

Orthonormal Matrix Codebook Design for Adaptive Transform Coding

Rashmi Boragolla, and Pradeepa Yahampath

Department of Electrical and Computer Engineering, University of Manitoba
75 Chancellors Circle, Winnipeg MB R3T 2N2, Canada
borbnwmr@myumanitoba.ca, pradeepa.yahampath@umanitoba.ca

Abstract

A novel algorithm for designing optimized orthonormal transform-matrix codebooks for adaptive transform coding of a non-stationary vector process is proposed. This algorithm relies on a block-wise stationary model of a non-stationary process and finds a codebook of transform-matrices by minimizing the end-to-end mean square error of transform coding averaged over the distribution of stationary blocks of vectors. The algorithm, which belongs to the class of block-coordinate descent algorithms, solves an intermediate minimization problem involving matrix-orthonormality constraints in a computationally efficient manner by mapping the problem from the Euclidean space to the Stiefel manifold. As such, the algorithm can be broadly applied to any adaptive transform coding problem. Preliminary results obtained with inter-prediction residuals in an H265 video codec are presented to demonstrate the advantage of optimized adaptive transform codes over non-adaptive codes based on the standard DCT.

1 Introduction

It is well-known that the optimal transform for coding stationary Gaussian sources based on the mean squared error (MSE) is the Karhunen-Loève transform (KLT) [1, 2]. However, in real-world applications where the source data is highly non-stationary, the tendency has been to use generic fixed-transforms such as the discrete cosine transform (DCT). Nonetheless, adaptive transform coding (ATC) remains an active research area, and recently, some forms of ATC have also been incorporated into newer video coding standards such as H265. ATC methods reported in the literature are very diverse and typically specialized to signals with certain characteristics, such as images containing various directional properties [3, 4], and video residuals produced by intra-prediction [5–8] or motion-compensation (MC) [9–14]. Others have considered the design of more general transforms using sparsity constraints [15, 16]. However, such approaches do not take into account the effect of quantization errors due to finite rate compression.

In contrast to previous work, we present a novel algorithm for designing a fixed-size *codebook* of orthogonal transform matrices based on a fairly general block-wise stationary model of a non-stationary vector process that exhibits local stationarity properties. In this case, the best transform matrix from the codebook is adaptively chosen for each stationary block. The matrix codebook is designed to minimize the block-MSE, averaged over an ensemble of blocks representing the non-stationary process. The codebook optimization is carried out by applying the algorithm to a training

set of blocks. The only requirement to use this algorithm is the availability of a differentiable expression for the quantization MSE of the transform coefficients expressed in terms of the transform matrix. To this end, we make the general assumption that the probability distribution of the transform coefficients is solely characterized by the mean and the variance (e.g., Gaussian and Laplace distributions). While we have focused on uniform quantization which has been widely used in practice, non-uniform quantization may also be incorporated. So far, we have only considered the design of non-separable transforms.

The proposed algorithm belongs to the class of *block coordinate-descent* (BCD) algorithms [17]. In contrast to other approaches previously considered to solve related matrix-codebook design problems [18–21], in our formulation we map the harder problem of minimizing an objective $\Theta(\mathbf{T})$ subject to an orthonormality constraint on the matrix argument \mathbf{T} , from the Euclidian space to a simpler unconstrained problem on the *Stiefel manifold* [22]. This approach, which explicitly minimizes the transform coding MSE with respect to the matrix codebook, is superior to other indirect methods used to enforce orthonormality constraints, such as the Givens parameterization where a set of rotation angles rather than orthogonal matrices have to be quantized [23]. Being a BCD algorithm based on the exact matrix-derivative of the objective function, our algorithm is guaranteed to converge to a (locally) optimal transform-matrix codebook.

We have so far tested the proposed algorithm on MC residuals obtained from an H265 video codec, which contains highly non-stationary segments. Results obtained with a substantial test set of videos showed that ATC based on proposed codebook designs consistently outperformed non-ATC based on the standard DCT. A selected set of results are presented in Section 4.

