of stationary sources," *IEEE Transactions on Information Theory*, vol. 48, no. 7, pp. 1935–1946, 2002.

[18] D.T. Pham, "Fast algorithms for mutual information based independent component analysis," *IEEE Transactions on Signal Processing,*, vol. 52, no. 10, pp. 2690–2700, 2004.

[19] G.H. Golub and C.F. Van Loan, *Matrix computations*, Johns Hopkins Univ Pr, 1996.

[20] M. Gastpar, P.L. Dragotti, and M. Vetterli, "The distributed Karhunen-Loève transform," *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5177–5196, 2006.

[21] VK Goyal and J. Kovacevic, "Generalized multiple description coding with correlating transforms," *IEEE Transactions on Information Theory,*, vol. 47, no. 6, pp. 2199–2224, 2001.