Towards Optimality in Multiterminal Transform Coding *

Emrah Akyol and Kenneth Rose

University of California, Santa Barbara, CA

Email:{eakyol, rose}@ece.ucsb.edu

Abstract

This paper is concerned with transform coding of correlated sources in conjunction with variable rate quantization at high resolution. The approach builds on our prior work on optimality conditions for transform coding in the point-to-point setting. The first contribution involves transform coding with decoder side information. In this setting, side information is only available to the decoder, whereas the encoder knows the joint statistics. The necessary and sufficient condition for optimality of a unitary transform in the side information setting is derived, namely, such transform minimizes a conditional divergence-based measure of inter-dependence of the transform coefficients, given the side information. This optimality result subsumes prior, known results that were restricted to the Gaussian case, where the conditional Karhunen-Loeve transform is optimal. The second contribution involves distributed transform coding, where two correlated sources are to be transform coded separately, but decoded jointly. The necessary and sufficient condition for optimality of unitary transforms in the distributed coding setting is derived. It is then specialized to produce closed form optimal transforms for specific source densities, including the case of jointly Gaussian sources.

1   Introduction

Transform coding is a computationally attractive approach to source coding, and is widely used in audio, image and video compression. The transform coding problem has been studied extensively. In their seminal paper, Huang and Schulthesis have shown [1] that if the vector source is Gaussian and the bit budget is asymptotically large, then the Karhunen-Loève transform (KLT) and its inverse are an optimal pair of transforms for fixed-rate coding. A more recent paper by Goyal, Zhuang and Vetterli strengthened that result by showing the optimality of KLT for Gaussian sources without recourse to high resolution assumptions [2].

In [3], the "popular faith" in the optimality of KLT was challenged and it was demonstrated by examples that KLT can be suboptimal for both fixed and variable rate quantization, at asymptotically high rate (with high resolution approximations).

Moreover, a sufficient condition for optimality of KLT was derived: if KLT generates independent coefficients then it is the optimal transform for variable rate coding. This condition is obviously satisfied in the Gaussian case. In [4], a significant extension of the known range of optimality of KLT was obtained: KLT, in conjunction with variable rate high resolution coding, is not only optimal for Gaussians but also for the broader family of Gaussian vector mixtures.

In our recent work, [5], we derived a necessary and sufficient condition for optimality of a transform in conjunction with variable rate coding at high resolution. Specifically, we showed that the optimal transform is the one that minimizes the divergence between the joint distribution of the resulting transform 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 resolution), but it also determines the optimal transform when it is not KLT. Since minimizing a measure of dependence is closely related, at the high level, to the objective of a variant of the well studied problem of source separation, we leveraged the rich reservoir of numerical algorithms, most importantly relating to independent component analysis [6], in order to derive an algorithm to optimize the transform.

In the present work, we study the transform coding problem in the broader scope of network settings. First, transform coding with side information is analyzed. Most of the optimality results involving transformation in distributed settings were restricted to Gaussian sources. Linear approximation and rate-distortion analysis (asymptotic delay) for Gaussian sources were studied in [7] which led to the optimality of the *conditional KLT*. Transform coding of Gaussian sources with high resolution variable rate scalar quantizers was studied in [8], where it is shown that conditional KLT is optimal for jointly Gaussian source and side information. The objective of this paper is to derive general optimality results without recourse to Gaussian assumptions. We present the necessary and sufficient condition for optimality of a unitary transform in the side information setting, as well as specializing the result to demonstrate the optimality of conditional KLT for a family of non-Gaussian sources. Finally, we consider the distributed transform coding setting, where two correlated sources are to be encoded separately, via transform coding. This setting has also been studied in [7], where optimal transforms for linear approximation and asymptotical compression (Shannon sense) were derived for Gaussian sources.

The paper is organized as follows: in Section 2, we present preliminaries, including review of prior work on the point-to-point setting. Transform coding in the decoder side information setting and in the distributed settings, is studied in Sections 3 and 4, respectively. Conclusions are presented in Section 5.