2 Adaptive Transform Coding Setup

First, consider the standard transform coding problem, as applied to a stationary random vector $\mathbf{X} \in \mathbb{R}^k$. In the encoding process, rather than directly quantizing \mathbf{X} , it is first linearly transformed to obtain $\mathbf{Y} = \mathbf{T}\mathbf{X}$, and the transform coefficients $\mathbf{Y} = (Y_1, \dots, Y_k)^T$ are then scalar quantized and entropy coded, where $\mathbf{T} \in \mathbb{R}^{k \times k}$ is the orthonormal transform matrix. In this work, we assume that uniform scalar quantizers are used. Let the step-size of the quantizer for the transform coefficient Y_i be $\Delta_i, i = 1, \dots, k$ and denote the quantized version of \mathbf{Y} by $\hat{\mathbf{Y}} = (\hat{Y}_1, \dots, \hat{Y}_k)^T$. The reconstructed source vector is thus $\hat{\mathbf{X}} = \mathbf{T}^{-1}\hat{\mathbf{Y}}$. Assume that each transform coefficient is zero mean and that the quantization MSE is a function of the quantizer’s input variance $\sigma_{Y_i}^2 = E[Y_i^2]$. Accordingly, denote the quantization MSE of the coefficient Y_i by $\theta(\sigma_{Y_i}^2, \Delta_i) = E(Y_i - \hat{Y}_i)^2$. Let the variances of the transform coefficients be $(\sigma_{Y_1}^2, \dots, \sigma_{Y_k}^2)^T$. These variances are the diagonal elements of $\mathbf{C}_Y = \mathbf{T}\mathbf{C}_X\mathbf{T}^T$, where $\mathbf{C}_X = E[\mathbf{X}\mathbf{X}^T]$ is the covariance matrix of \mathbf{X} . The MSE of transform coding \mathbf{X} is given by $\Theta(\mathbf{T}, \mathbf{C}_X) = E\|\mathbf{X} - \hat{\mathbf{X}}\|^2 = \sum_{i=1}^k \theta(\sigma_{Y_i}^2, \Delta_i)$, where we make it explicit the dependence of Θ on \mathbf{C}_X through the variances $\sigma_{Y_i}^2$. We will assume that $\Delta_i, i = 1, \dots, k$ are given. In this case, the optimal \mathbf{T} can be found by solving

$$\underset{\mathbf{T}}{\text{Minimize}} \Theta(\mathbf{T}, \mathbf{C}_X) \quad (1)$$

subject to $\mathbf{T}^T \mathbf{T} = \mathbf{I}_k$ (orthonormality constraint),

where \mathbf{I}_k is the $k \times k$ identity matrix. It is easy to demonstrate that if \mathbf{X} is Gaussian, regardless of Δ , the KLT is the solution to (1) [1, 2].

Next, consider coding a sequence of non-stationary random vectors. In this case, the optimal transform may vary with the input vector, and adaptive coding can significantly improve the rate-distortion performance. We model the non-stationary process by a *block-wise* stationary vector process and encode each stationary block of vectors using a single transform optimized to local statistics. To state the problem more concretely, let $B \in \mathbb{S}_B$ be a locally stationary block of vectors in the non-stationary process, where \mathbb{S}_B is the set of all possible (ensemble of) blocks. Also, let \mathbf{C}_B be the covariance matrix for $\mathbf{X} \in B$, and \mathbb{S}_C be the ensemble of covariance matrices. Obviously, we can determine the optimal transform matrix for each \mathbf{C}_B in the ensemble by solving (1) on a per-block basis. For example, if \mathbf{X} is a Gaussian vector, then the optimal transform for stationary block B is the KLT of \mathbf{C}_B [1, 2]. Instead, we use a codebook of orthonormal transform matrices $\mathcal{T} = \{\tilde{\mathbf{T}}_1, \tilde{\mathbf{T}}_2, \dots, \tilde{\mathbf{T}}_N\}$ and the best transform matrix for a given block B is selected from the codebook as $\mathbf{T}^* = \arg \min_{\mathbf{T} \in \mathcal{T}} \frac{1}{|B|} \sum_{\mathbf{X} \in B} \|\mathbf{X} - \hat{\mathbf{X}}(\mathbf{T})\|^2$ where $\hat{\mathbf{X}}(\mathbf{T})$ is the reconstructed version of source vector \mathbf{X} using the transform \mathbf{T} and $|B|$ is the number of vectors in B .

3 Transform Matrix Codebook Optimization

Let us consider the problem of designing the optimal codebook \mathcal{T} for a given block ensemble \mathbb{S}_B . We note that the design problem essentially involves partitioning \mathbb{S}_B into N non-overlapping subsets and assigning a single transform matrix to each subset, such that the MSE of transform coding, averaged over the ensemble is minimized. As emphasized in (1), the MSE of block B is a function of the covariance matrix \mathbf{C}_B . Thus, an equivalent codebook design problem is to partition the covariance matrix ensemble \mathbb{S}_C into N non-overlapping subsets $\mathcal{G} = \{\Omega_1, \dots, \Omega_N\}$, where $\Omega_i \subset \mathbb{S}_C$, and to determine the optimal codeword $\tilde{\mathbf{T}}_i$ for each Ω_i , $i = 1, \dots, N$. Assume that \mathbf{C}_B is distributed over \mathbb{S}_C according to some probability density function. Then, the MSE of encoding all stationary blocks $B \in \mathbb{S}_B$ using the transform codebook \mathcal{T} is