2 Preliminaries

2.1 Notation

In general, lowercase letters (e.g., c) denote scalars, boldface lowercase (e.g., $\boldsymbol{x}$) vectors, upper-case (e.g., C, X) matrices and random variables, and boldface uppercase (e.g., $\boldsymbol{X}$) random vectors. Throughout this paper, vectors are N-tuples, and matrices have size $N \times N$. Let I denote the identity matrix, and let $\mathcal{U}$ denote the set of real unitary matrices. Without loss of generality we will assume zero-mean sources. Hence, $R_X = \mathbb{E}(\boldsymbol{X}\boldsymbol{X}^T)$ denotes both the autocorrelation matrix and covariance ma-

trix of $\boldsymbol{X}$, and $R_{XZ} = \mathbb{E}(\boldsymbol{X}\boldsymbol{Z}^T)$ denotes the matrix of its cross-covariance with $\boldsymbol{Z}$. We follow standard notation for information-theoretic quantities (see e.g., [9]). A scalar quantizer Q is a mapping $Q : \mathbb{R} \rightarrow \mathcal{C} \subset \mathbb{R}$, where $\mathcal{C}$ is a countable set called the codebook. We restrict this paper to variable rate analysis, and the rate needed to describe random variable Y after quantization by quantizer Q is determined by the entropy, $H[Q(Y)]$. A transform coding scheme is a structured vector quantizer where the random vector $\boldsymbol{X}$ is transformed into $\boldsymbol{Y} = U\boldsymbol{X}$ by a unitary matrix $U \in \mathcal{U}$, and then each component Y_i is quantized by a scalar quantizer Q_i. The total rate of the transform coder is

$$R_T = \sum_i H(Q_i(Y_i)) \tag{1}$$

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

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

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

2.2 High rate approximation

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 [10]. Therefore, we use uniform quantizers throughout the paper. Let Δ_i be step size for the ith transform coefficient. Under the mild assumption that the density of a scalar random variable is Riemann integrable, at sufficiently high rate the distribution within a quantization interval is uniform [10]. This assumption results in quantization noise that is uniformly distributed over $(-\Delta_i/2, \Delta_i/2)$. Thus, at high resolution the distortion D_i is approximated as $D_i = \frac{\Delta_i^2}{12}$. We present the following lemma without proof.

Lemma 1 ([4, 11]). *At high resolution, subject to the distortion constraint, D, the average bit rate $R(D)$ is minimized if the quantizer step sizes are equal and given by*

$$\Delta_i^* = \sqrt{\frac{12D}{N}}, 1 \leq i \leq N, \tag{3}$$

regardless of the unitary transform U and the density of source $\boldsymbol{X}$.

Note that, unlike the optimal step size Δ, which is invariant to the source statistics, the optimal transform U depends on the source. The following straightforward auxiliary lemma (a consequence of high resolution quantization) relates the differential entropy of a continuous random variable with the entropy of its reproduction after uniform quantization at high resolution:

Lemma 2 (e.g., [9]). *If density $f_X(\cdot)$ 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 $\Delta \rightarrow 0$:*

$$H(Q(X)) + \log \Delta \rightarrow h(X) \tag{4}$$

2.3 On the Question of KLT Optimality

KLT : The matrix R_X is real, symmetric, and positive semidefinite, with eigenvalues $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_N \geq 0$. It allows thus a diagonalization given by

$$R_X = S_X^T \Lambda_X S_X \tag{5}$$

where $S_X \in \mathbb{R}^{N \times N}$ is a unitary matrix whose columns are the eigenvectors of the matrix R_X ordered by decreasing eigenvalues , and $\Lambda_X \in \mathbb{R}^{N \times N}$ is diagonal with entries $\lambda_1, \lambda_2, ..., \lambda_N$. The matrix S_X is called Karhunen Loeve transform (KLT) of $\boldsymbol{X}$.

Clearly, KLT generates uncorrelated coefficients. It is well known that KLT is optimal for "zonal sampling" or "linear approximation": if the source estimate is approximated by expansion from a pre-determined subset of the transform coefficients, then the KLT (if unique) minimizes the approximation error. 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 high resolution approximations [2].

Note that KLT is not necessarily unique. As example, when $R_X = I$, any unitary transform U "diagonalizes" R_X as $UIU^T = I$. 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 [3] and is reproduced here.

Theorem due to Effros-Feng-Zeger (EFZ) [3]: If a KLT produces independent transform coefficients, then it is optimal for variable-rate transform coding at high resolution.

Note that the sufficient condition of EFZ is not necessary for optimality. 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 [4].

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

Theorem due to Jana-Moulin (JM) [4]: 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.

2.4 Optimality Condition for the Single Terminal Setting

The state of the art, summarized so far, leaves several natural follow-up questions 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 section, we review our recent result that provided, for the point-to-point setting, a necessary and sufficient condition for optimality of a transform, regardless of whether or not it is KLT. In other words, this condition naturally determines if KLT is optimal, and when KLT is suboptimal, it provides the means to identify the non-KLT optimal transform. In fact, this condition leads to an algorithm to optimize the transform.

Single Terminal Transform Optimality Theorem [5]: Unitary transform U^* is optimal if and only if the following is satisfied:

$$U^* = \operatorname*{argmin}_{U \in \mathcal{U}} \mathcal{D}(f_Y || \prod_{i=1}^{N} f_{Y_i}) \tag{6}$$

where $\boldsymbol{Y} = U\boldsymbol{X}$ and $\mathcal{D}$ is the divergence.

Remark 1. *This result subsumes the EFZ theorem [3], as well as the known optimality of KLT for jointly Gaussian sources at high resolution variable rate coding [10], as represented by the extreme special case where KLT yields independent coefficients.*

Remark 2. *This theorem essentially states that the optimal transform is the one that minimizes a measure of 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 do not capture well the overall dependence.*

3 Transform Coding with Side Information

We consider the setting where a random variable $\boldsymbol{Z}$, which is correlated with source $\boldsymbol{X}$, is only available to the decoder, as shown in Figure 1. The encoder does not have access to the realization of the side information $\boldsymbol{Z}$, but knows the joint statistics of $\boldsymbol{X}$ and $\boldsymbol{Z}$. The decoder first decodes the entropy coded transform coefficients $\hat{\boldsymbol{Y}}$ using the side information $\boldsymbol{Z}$, then applies the inverse transform $U^{-1} = U^T$ to obtain the reconstruction $\hat{\boldsymbol{X}} = U^T\hat{\boldsymbol{Y}}$ that minimizes $\mathbb{E}\{||\boldsymbol{X} - \hat{\boldsymbol{X}}||^2\}$.

This problem has been studied extensively from an asymptotic (in the Shannon sense) point of view [12, 13]. Specifically of interest here is the seminal result of Slepian-Wolf that source X can be losslessly reconstructed at the decoder at rate $H(X|Z)$ when the side information Z is only available to the decoder [12]. Recently, researchers considered the distributed transform coding problem [7, 8]. Specifically, [7] studied optimal transforms for linear approximation and asymptotic (in delay) compression. Conditional KLT was introduced and shown to be the optimal transform in the case of jointly Gaussian sources. A high resolution analysis of transform coding with scalar quantizers in conjunction with decoder side information was provided in [8]. It was shown that for Gaussian sources conditional KLT and uniform quantizers are optimal for variable rate coding at high resolution. In this section we derive general optimality conditions without restriction to Gaussian sources. Let us start with relevant definitions.