$$\bar{\Theta}(\mathcal{T}, \mathcal{G}) = \sum_{i=1}^N E \left[\Theta(\tilde{\mathbf{T}}_i, \mathbf{C}_B) | \mathbf{C}_B \in \Omega_i \right] P(\Omega_i), \quad (2)$$

where the expectation is taken over the distribution of \mathbf{C}_B . We wish to find the codebook \mathcal{T}^* which minimizes (2) with respect to the sets $(\mathcal{T}, \mathcal{G})$. Specifically, we seek to solve

$$\underset{\mathcal{T}, \mathcal{G}}{\text{Minimize}} \bar{\Theta}(\mathcal{T}, \mathcal{G}) \quad (3)$$

subject to $\tilde{\mathbf{T}}_i^T \tilde{\mathbf{T}}_i = \mathbf{I}_k$, $i = 1, \dots, N$.

The direct solution of this constrained minimization problem appears intractable. The problem is reminiscent of codebook design in vector quantization, where a BCD algorithm, commonly referred to as the generalized Lloyd's algorithm, is used to solve a similar problem [1]. However, in that case there are no constraints on code vectors. In this paper, we present a BCD algorithm incorporating orthogonal constraints on the matrix codebook to solve (3), where one alternates between the solutions to two sub-problems described below, such that the algorithm converges to a solution of (3).

3.1 Sub-problem 1: Optimal partition \mathcal{G} for a fixed codebook \mathcal{T}

It is straightforward to argue that, for a fixed \mathcal{T} , the optimal \mathcal{G} that minimizes $\bar{\Theta}(\mathcal{T}, \mathcal{G})$ is given by $\mathcal{G}^* = \{\Omega_1^*, \dots, \Omega_N^*\}$, where

$$\Omega_i^* = \{\mathbf{C}_B \in \mathbb{S}_{\mathcal{C}} : \Theta(\tilde{\mathbf{T}}_i, \mathbf{C}_B) < \Theta(\tilde{\mathbf{T}}_j, \mathbf{C}_B) \forall j \neq i\} \quad (4)$$

with ties broken suitably. Note that the orthogonality constraint is irrelevant to this sub-problem.

3.2 Sub-problem 2: Optimal codebook \mathcal{T} for a fixed partition \mathcal{G}

Given the partition set \mathcal{G} , $\tilde{\mathbf{T}}_i$ only affects the i -th term of the sum in (2). Thus, the codebook $\mathcal{T}^* = \{\tilde{\mathbf{T}}_1^*, \tilde{\mathbf{T}}_2^*, \dots, \tilde{\mathbf{T}}_N^*\}$ that minimizes $\bar{\Theta}(\mathcal{T}, \mathcal{G})$ is given by

$$\tilde{\mathbf{T}}_i^* = \arg \min_{\mathbf{T} \in \mathbb{R}^{k \times k}} E[\Theta(\mathbf{T}, \mathbf{C}_B) | \mathbf{C}_B \in \Omega_i] \text{ subject to } \mathbf{T}^T \mathbf{T} = \mathbf{I}_k, \quad (5)$$

$i = 1, \dots, k$. While the solution to this constrained minimization problem does not appear to be straightforward, we note that the solution space, the set of all $k \times k$ orthogonal matrices, is the compact manifold referred to as the *orthogonal group* $O(k)$ [22]. Therefore, rather than solving (5) as a constrained minimization problem on the Euclidian space $\mathbb{R}^{k \times k}$, we can equivalently solve it as an unconstrained minimization problem on $O(k)$. Optimization on manifolds is a widely studied problem, see [22] and references therein. We have used a low-complexity, modified steepest descent algorithm on $O(k)$ [24] to solve (5). This algorithm requires that the objective function in (5) be differentiable. That is, $\Theta(\mathbf{T}, \mathbf{C}_B)$, which is the quantization MSE associated with coding a vector with a covariance matrix \mathbf{C}_B using the transform matrix \mathbf{T} , must be differentiable with respect to $\mathbf{T} \in \mathbb{R}^{k \times k}$. We identify in the following two models for $\Theta(\mathbf{T}, \mathbf{C}_B)$ satisfying this requirement, both analytically simple enough to be incorporated in to the manifold steepest-descent algorithm.

First, let $\mathbf{Y}(\mathbf{T}) = \mathbf{T}\mathbf{X}$ be the coefficient vector obtained by transforming $\mathbf{X} \in B$ using the transform matrix \mathbf{T} , and let the variances of these transform coefficients be $\sigma_{Y_j}^2(\mathbf{T})$, $j = 1, \dots, k$, which are the diagonal elements of the matrix $\mathbf{T}\mathbf{C}_B\mathbf{T}^T$.