Conditional KLT[1]: A conditional KLT of $\boldsymbol{X} \in \mathbb{R}^N$ with respect to $\boldsymbol{Z} \in \mathbb{R}^K$ is the linear transform characterized by the matrix $S_{X|Z}$ that satisfies

$$S_{X|Z}^T R_E S_{X|Z} = \operatorname{diag}(\lambda_{E,1}, \lambda_{E,2}, ..., \lambda_{E,N}) \tag{7}$$

---

[1]This definition differs from the one appearing in [7] in that it makes explicit the relation to prediction error covariance.
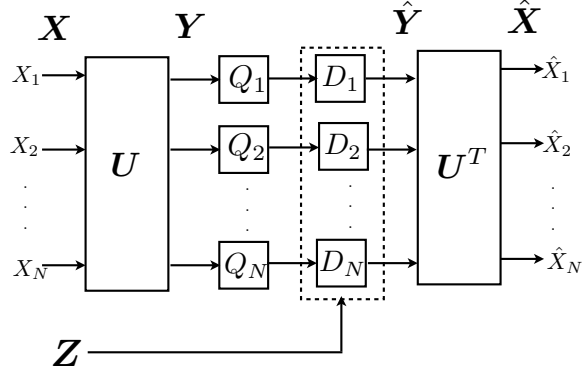
Figure 1: Transform coding with decoder side information

where R_E is the covariance of the prediction error of the best linear predictor of $\boldsymbol{X}$ from $\boldsymbol{Z}$

$$R_E = R_X - R_{XZ} R_Z^{-1} R_{XZ}^T \tag{8}$$

We summarize a few straightforward observations and properties of conditional KLT (some directly inherited from the standard KLT):

Remark 3. *Conditional KLT is a unitary transform and it is not unique.*

Remark 4. *Components of the vector $\boldsymbol{Y} = S_{X|Z} \boldsymbol{X}$, i.e., the transform coefficients are conditionally uncorrelated given $\boldsymbol{Z}$. For jointly Gaussian source $\boldsymbol{X}$ and $\boldsymbol{Z}$, the transform coefficients are conditionally independent given $\boldsymbol{Z}$, i.e., $Y_i \leftrightarrow \boldsymbol{Z} \leftrightarrow Y_j$ form a Markov chain for $i \neq j$.*

The optimality of uniform quantization followed by Slepian-Wolf encoding was shown in [8]. The following lemma states the optimality of identical uniform quantizers when transform coding is performed within the decoder side information setting.

Lemma 3. *Subject to the distortion constraint, D, the average bit rate is minimized if the quantizer step sizes are equal and given by*

$$\Delta_i = \Delta = \sqrt{\frac{12D}{N}}, \forall i \tag{9}$$

regardless of the unitary transform U, the joint density of source $\boldsymbol{X}$ and side information $\boldsymbol{Z}$.

Proof. The proof follows the footsteps of the proof of Lemma 1 (optimality of uniform quantization for transform coding). □

Our main result for this setting, given in the following theorem, states the necessary and sufficient condition for optimality of a unitary transform.

Theorem 1. *Unitary transform U^* is optimal, for source $\boldsymbol{X}$ and side information $\boldsymbol{Z}$, if and only if the following is satisfied:*

$$U^* = \operatorname*{argmin}_{U \in \mathcal{U}} \mathcal{D}(f_{Y|Z} || \prod_{i=1}^{N} f_{Y_i|Z}) \tag{10}$$

where $\boldsymbol{Y} = U\boldsymbol{X}$ and $\mathcal{D}$ is the conditional divergence.

Proof. Using Lemma 3 we have the following for total distortion, and for total rate,

$$D_T = \frac{N\Delta^2}{12}, \ R_T = \sum_i H(Q(Y_i)|\boldsymbol{Z}) \tag{11}$$

assuming an ideal entropy coder that achieves the conditional entropy due to the asymptotic optimality of Slepian-Wolf encoding [12]. 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 2, we can rewrite R_T as

$$R_T = -\sum_i \int f_{Y_i,Z}(y_i, \boldsymbol{z}) \log f_{Y_i|Z}(y_i, \boldsymbol{z}) dy_i d\boldsymbol{z} + \log \Delta \tag{12}$$

where $f_{y_i,Z}$ is the joint density of the ith transform coefficient and the side information $\boldsymbol{Z}$. Since the quantization intervals are fixed, the optimal transform must minimize the first term, hence the cost function:

$$J = -\sum_i \int f_{Y_i,Z}(y_i, \boldsymbol{z}) \log f_{Y_i|Z}(y_i, \boldsymbol{z}) dy_i d\boldsymbol{z}$$

$$= -\int f_{YZ}(\boldsymbol{y}, \boldsymbol{z}) \left[\sum_i \log f_{Y_i|Z}(y_i, \boldsymbol{z}) \right] d\boldsymbol{y} d\boldsymbol{z} \tag{13}$$

Noting that differential entropy is invariant to unitary transformations, $h(U\boldsymbol{X}|\boldsymbol{Z}) = h(\boldsymbol{X}|\boldsymbol{Z})$, we write the conditional entropy $h(\boldsymbol{Y}|\boldsymbol{Z})$ as

$$h(\boldsymbol{Y}|\boldsymbol{Z}) = h(\boldsymbol{X}|\boldsymbol{Z}) = -\int f_{XZ}(\boldsymbol{x}, \boldsymbol{z}) \log f_{X|Z}(\boldsymbol{x}, \boldsymbol{z}) d\boldsymbol{x} d\boldsymbol{z} = C \tag{14}$$

where C is used to emphasize that the joint entropy is determined by the source and the side information and is hence constant with respect to the transform. Subtracting the constant C from both sides of (13), and noting that minimizing J is equivalent to minimizing $J - C$, we obtain

$$J - C = -\int f_{YZ}(\boldsymbol{y}, \boldsymbol{z}) \left[\sum_i \log f_{Y_i|Z}(y_i, \boldsymbol{z}) \right] d\boldsymbol{y} d\boldsymbol{z} + \int f_{YZ}(\boldsymbol{y}, \boldsymbol{z}) \log f_{Y|Z}(\boldsymbol{y}, \boldsymbol{z}) d\boldsymbol{y} d\boldsymbol{z}$$

$$= \mathcal{D}(f_{Y|Z} || \prod_{i=1}^N f_{Y_i|Z})$$

$\square$

Corollary 1. *Conditional KLT is optimal for transform coding of jointly Gaussian source and side information as well as a source distributed as jointly Gaussian conditioned on the (possibly non-Gaussian) side information.*

Proof. Remark 4 states that, conditional KLT renders the transform coefficients conditionally uncorrelated given the side information which in turn implies conditional independence for a conditionally Gaussian source. This guarantees that the conditional divergence of the optimality condition is zero. $\square$

Corollary 2. *Conditional KLT is optimal for sources that are conditionally GSVM given the (possibly non-Gaussian) side information.*

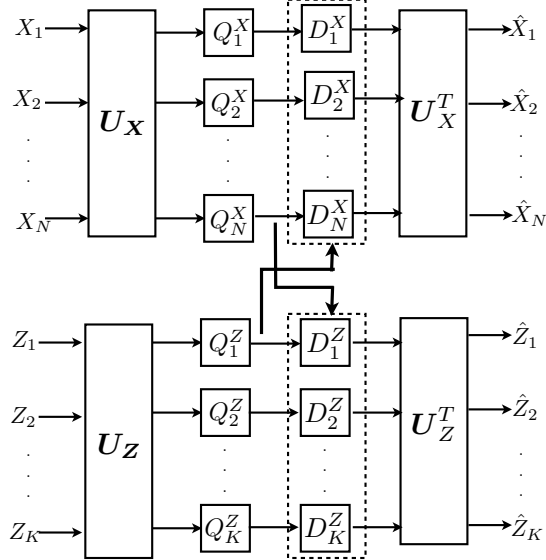The proof can be obtained by using the proof of JM Theorem and Theorem 1, and is omitted here for brevity.