1) *High-rate Gaussian model*: Given a target rate (quantizer output entropy) R_0 bits/vector, the minimum MSE (MMSE) of transform coding a Gaussian vector is [1, 2]

$$\Theta(\mathbf{T}, \mathbf{C}_B) = \frac{k\pi e}{6} \left(\prod_{j=1}^k \sigma_{Y_j}^2(\mathbf{T}) \right)^{1/k} 2^{-\frac{2R_0}{k}}. \quad (6)$$

Since R_0 is only a scaling factor, it can be ignored. In this case, we can show that

$$D_{\mathbf{T}}(E[\Theta(\mathbf{T}, \mathbf{C}_B)]) = E \left[\sum_{i=1}^k 2\mathbf{e}_i \mathbf{e}_i^T \mathbf{T} \mathbf{C}_B \left(\prod_{j=1, j \neq i}^k \mathbf{e}_j^T \mathbf{T} \mathbf{C}_B \mathbf{T}^T \mathbf{e}_j \right) \right],$$

where $D_{\mathbf{T}}(f(\mathbf{T})) \in \mathbb{R}^{k \times k}$ denotes the matrix of derivatives of function $f : \mathbb{R}^{k \times k} \rightarrow \mathbb{R}$ with respect to the elements of $\mathbf{T} \in \mathbb{R}^{k \times k}$, see [24, Eqn. (2)] for the definition, and \mathbf{e}_i is the i^{th} column of \mathbf{I}_k .

2) *Laplacian model*: An alternative unimodal pdf often used as a model for image transform coefficients is the Laplace pdf. We can show that, the MSE of quantizing a mean-zero Laplace variable with variance σ^2 using a uniform quantizer with a dead zone $(-\frac{z}{2}, \frac{z}{2})$ and quantization step-size Δ is

$$\theta(\sigma^2, \Delta, z) = 2b^2 - e^{-z/2b} \left(\frac{z^2 - \Delta^2}{4} + zb + \Delta b \frac{(e^{\Delta/b} + 1)}{(e^{\Delta/b} - 1)} \right),$$

where $b = \sqrt{\frac{\sigma^2}{2}}$, and we have assumed that the quantizer has infinite-support. This assumption is reasonable when variable rate coding is used. In this case, we have

$$\Theta(\mathbf{T}, \mathbf{C}_B) = \sum_{j=1}^k \theta(\sigma_{Y_j}^2, \Delta_j, z_j), \quad (7)$$

and it can be shown that

$$D_{\mathbf{T}}(E[\Theta(\mathbf{T}, \mathbf{C}_B)]) = E \left[\sum_{j=1}^k \left\{ 4b_j - \frac{e^{-z_j/2b_j} z_j}{2b_j^2} \left(\frac{z_j^2 - \Delta_j^2}{4} + z_j b_j + \Delta_j b_j \frac{(e^{\Delta_j/b_j} + 1)}{(e^{\Delta_j/b_j} - 1)} \right) \right. \right. \\ \left. \left. - e^{-z_j/2b_j} \left(z_j + \frac{\Delta_j b_j e^{2\Delta_j/b_j} - \Delta_j b_j + 2\Delta_j^2 e^{\Delta_j/b_j}}{b(e^{\Delta_j/b_j} - 1)^2} \right) \right\} \frac{\mathbf{e}_j \mathbf{e}_j^T \mathbf{T} \mathbf{C}_B}{2b_j} \right].$$

Steepest-descent minimization of (5) on $O(k)$

The constrained minimization problem (5) can be equivalently stated as an unconstrained minimization problem on $O(k) = \{\mathbf{T} \in \mathbb{R}^{k \times k} : \mathbf{T}^T \mathbf{T} = \mathbf{I}_k\}$ which is the set of all $k \times k$ real orthogonal matrices. Specifically,

$$\tilde{\mathbf{T}}_i^* = \arg \min_{\mathbf{T} \in O(k)} E[\Theta(\mathbf{T}, \mathbf{C}_B) | \mathbf{C}_B \in \Omega_i]. \quad (8)$$

The modified steepest-descent algorithm [24, Algorithm 15] can be used to minimize a differentiable function on the complex Stiefel manifold $St(k, n) = \{\mathbf{T} \in \mathbb{C}^{k \times n} : \mathbf{T}^H \mathbf{T} = \mathbf{I}_k\}$, the set of all $k \times n$ complex matrices whose columns are orthonormal vectors. This algorithm, which uses the Armijo's step-size rule, almost always converges to a local minimum. As $O(k)$ is a special case of $St(k, n)$, we adapted [24, Algorithm 15] to solve (8). Due to space limitations, we will not elaborate on specifics here, but refer the reader to [24].