Figure 2: Distributed transform coding problem

4   Distributed Transform Coding

We consider the setting where two correlated sources are separately encoded via transform coding as shown in Figure 2. Vector sources $\boldsymbol{X} \in \mathbb{R}^N$ and $\boldsymbol{Z} \in \mathbb{R}^K$ are transformed by unitary transforms $U_X, U_Z \in \mathcal{U}$ to produce $\boldsymbol{Y} = U_X \boldsymbol{X}$ and $\boldsymbol{T} = U_Z \boldsymbol{Z}$. Transform coefficients are then quantized with high resolution scalar quantizers $(Q_1^X, ..., Q_N^X$ and $Q_1^Z, ..., Q_K^Z)$ followed by entropy coding. The decoder first decodes the entropy coded quantized coefficients, $\hat{\boldsymbol{Y}}$ and $\hat{\boldsymbol{T}}$ and then generates $\hat{\boldsymbol{X}} = U_X \boldsymbol{Y}$ and $\hat{\boldsymbol{Z}} = U_Z \boldsymbol{T}$ that minimize the weighted error distortion metric:

$$\alpha \frac{1}{N} \mathbb{E}||\boldsymbol{X} - \hat{\boldsymbol{X}}||^2 + (1 - \alpha)\frac{1}{K}\mathbb{E}||\boldsymbol{Z} - \hat{\boldsymbol{Z}}||^2 \qquad (15)$$

where $\alpha \in [0, 1]$. The following lemma states the intuitive fact that conditioning entropy on the high resolution reconstruction of a random variable is equivalent to conditioning it on the original source.

Lemma 4. *Let continuous random variables X, Y have a Riemann integrable joint density. Let $Q(X, D)$ denote the quantized version of X with the optimal variable rate quantizer whose MSE distortion is D. Then,*

$$\lim_{D \to 0} h(Y|Q(X, D)) = h(Y|X) \qquad (16)$$

The proof can be obtained by considering the conditional entropy given discretized X as a Riemann sum where sampling of the real line is governed by the high resolution quantizer.

Theorem 2. *The optimal quantizers for distributed transform coding are uniform with step sizes Δ_X and Δ_Z for all transform coefficients of $\boldsymbol{X}$ and $\boldsymbol{Z}$. Also optimal step sizes do not depend on the transforms and are given by:*

$$\Delta_{X,i} = \Delta_X = \sqrt{\frac{6ND}{\alpha}}, \Delta_{Z,j} = \Delta_Z = \sqrt{\frac{6KD}{1 - \alpha}}, \forall i, j \qquad (17)$$

Proof. We consider optimality of one encoder while fixing the other. If we fix the second encoder, the decoder has access to $\widehat{\boldsymbol{T}}$. Using Lemma 4, we have

$$h(\boldsymbol{Y}|\widehat{\boldsymbol{T}}) \approx h(\boldsymbol{Y}|\boldsymbol{T}) = h(\boldsymbol{Y}|\boldsymbol{Z}) \text{ as } D \to 0 \tag{18}$$

Hence, the required rate is the same as if $\boldsymbol{Z}$ were available to the decoder. Using this and Lemma 4, we deduce that the optimal quantizers are uniform. Let $J = D + \gamma R$ where

$$D = \frac{\alpha}{12N} \sum_{i=1}^{N} \Delta_{X,i}^2 + \frac{1-\alpha}{12K} \sum_{j=1}^{K} \Delta_{Z,j}^2 \tag{19}$$

$$R = \sum_{i=1}^{N} h(\boldsymbol{Y}_i|\boldsymbol{Z}) - \log \Delta_{X,i} + \sum_{j=1}^{K} h(\boldsymbol{T}_j|\boldsymbol{X}) - \log \Delta_{Z,j} \tag{20}$$