3.3 Complete CD Algorithm for Transform Codebook Design

Given: A training set of covariance matrices \tilde{S} representing \mathbb{S}_C , an initial codebook of orthonormal matrices $\mathcal{T}^{(0)} = \{\tilde{\mathbf{T}}_1^{(0)}, \dots, \tilde{\mathbf{T}}_N^{(0)}\}$, a tolerance parameter $\epsilon > 0$, and maximum allowed iterations M . Set iteration index $t = 1$.

1. Given $\mathcal{T}^{(t-1)}$, partition \tilde{S} into N subsets $\{\Omega_1^{(t)}, \dots, \Omega_N^{(t)}\}$ according to (4).
2. Given $\{\Omega_1^{(t)}, \dots, \Omega_N^{(t)}\}$, find the optimal transform codebook $\mathcal{T}^{(t)}$ by solving (8).
3. Estimate (by sample averaging) $\bar{\Theta}^{(t)}$, see (2). If $\frac{\bar{\Theta}^{(t-1)} - \bar{\Theta}^{(t)}}{\bar{\Theta}^{(t-1)}} \leq \epsilon$ or $t > M$ stop; Otherwise, let $t \rightarrow t + 1$ and repeat from 1.

4 Experimental Results

4.1 A toy example

While the KLT is the optimal transform for coding Gaussian vectors, it is known that the KLT is not optimal for vectors from a Gaussian mixture [25]. Using the proposed algorithm, we optimized a single transform matrix for 2-dimensional vectors drawn from a Gaussian mixture with 3 mean-zero components whose covariance matrices are

$$\mathbf{C}_1 = \begin{bmatrix} 1.54 & -1.84 \\ -1.84 & 2.62 \end{bmatrix}, \mathbf{C}_2 = \begin{bmatrix} 0.46 & 0.40 \\ 0.40 & 0.70 \end{bmatrix}, \text{ and } \mathbf{C}_3 = \begin{bmatrix} 2.22 & 0.77 \\ 0.77 & 0.38 \end{bmatrix}.$$

Tables 1 compares the SNR and the entropy estimated from a test set of 3×10^6 vectors. The SNR is defined as $10 \log_{10} \frac{\sum \|\mathbf{x}\|^2}{\sum \|\mathbf{x} - \hat{\mathbf{x}}\|^2}$ with sums taken over the test set. The step-size Δ was chosen such that the theoretically expected rate is 0.6 bits/sample. *High-rate* and *Laplace* respectively refer to transforms optimized using the high-rate Gaussian model (6) and the Laplacian model (7). Note that the KLT has been computed from a single covariance matrix, estimated from the complete dataset. While the result itself is not surprising, our design algorithm does find a transform matrix noticeably better than the KLT.

4.2 Preliminary Results for Motion-compensated (MC) Video Residuals

This section presents experimental results obtained with MC (inter-frame prediction) residuals of video sequences. Since our transform codebook design algorithm in its current form is applicable only to non-separable transforms, we considered transform coding of 4×4 pixel blocks in our experiments to keep the computational complexity low. Thus we would encode 16-dimensional vectors using a 16×16 transform matrix. For the purpose of generating residual frames to be used in the training of transform codebook design, we used the HM test model, [26] and a set of 9 standard CIF resolution (352×288), 30 fps gray-scale video sequences (*Bus, Coastguard, Crew, Football, Foreman, Mobile, Soccer, Stefan, Tennis*), containing a diverse range of motion characteristics. To apply the block-wise stationary model, we divided each residual frame

Table 1: SNR and entropy (bits/sample) for coding a 3-component Gaussian mixture.

Transform	SNR (dB)	Entropy
KLT	3.21	0.71
DCT	3.69	0.63
High-rate	4.0	0.59
Laplace	4.0	0.59

into 16×16 non-overlapping blocks (*spatially stationary* blocks) and each such block was divided into 4×4 -pixel transform coding blocks (16-dimensional vectors). We then considered a set of time-aligned spatially stationary blocks in 8 adjacent frames to be a locally stationary block of 16-dimensional vectors, and estimated a single covariance matrix for each such *spatio-temporal* stationary block. The total training set thus consisted of 26639 local covariance matrices, which was used as the input to the transform matrix codebook design algorithm. Various codebook designs were tested using a separate set of 7 video sequences (See Table 2.)

In our experiments we used the same step-size Δ for all transform coefficient in a vector and set the dead zone to $(-\frac{\Delta}{2}, +\frac{\Delta}{2})$. In general, Δ determines the bit rate. However, when the Laplacian MSE model is used, Δ required to achieve a given bit rate has to be used in the transform codebook design algorithm, see (7). This can be achieved by designing transform codebooks for many Δ values and selecting the Δ and the transform codebook that has (approximately) the desired rate. We however found that, a transform codebook designed for an appropriately chosen value of Δ remains nearly optimal for a wide range of Δ values. We further elaborate on this point below. Since the DCT will likely be good for some stationary blocks, we included the DCT as an additional codeword, after designing a codebook. In the following, we have used the estimated entropy of the quantization indices as a proxy for the rate of variable-length coding. The total rates reported for adaptive coding include the rates of both the quantized transform coefficients and the codebook indices identifying the transform matrix for each stationary block.

Table 2, presents the gains in peak signal-to-noise ratio (PSNR) and rate achieved by optimized codebooks relative to the DCT, expressed in terms of Bjøntegaard-Delta (BD) PSNR and BD-bit-rate [27]. Note that, a positive BD-PSNR indicates a gain in PSNR compared to the DCT while a negative BD-rate indicates a rate saving. To further illustrate the advantage of using adaptive transforms over the DCT, we show in Fig. 1, the histograms of codeword usage in Football and Ice sequences.

As an example of how adaptive transform coding and non-adaptive DCT compares over a single sequence, Fig. 2 shows the absolute PSNR (not BD-PSNR) of the MC residual for the Ice sequence, where to avoid clutter in the figure, we compute and show the average PSNR for groups of 8 consecutive frames. As one would expect, the difference between the codebook optimized with the high-rate Gaussian model and the finite-rate Laplacian model diminishes as the rate increases. However, in all our experiments with MC residuals, it was observed that the Laplacian model always yielded a better codebook. This is because at low rates, the high-rate Gaussian model

Table 2: BD-PSNR and BD-Rate gains achieved by transform matrix codebooks over the standard DCT. *High-Rate* and *Laplace* refer to quantization MSE model used for codebook optimization.

Sequence	Codebook size N											
	3				6				9			
	BD-PSNR (dB)		BD-Rate (%)		BD-PSNR (dB)		BD-Rate (%)		BD-PSNR (dB)		BD-Rate (%)	
	High-Rate	Laplace	High-Rate	Laplace	High-Rate	Laplace	High-Rate	Laplace	High-Rate	Laplace	High-Rate	Laplace
Akiyo	0.096	0.148	-2.86	-5.10	0.122	0.185	-3.60	-5.39	0.135	0.192	-3.94	-5.43
City	0.096	0.233	-1.44	-3.80	0.180	0.332	-2.75	-5.14	0.180	0.343	-2.74	-5.35
Flower	-0.009	0.152	0.15	-1.44	0.048	0.221	-0.41	-2.04	0.048	0.211	-0.39	-1.88
Hall Monitor	0.137	0.250	-3.44	-6.49	0.187	0.313	-4.46	-7.46	0.185	0.323	-4.46	-7.65
Ice	0.157	0.393	-2.91	-7.31	0.180	0.467	-3.27	-8.27	0.179	0.465	-3.29	-8.28
Mother daughter	0.224	0.290	-5.83	-8.66	0.259	0.367	-6.61	-9.08	0.264	0.379	-6.74	-9.58
Waterfall	0.152	0.239	-1.86	-3.53	0.281	0.356	-4.10	-5.00	0.292	0.354	-4.30	-5.17
Average	0.122	0.244	-2.60	-5.19	0.180	0.320	-3.60	-6.05	0.183	0.324	-3.69	-6.19

can be quite inaccurate or even outright invalid [2]. Given that the codebooks are designed off-line, the slight complexity increase associated with the use of Laplacian model (compare (6) and (7)) would be of no consequence. An interesting issue however is the robustness of a codebook designed with the Laplacian model for a specific Δ : how different are codebooks designed for significantly different Δ values? In HM test model (H265 codec), the parameter QP whose value can range from 0 to 51, is used to set the bit-rate and hence the quantization step-size. Our experiments showed that, a single Laplace codebook optimized for QP=34 is nearly as good as the codebooks optimized for each QP value in the entire range.

5 Future Work

There exist other descent-type algorithms for optimization on manifolds, which can be explored, see [22]. While these algorithms may converge faster, they perform descent steps on geodesics of the manifold which can be computationally quite complex. The algorithm we have used performs descent steps on tangent spaces of the manifold which is much simpler, but results in slower convergence. It seems possible to extend our codebook design procedure to use the Newton-type algorithm for manifold optimization presented in [24], which also takes descent steps on tangent spaces, but will likely converge faster. Finally, while we have considered only non-separable transforms in this paper, an extension to separable transforms is being currently developed.

References

- [1] A. Gersho and R. M. Gray, *Vector quantization and signal compression*, Kluwer Academic Publishers, Norwell, MA, USA, 1991.
- [2] V. K. Goyal, J. Zhuang, and M. Veiterli, “Transform coding with backward adaptive updates,” *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1623–1633, Jul. 2000.
- [3] C.-L. Chang and B. Girod, “Direction-adaptive discrete wavelet transform for image compression,” *IEEE Trans. Image Process.*, vol. 16, no. 5, pp. 1289–1302, May 2007.
- [4] B. Zeng and J. Fu, “Directional discrete cosine transforms—A new framework for image coding,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 3, pp. 305–3013, Mar. 2008.