Then, $\frac{\partial J}{\partial \Delta_{X,i}} = 0$ and $\frac{\partial J}{\partial \Delta_{Z,j}} = 0$ yield (17). Note that the cost function J is indeed convex with respect to $\Delta_{X,i}$ (and $\Delta_{Z,j}$), $\frac{\partial^2 J}{\partial \Delta_{X,i}^2} = \frac{2\alpha}{12N} + \frac{1}{2\Delta_{X,i}^2} \geq 0$. $\qquad \square$

The following theorem presents our main result within this setting:

Theorem 3. *Unitary transforms $\boldsymbol{U}_X^*$ and $\boldsymbol{U}_Z^*$ are optimal for distributed transform coding of sources $\boldsymbol{X}$ and $\boldsymbol{Z}$, if and only if the following is satisfied:*

$$U_X^* = \underset{U_X \in \mathcal{U}}{\operatorname{argmin}} \, \mathcal{D}(f_{Y|Z} || \prod_{i=1}^{N} f_{Y_i|Z}), \ \ U_Z^* = \underset{U_Z \in \mathcal{U}}{\operatorname{argmin}} \, \mathcal{D}(f_{T|X} || \prod_{i=1}^{N} f_{T_i|X}) \tag{21}$$

where $\boldsymbol{Y} = U_X \boldsymbol{X}$, $\boldsymbol{T} = U_Z \boldsymbol{Z}$ and $\mathcal{D}$ is the conditional divergence.

Proof. The proof uses ideas from the proof of Theorem 2. Note that distortion does not depend of the transforms U_X, U_Z and only the rate does. The total rate expression (20) has two terms which are identical to that of the side information setting. Hence, following the same steps used in proving Theorem 2, we have Theorem 3. $\qquad \square$

Corollary 3. *The optimal unitary transform for jointly Gaussian sources is given by the conditional KLTs: $U_X^* = S_{X|Z}$ and $U_Z^* = S_{Z|X}$.*

Remark 5. *The above results can be trivially extended to more than 2 sources or to a mixed setting where some of sources are directly available to the decoder and others are encoded through transform coding.*

5 Conclusion

In this paper, we studied the high resolution, variable rate transform coding problem for network settings. Two specific settings were considered: decoder side information and distributed transform coding. In the side information setting, known results were restricted to Gaussian sources. We derived optimality results for general, non-Gaussian sources and side information. Specifically, the optimal unitary transform minimizes the conditional divergence between the joint density of transform coefficients and the product of their marginals, given the side information. For the distributed transform coding problem, we derived the necessary and sufficient condition

for optimality of a unitary transform. We also specialized the results to prove the optimality of conditional KLT for various classes of sources. The scope of this paper is limited to unitary transforms, as is the case of most prior work in this area. Nevertheless the basic ideas in this paper can be extended (albeit nontrivially) to non-unitary transforms, and to other transform coding settings which are the subjects of ongoing investigation.

References

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

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

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

[4] 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 Trans. on Inf. Th.*, vol. 52, no. 9, 2006.

[5] E. Akyol and K. Rose, "A necessary and sufficient condition for transform optimality in source coding," in *Proc. of IEEE Int. Symp. on Inf. Th.*, 2011.

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

[7] M. Gastpar, P.L. Dragotti, and M. Vetterli, "The distributed Karhunen-Loève transform," *IEEE Trans. on Inf. Th.*, vol. 52, no. 12, pp. 5177–5196, 2006.

[8] D. Rebollo-Monedero, S. Rane, A. Aaron, and B. Girod, "High-rate quantization and transform coding with side information at the decoder," *Signal Processing*, vol. 86, no. 11, pp. 3160–3179, 2006.

[9] T. Cover and J. Thomas, *Elements of Information Theory*, J.Wiley NY, 1991.

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

[11] S. Mallat and F. Falzon, "Analysis of low bit rate image transform coding," *IEEE Transactions on Signal Processing*, vol. 46, no. 4, pp. 1027–1042, 1998.

[12] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Transactions on Information Theory*, vol. 19, no. 4, pp. 471–480, 1973.

[13] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. on Inf. Th.*, vol. 22, no. 1, 1976.