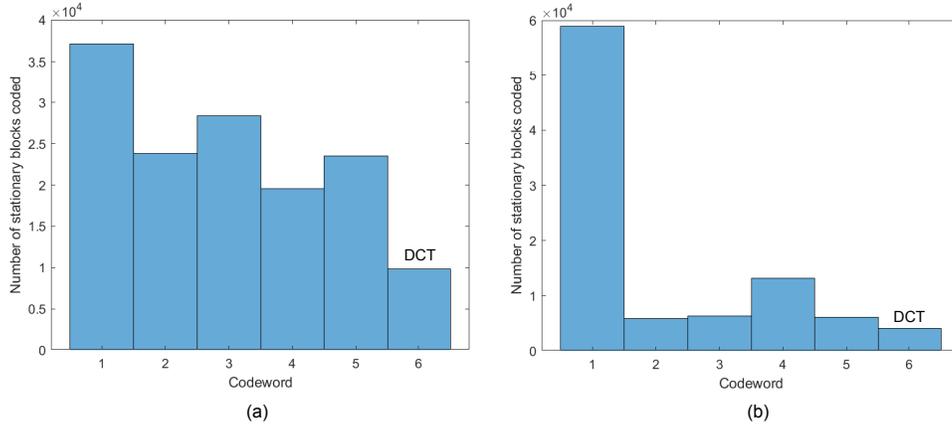


Figure 1: Histograms of transform-matrix codeword usage in (a) Football and (b) Ice sequences. The size of the codebook $N = 6$. The first 5 codewords have been optimized using the proposed algorithm, using the Laplacian MSE model (codeword 6 is the standard DCT).

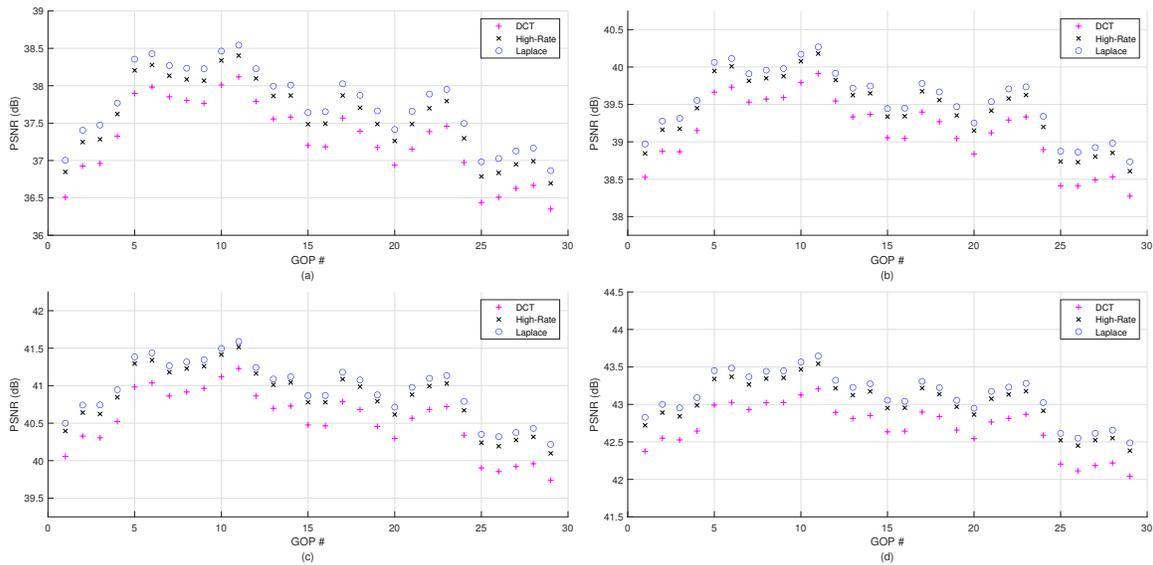


Figure 2: PSNR of coding the Ice sequence using adaptive transforms (codebook size $N = 6$) and the DCT (non-adaptive). (a) 0.45 bits/pixel, (b) 0.75 bits/pixel, (c) 1.02 bits/pixel and, (d) 1.44 bits/pixel. PSNR has been computed for groups of 8 consecutive frames.

- [5] Y. Ye and M. Karczewicz, “Improved H.264 intra coding based on bi-directional intra prediction, directional transform, and adaptive coefficient scanning,” in *IEEE DCC*, 2008, pp. 2116–2119.
- [6] J. Han, A. Saxena, V. Melkote, and K. Rose, “Jointly optimized spatial prediction and block transform for video and image coding,” *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1874 – 1884, Apr. 2012.
- [7] X. Zhao, J. Chen, M. Karczewicz, A. Said, and V. Seregin, “Joint separable and non-separable transforms for next-generation video coding,” *IEEE Trans. Image Process.*, vol. 27, no. 5, pp. 2514–2525, May 2018.
- [8] F. Zou, O. C. Au, C. Pang, J. Dai, X. Zhang, and L. Fang, “Rate-distortion optimized

- transforms based on the Lloyd-type algorithm for intra block coding,” *IEEE J. Sel. Topics Signal Process.*, vol. 7, no. 6, pp. 1072 – 1083, Dec. 2013.
- [9] T. Biatek, V. Lorcy, and P. Philippe, “Transform competition for temporal prediction in video coding,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 3, pp. 815–826, Mar. 2019.
- [10] K. N. Ngan M. Wang and L. Xu, “Efficient H.264/AVC video coding with adaptive transforms,” *IEEE Trans. Multimedia*, vol. 16, pp. 913–946, Jun. 2014.
- [11] X. Cao and Y. He, “Singular vector decomposition based adaptive transform for motion compensation residuals,” in *Proc. IEEE Int. Conf. on Image Process. (ICIP)*, Oct. 2014, pp. 4127–4131.
- [12] Z. Gu, W. Lin, B. Lee, and C. T. Lau, “Rotated orthogonal transform (ROT) for motion-compensation residual coding,” *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4770–4781, Jun 2012.
- [13] F. Kamisli and J. S. Lim, “1-D transforms for the motion compensation residual,” *IEEE Trans. Image Process.*, vol. 20, no. 4, pp. 1036–1046, Oct. 2010.
- [14] C. Lan, J. Xu, G. Shi, and F. Wu, “Exploiting non-local correlation via signal-dependent transform (SDT),” *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 7, pp. 1298–1308, Nov. 2011.
- [15] O. G. Sezer, O. G. Guleryuz, and Y. Altunbasak, “Approximation and compression with sparse orthonormal transforms,” *IEEE Trans. Image Process.*, vol. 24, no. 8, pp. 2328–2343, Aug. 2015.
- [16] J.-W. Kang, M. Gabbouj, and C.-C. J. Kuo, “Sparse/DCT (S/DCT) two-layered representation of prediction residuals for video coding,” *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2711–2722, Jul. 2013.
- [17] D. P. Bertsekas, *Nonlinear Programming*, Athena Scientific, Belmont, MA, 2nd edition, 1999.
- [18] R. D. Dony and S. Haykin, “Optimally adaptive transform coding,” *IEEE Trans. Image Process.*, vol. 4, no. 10, pp. 1358–1370, Oct. 1995.
- [19] H. Caglar, S. Güntürk, B. Sankur, and E. Anarim, “VQ-adaptive block transform coding of images,” *IEEE Trans. Image Process.*, vol. 7, no. 1, pp. 110–115, Jan. 1998.
- [20] M. Effros, P. A. Chou, and R. M Gray, “Weighted universal image compression,” *IEEE Trans. Image Process.*, vol. 8, no. 10, pp. 1317–1329, Oct. 1999.
- [21] C. Archer and T. K. Leen, “A generalized Lloyd-type algorithm for adaptive transform coder design,” *IEEE Trans. Signal Process.*, vol. 52, no. 1, pp. 255–264, Jan. 2004.
- [22] J. H. Manton, “Geometry, manifolds, and nonconvex optimization,” *IEEE Signal Process. Mag.*, vol. 37, no. 5, pp. 1605–1619, Sep. 2020.
- [23] M. A. Sadrabadi, A. K. Khandani, and F. Lahouti, “A new method of channel feedback quantization for high data rate MIMO systems,” in *IEEE GLOBECOM*, 2004, pp. 91–95.
- [24] J. H. Manton, “Optimization algorithms exploiting unitary constraints,” *IEEE Trans. Signal Process.*, vol. 50, no. 3, pp. 635–650, Aug. 2002.
- [25] M. Effros, H. Feng, and K. Zeger, “Suboptimality of the Karhunen-Loève transform for transform coding,” *IEEE Trans. Inf. Theory*, vol. 50, no. 8, pp. 1605–1619, Aug. 2004.
- [26] Joint Collaborative Team on Video Coding (JCT-VC), “HEVC HM reference software,” <https://vcgit.hhi.fraunhofer.de/jvet/HM>.
- [27] G. Bjøntegaard, *Improvements of the BD-PSNR Model, document VCEG-AI11, ITU-T SG16/Q6*, 35th VCEG Meeting, Berlin, Germany, Jul. 2008.